


# Exercise 1: Detect patterns

 How can I print an exercise to PDF format?

## 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 provides slightly different information about these patterns.

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

The Outlier Analysis tool 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 that are significantly dissimilar from their neighbors.

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

## Scenario

The Supplemental Nutrition Assistance Program (SNAP) is a federal program in the United States that helps families buy nutritional food to maintain their health and well-being. In this exercise, you will complete a hot spot analysis and an outlier analysis to find meaningful patterns of high and low SNAP participation. This information can help decision makers distribute resources more efficiently and equitably, ensuring that healthy food is accessible to all SNAP recipients.

**Note:** The exercises in this course include View Result links. Click these links to confirm that your results match what is expected.

**Estimated completion time in minutes: 45 minutes**

Expand all steps ▼

Collapse all steps ▲

- 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://links.esri.com/Section04/Data> and download the exercise data ZIP file.

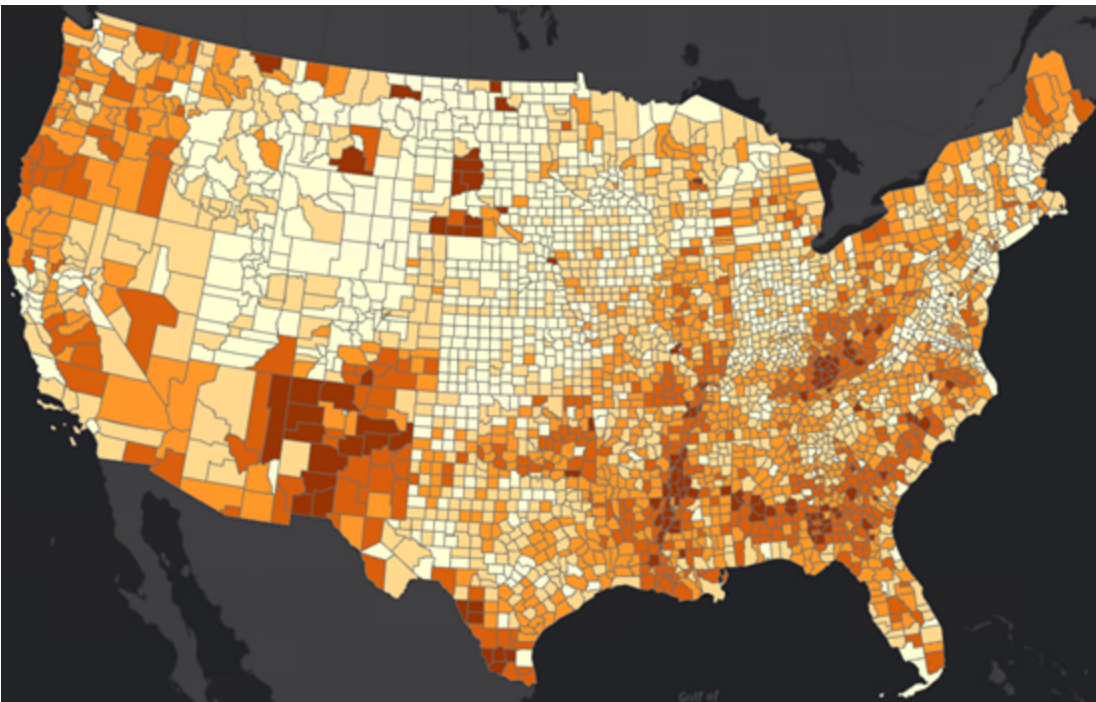
**Note:** The complete URL to the exercise data file is <https://www.arcgis.com/home/item.html?id=c8cc198207034a9f957df448527f21e3>.

- c Extract the files to the EsriTraining folder on your local computer.

## - Step 2: Open an ArcGIS Pro project

In this step, you will open the ArcGIS Pro project that you downloaded.

- a Start ArcGIS Pro.
- b If necessary, sign in using the provided course ArcGIS account.
- c On the Start page, click Open Another Project.  
**Note:** If you have configured ArcGIS Pro to start without a project template or with a default project, you will not see the Start page. On the Project tab, click Open, and then click Open Another Project.
- d In the Open Project dialog box, browse to the PatternDetection\_SpaceTime folder that you saved on your computer.
- e Click PatternDetection\_SpaceTime.aprx to select it, and then click OK.



