

# Data Visualization Project Summary

{{whoami|}} Name: Mussa Yousef  
Module: Data Visualization  
Email: [mussa.yousef.1@city.ac.uk](mailto:mussa.yousef.1@city.ac.uk) {{|whoami|}}

{{task|}}

You should complete this datavis project summary document and submit it, along with any necessary supplementary files to Moodle by **Sunday 18th April, 5pm UK time (16:00 GMT)**. Submissions will be awarded up to **80 marks** towards your coursework assessment total.

You are also encouraged to regularly commit and push changes to your datavis project throughout the term as you develop your project.

{{|task|}}

{{questions|}}

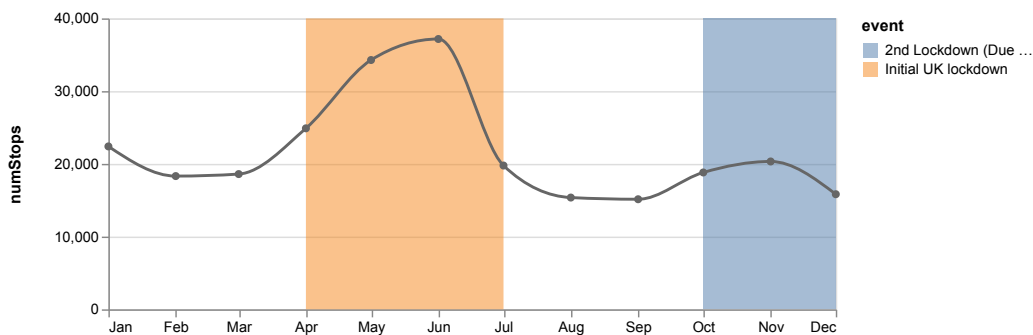
How has the coronavirus impacted policing in London? More specifically:

1. Did the lockdown 2020 affect the number of stop and searches in London?
2. What trends could be found in Policing stop by race and age throughout the year of 2020?
3. What crimes committed was given most arrests through 2020 and what trends could be seen for this in the demographic of people arrested during 2020?
4. Which boroughs in London saw the most crimes committed? Was this different from when London was in lockdown compared to out of Lockdown?

