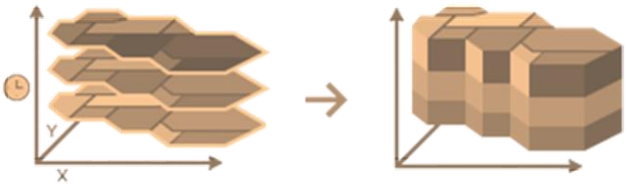


"Lost in Spatial Data"

Written by Sumeeta Srinivasan, updated October 2025 by Sumeeta Srinivasan

- Descriptive spatial statistics in ArcGIS: Mean Centroid, Standard Deviation Ellipses
- Spatial autocorrelation: Global Moran's I and Local Moran's I
- Emerging hotspots in space and time
- Local Outlier analysis in space and time
- Visualization in 3D

Estimated time: 4 hours

Contents

In this exercise you will work with spatial statistics tools first and then space time statistics in ArcGIS.

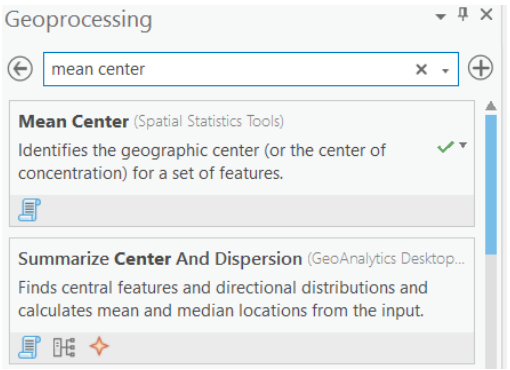
Data:

- SF Crime data in the **geodatabase SFCrime.gdb**[1] from the City of San Francisco for 2013, 2017, 2021 and 2024 with three types of crimes (drug offences, vehicle theft and assaults) as points in the feature dataset crime. A point feature class for all crimes 2018-2025.

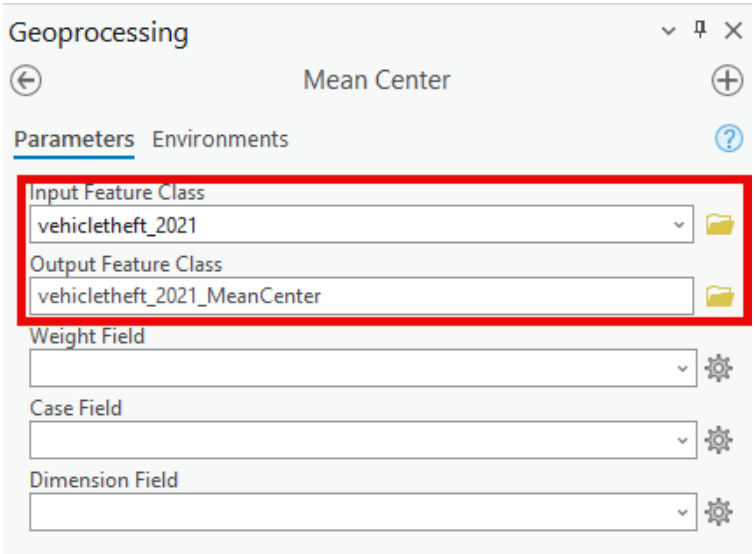
1.0 Descriptive Summaries: Standard Deviational Ellipse and Centroid

The **mean center**[2] is the average x and y coordinate of all the features in the study area weighted by the value of the variable that you are trying to measure the mean center for (the number of crimes in a block group for example). **Standard distance**[3] measures the degree to which features are concentrated or dispersed around the points. One way of describing the spatial distribution of a set of points is to calculate the standard distance separately in the x and y directions. An ellipse composed of these two distances is referred to as the **standard deviational ellipse**[4]. The ellipse allows you to see if the distribution of features is elongated and thus has a particular orientation.

If the underlying pattern in the data is **normally distributed** about the mean, the one standard deviation ellipse polygon will cover approximately 68 percent of the data in the cluster. Two standard deviations will contain approximately 95 percent of the data, and three standard deviations will cover approximately 99 percent of the data in the cluster.



1. Copy the geodatabase to your own folder for this exercise
2. Start up a new project in ArcGIS Pro with a blank map and save it to the folder for this exercise.
3. In ArcGIS Pro, to calculate the mean center, in the Analysis menu click on Tools and then Toolboxes in the window that opens and expand "**Spatial Statistics Tools**" and then "**Measuring Geographic Distributions**" and then double click on "**Mean Center**" to get a new window. (Or simply search for Mean center)



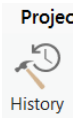
4. Input the feature class to be one of the layers from the crime feature dataset: **whichever type of crime or year combination you want from the crime feature dataset**, output the

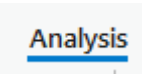
[1] https://datasf.org/opendata/
[2] https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/mean-center.htm
[3] https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/standard-distance.htm
[4] https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/directional-distribution.htm

feature class to the **gdb** in your folder and give it a name that makes sense to you later, there is no weight field for the point shapefiles (each point is a crime). Click OK to get a mean center for the crime.

Note that if your input was polygons, you may have different values for each polygon in terms of the number of crimes. Then you would weight it by that attribute.

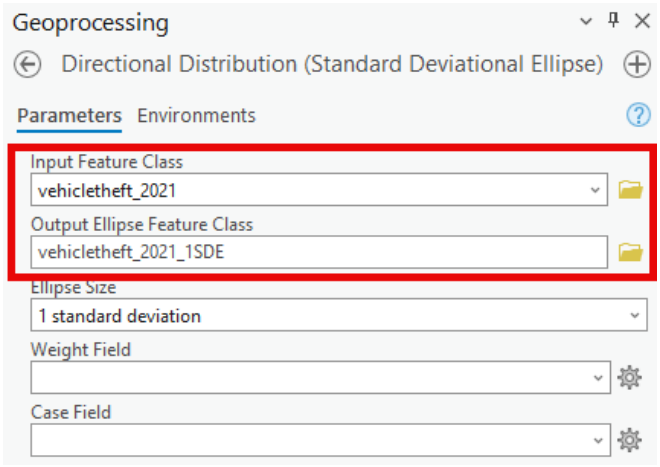
5. Repeat this for another type of crime in the **same year** or the **same crime** in a different year.

Recall that **History** 

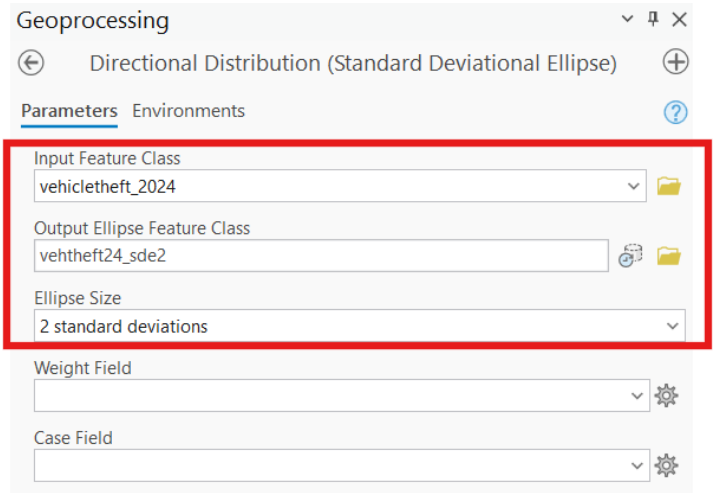
in the Analysis ribbon 

will show you the progress and results (if any) of your tool calculations.

