FIT5147 Data Visualization S1

The influence of Age, Gender and Ethnicity on London Police forces Stop and Search Actions

DVP

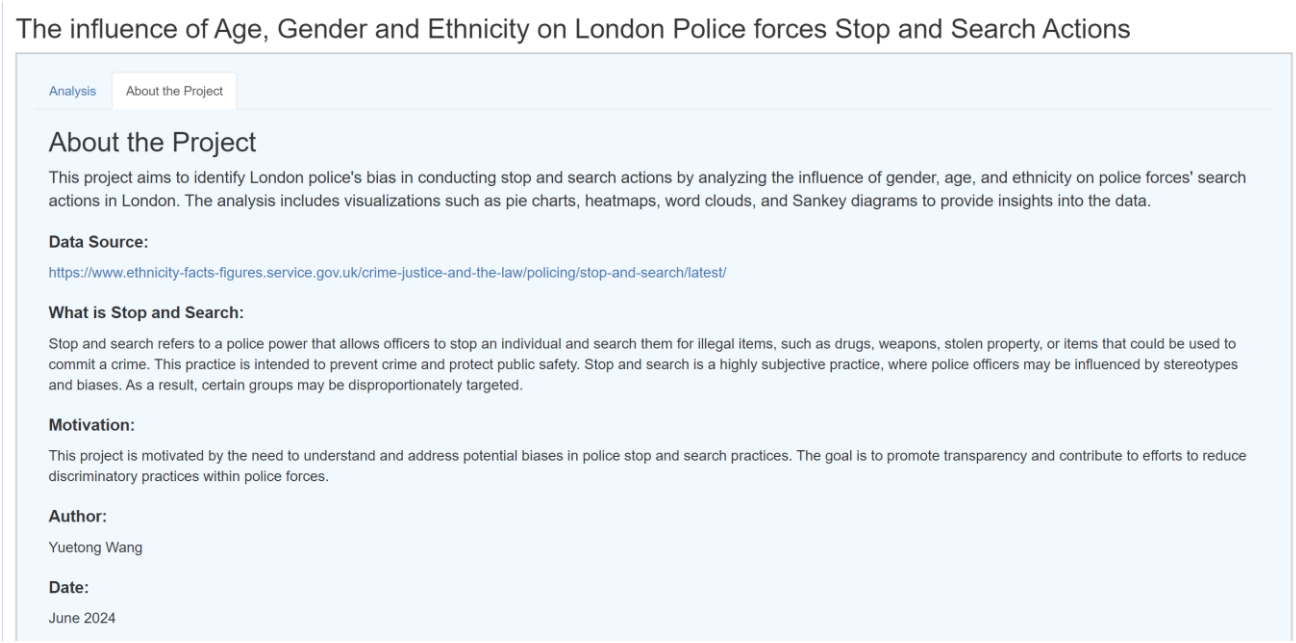
Yuetong Wang
2024-6-1
33581975

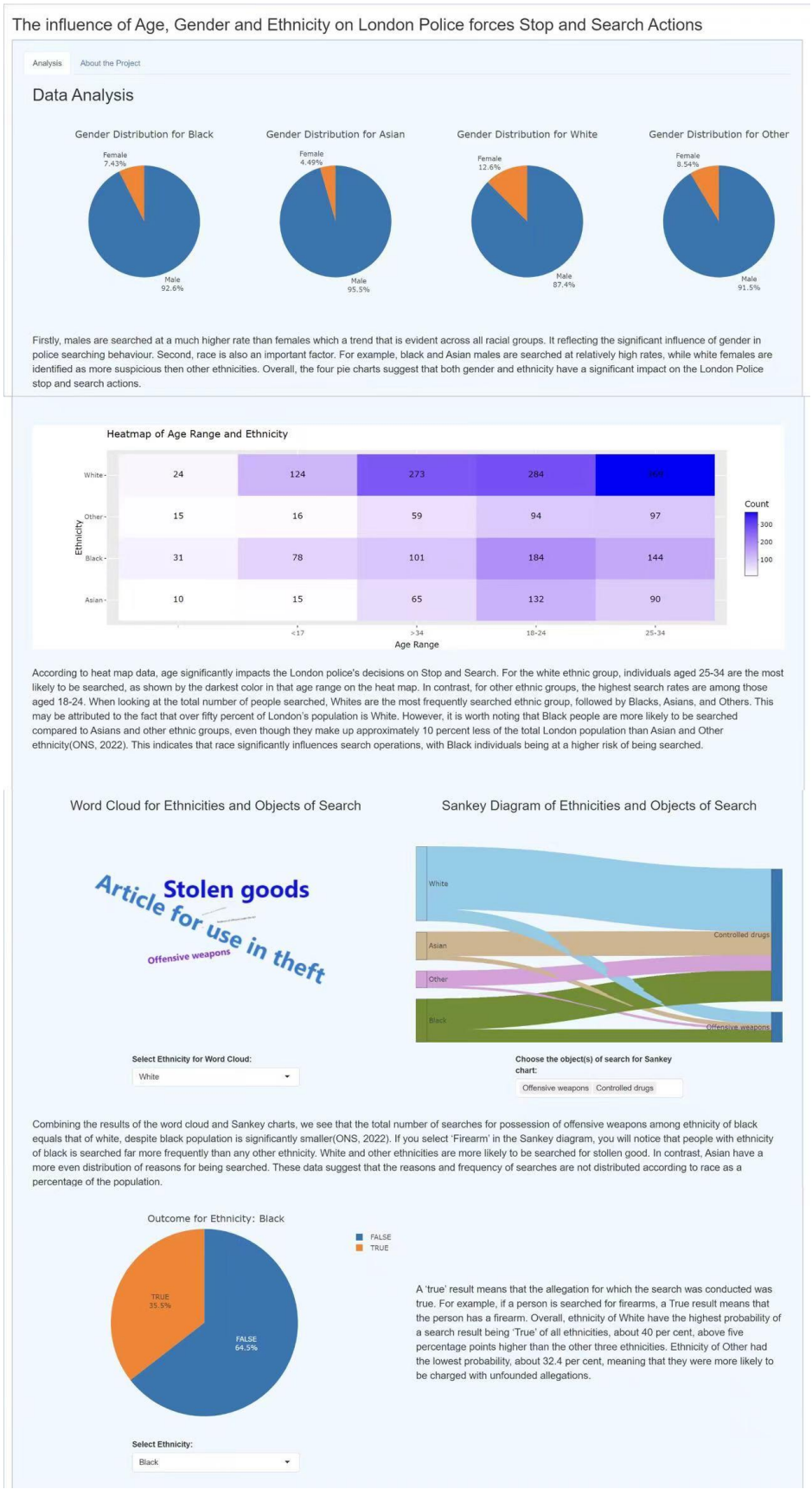
1. Introduction

Stop and Search is a policy set up by the United Kingdom government. It empowers UK police officers to stop a person and conduct a full body search, which may include the removal of more than just outer clothing, by stating they are suspicious of violating the law (GOV.UK, 2011). This DVP project is aiming to identify the bias contained in the Metropolitan Police and City of London police forces' Stop and Search action in the London area. Specifically discussing the influence of age, gender, and ethnicity on police officers' search reason (object of search) and search outcomes. The results of the study show that there are significant differences in police officer's subjective decision on conducting stop and search of individuals with different demographic characteristics.

The intended audience for this project includes policymakers, law enforcement officials as well as the general public. Policymakers can use these insights to drive potential reformation of the policy to ensure fairer police practices. Police forces can better understand their impact with their behaviours being visualised. The general public, particularly in the London area, can gain a clearer understanding of bias and speak out for themselves with data.

Data used in this project is originally downloaded from the UK police data then merged, categorized and cleaned.





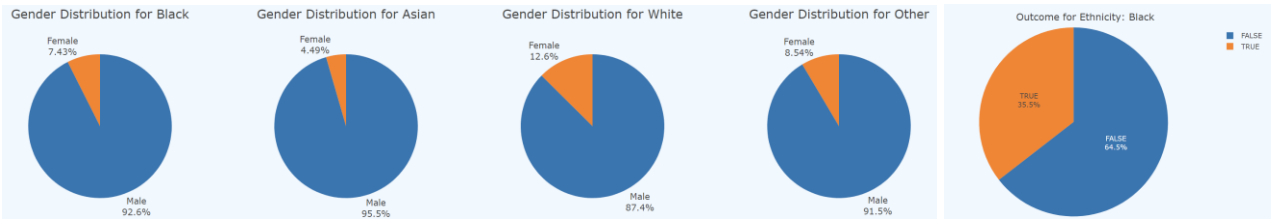
(Overview of the project)

2.Design Process

Graph choice:

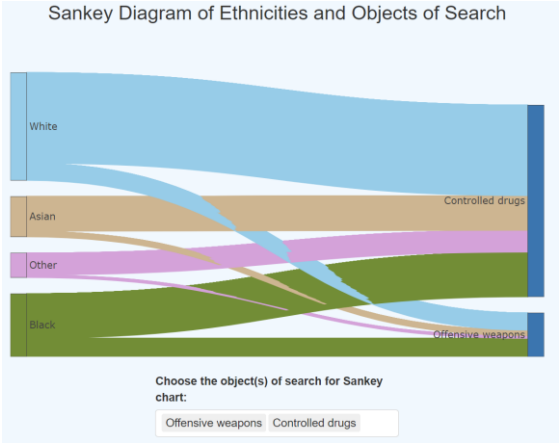
Graph choices are based on a graph webpage suggested in class: From Data to Viz (Healy, 2024).
I identified the data type first than based on the webpage's advice choose effective graph type.

Pie charts:



Ethnicity and Outcome are categorical data; Search count is numerical data. Pie charts can therefore be effectively used to show the distribution of gender and search results with different chars representing distinct ethnicities.