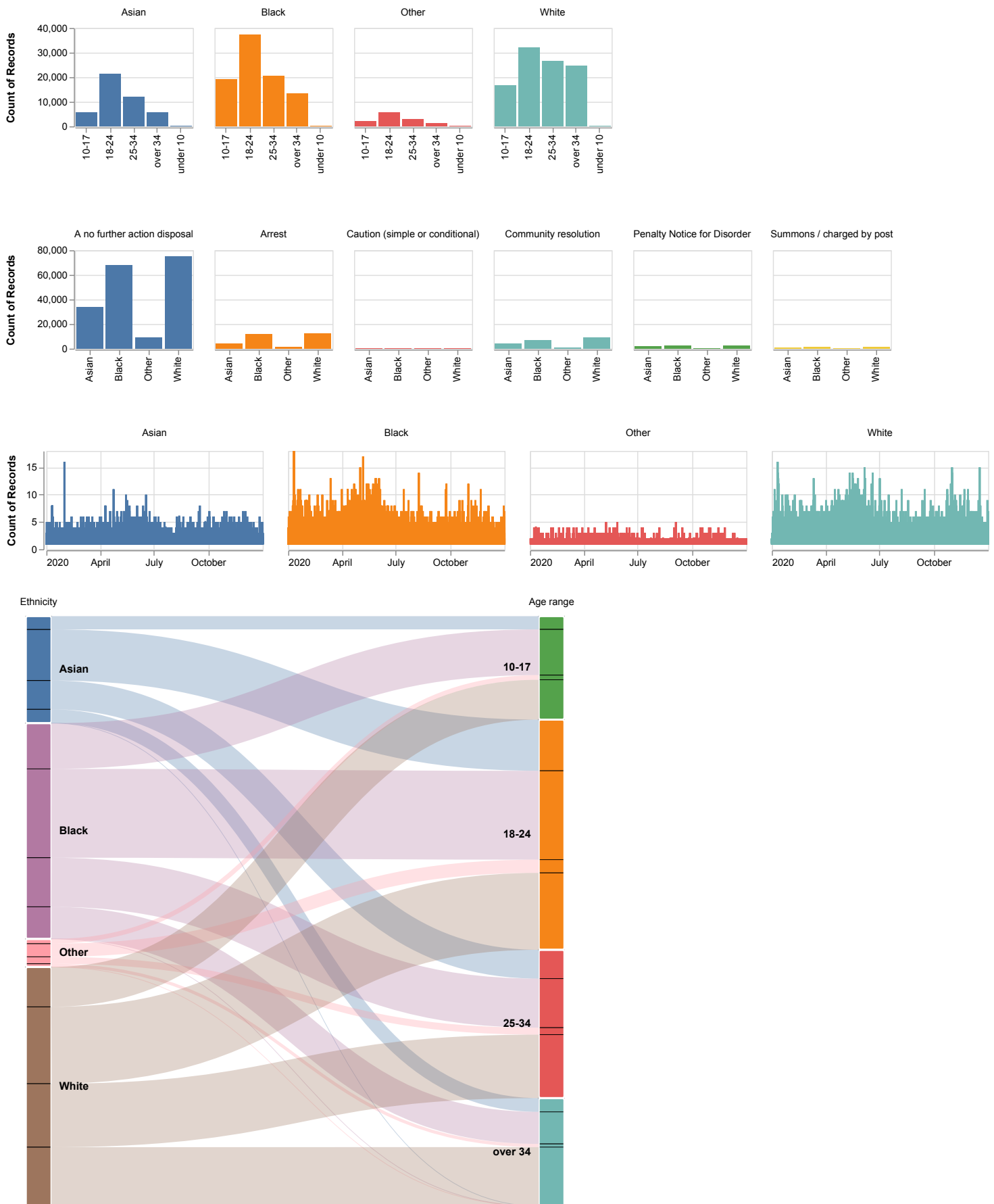
{{|questions|}}

{{visualization|}}

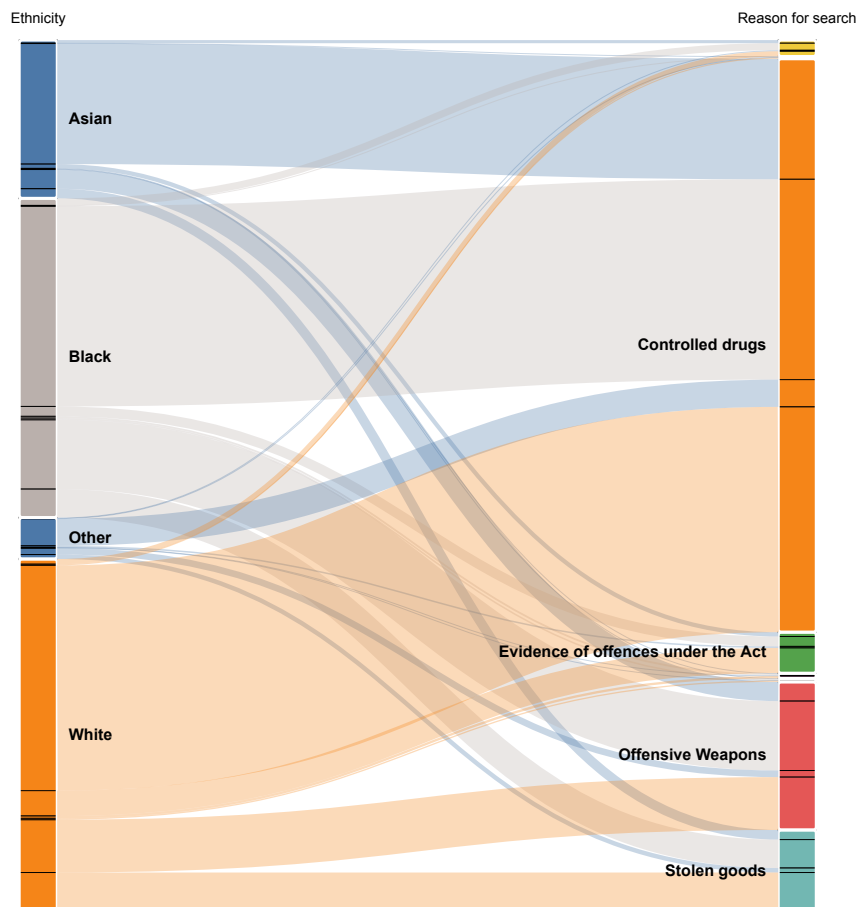
