**Due**

Section 0101: 10AM, March 26, 2021, Friday, ELMS

Section 0101: 12PM, March 26, 2021, Friday, ELMS

Submit a Word document with your answers on ELMS.

**Question 1. (2 points)** Given a point dataset having 5 cities with coordinates Euclidean Space (miles) and populations (10,000) listed in the table:

|  |  |  |  |
| --- | --- | --- | --- |
| **City name** | **X coordinate** | **Y coordinate** | **Population** |
| A | 5.2 | 3.1 | 12.2 |
| B | 10.4 | 2.4 | 6.9 |
| C | 3.4 | 2.8 | 16.7 |
| D | 3.1 | 8.6 | 40.5 |
| E | 1.8 | 12.5 | 31.4 |

Calculate the **mean center**, **central feature**, and **weighted mean center** (in terms of population) of the 5 cities. You can use calculator and/or Excel but not ArcGIS to complete this question.

**N = 5**

**Mean center:**

**X̄ = (5.2 + 10.4 + 3.4 + 3.1 + 1.8) / 5 = 23.9 / 5 = 4.78**

**Ȳ = (3.1 + 2.4 + 2.8 + 8.6 + 12.5) / 5 = 29.4 / 5 = 5.88**

**Total weight = 12.2 + 6.9 + 16.7 + 40.5 + 31.4 = 107.7**

**Weighted mean center:**

**X̄ = (5.2\*12.2 + 10.4\*6.9 + 3.4\*16.7 + 3.1\*40.5 + 1.8\*31.4) / 107.7= 374.05 / 107.7 = 3.47**

**Ȳ = (3.1\*12.2 + 2.4\*6.9 + 2.8\*16.7 + 8.6\*40.5 + 12.5\*31.4) / 107.7 = 891.94 / 107.7 = 7.82**

**Central feature:**

function centralFeature(coordinates) {

    let minDistance = -1;

    let distance = 0;

    let cfCoordinates;

    for (let i = 0; i < coordinates.length; i++) {

        current\_X = coordinates[i][0];

        current\_Y = coordinates[i][1];

        // Calculate the total distance for the current [X, Y]

        for (let j = 0; j < coordinates.length; j++) {

            X = coordinates[j][0];

            Y = coordinates[j][1];

            distance += Math.sqrt(((X - current\_X) \*\* 2) + ((Y - current\_Y) \*\* 2));

        }

        // If minVal is -1 then there has not been an initial comparison yet

        if (minDistance == -1 || distance < minDistance) {

            minDistance = distance;

            cfCoordinates = [current\_X, current\_Y];

        }

        // Reset sum and check the rest of array

        distance = 0;

    }

    return cfCoordinates;

}

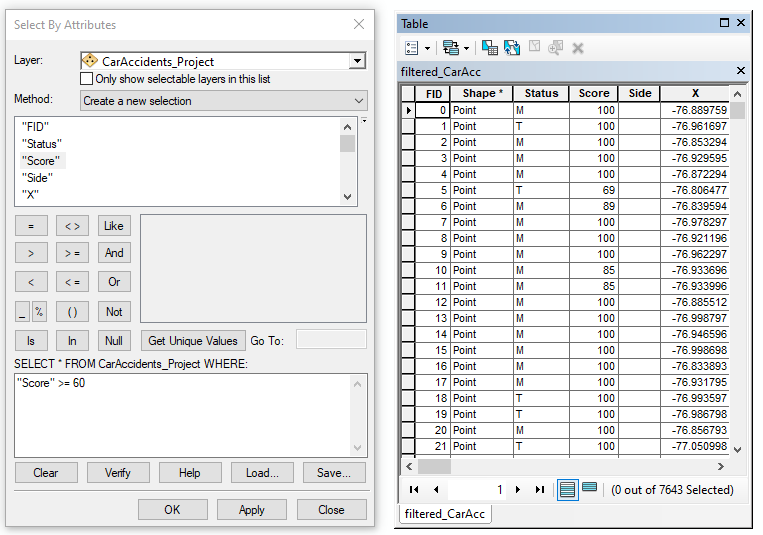
// Written by Carlos Fernandez



**Question 2. (5 points)** For the Quadrat Count Analysis handout (available on ELMS), we determined whether the points follow a Poisson distribution (random pattern). For this homework question, using the car accidents data (provided in the Point Data Analysis lab), ArcGIS and calculator (or Excel) to calculate and determine whether the car accidents follow a perfectly dispersed spatial pattern. Report the z-score and p-value of the pattern.

