GEOM90008 Foundations of Spatial Information

Prof. Dr.-Ing. habil. Stephan Winter

Department of Infrastructure Engineering
The University of Melbourne
Victoria 3010, Australia

email winter@unimelb.edu.au
url http://www.ie.unimelb.edu.au/

THE UNIVERSITY OF MELBOURNE

ASSIGNMENT 2, MODULE 4: POINT ANALYSIS: HOT SPOT MAPPING AND INTERPOLATION

1. OBJECTIVE

The aim of this module is to obtain an understanding of

- Hot spot mapping: Spatial analysis on point features in terms of finding patterns and trends
- Interpolation: Generating a surface map using few control points

2. INTRODUCTION

Over the last few years with the growing number of cars and other forms of transport modes, the number of road accidents has increased. This becomes a major issue for roadway and traffic safety. To understand the patterns and trends of road accidents, authorities (e.g., VicRoads and RACV) are in a process to analyze the incident data at different resolution with the different severity level. In this context, it is essential to find the most sensitive accident-prone zones and to find the accident patterns.

Thus, in this exercise, you will use a sample of roadway accident data to find the pattern and delineate the most sensitive zones using a GIS-based hotspot analysis. You will also learn how to generate a continuous surface map that shows a variation in accident magnitude across Melbourne CBD and its surroundings.

3. DATA REQUIRED

In this exercise, you will mainly use road crash data over the last five years supplied by VicRoads, which is the state government agency tasked with managing roadway traffic and public transport infrastructure. You will require:

- zipped data from this directory (hot spot accident data folder), which contains a study area (“mask.shp”), crash data (“Crashes_Last_Five_Years_Melbourne.shp”), and administrative blocks (“SA1_2011_AUST.shp”).

3.1 Preparing the dataset

Open ArcGIS Pro. From the Catalog pane, right-click on *Folders*, and choose *Add Folder Connection*. In the opened dialogue box, browse the respective folder where you saved your data set and then click *OK*. In this way, you get access to all the layers from the Catalog pane, make it easier to manage your data and their corresponding metadata. Finally, add all the layers to your table of contents.

Step 1: Retrieve selected information

Now go to the “Crashes_Last_Five_Years_Melbourne.shp” and open the attribute table (right-click on the layer → Attribute Table). Skim through the table to get an idea about the different types of information stored in the table including the severity of an accident, its location, time, number of affected passengers or people, or license status. For this module, we are interested only in high-risk accidents. Hence, we will extract only those crash data points with ‘severity *fatal’* or ‘*serious injury’*.

To do that, we will run a simple query lookup. To do this, from the ribbon, go to the *Table’s View* tab, and choose *Select By Attributes.* In the corresponding pane, choose the appropriate layer (Crashes_Last_Five_Years_Melbourne) as the input rows. Click on the new expression to select a subset of record.

In the left drop-down box, choose SEVERITY as the queried field. The relationship should be ‘is equal to’. For the last box choose 'Fatal accident'. Next, add another clause and choose ‘Or’ as the connecting condition. This time choose 'Serious injury accident' as the required value for the last box. This query can also be written in SQL format. To do so, simply turn on the SQL switch button. Here, the query can be expressed with the following statement "SEVERITY" = 'Fatal accident' OR "SEVERITY" = 'Serious injury accident'. Use the verify tool to validate your SQL expression.

Click *Run*. Then go to the attribute table and click on show selected records. Export the selected records as a shapefile (name it “fatal_serious.shp”).

Tip on how to convert exported selected records to .shp file.

First export the selected records as a table. To do so, simply right-click on the layer with selected records that contains information related to either fatal accidents or serious injury and from the *Data* menu, choose *Export Table*. This will add a new tabular layer to your table of contents.

Next, use *Tools > Data Management Tools > XY Table to Point tool* to convert the table to shapefile namely “fatal_serious.shp”. Select proper X and Y as Longitude and Latitude and choose *Run*. Add the shapefile to your map. Finally, remove Crashes_Last_Five_Years_Melbourne and the tabular layer of events.

Note that some of the layers that you have added to your map are in WGS84 coordinate system. To make sure which layers are not projected, go to the layer names in the table of content. Right-click on the layer name, click *Properties*. Next, open the *Source* tab and from the *Spatial Reference* drop-down see the corresponding geographic coordinate system of each layer.

