POLICE: Stop & Search Data

Information Visualization

Ana Terović Fani Sentinella-Jerbić

October 2022

1 Introduction

Police are arguably one of the pillars of a peaceful and righteous modern life. They are responsible for various types of law enforcement in order to protect the citizens of modern society.

Sometimes, in order to protect the rights of the majority, police need to violate the rights of certain individuals, such as the right to privacy and freedom of movement. These types of operations are often referred to as "stop and search" processes which should be carried out only if the violation is reasonable, appropriate, and legal. However, certain police forces are abusing their authority by stopping and searching people in a way that is discriminatory, inefficient, and a waste of public money [2]. This topic has gained popularity in recent years, inspired by the murder of George Floyd in 2020, causing the rise of civil unrest and media coverage of police brutality and racism.

Through this assignment, we are trying to approach this issue from a data-driven standpoint. The data we are using is provided by the London Metropolitan Police and includes all stop and search actions from the 1st of June 2021 to the 1st of July 2022. We try to offer a visualization for navigating this complex topic, which could be used by the police professionals in order to improve their operations.

2 Domain

The domain of our visualizations would be considered law enforcement. More specifically, crime statistics. We would like our work within this domain to be used in an educational manner. By informing criminal justice and law enforcement professionals of their biases, gaps, and errors in past operations, we believe we could improve their operations in the future. Ultimately, this would lead to better crime prevention and protection of the general public.

3 Users

Our users would be law enforcement professionals. This includes a wide range of different roles within the profession. We aim to educate not only police officers whose work is tightly related to the stop and search processes, but also other workers of the police force, because certain biases and errors are transferable to other areas of policing.

Most workers of the police force need to pass certain medical, literal and numerical tests in order to gain their positions. This means we can assume our users hold basic digital, cognitive, and literal skills and abilities. Additionally, these people are trained to deal with stressful and sensitive situations in a rational manner, so we can expect them to react to racist, ageist and other controversial topics more calmly than an average person, or at least to be aware of the need to do so.

4 Data

As previously stated, we have chosen Stop and Search Dataset which includes stop and search actions by the Metropolitan Police, London. The dataset is multidimensional, consisting of 16 attributes and 16685 samples.
The attributes are,

- **type**, indicating the type of search action that was done by the police, ie. Peron search, Vehicle search,
- **date**, date and time the search was done,
- **part of a policing operation**, a boolean value indicating whether or not the search was done under a policing operation,
- **policing operation**, if the search was done under a policing operation this variable describes what kind of a operation,
- **latitude**, latitude at which the stop and search was done,
- **longitude**, longitude at which the stop and search was done,
- **gender**, gender of the person on whom the stop and search was done,
- **age range**, age range of the person on whom the stop and search was done,
- **self-defined ethnicity**, ethnicity of the person on whom the stop and search was done,
- **officer-defined ethnicity**, ethnicity of the officer who did the stop and search,
- **legislation**, reason for which the stop and search was done,

- **object of search**, what the police looked for during the stop and search,
- **outcome**, outcome of the stop and search,
- **outcome linked to object of search**, what the police did with the object of search,
- **removal of more than just outer clothing**, what was removed from a person other than their outer clothing in order to do the stop and search.

Attributes are mostly categorical, like Type, Age Range, Self-defined ethnicity. There is also continuous data like Date, binary data like Part of a policing operation and discrete data like Longitude and Latitude.

Certain attributes we will not be taking into account for our visualizations because they don't contain any useful information. Those are Part of policing operation, Outcome linked to object of search, and Removal of more than just outer clothing.

5 Goals and Tasks

For our visualization we have set out 3 main goals;

- comparison based on ethnicity, gender, and age,
- time analysis,
- geographical analysis.

5.1 Comparison based on ethnicity, gender, and age

When we look at the news, there are often examples and discussions about discrepancies between different ethnic groups and genders. In 2020, police killed an African-American, George Floyd, causing one of the biggest protests against police brutality in recent history. This and many other examples give us interest in exploring the given dataset in terms of how a person's ethnic background affects policing decisions.

When it comes to gender, the problems usually arise from the expected roles which seem to affect all the aspects of life for both genders. Because of this, it would be interesting if somehow gender also affects the possibility of a person being stopped and searched.

Age is another attribute we thought might be interesting to explore. For example, maybe younger people have a higher chance of being suspected since they are less mature and therefore might seem more prone to trouble.