Hint: a goodness-of-fit test has to be used. On ELMS Homework module, see the slides for the methods together with the test table, be careful to choose a correct method. The procedure is similar with the Handout example. You need to show the steps and details of your calculation, and report the z-score, p-value and the conclusion of the pattern.

1. Car accident features with poorly geocoded coordinates are removed using SQL query. The selected features with valid coordinates were exported into a new shapefile called **filtered\_CarAcc.**



1. In order to determine the proper quadrat size for the car accident data, and are obtained by taking the , , , and (measured in meters). The layer used for this spatial extent is **PrinceGeorgesCounty\_Project.shp.** The **Identify** tool was used to located the following coordinates:

= 7,643

= 392,545.938 m

= 428,580.105 m

= 96,479.279 m

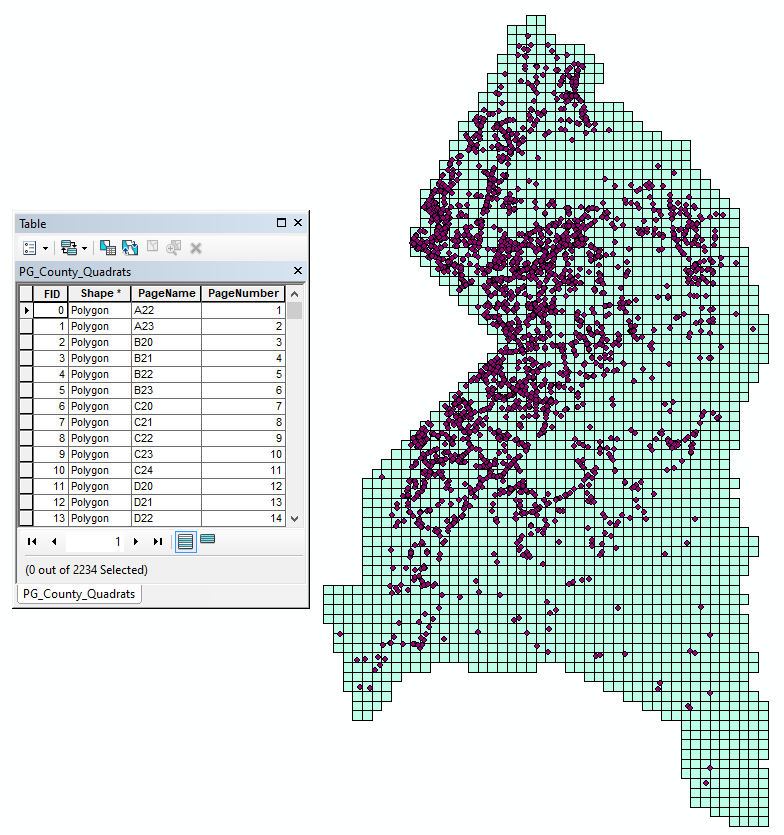
= 162,638.347 m

= 428,580.105 - 392,545.938 = 36034.167 m

= 162,638.347 - 96,479.279 = 66,159.068 m

= 789.8325 m

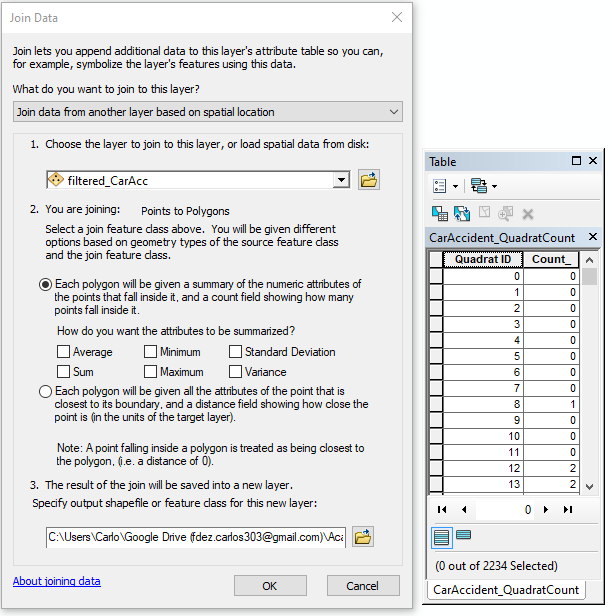
1. Using **ArcToolbox** -> **Cartography Tools** -> **Data Driven Pages** -> **Grid Index Features**, a layer of rectangular cells in the form of polygons was based on the **PrinceGeorgesCounty\_Project.shp** file. Each cell width and height parameter are 789.8 m.