Step 2: Projection

In order to carry out a hotspot mapping and interpolation operation, you should project all the layers in the same coordinate system. To perform the projection operation, go to *Tools* and search for *Project (Data Management)*.

In the popped-up pane, use one of your layer shapefiles as input data set. Choose an appropriate name and folder to keep your output feature class. To select the output coordinate system, click on the globe button and search for GDA 1994 MGA Zone 55. After finding the coordinate system, click *OK*. Finally, *Run* the project tool. This will create a new projected feature class. Repeat the same operation for the remaining layers (which are not yet projected).

Tip: To run the same operation, again and again, you can perform a batch processing in model builder.

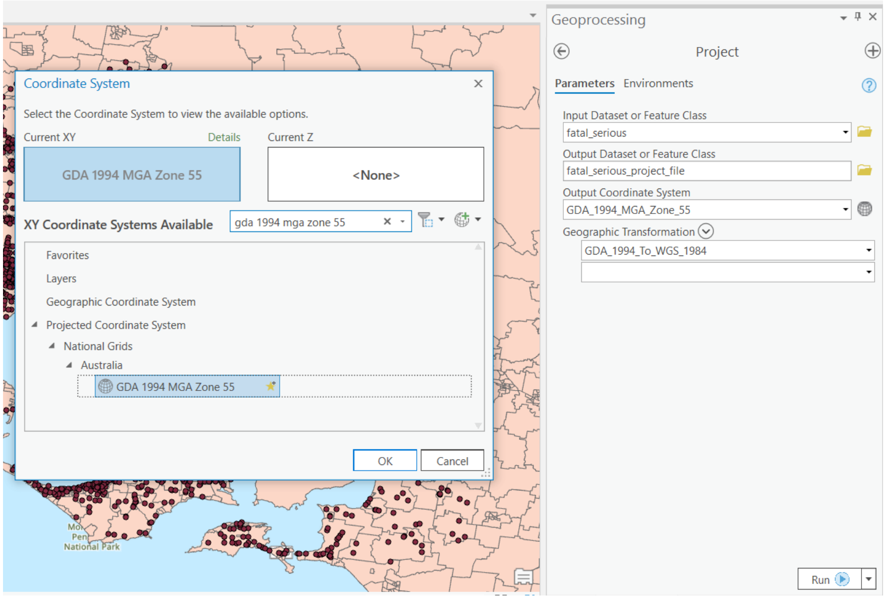

Figure 1 - The projection operation

Step 3: Clipping

Next, we will select our study area. Since the extent of the crash data set is quite large and we are interested only in the CBD and its surroundings we will narrow down our focus. This is already done in the form of the “mask.shp” layer, which specifies the extent of your study area for this module; however, you are encouraged to do this same task at home on a larger extent and see the variation in your results. Use this layer to **clip** all the data sets *(input: “fatal_serious.shp”, input feature: “mask”, output: “fatal_serious_clp.shp”).*

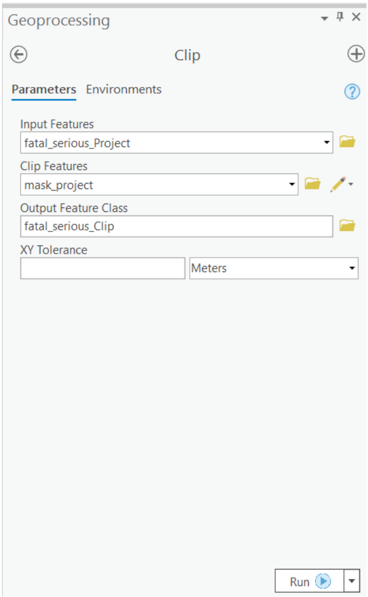

Figure 2 - Clip tool

3.2 Aggregate your incident data

Now you are ready to perform a hotspot analysis. The hot spot analysis tool assesses whether high or low values in any data set (e.g., the number of crimes, number of accidents, accident severity, dollars spent on sporting goods, or number of medical care per household) cluster spatially. The field containing those values is your analysis field. For point incident data, however, you may be more interested in assessing incident intensity than in analyzing the spatial clustering of any particular value associated with the incidents. In that case, you will need to aggregate your incident data before the analysis. In this module, we will aggregate the accident data within a given spatial constraint of 40 m especially in the urban canyons, assuming the location of the accidents reported involves 40 m spatial uncertainty captured by GPS devices.

In order to aggregate some of your crash data points, we will use two functions called **Integrate** and **Collect Events**.