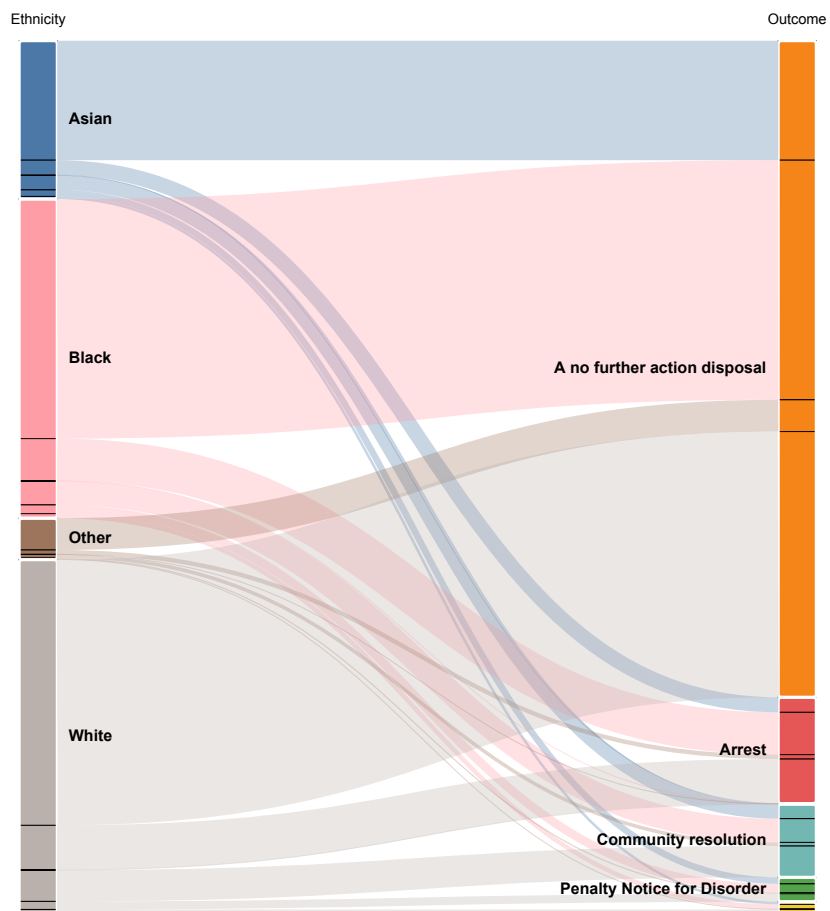
## 1. Did the lockdown 2020 affect the number of stop and searches in London?