An idea for visualization is using bar charts to showcase how many stop and searches were done in comparison with different groups, visible in 1. Bar charts

are a good visualization tool for quantitative datatype since they use length as a means of displaying data. An interactive feature would be layering the data from the dataset with data about what the actual percentage of groups are in the population. For London population this is visible in figure 2. This way, we can visualize the deviation from what is supposed to be a "normal" distribution.

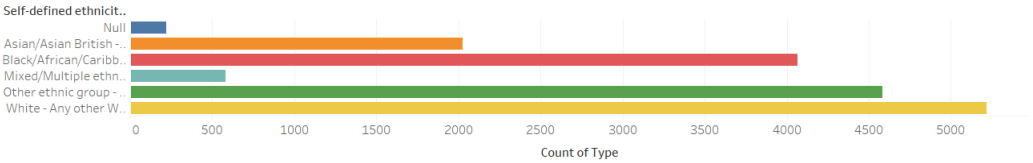

Figure 1: Comparison of the amount of stop and searches done by ethnicity

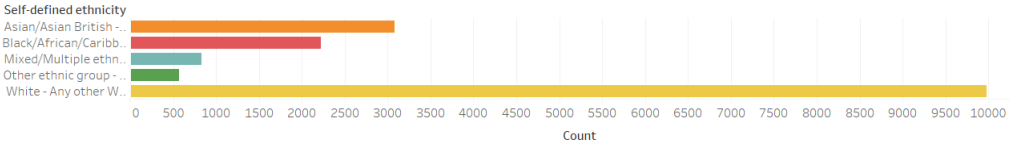

Figure 2: Comparison of the amount of people living in London by ethnicity

Alternatively, we had the idea of using a more intriguing method called "village of 100 people". The idea is to color the number of people's icons according to the percentage of the group in the data. This helps cognitively to put the numbers into perspective. For the sake of simplicity this time we visualized it as the "village of 10 people" in the figure 3.

Figure 3: "Village of 10 people" visualization of the London population and the search population

To comment on the color palettes; for this initial phase we used some default

colours but in the real visualizations we would be careful to use qualitative colormaps which are used to represent information which does not have an order. This is specially important in the case of ethnicity in order to sensitively approach this topic.

5.2 Geographical Analysis

As our data also contains geographical information about operations, we decided to tackle spatial analysis of the recorded searches, as there could also be interesting takeaways. We must note that we are aware of the location anonymization process performed by the Metropolitan Police before the publication of data but we believe this process doesn't change the data in such a large manner that one can't be able to make conclusions based on it.

We would use a geographical visualization with stop and search operations plotted as dots on the map of London. Users would have the option to zoom and move around the map in order to better inspect various areas. We expect the stop and search data when plotted to look something like the figure 4. The goal here is to provide the user with a tool which could help with future planning of stop and search operations.

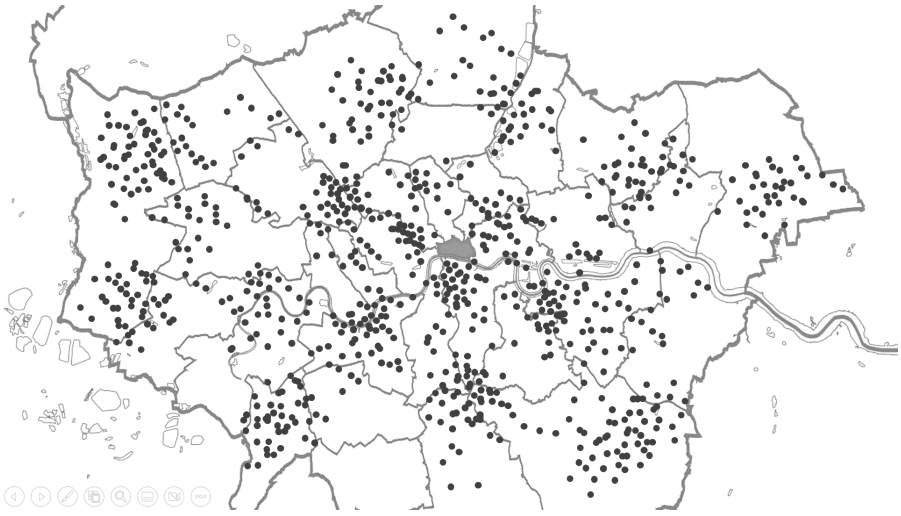

Figure 4: Geographical map of searches

What we would like to note here is that the user could make one common logical error when looking at a map like this. Seeing the operations are somewhat clustered, one could think the clustered areas are places where crime is most prominent. This, however, might not be the case. For example, it could be that the areas where there is the biggest number of searches are just where police officers are most often on duty. This error is often referred to as the survivorship

bias [1]. To solve this issue, we could take different approaches. For example, visualizing the searches with different hues for different outcomes. This may at least show where the real crimes are actually happening. Or by including the location of the police stations in the visualization. We can expect some searches to be clustered around them, such as shown on the figure 5. This could make the police officials consider relocating future operations to further areas. Whatever the approach we take, it is important that we somehow inform the user to avoid that logical error.

