Analyzing Arrest Patterns and Racial Representation in Toronto*

Akshat Aneja

September 23, 2024

The city of Toronto has an extremely diverse population in terms of its ethnicity and culture. This paper examines arrest data from the Toronto police service between 2020 and 2021. It focuses on the distribution of arrests by ethnic identity, age and booking status. This is done to check whether policing practices exhibit systemic bias. By cleaning and analyzing the records, trends are explored over major ethnic groups. Despite concerns over racial bias, the data suggests that arrests are influenced by objective factors, though there still may be disparities due to socio-economic status. These findings may help in providing insights that may enhance transparency in policing practices.

Table of contents

*Code and data supporting this analysis is available at: https://github.com/Akshat211202/Analysis-of-bias-in-arrests/

1 Introduction

In recent years, debates surrounding racial bias in policing have intensified, particularly regarding its disproportionate impact on marginalized communities. This issue is especially pertinent in Toronto, a culturally and ethnically diverse city. Discussions on police reform often reflect historical connections to slavery and contemporary racial disparities, raising concerns about whether law enforcement treats marginalized communities differently, potentially leading to racial profiling.

A report from Toronto Police Services indicates that Black, East/Southeast Asian, Middle Eastern, and Latino individuals are overrepresented in use-of-force incidents relative to their population in enforcement actions. Despite efforts by Toronto Police to promote a more equitable society, these findings suggest the continued presence of racial bias in arrests.This paper analyzes arrest data from 2020 and 2021 to explore interactions between police and various ethnic groups. Disparities in arrest rates may reflect broader socio-economic factors such as poverty and unemployment, which disproportionately affect marginalized communities and contribute to higher involvement in criminal activities.

Arrest data provides a quantifiable view of law enforcement actions and their correlation with ethnicity and age. Section 2, outlines the data cleaning process using R, followed by an analysis of arrest distributions across age groups to identify potential trends. Age is a critical factor in this analysis, as it may reveal patterns of concern. Section 3, presents findings and discusses their implications. The study seeks to contribute to a deeper understanding of policing in Toronto, providing insights for policy reforms and community engagement efforts

2 Data

This data is sourced from the Toronto Open Data Portal `opendatatoronto` (Gelfand 2022). The following data is used for an in-depth analysis of arrests and strip searches which involve various ethnic and identity-based groups in the city of Toronto. The information is gathered by the Toronto Police Service (TPS) (Data 2022) and is collected under the authority of the Police Services Act. Its purpose is to improve the relationship between the police and the community by being more transparent with the public. The data covers the period from 2020 to 2021.

The data underwent cleaning and analysis using the R programming language (R Core Team 2023). 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 dataset consisted of 32,000 arrest records from 2020 and 2021, with 26 variables available for analysis. For this study, we concentrated on specific variables, including Arrest Year, Perceived Race, Age Group, Youth at Arrest, Booked, Search Reason Cause Injury, Search Reason Assist Escape, Search Reason Possess Weapons, and Search Reason Possess Evidence. To streamline the data, we combined similar columns—Search Reason Cause Injury, Search Reason Assist Escape, Search Reason Possess Weapons, and Search Reason Possess Evidence—into a single column called "Search Reason." utilizing the `any()` function (Wickham et al. 2022).

2.1 Total number of arrests by age group for the year 2020 and 2021

It is crucial to analyze the distribution of arrests across different age groups to understand potential patterns and disparities. Table 1 presents a summary of arrest counts for various ethnic groups across all age categories in 2020 and 2021. The age groups included in the dataset are: Under 17, 18 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, and 65 years and over.

Given the complexity and extensive nature of the dataset, which involves nine age groups across eight ethnicities, summarizing the entire dataset in a comprehensible manner is challenging. To address this and ensure clarity, we have selected a subset of the data focusing specifically on individuals aged 25 to 34 years. This group was chosen as it represents a significant proportion of the data and allows for a more straightforward comparison of arrest trends across ethnic groups while avoiding information overload. By narrowing the scope, the analysis becomes more manageable and the insights more focused.

