Spatial Data Science MOOC

Exercise

Explore data patterns using space-time pattern mining

Section 4 Exercise 2

November 9, 2020

esri
THE SCIENCE OF WHERE

Explore data patterns using space-time pattern mining

Time to complete
80 minutes

Introduction

Space-time pattern mining tools use statistics to incorporate the spatial and temporal aspects of your data to help you understand its spatiotemporal trends.

The initial step in space-time pattern mining is creating a space-time cube. A space-time cube aggregates your data into a multidimensional cube data structure (netCDF). The cube aggregates the data into space-time bins with the x and y dimensions representing space and the t dimension representing time.

Three tools in the Space Time Pattern Mining toolbox can be used to create a space-time cube. The tool that you use depends on your data and spatial question.

Tool	When to use	Example
Create Space Time Cube By Aggregating Points	Analyzing incidents that change location over time	Crime incidents in a state
Create Space Time Cube From Defined Locations	Analyzing fixed locations with attributes that change over time	County tax rates
Create Space Time Cube From Multidimensional Raster Layer	Analyzing multidimensional data that changes over time	Rate of precipitation

After you create a space-time cube, you can analyze the statistically significant patterns in your data using the following space-time pattern mining tools.

Tool	When to use	More information
Emerging Hot Spot Analysis	Identifying trends in the clustering of point counts or attributes in a space-time cube using the Getis-Ord Gi* statistic and the Mann-Kendall trend test	ArcGIS Pro Help: *How Emerging Hot Spot Analysis works*
Local Outlier Analysis	Identifying statistically significant clusters of high and low values as well as outliers that have values that are statistically different from their neighbors in space and time using the Anselin Local Moran's I statistic	ArcGIS Pro Help: *How Local Outlier Analysis works*
Time Series Clustering	Identifying locations in a space-time cube that are most similar and partitions them into distinct clusters, where members of each cluster have similar time-series characteristics	ArcGIS Pro Help: *How Time Series Clustering works*

The Space Time Pattern Mining toolbox also includes the Time Series Forecasting toolset. This toolset includes four tools that you can use to forecast and estimate future values of a space-time cube and evaluate and compare different forecast models at each location in a space-time cube. To learn more about the Time Series Forecasting toolset, see ArcGIS Pro Help: *An overview of the Time Series Forecasting toolset.*

In this exercise, you will create a space-time cube to use in an Emerging Hot Spot Analysis and Local Outlier Analysis.

Exercise scenario

You are assisting in outreach efforts for the Supplemental Nutrition Assistance Program (SNAP). You completed a Hot Spot Analysis and Outlier Analysis to identify areas with statistically significantly high SNAP participation in the contiguous United States during 2016. To continue your analysis, you will use space-time pattern mining tools to determine how these participation rates have changed over time.

Step 1: Open an ArcGIS Pro project

In this step, you will open the ArcGIS Pro project package that you downloaded for the first exercise in this section.

Note: If you did not download the project package, please follow the instructions in the Download The Exercise Data Files step from the previous exercise.

a. Start ArcGIS Pro.

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

c. In the bottom-left corner of the ArcGIS Pro 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.

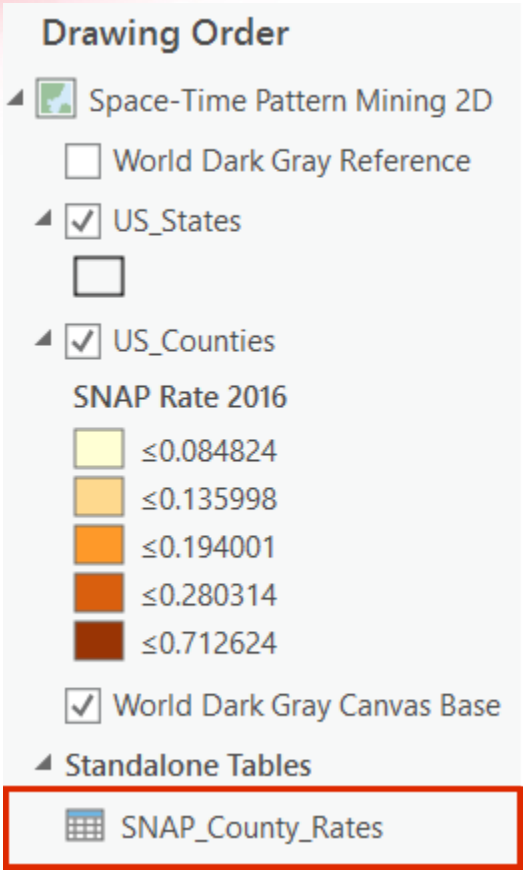
d. Browse to the PatternDetection_SpaceTime folder that you saved on your computer.

e. Click PatternDetection_SpaceTime.ppkx to select it, and then click OK.

f. In the Catalog pane, expand Maps.

g. Under Maps, right-click Space-Time Pattern Mining 2D and choose Open.

Similar to the Pattern Detection map, the Space-Time Pattern Mining 2D map illustrates 2016 SNAP participation rates for counties in the contiguous United States. This map also includes a table of SNAP participation rates from 2000 to 2016 in the Contents pane.



Step 2: Create a space-time cube

Before analyzing the spatial and temporal patterns of this data, you must create a space-time cube.

You are analyzing fixed locations (counties) with an attribute (SNAP rates) that changes over time. You will use the Create Space Time Cube From Defined Locations tool to create the space-time cube.

a. In the Geoprocessing pane, search for and open the **Create Space Time Cube From Defined Locations** tool.

Hint: If you closed the Geoprocessing pane: Analysis tab > Geoprocessing group > Tools.

The input to this tool can be a feature class with repeating shapes or a feature class with one set of shapes and a related table containing the attributes recorded over time. You will use a feature class of the counties and a related table of SNAP participation rates for each county from 2000 to 2016.

b. For Input Features, choose US_Counties.

