Spatial Analysis of Los Angeles Crime

Xinyi Yan

Introduction

Crime is usually an important factor that affects the quality of life and the development of the city. Crime analysis allows law enforcement agencies to get a better understanding of the patterns of crimes and enables them to allocate their resources more efficiently. For this study, we aimed to explore the distributions of age and gender of victims and the rate of each crime in Los Angeles(LA) in 2017, and assess the level of spatial clustering for a specific type of crime. In addition, we also identified the most likely crime districts and provided the related visualizations. The dataset used in this project is provided by the Los Angeles Police Department, and the cleaned version can be found on Kaggle[1].

Method

Data

This dataset is transcribed from original crime reports provided by the Los Angeles Police Department, from 2010 to September 2017. The location, type, and reporting district of the crimes, the age, and gender of the victims were reported. A total number of 61709 crimes were recorded in the year of 2017. A separate shapefile for the reporting districts is also provided for a better understanding of the neighborhoods in LA. Geographic coordinates of each district were also obtained to approximately represent the geographical centroid for each district.

Spatial Autocorrelation

To examine the potential spatial dependencies, we assumed a quasi Poisson intercept model for the occurrence of robbery, under the null hypothesis that the expected number of robberies in each district is proportional to the number of all types of crimes in that area.

.

$$Y_i \sim Poisson(E_i \; exp(\beta_0))$$

Where Y_i is the observed number of robberies in each district, and E_i is the expected number of robberies if we assume there is no spatial correlation. We used both Moran's I and Geary's c statistics to assess the spatial dependencies through the model's residuals. Two weight matrices were created using binary(B style) and row standardized (W style) coding schemes respectively. Both Moran's I and Geary's c tests provide corresponding p-values based on permutation, and those p-values indicate the significance level of spatial correlation among residuals from our Poisson model. A significance level was set to be 0.05.

Cluster Detection

To detect the districts with higher intensity of robbery, compared to other districts, we used the spatial scan statistic method with SatScan software. Again, we also assumed a Poisson model distribution for the number of robberies with a population consists of all types of crimes. The null hypothesis is that there are no significant clusters. Likelihood ratio tests were conducted and p-values were obtained for scanning windows with likelihood ratios greater than expected. The p-values were calculated by Monte Carlo

simulation. The most likely cluster is detected if the corresponding p-value is smaller than 0.05. The maximum population size was set to be 20%, and the number of simulations was 500.

Result

Descriptive Statistics

In the year of 2017, there were a total number of 61709 crimes committed in 1129 districts. Among the 61709 crimes reported, the top five types of crimes are 'Vehicle-stolen', 'Battery-simple assault', 'Burglary from vehicle', 'Burglary' and 'Theft plan-petty($950 and under')' (Figure 1). The most committed crime in 2017 is stolen vehicles with a number of 5155. We observed that the majority of the victims were females around 25 years old. Male victims were about half the number of female victims and with the same pattern in age.