According to the attribute table for the grid index layer the number of quadrats observed are 2234.

1. Now, the number of average number of points in a quadrat for a perfectly dispersed spatial pattern can be calculated:

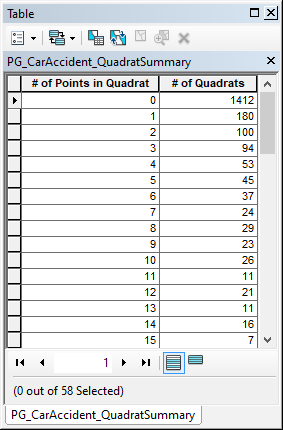
1. A spatial join will now be performed on **PG\_County\_Quadrats.shp** and **filtered\_CarAcc.shp** in order to obtain the number of accidents for each individual polygon (quadrat).

****

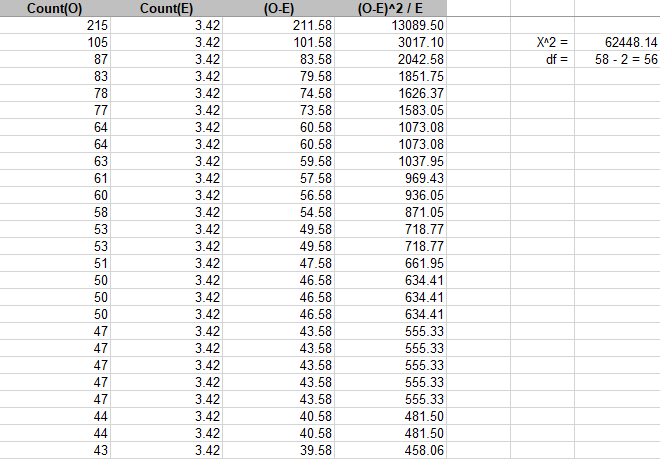
1. The **PG\_CarAccident\_QuadratSummary** table was exported to Microsoft Excel in order to begin the chi-square test process. Data was exported using **ArcToolbox > Conversion Tools > Excel > Table To Excel.** This reveals 58 categories of quadrat counts giving a df value of 58 - 2 = 56 degrees of freedom.

The car accidents follow a perfectly dispersed spatial pattern defined by the average number of points per quadrat

The car accidents do not follow a perfectly dispersed spatial pattern defined by the average number of points per quadrat.



1. Calculate the chi-square value



This yields an extremely large chi-square value that returns a p-value < 0.001 meaning that we can reject our null hypothesis of the car accident data being perfectly dispersed.

**Question 3. (8 points).** From the point data analyses lectures and labs,we conducted the spatial analyses of car accidents by doing data conversion, geocoding, database query, and point data analysis.

For this question, you need to write a report with at least **FOUR** pages to summarize your work to discuss the spatial patterns of car accidents. The 4-page report should be no more than 1.5 line spaced, font size 10, at most 2 figures (others can be placed in the appendices), excluding the cover page. The report should

1. include 1 method of central tendency, 1 method of dispersion, 1 method of cluster/density, one method of spatial autocorrelation,
2. briefly describe the parameters of each method (but not to explain the methods in details),
3. describe and discuss the spatial pattern(s) detected with each method, and
4. make conclusion on the spatial patterns of the accidents.

Refer the first lecture slides for the steps of doing a spatial analysis project.

You can choose the methods that you have learned in the databases and point analyses lectures and labs. Lab results and/or homework questions can be used. You are highly encouraged to collect and analyze your OWN point data for this report and/or implement new method(s). A high-quality report is expected.

**Spatial Pattern Analysis Report**

Carlos Fernandez

GEOG473 Section 0101

Generally speaking, spatial pattern analysis focuses on describing observed patterns in data across space. This type of analysis also tests whether an observed pattern can be confirmed or rejected from an expected null hypothesis. The point data analysis labs cover spatial pattern methods including central tendency, dispersion, clustering, density and spatial autocorrelation. The car accident data used for the point data labs is named **filtered\_CarAcc.shp** which encapsulates the occurrences of car accidents within the borders of Prince George’s County. This shapefile excludes poorly geocoded records with scores less than 60 and a **Status** attribute of U (Unmatched). Upon first glance of the point layer data, it is apparent that there are clustering behavior occurring mainly on the western border of Prince George’s County. For this lab, the purpose of these spatial pattern analyses is to discover valid statistical descriptions of the car accident data which may not be immediately interpretable through visual deduction.

