No Evidence of Bias in Arrests Made by the Toronto Police Service*

Aviral Bhardwaj

January 24, 2024

The ethnic identity of citizens arrested by the police has consistently been a topic of discussion, often highlighting the pronounced skewness of certain ethnic identities. This paper seeks to analyze data from the Toronto Open Data Portal to investigate whether there is any bias in arrests made by the Toronto Police Service between 2020 and 2021, while accounting for factors such as age, booked after arrest, and search on arrest. While the population distribution of various ethnic identities varies significantly, it was found that ….. Further analysis can provide valuable insights for policymakers, law enforcement agencies, and community advocates in fostering a more equitable society while emphasizing the importance of promoting transparency and accountability in policing.

Table of contents

*Code and data supporting this analysis is available at: https://github.com/Aviral-03/Ethnic-Identity-Arrests-Toronto

1 Introduction

In the vibrant metropolis of Toronto, the intersection between law enforcement practices and demographic characteristics holds profound implications for community dynamics and public policy. Toronto, named Canada's most diverse city, encounters numerous challenges arising from the coexistence of ethnically diverse populations within the same neighborhoods. According to an article published by Georgetown University, the "ongoing debate surrounding police reforms is heavily influenced by historical connections with slavery and contemporary racial disparities" (Jackson (2022)). This statement underscores the contentious nature of whether the race identity affiliation of police officers affects their policing practices, despite the expectation for fair and unbiased treatment.

In a recent article published by the Toronto Police Services, it is highlighted that "Black, East/Southeast Asian, Middle Eastern, and Latino people were over-represented in reported use of force incidents compared to their presence in the enforcement action population" (Fanfair (2022)). This indicates that, despite Toronto Police Services working towards a more equitable society, there are still hints of racial bias in the arrests made by the Toronto Police Services.Conversely this paper undertakes a comprehensive analysis of arrest data of various race and ethinic iden spanning the years 2020 and 2021 in the city of Toronto, seeking to unravel the intricate patterns that shape the city's legal landscape.

Arrest data serves as a vital lens through which we can gain insights into the dynamics of law enforcement activities and their correlation with ethnicity and age. In the data section we clean and process our data. Following which in our discussion and results our inquiry begins with a broad exploration of the distribution of arrests across various age groups, aiming to discern overarching trends and fluctuations. The age spectrum, a fundamental dimension of this analysis, reveals intriguing patterns that underscore the nuanced nature of arrests in Toronto. This paper aims to contribute to understanding the dynamics of arrests in Toronto, providing insights that can inform targeted interventions, policy considerations, and community engagement efforts.

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-

Table 1: Total number of arrests vs Booked for the year 2020, and 2021

Ethinic Identity	2020	2021	Ethinic Identity	2020	2021
White	7236	7188	White	3947	3673
Black	4657	4524	Black	2699	2524
Unknown or Legacy	1235	1144	Unknown or Legacy	641	570
East/Southeast Asian	733	900	East/Southeast Asian	393	432
South Asian	722	723	South Asian	407	373
Middle-Eastern	593	673	Middle-Eastern	303	385
Indigenous	534	497	Indigenous	310	258
Latino	315	324	Latino	176	183

ing 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 comprised 32,000 arrest records spanning both 2020 and 2021, encompassing 26 variables of analyzable data. However, for the analysis in this paper, we narrowed our focus to 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. Additionally, we consolidated similar columns, namely Search Reason Cause Injury, Search Reason Assist Escape, Search Reason Possess Weapons, and Search Reason Possess Evidence, into a unified column labeled "Search Reason." utilizing the `any()` function (Wickham et al. 2022).

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

Since our analysis primarily revolves around the distribution of arrests for each ethnic identity, we extracted all ethnic identities from our data. These ethnic identities include Black, White, South Asian, East Asian, Middle Eastern, Latin American, Southeast Asian, and Indigenous. The data for the years 2020 and 2021 are summarized in Table 1 (a) (Table 1) and Table 1 (b), respectively.