Figure 1: Top 10 Crime Recorded in 2017

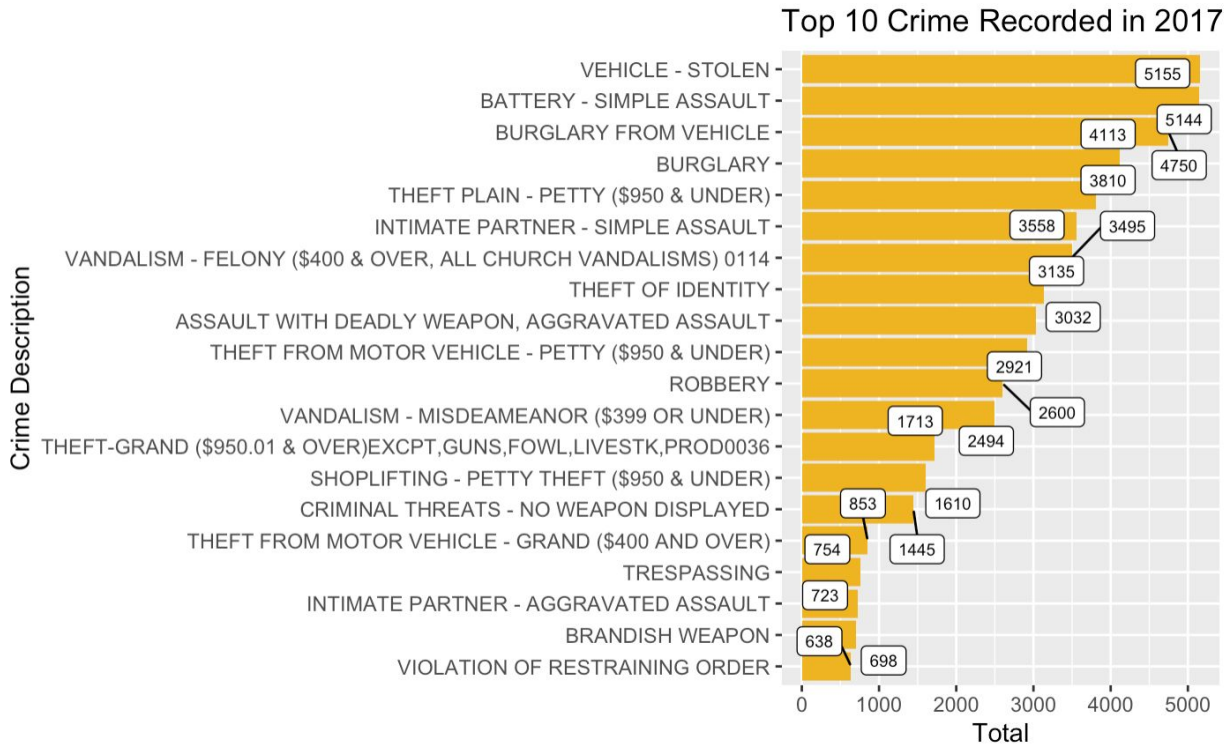
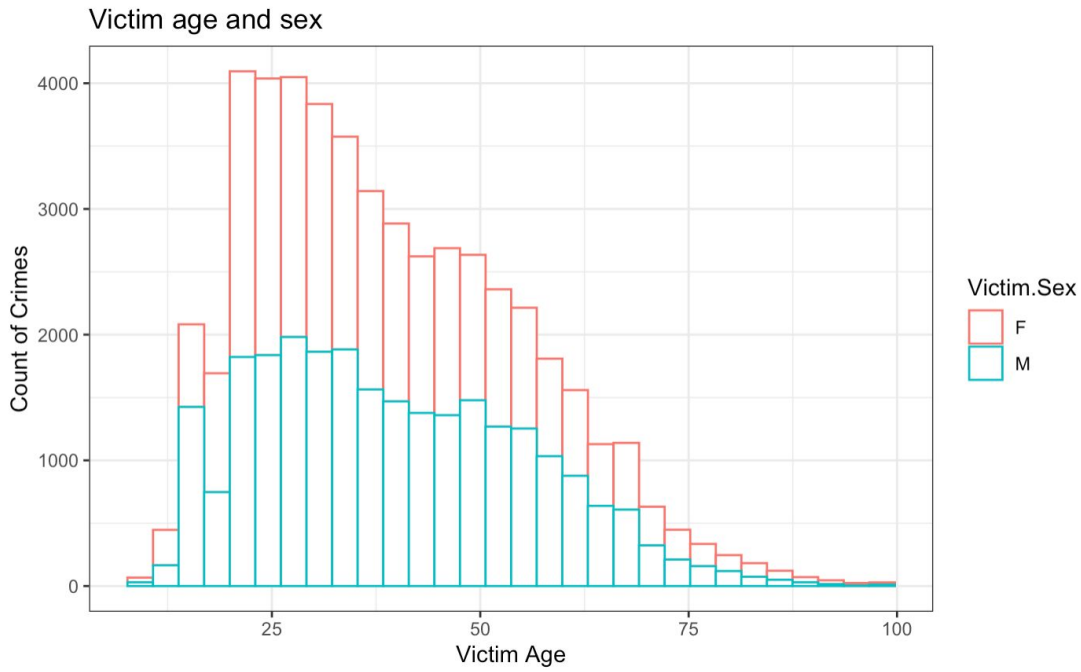


Figure 2: Distribution of victims' age grouped by sex



Spatial Autocorrelation

In 2017, a total number of 2600 robberies were committed. Robberies were observed to be more prevalent within districts located in the north regions and the middle east regions as well (Figure 2). The Standardized Mortality Ratios(SMRs), defined as the ratio of observed counts and expected counts, are mapped in figure 3. Darker color represents a higher ratio of the observed number of robberies compared to the expected number of robberies in each district that is proportional to the number of all types of crimes in that area. Again, the rates were observed to be more prevalent in the north and middle east regions.

