

# Spatiotemporal Analysis of London Crimes

GABRIELLA PAULINE  
DJOJOSAPUTRO  
School of Information Systems, Singapore  
Management University

TAN HUIYONG  
School of Information Systems, Singapore  
Management University

TOH MEI SIEW  
School of Information Systems, Singapore  
Management University

## ABSTRACT

With the limited police resources and possible adverse impact when crime occurs, crime analytics is required. Crime occurrence was found to have spatial patterns, and predictive analytics should be possible. However, mixed results were obtained in the research to determine whether predictive policing results in lower crime rates. Hence, this project aims to use the analytics to determine areas with higher risk of crime and to discover the underlying factors to the increased risk in the form of R Shiny app. Literature review was performed to improve on the existing methodologies and the project focuses on analysing crimes in London using Geographically weighted regression (GWR) and Spatiotemporal clustering, supported by Exploratory Data Analysis (EDA) and Exploratory Spatial Data Analysis (ESDA). The result from GWR suggests a possibility of spatial patterning during visualizing the residual of the linear model through a choropleth. Boroughs with over-prediction (negative residuals) are neighbouring boroughs, suggesting a possibility of spatial patterning. Conversely, spatiotemporal clustering is able to point out the unique features of each cluster that are derived.

This project hopes to improve the analysis in the future. Smoothing, as explored in ESDA, on crime rate can be used to improve the result for clustering as it could smooth out outliers. This project also hopes to perform GWR modelling on the ward level once demographics datasets are available. The gwss function can be further explored to understand the statistical significance of each variable on the local level.

## 1. INTRODUCTION

Crime datasets are made available on the ward and borough level of London whereas, population datasets and demographic datasets are available on the borough level. There are a total of 32 boroughs in Greater London, excluding the

City of London. Using the geo-spatial polygons, the datasets are consolidated to develop a Shiny application which aims to provide users actionable insights based on the following analyses:

- Exploratory Data Analysis (EDA) of crime rate over time
- Visualisation of crime rate across London
- Exploratory Spatial Data Analysis (ESDA) of hot spots of crime rate, with a visual map of Greater London
- Clustering of boroughs based on different technique
- Geographically weighted regression (GWR) to support predictive analysis

With data-driven insights, it can further inform preventative measures such as warnings and allocation of police force resources and influence ward planning policies.

## 2. MOTIVATION AND OBJECTIVES

With the limited police resources and possible adverse impact when crime occurs, analytics on crime has been done as far back as in the 1800s [6]. Crime occurrence was found to have spatial patterns, and thus predictive analytics should be possible. However, mixed results were obtained in the research to determine whether predictive policing results in lower crime rates [12]. Thus, it is more beneficial to use the analytics to determine areas with higher risk of crime and to discover the underlying factors to the increased risk.

Traditionally, crime analysis is done manually or through a spreadsheet program [15]. It is of interest for this project to better understand the theories and practical application of such analysis. Thereafter, this project would provide the users an easier way to do the analysis using a web application.

## 3. REVIEW AND CRITIC ON PAST WORKS

This project draws inspiration from a previous work by students of SMU. In Simple Geo-Spatial Analysis Using R-Shiny [9], the work covered a wide range of analysis, including EDA, ESDA, clustering and GWR.

The ESDA is presented in a very clear and concise manner and the user has a wide variety of variables to pick from.

However, users will not be able to select multiple variables. While the colours reveal the underlying spatial clusters and outliers, it lacks quantitative data at borough level visualization. Furthermore, the visualization lacks comparison of the analysis with the pristine data and this diminishes the user's appreciation of the analysis and they are unable to observe the visual difference.

On the clustering tab, the `dtwclust` package provides a sample Shiny app that can be accessed using interactive clustering function or in `shinyapps.io` website. This app provides a way to get the time series by cluster visualisation easily in a faceted plot after taking the clustering arguments from the user. However, it does not provide a way to evaluate the clusters formed visually other than using the time series plot itself. The metrics are provided in a text format, and it can only be computed for the specified number of clusters in the initial run. To see the optimum number of clusters, the users need to do trial and error using different number of clusters and manually compare whether the value for the evaluation metrics increase or decrease. Besides, because this package is more focused on time series clustering, the geo-spatial relation in the cluster members is not shown in the Shiny app.

