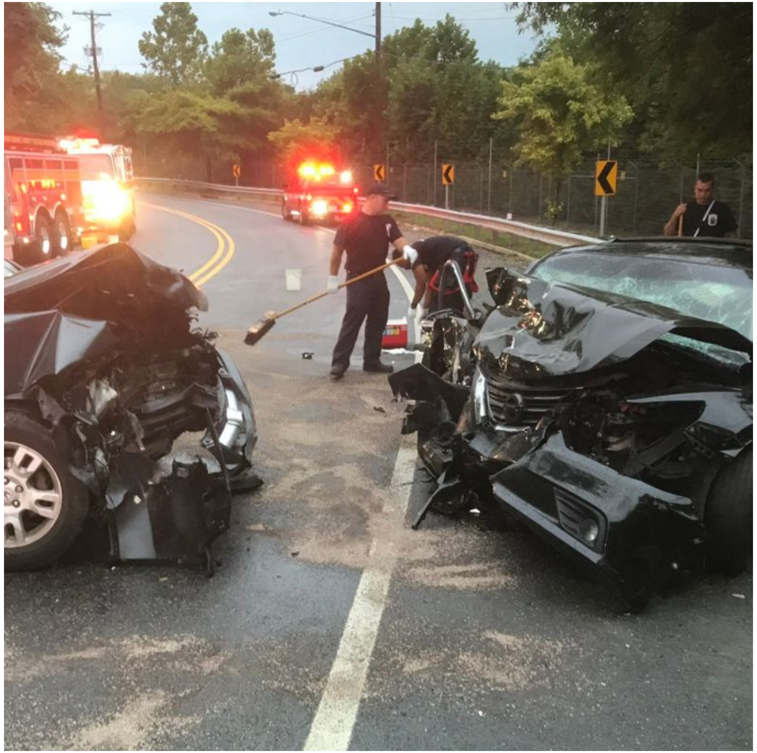
Identifying High-Risk Areas for Traffic Collisions in Montgomery, Maryland Using KDE and Spatial Autocorrelation Analysis

Stanislav Liashkov
University Of Colorado Boulder, Master’s of Science in Data Science

Agenda

1. Abstract
2. Intro
3. Data
4. Related work
5. Methods
6. Results
7. Discussion
8. Conclusion
9. References



Abstract

According to World Health Organization, motor vehicle crashes remain t**he leading cause of deaths for children and young people aged 5-29 years** in on 2023, though number of fatalities from collision gradually decline due to improved research and road safety policies. global status report on road safety 202y3

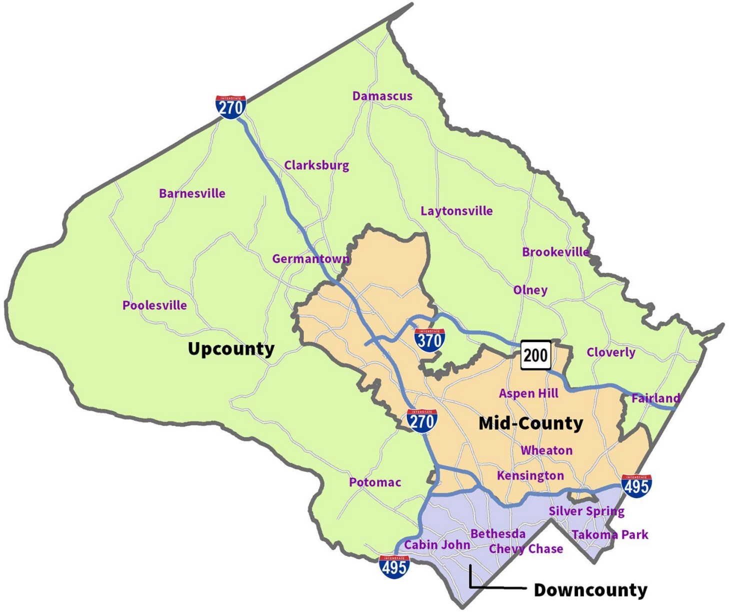
This observational study focuses on identifying dangerous high-risk areas for traffic collisions that may require further detailed investigations with domain experts and targeted interventions to reduce motor vehicle crashes in *Montgomery County, Maryland, USA.*

Using KDE and spatial autocorrelation analysis, we estimate collision densities and identify hotspots for targeted interventions. Our findings reveal significant spatial clustering of traffic collisions, with distinct patterns in densely populated urban areas and rural regions

Intro

Situated adjacent to Washington D.C., **Montgomery County** stands as the most populous county **in the U.S. state of Maryland**, with a population of 1,062,061 as recorded in the 2020 census, reflecting a 9.3% growth since 2010. The county seat is Rockville, while Germantown holds the distinction of being the most populous municipality within the county.

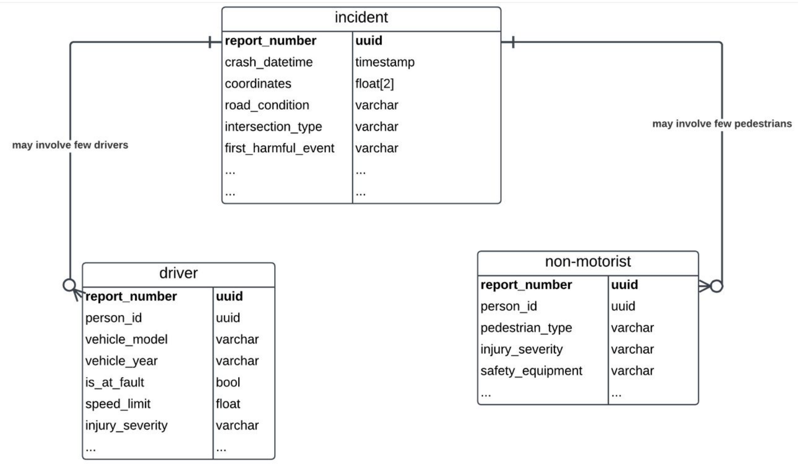
Montgomery County forms a vital part of both the Washington metropolitan area and the Washington–Baltimore combined statistical area.



Data

The dataset was compiled from open data on traffic collisions in Montgomery County, Maryland. This data was collected by the Automated Crash Reporting System (ACRS)

The dataset includes detailed information (time, location, severity, type of vehicle, circumstances etc) about motor vehicle accidents occured in country within the range of **2015 to 2024** and reported to police. There were documented around **106, 000 road accidents,** including **2,547 severe incidents** that caused fatalities or serious injuries as result of collision.
The dataset consists of three interconnected tables: *incidents, drivers, and non-motorists*



