

Visualising London's Crime Data: Uncovering crime's seasonal relationship and patterns

HO Jue Hong, Jazreel THO Wei Wen, NGOH Yi Long

Abstract—Over the years, violent crime rates have been increasing at an alarming rate, from approximately 140k violent crimes in 2010 to 200k in 2017(City of London & Greater London). Both police forces, Metropolitan Police and the City of London Police have aimed to try reducing the number of crimes but to no avail. Hence, this paper would like to help shed light and find the different correlation between different crimes in different seasons. Furthermore, we would like to find out the danger level in different neighbourhoods and to help visualise the police station around each neighbourhood in the aid of alleviating this growing problem.

1 INTRODUCTION

Crime rates have remained high over the years in London (Fig. 1).

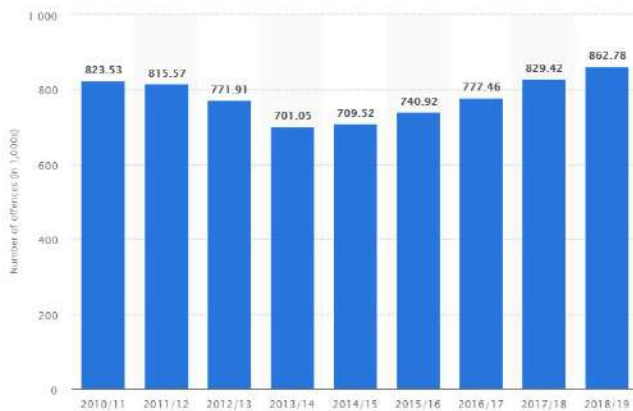


Figure 1: Increasing crime rates in London

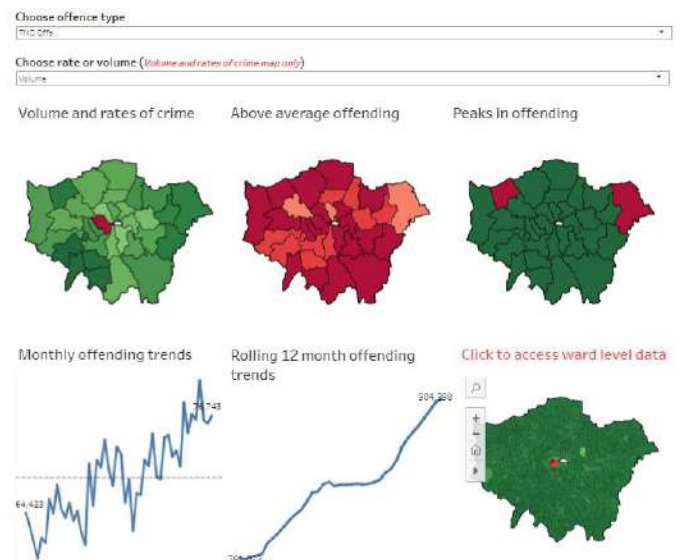
From the year 2010/11 it has even increased from around 823,000 to 862,000 by year 2018/19. Number of reported crimes have increased by 2.2% from last year but it is at an alarming increase of 20% from 2014/15. There has been police budget cuts in recent years and this has resulted in a decrease in the number of police officers in the service. Even police officers themselves are feeling unsafe to be patrolling and do stop and search on people on the streets. Keeping these points in mind, we are hoping to be able to visualize the crime hotspots in London. In addition, mapping the police station locations to compare the crime rates and police surveillance to have a better overview on areas which may require increased security. Related works used as inspiration for the project. In Section 3, we will discuss some of the data collection and preparation efforts for the project.

