Spatial statistics of Health Data in ArcGIS Pro – Clustering of values to search for spatial patterns

In this tutorial you are going to become familiar with clustering techniques to understand spatial patterns of infectious disease.

**OBJECTIVES**

* **Utilize exploratory data analysis techniques to search, characterize, and describe the spatial distribution of infectious disease incidence data**
* **Utilize widely used methods to characterize spatial associations at global and local levels (Getis-Ord G Statistics, Global Moran’s I, Getis-Ord G\*, Anselin Local Moran’s I)**

Requirements:

1. Data from the Practice Exercises 4 folder. Sources:
   1. Lyme disease incidence data was derived from the following websites:
      1. Lyme disease case data between 2000-2017 (<https://www.cdc.gov/lyme/stats/survfaq.html>)
      2. Population estimates (<https://www.census.gov/programs-surveys/popest/data/data-sets.All.html>)
         1. Population Estimates of Resident Population for Counties and States: April 1, 2000 to July 1, 2010
         2. County Population Totals and Components of Change

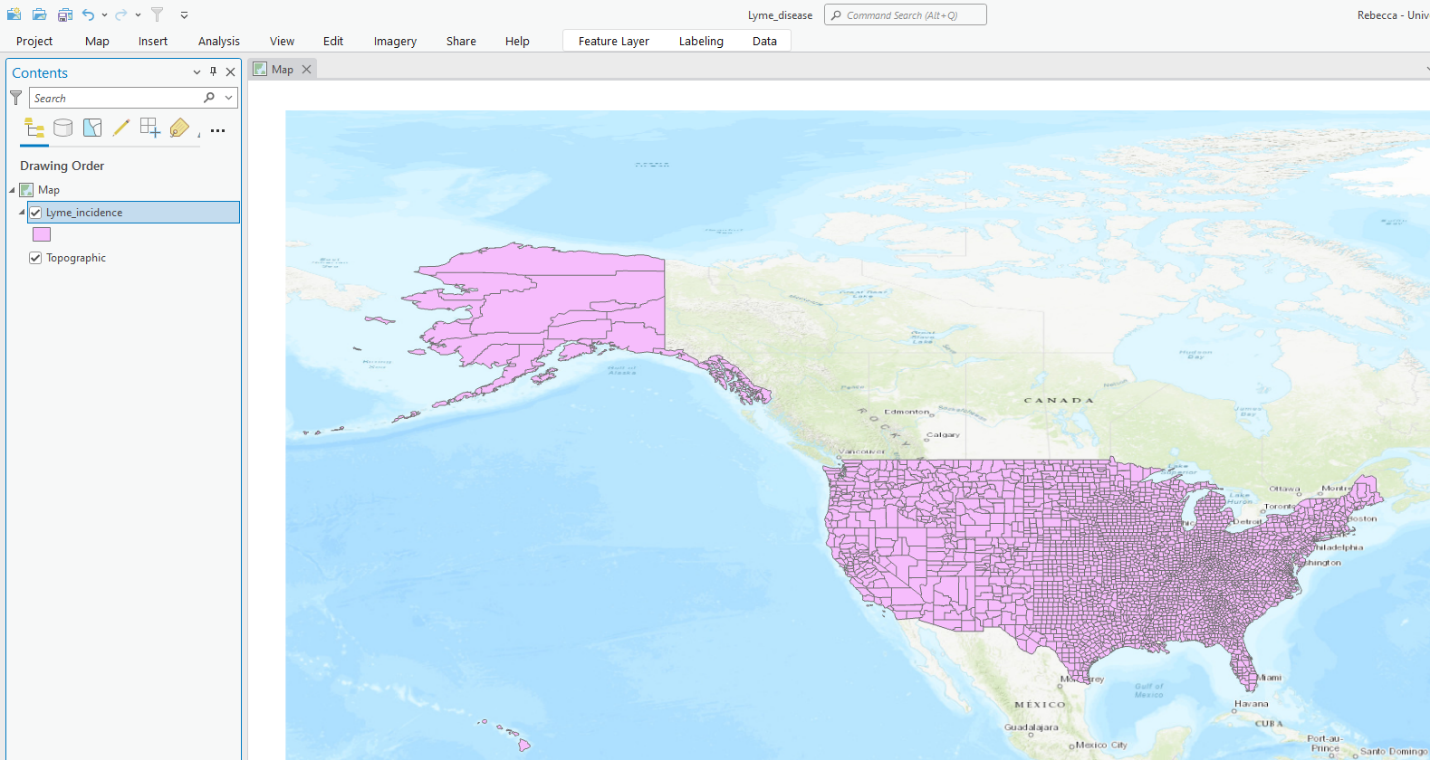
Section 1: Global Level

Understanding how disease spreads is an important topic for health geospatial researchers. Tobler’s first law of geography is “everything is related to everything else, but near things are more related than distant things”. This lab begins searching for spatial autocorrelation of data. Strong spatial autocorrelation means that adjacent geographic objects are strongly related to one another (positively or negatively).

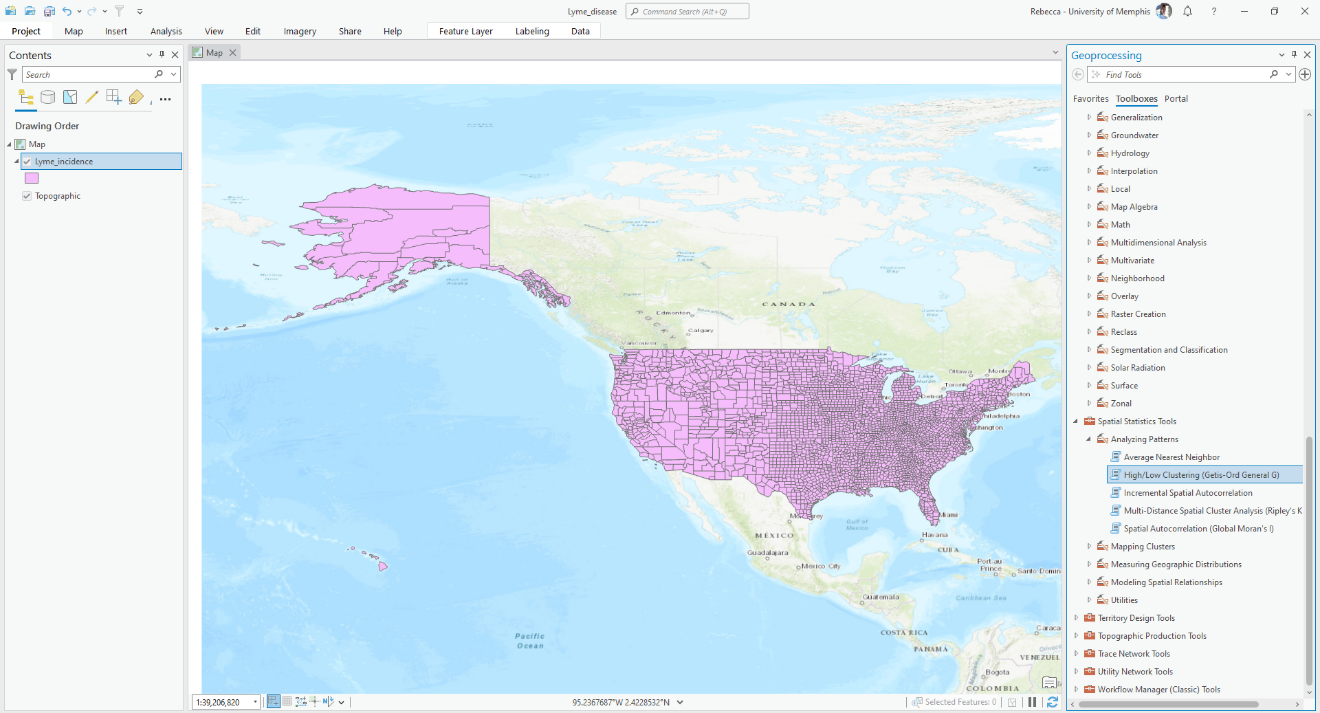
## Section 1.1: Calculate spatial autocorrelation using Getis-Ord General G

For the first part of this tutorial, you will calculate the Getis-Ord General G to determine if the data is clustered and the cluster’s pattern type.

1. Open the **Lyme\_disease.aprx** map file by navigating to the Practice Exercises 5 folder and double-clicking the file.



