Police Arrests and Ethnic Background

Aviral Bhardwaj

January 21, 2024

1 Introduction

2 Data

This data is sourced from the Toronto Open Data Portal `opendatatoronto` (Gelfand 2022).
The following data is utilized for a more in-depth analysis of arrests and strip searches involving
various ethnic and identity-based groups in Toronto. The information is gathered by the
Toronto Police Service (TPS) and is collected under the authority of the Police Services Act. Its
purpose is to enhance comprehension of the relationship between the police and the community.
The data covers the period from 2020 to 2021.

The data underwent cleaning and analysis using the R programming language (R Core Team
2022). Cleaning was performed with the `tidyverse` package (Wickham et al. 2019), involv-
ing the removal of unnecessary columns for the analysis. Subsequently, analysis was con-
ducted utilizing the `dplyr` package (Wickham et al. 2022), with the data grouped by ethnic
and identity-based categories. Visualization was then executed using the `ggplot2` package
(Wickham 2016), depicting the number of arrests and strip searches across various ethnic and
identity-based groups.

The raw data included 32,000 arrests data with 26 variables of analyzable data, that included
variables such as Arrest year, Arrest_Month, Perceived Race, Age Group, Youth at Arrest
and Booked which we will be using for the paper.

```
cleaned_data <- read.csv("../data/cleaned_data.csv")

table <- cleaned_data |>
  group_by(arrest_year, perceived_race, search_reason) %>%
  summarise(n = n(), .groups = "drop") %>%
  arrange(desc(perceived_race)) %>%
  pivot_wider(names_from = arrest_year, values_from = n, values_fill = 0)
```

```
table <- pivot_longer(table, cols = c("2020", "2021"), names_to = "year", values_to = "val

ggplot(table, aes(x = perceived_race, y = value, fill = factor(search_reason))) +
  geom_bar(stat = "identity", position = "fill") +
  facet_wrap(~ year, scales = "free_y") +
  labs(title = "Percent Stacked Bar Chart Based on 'booked'",
       x = "Perceived Race",
       y = "Percentage",
       fill = "Search Reason") +
  theme_minimal() +
  scale_fill_discrete(labels = c("Not Searched", "Searched")) +
  theme(legend.position = "right") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```
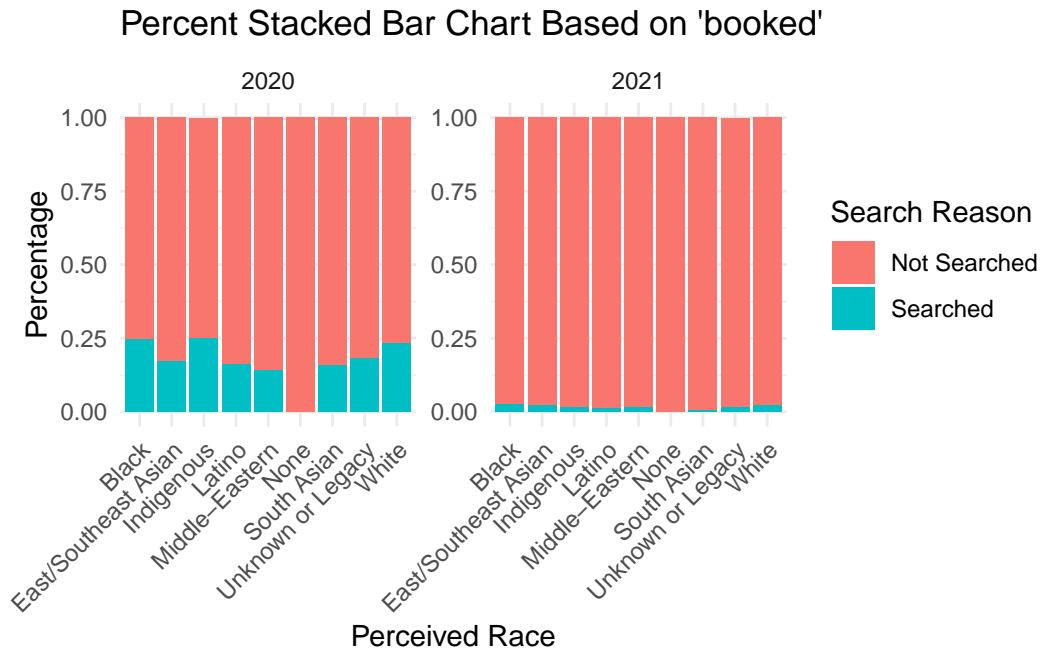

Table 1: Number of arrests by ethinic background for the year 2020, and 2021

Ethinic and Identity Based Groups	2020	2021
White	7236	7188
Black	4657	4524
Unknown or Legacy	1235	1144

Ethnic and Identity Based Groups	2020	2021
East/Southeast Asian	733	900
South Asian	722	723
Middle-Eastern	593	673
Indigenous	534	497
Latino	315	324
None	1	1

Table 2: Arrests Booked by ethnic background for the year 2020, and 2021

Ethnic and Identity Based Groups	2020	2021
White	3947	3673
Black	2699	2524
Unknown or Legacy	641	570
East/Southeast Asian	393	432
South Asian	407	373
Middle-Eastern	303	385
Indigenous	310	258
Latino	176	183