With available dataset being able to provide details of the individual crimes committed, it will be a good visualising aid to be able to understand these data through the use of dashboards. Hence, this paper will be presenting our idea of an interactive visualisation platform, specially designed to enable users to have a better understanding of their crime statistics down to their neighbourhood level. This research paper consisting of X main sections, documenting our thought process for the dashboard designs and flow, which we hoped to present an interactive yet informative platform for users to have a clearer picture of the London's crimes from the year 2017 to 2019. Section 1 and 2 will have details of our motivation and also starting research of design ideas to implement into our own dashboard. Section 3 documents our data collection and preparation process to achieve a cleaned dataset for us to implement with our dashboard. Section 4 details our walkthrough of the designs

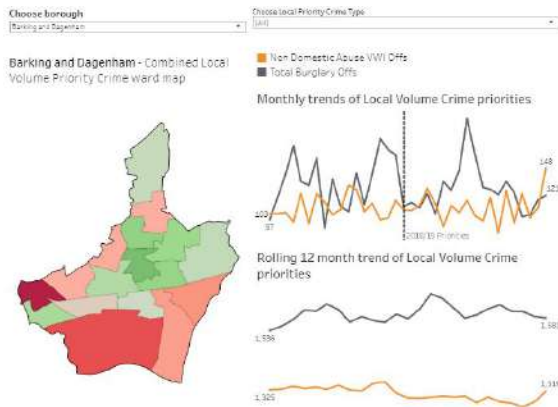
we will be implementing on our dashboard. Section 5 and 6 states the our application architecture where we showcase the different libraries to create multiple visualisation that we have implemented into our dashboard. Section 7 shows some of the analytical results that can be achieved from the usage of our dashboards. Lastly, section 8 and 9 will conclude some of our learning points from creating this dashboard and also some possible directions this project can take on in future for further implementations.

2 RELATED WORKS

Before starting on our visual analytics project, we researched on current analysis that has been done by the UK government. The UK government has uploaded a crime dashboard online that tracks local crimes by borough and ward. There are 3 analysis tab: Track local crime priorities, Borough volume and trends and ward local level priorities.



The above is the Borough Volume and Trends that shows an overview of the crime dashboard by borough uploaded by the UK government. There are 6 charts being shown: Crime rate across boroughs, above average offending, peaks in offending choropleth map, monthly offending trends line chart, rolling 12 month offending trends line chart and ward level choropleth map. Users are able to select the offence type and metrics to analyse the boroughs.



Users are then able to drill down to the ward level, which is similar to what we are analysing for our project - analysing crime data in London wards/neighbourhoods. There are 3 charts - choropleth map that shows the crime rate in the selected borough, broken down into wards, Monthly trends of local volume crime priorities and rolling 12 month trend of Local Volume Crime Priorities.

Fig. 1. Yelp Insights : Data Visualization by Shirley Lei

3 DATA COLLECTION AND PREPARATION

3.1 DATA COLLECTION

Data collection is an important step before we are able to proceed to creating our dashboard. We started by searching for data which consists of the London crime types and coordinates of the actual crimes. We managed to retrieve this dataset for the year 2017 to the most current from data.police.uk [1]

Our planned dashboard was to have a map which we are able to break it down into its neighbourhoods, we sourced for Shapefiles [2] which would enable us to plot London's geographical landscape.

The last dataset we will need to collect would be the coordinates of the police station within London [3] as to plot it onto our map to allow better understanding the police presence around each of the neighbourhoods.

3.2 DATA PREPARATION

After getting the relevant data that we needed to work with, we had to prepare our data so that we are able to utilise it into our dashboard. To plot the crimes onto the map on our dashboard, we have to map the crime data to the coordinates of the neighbourhoods, then aggregate the number of crimes in each neighbourhood. This would allow us to create a heatmap to show the density of the crimes in each neighbourhood.

Before we could join our London data crime file with our spatial files, we had to convert our data to a Simple Features (sf) object to be used as a geospatial object to be manipulated for our dashboard. Converting the data crime file by its longitude and latitude required us to convert the Coordinate Reference System (CRS) information to align with our London ward shapefile. .

We planned to have a dashboard which shows the seasonal changes of the crimes in each neighbourhood. The date column of the data crime file was in a 'yyyy-mm' format, therefore we added a 'YEAR'

and a 'MONTH' column to the data. This would allow us to split the dataset by months or year if we needed to help in our analysis.

After we have prepared the data, we offload the relevant data into Comma-separated Values (csv) files so that we were able to read the prepared data straight into our R Shiny application without having to wait a longer loading time when operating our application. This would also enhance the overall performance of reactivity as it would take a shorter loading time to render the results on the user interface as the user toggles with the panels.