We examined the level of clustering through the residuals of the quasi Poisson intercept model using Moran's I and Geary's c statistics. From table 1, we see that Moran's I test statistics are 0.13 for both binary and row standardized neighbor weight schemes, and the corresponding p-values are both smaller than 0.05. Geary's c test statistic is 0.90 for B style and 0.86 for W style, and the corresponding p-values are both smaller than 0.05 as well. Thus, we have enough evidence to reject the null hypothesis that there is no clustering in the SMRs by the district. We have evidence of spatial clustering from both Moran's I and Geary's c test statistics.

Table 1: Moran's I and Geary's c test statistics using 'B' and 'W' weight schemes

	Moran's I		Geary's c	
	Test statistic	P-value	Test statistic	P-value
B Style	0.13189	<0.00001	0.89545	0.00006
W Style	0.13206	<0.00001	0.85673	<0.00001

Figure 2. Observed number of robberies

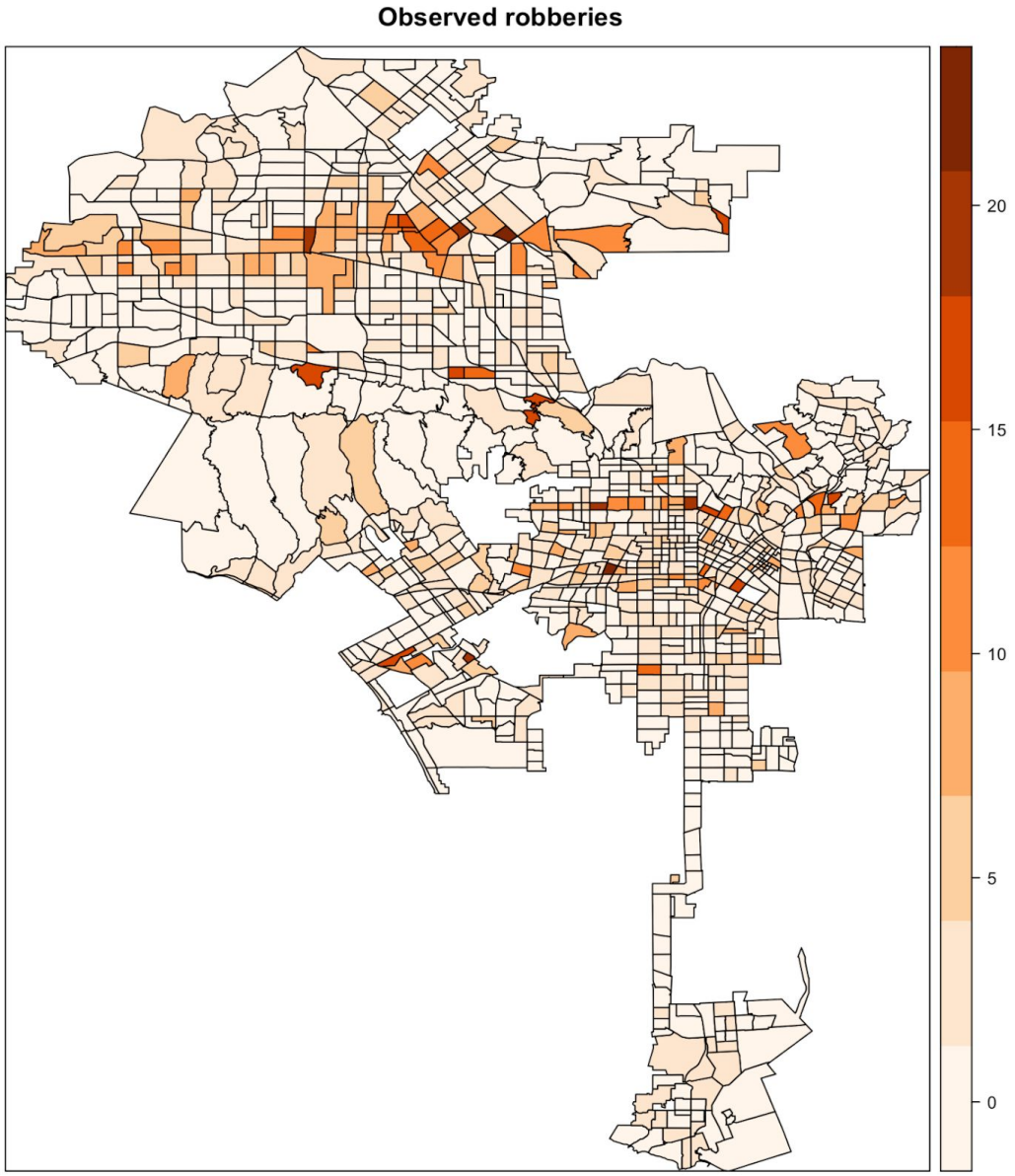
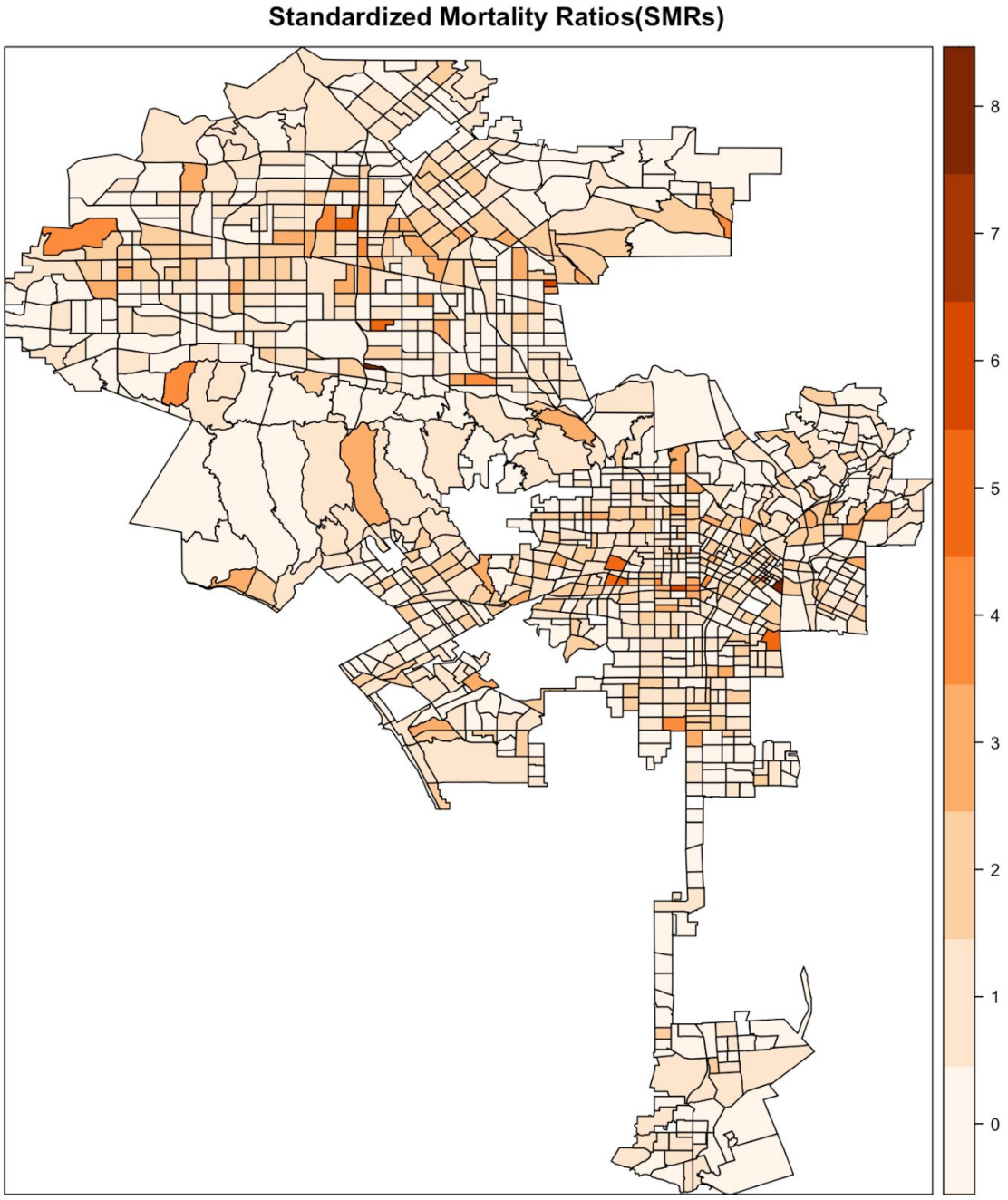