Sankey chart:



Ethnicity and Object of search are two categorical variables from the dataset. Therefore, Sankey char can be used to effectively visualize the flow and relationship between these two variables.

Word cloud:

Ethnicity and Object of Search are both categorical variables, and by fixing one and showing the frequency distribution of the other, these relationships can be effectively visualised using Word Cloud.

All the above graphs except the pie chat for outcome came from the combination of sheet 2 3 4 from the

five-sheet design (see appendix). Each graph illustrates at least one of the factors that influence the police forces' stop and search actions.

Colour choices:
This design mainly uses distinct colour hue to assist with quick distinguishment between different categories. For example, in the pie chart blue represent male and Orange represent female. They both are strong, easily recognizable colour that stands out in a pie chart. Orange is used to contrast with blue. It is warm and bright, making it easily distinguishable from the blue segment. Colour in Sankey chart has lower colour construct than other graphs to reduce visual fatigue caused by complex graphics. The heat map mainly uses the change of colour brightness for assisting the viewer visualise the search frequency. Bright colours indicate fewer values and dark colours indicate more values.

Typographic:
In my project, a uniform font is used for all text elements to maintain consistency and readability and to ensure visual coherence throughout the page. Headings are relatively large to make it easy for users to navigate through the different sections. Some images take up the full width of the page and others take up half the page width for easy side-by-side comparisons. All images and text are centrally aligned which keeps the page clean and organised, making it easier for users to focus on the content.

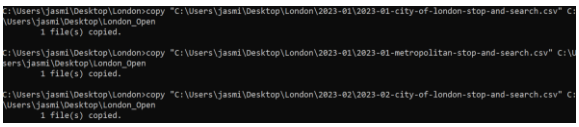
Narrative:
This project mainly uses simple language to describe findings based on visualisations ensure the target audiences' understanding.