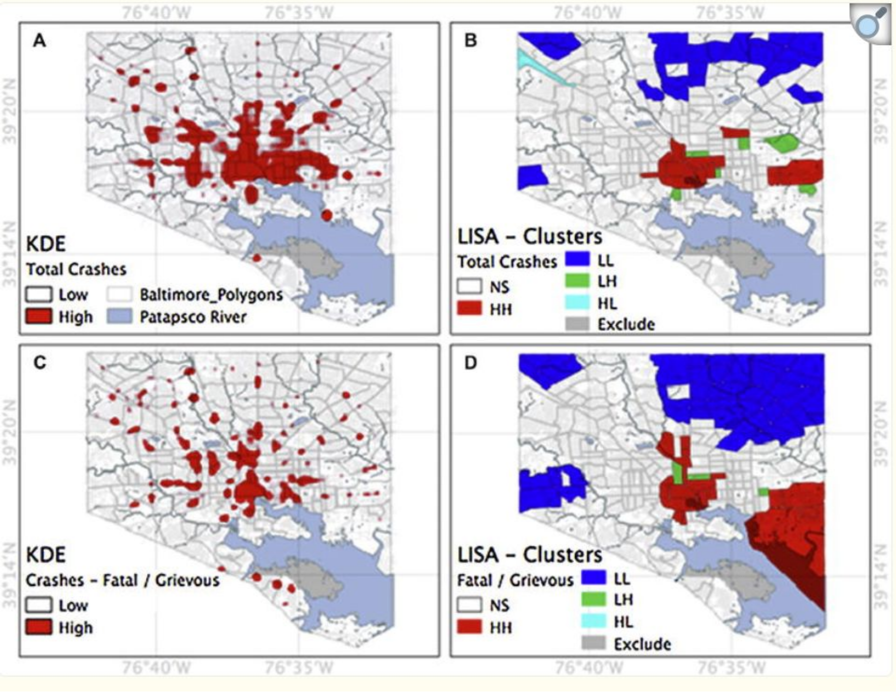
Related Work - Dezman et al. 2016

(paper: Hotspots and causes of motor vehicle crashes in Baltimore, Maryland: A geospatial analysis of five years of police crash and census data)

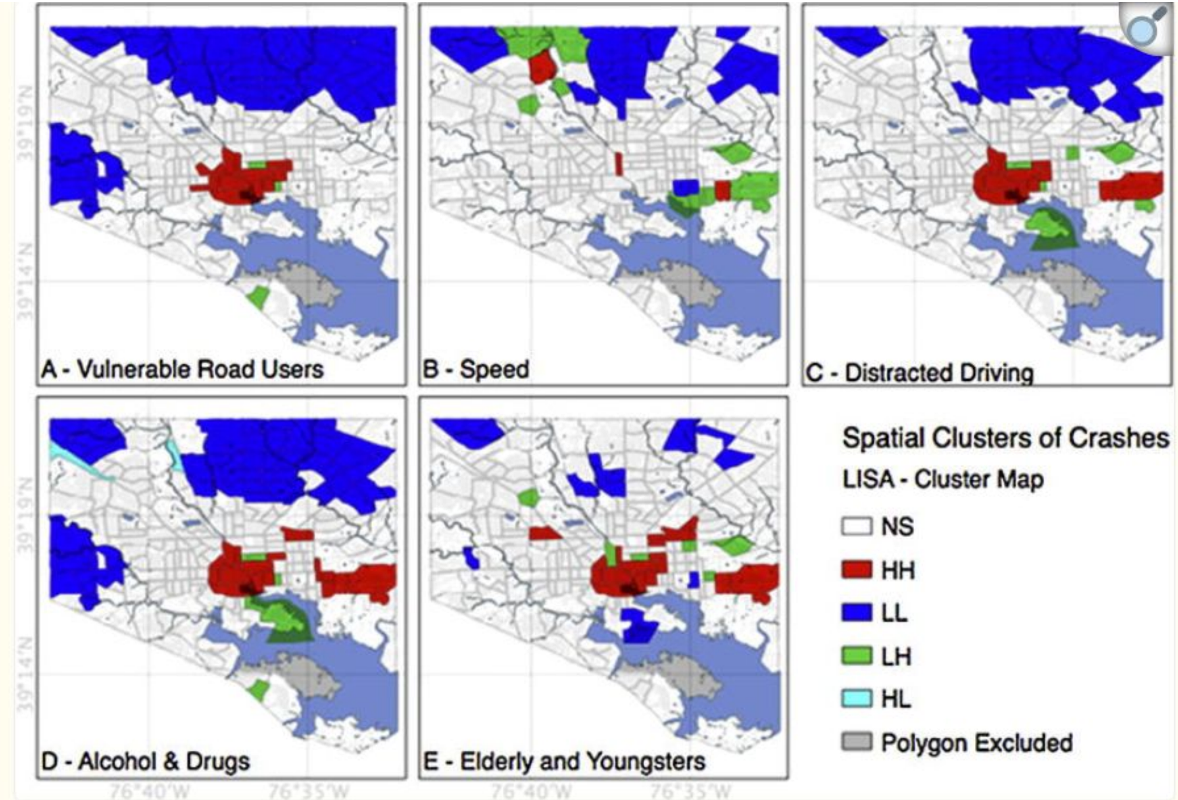
Dezman et al. (paper) analyzed traffic collisions in Baltimore city, Maryland from 2009 to 2013 utilizing *ARIMA* for time-series analysis and spatial autocorrelation techniques *Moran'I* and *Local Indicators of Spatial Autocorrelation (LISA) and Kernel Density Estimation (KDE)* for spatial visualization of density of crashes.

Their study demonstrated that most collisions occur in urban areas and at intersections with significant pedestrian activity and that the various modifiable risk factors have different spatial distribution and can be addressed in targeted manner.

We utilized their approach to analyze spatial association and patterns in collisions in Montgomery.



Related Work - Dezman et al.

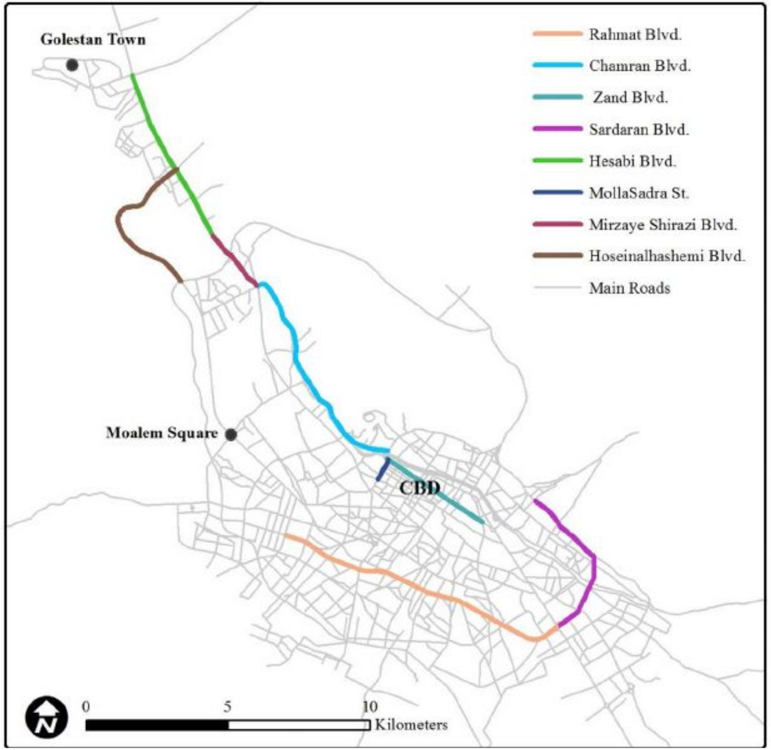


Related Work - Ali Soltani et al. 2017

(paper :Exploring Spatial Autocorrelation of Traffic Crashes based on Severity)

Researches Ali Soltani et al. have applied methods from spatial statistics such as Getis Ord Gi* statistics and Moran's I to detect hotspots of road crashes in Shiraz city, Iran and found that collisions produced significant clustering.

