**Can streetlight decrease crime?**

* **A case study in Los Angeles**

**Abstract**

Understanding the indicators of crime data is important since the increased crime rates year by year in certain cities in United States. Except the social indicators that widely used as indicator to predict crime rates, streetlight has been a feature that used to prevent crime. Previous research have been implemented to exam the spatial association between streetlight density and the number of crime; machine learning algorithms has been increasingly used to understand crime patterns and predict future crime occurrences, which provided more credible evidence for researchers to optimize scarce public safety resources, such as improving streetlights in disadvantaged neighborhoods.

**Introduction:**

Los Angeles is the largest city in California. This city has a long history to decrease the crime rate. The crime in Los Angeles has varied throughout time, reaching peaks between 1970 to 1980s. Los Angeles was once a notoriously dangerous region of the City of Los Angeles which has had an extensive history of gang violence started in the 1920s [5]. Thus, even though the crime rate is continuing to decline, research on Los Angeles crimes and their predictors is still needed in order to prevent and reduce crimes in this area. According to the past experience, social factors such as race [3, 12], education [6], income [11], and unemployment rate [10] are the most common variables that have been taken into account and proved related to the crime.

In terms of the research methods, the use of spatial analysis in crime analysis research is now common since it can help us clarify crime concerns with visual information [14]. A key hypothesis behind spatial crime pattern analysis is that crime rates exhibit correlations with environmental settings [14]. Recently, researchers have found that street lighting is an important factor in affecting the probability of crime occurrence [2]. Nighttime lighting is an important public service that impacts human activities and promotes transportation and pedestrian safety. When evaluating the relationship between street lighting and criminal rate, on the one hand. It has been showed that urban layout and structure could contribute to the decrease of criminal behavior caused by street light [7]. On the other hand, some studies found out that the improved street lighting has no relationship with the number of crimes [8] or even increase the number of crimes in some circumstances [13].

Thus, in order to further decrease the criminal rate, more researches needed to be done. Nowadays, with the development of big data and machine learning, some of researchers would like to use machine learning algorithms to understand crime patterns and predict future crime occurrences. Various machine learning techniques have been successfully applied in fraud detection, violent crime, traffic violence, and cybercrime [9]. Thus, the purpose of this research is using spatial analysis and machine learning tools to analyze the relationship between street lighting and the crimes.

**Data:**

All of the datasets in this research are publicly available. For the street lighting location data, it comes from Bureau of Street Lighting – City of Los Angeles [1]. The L.A Neighborhood Boundaries data, crime data, and Demographics are all come from the GeoHub [4], which is the City’s new public platform for visualizing and downloading Open Data. Detailed data features are listed below (Table 1). For some of the columns that I kept in crime data need more explanation. The crime data has classify the crimes into four categories. Crime Code 1 is the primary and most serious one. Crime Code 2, 3, and 4 are respectively less serious offenses. Lower crime class numbers are more serious. Using these variables can help us further develop the detailed spatial relationship between the number of crimes and the number of streetlights.

In terms of the data processing, the number of streetlight and the number of crimes will be calculated based on the neighborhood boundaries. For the demographic variable, I would like to use the black people density as one of the indicators. Based on the analysis above, I will further decide it is necessary or not to add some other factors (e.g., education level, unemployment rate, and income).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Dataset name** | **URL** | **Number of rows** | **Number of columns** | **Number of relevant columns** | **Number of valid rows (not NaN on relevant columns)** | **Data type for each relevant column** |
| Streetlight locations within the City of Los Angeles | <https://data.lacity.org/d/9ei6-svt8?category=A-Well-Run-City&view_name=Streetlight-locations> | 217639 | 12 | 2 | 217639 | geometry  : point  status:  TEXT |
| LA Times Neighborhood Boundaries | https://geohub.lacity.org/datasets/d6c55385a0e749519f238b77135eafac\_0 | 114 | 3 | 2 | 114 | geometry: POLYGON  Name: TEXT |
| Crime Data from 2010 to 2019 | https://data.lacity.org/A-Safe-City/Crime-Data-from-2010-to-2019/63jg-8b9z | 220029 | 28 | 220029 | 7 | From Crm Cd 1 to  Crm Cd 4: TEXT  Lat and Lon: Number  Date\_rptd:  [Floating Timestamp](https://dev.socrata.com/docs/datatypes/floating_timestamp.html) |
| Demographics of Neighborhood Councils | https://geohub.lacity.org/datasets/demographics-of-neighborhood-councils | 97 | 77 | 3 | 97 | geometry: POLYGON  Total population:  Number  Black population: Number |