*Step 2e\*\*\*: Open an ArcGIS Pro project.*

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 2019.

## - 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 that is named SNAPRate, which is the percentage of

eligible people who participate in the SNAP program in each county.

- a In the Geoprocessing pane, under the search field, click the Toolboxes tab.

**Note:** If you closed the Geoprocessing pane, from the Analysis tab, in the Geoprocessing group, click Tools.

- b Expand Spatial Statistics Tools, and then expand Mapping Clusters.
- c Click Optimized Hot Spot Analysis.

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

- d In the Geoprocessing pane, set the following parameters:

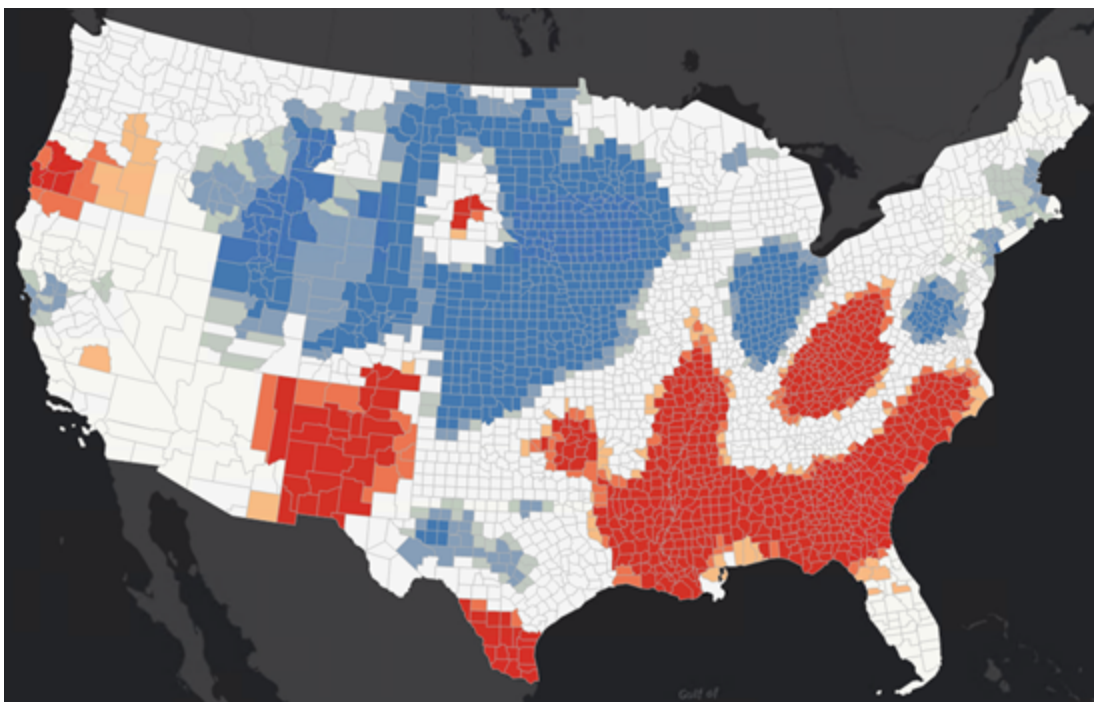
- Input Features: US\_Counties
- Output Features: **SNAPHotSpots**
- Analysis Field: SNAPRate

- e 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 whether 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 the Incremental Spatial Autocorrelation tool, go to [ArcGIS Pro Help: Incremental Spatial Autocorrelation \(Spatial Statistics\)](#).

The fixed-distance band method can be used to define feature neighborhoods. Other methods are available with the traditional Hot Spot Analysis (Getis-Ord Gi\*) tool. To learn more about this tool, go to [ArcGIS Pro Help: Hot Spot Analysis \(Getis-Ord Gi\\*\) \(Spatial Statistics\)](#).

- f 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 not the result of random chance.

- **Step 4: Review the analysis parameters**

Next, you will review the analysis details to ensure that the parameters were appropriate for your analysis of SNAP participation.