The result of their study confirmed crashes hotspots generally appeared on arterial roadway which tend to have higher car speed/volume and more travel lanes

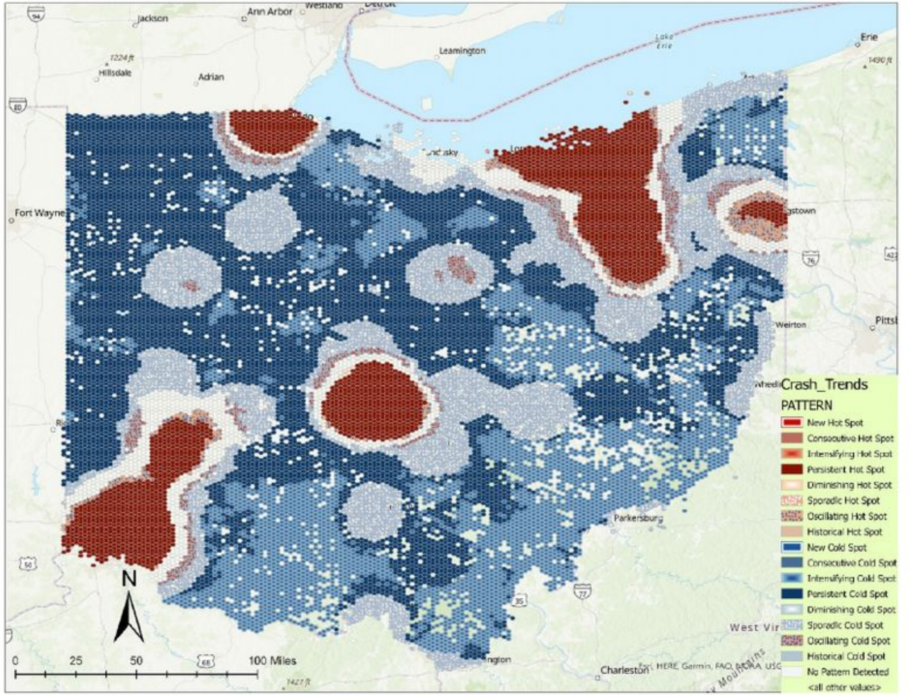


Related Work - Md Saiful Alam et al. 2023

(paper: Spatial pattern identification and crash severity analysis of road traffic crash hot spots in Ohio)

They utilized Getis Ord Gi*, the crash severity index and Moran's I to evaluate distribution of road traffic crashes in Ohio.

As result of their research, they discovered significant hotspots in big cities of state: Cleveland, Cincinnati, Toledo, and Columbus and demonstrated the efficiency of spatial analysis applied to identifying dangerous areas prone to vehicle crashes.



Methods - Moran’s I

Moran’s I - measure of spatial autocorrelation in data. It is used to test null hypotheses about spatial Randomness in data.

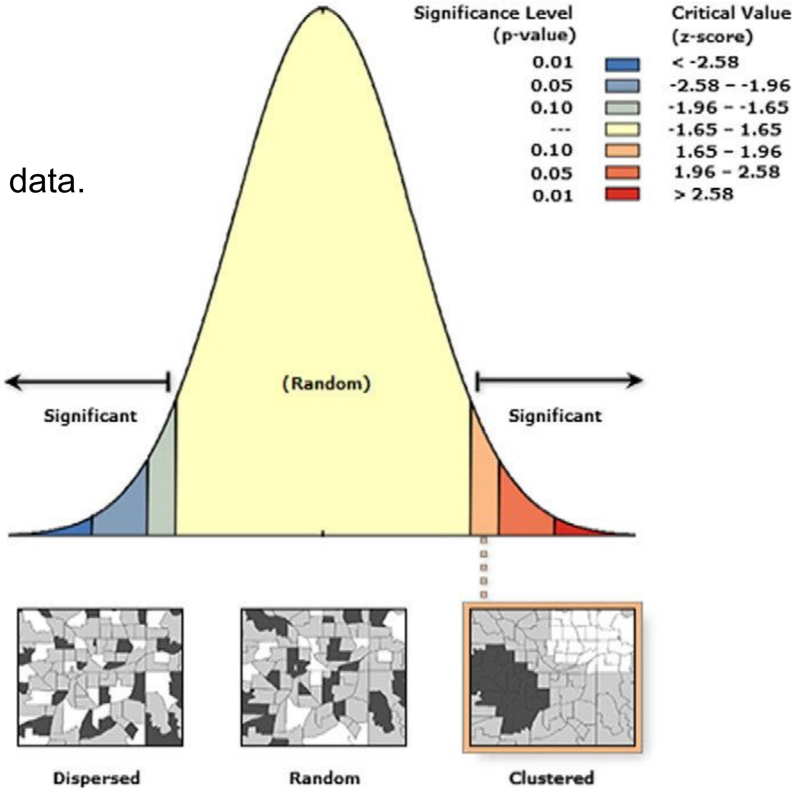
If it is significant then there is spatial autocorrelation (clustering) in data.

Global Moran's I is a measure of the overall clustering of the spatial data. It is defined as

$$I = \frac{N}{W} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

where

- N is the number of spatial units indexed by i and j;
- x is the variable of interest;
- $\bar{x}$ is the mean of x;
- w_{ij} are the elements of a matrix of spatial weights with zeroes on the diagonal (i.e., $w_{ii} = 0$);
- and W is the sum of all w_{ij} (i.e. $W = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$).

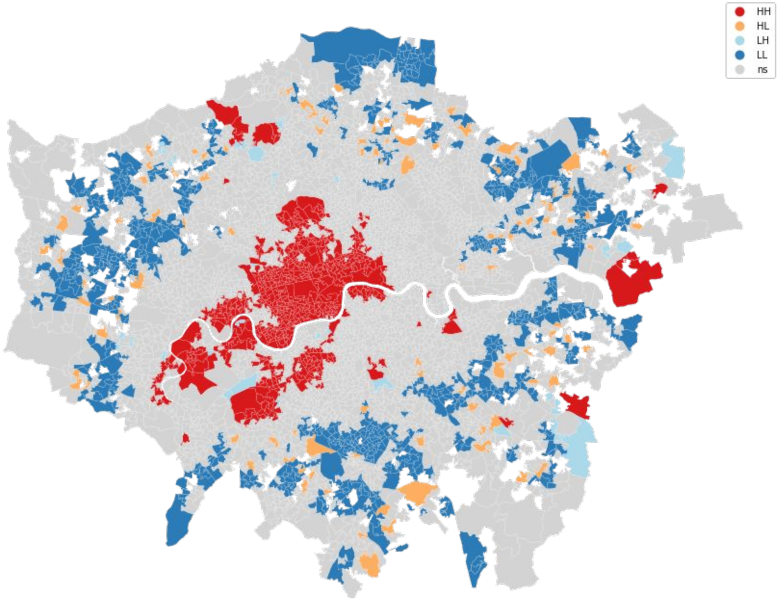


Methods - Local Indicators of Spatial Autocorrelation(LISA)

LISA - statistical method used to identify local clusters and spatial outliers. Use local statistics e.g. Local Moran.