```
ggplot(cleaned_data, aes(x = perceived_race, y = value, fill = factor(booked))) +
  geom_bar(stat = "identity") +
  facet_wrap(~ year, scales = "free_y") +
  labs(title = "Stacked Bar Chart Based on 'booked'",
       x = "Perceived Race",
       y = "Value",
       fill = "Booked") +
  theme_minimal() +
  scale_fill_discrete(labels = c("Not Booked", "Booked")) +
  theme(legend.position = "right") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```
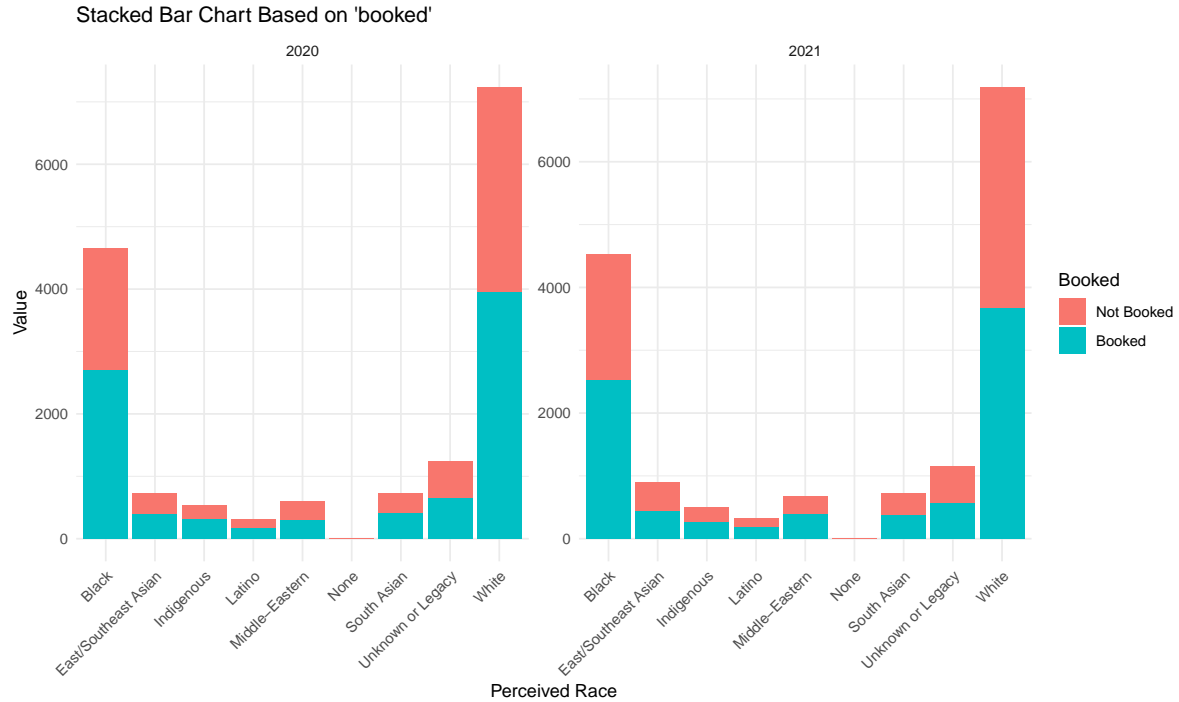
Stacked Bar Chart Based on 'booked'

