

A Case Study of the Metropolitan Police: Is Policing Fair in London?

Maeve Hutchinson

ID: 200051807

Department of Computer Science

City, University of London

December 2021

Abstract—This report investigates policing in London, focussing on the treatment of people by the Metropolitan Police based upon ethnicity. It was found that the rate of police force incidents per month per 1000 people was significantly higher for black people than people who are not black ($t = 13.7, p = 1.66 - 24 \times 10^{-24}$) and that black and mixed race people are less likely to think that the police treat people fairly regardless of who they are ($p = 5.86 \times 10^{-80}$). It was also found that police force usage is increasing over time, and it is unrelated to rates of crime in the city.

I. INTRODUCTION

In light of the March 2021 murder of Sarah Everard in London by a serving Metropolitan Police officer [1], policing in London is being scrutinised now more than ever. This recent scandal, combined with the global ‘Black Lives Matter’ movement — spurred by the murder of George Floyd, a black man, by a police officer in Minneapolis in May 2020 [2] — has placed the fairness and integrity of policing into question globally, in particular on the grounds of race [3].

Thus, this report aims to gain an understanding of the fairness of policing in London based on ethnicity, focussing on the use of force by the Metropolitan Police.

II. ANALYTICAL QUESTIONS AND DATA

A. Analytical Questions

This report will investigate the fairness of policing in London on three fronts: firstly, exploring if people feel like they are being treated fairly; secondly, investigating the reality of police treatment; and finally asking if there is any justification for poor treatment by the police. Thus, The following questions will be analysed:

- Do people feel that the Metropolitan Police treat people fairly, regardless of race? Is this perception different for different ethnic groups?
- Do the Metropolitan Police use force fairly across different ethnic groups?
- Are there trends in the use of force by the Metropolitan Police over time?
- Is there a link between crime levels and the use of force by the Metropolitan Police?

B. Data

The datasets used to answer these questions are as follows:

- Use of Force Dataset, Metropolitan Police [4]
- Police Perception Demographics, from The Mayor’s Office for Policing And Crime [5]
- 2016 Population Estimates by Ethnicity, Office for National Statistics [6]
- Crime Data, Metropolitan Police [7]

The Use of Force data is the key dataset in this investigation, as it provides information about how the police actually treat people. Each row in the dataset represents an instance of a single police officer using force against a subject.

It contains demographic information about the subject, including ethnicity, so differences in police treatment across different ethnicities can be analysed. However, it is important to note that as opposed to the other datasets, these are not self-described identities, but the ethnicity that the police officer thinks that the subject is. In this instance as the ethnicity assumed by the police officer is perhaps more important than a self-described identity, as it allows us to see how police officers treat people based upon their perceived ethnicity. Also, due to the broad nature of the ethnic groups, it is unlikely to be hugely inaccurate. There are also columns pertaining to subject injuries, use of tasers, and use of firearms, so the severity of the incidents can be explored.

The dataset contains the time of each incident, so time trends can be analysed and modelled. However, only began being collected in its current format in April 2017, this is where the analysis will begin, and it will cover a time period of 4 full years until April 2021. This is because data after April 2021 is still subject to change.

The Police Perception dataset is in the form of survey responses: residents of London were asked 6 questions about the police, and the proportion of respondents who agree with each statement is reported. The survey has been done every 3 months from the beginning of 2015, with the latest survey being taken in September 2019, so it lends itself to looking at trends in police perception over time. The proportion of respondents is also split by ethnic group, the difference in police perception across ethnic groups can be explored, which is key to this investigation.

The 2016 Population Estimates by Ethnicity dataset provides the population of each ethnic group in London. This time was chosen as population data by ethnicity is not produced often, and 2016 is the closest dataset in time to the use of force data. These figures will be used to provide context.

The final data set is the Crime Dataset. It provides the monthly number of crimes by crime category in each borough. It will be truncated so that the time range matches up to the use of force data. This dataset will be used to see if there is a link between crime and police use of force, to try and answer if there is any justification for the force usage.