3.Implementation

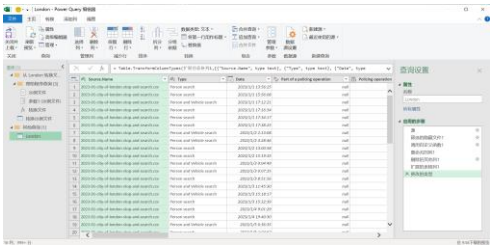
3.1 Technical Implementation

Before implementing the DVP project, five sheet design has been conducted for clear guideline. The first design sheet is used for brainstorming and is responsible for coming up with ideas and screening them (Github, 2024). The second, third and fourth design sheets are the process of trying out different combinations (Github, 2024). The fifth design sheet shows the general direction and layout of the final data visualisation project (Github, 2024). More details will be provided in the appendix. The final design of this DVP did not follow the fifth design table exactly. On the one hand, I eliminated the drop-down selection on the pie charts that illustrate gender factors because the pattern of data in the four pie charts do not show significant differences, and displaying them side by side would allow for clearer comparisons. On the other hand, when I actually conducted the DVP, I found that the analysis on the outcome was missing in the fifth design table. Therefore, I added a pie chart on outcome in the final design.

Original data is merged using excel Power query and cmd from 12 folders including 24 CSV files. The final data set used contains 2206 lines and 17columns.



(Open folders in cmd)



(Merged in Power Query)

From a technical perspective, this project is implemented by R studio assist with some packages.

1) Used the data.table package to export the necessary variables from the original CSV file into several smaller ones, which optimizes the runtime for the final project.

```
#London_Gender <- London_Data[, c("Gender","Officer.defined.ethnicity")]
#fwrite(London_Gender, "London_Gender.csv")
#London_Objects <- London_Data[, c("Officer.defined.ethnicity", "Object.of.search", "search_cases")]
#fwrite(London_Objects, "London_Objects.csv")
#London_Age_Eth <- London_Data[, c("Age.range", "Officer.defined.ethnicity")]
#fwrite(London_Age_Eth, "London_Age_Eth.csv")
```

2) Read prepared files into R Studio

```
# Read file
#London_Data <- read.csv("modified_merged_london.csv")
London_Gender <- read.csv("London_Gender.csv")
London_Objects <- read.csv("London_Objects.csv")
London_Age_Eth <- read.csv("London_Age_Eth.csv")
London_Police <- read.csv("UKPolice_Force_Areas.csv")
London_Outcome <- read.csv("London_Outcome.csv")
```

3) Use ggplot2, wordcloud2 and plotly produce interactive graphs before define UI and Server.

```
# Define initial plots

# Sankey chart
create_sankey_plot <- function(data) {
  # Create unique nodes
  nodes <- unique(c(data$Officer.defined.ethnicity, data$Object.of.search))
  node_indices <- setNames(seq_along(nodes) - 1, nodes)

  # Map sources
  data$source <- node_indices[data$Officer.defined.ethnicity]
  data$sink <- node_indices[data$Object.of.search]

  # Define colors
  node_colors <- c("Black" = "#6b8e23", "Asian" = "#d2b48c", "White" = "#87ceeb", "Other" = "#dda0dd")

  # Create a color vector for all nodes
  colors <- rep("#1f77b4", length(nodes))
  colors[match(names(node_colors), nodes)] <- node_colors

  # Create the plot using plot_ly
  plot_ly(
```

4) For building UI, HTML is used for common structure, fluid page is used for graph and word positioning.

```
# Define UI
ui <- fluidPage(
  tags$head(
    tags$style(
      HTML("
        /* Add Frame */
        .page-container {
          border: 2px solid #ccc;
          padding: 20px;
          background-color: #f0f8ff;
        }
        /* Center the word cloud */
        #wordcloud {
          margin: 0 auto;
        }
        /* Center the selectInput and selectizeInput */
        .input-container {
          display: flex;
          justify-content: center;
        }
```

```
titlePanel("The influence of Age, Gender and Ethnicity on London Police forces Stop and Search Actions"
div(class = "page-container",
    tabsetPanel(
      tabPanel("Analysis",
               fluidPage(
                 titlePanel("Data Analysis"),
                 tags$style(
                   HTML(".fluid-row {
                      display: flex;
                      justify-content: center;
                      align-items: center;
                   }")
                 ),
                 fluidRow(
                   column(3, div(class = "plot-container", plotlyOutput("pieChartBlack"))),
                   column(3, div(class = "plot-container", plotlyOutput("pieChartAsian"))),
                   column(3, div(class = "plot-container", plotlyOutput("pieChartWhite"))),
                   column(3, div(class = "plot-container", plotlyOutput("pieChartOther")))
```

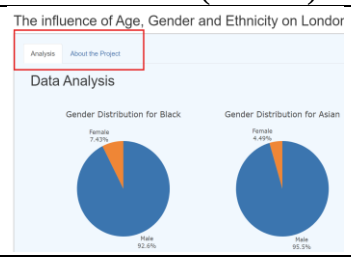
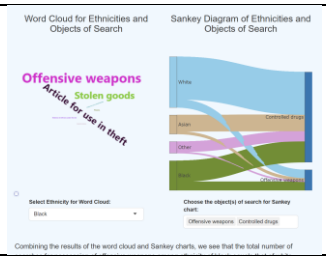
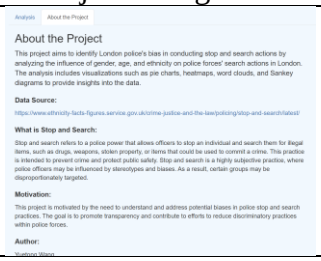
5) Generate output through server

```r
# Define server logic
server <- function(input, output, session) {

  # Filter data based on selected ethnicity for word cloud
  filtered_wordcloud_data <- reactive({
    subset(London_objects, Officer.defined.ethnicity == input$cloudEthnicity)
  })

  # Filter data based on selected object of search for Sankey chart
  filtered_sankey_data <- reactive({
    subset(London_objects, Object.of.search %in% input$sankeySearch)
  })

  # Create word cloud based on filtered data
  output$wordcloud <- renderWordcloud2({
    create_wordcloud(filtered_wordcloud_data())
  })

  # Create Sankey chart based on filtered data
```