- 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.

**Tip:** 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 Click Messages, and then review the geoprocessing details to answer the following question.

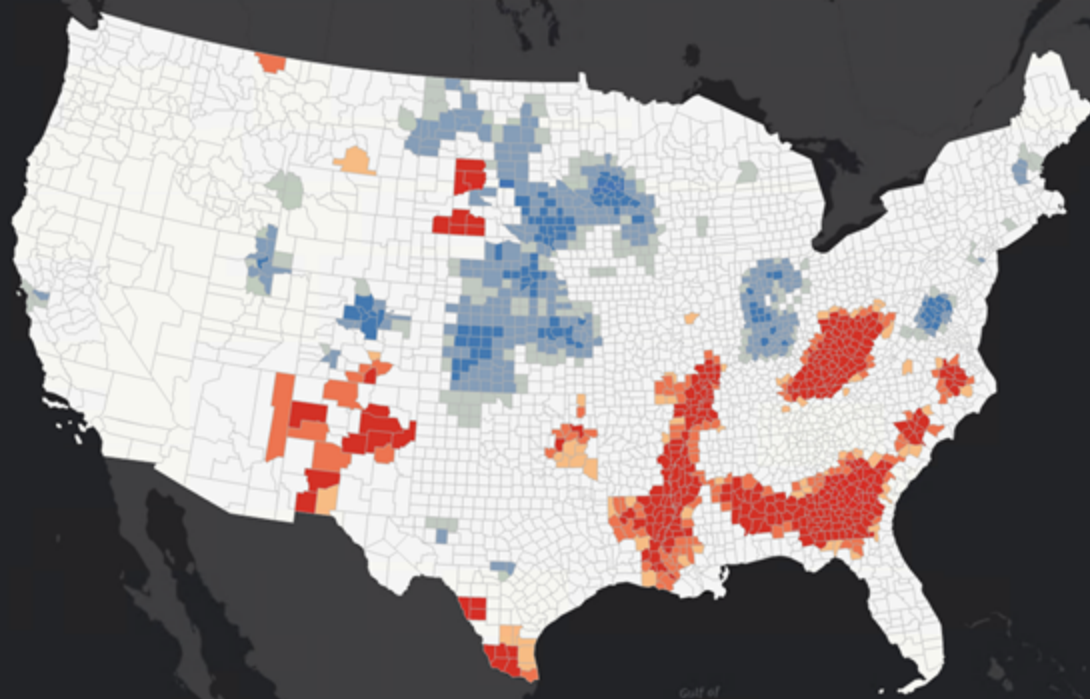
? Which distance band was chosen for this analysis?

- Answer

151156 meters

The tool chose a default distance band of approximately 150 kilometers based on the average distance to 30 nearest neighbors. This default value 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 150-kilometer distance band is too large because you want to analyze more local patterns in SNAP participation. You will reduce the distance band to 75 kilometers 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, in the left-hand field, type **75**.
- f In the field to the right of 75, click the down arrow and choose Kilometers.
- g Click Run.



*Step 4g\*\*\*: Review the analysis parameters.*

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

In this step, you will interpret the results of your statistical analysis.

- 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.

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


- Answer

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

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.

Completing an outlier analysis will help you identify features that have values that are statistically significantly different from neighbors' values. This analysis will provide additional insight into the spatial patterns of the data.

- a In the top-left corner of the Geoprocessing pane, click the Back button , and then search for **outlier**.
- b Click Optimized Outlier Analysis (Spatial Statistics Tools).

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, go to ArcGIS Pro Help: Cluster and Outlier Analysis (Anselin Local Moran's I) (Spatial Statistics).