The first of the spatial pattern methods to be analyzed is central tendency. Among the types of central tendency that were performed for this lab were mean center and weighted mean center. Under **Spatial Statistics Tools** > **Measuring Geographic Distributions** > **Mean Center**, this central tendency produced a new geographically centered point relative to the surrounding car accident data’s distances from it. This calculation provides a new data point rather than using one of the existing points as the new mean center. The parameters used for the mean center include an input feature class which is **filtered\_CarAcc.shp** and an optional weight field that was omitted since this central tendency method only relies of projected data to accurately measure distances. In addition to mean center, the weighted mean center was also calculated for Prince George’s County’s car accident data. The parameters used for this method are similar to the mean center in which the **filtered\_CarAcc.shp** serves as the input feature class; however, the optional weight parameter is now accounted into the calculation. Specifically, the **Level** attribute of the car accident data was selected for the **Weight Field** parameter. These two central tendency methods produced geographically-centered points that were located within close proximity of one another. Figure 1a (appendix) illustrates how these points are also located within the western border of Prince George’s County where clustering behavior is observed. Using ArcMap’s **Measure** tool, the placement of points for these central tendency methods were measured to be 207.12 m. The weighted mean center was located northwest of the mean center which suggests that car accidents with higher **Level** attribute values occur along the upper northwestern border of Prince George’s County. The placement of the mean center can also suggest that larger number of car accidents, not account for weights, occurred along the southwestern border.

Furthermore, the final central tendency method used for the point data analysis lab was the central feature. This method identifies the most geographical central location that contains the shortest total distance to all other features. Specifically, this method determined the car accident that was most centrally located in the input feature data. Unlike mean center and weighted mean center, the central feature does not create a newly calculated point; instead, it actually selects a single car accident feature that already exists and use it to represent the central feature. The central feature accepts an **Input Feature Class** and a **Weight Field**, similar to the previous central tendency methods, but a **Distance Method** parameter which specifies how the distances from each point are calculated. Two central features were calculated with the **Distance Method** being set to Euclidean distance and Manhattan distance, respectively. Figure 2a (appendix) also illustrates the how the newly selected central features differ when calculating central tendency with differing distance methods. These two central features were measured to be 1508.7 m from one another which confirms the importance of choosing a suitable distance method since it has the potential to return various features.

In addition to analyzing central tendency, the spatial dispersion of the car accident data also yields very important insights. Particularly, the car accident data was calculated and determined whether or not it follows a perfectly dispersed pattern. The procedure for determining this pattern of spatial dispersion also included the use of a goodness-of-fit test which was the chi squared method. The quadrat count analysis was used to measure the density of the points. Poorly geocoded features that contained X and Y attribute values of zero were filtered from the data using a SQL selection query. The selected features were exported to a new file where 7,643 properly geocoded car accidents can now be observed. The size of the quadrats was to be properly calculated in order to yield accurate results. Using **ArcToolbox > Cartography Tools > Data Driven Pages > Grid Index Features** on **PrinceGeorgesCounty\_Project.shp,** a quadrat size (in meters) of was used for the cell size width and height.