Table 1. Data Summaries

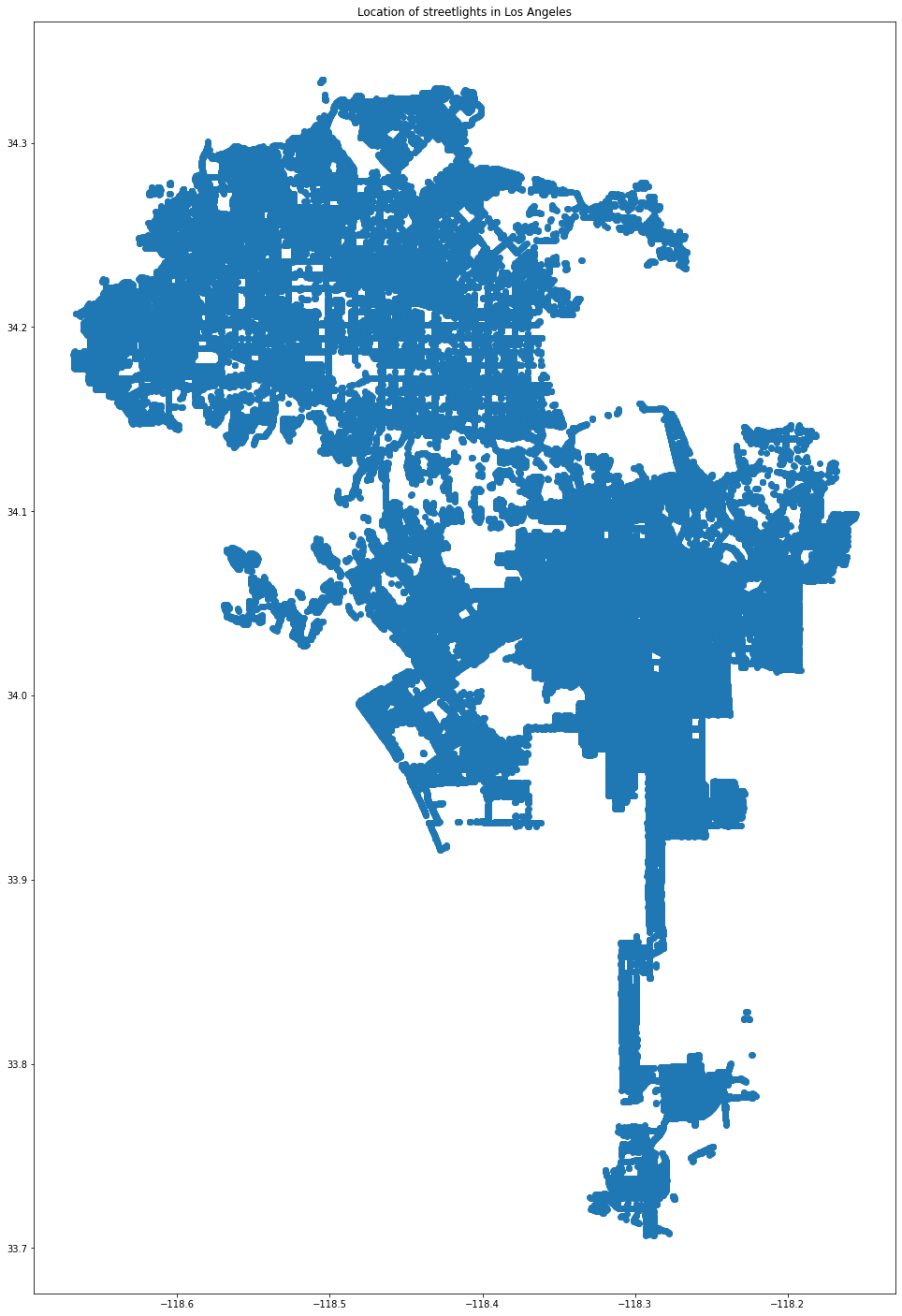


Figure 1. Spatial location of streetlights in Los Angeles

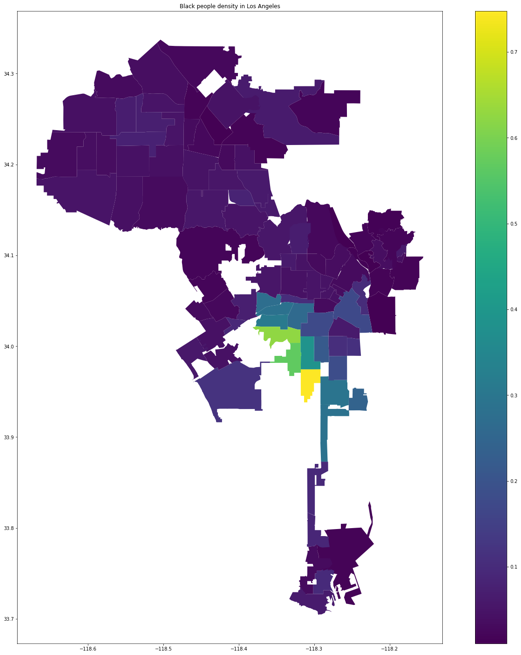


Figure 2. Black people density in Los Angeles

Figure 1 indicates the location of streetlights in Los Angeles. Figure 2 is the black density map which is plotted based on different neighborhood boundaries. For the black density map, we can see that areas around central Los Angeles have the higher value of black density. Since there are 220029 rows in crime data, I will only plot the number of crimes on different neighborhood boundaries in the future analysis.