Table 1 (a) (Table 1) provides a summary of our data, revealing the number of individuals arrested from each prominent ethnic group in the years 2020 and 2021. Table 2 (b) shows, out of all the people arrested from each ethnic identity, how many were actually booked under a criminal case.

Table 2: 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	2077	1864
Black	Aged 25 to 34 years	1856	1797
Unknown or Legacy	Aged 25 to 34 years	431	380
East/Southeast Asian	Aged 25 to 34 years	201	264
South Asian	Aged 25 to 34 years	243	254
Middle-Eastern	Aged 25 to 34 years	182	241
Indigenous	Aged 25 to 34 years	189	154
Latino	Aged 25 to 34 years	93	121
None	Aged 25 to 34 years	1	0

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

It would be interesting to explore the age group distribution of arrests. Table 2 (Table 2) exactly presents that. A data summary indicating the count of different age groups of individuals from prominent ethnic groups in the years 2020 and 2021. All the age groups were the following: Ager 17 years and under, Aged 18 to 24 years, Aged 25 to 34 years, Aged 35 to 44 years, Aged 45 to 54 years, Aged 55 to 64 years, Aged 65 years and over.

Due to the extensive nature of the dataset, encompassing nine different age groups across eight ethnicities, summarizing the information for clarity posed a challenge. Consequently, we have opted to showcase a subset of the data specifically focusing on the age group of 25 to 34 years for improved readability.

3 Results

We would like to acknowledge that opendatatoronto (Gelfand 2022) provides the distribution of the population by ethnicity. However, due to the vast size of Toronto and its numerous wards, it has been challenging to compare all the data and cross-verify it with our crime rate data comprehensively. Nevertheless, a more detailed analysis can be conducted by comparing specific ethnic groups with each other.

Now that we have cleaned the data, 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 1), Figure 2 (Figure 2), Figure 3 (Figure 3), and Figure 4 (Figure 4) was created using `ggplot` (Wickham 2016), displays the information that can be used as compare the trends as described above.

3.1 Number of Arrests in year 2020-2021

With reference to Table 1 (Table 1) and the graph (see Figure 1) plotted the data shows a non-uniform distribution of arrests across ethnic groups, with the Black ethnic group and White group exhibiting the highest numbers of arrests in both 2020 and 2021. Furthermore, the data indicates a decrease in the number of arrests for the top two ethnic groups from 2020 to 2021. However, a contrasting trend emerges for other ethnic groups, where the number of arrests has increased over the same time period.