6. In Toolbox, expand "**Spatial Statistics Tools**" and then "**Measuring Geographic Distributions**" and then double click on "**Directional Distribution**" and create a standard deviational ellipse for the crime feature classes that you estimated the mean center for. Create an ellipse for **1 standard deviation** for the same **two** feature classes for which you calculated the mean center.



7. (Optional) Repeat step 6 for the 2nd standard deviation ellipse for the same feature classes.



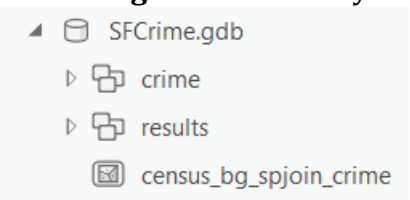
How are the patterns different?

Question 1 (0.5 point)

Compare the mean center and standard deviational ellipses of **two different crimes in the same year** or the **same crime** in **two different years**. Attach a screenshot that shows the centroids and SDE for both points of time or different crimes or for crime compared to population in one map and describe in a few sentences how they help you in summarizing the crime patterns that you see in San Francisco and how they contrast with either the other crime or the same crime in another year.

Data used in this part of the lab

- SF Crime data by block group **feature class census_bg_spjoin_crime** in the geodatabase **SFCrime.gdb** from the City of San Francisco for the years 2013, 2017, 2021 and 2024.

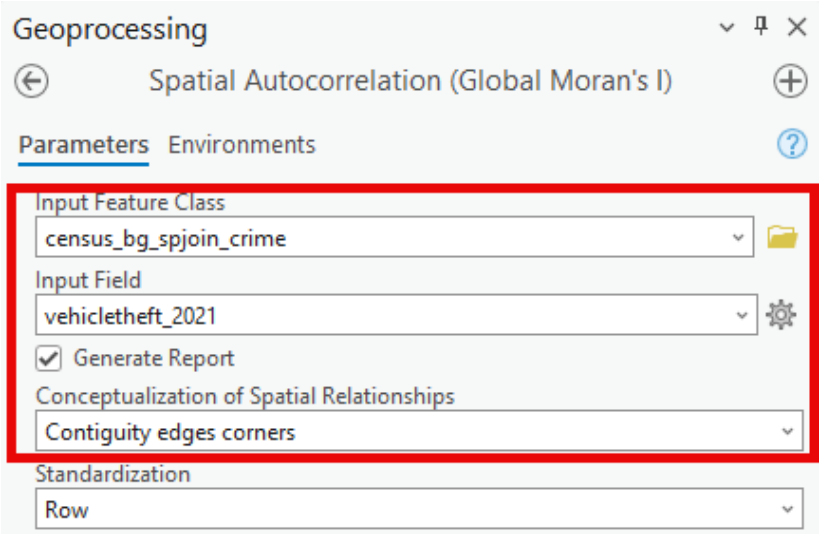


1.2 Descriptive Summaries: Global and Local Moran

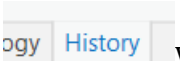
Another summary that is useful in looking for overall (or global) clustering is the Moran's coefficient. However, unlike the previous section we cannot use the original points since they don't have an attribute that measures **value** at that location. Instead, you will use the feature class census_bg_spjoin_crime which includes attributes that are counts of the number of crime points in each **census block group**. The crime variables are for assaults, drug offences and vehicle theft for the years 2013, 2017, 2021, 2024. These are the names of some of the attributes:

assault_2017	drugoffence_2013	drugoffence_2017	vehicletheft_2013	vehicletheft_2017	assault_2021	vehicletheft_2021	drugoffence_2021	assault_2024

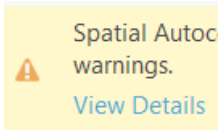
8. In the Geoprocessing Toolboxes, expand "**Spatial Statistics Tools**" and then "**Analyzing Patterns**" followed by "Spatial Autocorrelation (Moran's I)".
9. Follow the screenshot to calculate the Moran's I for the same variables (two different crimes in the same year or the same crime in two different years) that you analyzed in the previous section (but now for a different spatial unit – block group). Make sure that the generate report check box is checked.



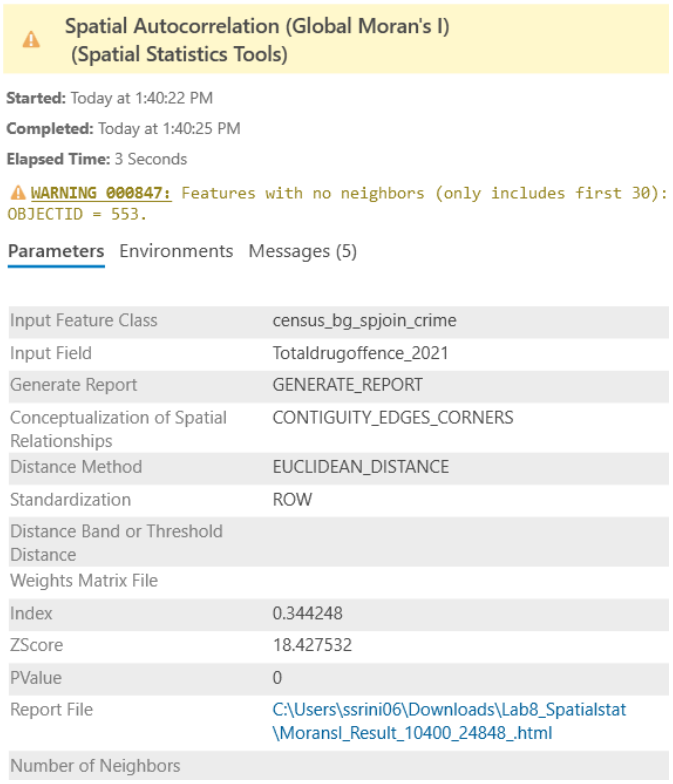
10. The results are saved to a HTML report file that can be opened (click on it) from the



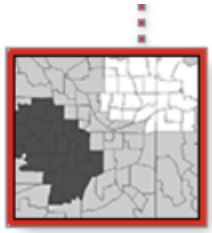
window. You can also find it in your folder for this exercise (if you checked the generate report box). You can also see it in the



View Details link that will show up after the process runs. Note that this is warning you that some locations have no neighbors (islands).



Attach the report to your analysis. What do the results suggest about the attribute? Also compare it to the **other attribute** for which you generated Moran's I. Does the crime that you chose appear to be significantly clustered for that year?

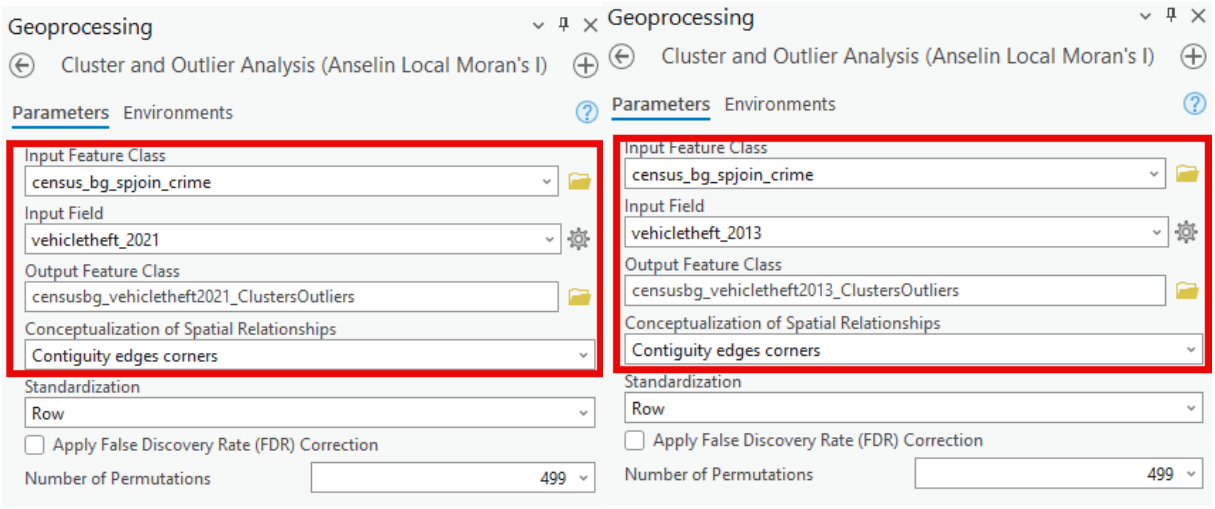


Clustered

Next, look at **<u>where</u>** there is **clustering** of crime. Unlike the Moran's I which results in one number the Moran's Index (and a z score to assess statistical significance), the local version of Anselin Local Moran's Index, should result in a feature class with a legend showing you where the clusters