3.3 LIMITATIONS

There are a few limitations that we faced throughout the project. Firstly, we weren't very comfortable with .shp files because we are unfamiliar with this new concept, it took us quite awhile to understand how a geometry point is made with its polygon boundaries. Coupled with the limited examples and explanation on the web, we were unsure of how to plot the files via Rstudio. Though there were well-made libraries in R, we were unable to fully show the potential of each library used even with the proper documentation of each libraries because there were close to no examples that show us how to use each function in the specified library.

Secondly, the compiler error was very vague, with each line of error given we had to spend hours to resolve a single problem. We did try to Google the error message to resolve the problem, but we quickly realised that a lot of other users face similar problems without a concrete solution to resolve the problem. Hence through guesswork and trial an error, we are able to resolve the given problems that we faced. It was really tedious and tough because of these errors.

All in all, we were inexperienced and had close to zero experience with R, hence the data preparation and data exploration took a very long time and because of that we feel that it was likely and possible of the omission of certain findings that would have been uncovered if not of our inexperience.

4 DESIGN CONSIDERATION

4.1 CHOROPLETH HEATMAP

Before the selection of Heat Map, we researched various visualization techniques and ideas by using different libraries in R studio to plot a suitable map. Preliminary ideas included density heatmap, hexagon binning of 1 kilometer size to try and visualize the spread of the given crime rates. The use of such visualization method does not conform to our given principles which was the clarity of the data being displayed.

Example 1:



From the given image we are able to see how clustered the data points are hence it will impact the clarity of the plot, and users might not be able to understand what we would be trying to drive across.

Figure 1: Individual Crime Plot

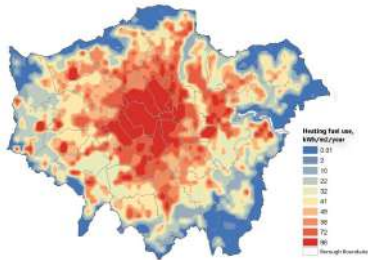
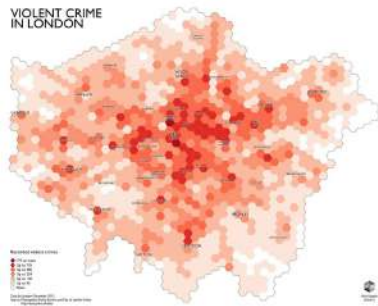


Figure 2: Density Heat Map neighbourhood in the borough. Hence we decided against using the density heatmap as it does display the insights of the given data.

Example 2:

By using a density heat map it does not bring across important point of crime prevalence in certain parts of London. It will be probable for us to see an entire borough being shaded in red because of the high crime rate in that borough but in fact, there might be just spots in the borough and there might actually be safe neighbourhood in the borough.



Example 3:

The last and final example would be the thought of using hexagon binning of a certain radius in the given plot. The problem with this plot was that we already had each neighbourhood .shp file hence by binning data in such a way it does not display the intended Figure 3: Hexagon

Binning data correctly. The bins would cross each territory and it might be under a different police jurisdiction and hence rendering the boundaries completely useless. Any given analysis derived would have little to no actionable action, therefore leading us to scrape the idea of using hexagon binning plot.

In conclusion we decided to use the choropleth heatmap in where the boundaries are set by each neighbourhood boundaries hence displaying a more coherent graph where users are able to understand quickly.



Heat Map

With the data points plotted using choropleth map, we are able to show effective crime rates in each neighbourhood and it does not over generalise the given data or under represent the given crime numbers.