1. Next, you will have to set the data source for each layer in the map file as you did in Exercise 2. A short example of steps is included below. Do this for each layer.
   1. Right-click the **Lyme\_incidence** layer->Choose Properties->Select Data->Click Set Data Source
   2. Navigate to the **Lyme\_disease.gdb** geodatabase in the data folder for this exercise
   3. Choose the **Lyme\_incidence** layer
   4. Click **OK**
2. Next, you will want to click **Analysis** in the top panel of the ArcGIS Pro environment. Click **Tools** to open the Geoprocessing toolbox on the left side of your ArcGIS Pro environment. Navigate to **Spatial Statistics Tools ->Analyzing Patterns->High/Low Clustering (Getis-Ord General G).**



1. Set the input feature class to **Lyme\_incidence** and the input field to **AverageInc** (Average incidence rates for the years 2000-2017). Confirm that the conceptualization of distance is set to **Contiguity edges only**. Check the box next to **Generate Report**. This will add graphical outputs to your results window HTML format. Leave all other fields blank and click **Run**.

A screenshot of a computer

Description automatically generated

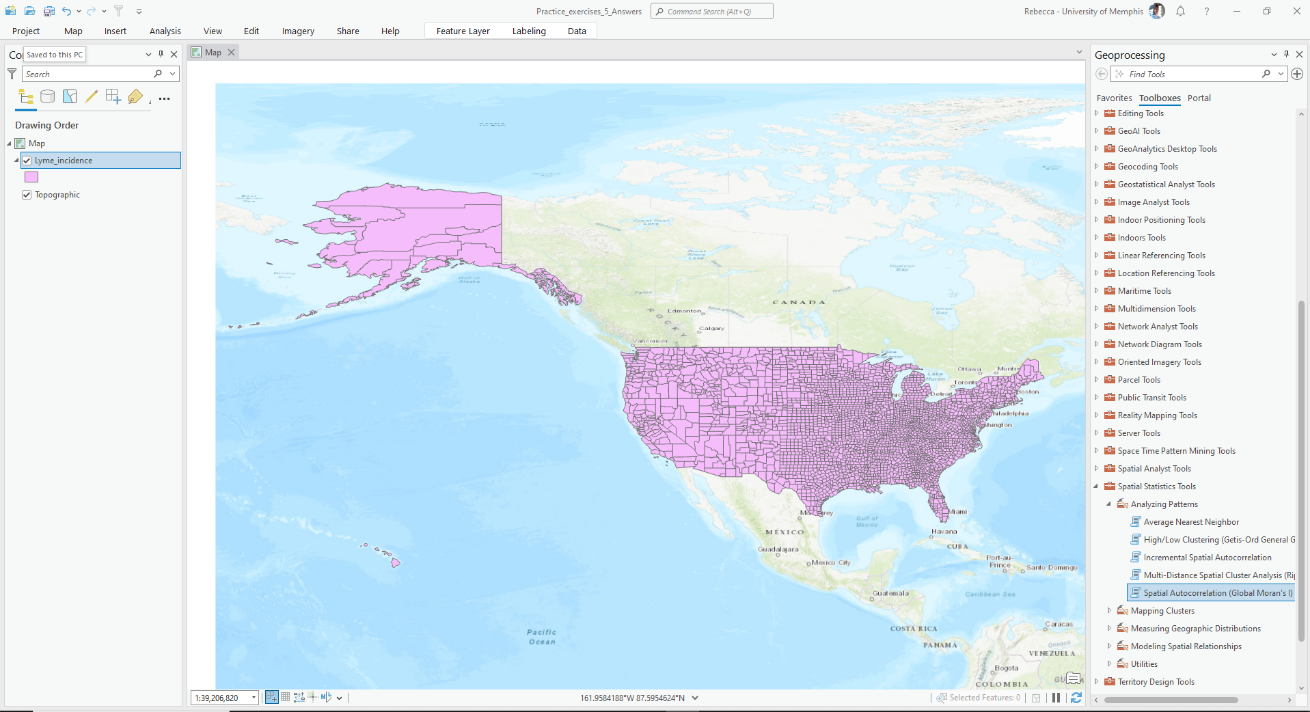
1. Repeat step 4 with input as **Inciden\_15** (Lyme disease incidence in 2015), **Inciden\_16** (Lyme disease incidence in 2016), and **Inciden\_17** (Lyme disease incidence in 2017).