(high values surrounded by high values or low values surrounded by low values) and outliers (high surrounded by low or low surrounded by high values) are.

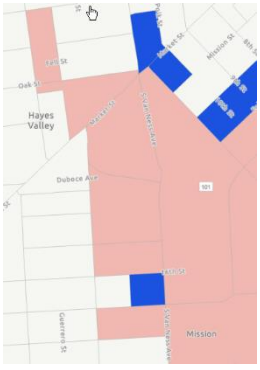
11. In Geoprocessing Toolbox expand **Spatial Statistics**, then **Mapping Clusters** and then double click on **Cluster and Outlier Analysis**. (See screenshot that follows). Again, select the same variable as you did for Global Moran's I spatial autocorrelation.
12. **Do this again** for the other crime or other year that you want to compare it with.



13. Open the attribute table for the new feature class that is added to the **Contents** window.

Notice that you get new attributes associated with each location (with prefix LmiIndex, LMi ZScore, LMi PValue). A high positive Z score for a feature indicates that the surrounding features have similar values (either high or low). A group of adjacent features having high Z scores indicates a cluster of similarly high or low values. A low negative Z score for a feature indicates the feature is surrounded by dissimilar values—that is, if a feature gets a negative Z score, it has a different value than its neighbors (a high value relative to a neighborhood that has low values or a low value relative to a neighborhood that has high values).

LMiIndex RS	LMiZScore RS	LMiPValue RS	COType RS	NNeighbors RS
0.019662	0.432735	0.012	LL	7
0.010624	0.509201	0.014	LL	8
0.0287	0.438952	0.028	LL	7
0.0287	0.436075	0.03	LL	7



The COtype tells you if it is part of a cluster or outlier. The default symbolization in the Contents is to point out whether you have a HH – high value (of the crime) location surrounded by High values (of the crime) or LL – low surrounded by low values; HL – high surrounded by low or LH – low surrounded by high values.

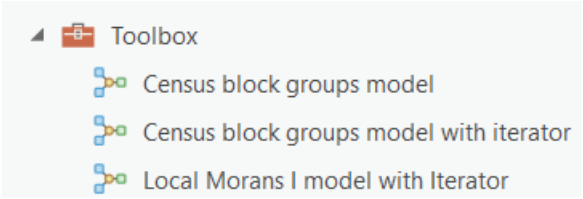
14. Where are the HH or HL locations? (Use the basemap to find neighborhood names which are part of HH clusters).

Question 2 (1 point)

Discuss in a paragraph what Moran's I and Local Moran's I show about the patterns of crime(s) in San Francisco. Use screenshots to illustrate your paragraph as you compare **either two** crimes in the same year or the same crime in **two** different years.

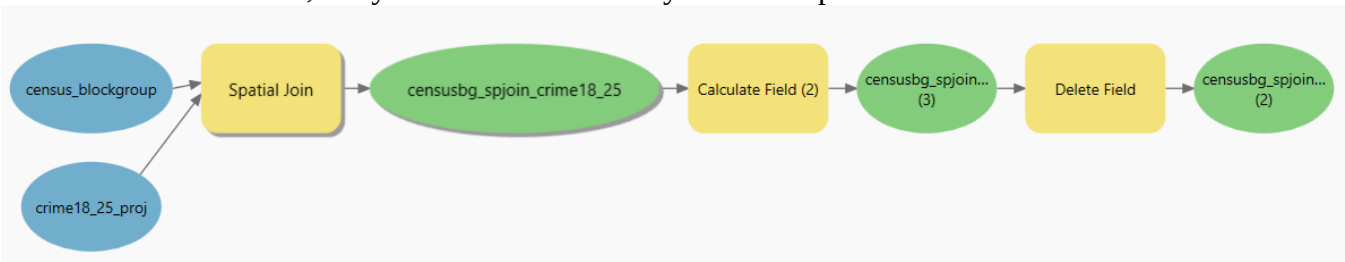
Question 3 (2.5 points)

Open the **first three** models in the Toolbox that are in the SFcrime geodatabase in Modelbuilder (right click and edit). Examine them and try to understand what they would do.

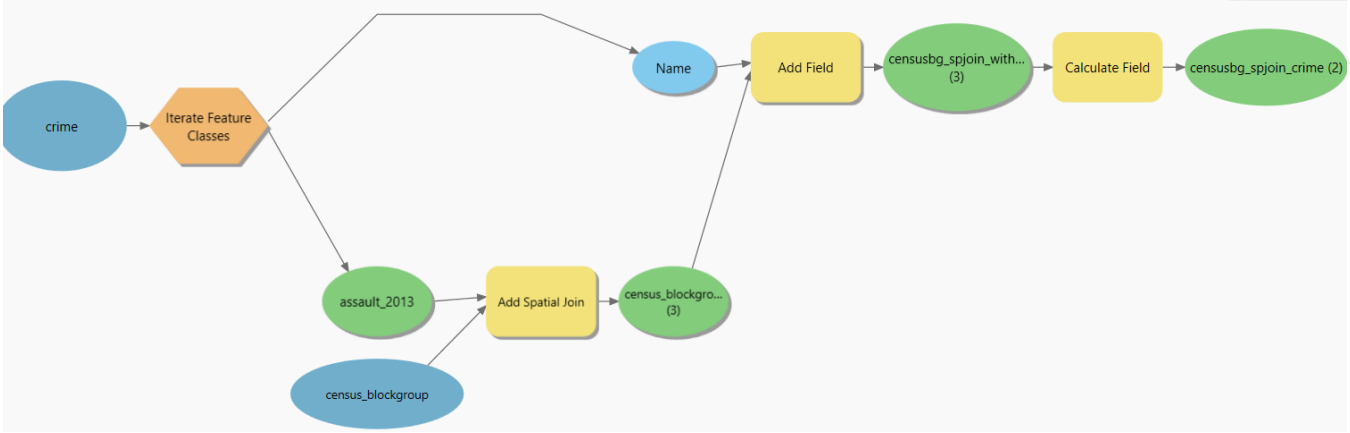


- What are the resulting layers in each model going to look like – point line or poly?
- What attributes will the resulting layers have – (give two or three examples)
- In a sentence or two, explain what each model will do