Figure 4: Choropleth

4.2 BAR GRAPH

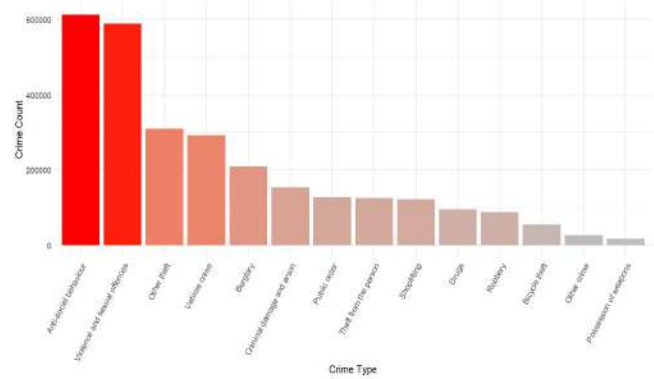


Figure 5: Bar Graph by Overall Crime per Crime type. In the graph above we are representing crime count by each crime type. The bar graph helps us to quickly compare crime counts between the different crime types as it reveals the highs and lows at a single glance. The bar graph is also especially effective when we have numerical data that splits nicely into different categories to identify trends within the given data.

Furthermore, by colourising each bar in the plot we are able to show important information, as the higher the crime count the more alarming it should be hence, us using the colour red to indicate this. In conclusion, the use of the bar graph is the most appropriate as compared to the other charts.

4.3 LINE GRAPH

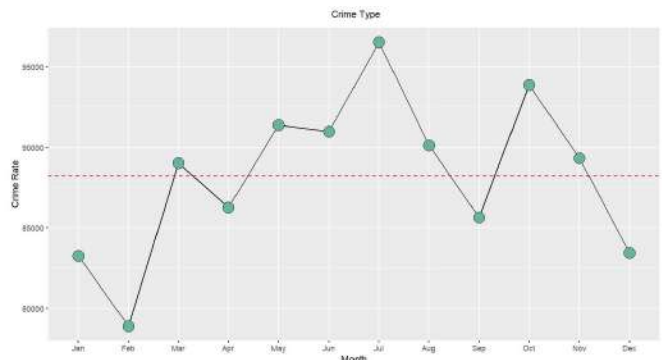


Figure 6: Crime count by months

The line charts are similar to the bar charts as one of the most frequently used charts. The line chart helps us to connect individual numeric data points in this case the sum of crimes in each month. This helps the user to visualize a sequence of values as it displays the trends over a period of time, this is the simplest way and the quickest way the users are able to understand the data shown without having to think too much.

In conclusion we chose to use the line graph to display this series of data to let users absorb information quickly without having to think too much.

4.4 SEASONAL LINE GRAPH

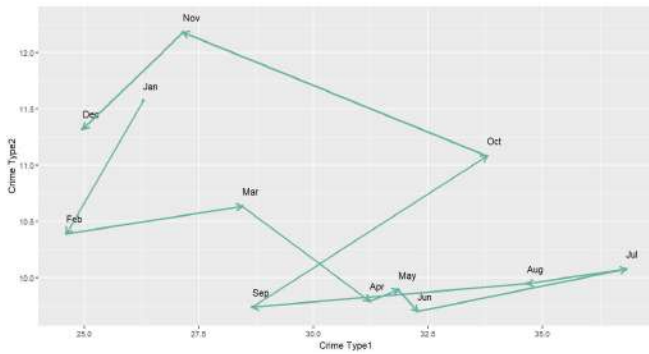


Figure 7: Line Chart for Seasonal Patterns

The graph above showed the correlation between the count of the two crime type with a third additional variable which was the month. It was hard and difficult to plot a graph with 3 different axis and yet at the same time maintaining clarity where users are able to understand the graph without any difficulties. Thus we decided to use the line graph and attempt to display the relationship between the two crimes. Each point represents a certain month and the crime count of each crime type. By joining them in sequence it creates a zig-zag line graph.

Given the line graph's strength we are immediately able to tell the trend of the two given crimes. Crime 1 and Crime 2 has a correlation and higher occurrence in October. Whereas there is a lull period from April to September where the correlation between these two crime time drops, and Crime type 1 occurs more frequently without crime type 2.