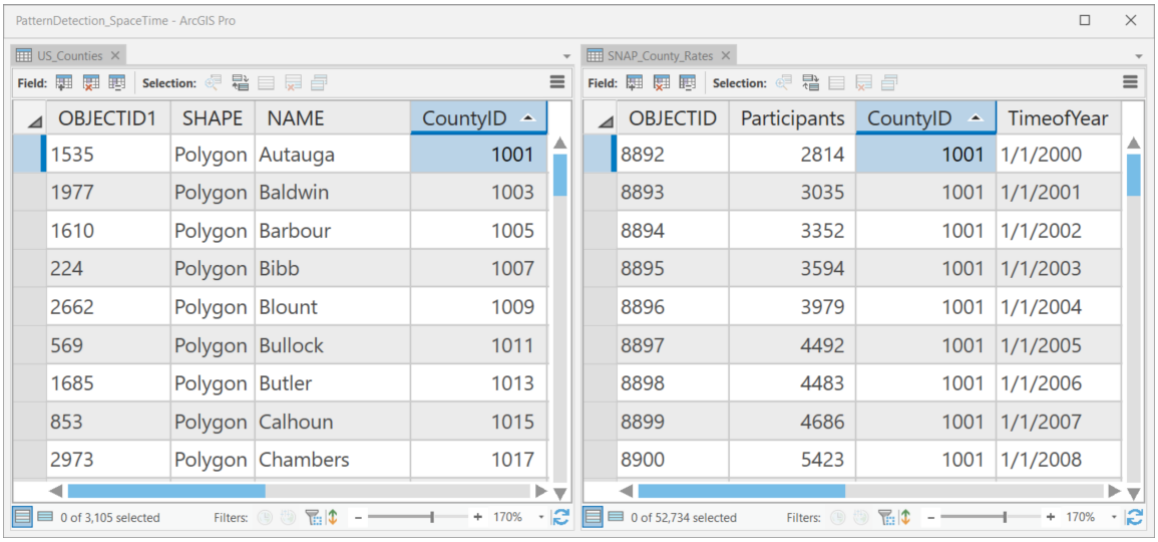
c. For Output Space Time Cube, type **US_SNAPRate_1Yr_STN**.

d. For Location ID, choose CountyID.

e. For Related Table, choose SNAP_County_Rates.

f. For Related Location ID, choose CountyID.

CountyID is a field in the US_Counties feature class and in the SNAP_County_Rates table. It includes the same values, which are used to relate the yearly participation rates from the SNAP_County_Rates table to the geographic location of the counties in the US_Counties feature class.



g. For Temporal Aggregation, leave the option unchecked.

Temporal Aggregation is used when the data was collected in more frequent time intervals than your analysis requires. For example, the data was collected monthly, but you are analyzing yearly trends. This analysis is looking at yearly trends in SNAP participation, and the data was collected yearly, so you will not need a temporal aggregation.

h. For Time Field, choose TimeofYear.

Time Field indicates the field from the related table that contains the timestamp associated with each record in the table.

i. For Time Step Interval, type **1** and update the second parameter to Years.

Because you did not need a temporal aggregation, the Time Step Interval should match the time intervals of your data (yearly).

j. For Time Step Alignment, leave the default.

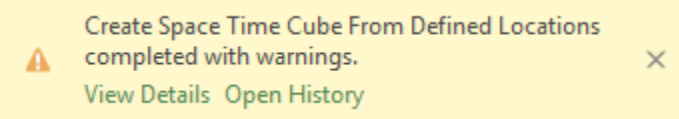
Time Step Alignment determines where aggregation, if necessary, will begin and end. You will not need to aggregate the data for this cube. For more information about time step alignment, see ArcGIS Pro Help: *How Creating a Space Time Cube works*.

k. For Variables, set the Field to SNAPRate.

l. Set Fill Empty Bins With to Space-Time Neighbors.

Variables are the fields with the attribute values that change over time. If your data is missing a value for a specific bin, you can specify what to do with that bin. This decision depends on your data and analysis question. If you are analyzing crime and there is an empty bin, that may mean no crimes took place at that location during that specific time, or it could mean that no data was collected. If the empty bin represents an absence of crime, the bin should be filled with a value of zero. If the empty bin represents the absence of data, the bin should be dropped or filled by averaging its spatial and/or temporal neighbors.

m. Click Run.



The tool creates a netCDF file that will not display in the map, but it is saved to the location that you specified in the Output Space Time Cube parameter. You can find the file path for the cube in the Output Space Time Cube parameter after running the tool.

The tool message window will inform you that the tool completed successfully but with warnings. You will review the warnings to gather more information.

n. Click View Details.

The warning indicates that there were 28 records with null values, which correlates to 28 empty bins in the cube. In the tool parameters, you chose to fill missing bins using space-time neighbors. You can review the tool messages to confirm that values were estimated for these bins.

o. In the tool message window, collapse the Error And Warnings header and the Parameters header.

p. Expand Messages, and then scroll to the Locations section.

Locations	3105
% of locations with estimated observations	0.29
- Total number	9
Total observations	52785
% of all observations that were estimated	0.05
- Total number	28

According to Total Observations, 28 of the 52,785 observations were estimated. The number of estimated observations aligns with the number of null values (28), indicating that the tool was able to successfully estimate values for each bin. The number of estimated locations is nine, which means that some of the 28 observations (or bins) were at the same location but different years.

q. Close the tool message window.

Step 3: Visualize a space-time cube in 2D

You can visualize the results of the space-time cube in 2D and 3D. Each visualization provides different insight into your data.

a. In the Contents pane, turn off the US_Counties layer.

b. In the Geoprocessing pane, click the Back button, and then search for and open the **Visualize Space Time Cube In 2D** tool.

c. For Input Space Time Cube, click the Browse button.

d. Browse to US_SNAPRate_1Yr_STN.nc and click OK.

Note: You can find the Input Space Time Cube (netCDF file) from the file path that you specified in the Output Space Time Cube parameter from the previous geoprocessing tool.

e. For Cube Variable, choose SNAPRATE_NONE_SPACE_TIME_NEIGHBORS.

The Cube Variable is the variable, or variables, that you specified when creating the cube.