Table 1: Arrests Booked by ethinic background for the year 2020, and 2021

Ethinic and Identity Based Groups	Age Group	2020	2021
White	Aged 25 to 34 years	4179	4112
Black	Aged 25 to 34 years	3234	3098
Unknown or Legacy	Aged 25 to 34 years	804	842
East/Southeast Asian	Aged 25 to 34 years	714	497
South Asian	Aged 25 to 34 years	614	512
Middle-Eastern	Aged 25 to 34 years	553	439
Indigenous	Aged 25 to 34 years	338	384
Latino	Aged 25 to 34 years	356	272
None	Aged 25 to 34 years	0	1

2.2 Arrests by Ethnic and Identity-Based Groups for Year 2020 and 2021

Our analysis primarily focuses on the distribution of arrests across different ethnic identities. We extracted data for the following ethnic groups: Black, White, South Asian, East Asian, Middle Eastern, Latin American, Southeast Asian, and Indigenous.

The data for 2020 and 2021 is summarized in Table 2 and Table 3. Table 2 presents the number of individuals arrested from each ethnic group during these years, while Table 3 details how many of those arrested were subsequently booked under criminal charges.

Table 2: Year 2020 vs Year 2021 Total number of arrests

Race	2020	2021
White	14116	13607
Black	8878	8648
Unknown or Legacy	2444	2612
East/Southeast Asian	2361	2054
South Asian	1871	1742
Middle-Eastern	1730	1507
Indigenous	935	999
Latino	960	808

Table 3: Year 2020 vs Year 2021 Total number of Booked

Race	2020	2021
White	7185	7075
Black	4885	4787
Unknown or Legacy	1288	1149
East/Southeast Asian	1037	1057
South Asian	901	876
Middle-Eastern	741	891
Indigenous	583	491
Latino	476	489

It is important to note that the OpenDataToronto (Gelfand 2022) provides population distribution data by ethnicity. However, due to the size and diversity of Toronto and its numerous wards, it has been challenging to cross-verify this data with crime rate statistics in a comprehensive manner. Further, a more detailed analysis could be conducted by comparing specific ethnic groups directly.

Now that the data has been cleaned, we can begin to analyze the data. The following section will provide a summary of the data, and the trends that emerge from the data. Figure 1,

Figure 2, and Figure 3, was created using `ggplot` (Wickham 2016), displays the information that can be used as compare the trends as described above.

2.3 Total Arrests in 2020-2021

Referring to Table 2, Table 3 and Figure 1 the data reveals a non-uniform distribution of arrests across ethnic groups, with the Black and White populations showing the highest numbers of arrests in both 2020 and 2021. Additionally, the data indicates a decline in arrests for these two ethnic groups from 2020 to 2021. In contrast, arrests for other ethnic groups, such as South Asian, East Asian, and Indigenous populations, increased over the same period. These opposing trends highlight the divergent patterns of arrest across different ethnic groups during this time.