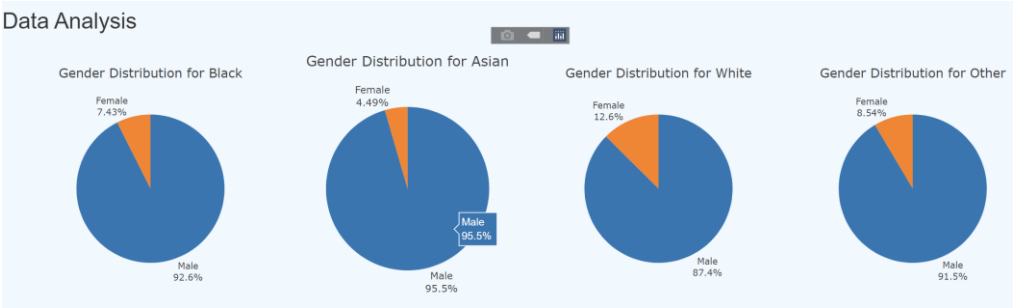
3.2Interactive Narrative Visualisation Implementation

3.2.1 Tabs

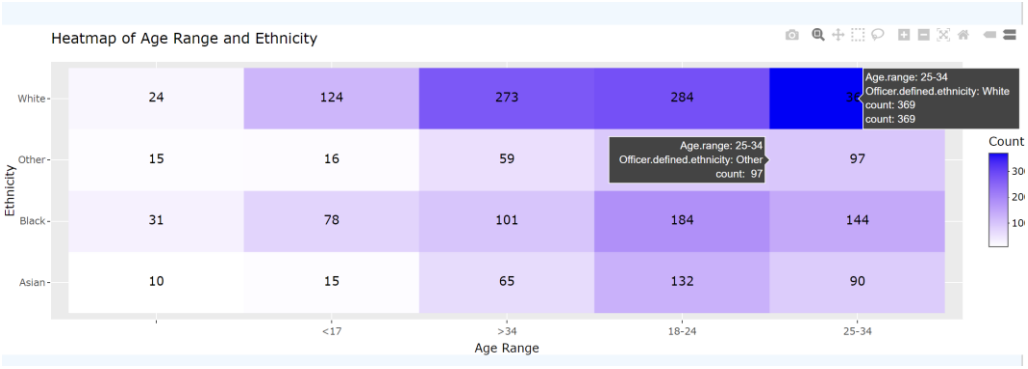
There are two main tabs within the app. One for data visualisation with explanation and one for project background and details.

Tab location (red box)	Visualisation	Project background
		

3.2.2User friendly design

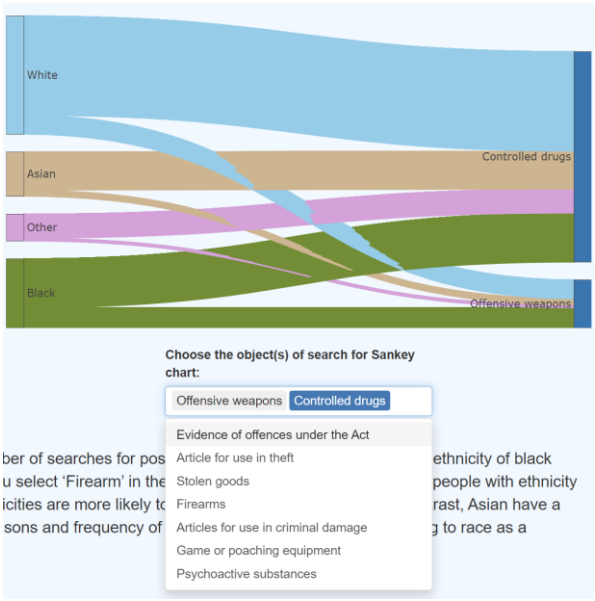


The graph that user hove over will be slightly emphasized and labels are clearly provided for each graph.

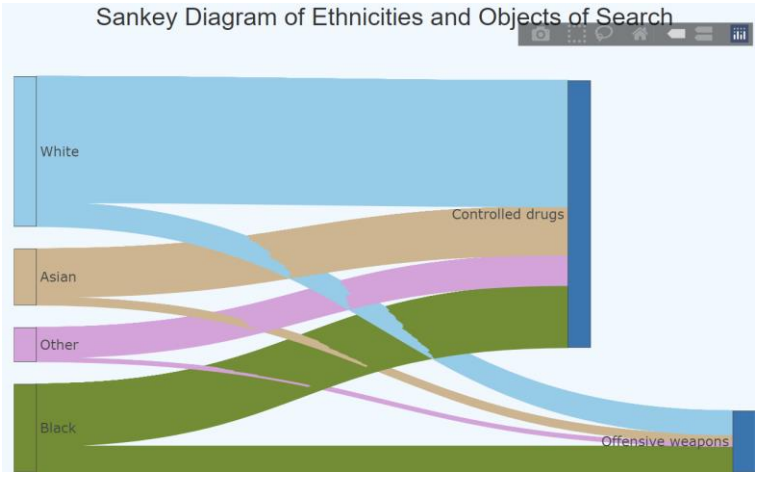


In heat map specifically, user can choose to show two chosen labels for comparison.

A list of other functions is also provided by clicking the icon on the top right corners of the graphs. From left to right respectively: Save graph, emphasise chosen section, Drag the graph, Box select, Lasso select, Zoom in/out, Auto Scale, Reset graph, Show label, Compare labels.