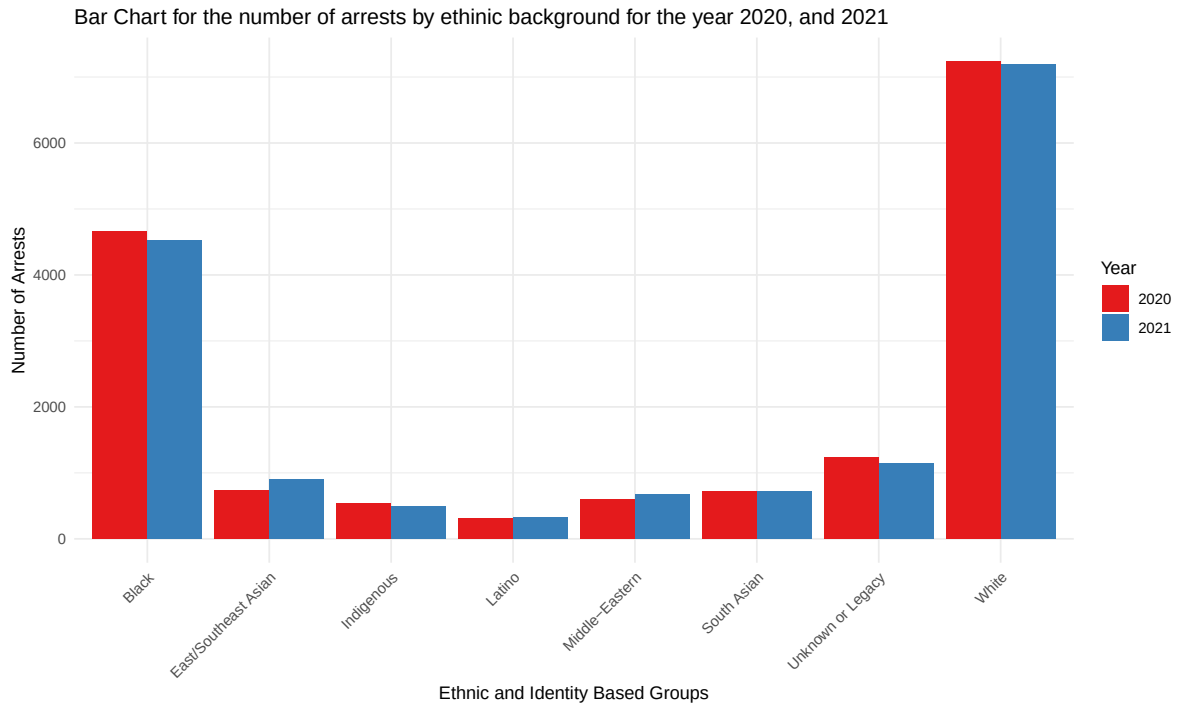

Figure 1: Number of arrests by ethinic background for the year 2020, and 2021

Referring to Table 1 (Table 1) for the proportion of arrests that were booked, we plotted a stacked bar graph (see Figure 2), with y-axis representing the total percentage. The data illustrates a consistent distribution of the number of individuals booked after arrests across various ethnic groups. Notably, there is a higher count of individuals booked after arrest compared to those who were arrested but not booked for a criminal case. Additionally, it is observed that, across all ethnic groups, the percentage of individuals booked falls within the range of 50% to 55%, and this pattern remains consistent in the year 2021.

However, it is noteworthy that the Black ethnicity exhibits the highest number of individuals booked after arrest in both the years 2020 and 2021.