Note: The Integrate tool modifies the input dataset by changing the locations of the input features. To preserve the original data, it is important to make a copy of the original input data before proceeding.

Step 4: Copying the data set

Search for the 'Copy Features' tool and choose Copy Features (Data Management) to open the tool. Use "fatal_serious_clp.shp" as the input features, and create an output feature class as "fatal_serious_cpy.shp". Click OK. This will now create a duplication of the original "fatal_serious_clp.shp" layer with a new name as "fatal_serious_cpy.shp", which will appear in your left-hand pane, the Table of Contents.

Step 5: Aggregating the incident points with sample counts

This operation consists of two steps, running an *Integrate* operation followed by running *Collect Events*.

Step 5.1: Run Integrate

Search for the 'Integrate' in tools and choose the *Integrate (Data Management)* to open the tool's pane. Next, use fatal_serious_cpy as the input feature. Type **40** in XY Tolerance field. Don't forget to select the proper unit (meter). This will snap all the incident points within 40 m of a queried incident point with a varied weightage.

Step 5.2: Run Collect Events

Search for the 'Collect Events' tool and choose the *Collect Events (Spatial Statistics)* to open the tool. The input is "fatal_serious_cpy.shp", and the output should be the weighted point feature class: fatal_serious_agg. This will create a new feature class named as fatal_serious_agg with aggregated event written in ICOUNT field. Open the attribute table and have a look at how the incident points are snapped together and how ICOUNT varies in your table.

Tip: There are some other ways to aggregate your data.

Tip: If you want to aggregate a point incident data on a polygon feature class you may use Spatial Join operation or Create Fishnet tool followed by Spatial Join.

3.3 Hot spot analysis

Now you are ready to perform a hot spot analysis on the accident data. There are several ways to perform a hot spot analysis. In this module you will use the Getis-Ord Gi* spatial statistics measure to perform the hot spot mapping. You will learn more on spatial statistics in advanced subjects; for now only this much: a statistically significant point (or region) would show either a high (positive) *Z* score with low *p*-value (a so-called “hot” spot) or low (negative) *Z* score with low *p*-value (a “cold” spot). The observation of a hot or cold spot rejects the null hypothesis of random distributions and supports the assumption of a spatial pattern across the study area. On the other hand, a *Z* value close to zero indicates that the distribution is a result of random chance.

In hot spot analysis, each point is evaluated with respect to its neighbour points. Thus, the aspect of the neighborhood is critical for an effective hot spot mapping. This is expressed as a **conceptualization of spatial relationship**. In this module, we will use a Fixed_Distance_Band (default option) as the conceptualization of a spatial relationship, which maintains the same scale of the neighborhood for all the points in your study area. To select the appropriate neighbourhood distance (scale of the particular hot spot analysis), we will perform an optimization operation.

Step 6: Incremental spatial autocorrelation

From tools, search for *Incremental Spatial Autocorrelation*. Use the following information to fill the dialogue box.

- Input feature class: fatal_serious_agg
- Input field: ICOUNT
- Number of distance bands: 10
- Beginning distance: 1000
- Distance increment: 1000
- Output report file: HSDreport.pdf (you may choose any other names)
- Leave other options as they are.
- Click *Run*

The analysis may take a while. Then go to the respective folder where you saved the HSDreport.pdf. Open the pdf to see how *Z*-score changes with varying the neighbourhood distance. You will select the distance corresponding to the *Z* peak.