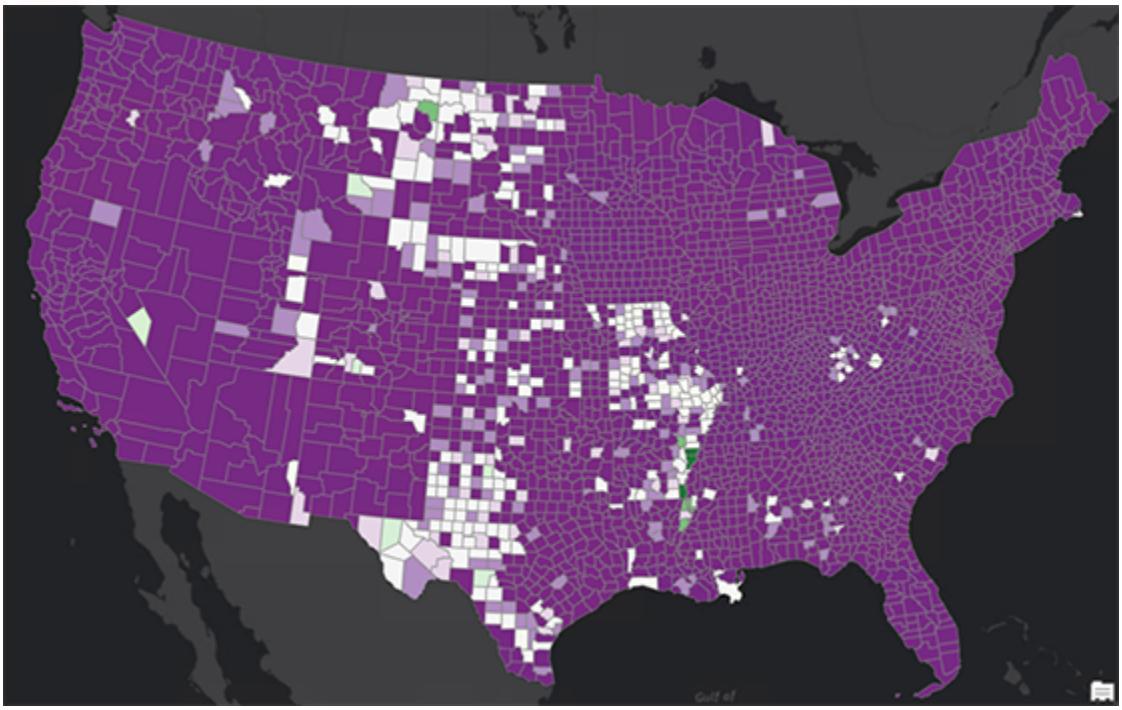
f. For Display Theme, choose Trends.

Note: Display Theme options vary based on the space-time pattern mining analyses that you have completed using the cube. For a list of all possible display themes, see ArcGIS Pro Help: Visualize Space Time Cube in 2D.

g. Check the Enable Time Series Pop-Ups box.

h. For Output Features, type **US_SNAPRate_1Yr_STN_Trends2D**.

i. Click Run.



Trends uses the Mann-Kendall statistic to identify areas that have increased (purple) or decreased (green) SNAP participation rates over time. The varying shades of these colors indicate how confident that you can be that these increased and decreased trends are not random and represent meaningful patterns.

j. Zoom to an area on the map, and then click one of the counties.

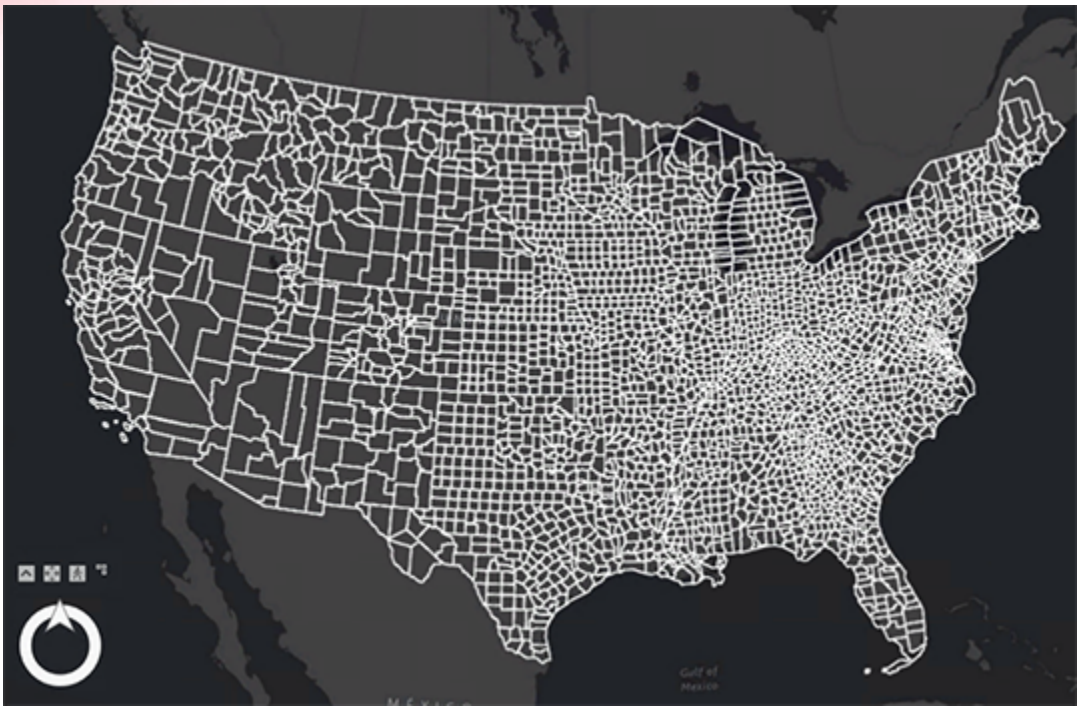
A time-series pop-up window appears. Time-series pop-ups include a chart that visualizes the time-series trends for each county. You can use the chart to better understand the overall trend of a county's SNAP participation and identify its yearly SNAP participation rates.

k. Close the pop-up window.

Step 4: Visualize a space-time cube in 3D

To visualize the cube in 3D, you will use a scene. Scenes allow you to visualize your data and analyze geographic information in an interactive 3D environment.

a. In the Catalog pane, under Maps, right-click Space-Time Pattern Mining 3D and choose Open Local View.

