Spatial Data Science MOOC

Exercise

Detect patterns

Section 4 Exercise 1

03/2020

esri

THE SCIENCE OF WHERE™

Detect patterns

Time to complete
45 minutes

Introduction

Statistical cluster analysis can help you minimize the subjectivity in your maps by identifying meaningful clusters in your data. The Hot Spot Analysis and Outlier Analysis tools use statistics to detect spatial patterns in your data, but each provide slightly different information about these patterns.

Hot Spot Analysis uses the Getis-Ord Gi* statistic to identify statistically significant spatial clusters of high values (hot spots) and low values (cold spots).

Outlier Analysis uses the Anselin Local Moran's I statistic to identify statistically significant clusters of high and low values and to detect spatial outliers, or features with values significantly dissimilar from their neighbors.

ArcGIS provides traditional and optimized statistical cluster analysis tools. The optimized statistical cluster analysis tools interrogate your data to provide smart default values, optimizing the analysis workflow. The traditional statistical cluster analysis tools allow you more flexibility in defining the spatial relationships in your data, providing you with more control of your analysis. In this exercise, you will use the optimized statistical cluster analysis tools to explore the spatial patterns in the data.

Exercise scenario

Supplemental Nutrition Assistance Program (SNAP) is a federal program that helps families buy nutritional food to maintain their health and well-being. In this exercise, you will complete a Hot Spot Analysis and Outlier Analysis to find meaningful patterns of high and low SNAP participation. This information can help decision makers distribute resources in a more efficient and equitable way, ensuring that healthy food is accessible to all SNAP recipients.

Step 1: Download the exercise data files

In this step, you will download the exercise data files.

a. Open a new web browser tab or window.

b. Go to https://bit.ly/35zFGfG and download the project package file.

Note: The complete URL to the exercise data file is https://www.arcgis.com/home/item.html?id=f2cedb26fced4bc485b3631d95b0e769.

Step 2: Open an ArcGIS Pro project

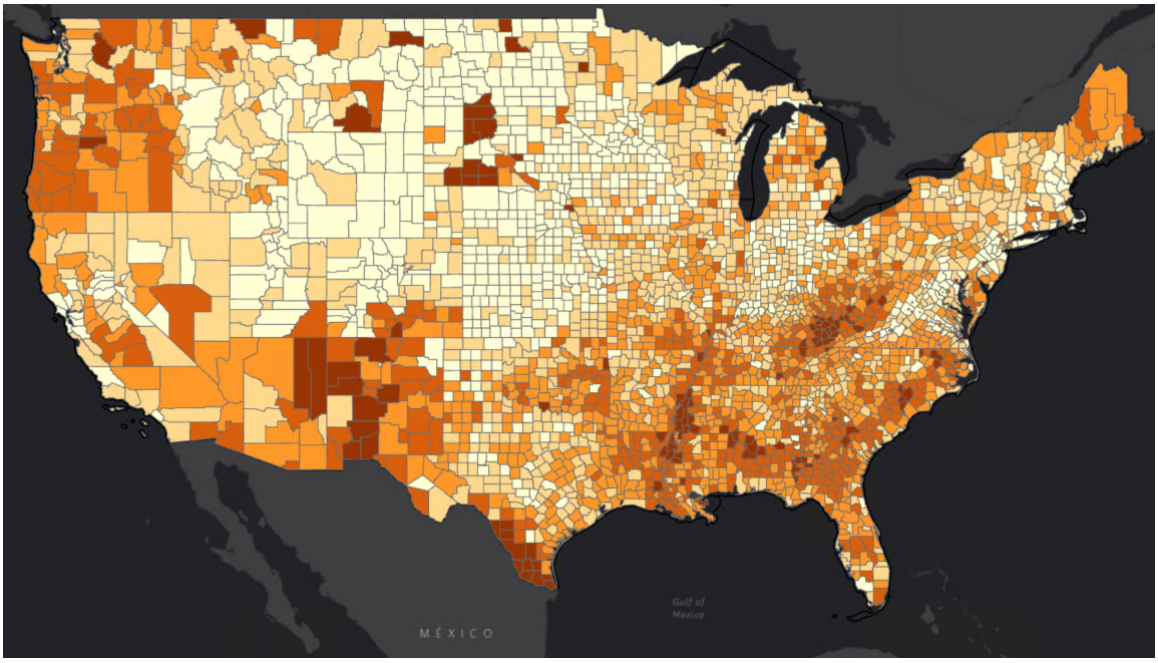
a. Start ArcGIS Pro.

b. If necessary, sign in using the provided course ArcGIS account.

c. Click Open Another Project.

d. Browse to the PatternDetection_SpaceTime.ppkx project that you saved on your computer.

e. Click OK.