**Section 1.1 Task:** Create a table that includes the values for the Observed General G, Expected General G, Variance, Z-score, p-value, and Pattern Type for each input (Average, 2015, 2016, and 2017).

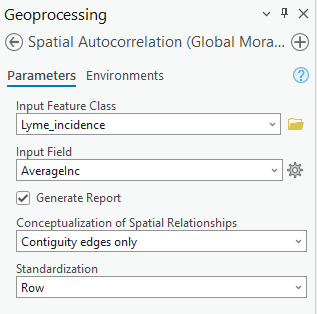
## Section 1.2: Calculate spatial autocorrelation using Moran’s I

For the first part of this tutorial, you will calculate Moran’s I to determine if the data is clustered and the cluster’s pattern type.

1. In the Geoprocessing toolbox on the left side of your ArcGIS Pro environment, navigate to **Spatial Statistics Tools ->Analyzing Patterns->Spatial Autocorrelation (Global Moran’s I).**



1. Set the input feature class to **Lyme\_incidence** and the input field to **AverageInc** (Average incidence rates for the years 2000-2017). Confirm that the conceptualization of distance is set to **Contiguity edges only**. Check the box next to **Generate Report**. This will add graphical outputs to your results window HTML format. Leave all other fields blank and click **Run**.



1. Repeat step 2 with input as **Inciden\_15** (Lyme disease incidence in 2015), **Inciden\_16** (Lyme disease incidence in 2016), and **Inciden\_17** (Lyme disease incidence in 2017).

**Section 1.2 Task 1:** Create a table that includes the values for the Moran’s I (Index), Expected I (Index), Variance, Z-score, p-value, and Pattern Type for each input (Average, 2015, 2016, and 2017).

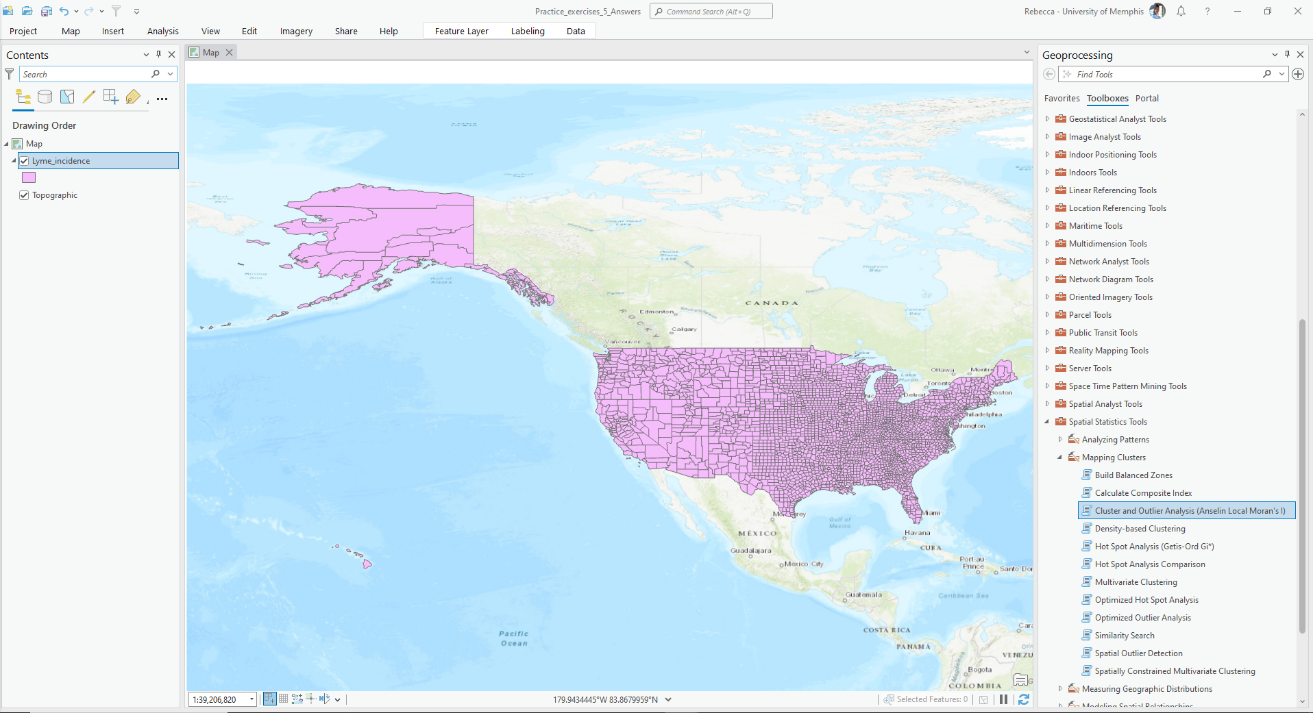
**Section 1.2 Task 2:** Compare and describe the Getis-Ord General G and Moran’s I results.