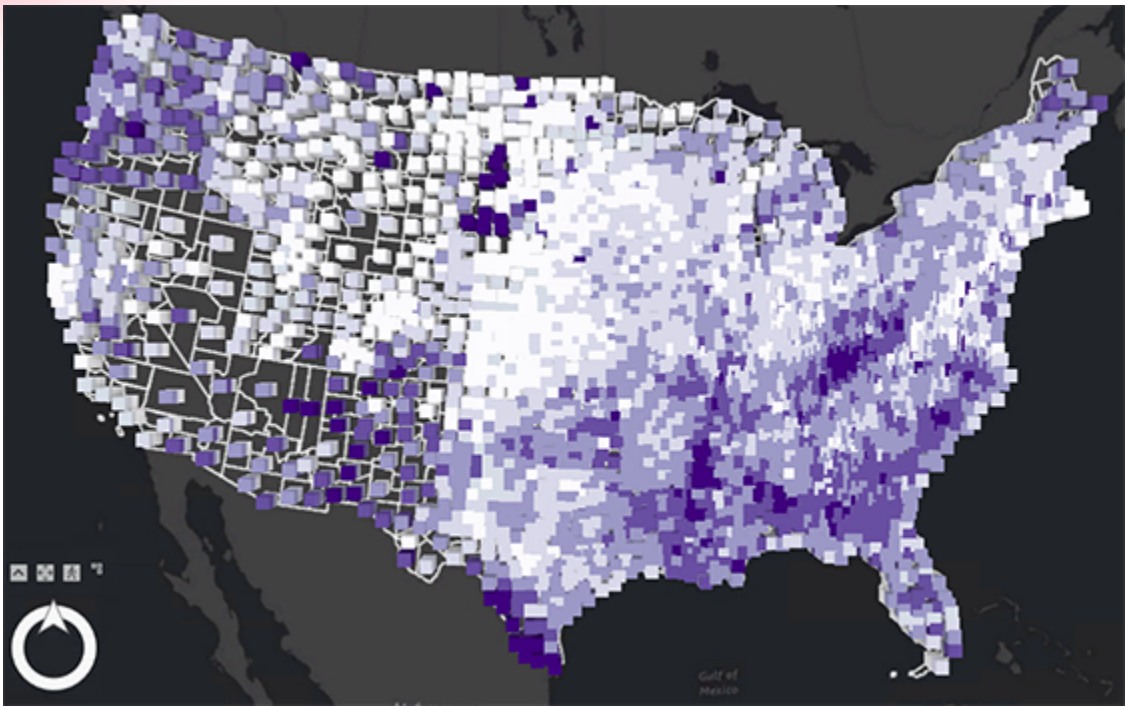


A scene zoomed to the contiguous United States appears.

b. In the Geoprocessing pane, search for and open the **Visualize Space Time Cube In 3D** tool, and then set the following parameters:

- Input Space Time Cube: US_SNAPRate_1Yr_STN.nc *(Hint: Use the Browse button)*
- Cube Variable: SNAPRATE_NONE_SPACE_TIME_NEIGHBORS
- Display Theme: Value
- Output Features: **US_SNAPRate_1Yr_STN_Value3D**

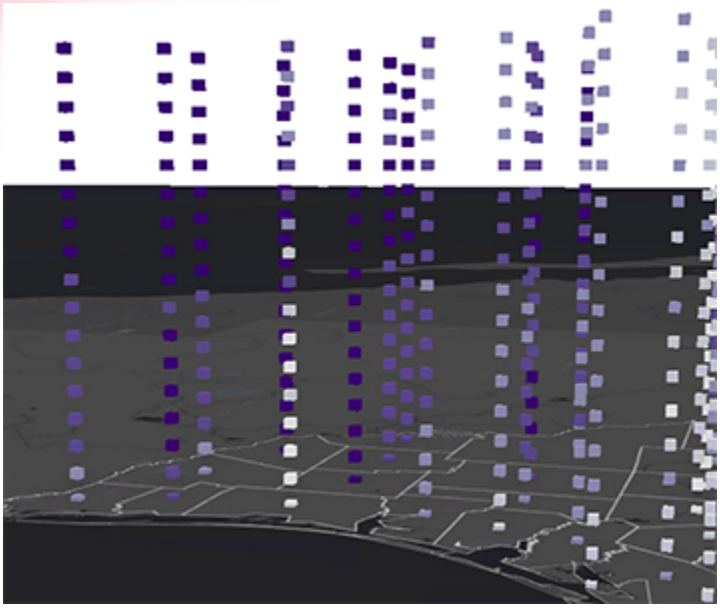
c. Click Run.



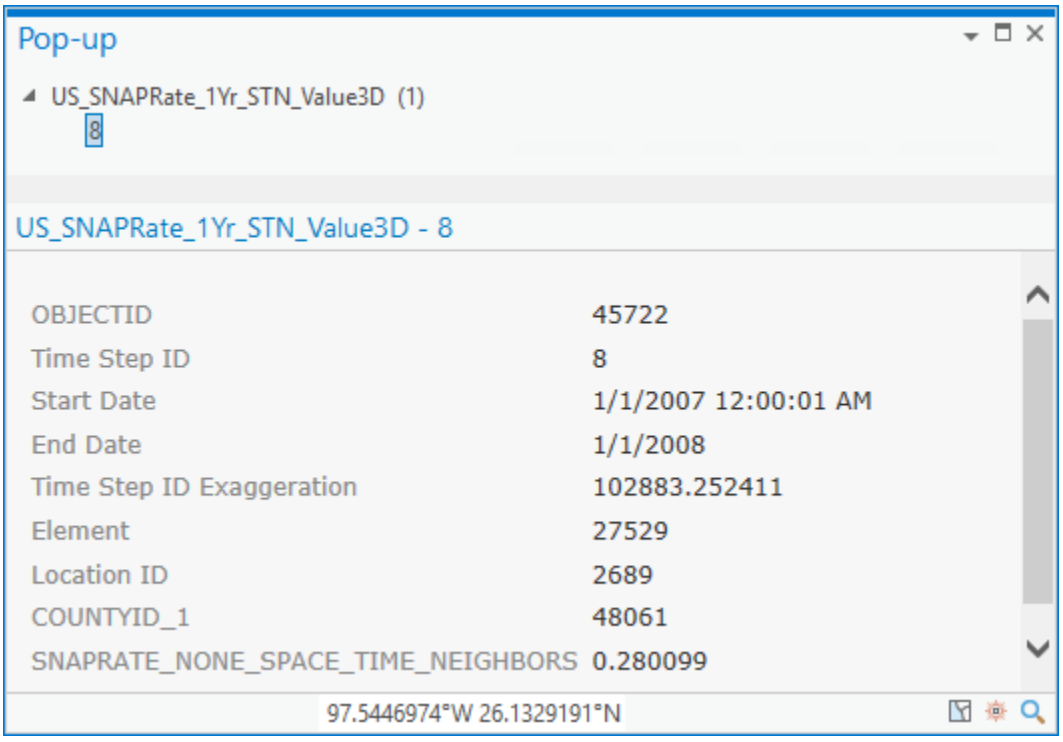
Note: It may take a few minutes for the 3D layer to appear. Processing time for 3D visualization can be affected by a few factors, including your computer's graphics card and the number of features to display. If the layer does not display in a few minutes, try zooming in to the screen so that there are fewer features to display at one time.