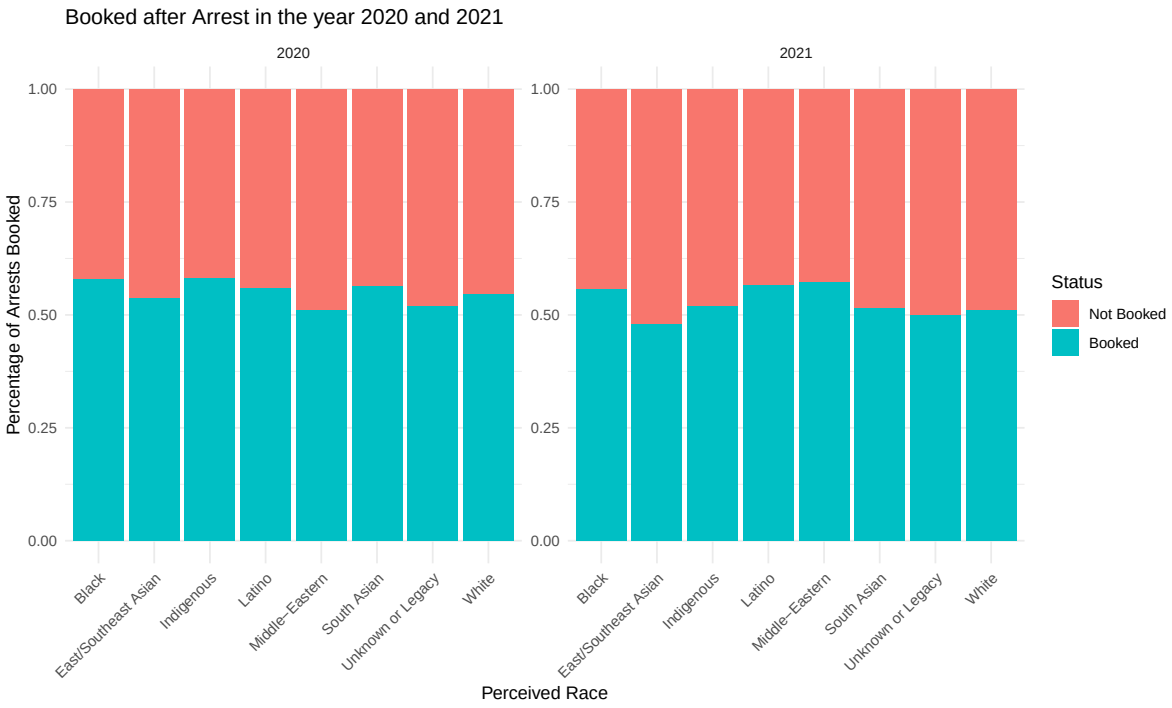

Figure 2: Number of arrests booked by ethinic background for the year 2020, and 2021

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

As depicted in Figure 3 (see Figure 3), the data reveals a non-uniform distribution of arrests across age groups, with the age groups of 25 to 34 years and 35 to 44 years exhibiting the highest number of arrests in both 2020 and 2021. Furthermore, there is a notable decrease in the number of arrests for these top two age groups from 2020 to 2021. In contrast, a different trend emerges for other age groups, where the number of arrests has increased over the same time period.

Examining Figure 4 (see Figure 4), which illustrates the dot plot for the number of arrests by age group in the year 2020 and 2021 for the top two ethnic groups, “Black” and “White,” it becomes evident that the age group 25 to 34 years represents the highest number of arrests for individuals of Black ethnicity. The trend for Black ethnicity indicates a decreasing curve from this age group onward. In contrast, for individuals of White ethnicity, the peak in arrests is observed in the age group 35 to 44 years. It is noteworthy that from age 45 to 65, the difference in the number of arrests between both ethnicities starts to converge.