# Section 2: Local Level

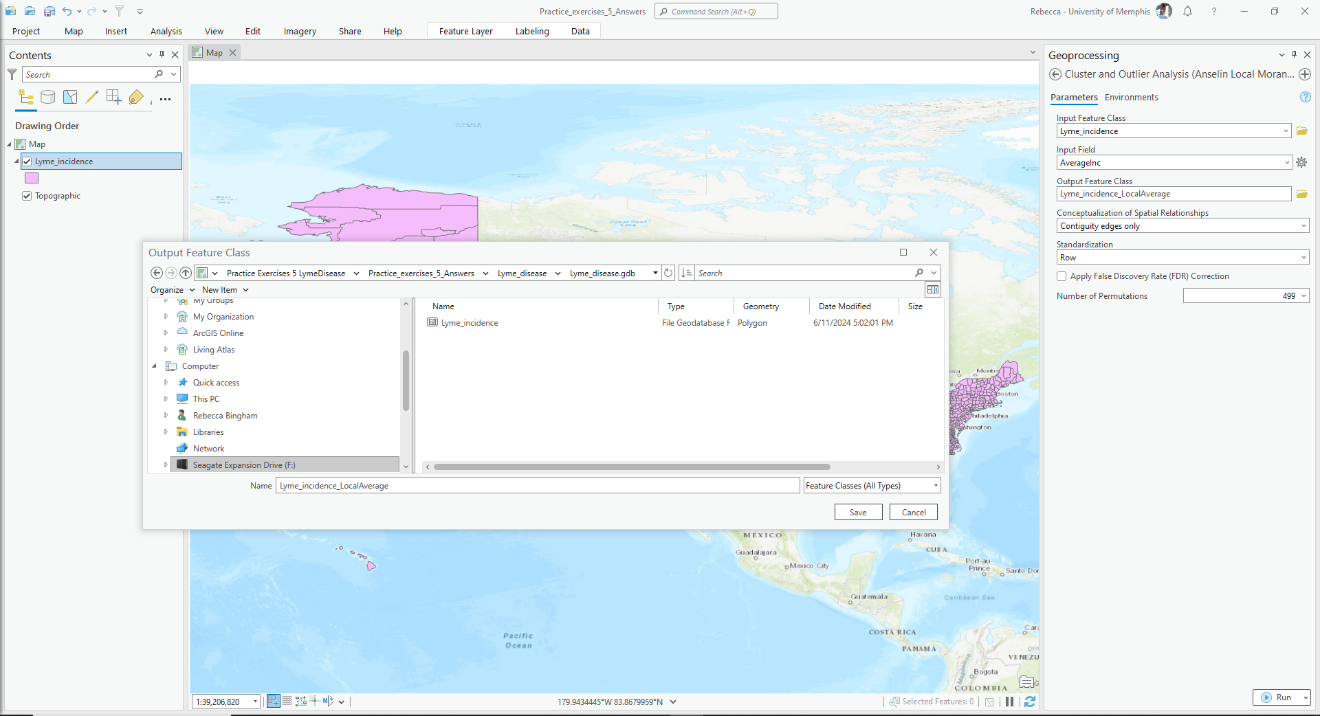
## Section 2.1: Calculate spatial autocorrelation using Anselin Local Moran’s I

For the first part of this section, you will calculate Anselin Local Moran’s I to determine if the data is clustered and the cluster’s pattern type.