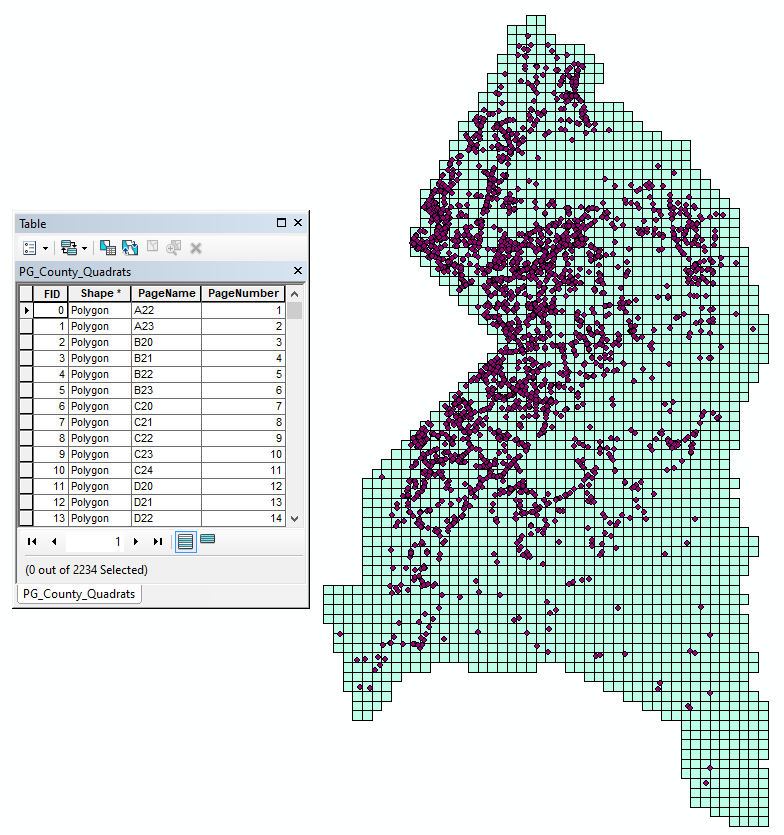


Figure 1. An implemented grid-indexed layer representing the quadrats for car accidents in Prince George’s County

Although they appear seemingly small, the calculated 789.8325m x 789.8325m quadrats total to 2,234 individual quadrats. Since this dispersion analysis is concerned with determining if the points follow a perfectly dispersed spatial pattern then the average amount of car accidents per quadrat can be calculated. After spatially joining **filtered\_CarAcc.shp** to **PrinceGeorgesCounty\_Project.shp**, a new layer called **CarAccident\_QuadratCount.shp** was created. This shapefile contained a **Count\_** attribute for the number of car accidents failing within each quadrat. This **Count\_** attribute was summarized with **Right-Click > Summarize…** which returned a summary of points within each quadrat.

Using Microsoft Excel, an exported version of **CarAccident\_QuadratCount.shp** was used for calculating the goodness-of-fit test. Specifically, the chi-square test was used for determining whether the car accident data was perfectly dispersed among all the quadrats. The expected value of the points was represented as . The variable represents the observed number of points. The screenshot of the chi-squared parameters and calculations can be seen in Figure 3a (appendix). The sum of the column represents . The degrees-of-freedom value was calculated by summarizing the **Count\_** attribute in **CarAccident\_QuadratCount.shp** which was then exported into a database file called **PG\_CarAccident\_QuadratSummary.dbf**. Figure 3a indicates a total quadrat count value of 58 which makes the degrees-of-freedom value 58 -2 = 56 along with the value of which was determined to be an extremely large value that corresponds with a -value < 0.001. This allows the results to reject the null hypothesis of the car accident spatial data exhibiting a perfectly dispersed spatial pattern.

While the results dismiss the possibility of a perfectly dispersed spatial pattern there are also drawbacks to the quadrat count analysis method. The quadrat count method is actually a measure of dispersion rather than a measure of pattern because it analyses the density of points rather than their spatial positioning relative to each other. Additionally, variations within regions of the car accident data are not recognized since this analysis only returns a single value for the entire distribution. Perhaps the biggest drawback of the quadrat count analysis is how the grid size can affect the overall calculations. The grid index layer created for the car accident data in ArcMap does not perfectly encapsulate the irregular shape of Prince George’s County’s borders. Consequently, the calculations will be affected by any quadrats not within Prince George’s County. These values contribute to a larger number of quadrats recording zero accidents which affects the data’s results. The chi-square value is an extremely large which does not have a corresponding p-value found in most available chi-square test distribution tables. This can suggest that the data could may still not be perfectly dispersed with a more precise grid index layer.

Methods of determining cluster and density were used to analyze the car accident data. Among the applicable methods, Kernel-density estimation (KDE) and Mean nearest-neighbor index were used to determine density and clustering behaviors, respectively. The KDE calculated the density of car accidents in a neighborhood around those car accidents. The calculation involves dividing the number of points in a given kernel and then divided by the total size of kernel . In a GIS application, the parameters for calculating the Kernel density in ArcMap included the **Population Field** option that denotes population values for each car accident point was left with a value of NONE. The rest of the parameters were also left with their default values. Furthermore, Mean Nearest-Neighbor index was used to measure the distance between the centroid of each car accident point and its nearest neighbor’s centroid location. The value is represented as a ratio where represents the observed mean distance between each car accident and its neighbor and represents the expected mean distance for the car accidents given a random pattern.

The findings of the KDE calculations found in the Figure 2 below illustrate the density of the car accidents around each output raster cell. The density levels vary across the Prince George’s County. Dispersed points that do not have any other car accidents within a relatively close proximity do not visually appear in the cluster. The results for the average nearest-neighbor appears similar to the Kernel Density output with key differences being the proximity of the density locations. For example, the average nearest neighbor output contains smaller individual areas of density while the Kernel density output’s areas of density overlap with one another.