In conclusion we chose to use the line graph once again as it is able to display the findings with clarity and precision

4.5 CORRELATION PLOT

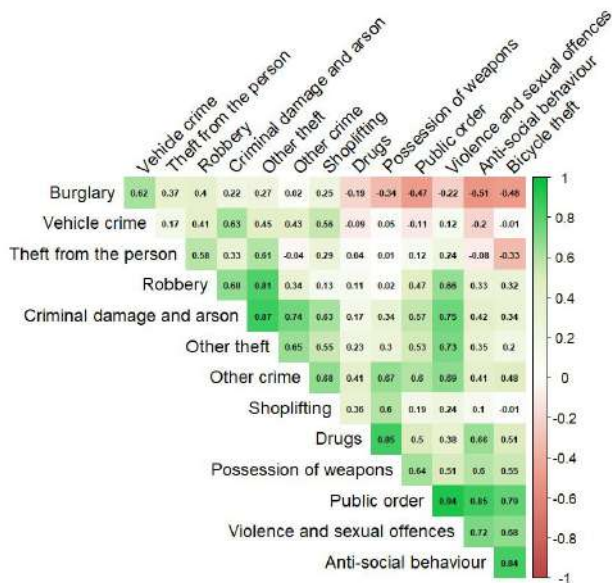


Figure 5: Crime Correlation Plot

The graph above is a called a correlation plot. It shows the correlation between all the crimes and it help users to choose the two different crimes that they would want to further understand. The numbers are the coefficient of correlation, we do understand that

some users do not understand this coefficient, thus we colour coded the box the coefficient is in. With the darker shade of colours showing strong correlation. Red being negatively correlated and Green being positively correlated. The paler colours shows that these crimes occur randomly and have close to no relations.

The colours were chosen in specification to help find positive correlated crimes instead of negatively correlated crimes. Thus green gives people a safe feeling, red gives people the feeling of rejection or danger and white gives people a neutral feeling. Just with the colours of the plot the users are more likely to immediately look at the Greener colours and find out more about the related crimes.

In conclusion, each graph was selected with a high attention to detail and how it would help the user to perceive the relevant information with the greatest ease and clarity.

5 METHODS

5.1 APPLICATION ARCHITECTURE



We developed our dashboard in R, with Shiny framework. It is an interactive dashboard that allows users to analyse by the desired neighbourhood, crime type and time period. In order to transform and merge our data, we used the following packages: dplyr, tidyverse, lubridate, tidyr and sf.

To display the respective charts in our dashboard, we used the following packages - ggrepel, ggplot2, dplyr, tidyr and plotly.

6 APPLICATION DESCRIPTION

6.1 OVERVIEW TAB



Fig 1. Crime Dashboard - Overview tab

The overview dashboard displays 3 visualisations - Crime rate by Neighbourhood choropleth map, Crime rate by Crime Type bar graph and Number of crimes by Months line graph. Users are able to toggle the Date Period, Neighbourhood and Crime Type, whereby all 3 data slicers will be applied on all 3 graph.

The Crime rate by Neighbourhood choropleth map enables users to toggle between map layers via the layer filter on the top left corner of the map. The layers include the 3 basemap layers from Open Street Maps, a layer of aggregated data for the respective neighbourhood data and a point marker layer of the police counters' coordinates. The map is coloured in a manner whereby the darker shade of red a neighbourhood is, the higher crime rate it has. The Crime rate by

Crime Type bar chart shows the total number of crimes for each crime type in descending order. Similar to the choropleth map, the bar chart is coloured in a manner whereby the darker shade of red represents a higher the crime count for the respective neighbourhood.

The number of crimes by Months line graph shows the seasonal changes over months. The green dots are the aggregated average of all years selected for the respective month connected by a black line to show the trend across the months. the red dotted line shows the average crime count across all months. This allows users to gauge which month is above or below average, indicating when the selected crime is more prone to happen.

6.2 CRIME CORRELATION



Fig 2. Crime Dashboard - Overview tab

The overview dashboard displays 2 visualisations - Crime Correlation Matrix and Seasonal relationship line graph.

The correlation matrix displays the relationship among all crime types represented by a value between 1 and -1. As one glance, users can look at the colour of the boxes to determine the relationship between the 2 variables as well whereby a darker shade of green represents a stronger positive correlation, a lighter shade of green represents a weaker positive correlation, a darker shade of red represents a stronger negative correlation and a lighter shade of red represents a weaker negative correlation.