A 3D layer of SNAP participation rates appears in the scene. Each column represents the SNAP participation rates for a county over time. You can zoom and pan the scene to examine individual counties.

d From the Map tab, in the Navigate group, click Bookmarks and choose View Values.



e In the scene, click one of the bins.



A pop-up window appears that includes the county information and SNAP participation rate for the specified year.

Creating and visualizing a space-time cube transforms your spreadsheet of data into an interactive 3D view. You can easily see the SNAP participation rates for each county and how the rates have changed over time. Next, you will determine if there is statistically significant clustering in this data.

f. Close the pop-up window, and then save the project.

Step 5: Run an Emerging Hot Spot Analysis

You will use the Emerging Hot Spot Analysis tool to analyze a feature's spatial and temporal neighbors to determine if the feature and its neighborhood are statistically significantly different from the study area. This analysis will help you determine where and when there is statistically significant clustering of high and low SNAP participation rates.

a. At the top of the scene view, click the Space-Time Pattern Mining 2D map tab.

b. In the Contents pane, turn off the US_SNAPRate_1Yr_STN_Trends2D layer.

c. From the Map tab, in the Navigate group, click Bookmarks and choose Contiguous US.

d. In the Geoprocessing pane, search for and open the **Emerging Hot Spot Analysis** tool.

e. For Input Space Time Cube, browse to the US_SNAPRate_1Yr_STN.nc file.

f. For Analysis Variable, choose SNAPRATE_NONE_SPACE_TIME_NEIGHBORS.

g. For Output Features, type **US_SNAPRate_1Yr_STN_EHS**.

h. For Conceptualization Of Spatial Relationships, confirm that Fixed Distance is chosen.

i. For Neighborhood Distance, set it to **75** Kilometers.

If comparing the results of this tool to other Space-Time Pattern Mining or Pattern Detection tools, use the same Conceptualization Of Spatial Relationships and distance band to ensure that the analysis parameters are consistent.

j. For Number Of Spatial Neighbors, leave the field blank.

The Number Of Spatial Neighbors parameter is used in conjunction with the Conceptualization Of Spatial Relationships parameter to define each feature's neighborhood. It is typically used with the K Nearest Neighbors Conceptualization Of Spatial Relationships. If used with fixed distance, the assigned distance band is used to define neighborhoods unless

the band does not reach the specified number of spatial neighbors. In these cases, the distance band will extend to reach the specified number of spatial neighbors.

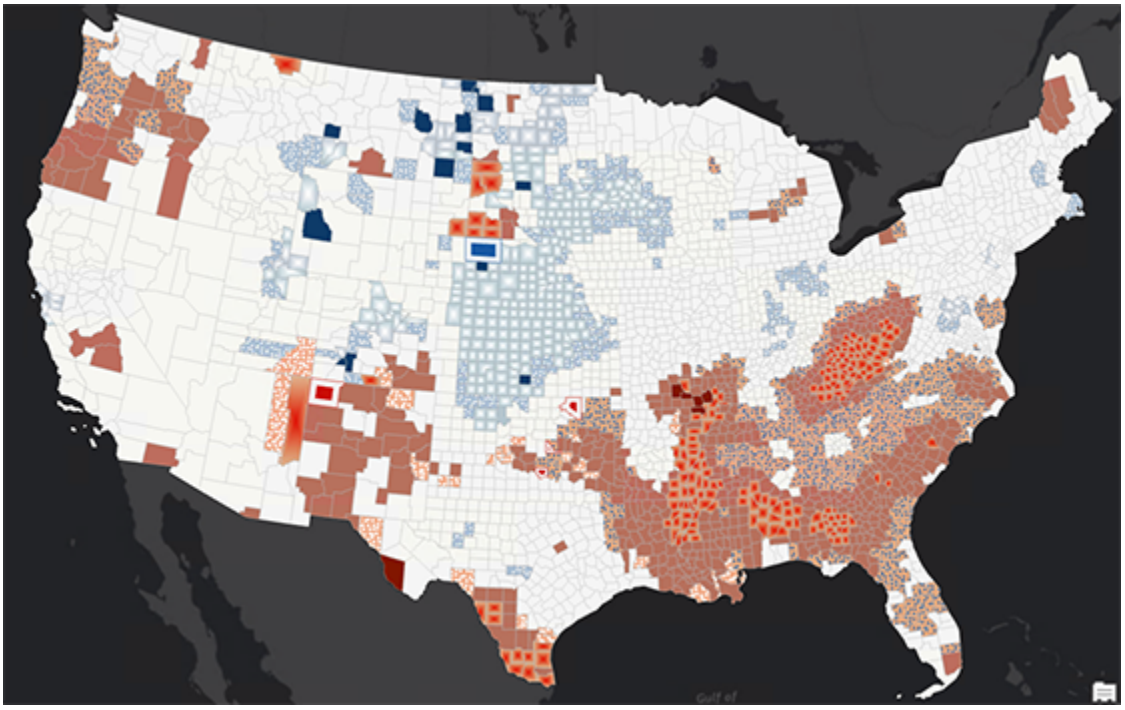
k. For Neighborhood Time Step, leave the default.

The Neighborhood Time Step defines the temporal neighborhood. Using a time step of 1 includes the feature, its neighborhood, and the feature and neighborhood from the previous year.

l. For Define Global Window, leave the default.

In Hot Spot Analysis, a feature and its neighborhood are compared to the study area. In Emerging Hot Spot Analysis, you can compare the feature and its neighborhood to the entire cube, bins within the neighborhood time step, or bins within the same time step. For more information, see ArcGIS Pro Help: *Emerging Hot Spot Analysis*.

m. Click Run.



Each location is evaluated based on its hot spot and cold spot results each year and then categorized (intensifying hot spot, persistent cold spot, and so on) to help you interpret the results. The layer symbology represents the different categories.

n. Open a web browser tab and go to ArcGIS Pro Help: *How Emerging Hot Spot Analysis works*.

o Review the Emerging Hot Spot Analysis result symbology, pattern name, and definition.

1. **What patterns can you detect from this analysis?**

The Emerging Hot Spot Analysis categories are based on patterns detected over time. You can visualize this information in 3D to see why each location was assigned a specific category. You can also use 3D visualization to learn more about a specific area of interest.

p Close the web browser tab and return to ArcGIS Pro.

q Save the project.

Step 6: Visualize Emerging Hot Spot Analysis in 3D

a At the top of the map view, click the Space-Time Pattern Mining 3D scene tab.

b In the Contents pane, turn off the US_SNAPRate_1Yr_STN_Value3D layer.

c From the Map tab, zoom to the Contiguous US bookmark.