1. In the Geoprocessing toolbox on the left side of your ArcGIS Pro environment, navigate to **Spatial Statistics Tools ->Mapping Clusters->Cluster and Outlier Analysis (Anselin Local Moran’s I).**



1. Set the input feature class to **Lyme\_incidence** and the input field to **AverageInc** (Average incidence rates for the years 2000-2017). Click the folder next to the Output Feature Class field input box. Navigate to the **Lyme\_disease.gdb** and name the output feature class **Lyme\_incidence\_LocalAverage.** Confirm that the conceptualization of distance is set to **Contiguity edges only**. Leave all other fields as is and click **Run**.



**Section 2.1 Task 1:** Provide a map of the clustering results for the average incidence of Lyme disease. Describe the pattern of spatial clustering including which areas exhibit clustering and which do not.

1. Repeat step 2 with input as **Inciden\_15** (name it **Lyme\_incidence\_Local2015**).

**Section 2.1 Task 2:** Provide a map of the clustering results for the 2015 incidence of Lyme disease. Describe any differences in the pattern of spatial clustering between this year and the average incidence.

1. Repeat step 2 with input as **Inciden\_16** (name it **Lyme\_incidence\_Local2016**).

**Section 2.1 Task 3:** Provide a map of the clustering results for the 2016 incidence of Lyme disease. Describe any differences in the pattern of spatial clustering between this year and the average incidence as well as any differences between this year and 2015.

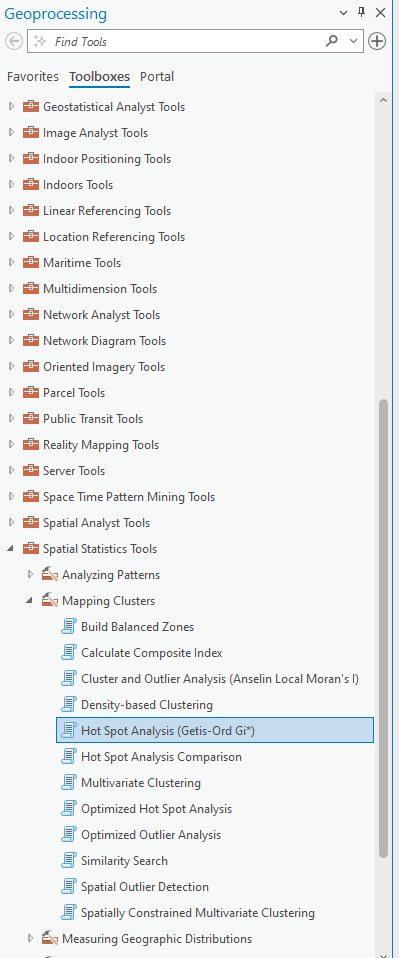
1. Repeat step 2 with input as **Inciden\_17** (name it **Lyme\_incidence\_Local2017**).

**Section 2.1 Task 4:** Provide a map of the clustering results for the 2017 incidence of Lyme disease. Describe any differences in the pattern of spatial clustering between this year and the average incidence as well as any differences between this year and the two years before.