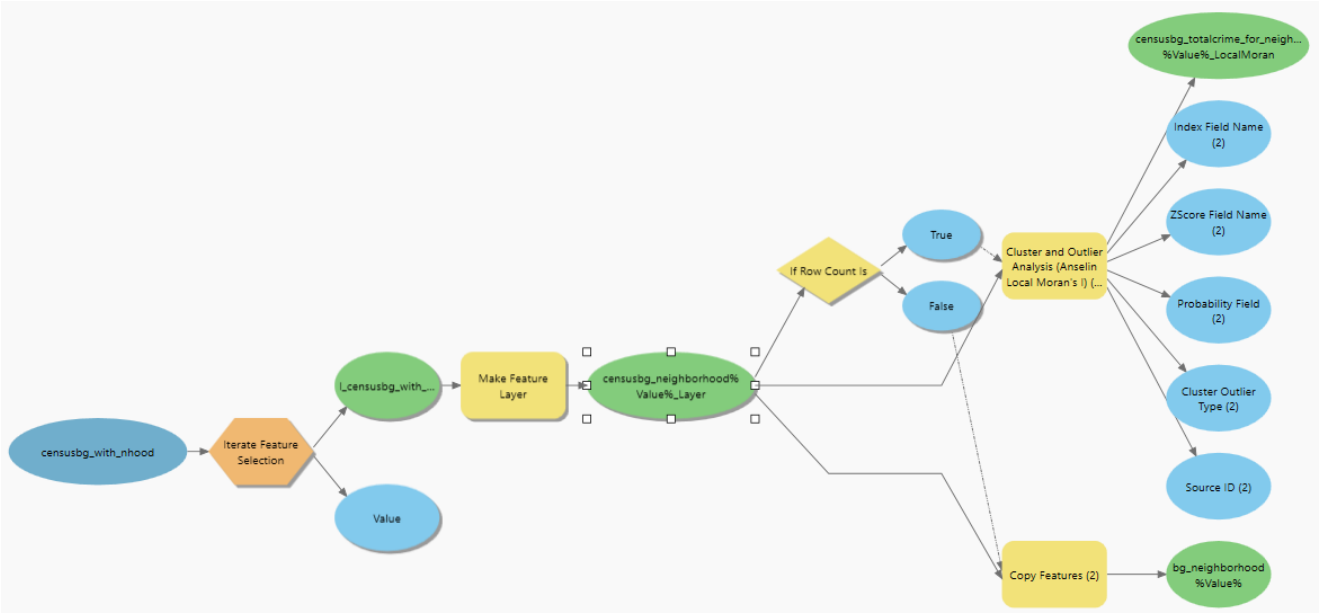
You can run the models, but you will have to modify the folder paths.



Model 1



Model 2

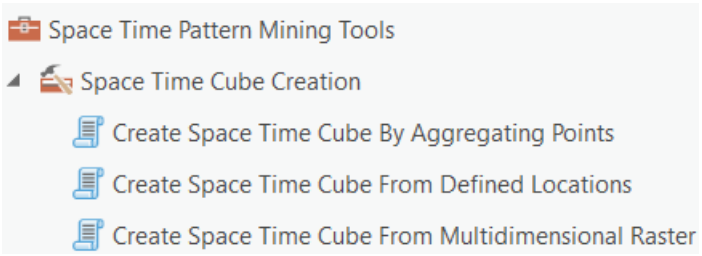


Model 3

2.0 Space-Time Analysis

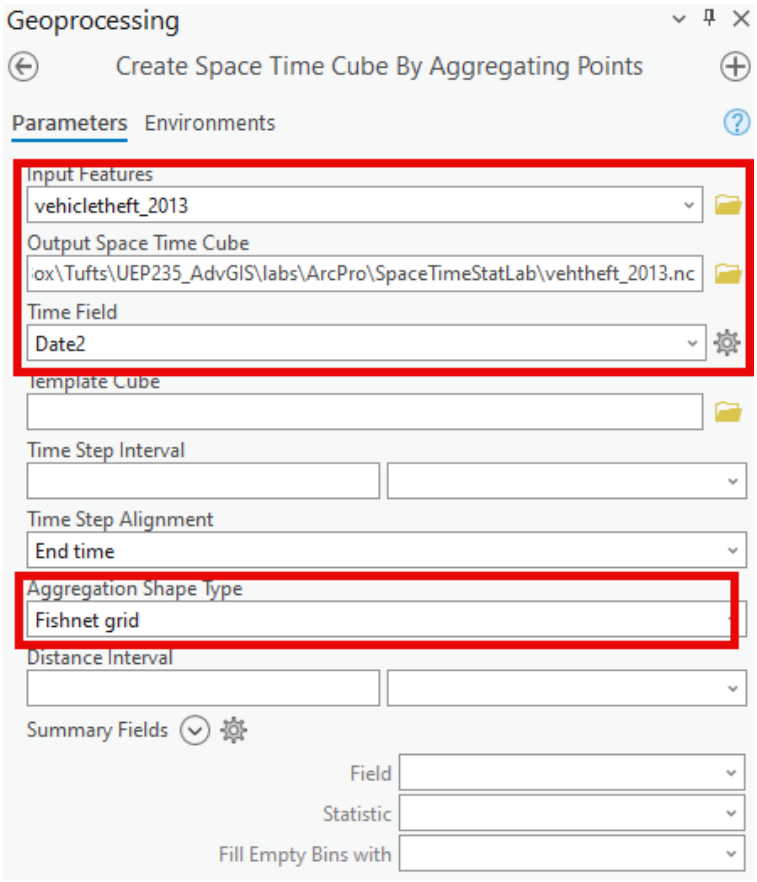
Data:

- Feature classes for crime points in the **results** feature data set in the geodatabase **SFCrime.gdb** from the City of San Francisco: these point layers have all the crimes in San Francisco for that year: T2018, T2019, T2020, T2021, T2022, T2023, T2024 or T2025. You could also use the crime points in the **crimes** feature dataset for 2013, 2017, 2021 and 2024 as you did in the earlier sections**. Doing it for <u>all crimes</u> will take longer to run than doing it for a subset of crimes!**



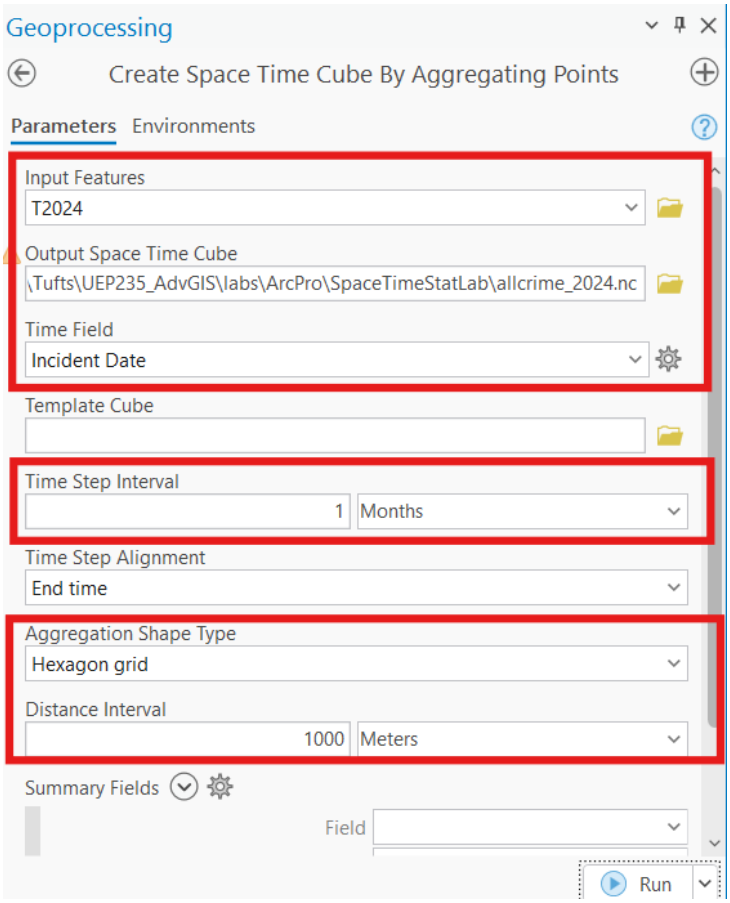
Now you will include **<u>time</u>** in trying to understand the clustering of these crimes.