Due to the number of objects of search for, in order to provide a clear view of the Sankey chart, a multi-selection drop-down menu is provided. To add another object of search, the user could click the selection box, then click on the choices list. To remove an option, the user could click on the chosen one, then click “Delete” on the keyboard. The Sankey chart will be automatically refreshed once the selection or deletion is complete. If the page is refreshed or the graph is reset, "Offensive Weapons" and "Controlled Drugs" are set as the default values.



Additionally, the target node on the right hand side of the Sankey chart can be dragged to provide a clearer view.

For word cloud and Pie chart, dropdown selection for ethnicities is provided.

3.2.3Word Explanation

Word explanation could be found under or beside each graph.

3.3 Using the Implementation

Overall, the visualisation project should be clear and easy to use. Please do not miss out on the “About the project” tab under the title next to the “Analysis” tab. (Location guide; see Section 5.2.1.)

To run this R Shiny app, I provide the code and the data used in the same folder. Please download the folder in your local PC’s working directory in order to run smoothly.

Please refresh the page after clicking on full screen; the graph and word explanation location will adjust your screen size after refreshing.

4. Conclusion

From the visualisations, we can conclude that age, gender, and ethnicity have significant impacts on police officer's stop and search actions. Gender-wise, Asian females are the least likely to be identified as suspicious, and white females, even if they are most likely to be identified as suspicious among females, still only account for 40 percent of overall white searches. This indicates that when the police conduct search actions, they are most likely to choose males. For ethnicities other than white, the most likely age group to be searched is 18-24, while for the white race it is 25-34. Despite having a smaller population than Asians, individuals identified as Black by officers are still more likely to be searched than Asians. They are also most frequently searched for possessing offensive weapons, whereas the primary search reason for White and Other ethnicities is possession of stolen goods. The reasons for searches are more evenly distributed in the Asian group. From the search outcome, guilty rate for people with white ethnicity is more than five precent above all the other ethnicity group.

However, there are still some limitations to this project. Firstly, the dataset used contains significant number of missing values. In terms of geographical scope, this analysis focuses only on the London area and the findings may not be generalisable to other parts of the UK, as regional differences and population demographic might defer. Additionally, ethnicity categorisation, which is based on officer-defined categories, may not accurately reflect an individual's self-identification, which can introduce bias Whilst the analysis reveals differences, it cannot be concluded that these biases necessarily exist as each individual is subjective in different ways. From an officer's perspective, any suspicious behaviour requires a search to maximise public safety. Further investigation such as policy analyses, will be required to understand the reasons for these results.

5. References (Including references used in code)

Github. (2024). *Five Design-Sheets*. Fds-Design.github.io. https://fds-design.github.io/

GOV.UK. (2011). *Police powers to stop and search: your rights*. GOV.UK.

https://www.gov.uk/police-powers-to-stop-and-search-your-rights

Healy, Y. H. and C. (2024). *From data to Viz | Find the graphic you need*. Www.data-To-Viz.com.

https://www.data-to-viz.com/#sankey

Holtz, Y. (2018). *Sankey Diagram for energy consumption*. Www.r-Graph-Gallery.com. https://r-graph-gallery.com/323-sankey-diagram-with-the-networkd3-library.html

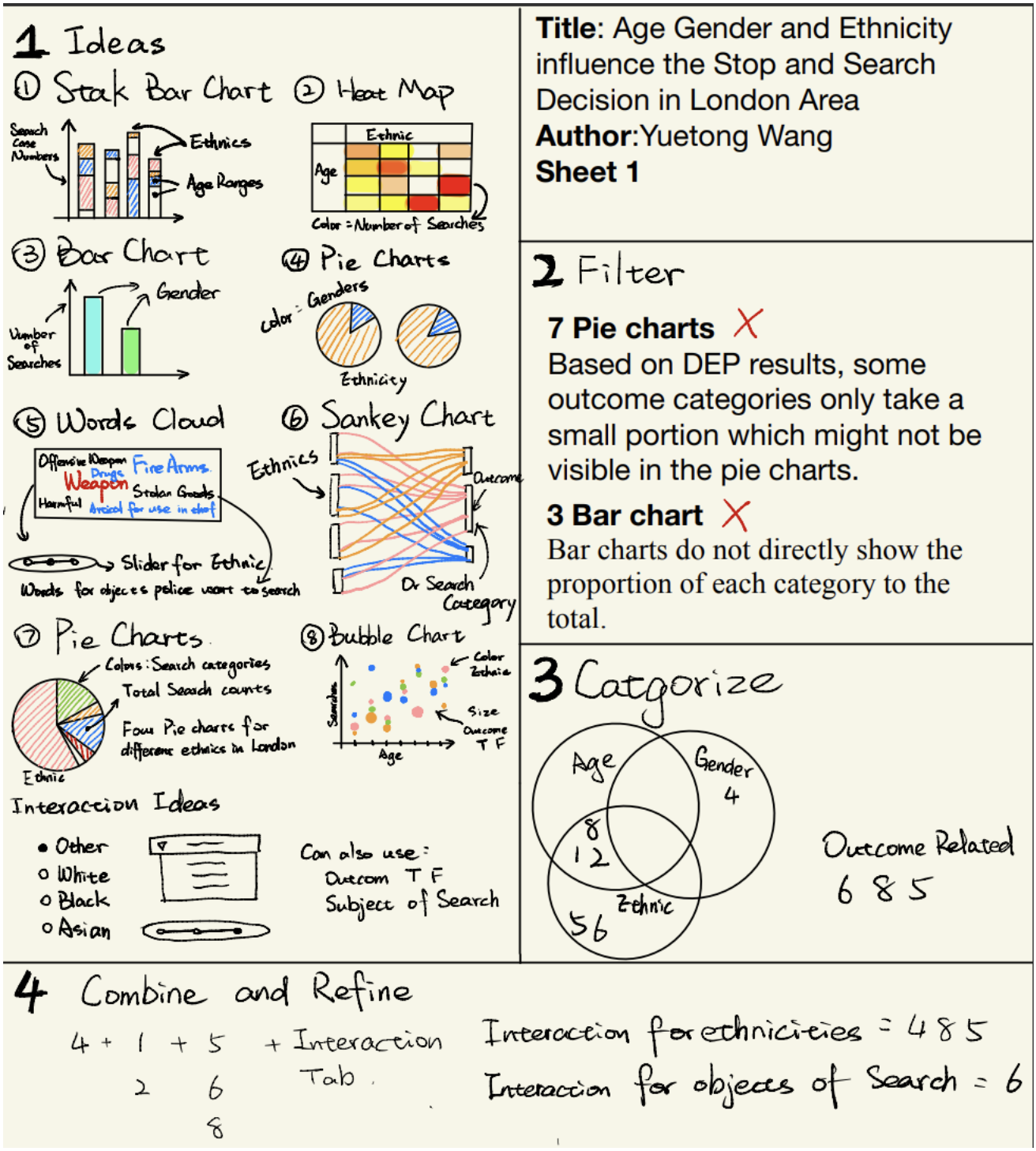
ONS. (2022, December 22). *Regional Ethnic Diversity*. Www.ethnicity-Facts-Figures.service.gov.uk. https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/regional-ethnic-diversity/latest/