**Methodologies:**

There are three methodologies that I would like to use in this project. Firstly, I would like to use the global Moran’s method to evaluate whether there is a spatial relationship in terms of the number of crimes, the number of streetlights, and the black density or not. Secondly, I will fit a line model to the data (the number of crimes and the number of streetlights) and a polynomial of second degree. In the meantime, I want to split the sample into training and testing data and evaluate the model on both sets. Then I will make an influencing plot for each of the 2 models and compare the models with the Likelihood Ratio test(set alpha to 0.05). Lastly, I will build linear model by adding black density as the second variable and create two separate models: random forest regressor and Stochastic Gradient Descent. These two models can then be compared based on their mean square error (MSE) and mean absolute error (MAE) on a test set. We will also be able to extract feature importance after the models are created to see what influenced their decisions. For the last step, there is still a possibility to add new variables based on the data analysis results in the first two steps.

**Deliverable:**

For the step 1, if Moran's I coefficients are positive and statistically significant at a p<0.05, it indicates that there is positive autocorrelation. The largest coefficient is indicating the most significant spatial correlation. For the step 2, a significance threshold of 5% will be assumed through. If p value is less than 5%, it indicates there is a significant relationship between the number of streetlights and the number of crimes. R2 and Adjusted R2 can be used to see which model fits better to the data. For the likelihood ratio test, the Null Hypothesis is that the more complex model is better than the simpler one. If the likelihood ration statistics is less than 0.05, it means this corresponds not being able to reject the Null Hypothesis at alpha 0.05. If it is larger than 0.05, it will be able to reject the Null Hypothesis at alpha 0.05. Thirdly, in terms of the random forest regressor and the Stochastic Gradient Descent, we need to test whether there is a overfitting problem first. According to the mean squared error (MSE) and mean absolute error (MAE) on the training and testing data (test size = 0.25), if the model performance on testing set is relative higher than model performance on training set, this might have overfitting problem. If there is limited or no overfitting problem on these two models, we can compare the value of the mean squared error (MSE) and mean absolute error (MAE) for testing data and the lower value indicates better performance.

***Link to GitHub repo:***

https://github.com/lanyu0322/PUS2020\_LYu/blob/master/Midterm/Midterm\_read\_data\_Lan.ipynb

**Bibliography:**

[1] bsl.lacity.org. (n.d.). *City of Los Angeles - Bureau of Street Lighting*. [online] Available at: https://bsl.lacity.org/ [Accessed 20 Oct. 2020].

[2] Farrington, D.P. and Welsh, B.C., 2002. *Effects of improved street lighting on crime: a systematic review*. London: Home Office.

[3] Gabbidon, S.L. and Greene, H.T., 2018. *Race and crime*. Sage Publications.

[4] Geohub.lacity.org. 2020. *City Of Los Angeles Hub*. [online] Available at: <http://geohub.lacity.org/> [Accessed 20 October 2020].

[5] Howell, J.C., 2015. *The history of street gangs in the United States: Their origins and transformations*. Lexington Books.

[6] Machin, S., Marie, O. and Vujić, S., 2011. The crime reducing effect of education. *The Economic Journal*, *121*(552), pp.463-484.

[7] Murray, A.T. and Feng, X., 2016. Public street lighting service standard assessment and achievement. *Socio-Economic Planning Sciences*, *53*, pp.14-22.

[8] Perkins, C., Steinbach, R., Tompson, L., Green, J., Johnson, S., Grundy, C., Wilkinson, P. and Edwards, P., 2015. What is the effect of reduced street lighting on crime and road traffic injuries at night? A mixed-methods study.

[9] Prabakaran, S. and Mitra, S., 2018, April. Survey of analysis of crime detection techniques using data mining and machine learning. In *Journal of Physics: Conference Series* (Vol. 1000, No. 1, p. 012046). IOP Publishing.

[10] Raphael, S. and Winter-Ebmer, R., 2001. Identifying the effect of unemployment on crime. *The Journal of Law and Economics*, *44*(1), pp.259-283.

[11] Rushton, J. P., & Templer, D. I. (2009). National differences in intelligence, crime, income, and skin color. *Intelligence*, *37*(4), 341-346.

[12] Sampson, R.J., Wilson, W.J., Hagan, J. and Peterson, R.D., 1995. Toward a theory of race, crime, and urban inequality. *1995*, pp.37-54.

[13] Welsh, B.C. and Farrington, D.P., 2008. Effects of improved street lighting on crime. *Campbell systematic reviews*, *4*(1), pp.1-51.

[14] Xu, Y., Fu, C., Kennedy, E., Jiang, S. and Owusu-Agyemang, S., 2018. The impact of street lights on spatial-temporal patterns of crime in Detroit, Michigan. *Cities*, *79*, pp.45-52.