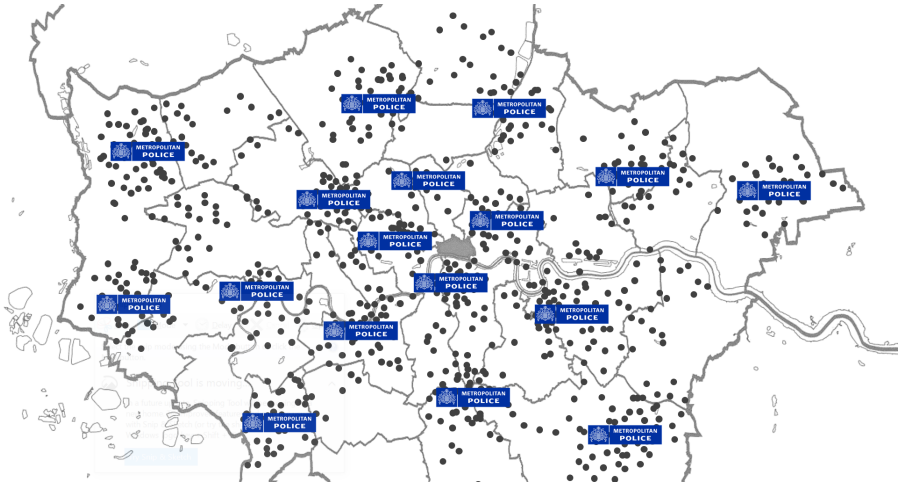

Figure 5: Clustering of searches based on locations of police stations

5.3 Time Analysis

The dataset provides the date and time the stop and search was done, giving us the ability to perform time analysis. The time period in which the data was collected is between 1st of June 2021 and 1st of July 2021. Since the dataset is limited to a one month period, we can't perform an in-depth analysis for longer periods of time. However, we would like to perform time analysis based on the time of the day and days of the week. For example, were the searches done more often at night and were the searches done more often during the weekend.

6 Possible extensions

One interesting extension would be overlapping of the goals we already mentioned. For example, we could improve the interaction with the geographical map by providing an option of visualizing different groups of suspects on the map by using color hues such as shown in the figure 6.

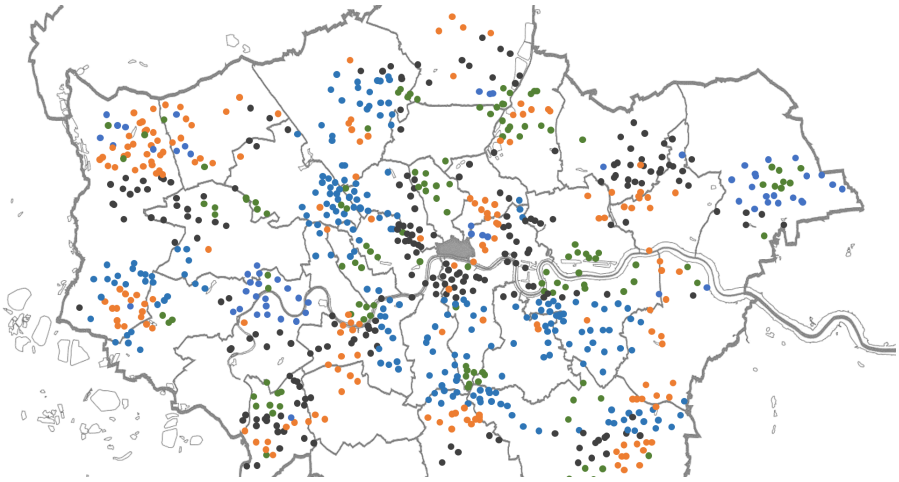

Figure 6: Geographical visualization of groups by color hues

References

[1] J. Ellenberg. *How not to be wrong: The power of mathematical thinking.* Penguin, 2015.

[2] Equality and Human Rights Commission et al. Stop and think: A critical review of the use of stop and search powers in England and Wales. 2010.