- c In the Optimized Outlier Analysis tool, set the following parameters:
  - Input Features: US\_Counties
  - Output Features: **SNAPOutliers**
  - Analysis Field: SNAPRate

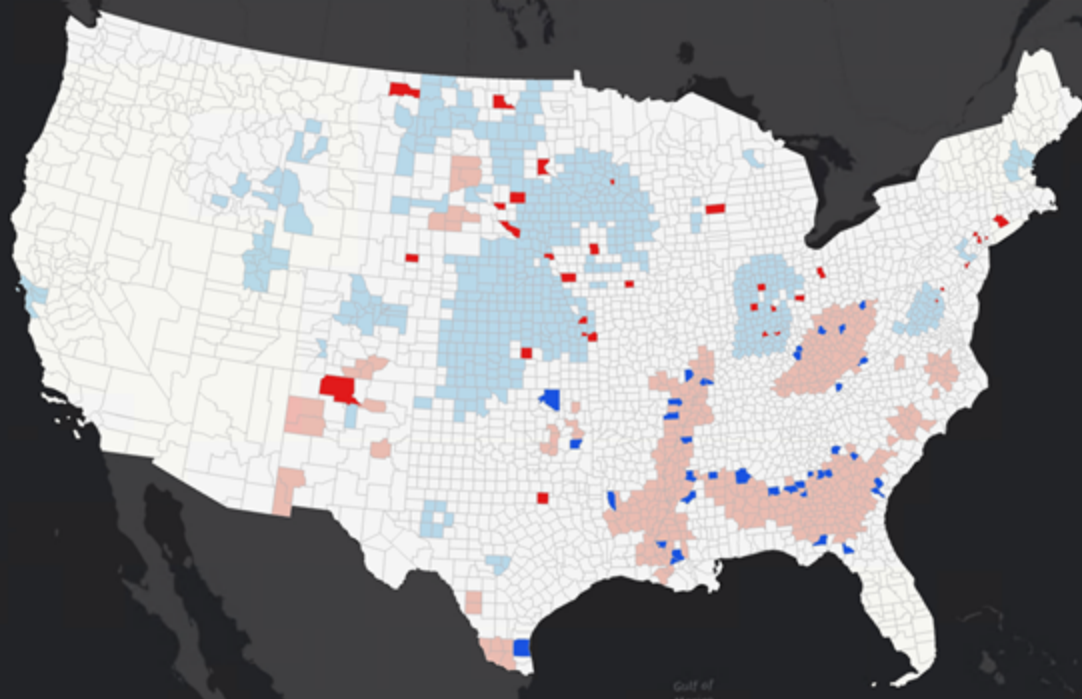
The Performance Adjustment field defines the number of permutations to create a random distribution. The tool will then compare your data's spatial distribution with the randomly generated values. To balance precision and processing time, you will leave the default. For more information about permutations,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

- d Expand Override Settings.

**Tip:** When you compare the results of a hot spot analysis and an outlier analysis, use the same distance band in the analysis.

- e Update the Distance Band fields to **75** Kilometers.
- f Click Run.





*Step 6f\*\*\*: Run an outlier analysis.*

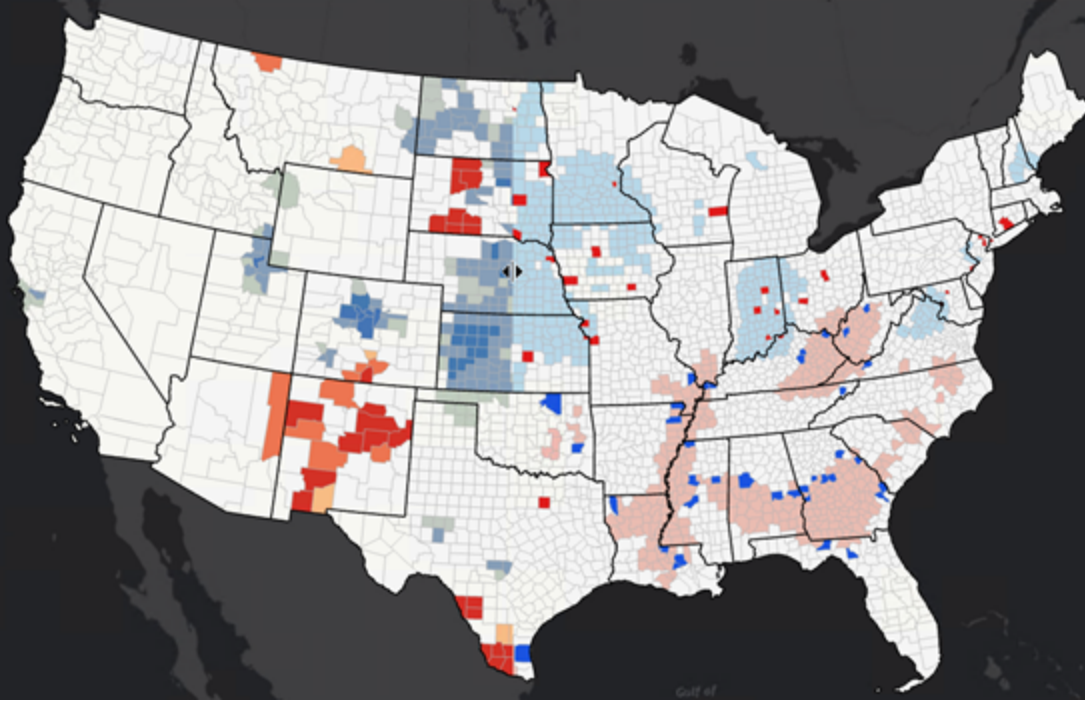
**Note:** The permutations in the Optimized Outlier Analysis tool compare your data values to a set of randomly generated values. Therefore, your results may vary slightly from the View Result graphic.