Features:

- **Cluster Detection**: Identifies local spatial clusters and classifies them into categories: High-High (HH), Low-Low (LL), High-Low (HL), and Low-High (LH).
- **Statistical Significance**: Provides pseudo p-values to assess the significance of detected clusters and outliers.
- **Localized Insight**: Highlights spatial heterogeneity by focusing on individual locations and their neighbors.



Results - Time Series

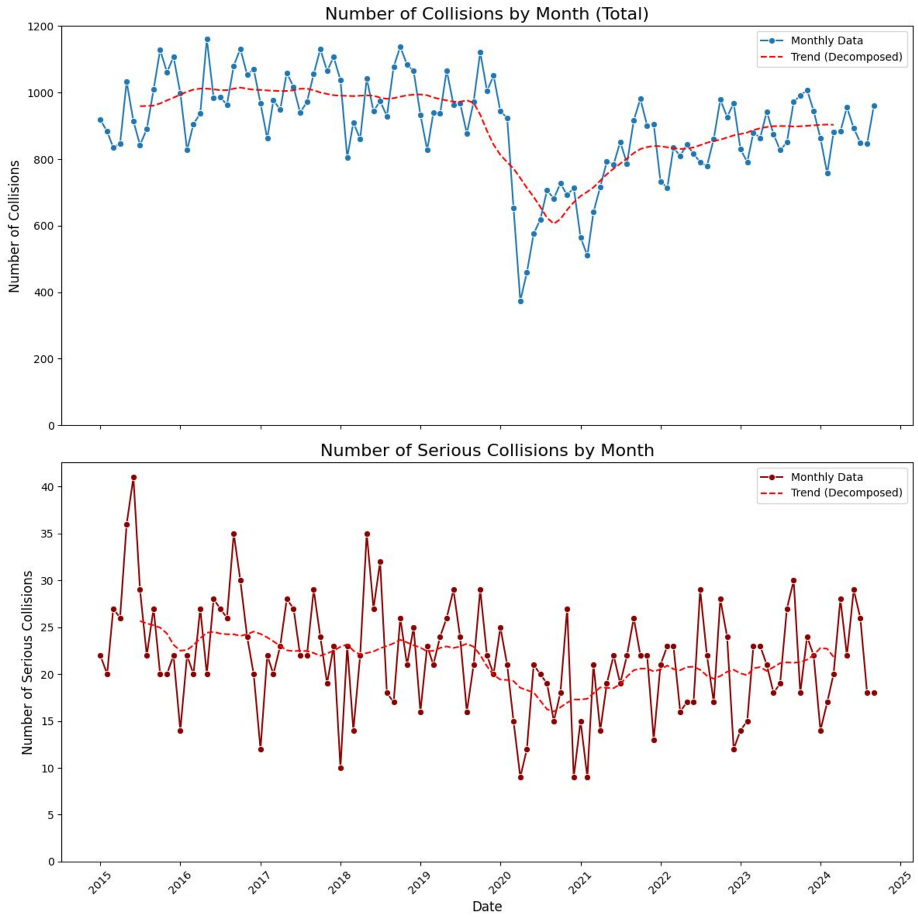
Observations:

- There is evident seasonality in collisions in Montgomery, MD

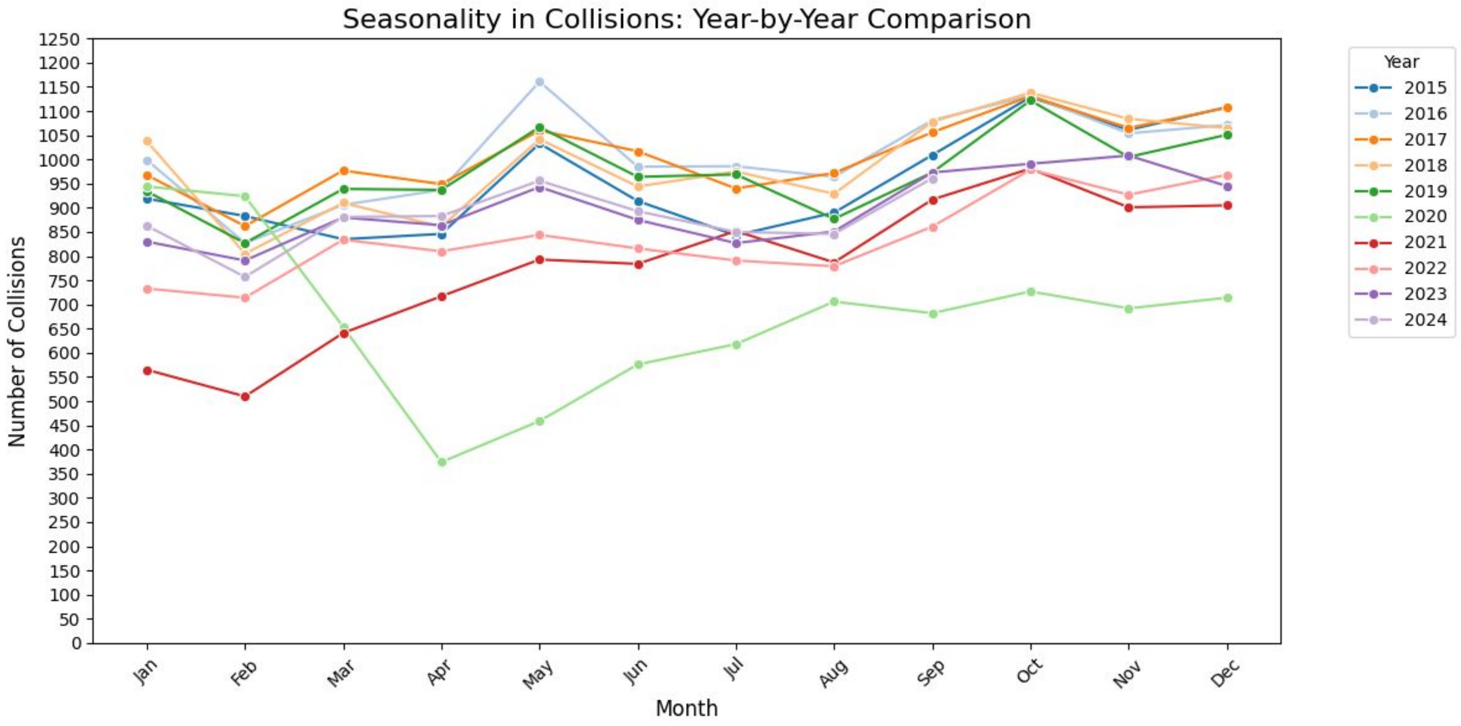
- Spring of 2020 is a structural break and significant reduction in collisions due to COVID19 lockdown

- Level of crashes is gradually getting back to pre-covid values.

- Violated assumptions for ANOVA: stationarity and homoscedasticity.