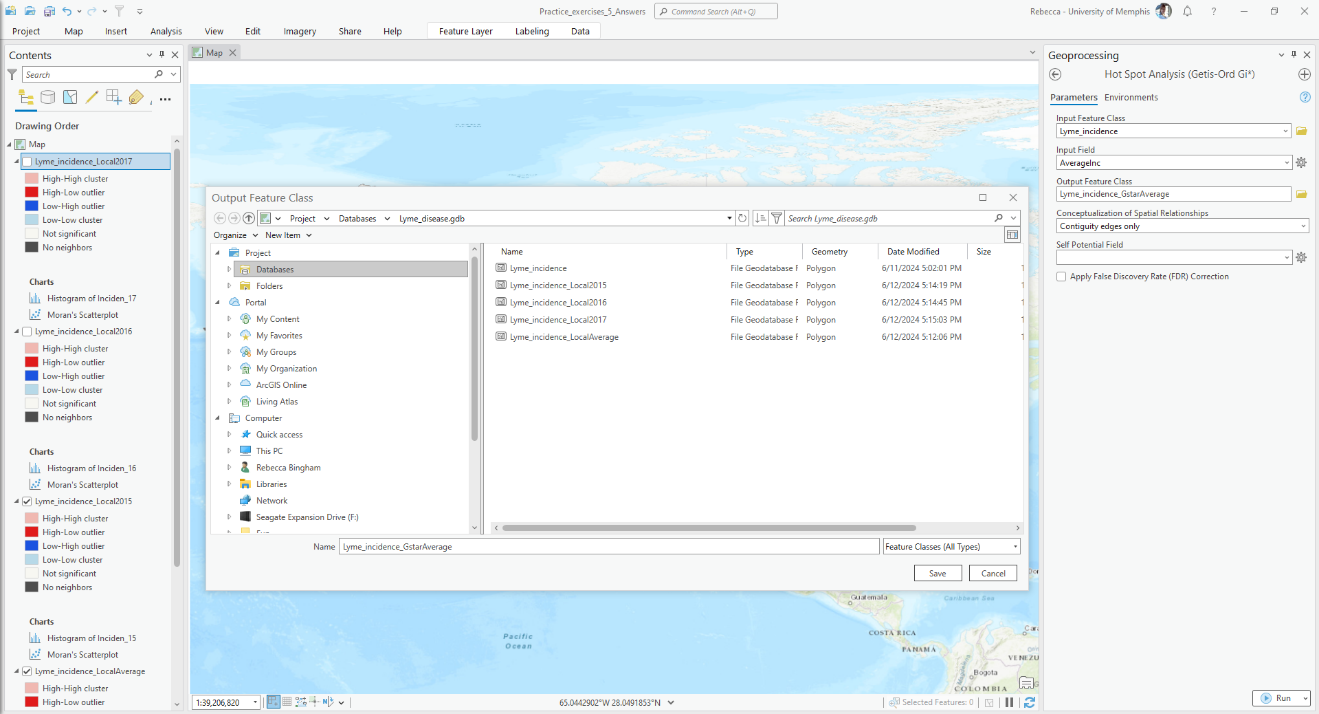
## Section 2.2: Calculate spatial autocorrelation using Getis-Ord Gi\*

For the first part of this section, you will calculate Getis-Ord Gi to determine if the data is clustered and the cluster’s pattern type.

1. In the Geoprocessing toolbox on the left side of your ArcGIS Pro environment, navigate to **Spatial Statistics Tools ->Mapping Clusters->Hot Spot Analysis (Getis-Ord Gi\*).**



1. Set the input feature class to **Lyme\_incidence** and the input field to **AverageInc** (Average incidence rates for the years 2000-2017). Click the folder next to the Output Feature Class field input box. Navigate to the **Lyme\_disease.gdb** (you should already be there) and name the output feature class **Lyme\_incidence\_GstarAverage.** Confirm that the conceptualization of distance is set to **Contiguity edges only**. Leave all other fields blank and click **Run**.



**Section 2.2 Task 1:** Provide a map of the GiZscore values for the average incidence of Lyme disease. Describe the pattern of spatial clustering including which areas exhibit clustering and which do not. (Note: you will have to go to the symbology and change the Field to **GiZScore Contiguity**).

1. Repeat step 2 with input as **Inciden\_15** (name it **Lyme\_incidence\_Gstar2015**).

**Section 2.2 Task 2:** Provide a map of the GiZscore values for the 2015 incidence of Lyme disease. Describe any differences in the pattern of spatial clustering between this year and the average incidence. (Note: you will have to go to the symbology and change the Field to **GiZScore Contiguity**).

1. Repeat step 2 with input as **Inciden\_16** (name it **Lyme\_incidence\_Gstar2016**).

**Section 2.2 Task 3:** Provide a map of the GiZscore values for the 2016 incidence of Lyme disease. Describe any differences in the pattern of spatial clustering between this year and the average incidence as well as any differences between this year and 2015. (Note: you will have to go to the symbology and change the Field to **GiZScore Contiguity**).

1. Repeat step 2 with input as **Inciden\_17** (name it **Lyme\_incidence\_Gstar2017**).

**Section 2.2 Task 4:** Provide a map of the GiZscore values for the 2017 incidence of Lyme disease. Describe any differences in the pattern of spatial clustering between this year and the average incidence as well as any differences between this year and the two years before. (Note: you will have to go to the symbology and change the Field to **GiZScore Contiguity**).

**Section 2.2 Task 5:** Compare your results obtained from Anselin Local Moran’s I and Getis-Ord Gi\*.