III. ANALYSIS

A. Police Force Data Preparation

The dataset that required the most preparation was the use of force dataset. This is primarily because the rows do not represent individual incidents, rather, each row represents when an officer has used force on a subject. This means that a subject could be represented multiple times if more than one officer was involved, which would skew the demographic proportions. Thus, the key data preparation was removing duplicate rows such that each row represents one use of force incident against an individual subject, but without losing the information about the type of force used on the subject.

First, the time of each incident was rounded to the nearest hour. Since the force usage is timed to the minute, one incident may not have exactly the same time across rows, so it was decided that rounding to the nearest hour would give rows pertaining to the same hour, the same time. This is not necessarily the most rigorous way to do this, and it may miss some duplicate rows, if the use of force incident was particularly long, or if the times of the duplicate rows fall either side of the half-hour mark. However, it will encompass most duplicates.

Then, the dataframe was grouped by the rows that would be the same for duplicate rows. These rows were the rounded date and time, the borough, location, and subject, age, ethnicity, and gender. All of the other columns had already been one-hot-encoded to 1s and 0s, so when grouping, the maximum value for a grouped row in a given column was taken so that information about the incident was preserved. It was found that approximately 17% of the raw dataset was duplicate rows. This process reduced the size of the data set from 531, 801 rows, to a set of 440,624 rows, each representing an incident.

B. Police Perception Data Preparation

For the police perception data, the number of respondents was not provided for months past January 2017, which would make it hard to analyse the significance of each survey. So for the years that the number of respondents was available, the trend was plotted, and it seemed to level off for each ethnicity. So, the number of respondents for each future survey for each ethnicity was imputed using the mean number of respondents from July 2015 to January 2017.

Also, only one question from the survey was directly relevant to this investigation, so only proportions for the question

'Agree the police treat everyone fairly regardless of who they are' were kept.

C. Use of Force Data Derivation

In order to compare the proportions of police violence incident by rate to population proportions, the use of force dataframe needed to be combined with the population data. First, counts of each ethnicity in the prepared use of force dataframe were taken and saved as a series. The 'Chinese' category was merged with the 'Asian' category, as it was unclear why 'Chinese' was a separate ethnicity. That series was then divided by the total number of rows in the dataframe and multiplied by 100 to give a percentage of the total incidents per race.

From the population data, the total population of London by ethnicity was taken. Each value, as with the use of force counts, was calculated as a percentage of the total population. This series was combined with the use of force counts by ethnicity series and plotted as a bar chart to see if any ethnicity was disproportionately affected by police force.

In order to perform a t-test to test the significance of the proportions being different, the dataframe was then grouped into 'Black' and 'Non-Black' groups. The dataframe was then grouped by month, and the incidents per month was converted to incidents per month per 1000 members of the population for each group.

D. Use of Force Time Series Data Derivation

To analyse the trends in police force over time, the data frame of incidents needed to be transformed into a time series. To do this, the data frame was grouped by day, and the number of rows for each day was counted. A new column was added containing the number of days since the beginning of the dataset. This was so that models could more easily be fitted over time.

E. Construction and Validation of Time Series Regression Models

The incidents per day were plotted against date, and there appeared to be two trends (Fig. 1). So, a KDE and Q-Q plot were examined (Fig. 2), which confirmed this - the histogram is very clearly bimodal. Thus, it was decided to only fit the most recent data, from September 2018 to April 2021. The start date was chosen by looking at the graphs. This produced a dataframe that was 943 rows.

First, a Poisson regression model was fitted as the data is in the form of counts, however this was a poor fit. So, Negative Binomial (NB) regression model was chosen, as it is very similar to a Poisson model but without the restrictive assumption that variance and mean are equal [8]. Once the NB model was fitted, it was compared to the previous Poisson model using the likelihood ratio, which was large enough to suggest that the NB was a much better fit.