Your ArcGIS Pro project includes a map of the counties in the contiguous United States. Each county is symbolized by the rate of the population that participated in SNAP during 2016.

Step 3: Run a hot spot analysis

The statistical cluster analysis tools analyze your data to detect patterns of high and low values. In this analysis, you will analyze the distribution of a value called SNAPRate, which is the rate of SNAP participation in each county.

a. From the Analysis tab, in the Geoprocessing group, click Tools.

b. In the Geoprocessing pane, under the search field, click Toolboxes.

c. Expand Spatial Statistics Tools, and then expand Mapping Clusters.

d. Click Optimized Hot Spot Analysis.

The Optimized Hot Spot Analysis tool opens in the Geoprocessing pane.

e. In the Geoprocessing pane, enter the following parameters:

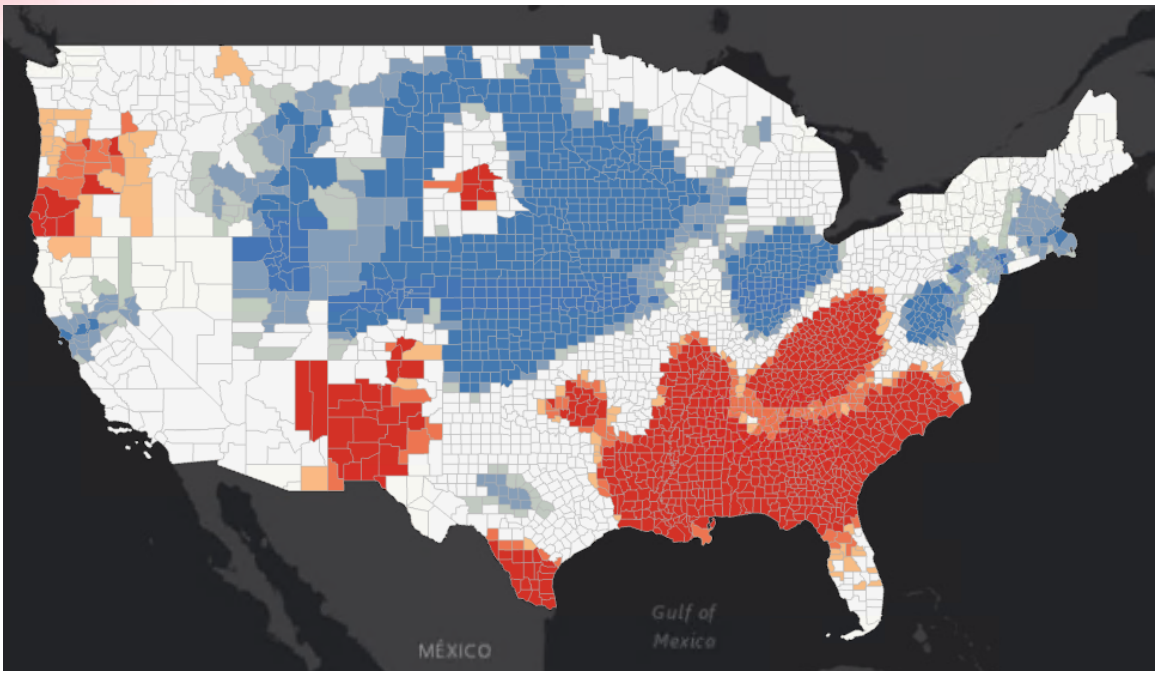
- Input Features: US_Counties
- Output Features: **SNAPHotSpots**
- Analysis Field: SNAPRate

f. Expand Override Settings.

The Optimized Hot Spot Analysis tool uses a fixed distance band to define each feature's neighborhood. If you do not specify a distance, the tool uses Incremental Spatial Autocorrelation to determine if there is a scale, or distance, at which clustering across the dataset is most pronounced. If the Optimized Hot Spot Analysis tool cannot identify a distance using this method, it will compute the average distance that would yield 30 neighbors for each feature. To learn more about Incremental Spatial Autocorrelation, see ArcGIS Pro Help: Incremental Spatial Autocorrelation (Spatial Statistics).