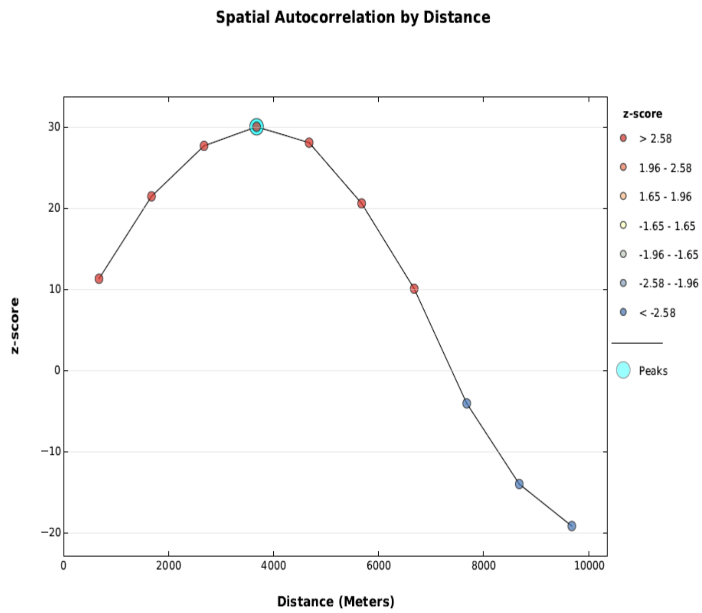

Figure 3 – Line graph summarizing spatial autocorrelation

The peak in this dataset occurs somewhere at 3800 m. Thus the fixed band distance value would be set as 3800 (m).

Step 7: Hot spot analysis (Getis-Ord Gi*)

Search for hot spot analysis and select the Getis-Ord Gi* tool and open it. Use the following information to fill the hot spot dialogue box.

Input feature class	Fatal_serious_agg
Input field	ICOUNT
Output feature class	..\fatal_serious_hotspot
Conceptualization of spatial relationship	Fixed distance band
Distance method	Euclidean distance
Standardization	Default
Distance band or threshold distance	3800

Leave other options as they are and click *Run*.

You can now see the hot spot map plotting the accident data points with three different colours (reddish: hot spot with different confidence levels, bluish: cold spot with different confidence levels, grey dots: not significant).

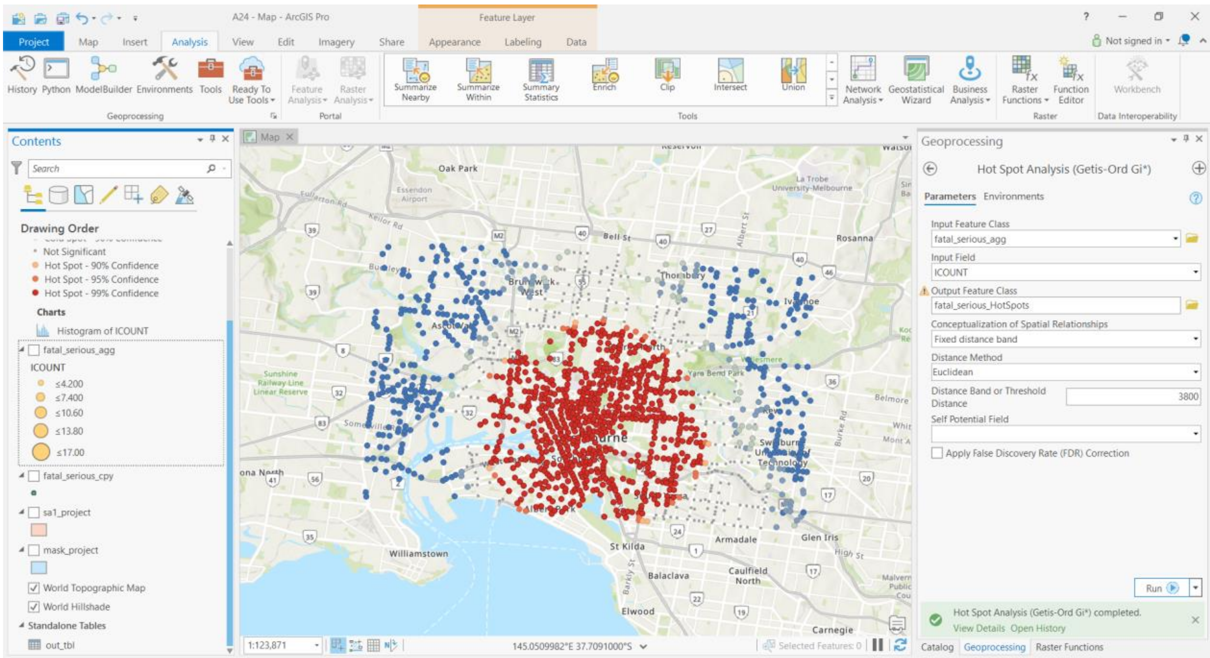

Figure 4 - Hot spot map plotting the accident data points

Step 8: Adding the statistical area layer

We aim to see which statistical areas are more sensitive and bear statistically significant accident events. You can now add the SA map. The default is SA1 which is the finest SA block size defined by the Australian Bureau of Statistics. We will change the scale of the SA blocks and convert it to SA2. For this step search for the **Dissolve** tool.