Plotly. (2024). *Pie*. Plotly.com. https://plotly.com/r/reference/pie/

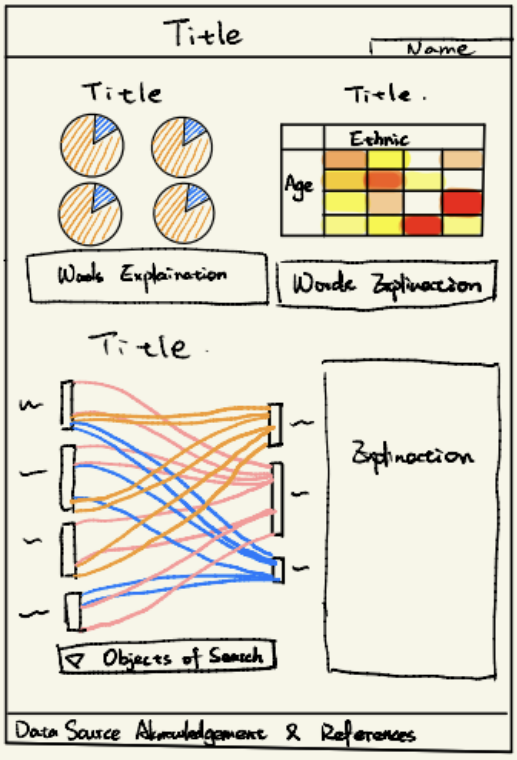
Rhoden-Paul, A. (2022, April 20). Stop and search: Ethnic minorities unfairly targeted by police - watchdog. *BBC News*. https://www.bbc.com/news/uk-61167875

Shiny. (n.d.). *html r shiny*. Bing. Retrieved June 2, 2024, from https://cn.bing.com/search?q=html+r+shiny+&qs=n&form=QBRE&sp=-1&lq=0&pq=html+r+shiny+&sc=8-13&sk=&cvid=8A242373BD6E4518924B759EEBFA34AB&ghsh=0&ghacc=0&ghpl=

6. Appendix


1 Ideas
① Stak Bar Chart
② Heat Map
Search Case Numbers
Ethnics
Age Ranges
Ethnic
Age
Color = Number of Searches
③ Bar Chart
④ Pie Charts
Gender
Number of Searches
color: Genders
Ethnicity
⑤ Words Cloud
Offensive Weapon
Fire Arms
Drugs
Weapon
Stolen Goods
Harmful
Article for use in theft
Slider for Ethnic
Words for objects police want to search
⑥ Sankey Chart
Ethnics
Outcome
Or Search Category
⑦ Pie Charts
Colors: Search categories
Total Search counts
Four Pie charts for different ethnics in London
Ethnic
⑧ Bubble Chart
Color Ethnic
Size Outcome T F
Searches
Age
Interaction Ideas
Other
White
Black
Asian
Can also use:
Outcom T F
Subject of Search
Title: Age Gender and Ethnicity influence the Stop and Search Decision in London Area
Author:Yuetong Wang
Sheet 1
2 Filter
7 Pie charts X
Based on DEP results, some outcome categories only take a small portion which might not be visible in the pie charts.
3 Bar chart X
Bar charts do not directly show the proportion of each category to the total.
3 Catgorize
Age
Gender
4
8
12
Ethnic
56
Outcome Related
6 8 5
4 Combine and Refine
4 + 1 + 5 + Interaction Tab
2 6
8
Interaction for ethniciteies = 4 8 5
Interaction for objeces of Search = 6

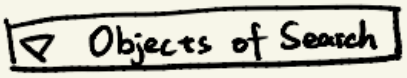
Layout



Title: Age Gender and Ethnicity influence the Stop and Search Decision in London Area
Author:Yuetong Wang
Sheet 2