Fixed distance band is one method available to define feature neighborhoods. Other methods are available with the traditional Hot Spot Analysis tool, Hot Spot Analysis (Getis-Ord Gi*). To learn more about the Hot Spot Analysis (Getis-Ord Gi*) tool, see ArcGIS Pro Help: Hot Spot Analysis (Getis-Ord Gi*).

g. Click Run.



The result of your analysis is a layer displaying hot spots in three shades of red and cold spots in three shades of blue. The varying shades correspond to three confidence intervals, indicating how confident you can be that these patterns are meaningful and not the result of random chance. You will review the analysis details to ensure that the parameters were appropriate for your question.

Step 4: Review analysis parameters

a At the bottom of the Geoprocessing pane, click View Details.

The Optimized Hot Spot Analysis tool message window appears, listing the tool's geoprocessing steps in detail.

If you choose the default values for the Optimized Hot Spot Analysis tool, review the geoprocessing details to identify the default parameter values. Ensure that these values are appropriate for the scale of your analysis.

b Review the geoprocessing details to answer the following question.

1. **What distance band was chosen for this analysis?**

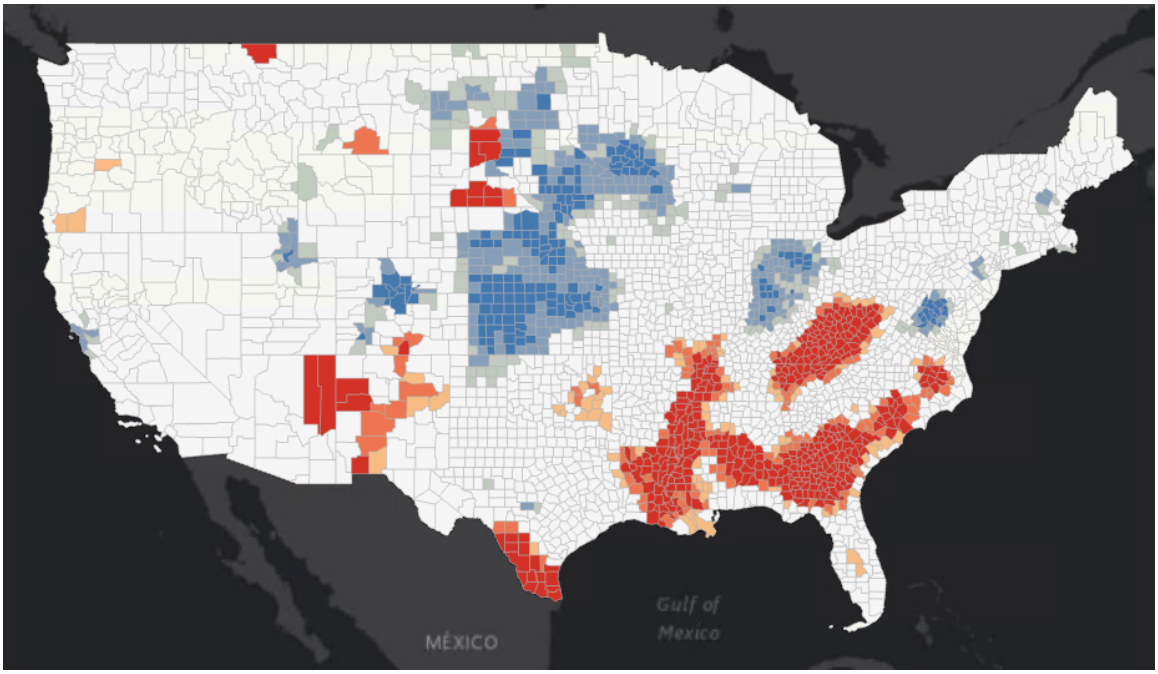
The tool chose a default distance band of approximately 200 kilometers (km) based on the average distance to 30 nearest neighbors. This default is a good place to start exploring your data, but it may not represent the scale at which you want to analyze patterns in your dataset. In this example, a 200 km distance band is too large because you want to analyze more local patterns in SNAP participation. You will reduce the distance band to 100 km to detect more local patterns in this county-level dataset.

c. Close the Optimized Hot Spot Analysis tool message window.

d. In the Geoprocessing pane, expand Override Settings, if necessary.

e. Under Distance Band, type **100** , and then next to Feet, click the down arrow and choose Kilometers.

f. Click Run.