Select SA1_2011_AUST (projected) and SA2_Name11 as the field based on which dissolve operation will run. Save the output as ..\SA2_2011_AUST.

Add SA2_2011_AUST layer under the hot spot map. Right-click on SA2_2011_AUST and choose *Properties*. From the *Display* tab, make sure the display field is set to SA2_NAME11. Also, tick *Show MapTips* and click *OK*. Now deselect all the layers except the hot spot layer and SA2_2011_AUST. Move your cursor over the study area and you can examine the names of different admin blocks (SA2) with different hot spot intensity. You can also add a route network on your ArcGIS Pro and examine which routes are more sensitive.

Step 9: Visualization and generating a continuous surface map

In some cases, you may need to generate a continuous surface map to see how an event (or attribute value) is varying across your study area. There are also cases where you have to estimate an attribute value at a given location using some known points. In such situations, we perform an *interpolation* operation. There are several ways to implement an interpolation operation such as inverse distance weight (IDW), Kriging, trend surface fitting, or nearest neighbour – you will learn these interpolation methods in detail in advanced subjects. In this module, we will just use IDW to generate a surface map.

A surface map acts as a heatmap (note: a “heatmap” is different from a “hot spot map”) which is helpful to a broader audience who can visually understand how an event (or attribute) is changing over space. In this module, we are particularly interested to see the variation of statistical significance (thus the sensitivity) across the study area in the form of a surface map.

For this, we will now perform an interpolation operation, namely *IDW (Spatial Analyst Tools)*. Fill the options as follows:

Input point features	fatal_serious_hotspot
Z value field (note here Z value means simply an attribute value. Don’t confuse this with the Z value in a statistical process)	GiZScore Fixed 3800
Output raster	..\idw_fatal_se1
Output cell size	Default
Power	2
Search radius:	Variable
Number of points	12
Maximum distance	Blank
Input barrier polyline features	Blank

Click *OK*.

This analysis may again take a while. Now you can see an interpolated surface map showing the variation of GiZscore (equivalent to *Z* score) across the study area.

