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

Aviral Bhardwaj

January 22, 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.

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

Table 1: 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
South Asian	722	723
Middle-Eastern	593	673
Indigenous	534	497
Latino	315	324
None	1	1

(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. Upon further examination the raw data contained values such as 'NA' under the Booked column and 'None' under the Search column. Using the `mutate` function (Wickham et al. 2022), we replaced the 'NA' and `None` values with 0 and 1 to summarize the data. 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." The process involved utilizing the `any()` function (Wickham et al. 2022), assigning a value of 1 to indicate the presence of any of these search motives.

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

Table 1 (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.

2.2 Booked after Arrest (Ethnic and Identity-Based Groups for Year 2020 and 2021)

Table 2 presents a data summary indicating the count of individuals from prominent ethnic groups who were "booked" in criminal cases post-arrest in the years 2020 and 2021. 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

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

Ethinic 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

Table 3: Arrests Booked by ethnic 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

opted to showcase a subset of the data specifically focusing on the age group of 25 to 34 years for improved readability.

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

Table 3 presents a data summary indicating the count of different age groups of individuals from prominent ethnic groups in the years 2020 and 2021. 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.

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.

3 Results

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 2, Figure 3, 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 and the graph 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.
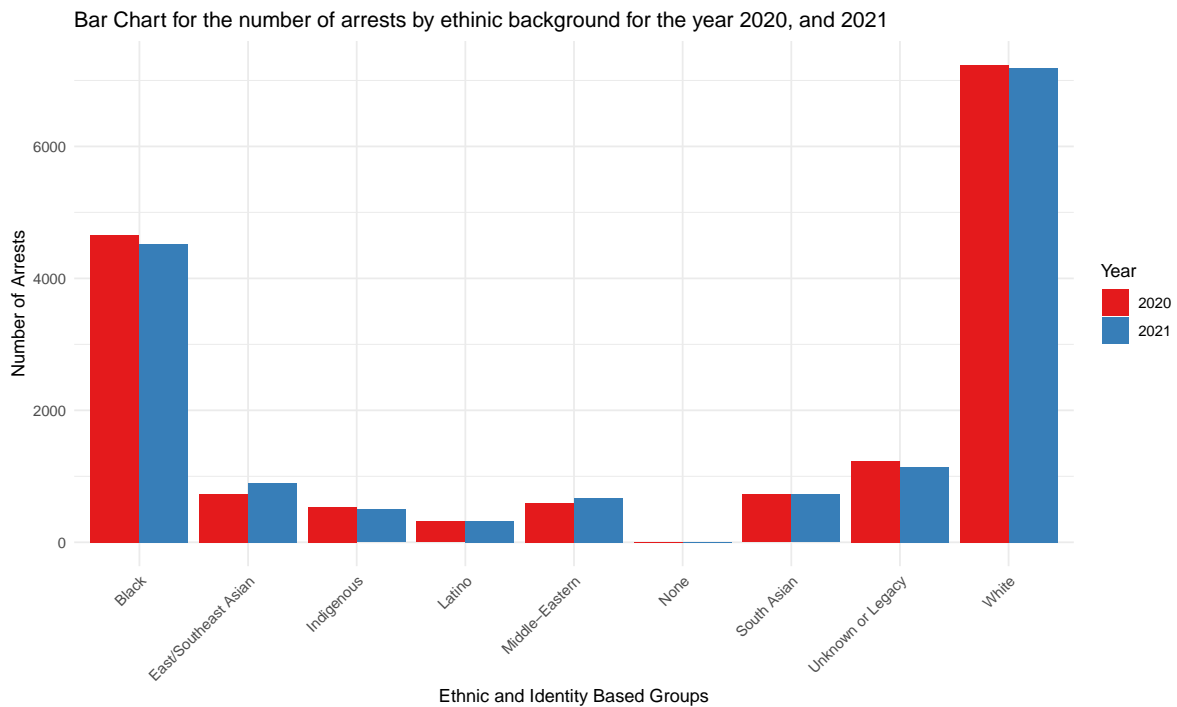