Note: ArcGIS Pro provides a History pane to access all the geoprocessing tools that you have used in this ArcGIS Pro project. You can double-click a tool to reopen it with the parameters used in that analysis. This workflow can save you time when running consecutive analyses that require the same tool with many of the same parameter values.

d From the Analysis tab, in the Geoprocessing group, click History.

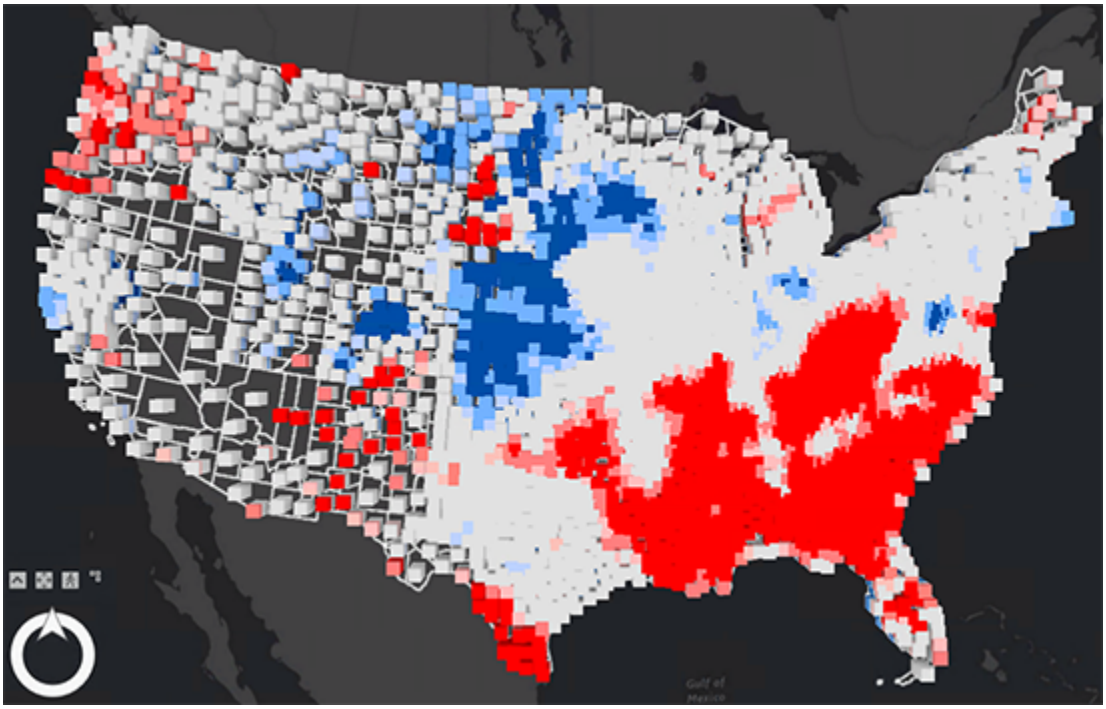
e In the History pane, double-click your previously run Visualize Space Time Cube In 3D tool to open it.

Your input space-time cube and cube variable will remain the same as the last run, but you will update the other two parameters.

f. In the Geoprocessing pane, update the following parameters:

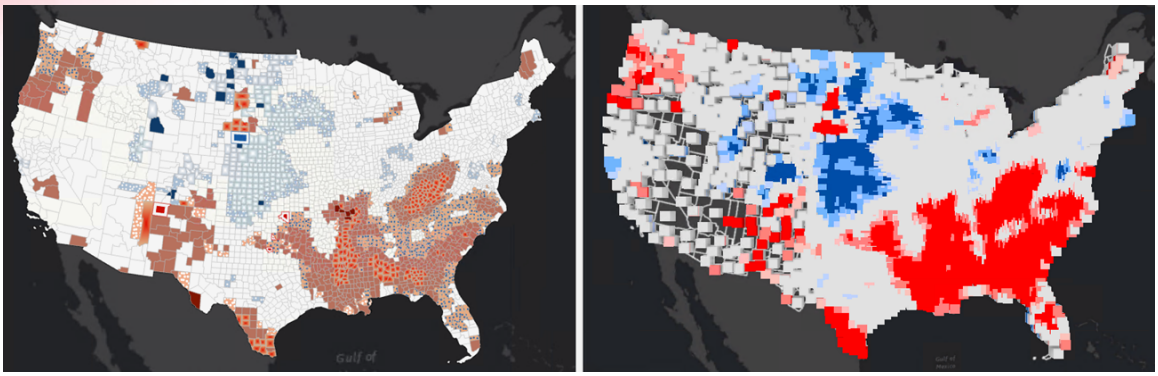
- Display Theme: Hot And Cold Spot Results
- Output Features: **US_SNAPRate_1Yr_STN_EHS3D**

g. Click Run.



A 3D layer of hot and cold spots appears in the scene. Each column indicates if there were clusters of high or low SNAP participation rates for each year. By comparing the 2D and 3D analysis results, you can gain a better understanding of the yearly trends that contributed to each category.

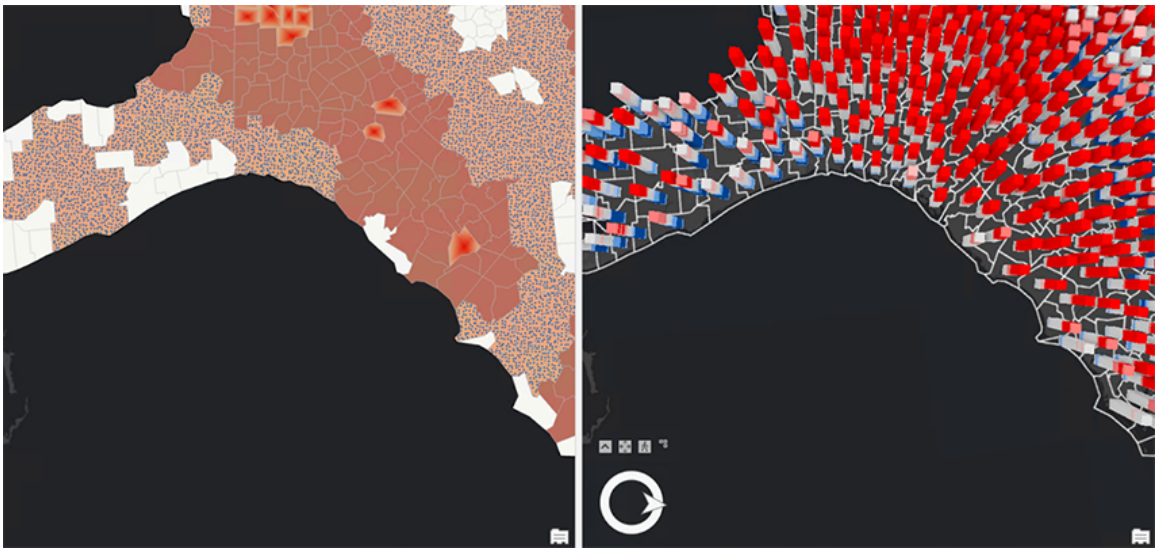
h. At the top of the scene view, right-click the Space-Time Pattern Mining 3D scene tab and choose New Vertical Tab Group.