15. Either expand the toolbox for Space Time Pattern Mining in Toolbox or search for **create space time cube** in the Tools.
16. Double click on **Create Space Time Cube by Aggregating Points**. Load the input features – one of the crime layers which are points. These are point level data with an **Incident Date (for crimes 2018 or later)** or **Date2 field (for 2013 and 2017)**. Click on Run to save the **netcdf** data layer (a multidimensional data format) to your own folder. See screenshots that follow. If you do not specify the time step interval, ArcGIS will calculate one that makes sure that you don't have empty space time bins. Note that the netcdf cannot be saved in the gdb – it should be saved to your folder.



Creating a netcdf for default time and distance interval for 2013 using a Fishnet grid aggregation

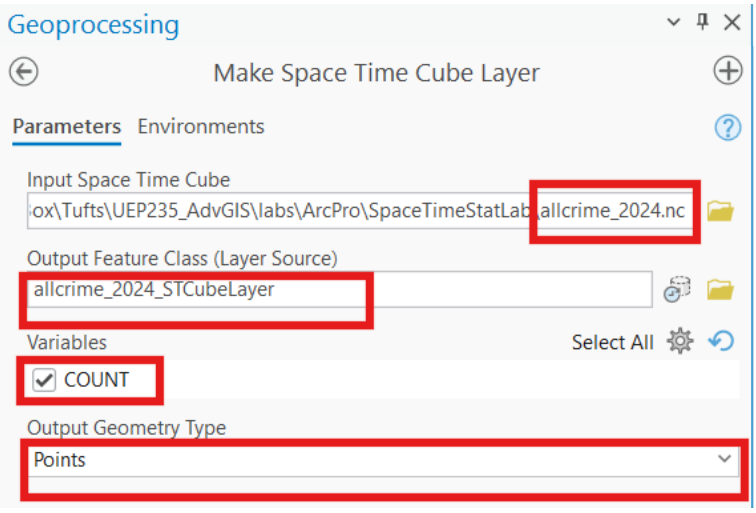
Note that the messages show the default that ArcGIS has selected for you. If you don't like what you see, you can change it to some other time step and/ or distance.



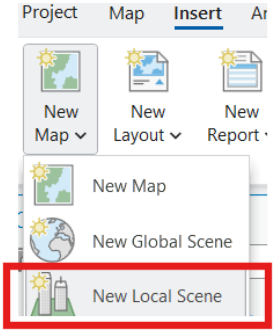
Creating a netcdf in 2024 for all crimes at 1000 meter distance interval, 1 month time step interval, and a Hexagon grid aggregation

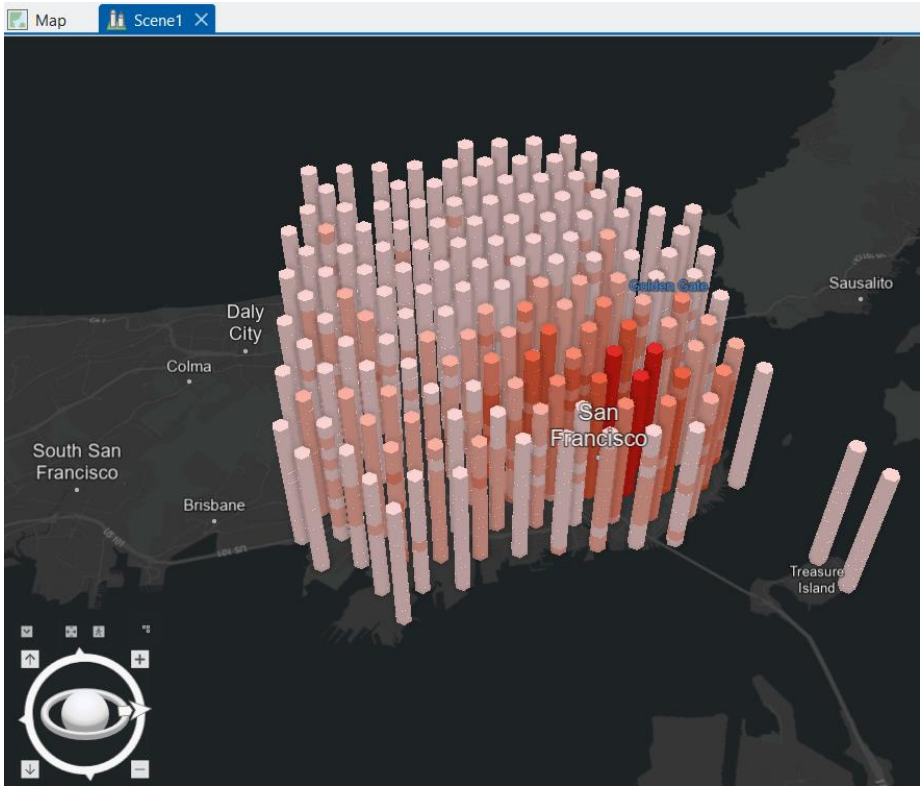
Also note that the netcdf **cannot be saved in the gdb** but should be stored in the folder for this exercise.

17. Search for Make Space Time Cube in 3D in the Geoprocessing Tools
18. In the window that follows include the netcdf and other details as shown in the screenshot that follows



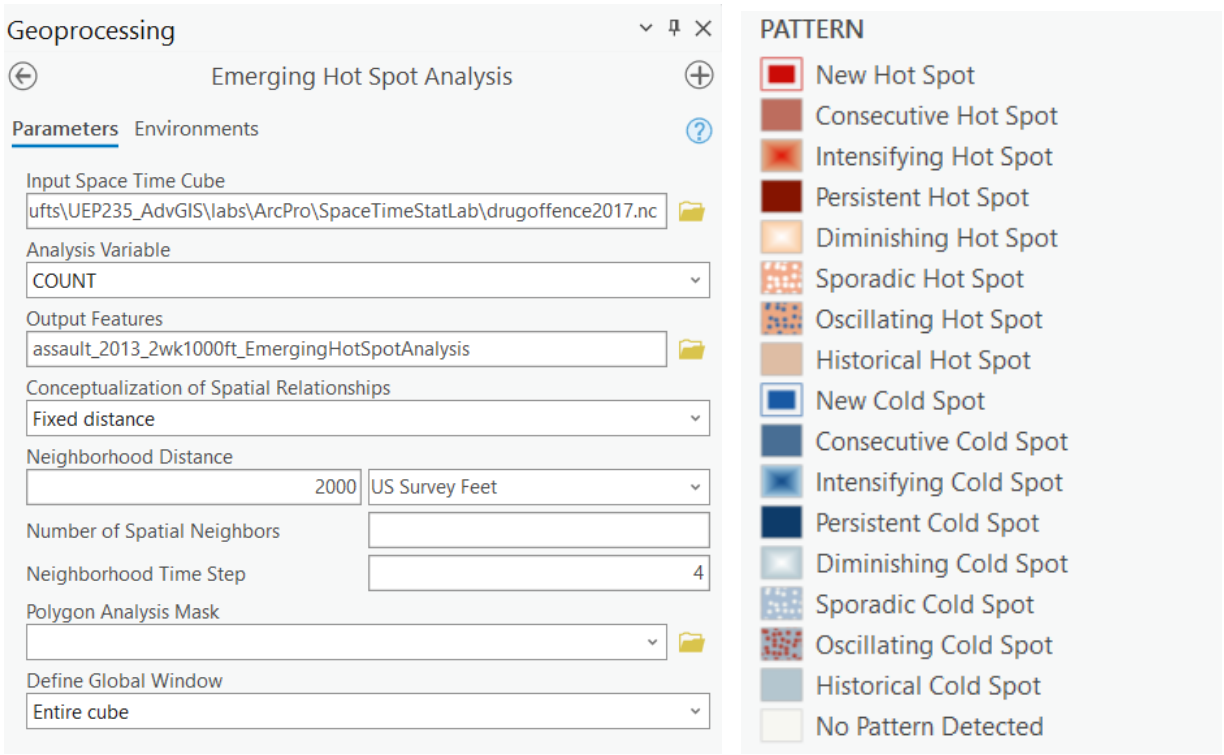
19. In the **Insert** pane, select new Map – **New Local Scene**.
Copy and paste the layer from the previous step (which will be in your gdb and in your Map) in this scene. Zoom to San Francisco and , you should see the counts of the crime in 3D.