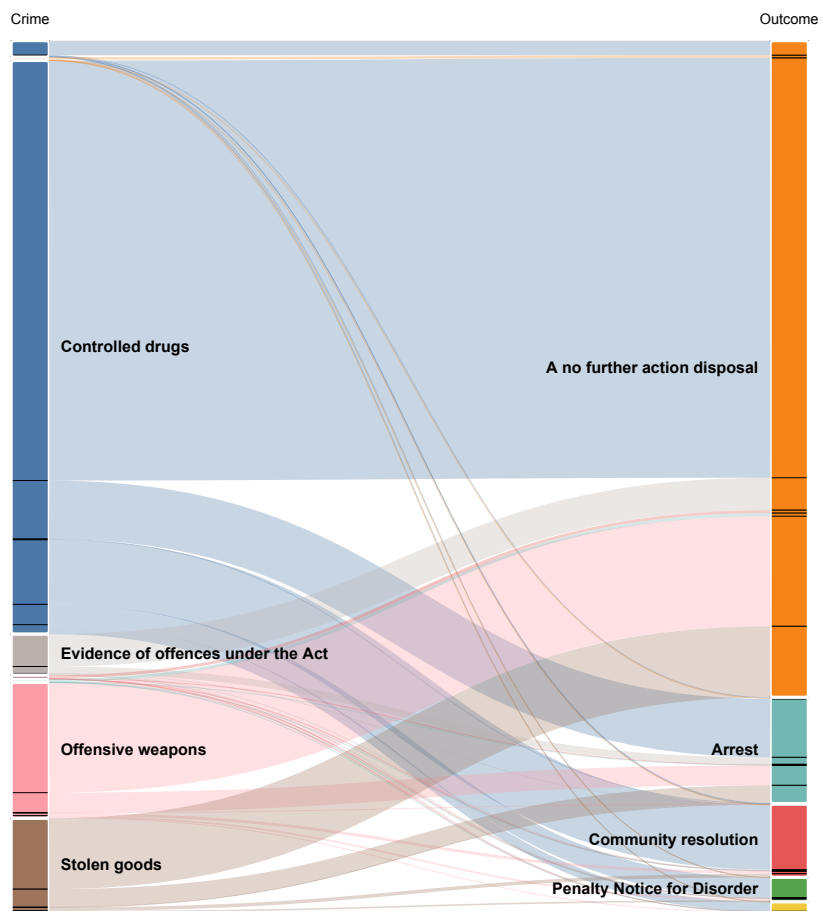


## 2. What trends could be found in Policing stop by race and age throughout the year of 2020?



3. What crimes committed was given most arrests through 2020 and what trends could be seen for this in the demographic of people arrested during 2020?





4. Which boroughs in London saw the most crimes committed? Was this different from when London was in lockdown compared to out of Lockdown?

{|visualization)}

{{insights|}}

For this project, we wanted to investigate trends of crime in London and how the metropolitan police carried out their duties in 2020. We segmented this down into 4 questions scrutinizing temporal and geospatial trends as well as the breakdown of the demographic of those stop and searched by the police. The following insights were found:

For our first research question, we present our visualization titled “highlighted”. We find that outside of Lockdown between January-March and July-September similarity in the number of stop and searches by the police around 16,000-19,000. We find a significant increase to stop and searches during the first lockdown almost up by 100%, the second lockdown followed the same trends however a much smaller increase at 25% in stop and searches during the second lockdown. Considering, the lockdown itself was one of the most unprecedented restrictions posed by the government to date, it takes time and effort i.e., policing to enforce this. This visualization shows that there was a potential over enforcement by the police but also shows that the police were far more prepared during the second lockdown. Furthermore, the difference in the increase of stops and searches shows that the government had effectively informed citizens.

We created four visualizations for our second question to be able to gain a spectrum of insights. From our first visualization, we can derive immediately that Black individuals aged between 18-24 were the most stopped throughout the year. Interestingly this age group is the highest stopped across all ethnic groups. Our second visualization further develops our insight, expressing stop and searches via outcomes. Here scrutinizing all outcomes, we find the same trends across all outcomes, although we find that majority of stop and searches were given an outcome of no further action, where white ethnics were given the most, closely followed by black ethnics. Our third visualizations translate this into a temporal visualization. White and Black ethnic stop and search recorded seems almost identical trends throughout the year, with slight peaks during the lockdown as expected seen in Q1. We find however higher peaks with black ethnic groups; however, these are not consistent. Our 4th visualization shows the same data as the first however here we can see the proportions of each ethnic group via age group. We also find that Black 18-24 year olds are have been 5,000 more times than the their white counterparts however compared for the 25-34 age range we find white ethnic groups to be stopped the most, 6000 more times than their black conterparts. This is an interesting finding, and could suggest, does age have an impact to ethnic crime rates.