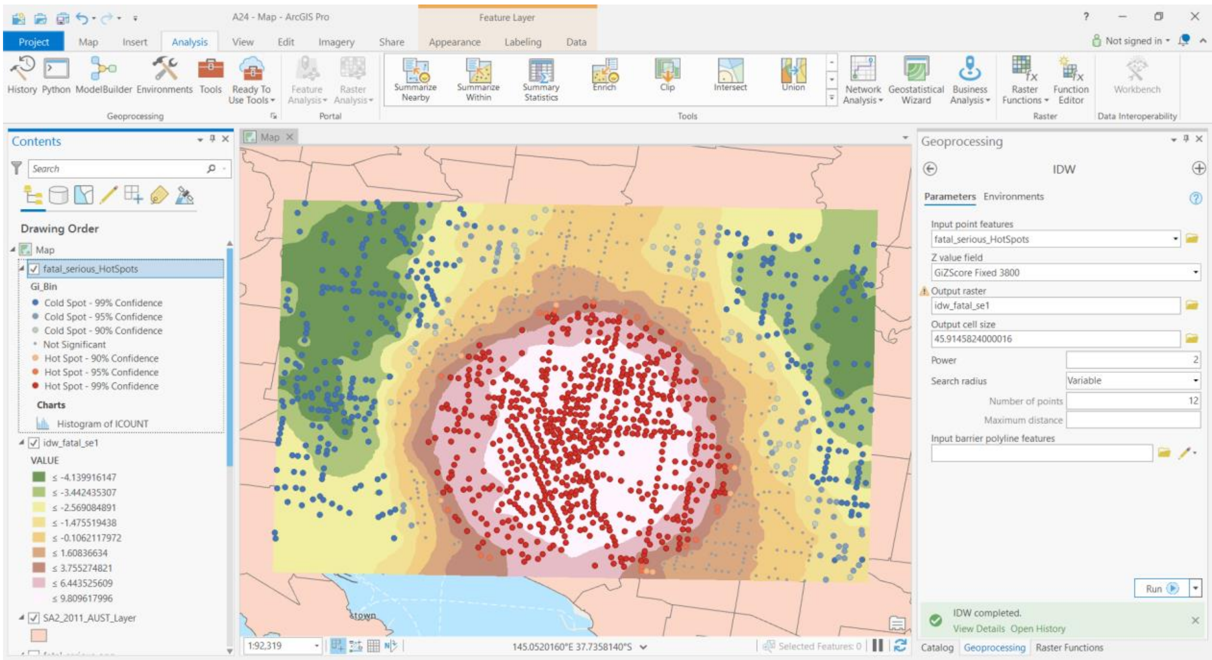

Figure 5 - Interpolated surface map showing the variation of GiZscore

Deselect the hot spot layer to better visualize your interpolated surface map. You can play around with different colour appearance by right-clicking on the interpolated layer (idw_fatal_se1) > Symbology. Choose *Classify* as the primary symbology and try different colour schemes. You can also reclassify your interpolated raster depending on the specific context (preferred visualization or type of problem you want to address). Now select the hot spot layer again along with the interpolated map with a proper transparency level.

Tip: To change the transparency, select the raster layer. Next, on the *Appearance* tab of the ribbon, find *Layer Transparency* and change the slider to increase or decrease the transparency value.

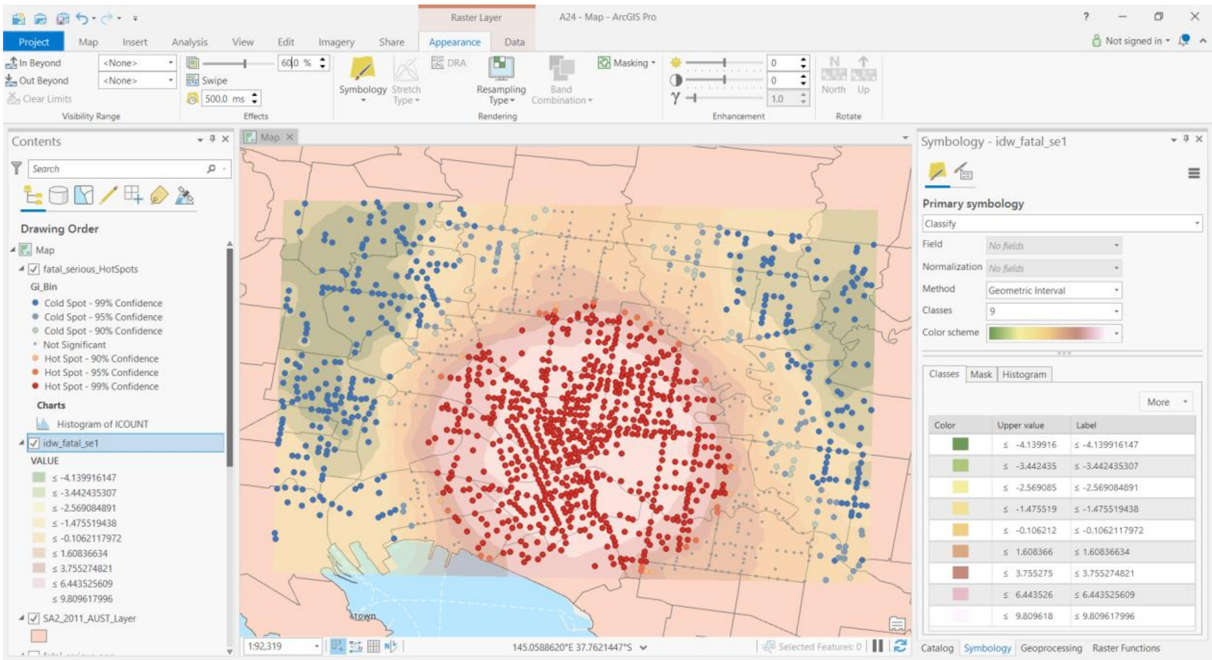

Figure 6 – Transparent raster interpolated layer with its 9 classes.

Further task: If you are further interested you can play around with different data set (say disease outbreak, emergency call, courier dispatch, crime report, purchase information, voting behaviour during an election). You can also investigate how a spatial pattern changes with changing the spatial scale (by creating a new mask with a larger extent). You can also try different other options to aggregate the incident points, for example, spatial join and see the spatial pattern across polygons (instead of points).

4. DELIVERABLE

Submit a document reporting on your findings, containing at least one map and its interpretation.

5. ASSESSMENT

The mini-module is worth 2 marks. To receive 1 mark (a "pass") there has to be a map and a reasonable interpretation, which is the expectation of an average submission. To receive 2 marks a well-designed map and a thorough interpretation beyond class average is expected. There are no fractions.