For GWR, users have to run the correlation analysis first to select the variables and re-select it on the GWR tab. An improvement will be to allow variables selected in the EDA tab to be automatically reflected in the GWR tab to reduce duplicated steps. With no action buttons, data is updated as users make changes to the input which is not ideal. Based on the initial results, users are not able to make multiple changes at once since the results will be constantly updated with each change. Lastly, other than using the GWR basic model, there are alternative functions available, such as `gwr.mixed` and `gwr.multiscale` to explore.

## 4. DESIGN FRAMEWORK

### 4.1 DATA PREPARATION

All the data wrangling and cleaning is performed using R. Crime data is extracted from 2 different datasets (2011-2018 and last 24 months) and consolidated to show the number of crimes for each crime type from 2011 to 2019. Instead of analysing the absolute number of crimes across boroughs, population data per borough is used to calculate crime rate (number of crimes per 100,000). The data is joined with other demographic datasets and polygonal shape data via the LAD shapefile from the London Datastore.

### 4.2 EXPLORATORY DATA ANALYSIS

The first stage of data exploration is to understand the crime rate across boroughs by crime type and year. Our shiny application provides the following for users to have a basic understanding of the crime data

Tab 1 provides a visualisation across space, with 2 choropleth maps side-by-side and showing crime rate based on 2 sets of users input (year, crime type). Tab 2 provides a visualisation over time based on the crime type, allowing users to understand the underlying trend and plausible seasonal patterns.

## 4.3 EXPLORATORY SPATIAL DATA ANALYSIS

The second stage of data exploration of spatial data should be performed using the Exploratory Spatial Data Analysis (ESDA) tools. These techniques allow spatial distributions, identify spatial outliers, discover patterns of spatial association, clusters or hot spots to be described or visualised [3].

### 4.3.1 ESDA METHOD

Local Indicators for Spatial Association (LISA), like local Moran's I is applied to detect clusters and/or outliers from the crime rate across the Boroughs in London. LISA provides statistics that evaluate the existence of clusters in the spatial arrangement of a given variable and they enable ESDA to be carried out. LISA also served as indicators of local pockets of nonstationarity, or hot spots, similar to the  $G_i$  and  $G_i^*$  statistics of Getis and Ord and to identify "outliers," as in Anselin's Moran scatterplot [1].

A spatial weighted matrix assigns weights to spatial neighbours, either Contiguity-based or Distance-based neighbours. Moran's I is one measure of spatial autocorrelation that uses the spatial weighted matrix and it is the correlation between two variables  $x$  and  $y$  of length  $n$ . The formula is

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

where  $w_{ij}$  is the weight between observation  $i$  and  $j$ , and  $S_0$  is the sum of all  $w_{ij}$ 's [14]

Conversely, Getis and Ord's  $G$ -Statistics distinguishes between hot and cold spots by identifying spatial concentrations. This statistic looks at neighbours within a defined proximity and the  $G_i$  value would be high if high values cluster together, vice-versa. The formula is

$$G_i^* = \frac{\sum_{j=1}^n w_{ij} x_j - \bar{X} \sum_{j=1}^n w_{ij}}{S \sqrt{\left[ \frac{n \sum_{j=1}^n w_{ij}^2 - \left( \sum_{j=1}^n w_{ij} \right)^2}{n-1} \right]}}$$

where  $x_i$  is the attribute value for feature  $j$ ,  $w_{ij}$  is the spatial weight between feature  $i$  and  $j$ ,  $n$  equals to the total number of features [2].  $G_i$  Statistics would utilise a distance-based weighted matrix, either fixed distance or adaptive proximity.

### 4.3.2 ESDA VISUALIZATION

The resulting spatial plots from the analysis would enable outliers and spatial clusters to be observed visually.