Figure 2.

Lastly, spatial autocorrelation was calculated using Moran’s I for the car accident data. Two versions of this method were used: Global Moran’s and Anselin Local Moran’s . The former version measures spatial autocorrelation based on the car accident locations and their **Level** while the latter version identifies any spatially significant spatial clusters of car accidents with high or low **Level** values. In essence, the local and global version of these spatial autocorrelation statistics are similar but differ in that Anselin Local Moran’s receives its own variance, z-value along with the expected value and variance of . Figure 4a (appendix) displays the summary for the Global Moran’s I. The p-value is high which means the null hypothesis of the car accidents being random cannot be rejected. The value indicates that the distribution of the car accidents is random. Figure 4b (appendix) also (appendix) shows the output layer for calculating the Local Moran’s I along with its corresponding attribute table. The **LMiIndex** attribute contains the Moran’s value and the **LMiZScore** contains the z-score. The positive value indicates that a given car accident has neighboring car accidents with similar high or low **Level** values. The negative values indicate a difference in **Level** values for neighboring car accidents.

Appendix

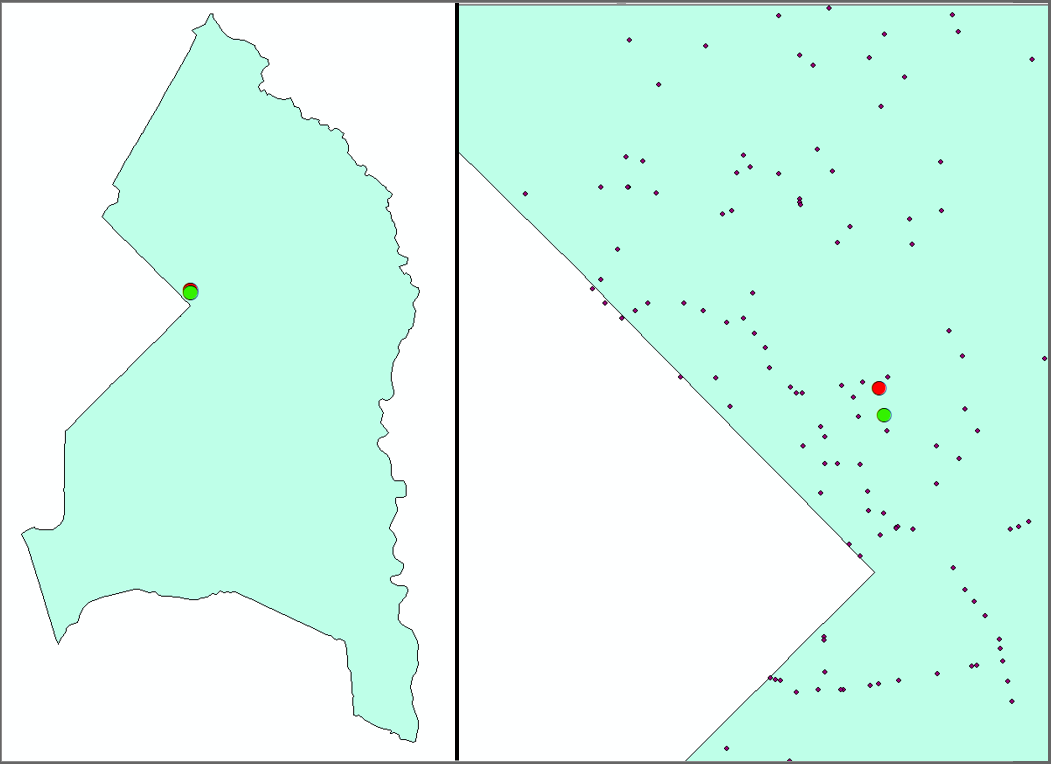


Figure 1a. Mean accidents (Red) and Weighted Mean Accidents (Green). The left picture is displayed with **filtered\_CarAcc.shp** turned off for better visibility the calculated points.