Results - Time Series (seasonality)



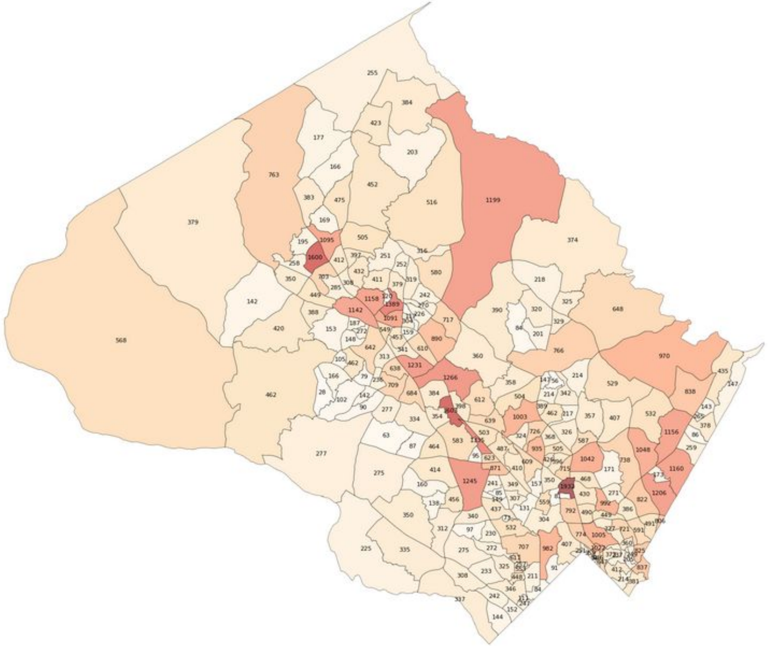
Results - Global Moran’s I for crashes

To analyze spatial autocorrelation in traffic collisions, each collision location was mapped to its corresponding census tract within Montgomery County.

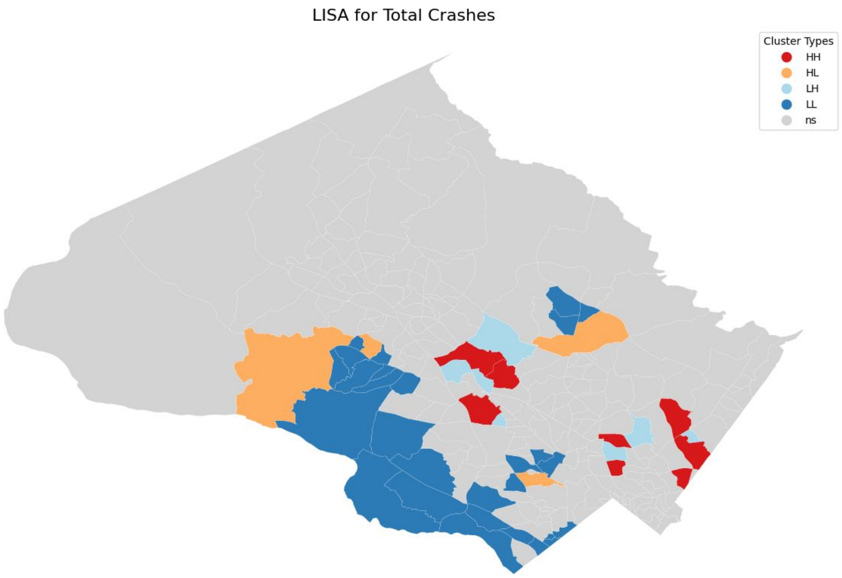
First, compute global Moran’s I to assess spatial autocorrelation. Use KNN(k=10) spatial weights matrix and significance level = 0.05. (this configuration used in subsequent steps too)

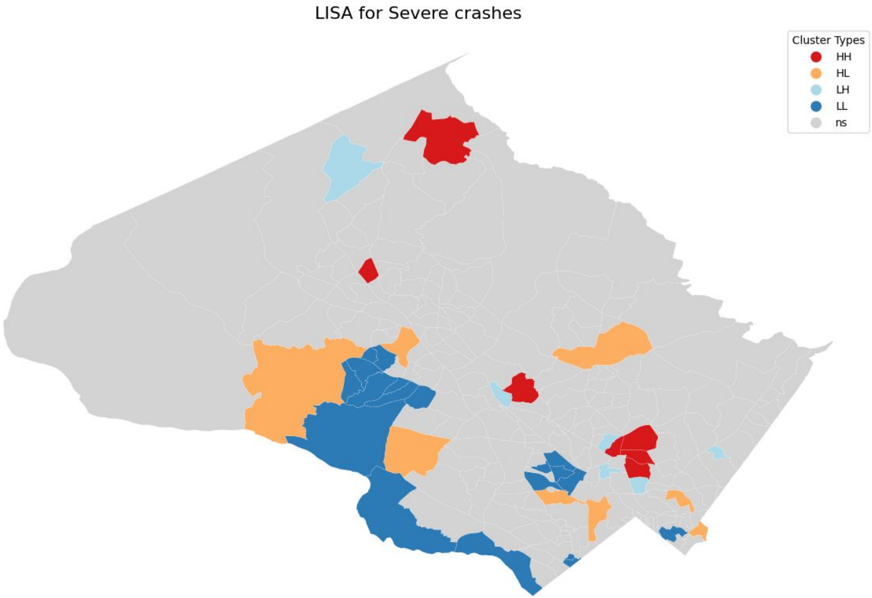
Results:

- **Moran’s I = 0.11**
- **P-value = 2.4e-5 : Spatial randomness REJECTED**
- **There is significant positive spatial autocorrelation between number of collisions in census tracts**