To further provide insights to LISA, 'geofacet' package is used. This visualization arranges a sequence of plots of geospatial data into a grid that preserves the original geo-

## 4.4 CLUSTER ANALYSIS

Geo-referenced time series clustering is appropriate for the data because it is a time series of crime rate in different geographical areas in London. It is performed using Dynamic Time Warping (DTW), a method commonly used for clustering time series data, which is available in the dtwclust package. The cluster results will be visualised using a map to see whether the clusters are related with their geospatial proximities to each other.

DTW uses dynamic programming to find the optimum warping paths to compute the distance of two time series [17]. Therefore, one point in time series can be mapped to a different point in time in another time series, as shown by the blue mapping lines in the following figure.

$$lcm(i, j) = \left( \sum_v |x_i^v - y_j^v|^p \right)^{1/p}$$

in the two series. The points in the optimum path are signified using the set  $\phi$ . The distance is equivalent to the sum of all weight in the optimum path ( $m_\phi$ ), normalised by the constant  $M_\phi$

The Shiny app also allows the users to choose several variations of DTW, including:

#### 4.4.2 CLUSTERING TYPE

### 4.4.3 CENTROIDS

Centroid is the central point of each dimension of each cluster, which means it is also a time series. Centroids affect how the items are assigned to the clusters. Available options in the Shiny app are:

Option	Description
Mean	Average along each dimension
Median	Median along each dimension
Shape Averaging	Average shape according to optimal mapping
DTW Barycenter Averaging	Evolving average between iterations
Soft DTW Centroids	Centroid calculation based on soft DTW
Partition Around Medoids	Using one of the time series in the data as centroid
Fuzzy c-means	Fuzzy version of means
Fuzzy c-medoids	Fuzzy version of medoids

#### 4.4.4 CLUSTER SIZE

Cluster size is one of the inputs required to do clustering, but the optimum number of clusters is rarely known beforehand because it is an unsupervised technique. To aid in determining the optimum number of clusters, visual analytics is used. For hierarchical clustering, cluster dendrogram can be used to evaluate the distance between and within the clusters, so the users can choose an optimum point to cut the tree.

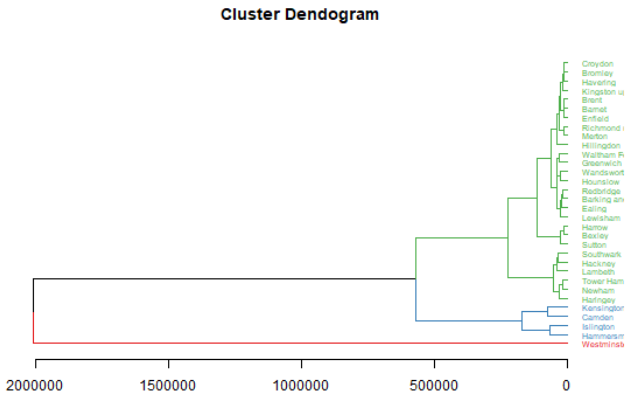


Figure 3: Hierarchical Clustering Dendrogram

Cluster Validity Indices (CVIs) are evaluation metrics to determine the purity or correctness of the clusters if ground truth is known [18]. Using a plot to visualise the metrics for different numbers of clusters, the users can easily update the number of clusters to their choice. Silhouette coefficient is used for partitional, which is to be maximised. Meanwhile, modified partition coefficient (MPC) is used for fuzzy clustering, in which higher values also correspond to better cluster purity.

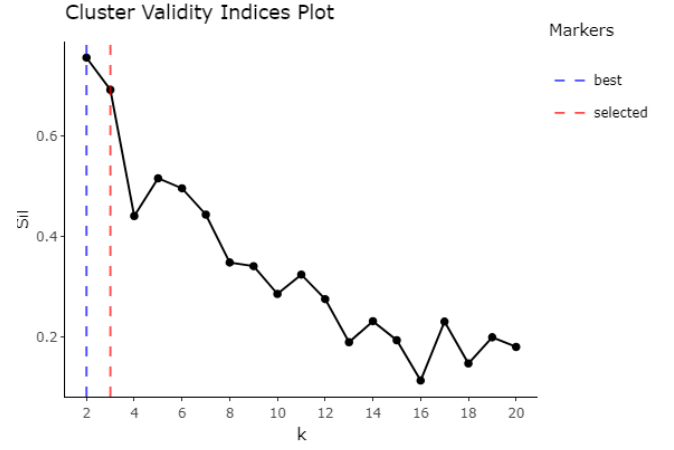


Figure 4: Cluster Validity Indices Plot

## 4.5 GEOGRAPHICALLY WEIGHTED REGRESSION

Geographically weighted regression (GWR) is used to explore spatial heterogeneity in the relationship between variables and can be found within the GWmodel R packages. A GWmodel is applicable for use cases where spatial data are not well described by the global form and for some regions, the localised fit provides a better description [5].