For our third visualization, our 3rd research question slightly intersects our 2nd research question however here we wanted to scrutinize how reasons for being stopped translated in the outcome of stops. Our first visualization deepens our 2nd visualization for our 2nd question however we visualize the proportions compared to ethnic groups. Our second visualizations we can see that majority of stop and searches was due to suspicion of drugs at 65%. We also find a large proportion of the reason for the search was due to an offensive weapon, where half of all stops were of black ethnics. Our third visualization analyses down the crime committed with the outcome. Here we find no further action shares a proportional relationship across all crimes committed. Focusing on arrests we also find the same proportional relationship where more than half of arrests come from controlled drugs, offensive weapons, and stolen goods respectively share almost the same proportions of arrests with 17% and 18%. interestingly 75% of the suspicion of Firearms were given no further action whereas only 20% of Firearms stops were arrested. This could be down to the wrong source of information from potential gangs or potentially an over-policing which could be linked to our previous findings of 18-24-year-old being the majority of stops.

Here we used geospatial analysis to be able to visualize our data segmented by boroughs and the number of crimes committed within those boroughs. Here we find that Westminster and tower hamlets had the most crimes committed throughout 2020. Considering through the end of March to June and October to December we were in lockdown. We filtered out each month and visualized this on the following two visualizations of London. We find the Westminster continuously had high crime rates in and out of lockdown however we find that majority of boroughs saw a surge of crimes committed when London went into Lockdown. We find peripheral boroughs of London level of crime stayed more consistent, boroughs like Kingston, Harrow, Havering, and Bexley crime rate barely changed when the lockdown was enforced by the government. We saw a greater change in inner-city crime rates variance in crime rates for inner boroughs. Croydon was the biggest difference, we saw a 55% increase in crime during Lockdown.

{{|insights}}

{{designJustification|}}

The recent light shed on the unreasonable treatment of black citizens in America by the police, which was transmitted to recent riots in London, has led me to focus on policing by demographic. Therefore, considering our target audience of these visualizations would be the general public who may need this information to demystify any false narratives. Furthermore, the route by which this visualization would be sourced to the public would be either the News or government/world organization. Our main aim to make sure the information we present is clearly understood and most importantly memorable. We considered recent research [2] which suggests that bar graphs are the most used within these channels. Bar plots are among the least memorable graphs, however, are amongst the easiest to understand. Given the need to

explore this data, our target audience, and potential mediums which would be used to publish this visualization we chose to use faceted Bar graphs to explore the demographic of those stopped and searched and their respective outcome aggregated. We used neutral colours to denote respective races while aligning our bar chart horizontally against each other allows easy comparison to be made. We didn't add interactions in our bar charts, rather utilising a faceted view allows us to control our visualization. We done this in parallel with the Data Feminism approach. "Understanding of emotion can expand our ideas about effective data Visualisation" [3] hence in this case we understand the uproar, confusion, and anger from our audience therefore, providing an interactive visualization was not needed here. Rather clear, concise, and easily absorbed visualization was our chosen route, which would address emotions and deflate narratives.

We utilised the Sankey diagram for a few of our visualizations for this project. Derived from engineering and used to visualize the efficiency of steam engines, the Sankey diagram which visualizes nodal connections and flows between two or more points. Our process of utilising the Sankey diagram consisted of several iterations. Considering Kindlmann and Scheidegger's Commutative Diagram [4], we initially had one form of data which was a full-scaled CSV, which was partly pre-processed. Here we found our visualization fairly restrictive, we aggregated our data to represent exactly what we needed to visualize and created a JSON file. In our first Sankey diagram for our research question 3, we represented the same information we represented in our ethnic-outcomes bar chart for our research question 2 which showed an identity transformation, which enriched our visualization while allowing our targeted audience to take away percentages as proportional flows. This added an elegant and attractive touch to our visualization which Tufte [5] states allows the visualization to tell a story about the data.