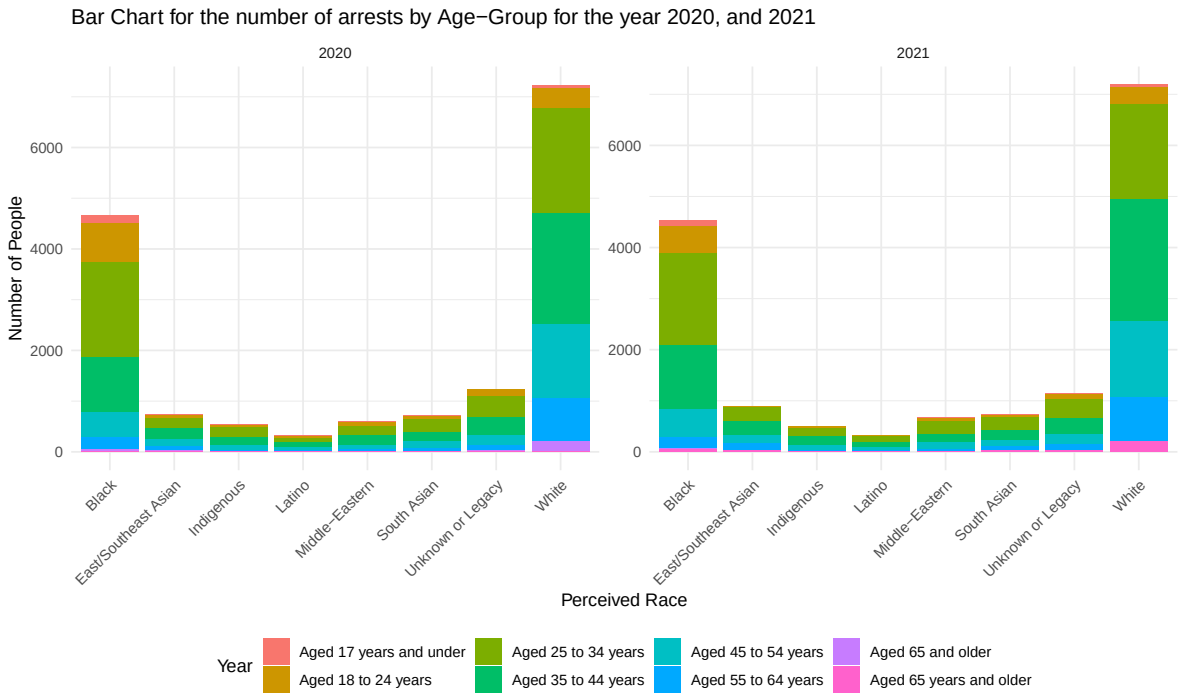

Figure 3: Number of arrests by Age-Group for the year 2020, and 2021

4 Discussion

References

Fanfair, R. 2022. "Race-Based Data Shows over-Policing." https://www.tps.ca/media-centre/stories/race-based-data-shows-over-policing/.

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

Jackson, Mariel. 2022. "Racial Identity in Policing Is More Complicated Than We Think." https://mccourt.georgetown.edu/news/racial-identity-in-policing/.

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.

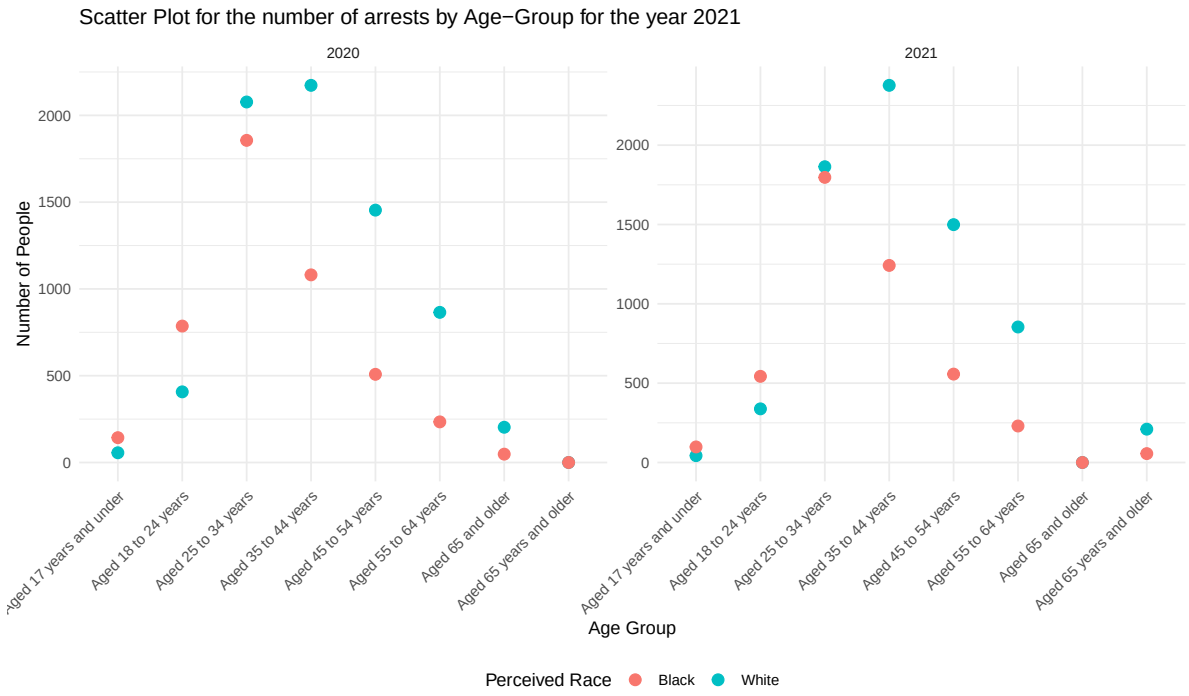

Figure 4: Scatter Plot for the number of arrests by Age-Group for the year 2020, and 2021