Results - LISA for All crashes/Severe only

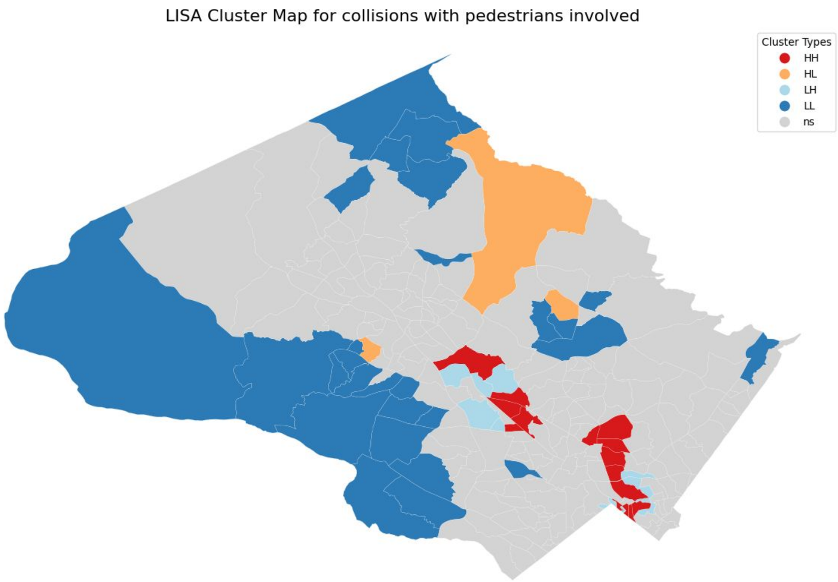


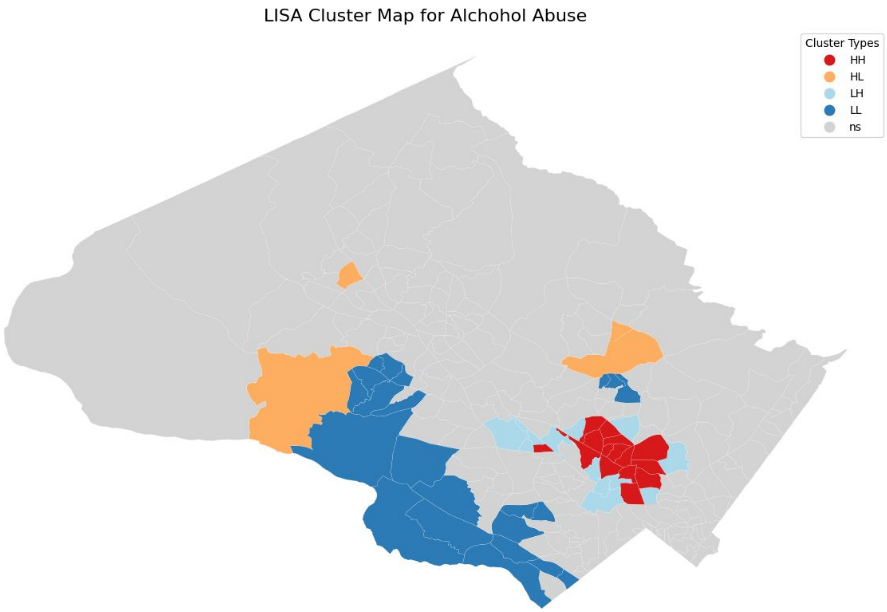


* **Severe collision** - is defined as an incident where either a driver or a non-motorist (e.g., pedestrian, cyclist) sustains a fatal injury or a suspected serious injury requiring hospitalization.

* (**Red** - High-high significant clusters, **Blue** - Low-Low significant clusters. Significance level = 0.05)

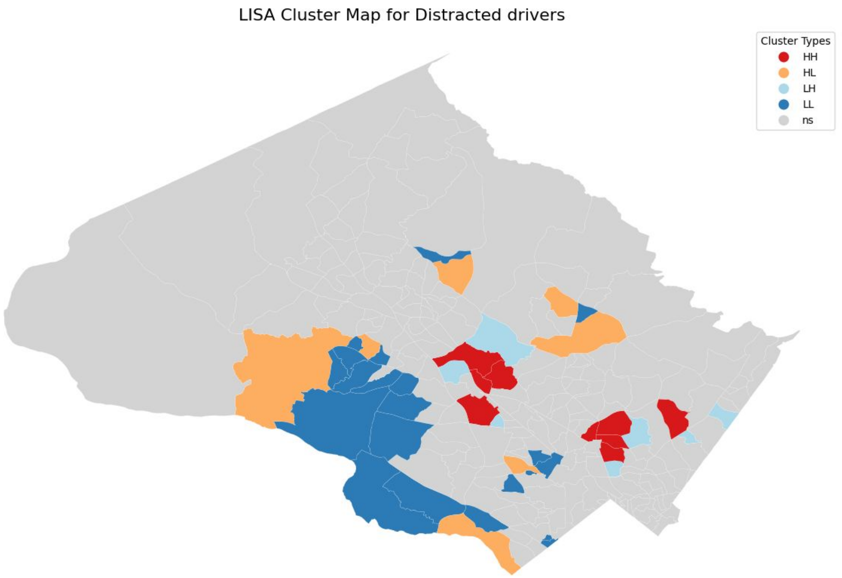
Results - LISA for crashes with pedestrians/Alcohol abuse

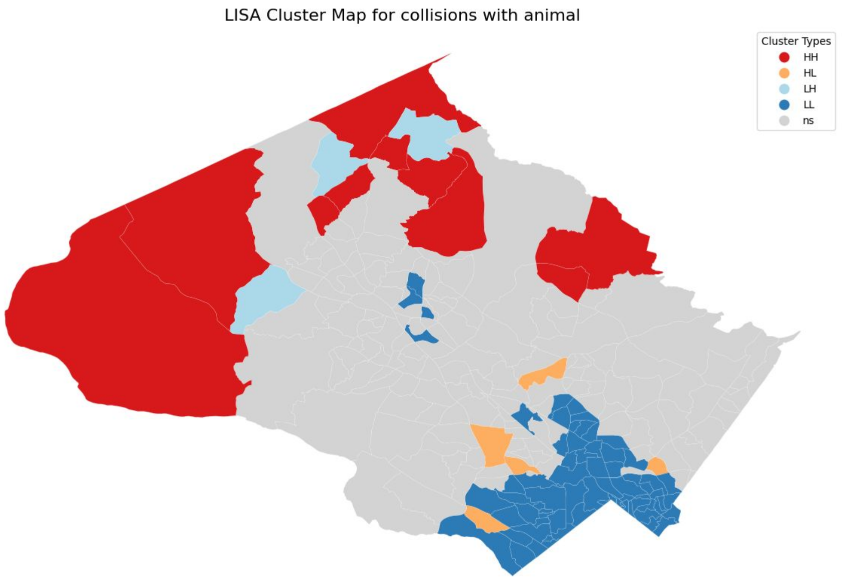




* (**Red** - High-high significant clusters, **Blue** - Low-Low significant clusters. Significance level = 0.05)

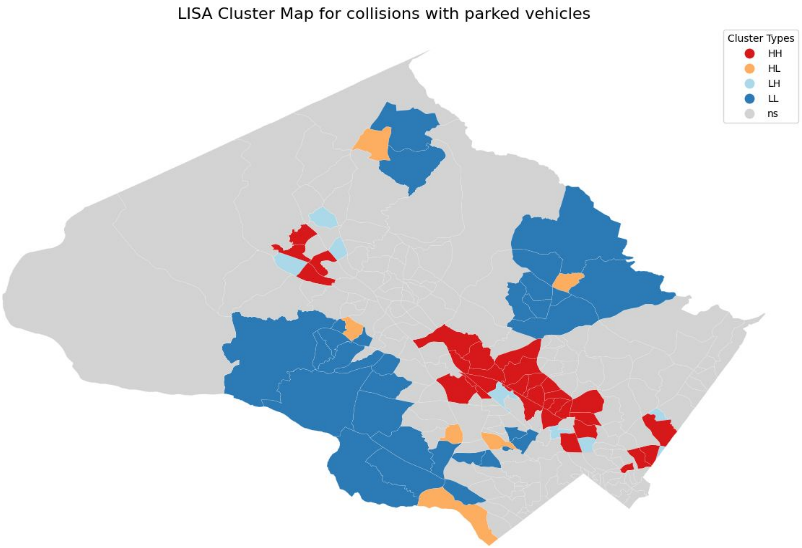
Results - LISA for distracted drivers/crashes with Animals