The Seasonal relationship line graph enables users to further analyse if there are any seasonal relationship between any 2 variables. The x axis represents the number of crimes for the first crime type selected and the y axis represents the total number of crimes for the second crime type selected. The arrows are pointed from January to February, all the way till December (depending on the data available) to show the changes from month to month.

7 RESULTS

7.1 OVERVIEW

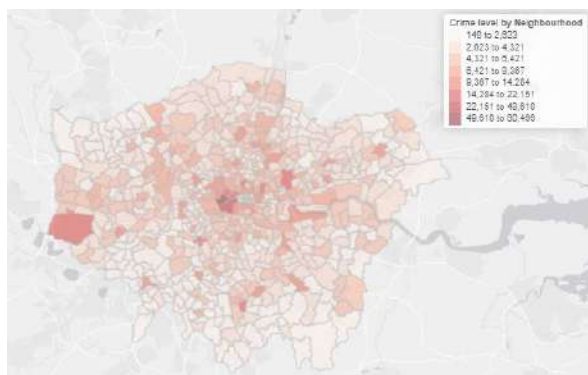


Figure 1: Heatmap of Crime by Neighbourhood

As an overview, in the years from 2017 to 2019, on the western part of London, there is a neighbourhood, Heathrow Villages, which is

high in the level of crime rate. The center part of London, the neighbourhoods surrounding the City of London, are also particularly prone to higher level of crime. As interesting as it may be, the City of London itself has a different territorial police force to the ones in the outer part of London, also known as Greater London.

7.2 ANALYSIS OF POLICE STATION TO CRIME RELATIONSHIP

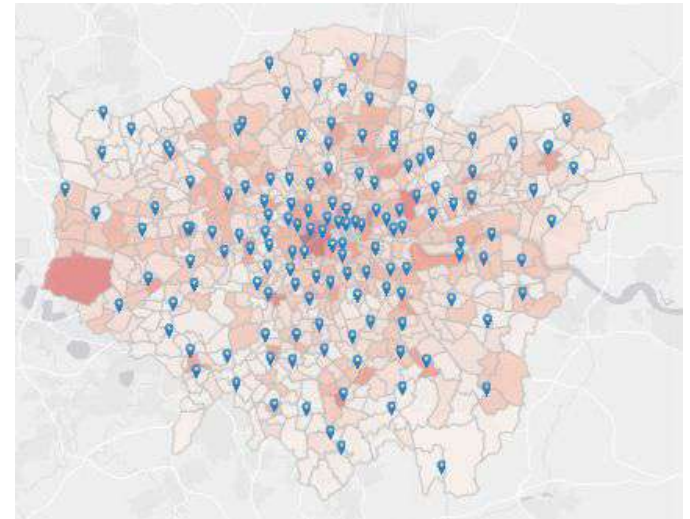


Figure 2: Heatmap of Crime by Neighbourhood with Police counters. We plotted the police station coordinates (blue markers) across London. Immediately, we are able to see that the western part of London, around the neighbourhood of Heathrow Villages, there is a lack in numbers of police stations in that area. Other than that area in London, we are unable to really tell how does the lack of police station would result in less crime as the central London has numerous police station, but yet high in crime rates. Through research, we have found out that although there are only 141 police stations for the residents to seek, London police officers are being dispatched in task forces outside of the physical police stations. The number of police officers that we found are related to the crime level around each neighbourhood. This could be the reason why there is not a strong relationship between the police stations and crime level in London.