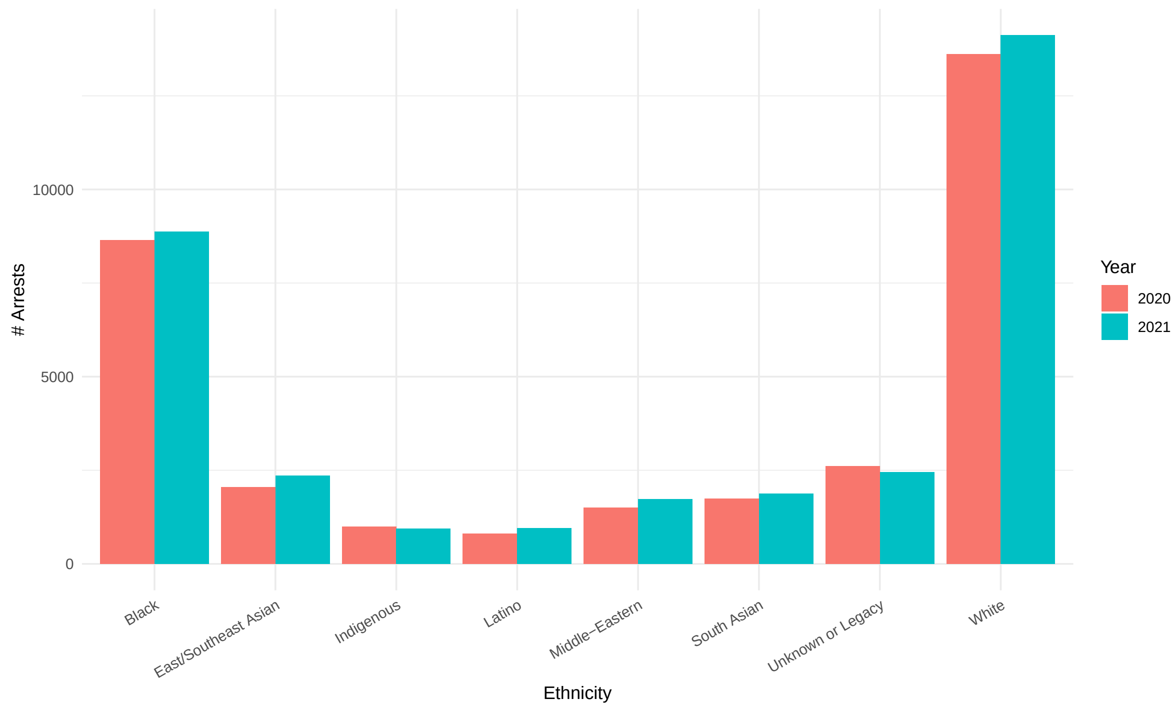

Figure 1: # arrests by Ethnicity in Year 2020 and 2021

Referring to Table 1 for the proportion of arrests that resulted in bookings, Figure 2 was plotted with the y-axis representing the percentage of individuals booked after arrest. The data demonstrates a consistent distribution across various ethnic groups, with significantly more individuals being booked following arrest than those who were not. Notably, the percentage of individuals booked ranges between 50% and 55% across all ethnic groups, a pattern that remains consistent in 2021.

However, it is worth highlighting that the Black population had the highest number of individuals booked after arrest in both 2020 and 2021.

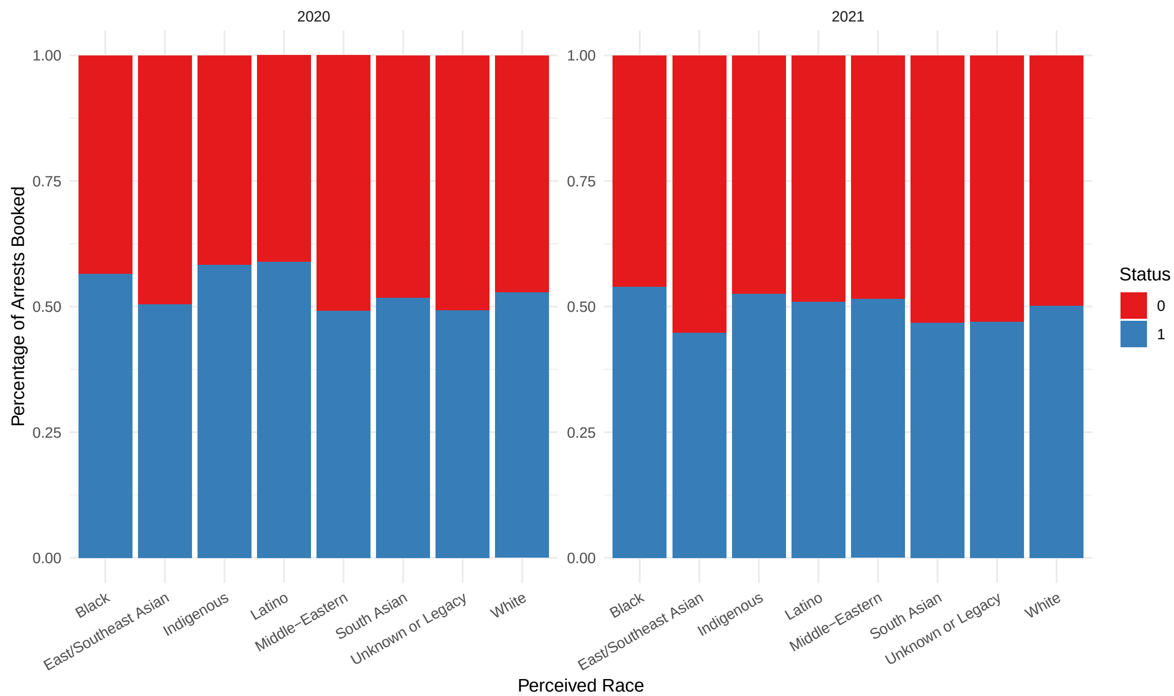

Figure 2: Booked after Arrest in the Year 2020 and 2021

2.4 Number of Arrests by Age Group for the year 2020-2021

As depicted in Figure 3, the data reveals a non-uniform distribution of arrests across age groups. The 25 to 34 and 35 to 44 age groups had the highest number of arrests in both 2020 and 2021. Additionally, the data indicates a notable decrease in arrests for these two age groups between 2020 and 2021. In contrast, other age groups experienced an increase in arrests during the same period, indicating differing trends across age demographics.