Reducing the size of the distance band identified more detailed patterns. This scale is more appropriate for this particular analysis.

Step 5: Interpret the results

a. In the Contents pane, locate the SNAPHotSpots and US_States layers.

b. Drag the US_States layer above the SNAPHotSpots layer.

c If necessary, expand the SNAPHotSpots layer to see the legend.

2. **What statistically significant spatial patterns can you detect from this analysis?**

_____

_____

_____

The results of this statistical analysis provide a measure of confidence that can help you identify areas with clusters of high SNAP participation. You can use this information to investigate these areas and their access to stores that accept SNAP and carry healthy foods.

d Save the project.

Step 6: Run an outlier analysis

Completing an Outlier Analysis will help you identify features with values that are statistically significantly different than their neighborhood. This analysis will provide additional insight into the spatial patterns of your data.

a If necessary, open the Geoprocessing pane.

Hint: From the Analysis tab, in the Geoprocessing group, click Tools.

b In the Geoprocessing pane, search for **outlier**.

c Open the Optimized Outlier Analysis tool.

Similar to the Optimized Hot Spot Analysis tool, the Optimized Outlier Analysis tool will interrogate your data to determine an appropriate neighborhood distance for your analysis. The traditional Outlier Analysis tool, Cluster and Outlier Analysis (Anselin Local Moran's I), gives you more control over the analysis parameters. To learn more about the Cluster and Outlier Analysis (Anselin Local Moran's I) tool, see ArcGIS Pro Help: Cluster and Outlier Analysis (Anselin Local Moran's I).

d In the Geoprocessing pane, enter the following parameters:

- Input Features: US_Counties
- Output Features: **SNAPOutliers**
- Analysis Field: SNAPRate

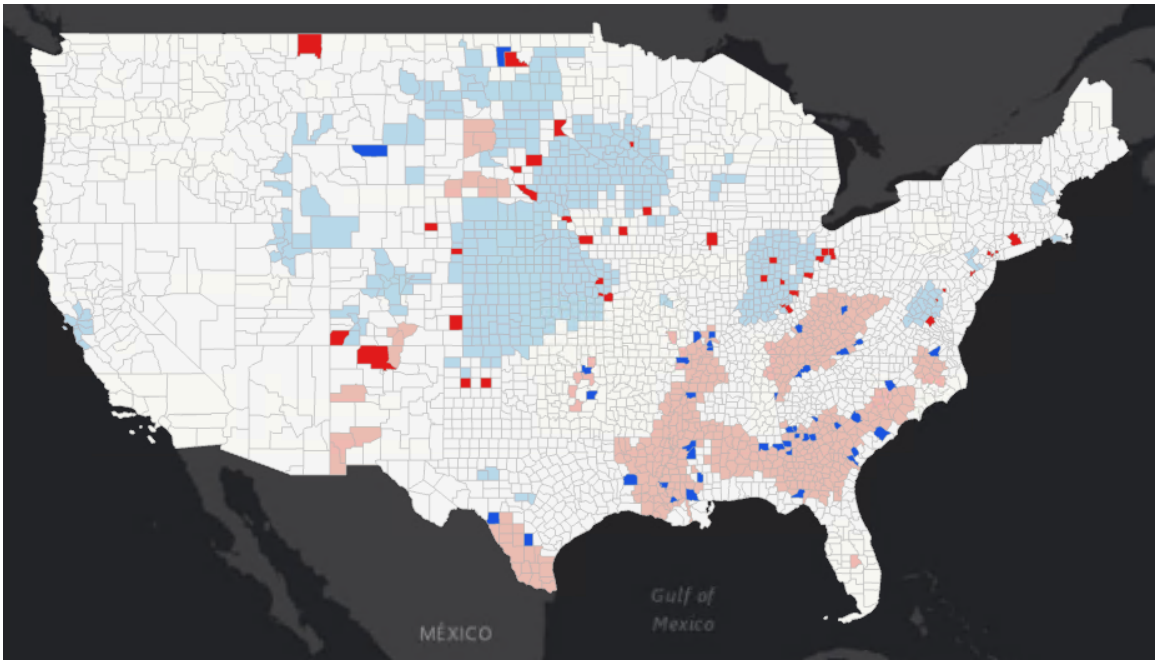
The Performance Adjustment field defines the number of permutations to create a random distribution. The tool will then compare your data's spatial distribution with the randomly generated values. To balance precision and processing time, you will leave the default. For more information regarding permutations, see ArcGIS Pro Help: How Cluster and Outlier Analysis (Anselin Local Moran's I) works.

e. Expand Override Settings.