7.3 ANALYSIS OF THE DIFFERENT CRIMES IN LONDON

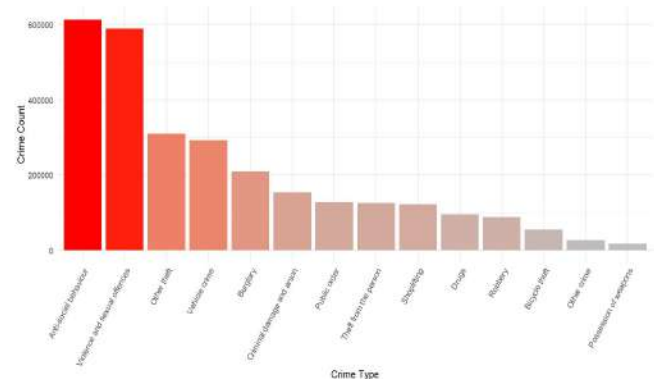


Figure 3: Overall Crime Count per Crime Type

This bar chart would allow us to see the crime counts breakdown in each neighbourhood as the user selects it interactively. From the total crime count seen above, the two crime types that are significantly

higher are the 'antisocial behaviour' and the 'violence and sexual offences' where it consisted of about 600,000 cases each. Both of these crime types are having a huge difference as compared to the other 12 crime types which shows an alarming issue on these two crime types.

7.4 ANALYSIS OF SEASONAL TRENDS

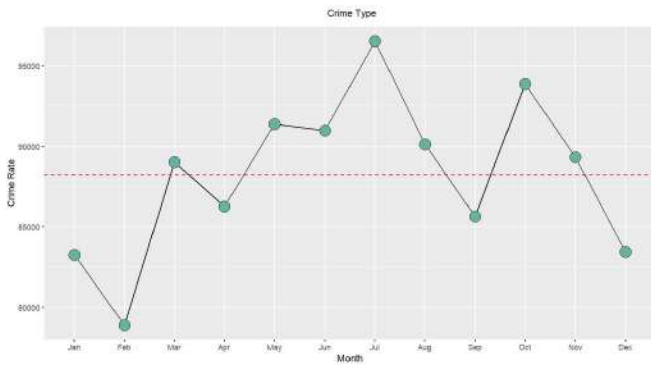


Figure 4: Crime Counts by Months

This line chart is splitting the crimes by months to analyse any seasonal patterns in crimes being committed in London. From the year 2017 to 2019 dataset that we have, it can be seen that in the month of July it has the highest average crime rate from all the neighbourhood in London.

In the month of December, January and February, it has one of the lowest crime rates throughout the whole year round. This could be due to the reason that it is during the winter period, where lesser residents would be out socialising and instead coped up at where. Thus, with lesser interaction, there would be lesser number of crimes during this period.

With this plot, we are able to differentiate the higher crime rate periods to the lower crime rate periods in London, possibly increasing the number of police force in the crime peak periods might help to tone down the crime numbers. This can also be an alert for travellers to avoid high peak crime periods in London. With the interactivity of toggling by the neighbourhood, it can help travellers choose the safer neighbourhoods to stay in.

7.5 ANALYSIS OF CRIME CORRELATION

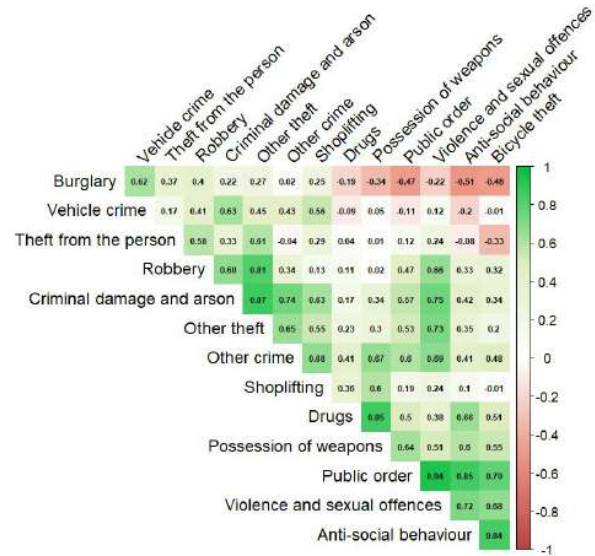


Figure 5: Crime Correlation Plot

Moving on to our next chart to see crime correlation in London. As to analyse the top two happening crimes, anti-social behaviour showed a very high positive correlation with multiple other crime types such as public order, violence and sexual offences, drugs and possession of weapons. In areas like this, would require higher attention as the rise in one of these few might result in the rest of the crime types to increase together.