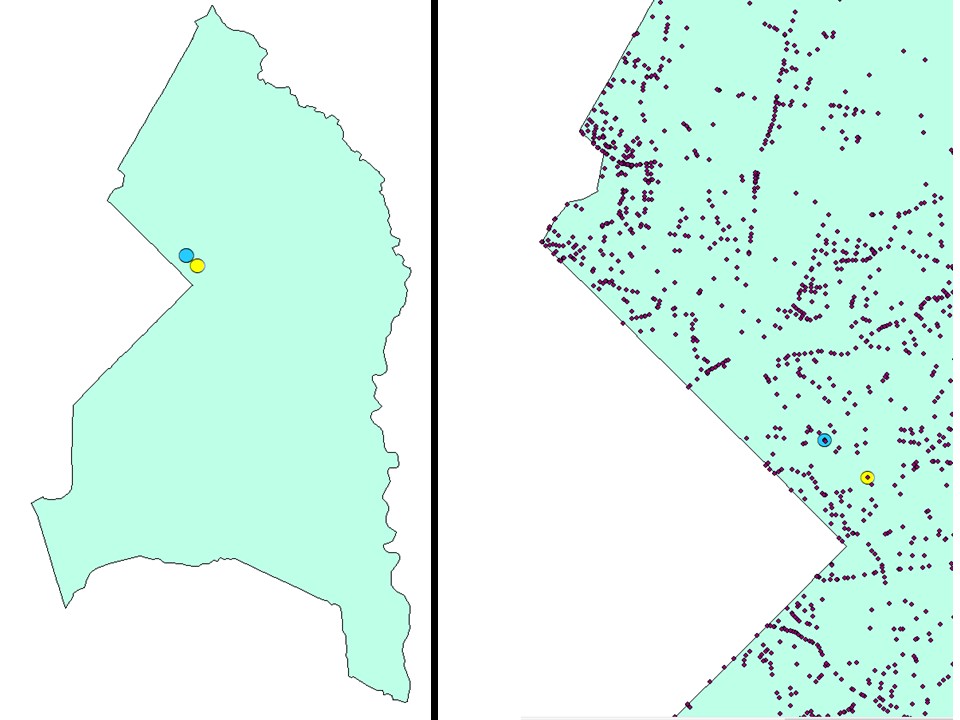


Figure 2a. Euclidean central feature (Yellow) and Manhattan central feature (Blue). The left picture is displayed with **filtered\_CarAcc.shp** turned off for better visibility the calculated points.

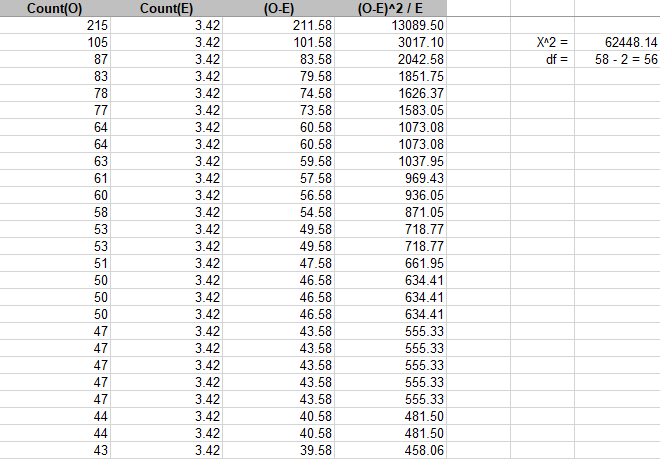


Figure 3a. The following screenshot shows a portion of the observed quadrat counts along with their expected values

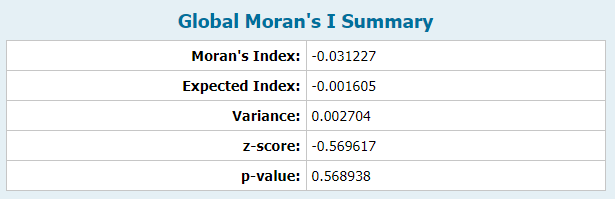


Figure 4a. Output summary for Global Moran’s I of the car accident dataset

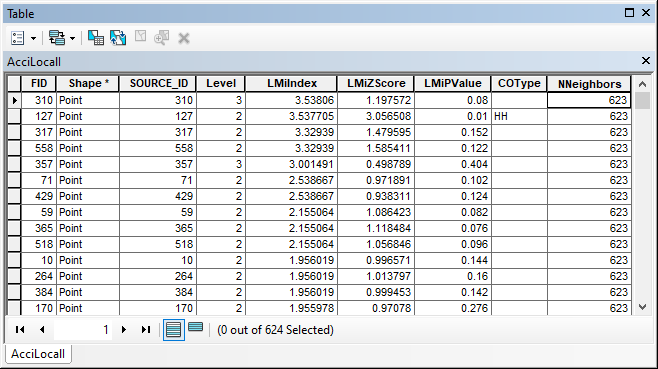


Figure 4a. Attribute table for Anselin Local Moran’s I of the car accident dataset

**Extra Question 1**

**(1 point)**

Although ArcGIS Desktop 10.# is widely used, ArcGIS Pro as an updated version has been available. The 400-level advanced GIS courses are also switching from 10.# to ArcGIS Pro. The major difference between 10.# and Pro is the user interface. Although not required but it is highly recommended to learn ArcGIS Pro.