The map and scene appear side by side. You can link these views so that moving one view will automatically update the other view.

i. From the View tab, in the Link group, click the Link Views down arrow and choose Center And Scale.

j. From the Map tab, zoom to the Hot Spots bookmark.



Note: Your views may differ slightly from the preceding graphic.

At this location, you can see the difference in yearly values between oscillating and consecutive hot spots.

This information provides additional insight into your data. Based on these patterns, you may want to investigate areas with oscillating hot spots to see if the increase in SNAP participation rates requires additional resources that can improve access to healthy foods.

Before you continue, you will unlink the map and scene and move the scene back so that each view is displayed separately.

k. To unlink your map and scene, from the View tab, in the Link group, click Link Views.

l. To no longer view the map and scene side by side, right-click the Space-Time Pattern Mining 3D scene tab and choose Move To Previous Tab Group.

m. Save the project.

Step 7: Run a Local Outlier Analysis

Completing a Local Outlier Analysis in conjunction with an Emerging Hot Spot Analysis will help you determine which areas are statistically significantly different from their neighbors over time.

a. View the Space-Time Pattern Mining 2D map.

b. In the Contents pane, turn off the US_SNAPRate_1Yr_STN_EHS layer.

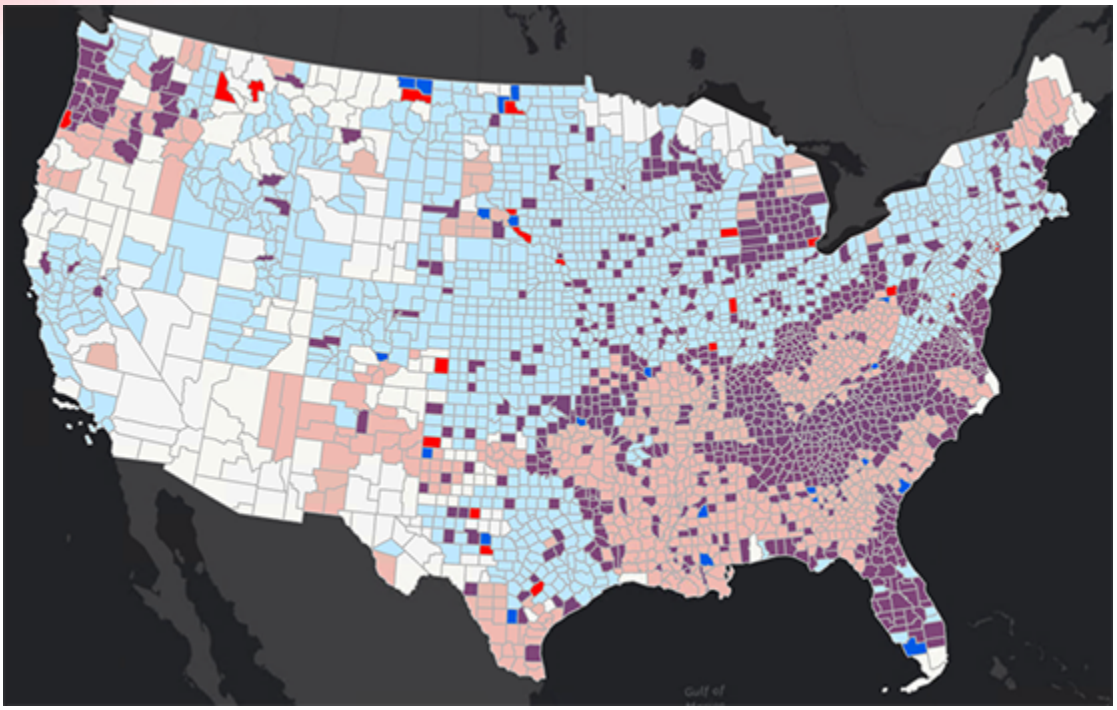
c. Zoom to the Contiguous US bookmark.

d. In the Geoprocessing pane, search for and open the **Local Outlier Analysis** tool, and then set or confirm the following parameters:

- Input Space Time Cube: US_SNAPRate_1Yr_STN.nc
- Analysis Variable: SNAPRATE_NONE_SPACE_TIME_NEIGHBORS
- Output Features: **US_SNAPRate_1Yr_STN_LO**
- Conceptualization Of Spatial Relationships: Fixed Distance
- Neighborhood Distance: **75** Kilometers

e. Leave the remaining defaults and click Run.

If you are comparing the results of this tool to other Space-Time Pattern Mining tools, use the same parameters to ensure that the analysis is consistent.



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

Each feature is evaluated and then categorized (High-High Cluster, High-Low Outlier, and so on) to help you interpret the results. Clusters indicate areas where the only statistically significant type throughout time has been a high or low cluster. Outliers indicate areas where the only statistically significant type throughout time has been a High-Low or Low-High Outlier. Multiple types indicate areas where there have been multiple types of statistically significant clusters and/or outliers throughout time.

You will visualize the cube in 3D to examine the outliers, determining what years and areas were statistically different from their neighbors.

Step 8: Visualize a Local Outlier Analysis in 3D

a. View the Space-Time Pattern Mining 3D scene.