A linear regression model utilises spatial data as a stationary process whereby the same model applies across the entire study area [4]. On the other hand, GWR incorporates the output and predictive variables of each region that fall within the neighbourhood of each target region. A matrix of weights specific to each region where regions nearer to the target region are given greater weight than regions further [4]. As such, the parameters estimates are local to each region, instead of a global constant across the study area.

There are several factors involved in building a GWR model [8]. They are i) bandwidth, ii) kernel used and lastly iii) p, which is the power of the Minkowski distance.

### 4.5.1 BANDWIDTH

Bandwidth is the distance (in metres) or the number of neighbours used for each local regression model. It can be either calculated as a fixed distance or adaptive based on the number of nearest neighbors. Bandwidth can be calculated using the gwr.bw function in the GWmodel R packages.

The adaptive bandwidth option adapts the number of neighbours required based on the density of the study area whereas a fixed bandwidth stays constant across each region in the study area. Depending on the make-up of the study area, an adaptive bandwidth is useful for study areas with uneven density in regions [13]. Bandwidths can be calculated via 2 approaches to determine the stopping rule: Cross-validation (CV) approach and Akaike Information Criterion corrected (AICc) approach.

The Shiny application allows users to choose between fixed

and adaptive bandwidth and the 2 approaches.

#### 4.5.2 KERNEL

Besides bandwidth, the kernel selection also influences the weighing matrix. The weights are determined using a kernel, which is a distance decay function that determines how quickly weight decreases as distance increases.

The Shiny application allows users to choose between the following kernels: Gaussian, Exponential, Bisquare, Tricube and Boxcar. The functions are as followed:

```
gaussian: wgt = exp(-.5*(vdist/bw)^2);

exponential: wgt = exp(-vdist/bw);

bisquare: wgt = (1-(vdist/bw)^2)^2 if vdist < bw, wgt=0 otherwise;

tricube: wgt = (1-(vdist/bw)^3)^3 if vdist < bw, wgt=0 otherwise;

boxcar: wgt=1 if dist < bw, wgt=0 otherwise
```

Figure 5: Kernel Functions of GWmodel

#### 4.5.3 MINKOWSKI DISTANCE

Distance is calculated using the Minkowski distance -  $p$  and there are 2 options - Euclidean distance with  $p=2$  and Manhattan distance with  $p = 1$ .

#### 4.5.4 TYPES OF GWMODEL

In the GWmodel package, it includes a variety of modeling methods, including gw.basic, gw.mixed and gtwr. This Shiny application allows users to explore the differences between the parameter estimates of a gw.basic and a gw.mixed model.

A gw.basic model is the basic model found in the GWmodel packages. On the other hand, a gw.mixed model implements a mixture of semi-parametric variables. This allows some predictor variables in the formula to be treated as global, instead of local.

In this Shiny application, users are able to select a variable to be treated as global and compare the difference in the parameter coefficients results. Typically, using gwss in the GWmodel package, the Monte Carlo approach can test for significant spatial variation in each coefficient based on the basic GW model [10]

The Monte Carlo approach is out of scope in this Shiny application and to be included as part of future works.

#### 4.5.5 OTHER PACKAGES

There are alternative packages utilised for geographically weighted regression, such as the spgwr, gwrr and McSpatial packages. The functions of these packages are similar to GWmodel.