When you compare the results of a Hot Spot Analysis and Outlier Analysis, use the same distance band in the analysis.

f. Update the Distance Band to **100** Kilometers.

g. Click Run.



Note: The permutations in the Optimized Outlier Analysis tool compare your data values to a set of randomly generated values. This means that your results may vary slightly from the preceding graphic.

The bright red and blue features represent spatial outliers. Features with high values surrounded by areas with low values are called High-Low outliers and are displayed in red. Features with low values surrounded by areas with high values are called Low-High outliers and are displayed in blue. The pink and light blue colors indicate clusters of features with

statistically significantly high values (pink) and statistically significantly low values (light blue). These clusters typically align with the hot and cold spots from the Optimized Hot Spot Analysis tool.

Step 7: Interpret the results

You have completed two different methods of pattern detection—Hot Spot Analysis and Outlier Analysis. You may be wondering, "Which method is more appropriate to use?" The answer is often both. Each method answers a different question and is valuable in understanding the patterns in your data. You can compare these analyses to gain additional insight into the spatial patterns of your data.

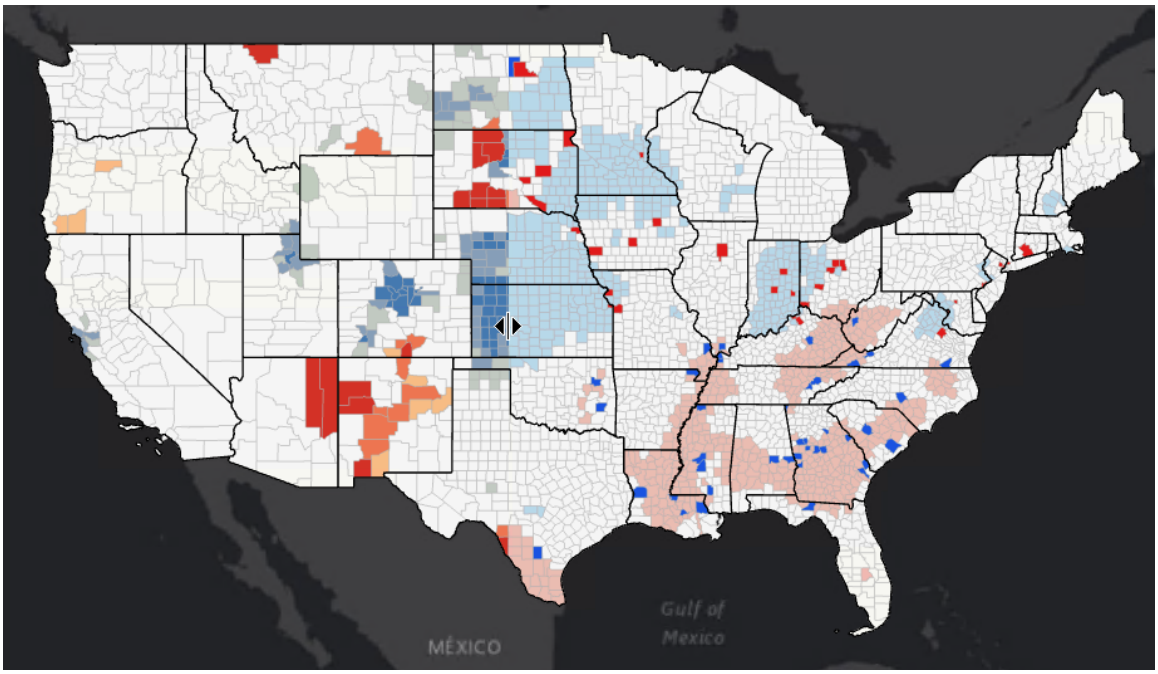
a. In the Contents pane, locate the SNAPOutliers and US_States layers.

b. Drag the US_States layer above the SNAPOutliers layer.

c. From the Appearance tab, in the Effects group, click Swipe.

d. In the Contents pane, select the SNAPOutliers layer.

e. In the map, click and drag to the left, to the right, or up and down to compare the Hot Spot Analysis and Outlier Analysis results.