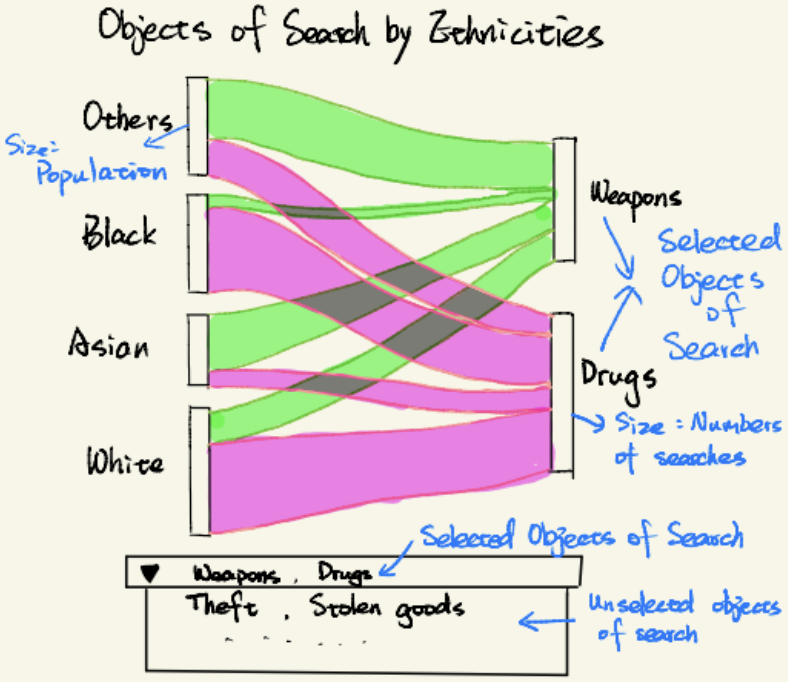
Operations

There will be only one interaction in this layout.



Users will be able to select objects of search, providing them with a detailed overview of how many people of different ethnicity's are being searched for different objects.

Focus



Discussion

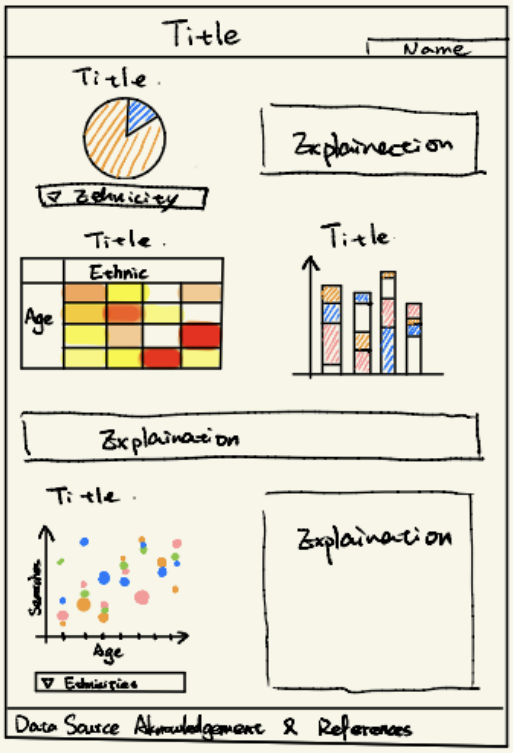
✓

The selection panal gives users enough freedom to select objects of search which simplifies the sankey chart and enhance readability.

✗

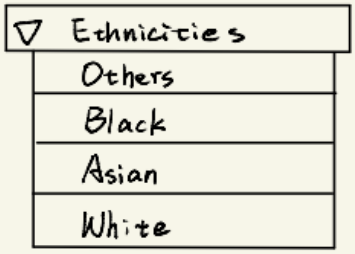
The portion of different ethnicities may not be obvious in sankey chart

Layout



Title: Age Gender and Ethnicity influence the Stop and Search Decision in London Area
Author:Yuetong Wang
Sheet 3

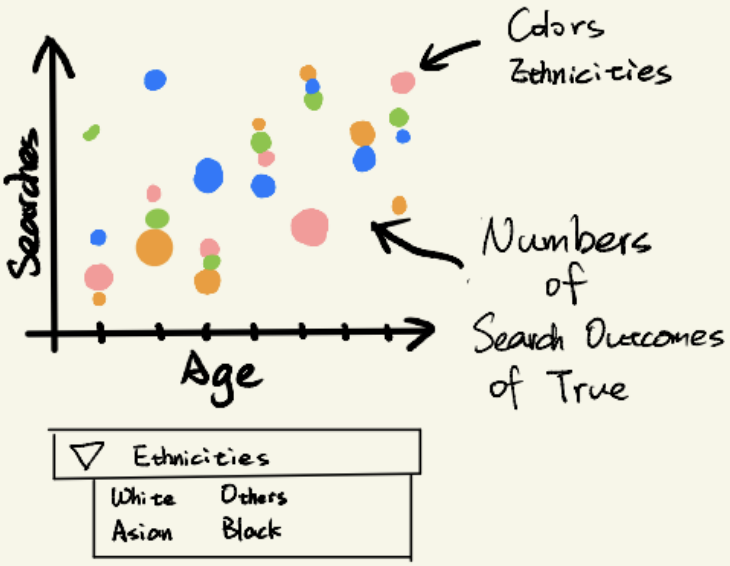
Operations



Offers two selection of four different ethnicities.

Focus

Age and Number of Searches by different ethnicities



Discussion

✓

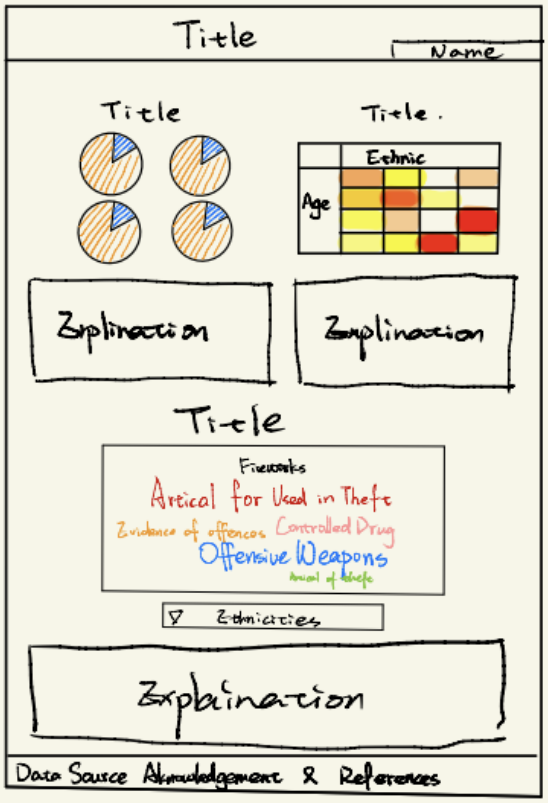
Two selection bars, save space and offer more flexibilities.