An advantage of GWmodel is the ability to provide 5 kernel functions, whereas spgwr can only provide 3. In terms of GW summary statistics, it provides a robust statistics summary with median, inter-quartile and range [5]

	GWmodel	spgwr
<b>Kernel Functions</b>	Box-car, Bi-square, Tri-cube, Gaussian, Exponential	Bi-square, Tri-cube, Gaussian
<b>Statistics Summary (Basic)</b>	Yes, with gw.ss Mean, Std. deviation, variance, skew, coeff of variation, covariance, Pearson's correlation	Yes, with gw.cov Mean, Std. deviation, std. error of mean, std.difference of global mean, std. difference of local mean, covariance, skew, Pearson's correlation
<b>Statistics Summary (Additional)</b>	Median, interquartile range, quantile imbalance, Spearman's rank correlation	No

Figure 6: Comparison between GWmodel and spgwr

## 5. DISCUSSIONS

### 5.1 ESDA

The map plot for the pristine crime rate shows that the Westminster area has the highest crime rate.

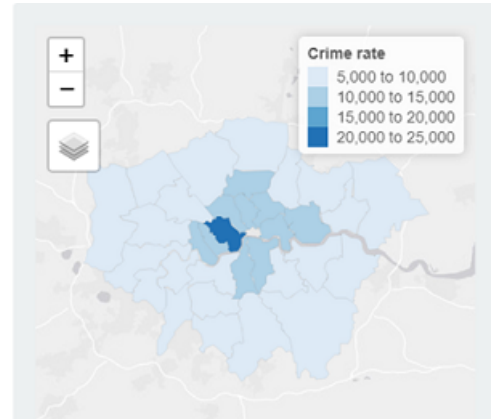


Figure 7: Simple spatial plot of crime rate per 100,000 residents in London

Nonetheless, when Moran's I and Gi statistics were compared, it became visually clear that the spatial clusters and outliers are quite different from the raw data; the Gi statistics is able to show the "hot" and "cold" spots of crimes while Moran's I shows the local outliers.

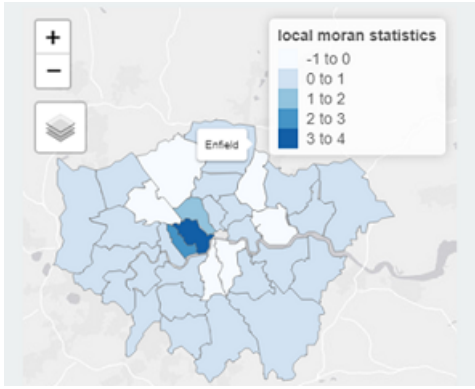


Figure 8: Gi value spatial plot of crime rate per 100,000 residents in London

## 5.2 CLUSTERING

To examine the geospatial proximities of cluster members, a map filled with different colours for each cluster assignment is used. For partitional clustering of Theft and Handling crimes using basic DTW distance metrics and partition around medoids as the centroid, the cluster assignment seems to be correlated with the boroughs' geospatial location. Cluster 1 and 2 is in the center of London, with Cluster 2 in the west side of London City. Meanwhile, Cluster 3 is in the outskirts area.

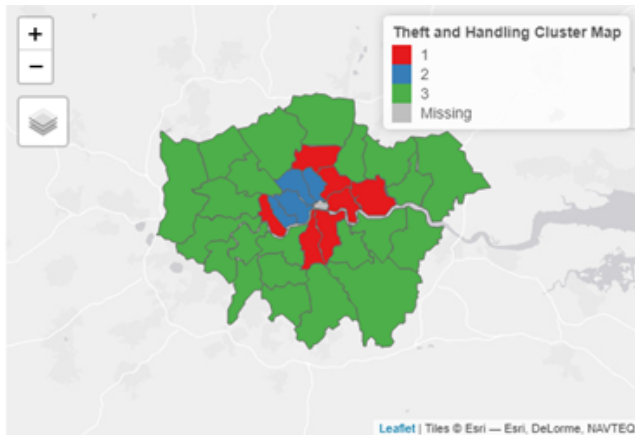


Figure 9: Spatial cluster of DTW distance metrics

The time series of the data can be used to describe the

characteristics of each cluster. It is faceted by the cluster assignment to provide clear separation between the clusters. Cluster 2 has the highest crime rate, with a decreasing trend from 2012 that reverted in 2016. Cluster 1 and 3 have a relatively flat trend, but Cluster 1 has a higher crime rate than Cluster 3.