Figure 1: Number of arrests by ethinic background for the year 2020, and 2021

3.2 Number of Arrests Booked in year 2020-2021

Referring to Table 2 and the plotted graph, 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.
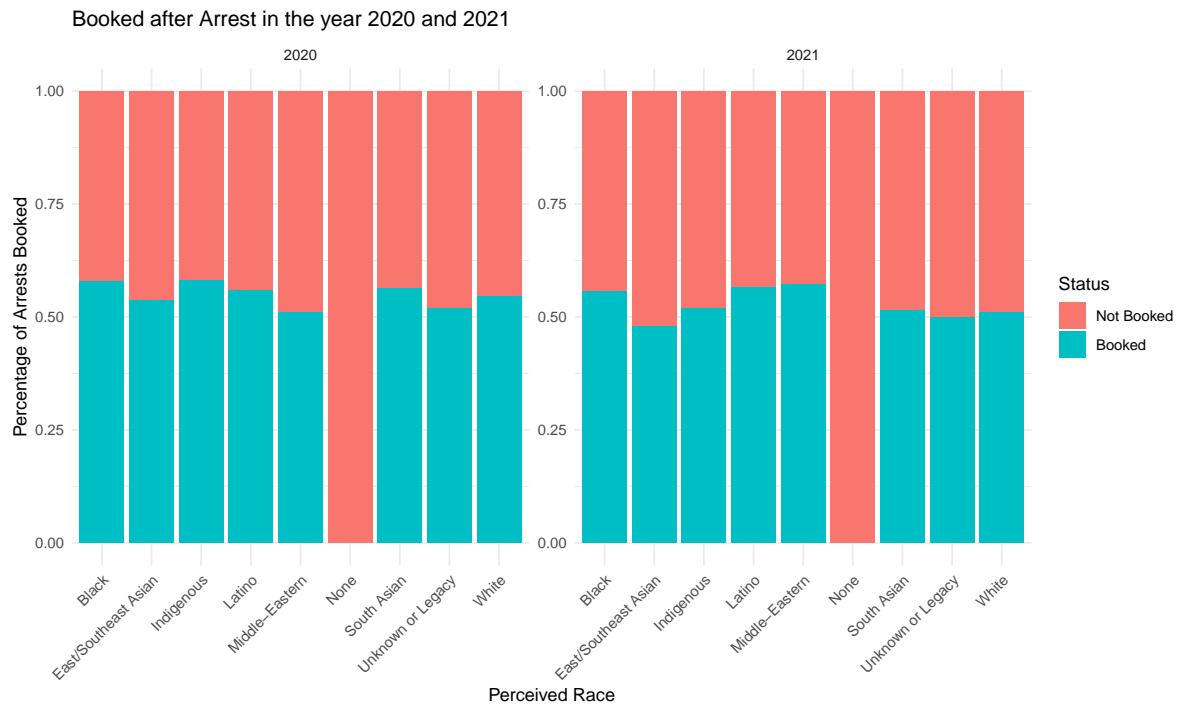

Figure 2: Number of arrests booked by ethnic background for the year 2020, and 2021

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

As seen in Figure 3, the data indicates a non-uniform distribution of arrests across age groups, with the age group of 25 to 34 years and age group 35 to 44 exhibiting the highest number of arrests in both 2020 and 2021. Furthermore, the data indicates a decrease in the number of arrests for the top two age groups from 2020 to 2021. However, a contrasting trend emerges for other age groups, where the number of arrests has increased over the same time period.
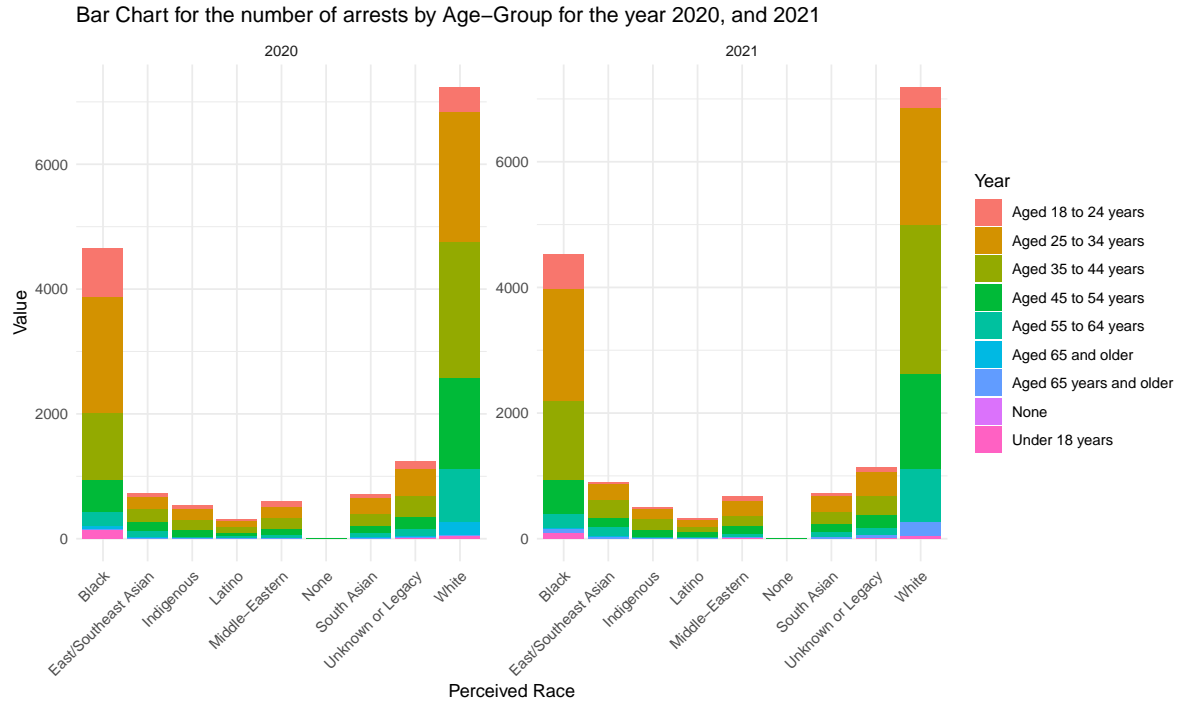
Figure 3: Number of arrests by Age-Group for the year 2020, and 2021