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 dark 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 spots and cold spots from the Optimized Hot Spot Analysis tool.

#### - **Step 7: Compare the results of analyses**

You have completed two different methods of pattern detection—a hot spot analysis and an 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 Feature Layer tab, in the Compare group, click Swipe.
- d In the Contents pane, click the SNAPOutliers layer.
- e Click the map, and then drag your pointer to the left, to the right, or up and down to compare the results from the Optimized Hot Spot Analysis tool and the Optimized Outlier Analysis tool.



*Step 7e\*\*\*: Compare the results of analyses.*

Using a hot spot analysis and an outlier analysis, you located statistically significant clusters of both high and low SNAP participation. This information can help in the allocation of SNAP resources to areas of higher food insecurities and also help identify areas where eligible residents may be underutilizing or not enrolling in the SNAP program. The results can help drive the decision to distribute resources more efficiently and equitably.

- 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.

#### - Step 8: Stretch goal (Optional)

Another way to run a hot spot analysis in ArcGIS Pro is by using R. The R-ArcGIS Bridge offers you the ability to tap directly into R from your current ArcGIS Pro project. You can then use the R-ArcGIS Bridge to combine the power of R and ArcGIS to solve spatial problems and for spatial data access and visualization. This stretch goal introduces you to the R-ArcGIS Bridge. You will use R to run a hot spot analysis and an outlier analysis.

**Note:** To complete the stretch goal, you must install the following software:

- R 3.5 or newer (<https://www.r-project.org/>)
- R-ArcGIS Bridge (<https://links.esri.com/RArcGISBridge>)

- a Use the following high-level steps to complete this analysis using ArcGIS Pro and RStudio:
  1. Install R, if necessary.
  2. Install R-ArcGIS Bridge, if necessary.
  3. Clone and activate a new arcgispro-py3 environment.



4. Install R-Essentials.
5. In ArcGIS Pro, in the Python Package Manager, make sure that you set your environment to `arcgispro-r`.
6. Explore your data.
7. Run a hot spot analysis using R.
8. Run an outlier analysis using R.

**Note:** For additional support, see the following sites:

- GitHub: R-Bridge-Tutorial-Notebooks (<https://github.com/R-ArcGIS/R-Bridge-Tutorial-Notebooks>)
- Esri: R-ArcGIS Support (<https://links.esri.com/RArcGIS/Support>)
- Esri: R-ArcGIS Bridge Overview (<https://links.esri.com/RArcGIS/Overview>)
- R Tutorial: Hotspot Analysis using Getis Ord Gi ([https://rpubs.com/heatherleeleary/hotspot\\_getisOrd\\_tut](https://rpubs.com/heatherleeleary/hotspot_getisOrd_tut))

- b** Use the Lesson Forum to post your questions and observations. Be sure to include the **#stretch** hashtag in the posting title.
- c** When you are finished, save the project and exit ArcGIS Pro.