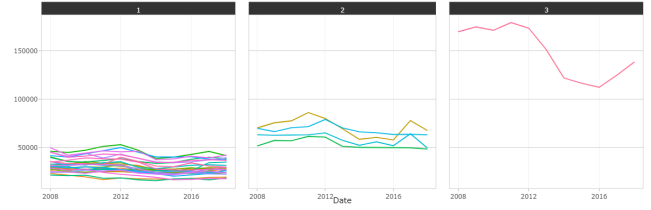


Figure 10: Faceted time series of Cluster 1 to 3

## 5.3 GEOGRAPHICALLY WEIGHTED REGRESSION

For 2017 crime rate on theft and handling, the 3 variables - population density, unemployment of White UK citizens and mean income are able to explain 62% of the crime rate across the 32 boroughs.

The boroughs with the top 20% percentile in crime rate are neighbours while boroughs with the bottom 20% of crime rate are concentrated mainly in the south.

### 2017 Crime Rate

for Arson and Criminal Damage, Theft and Handling

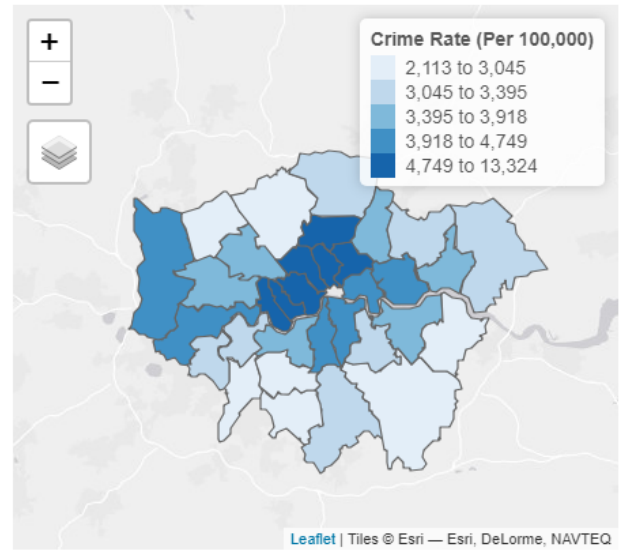


Figure 11: Choropleth of crime rate in London

An initial investigation will be visualising the residual of the linear model through a choropleth. Boroughs with over-prediction (negative residuals) are neighbouring boroughs, suggesting a possibility of spatial patterning [11]. Moran's I test rejected the null hypothesis that the residuals are randomly distributed.

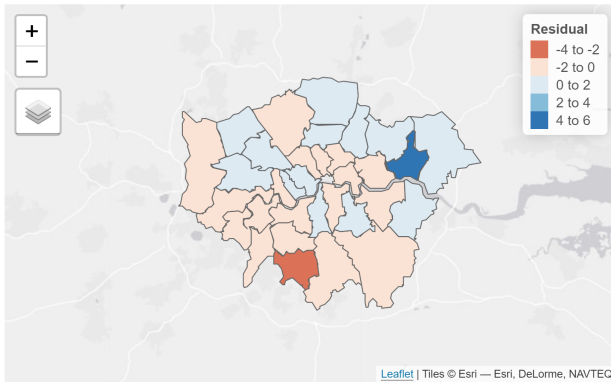


Figure 12: Choropleth of residual

With the GWmodel,  $R^2$  has improved slightly from 62% to 62.7%, showing that the localised model is able to explain crime data better.