4 Discussion

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.
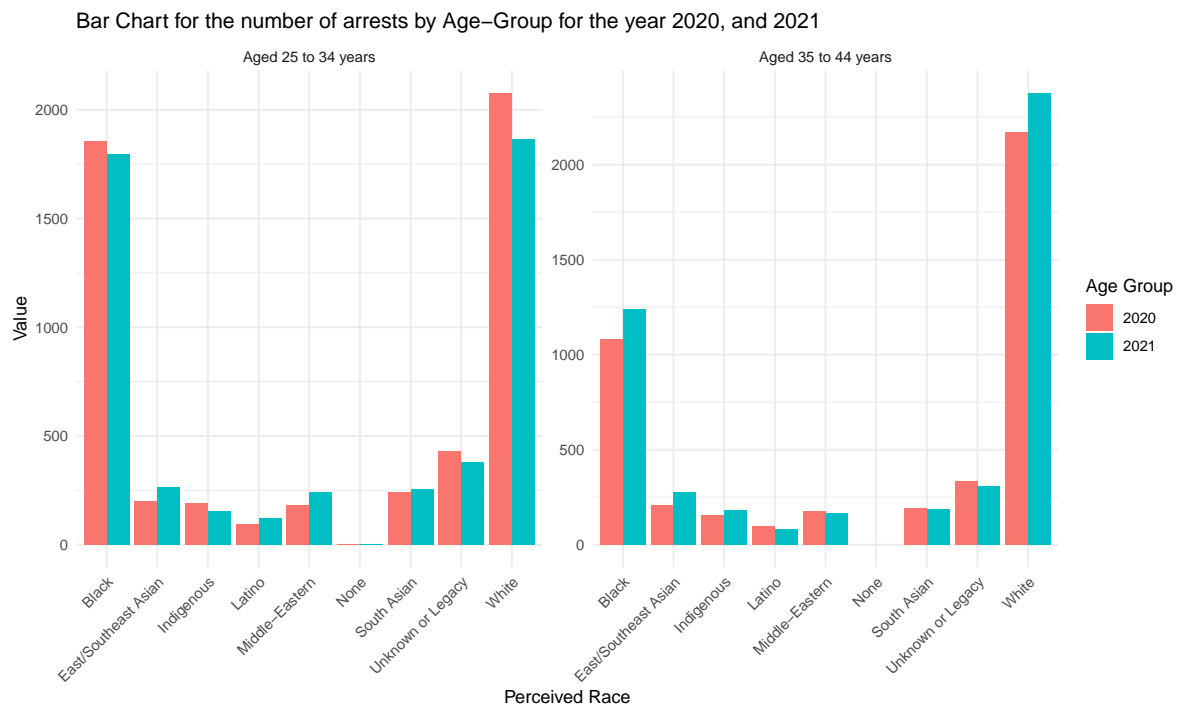
Figure 4: Number of arrests by Age-Group for the year 2020, and 2021