✗

No information about objects of searches.

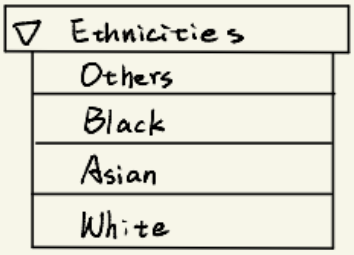
No significant differences of the size in the bubble chart.

Layout



Title: Age Gender and Ethnicity influence the Stop and Search Decision in London Area
Author:Yuetong Wang
Sheet 4

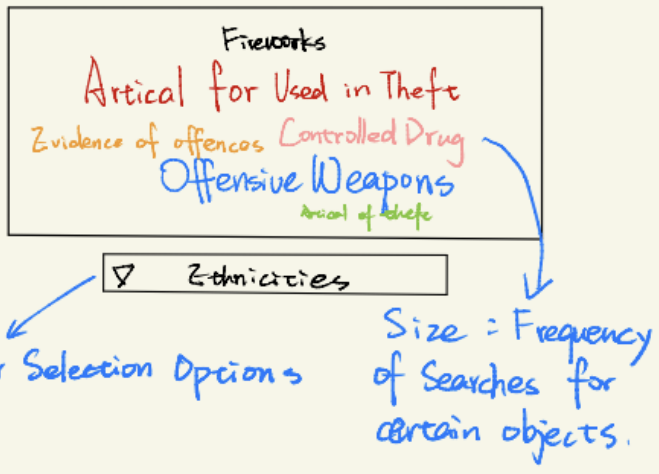
Operations



Offers selections of four ethnicities for wordclouds

Focus

Wordclouds for Objects of Search by Ethnicities



Discussion

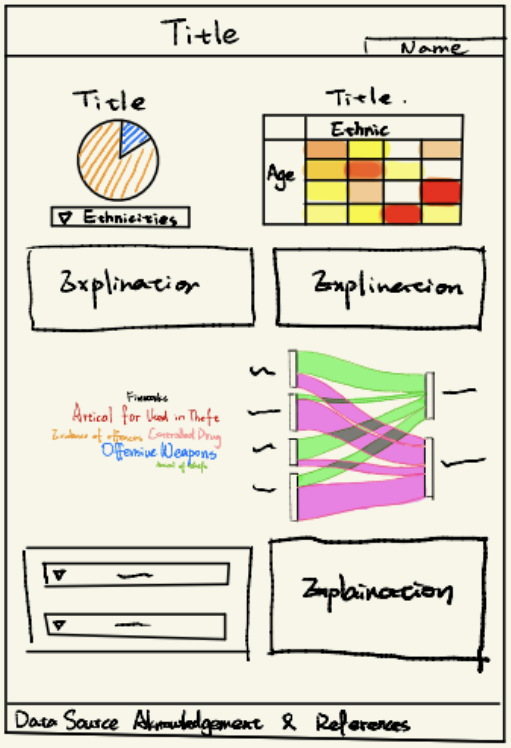
✓

Use heat map for identifying both age and ethnicities' effects on searches.

✗

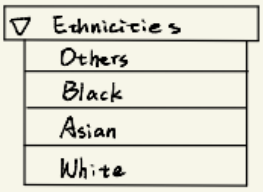
Does not provide a general view of the influence of ethnicities on objects of search.

Layout

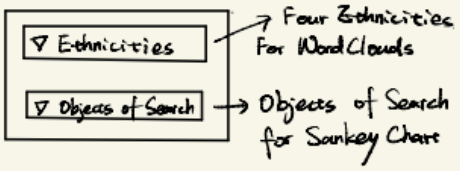


Title: Age Gender and Ethnicity influence the Stop and Search Decision in London Area
Author:Yuetong Wang
Sheet 5

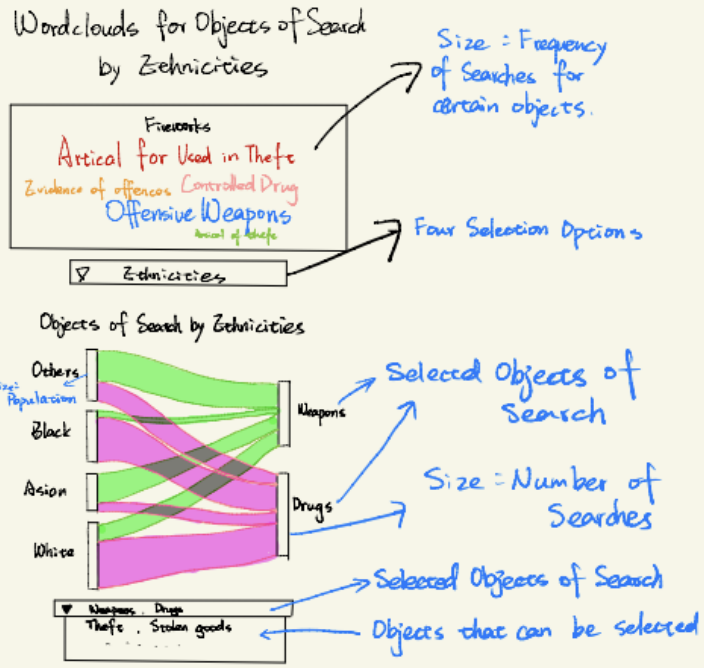
Operations

- 

Offers selections of four ethnicities for wordclouds

- 

Focus



Details

- Cleaned Data from Stop and Search London
- Use R Shiny fluid Page for layout
- Packages : R Shiny , ggplot2 wordclouds 2