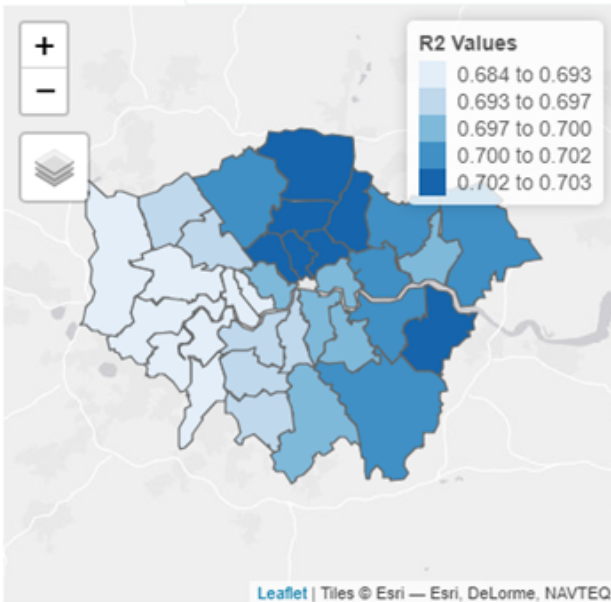


Figure 13: Improvement of  $R^2$  values after applying GWR

## 6. FUTURE WORKS

For geographically weighted regression, the lack of ward level data points, including population, limits the availability of crime rate data. In future, GWR modeling should be performed on the ward level once demographics datasets are available on the ward level. The gwss function can be further explored to understand the statistical significance of each variable on the local level.

The clustering can be improved by applying smoothing on the crime rate. Even though the rate has been normalised with the population, some boroughs still have significantly higher crime rates than others. Empirical Bayes smoothing

technique that is used in the ESDA portion could be used as a preprocessing function to smoothen the outliers before clustering. Additionally, comparison with other packages such as spatimeclus or tsclust can be performed.

## 7. ACKNOWLEDGMENTS

The authors would like to thank Professor Kam Tin Seong of Singapore Management University for his valuable advice and support on this project.

## References

- [1] Anselin, L. 1995. Local Indicators of Spatial Association—LISA. (Apr. 1995), 93–115.
- [2] ArcGis Pro How Hot Spot Analysis (Getis-Ord Gi\*) works. Esri software.
- [3] Dall'erba, S. 2009. Exploratory Spatial Data Analysis. *International Encyclopedia of Human Geography*. (Jul. 2009), 683–690.
- [4] Fotheringham, A.S. and Rogerson, P.A. 2009. *Geographically weighted regression*. Sage Publications, Inc.
- [5] Gollini, I. et al. 2015. GWmodel: an R package for exploring spatial heterogeneity. 63, 17 (Feb. 2015), 1–50.
- [6] Hunt, J. 2019. From Crime Mapping to Crime Forecasting: The Evolution of Place-Based Policing. 1, 281 (Apr. 2019).
- [7] Jason, L. 2019. What Is the Difference Between Hierarchical and Partitional clustering? Medium.
- [8] Kam, T.S. 2019. Hands-on Exercise 9: Building Hedonic Pricing Model with GWR Method. Exer9.
- [9] Li, J.D. et al. 2020. Simple Geo-Spatial Analysis using R-Shiny.
- [10] Lu, B. et al. 2013. The GWmodel R package: Further Topics for Exploring Spatial Heterogeneity using Geographically Weighted Models.
- [11] Medina, J. and Solymosi, R. 2019. Crime Mapping in R. Medium.
- [12] Meijer, A. and Wessels, M. 2019. Predictive Policing: Review of Benefits and Drawbacks. *International Journal of Public Administration*. 42, 12 (Feb. 2019), 1031–1039.
- [13] Meik, J. and Lawing, M. 2017. Considerations and Pitfalls in the Spatial Analysis of Water Quality Data and Its Association With Hydraulic Fracturing. (Jan. 2017), 227–256.
- [14] Paradis, E. 2020. Moran's Autocorrelation Coefficient in Comparative Methods.
- [15] RAND Corporation 2014. *Predictive Policing*. RAND Corporation.

- [16] Santos, R.B. 2016. Crime analysis with crime mapping. Sage Publications, Inc.
- [17] Sardá-Espinosa, A. 2019. From Crime Mapping to Crime Forecasting, The Evolution of Place-Based Policing.
- [18] Sardá-Espinosa, A. 2019. Package ‘dtwclust.’
- [19] Wang, W. and Zhang, Y. 2007. On fuzzy cluster validity indices. 158, 19 (Oct. 2007), 2095–2117.