**This Extra Question is absolutely NOT mandatory, you can choose to do it to earn ONE extra point in addition to the homework assignments**. There will be TWO Extra Points during this semester.

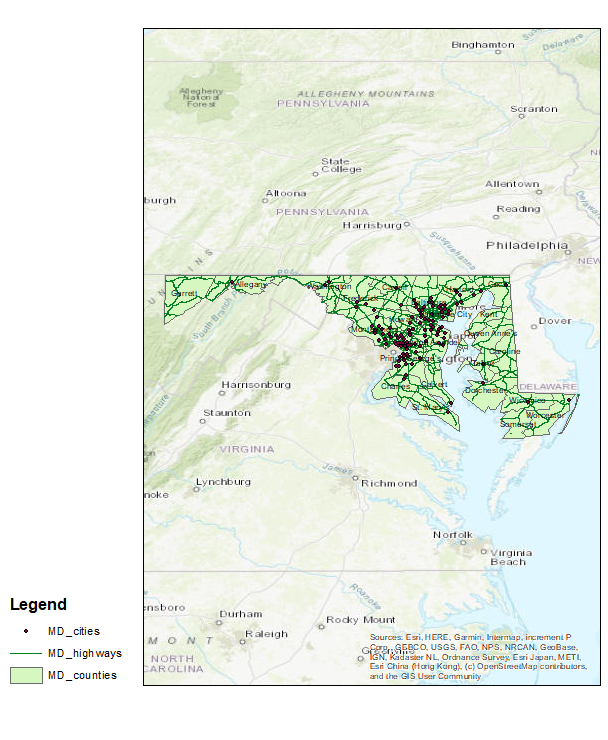
Visit GEOG473 course website on ELMS, under the module ArcGIS Pro, download the document *GEOG473 Extra Question 1 INTRO ARCGIS Pro.pdf* and the lab data *GEOG473 Extra Question 1 LAB DATA.zip*.

To access ArcGIS Pro, it is recommended to use the Virtual Lab Environment, or you can install it on your computer (contact the Instructor via [njzhou@umd.edu](mailto:njzhou@umd.edu) for the authorization code). Read *ArcGIS Desktop 1-year education license codes.pdf for details*.

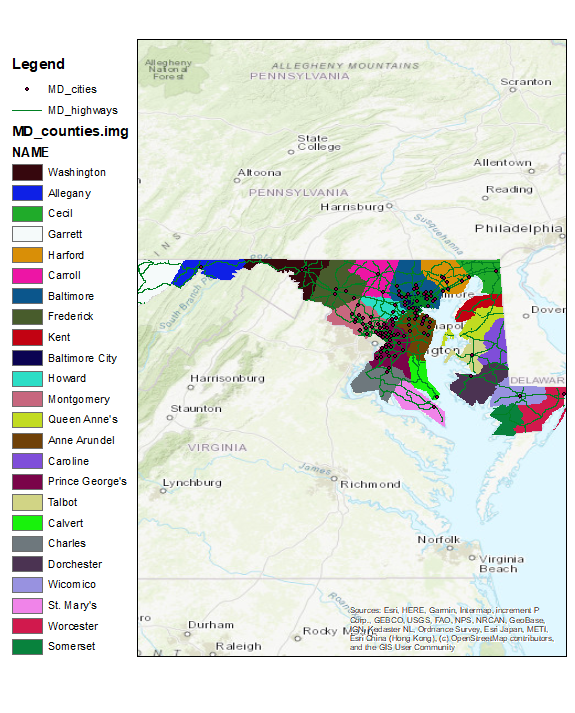
**What to do**: follow the instruction in the *GEOG473 Extra Question 1 INTRO ARCGIS Pro.pdf* and complete ALL the tasks.

**What to submit**: take a screenshot with ALL the layers visible and insert the screenshot into **this** document.

**Due**: the same as Homework 2. You won’t loose point **for this bonus question** even if you use late work policy.



**MD\_Counties.shp layer visible**



**MD\_Counties.img layer visible**