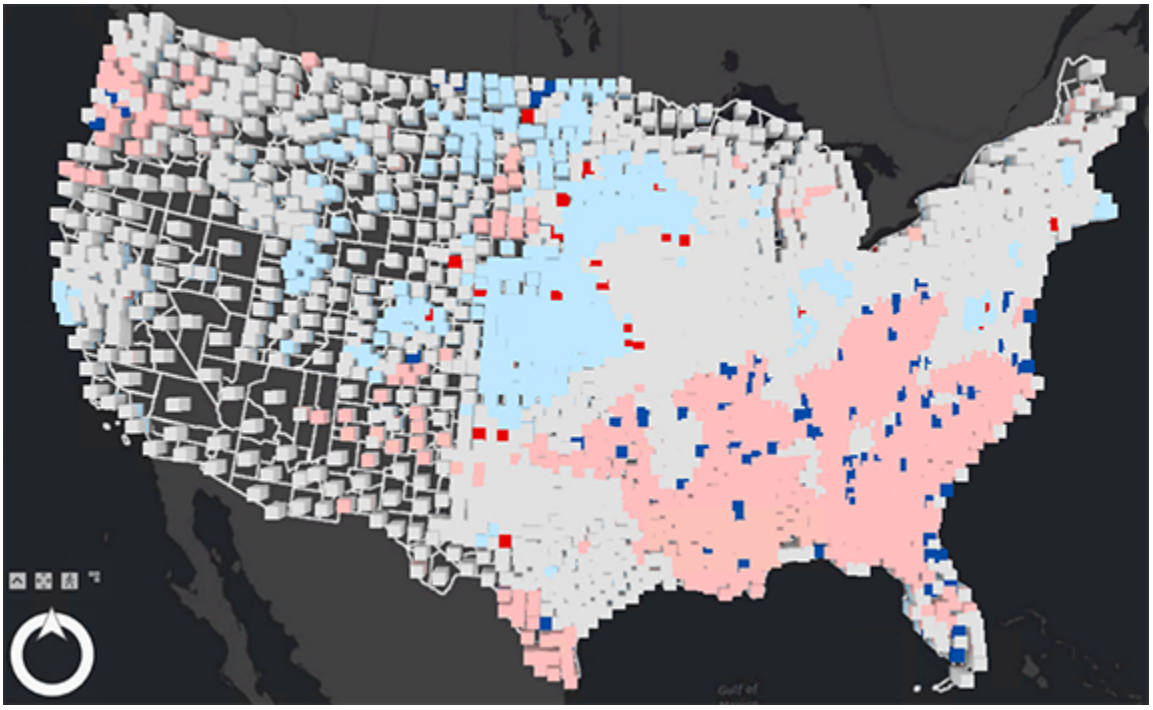
b. In the Contents pane, turn off the US_SNAPRate_1Yr_STN_EHS3D layer.

c. From the Map tab, zoom to the Contiguous US bookmark.

d From the History pane, double-click the previously run Visualize Space Time Cube In 3D tool to open it, and then update the following parameters:

- Display Theme: Cluster And Outlier Results
- Output Features: **US_SNAPRate_1Yr_STN_LO3D**

e Click Run.

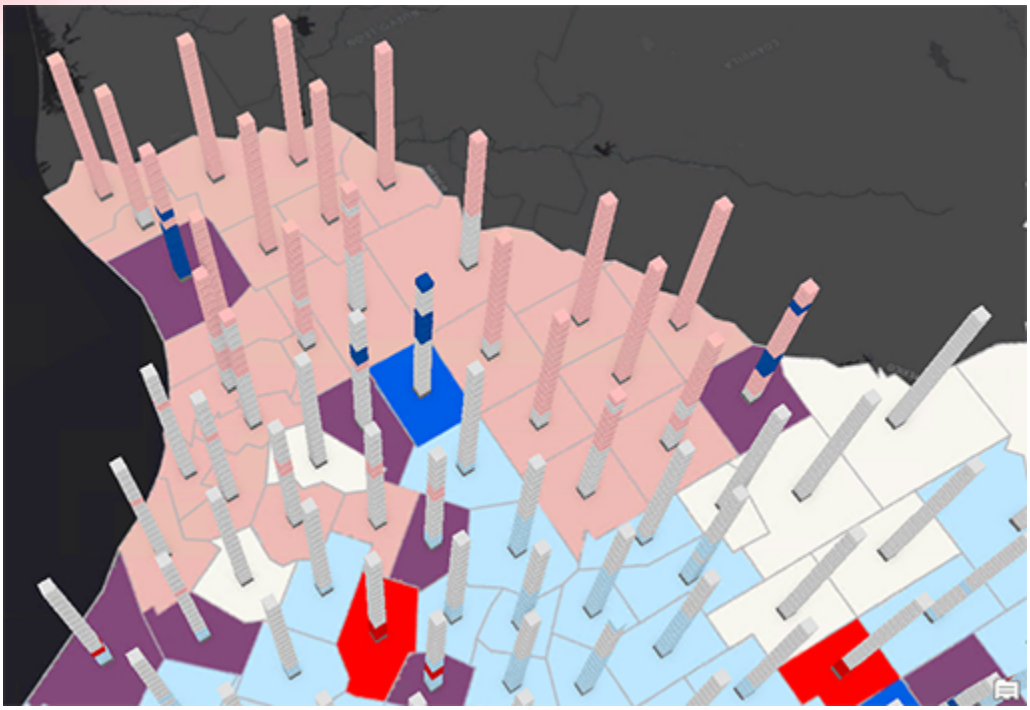


A 3D layer of cluster and outliers appears in the scene.

f Click the Space-Time Pattern Mining 2D map tab, and then in the Contents pane, right-click the US_SNAPRate_1Yr_STN_LO layer and choose Copy.

g Return to the Space-Time Pattern Mining 3D scene, and then in the Contents pane, right-click Space-Time Pattern Mining 3D and choose Paste.

h Zoom to the Outliers bookmark.



Adding the 2D Local Outlier Analysis result layer with the 3D Local Outlier Analysis layer can help you gain a better understanding of the yearly trends that contributed to each category. This information provides more insight into your data. For example, you may want to investigate these outliers in further detail to determine if they are the result of environmental factors, socioeconomic changes, outreach efforts, or a combination of the three.

You have completed an Emerging Hot Spot Analysis and Local Outlier Analysis to help you evaluate how SNAP participation rates have changed over time. 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.

Save the project and exit ArcGIS Pro.

Answers to Exercise Questions

1. What patterns can you detect from this analysis?

   The southeastern areas of the contiguous United States have a mixture of consecutive, intensifying, and oscillating hot spots. In this case, the oscillating hot spots indicate areas that started with low SNAP participation rates but in recent years have had high participation rates. The north-central areas of the United States have mostly diminishing cold spots, which means that the areas have significant clustering of low SNAP participation rates, but the intensity of that clustering has been decreasing. This pattern indicates that those areas are moving away from being cold spots and toward more of a random pattern.