To quantify the goodness-of-fit of the NB alone, the chi squared statistic at the 95% confidence level with 942 degrees of freedom was calculated, giving a value of 872, which

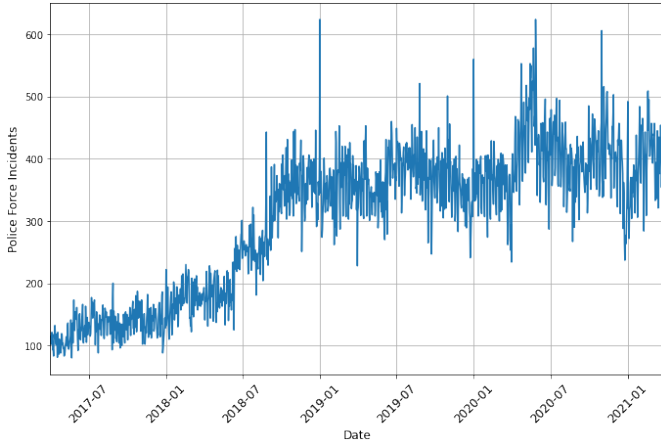


Fig. 1. Plot of police force incident counts per day over time.

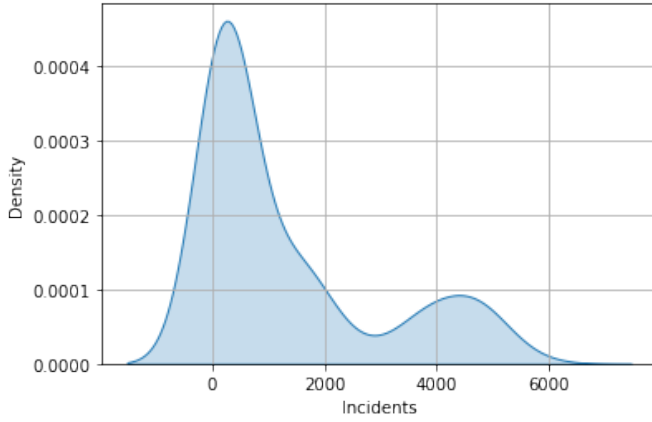


Fig. 2. Kernel density estimation plot of police force incident counts per day from April 2017 to April 2021.

is considerably larger than both the deviance and pearson chi squared of the model itself, which were both 17.8. This indicates that the NB model is a good fit [8]. 10-fold cross validation was implemented to estimate the generalisation of the model, and the validation log likelihood, deviation, and pearson chi squared were all very similar when the model was trained on the whole dataset.

The fit was plotted on top of the time series, and a Q-Q plot of residuals was examined to confirm the goodness-of-fit visually (Fig. 3). It can be seen that the Q-Q plot does have some skew, which is likely due to autocorrelation in the data. This was not examined further as the model was deemed a good enough fit, but it could be an area to further develop the model.

IV. FINDINGS, REFLECTIONS AND FURTHER WORK

The most interesting finding is the difference in the proportion of incidents affecting each ethnicity compared to the proportion of each ethnicity in London's population. The bar chart in figure 4 shows that black people are disproportionately

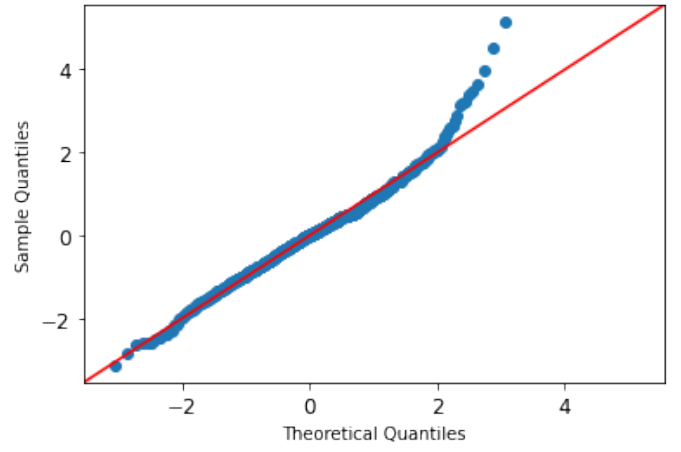


Fig. 3. Q-Q plot of the residuals of the Negative Binomial model fitted to police force incident counts per day from September 2018 to April 2021.

subjected to police force: whilst they only make up 12.5% of the population, they are the subjects of 34.8% of police force incidents. This suggests that the Metropolitan Police are perhaps not treating people fairly based upon ethnicity alone.