Next, you will look for emerging hotspots.

20. Search for the "Emerging Hot Spot Analysis" tool. See screenshot below for an example which uses a neighborhood distance of 2000 feet. You could experiment with different neighborhood time steps (I chose 4 weeks but these could persist over time so you could experiment with higher or lower time intervals). The resulting map should be interpreted based on the pattern legend that comes with it. The categories definitions follow the screenshot.



Category Definitions

Last time step is hot:
- New: the most recent time step interval is hot for the first time
- Consecutive: a single uninterrupted run of hot time step intervals, comprised of less than 90% of all intervals
- Intensifying: at least 90% of the time step intervals are hot, and becoming hotter over time
- Persistent: at least 90% of the time step intervals are hot, with no trend up or down
- Diminishing: at least 90% of the time step intervals are hot, and becoming less hot over time
- Sporadic: some of the time step intervals are hot
- Oscillating: some of the time step intervals are hot, some are cold

Last time step is not hot:
- Historical: at least 90% of the time step intervals are hot, but the most recent time step interval is not

Last time step is cold:
- New: the most recent time step interval is cold for the first time
- Consecutive: a single uninterrupted run of cold time step intervals, comprised of less than 90% of all
- Intensifying: at least 90% of the time step intervals are cold, and becoming colder over time
- Persistent: at least 90% of the time step intervals are cold, with no trend up or down
- Diminishing: at least 90% of the time step intervals are cold, and becoming less cold over time intervals
- Sporadic: some of the time step intervals are cold
- Oscillating: some of the time step intervals are cold, some are hot

Last time step is not cold:
- Historical: at least 90% of the time step intervals are cold, but the most recent time step interval is not

21. If you expand the messages in the **History** for emerging hotspots tool or you click on

View Details, you will also get definitions of the categories and summaries as shown in the screenshot. You may have to scroll down the messages to see this.

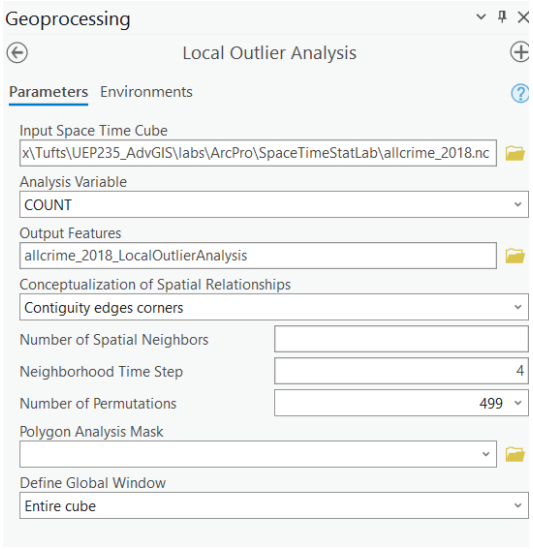
Summary of Results

	HOT	COLD
New	0	0
Consecutive	0	0
Intensifying	1	0
Persistent	7	0
Diminishing	0	0
Sporadic	88	0
Oscillating	0	0
Historical	0	0

All locations with hot or cold spot trends: 96 of 809

22. You should also try the local outlier analysis. See screenshot that follows. The results are interpreted just like Local Moran's I (HH is High surrounded by High crimes and LL is low surrounded by low number of crimes in the map. Looking at the 

View Details in the History will get you the number of clusters and outliers.



Location Summary

Category	# of locations	% of locations
Never Significant	309	10.01
Only high-high cluster	31	1.00
Only low-high outlier	85	2.75
Only low-low cluster	214	6.93
Only high-low outlier	129	4.18
Multiple Types	2320	75.13