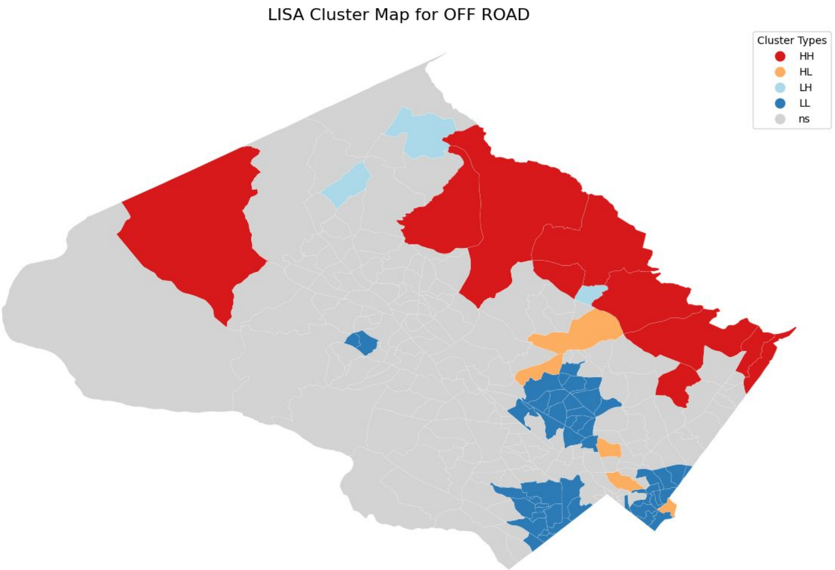




* (**Red** - High-high significant clusters, **Blue** - Low-Low significant clusters. Significance level = 0.05)

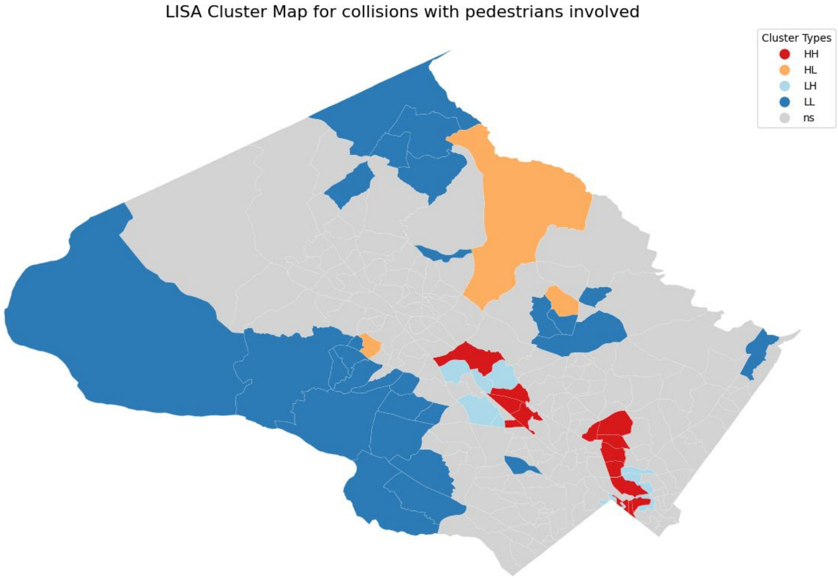
Results - LISA for crashes with parked vehicles/Off road accidents

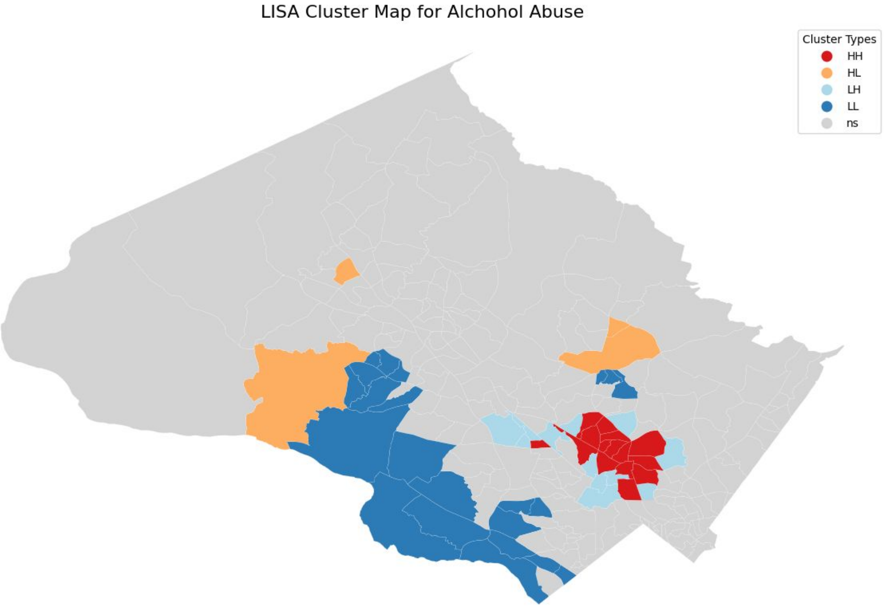




* (**Red** - High-high significant clusters, **Blue** - Low-Low significant clusters. Significance level = 0.05)

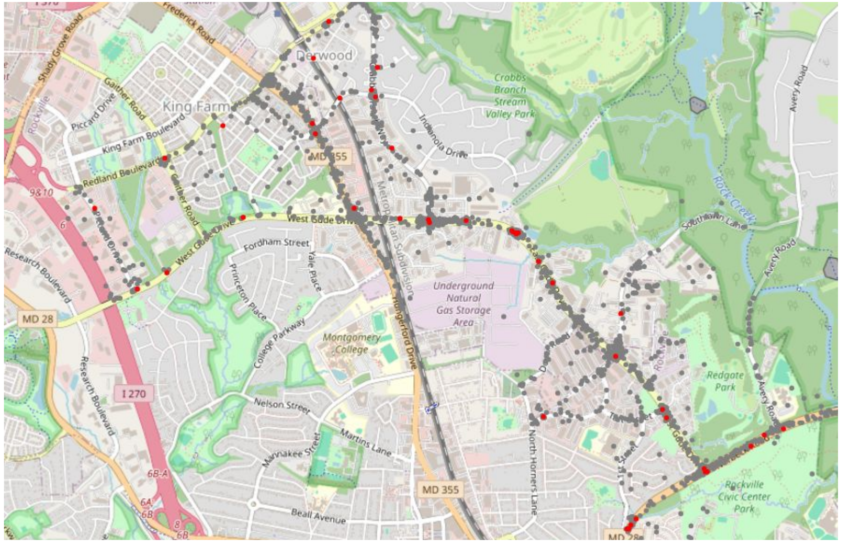
Results - LISA for crashes with pedestrians/Alcohol abuse