Using Hot Spot Analysis and Outlier Analysis, you located statistically significant clusters of high SNAP participation. This information can help in the allocation of SNAP resources to areas of higher food insecurities. The results can help drive the decision to distribute resources in a more efficient and equitable way.

f. At the top of the map view, next to Pattern Detection, click the X to close the map.

g. Save the project and exit ArcGIS Pro.

Stretch goal (Optional)

In this exercise, you completed a pattern detection using the geoprocessing tools available in ArcGIS Pro. This stretch goal provides the opportunity for you to try completing this pattern detection using the ArcGIS API for Python in an ArcGIS Notebook. Specifically, you will use the beta version of ArcGIS Notebook on ArcGIS Online.

ArcGIS Notebook on ArcGIS Online allows you to access ArcGIS geospatial analysis libraries through the ArcGIS API for Python and ArcPy, web GIS analysis tools, and hundreds of open source libraries. This is a beta version of the product that is still undergoing refinements. The beta is not publicly available, which means that you will only be able to access this product using the course ArcGIS account provided to you in this MOOC (user name ending with _sds).

The tasks for this stretch goal have been grouped to more clearly define the workflow that you are completing. If you need assistance, refer to the ArcGIS API for Python: Guide.

A sample notebook, titled Detect Patterns MOOC, is also available using the following link: https://www.arcgis.com/home/item.html?id=e57b99937032409ebb2778484676147a. You may need to sign in to your course ArcGIS account to open the sample notebook online.

Create a notebook in ArcGIS Online

1. Open a new private or incognito web browser window (https://bit.ly/2wAuJxO).
2. Go to www.arcgis.com and sign in using your course ArcGIS account user name (ending with _sds) and password.
3. From the ArcGIS Online organization, click Notebook.

Add map layers to the notebook

1. In the Notebook, from the Top Ribbon, click the Add button.
2. Search ArcGIS Online for the following layers:
    - US Counties MOOC owned by EsriTrainingSvc
    - US States MOOC owned by EsriTrainingSvc
3. Add the layers to the notebook, assigning each to a new variable.
4. Run the code cells.

Create a map in the notebook

1. In a new code cell, assign the `gis.map` map widget to a variable.
2. Use map widget properties to change the basemap (`.basemap`) and add the layers to the map (`.add_layer`).
3. Run the code cell.

Run a Hot Spot Analysis

1. From the Top Ribbon, click Analysis.
2. From the Analysis Tools pane, add Find Hot Spots to the notebook.
3. In the notebook, at the end of the tool syntax, replace the parenthesis with a question mark.
4. Run the code cell.
5. Replace the tool syntax with the syntax that is listed under the Signature.
6. Replace `analysis_layer` parameter with the variable that you used to define the US Counties MOOC layer.
7. Update the remaining parameters with the following values:
   - `Analysis_field=`**"SNAPRate"**
   - `Output_name=`**"SNAPHotSpots_<your initial's and today's date>"**
   - `Distance_band=`**100**
   - `Distance_band_unit=`**'kilometers'**
8. Assign the analysis result to a variable named **HS_result**.
9. Run the code cell.
10. In a new code cell, create a map in the notebook that displays `HS_result`.

Run an outlier analysis

You will run an outlier analysis using the same basic steps that you completed to run a hot spot analysis.

1. Add the Find Outliers tool to the notebook.
2. Replace `analysis_layer` with the variable that you used to define the US Counties MOOC feature layer.
3. Update the remaining parameters with the following values:
   - Analysis_field=**"SNAPRate"**
   - Output_name= **"SNAPOutlier_<your initial's and today's date>"**
   - Distance_band=**100**
   - Distance_band_unit=**'kilometers'**
4. Assign the analysis result to a variable named **OA_result**.
5. Run the code cell.
6. In a new code cell, create a map in the notebook that displays `OA_result` using **OA_result['outliers_result_layer']**.

Use the Lesson Forum to post your questions, observations, and syntax examples. Be sure to include the **#stretch** hashtag in the posting title.

Answers to Exercise Questions

1. What distance band was chosen for this analysis?

   195284 meters

2. What statistically significant spatial patterns can you detect from this analysis?

   Generally, the southeastern areas of the contiguous Unites States have statistically significantly high SNAP participation, and the north central areas of the contiguous Unites States have statistically significantly low SNAP participation.