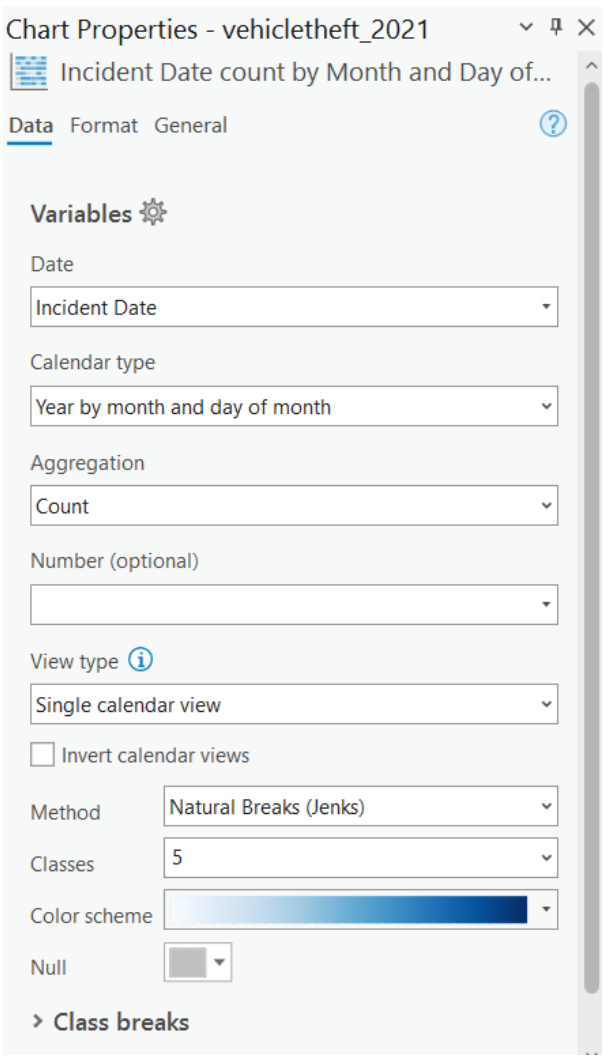
Entire Cube

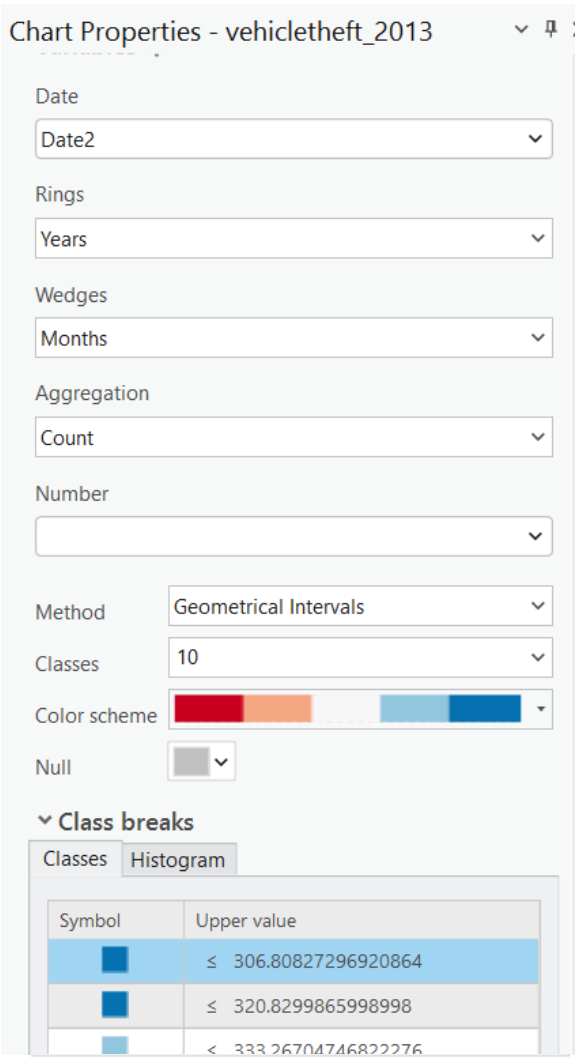
All locations with outliers:		2534 of 3088
Category	**# of bins**	**% of bins**
High-Low outlier	8180	2.94
Low-High outlier	12022	4.33
High-High cluster	20366	7.33
Low-Low cluster	63605	22.89
Not significant	173747	62.52

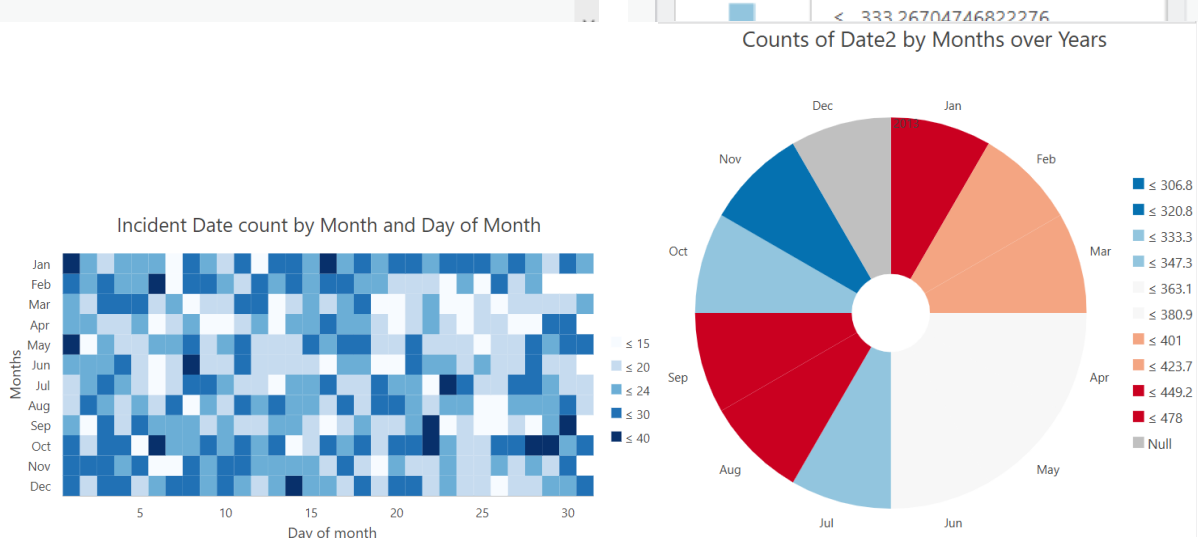
2.1 Space-Time Charts

Since the crime data have dates associated with them, it's possible to create Charts and animate them over time.

23. Insert a new map in your project and add in the crime point feature class you have been looking at, for example, assaults in 2021.
24. Right click on the layer in the Contents and select **Create charts**. Try out the **Calendar heat chart** and the **Data clock chart** for example. Are there any interesting patterns that you notice? Are there certain days of the week or months with more crime?





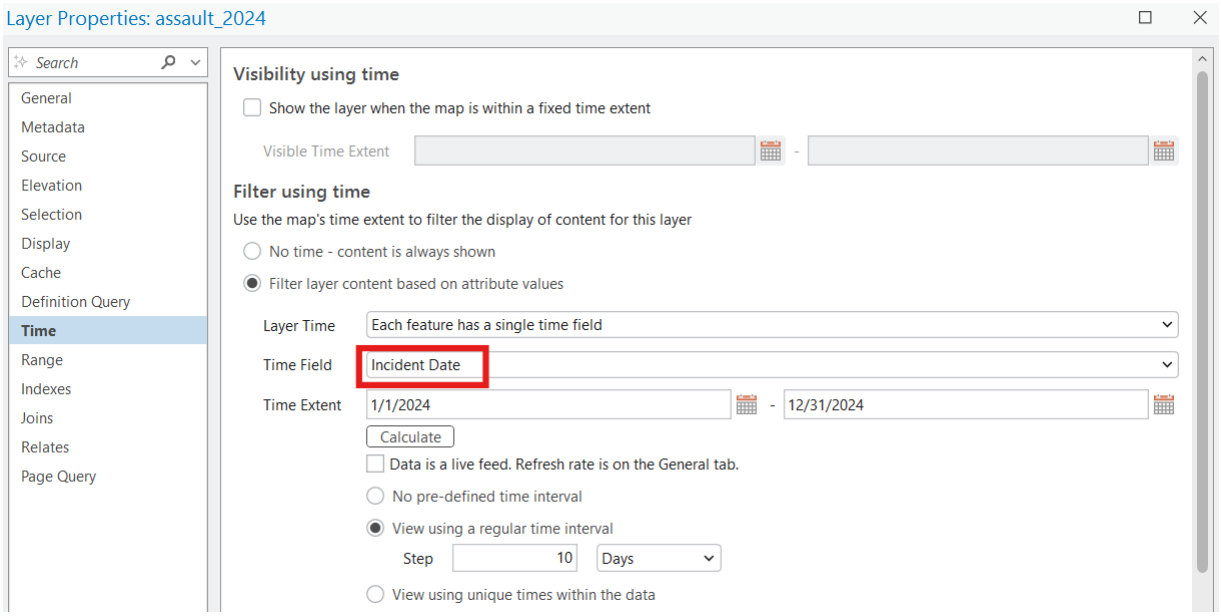


2.2 Space-Time Animation (optional)

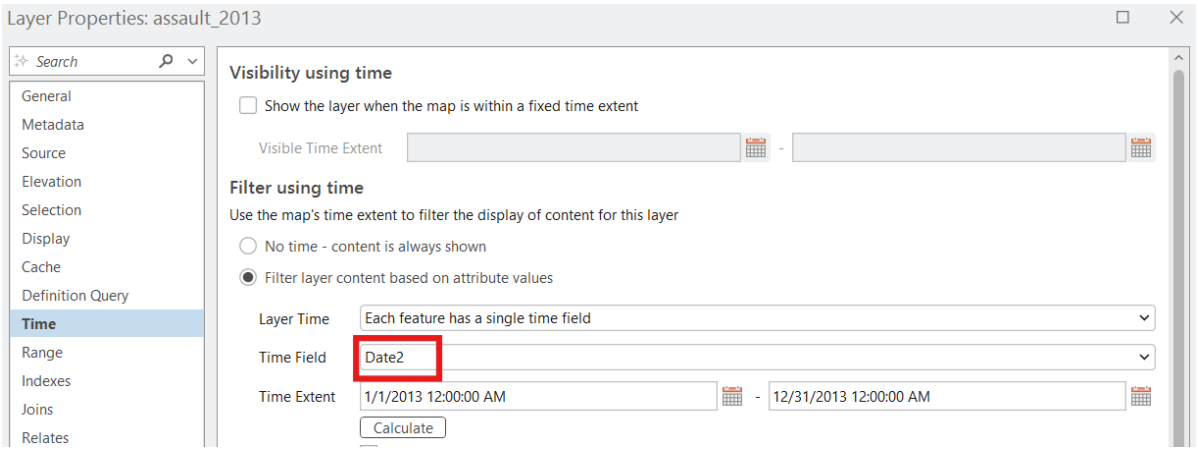
Next, activate the Time variable in the layer to create an animation:

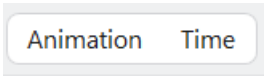
25. Right click on a layer which has a Date Field (for example the crime points) in the Contents and select properties.