3 Discussion

While this paper offers several important insights, it is important to acknowledge its limitations. A more detailed comparison of specific ethnic groups and further research are necessary for a deeper understanding of the data. Law enforcement agencies and community advocates can use these findings to develop targeted interventions and enhance community engagement efforts. Additionally, policymakers may consider using this analysis to inform necessary policy

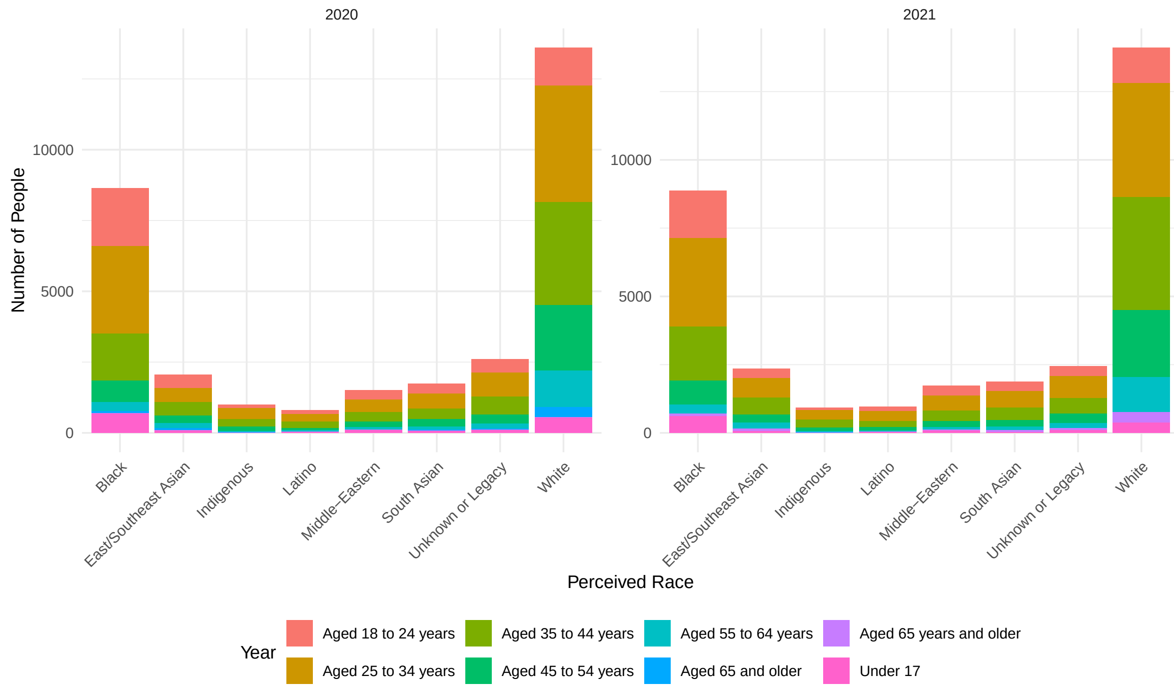

Figure 3: Number of arrests by Age-Group for the year 2020, and 2021

adjustments. The need for transparency and accountability in policing remains critical, as emphasized by the Toronto Police Services Board (Toronto Police Services Board 2019).

In conclusion, the analysis of Toronto's arrest data suggests that, despite concerns about racial bias, arrests appear to be influenced by objective factors rather than racial profiling. The reforms implemented by the Toronto Police Service have contributed to increased transparency and fairness within the police force. Although disparities may exist due to socio-economic factors, the data does not indicate systemic bias.

References

Data, Toronto Open. 2022. "POLICE RACE AND IDENTITY BASED DATA - ARRESTS AND STRIP SEARCHES." https://open.toronto.ca/dataset/police-race-and-identity-based-data-collection-arrests-strip-searches/.

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

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto Police Services Board. 2019. "Toronto Police Services Board's Race-Based Data Collection Policy." Report. Toronto Police Services Board. https://www.tpsb.ca/policies-by-laws/board-policies/177race-based-data-collection-analysis-and-public-reporting.

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.