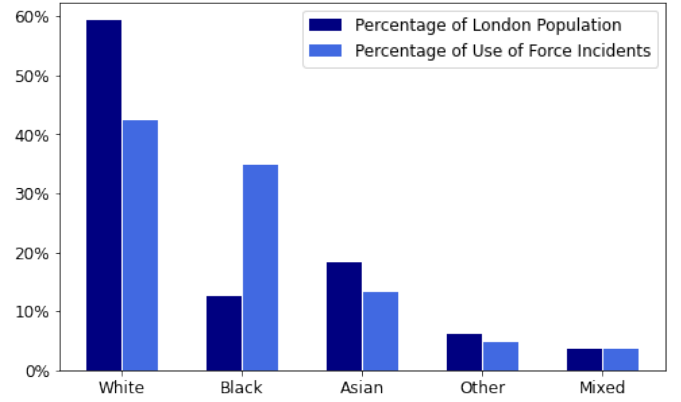


Fig. 4. Bar chart comparing the percentage of police force subjects by ethnicity to the corresponding percentage of London's population made up by that ethnicity.

This visualisation is reinforced by a t-test carried out on the number of incidents per month per 1000 people for black against non-black ethnicities. The results ($t\text{-statistic} = 13.7, p = 1.66 - 24 \times 10^{-24}$) suggest that there is significant difference between the level of police violence faced by black and non-black people at the 99% level.

The severity of the incidents across each ethnicity was examined and there was not found to be a significant difference in terms of the proportion of incidents causing injuries, involving tasers, or involving firearms. Thus, whilst black people are disproportionately affected by police use of force, the level of force used does not vary significantly based upon ethnicity, which does not suggest unfair treatment.

The police perception data aligns with this key finding.

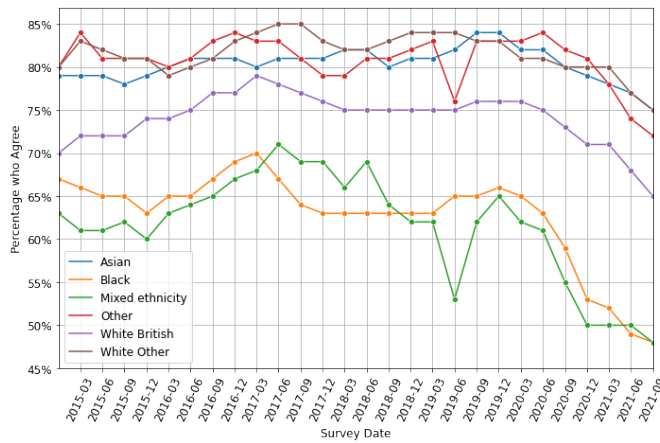


Fig. 5. Plot of the percentage of people, by ethnicity, who "Agree the police treat everyone fairly regardless of who they are" over time.

Figure 5 shows how the proportion of people per race agree that the police treat everyone fairly regardless of who they are. It can be seen that the black and mixed ethnicity groups are consistently lower than other ethnicities. A chi squared test was performed which confirmed that there is a statistically significant relationship between race, and agreeing that the police treat people fairly ($p = 5.86 \times 10^{-80}$). All ethnicities show a downward trend from March 2020, which interestingly roughly aligns with the 'Black Lives Matter' movement.

The NB regression model in figure 6 shows that police force incidents are increasing with time. So, it was investigated whether there is any justification for this increase. Under the assumption that police force is used as a response to crime, in particular violent crime, it could be argued that this increase is justified if it is linked to an increase in crime.

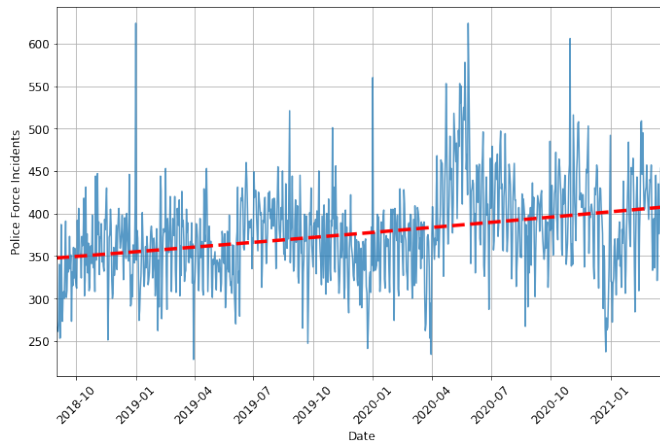


Fig. 6. Plot of police force incident counts per day over time fitted with a Negative Binomial regression.