26. In the new window click on Time to see the tab for it
27. Change the properties of the Layer time from “No time” to “Filter layer content based on attribute values” as shown in the screenshot. Click on OK.

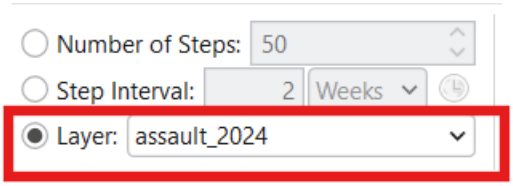


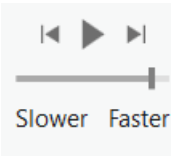
If you chose data for 2013 or 2017 the attribute for the Time field will be different:

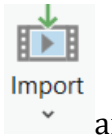


28. Notice that you get a new tab once you enable time on your layer 

. Click on the **Time** tab to see the menu associated with it in the main menu
29. You can change the step interval to 2 weeks (or more) or change the number of steps but make sure that the layer is the one you want to animate



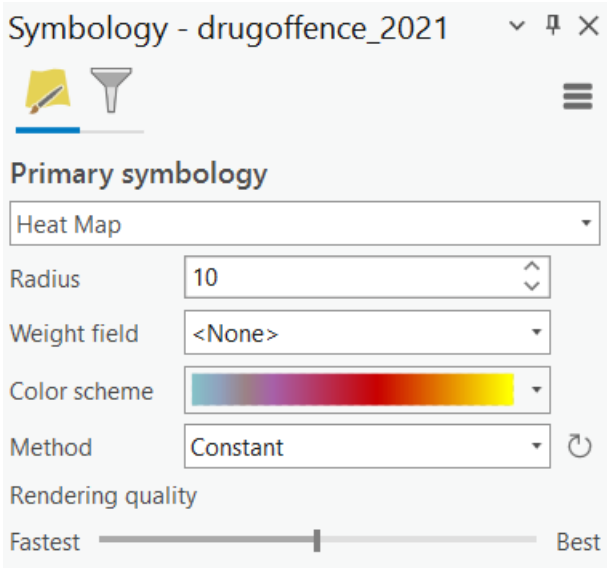
30. In the **Time** tab click on  in the Playback pane (you can decrease or increase the speed) to see the crimes over time 

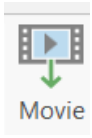
31. Click on **Import** 

and then **Time Slider Steps** 

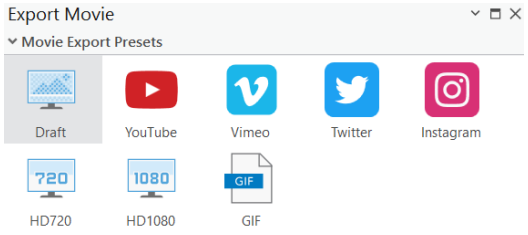
in the **Animation** tab

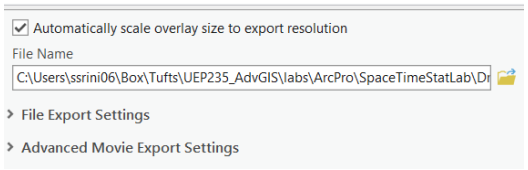
32. Right click in the **Contents** on the layer for the crimes that you want to animate and select **symbology**.
33. In the new window change it to Heat Map as shown in the screenshot. Notice the radius as well as color scheme can also be changed from the default.



34. Click on  to Play the animation in the Animation pane.
35. When the animation finishes you can export it 

as a movie to your folder or as a Draft (if you want a high-quality movie you can change this). (This could take a while)





Question 4 (2 points)

Tell a short story about a crime of your choice in San Francisco using the data from the gdb based on the space time visualizations that you created in ArcGIS for this section. Make sure to reference neighborhood names to make this a story that a resident could relate to. You can be creative! Attach screenshots of the space time maps showing hot spots and visuals like graphs you are using to write this story.

Samples from previous years:

*"The neighborhoods of xx, xx, and xx in xx San Francisco were always plagued with assault. To Sonam, a resident of neighboring xx, this was hardly a problem because xx, thought Sonam arrogantly, had historically been free of assault cases, relative to its surrounding communities. Today, however, interaction amongst neighborhoods increased and characteristics diffused to their bordering towns, leading an already apprehensive Sonam to the brink of anxiety as he feared for his town and community. As an immigrant, Sonam was aware that his xenophobic attitudes towards his neighboring communities was aggressively contradictory to his very existence in the US. However, he knew that the xx San Francisco community in xx had very recently become a hotspot of assault cases, a fact that haunted Sonam, especially because he was aware of the community's history of nonviolence. This paradox of being an immigrant yet being fearful of outside communities was a cause of great concern for Sonam. He felt that his views were starkly different than his peers', who tended to welcome and celebrate diversity. At the same time, he knew that people who echoed his views would likely not support him as he was the very definition of the thing they were so afraid of. Sonam spent much of his teenage and adult years wrestling with this paradox; much of his concern was rooted in the fact that he could not reconcile his existence with his beliefs—he always considered himself a "good" person but knew that his beliefs were often regarded as "regressive" and "bad" by popular media. How could one, thought Sonam, be a good person and have "bad" beliefs? But when he touched the ripe age of 40, he acquired a clarity that had been missing for the first four decades of his life. Sonam realized that one could not be a "good" person while lauding restrictive and regressive policies that had a real impact on people's lives. In the end, Sonam realized, he was, in fact, just an *****."*

By Namgay P. Tshering, 2020

"Sara first moved to San Francisco in 2010 as the new wife to a wealthy, long-time San Francisan. They were living in the Sea Cliff neighborhood known for its enormous homes and gorgeous ocean views, in his family's generationally owned home. Particularly important to Sara was the neighborhood's persistently low crime rate, a known consecutive cold spot, perfect for her dreams of raising a family. All was going well for Sara, until it wasn't. She was not able to get pregnant and she lost her job in the beginning of 2013. Her wealth meant she didn't have to work but now felt like she had no purpose in life. So, to combat her depression, she found herself walking and walking and walking. One day in early June, she found herself wandering through xx, completely lost in thought. She was generally heading towards the San Francisco Museum of Modern Art on 151 3rd St but was so consumed by her thoughts that she didn't realize she had walked into a persistent and intensifying hot spot area. As she looked up from the ground and had this realization, she saw an assault of a woman just a block ahead of her. She screamed and the perpetrator vanished, leaving the woman on the street. She went to help the woman on the street and found out from the police officers that arrived at the scene of the crime that it was the peak of assault-related crimes during the year, in that hot spot. With this information, Sara went on to help more and more women of assault through donations, volunteering, running care centers and more.
She found her new life's purpose."
By Eliza Jobin-Davis, 2024

(Sumeeta changed neighborhood names to xx to protect the privacy of the neighborhoods)