Figure 3. Standardized mortality ratios for robbery



Cluster Detection

We also performed cluster detection on the 1129 districts in LA, using the SatScan method of Kulldorff. The significance level was set to be 0.05, and the maximum population size was set to be 20%. The centroids for each district were estimated from the geographic coordinates that we obtained from the district shapefile. The cases are observed robbery cases, the expected cases are the averaged number of robberies in each district that is proportional to the number of all types of crimes in that area. For the most likely cluster, the corresponding p-value is 0.01(Figure 4), indicating a significant clustering pattern. Figure 5 maps the estimated most likely cluster in the color red. We observed that the most likely cluster consists of districts in the middle east of LA.

Figure 4: Monte Carlo Distribution

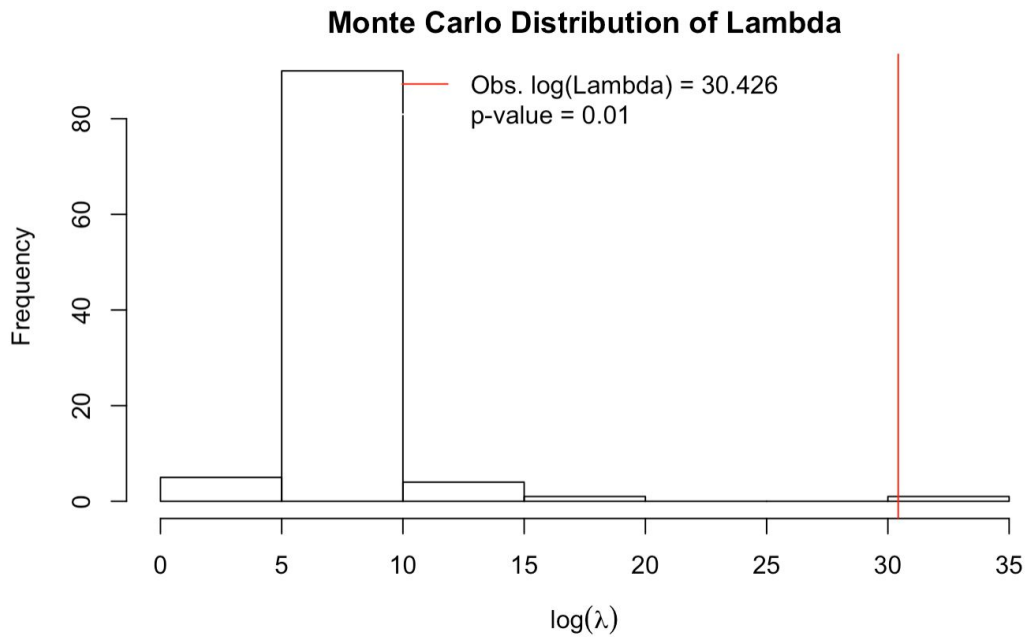
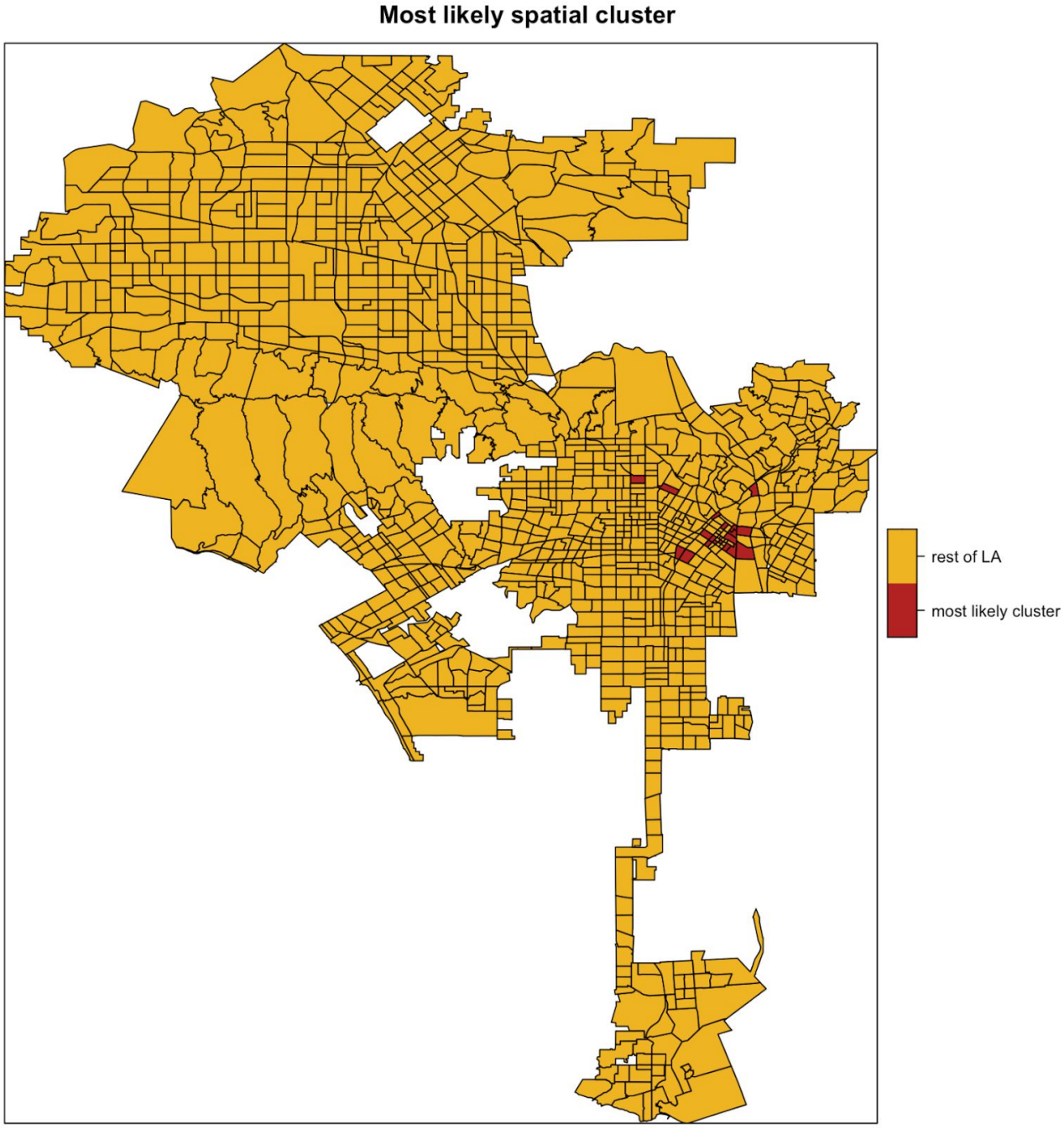


Figure 5. Most likely cluster via SatScan



Discussion

In this study of LA crime in 2017, we explored the types of most committed crimes and the age and gender distributions of victims. We found that stolen vehicle is the most committed crime, and the majority of the victims are females who were around 25 years old. We also assessed the level of clustering in the residuals from the quasi Poisson intercept model. We found evidence for clustering from both Moran's I and Geary's c test statistics. A cluster with higher intensity of robbery was also detected through the SatScan method of Kulldorf. This most likely cluster consists of districts in the middle east of LA.

However, it is also possible that the level of clustering can be explained by some unmeasured covariates which affect the mean function for the Poisson model. In addition, the choice of maximum population size for the SatScan method is arbitrary, and the significance level we chose is not proportional to the sample sizes, which can affect the power of our test.

In summary, our analysis provides evidence for spatial correlation and identifies middle east districts as areas with higher robbery intensity. Hopefully, our results can give directions for future allocation of police resources.

Reference

[1] City of Los Angeles. (2017, September). Crimes in Los Angeles. Retrieved March 10, 2021 from https://www.kaggle.com/cityofLA/crime-in-los-angeles.