Figure 1: Number of arrests by ethinic background for the year 2020, and 2021

References

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal.* https://CRAN.R-project.org/package=opendatatoronto.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.
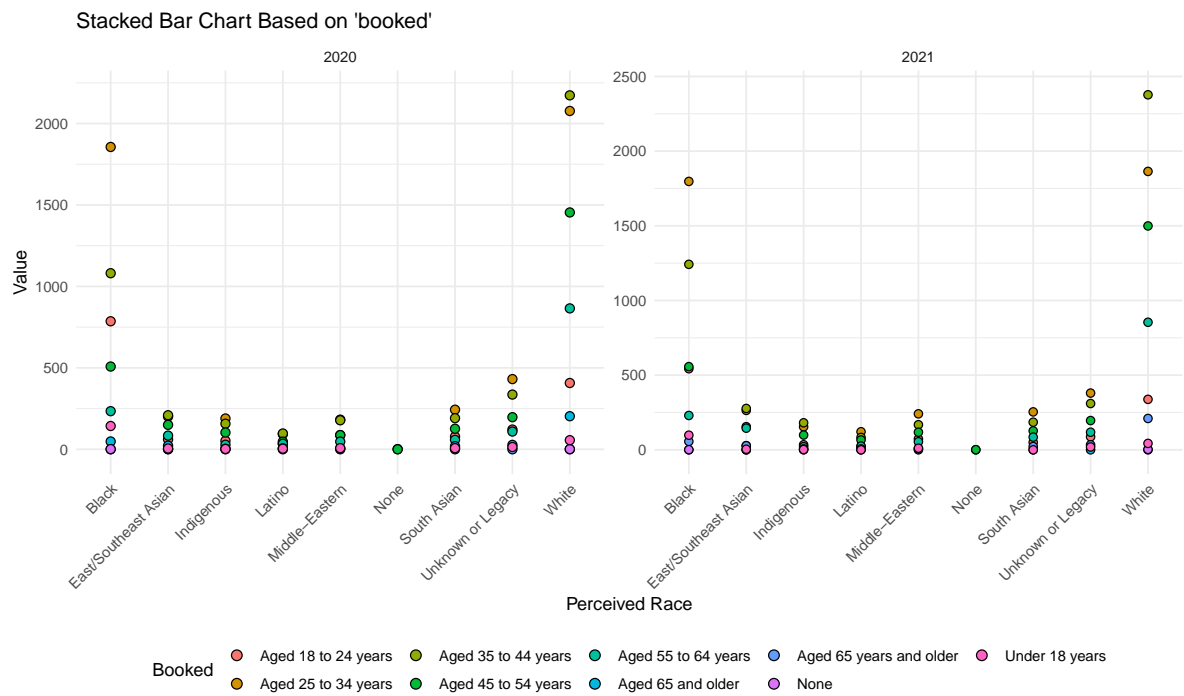
Figure 2: Number of arrests by ethinic background for the year 2020, and 2021
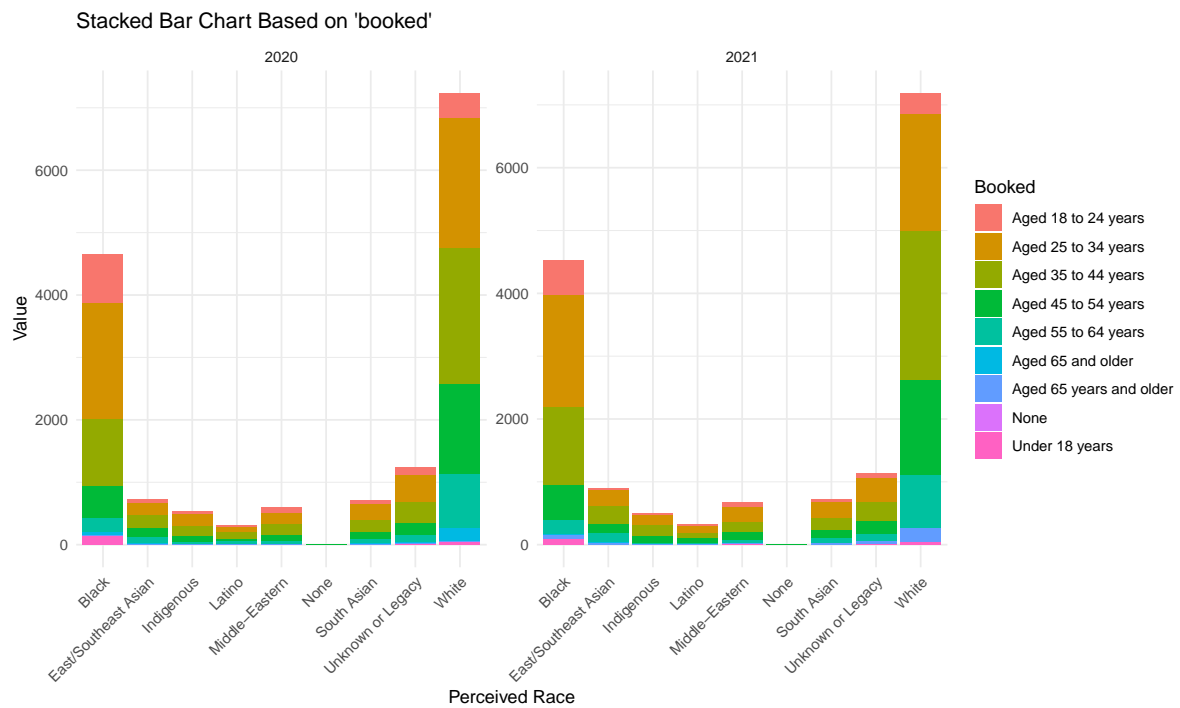
Stacked Bar Chart Based on 'booked'

Figure 3: Number of arrests by ethinic background for the year 2020, and 2021