So, the Pearson correlation coefficient was calculated between the number of police force incidents, and the total number of crimes per month. The value was $r = -0.019$

which suggests there is no significant correlation, as confirmed by figure 7. Next, the same coefficient was calculated between police force incidents, and violent crimes per month. The value was slightly higher at $r = 0.16$, but still not a significant correlation, which is again confirmed in figure 7. Thus, there is no relationship between police use of force and crime.

Further investigation could be done into the time dependence of this data, by incorporating more features into the model that could influence levels of police force. This would be able to provide some insight into not only the fact that police force is increasing but also why.

This investigation is limited in terms of assessing the whole domain of policing, because police force incidents are only one out of a range of ways that a police officer interacts with a community. An area for further investigation would be to explore the stop and search data, also provided by the Metropolitan Police, to see if there is also any ethnicity-based bias in that data. It would also be interesting to compare this data to rates of police force in the rest of the country, and other countries, especially the United States.

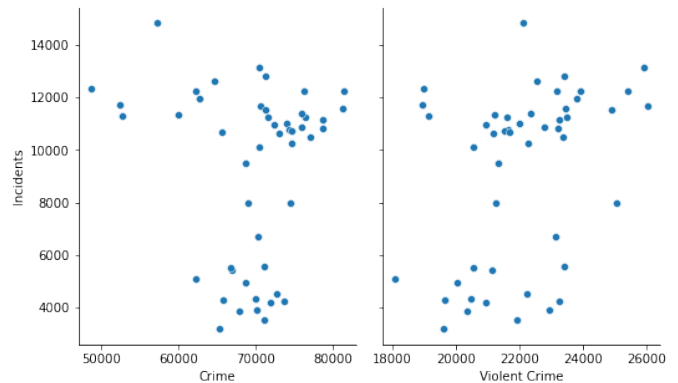


Fig. 7. Plot of police force incidents against crime (left) and violent crime (right) numbers in a given month.

REFERENCES

- [1] “Sarah Everard murder: Wayne Couzens given whole-life sentence,” BBC News, Sep. 30, 2021. Accessed: Dec. 23, 2021. [Online]. Available: <https://www.bbc.com/news/uk-england-london-58747614>
- [2] “George Floyd: What happened in the final moments of his life,” BBC News, Jul. 16, 2020. Accessed: Dec. 23, 2021. [Online]. Available: <https://www.bbc.com/news/world-us-canada-52861726>
- [3] L. Buchanan, Q. Bui, and J. K. Patel, “Black Lives Matter May Be the Largest Movement in U.S. History,” The New York Times, Jul. 03, 2020. Accessed: Dec. 23, 2021. [Online]. Available: <https://www.nytimes.com/interactive/2020/07/03/us/george-floyd-protests-crowd-size.html>
- [4] “Use of Force Dataset – London Datastore.” <https://data.london.gov.uk/dataset/use-of-force> (accessed Dec. 23, 2021).
- [5] “Public voice dashboard,” London City Hall, Aug. 24, 2017. <https://www.london.gov.uk/what-we-do/mayors-office-policing-and-crime-mopac/data-and-statistics/public-voice-dashboard> (accessed Dec. 23, 2021).
- [6] “Population characteristics research tables - Office for National Statistics.” <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationcharacteristicsresearchtables> (accessed Dec. 23, 2021).
- [7] “MPS Monthly Crime Dashboard Data – London Datastore.” <https://data.london.gov.uk/dataset/mps-monthly-crime-dahboard-data> (accessed Dec. 23, 2021).
- [8] J. M. Hilbe, Negative binomial regression, 2nd ed. Cambridge, UK; New York: Cambridge University Press, 2011.

WORD COUNTS

Section	Words
Abstract	107
Introduction	106
Analytical Questions and Data	597
Analysis	983
Findings, reflections and further work	584