

Police Arrests and Ethnic Background

Aviral Bhardwaj

January 19, 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), involving the removal of unnecessary columns for the analysis. Subsequently, analysis was conducted 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.

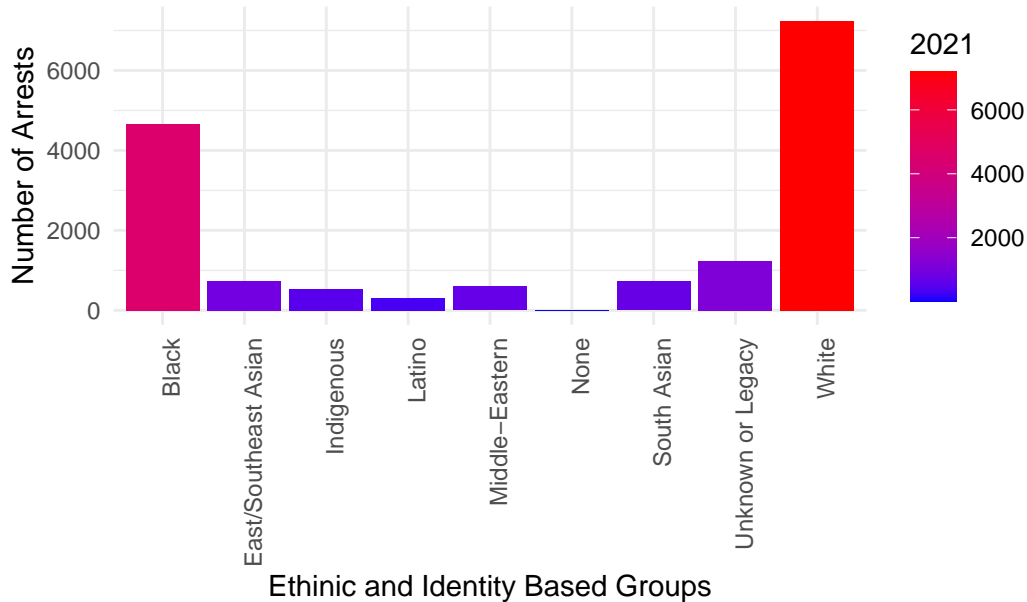
The raw data included 32,000 arrests data with 26 variables of analyzable data, that included variables such as Arrest year, Arrest_Month, Perceived Race, Age Group, Youth at Arrest and Booked which we will be using for the paper.

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

Ethnic and Identity Based Groups	2020	2021
White	7236	7188
Black	4657	4524
Unknown or Legacy	1235	1144
East/Southeast Asian	733	900

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

Number of arrests by ethnic 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 Golemund, 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>.