Our data is of London crime throughout 2020 hence, both temporal and geospatial data is presented in the spectacles of policing in London. Taking Tobler's first law of geography [6] "...near things are more related than distant things" literally. We decided to use a 2D geospatial mapping of the boroughs of London. Iterating 3 mapping projections that encapsulate our research questions. Our choice to use this mapping projection allowed us to discover clusters of where crime had risen during the lockdown, otherwise would not have been able to without this visualisation. We chose not to relax our boundaries and rather keep boundaries untouched, which allows our visualization to keep its rigid features. We used a continuous colour scheme as binning our colour choice would take away the ability to translate how much a borough's crime rate had changed per borough. Therefore, we utilised colour as a variable to distinguish this, making sure we keep our audience in mind when deciding on the colour we use. We looked at the "Colours in Culture" from Information is Beautiful [7]. Which states the shades from blue to yellow/white symbolises Loyalty, Unhappiness, Truce, and Freedom. Which is how we are aiming to appeal to our audience, making sure this visualization comes across empathetic was a strategic choice of colours.

{{designJustification}}

{{validation}}

Our visualizations allowed us to explore using geospatial and temporal which enabled clusters that otherwise would not have been able to be seen nor analysed looking at the raw form of our tabular data. Our geospatial visualization allows our audience to derive how certain parts of the London boroughs simultaneously saw an increase/decrease in crime rates. This could be linked to potential infrastructure between the cluster of boroughs or the demographic of people living within these boroughs may be similar. We were able to segment our temporal view of crime highlighting lockdown while similarly vertically placing our choropleth, and utilising horizontal faceting when comparing ethnic stop and search over 2020. This enables our audience to easily extract and compare results. Links can be made between ethnic groups which allow the audience to derive how peaks within ethnic groups are translated to the whole picture of stop and searches across London. Furthermore, we were able to see how other ethnic groups which may have had a low number of stop and searches offset those who have a peak throughout the year.

One major limitation of our choropleth visualization is our continuous colour scheme, which normalizes across each visualization. If you look at our London Choropleth 1 visualization, we find highly intensified blue colour boroughs denote high crime numbers. However, in our 2nd and 3rd London Choropleth, we find that because we have extracted our data for months of Lockdown, our minimum and maximum values have changed hence the change of colour, does not necessarily translate to an increase or decrease in crime rates in the borough. What we would have done differently here if given the opportunity, is taking the percentage of crime against the population of the boroughs during the lockdown and computing these visually. This way we standardize our crime rates against the population which we keep consistent through all our visualization.

The Sankey diagram, allowed us to express our data in a flow, depicting ethnic groups to reasons and outcomes of crime. Although this allowed a unique application of the visualization, the element of time and geographic location was eliminated. Given the nature of our data and research, this element is vital, understanding how lockdown may have impacted these reasons/outcomes of stop and searches would have added far more value. Although like the choropleth visualization, we could have filtered our data which linked to lockdown out, hence producing a separate visualization. This still doesn't describe our data in a way in which temporal trends could be extracted. Furthermore,

we incorporated two datasets in this research, crime data and stop and search data. However, our crime data wasn't as descriptive as our stop and search data which limited our scope to visualize effectively.

{|validation)}

{{references|}}

[1] Data set downloaded from - <https://data.police.uk/data/> - Between Jan2020 - Dec2020 Both Crime data and Stop and search data between January - December

[2] Borkin, M., Vo, A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A. and Pfister, H. (2013) *What makes a visualization memorable?* IEEE Transactions on Visualization and Computer Graphics, 19(12), pp.2306-2315.

[3] D'Ignazio, C. and Klein, L. (2020) *Data Feminism*. MIT Press

[4] Kindlmann, G. and Scheidegger, C. (2014) [An algebraic process for visualization design](#) IEEE transactions on Visualization and Computer Graphics 20(12), pp.2181-2190.

[5] Tufte, E. (2001) *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press.

[6] Tobler, W.. "Movie Simulating Urban Growth in the Detroit Region Author ( s ) :." (2005).

[7] Information is Beautiful - Colo<https://informationisbeautiful.net/visualizations/colours-in-cultures/>

{|references)}