The other happening crime type, violence and sexual offences, on the other hand, seems like it has a high positive correlation with a few other crime types which anti-social behaviour is not highly correlated to. These few are robbery, criminal damage and arson and other thefts and crimes.

Lastly, one interesting analysis is burglary. As compared to the rest of the crime types, burglary has the most negative correlation to most of the other crime type while keeping a high correlation to vehicle crime. This could show trends of areas with an abundance of both burglary and vehicle crime around the neighbourhoods in London and thus helping the police force to deploy more police security in those areas.

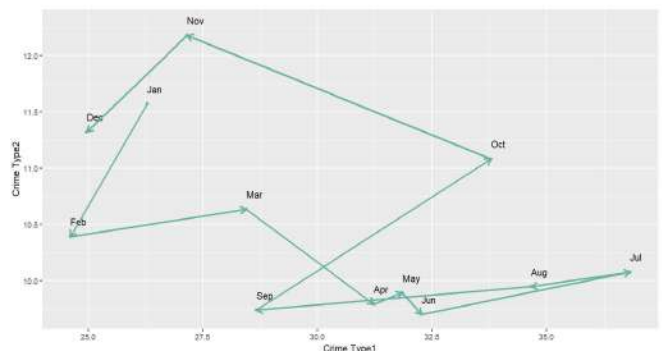


Figure 6: Line Chart for Seasonal Patterns

With the additional chart of this seasonal line graph, we are able to establish certain relationship between two respective crime type that the user can specify by a seasonal pattern. This allows the user to have extra information on the neighbourhood that they want to

research based on a few crime correlation that they are interested to find out based on the correlation plot.

The line chart creates a nice arrow directional view per month to allow ease of comparison of the crime correlation over the period selected by the user. With the help of both charts, it can allow user to understand better of the correlation of the crimes and also determine the seasonal pattern of the respective two crimes selected.

8 FUTURE WORKS

Our application can be used in a lot of different ways. Firstly, it can be used to explore different countries crime rate for the less developed countries who would want to visualise their own crime rates. This would help them understand what approach and direction to take through a data-driven approach.

This application could also be used by tourists to discover which areas are dangerous and which crimes are more prominent in that area hence being more prepared and aware of their surroundings.

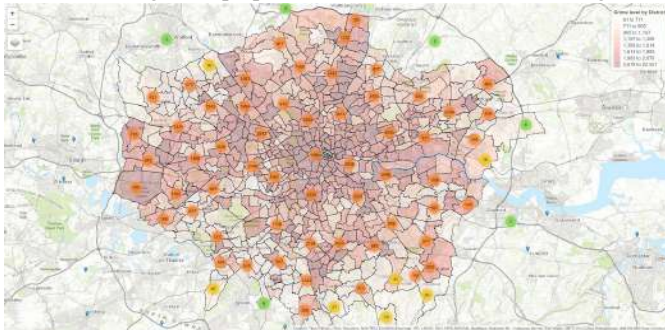


Figure 1: Individual crime plot

As we can see we are able to plot out each individual crime cases in both the City of London and greater London. Tourists are able to determine which street or area is plagued with crime hence avoiding unnecessary trouble.

With the addition of the dashboard, the above data can be plugged into a machine learning algorithm to predict future crimes. This can in turn be shown on the map for tourists to understand and be careful.

9 CONCLUSION

With this application, London's crime data can be analysed by studying the crime activity in each neighbourhood and time period, and better understand the correlation between crime types through publicly available data from the UK government data website and visualise these data via public free frameworks. This allows users to better understand the crime activity of all London neighbourhoods that can be used for multiple purposes such as tourists planning for trips, residents wanting to better understand the safety of their residence and the police force wanting to analyse which neighbourhoods are in need of more patrol etc.

ACKNOWLEDGMENTS

We would like to thank Dr. Kam Tin Seong for guiding us throughout this project.

REFERENCES

- [1] <https://data.police.uk/data/archive/>
- [2] <https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london>
- [3] <https://data.london.gov.uk/dataset/police-front-counter-access-times>