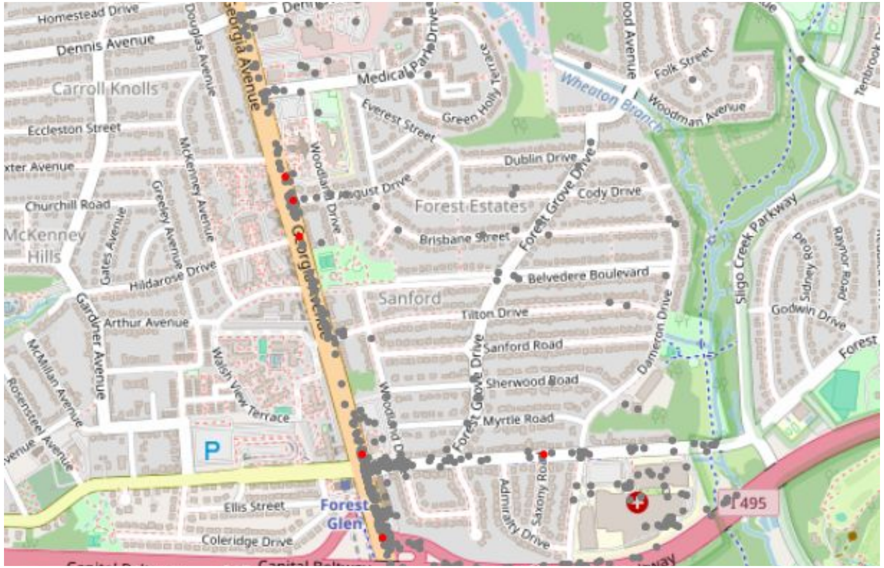
* (**Red** - High-high significant clusters, **Blue** - Low-Low significant clusters. Significance level = 0.05)

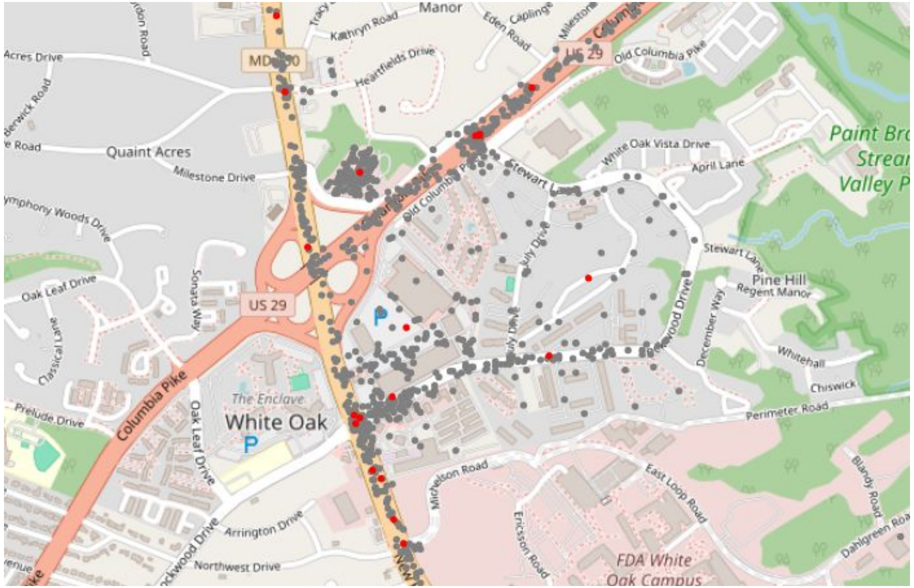
Results - Crashes distribution in significant HH clusters



* (gray points - all collisions, red points - severe collisions)

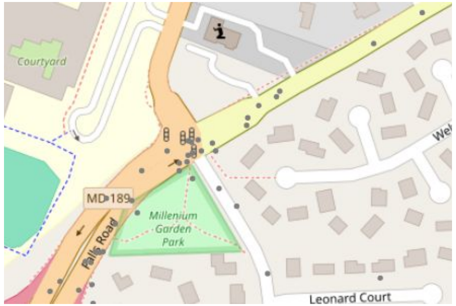
Results - Crashes distribution in significant HH clusters

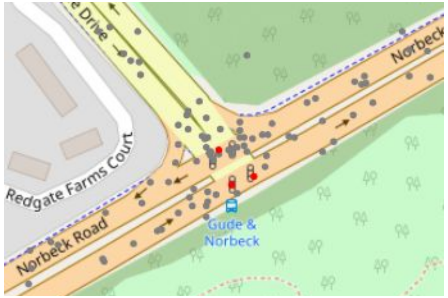




* (gray points - all collisions, red points - severe collisions)

Results - distribution of crash points (zoom-in)





* (gray points - all collisions, red points - severe collisions)

Discussion

- **Spatial Clustering of Collisions:**
  - Significant positive spatial autocorrelation found, indicating collision hotspots cluster together.
  - Urban areas (e.g., Rockville, Wheaton, Glenmont) and busy highways show high collision density.
  - Severe collisions follow similar clustering patterns in populated zones.
  - Key census tracts with high collision counts: 70390, 700903, 70121, 70150, 701001.

- **Modifiable Risk Factors:**
  - **Alcohol-Related Collisions:** Single HH cluster in Wheaton-Glenmont
  - **Collisions with Animals:** Clusters near forests/national parks; interventions like reduced speed limits or green bridges recommended.
  - **Off-Road Collisions:** Concentrated in rural, high-speed zones; road design improvements needed.
  - **Distracted Driving:** Significant clusters in Rockville, Wheaton, Glenmont, and Poolesville;

- **Urban Intersections and Pedestrian Crossings:**
  - High collision frequency at intersections and pedestrian crossings in densely populated areas.
  - Consistent with findings in Baltimore (Dezman et al.), emphasizing the need for targeted urban safety measures.

Conclusion

- Identified high-risk areas for traffic crashes using spatial autocorrelation analysis and kernel density estimation (KDE).
- Examined the spatial distribution of modifiable risk factors, including alcohol abuse, distracted driving, and collisions with animals.
- Provided guidance for policymakers and stakeholders to implement targeted interventions aimed at enhancing road safety in Montgomery country.

References

[1] Global Status Report on Road Safety 2023, World Health Organization. link

[2] Hotspots and causes of motor vehicle crashes in Baltimore, Maryland: A geospatial analysis of five years of police crash and census data. Z. Dezman et al. 2016, link

(University of Maryland School of Medicine, Baltimore, Maryland, USA; State University of Maringá, Maringá, Brazil; Duke University, Durham, North Carolina, USA)

[3] Wikipedia, Moran's I link

[4] Local Indicators of Spatial Autocorrelation (LISA), Luc Anselin, 2020 link (GeoDA software)

[5] Exploring spatial autocorrelation of traffic crashes based on severity. Ali Soltani, Sajad Askari, 2017, link

(Department of Urban Planning, Traffic and Transportation Research Centre (TTRC), Shiraz University, Iran)

[6] Getis–Ord Spatial Statistics to Identify Hot Spots by Using Incident Management Data, Praprut Songchitruksa and Xiaosi Zeng, 2010 link

[7] Spatial pattern identification and crash severity analysis of road traffic crash hot spots in Ohio, Md Saiful Alam and Nusrat Jahan Tabassum, 2023, link (University of Maryland, University of Toledo)

[8] Montgomery County, Maryland, USA. Wikipedia, link

[9] Automated Crash Reporting System (ACRS) Field reference guide, 2017, link

[10] Crash Reporting - Incidents Data from 2015 to 2024, data.gov, link

[11] Community report on tourism in Montgomery County, Visit Montgomery MD, 2022, link

[12] Empirical Bayes Index (mentioned in lecture), Luc Anselin, 2016, link

(University of Chicago)

[13] Laplacian Smoothing (Additive Smoothing), Wikipedia, link