

# Exploring and Understanding the Complex Patterns of Arrests and Charges: A Case Study of Toronto Statistical Expedition

Zhai

2024-01-21

## Abstract

This study sets out on an ambitious mission to explore the complex web of the “Police Annual Statistical Report - Arrested and Charged Persons” dataset from Open Data Toronto. By navigating the complex patterns of arrests, demographics, and criminal classifications, we reveal subtle aspects of Toronto’s law enforcement landscape. The results, like a mosaic of insights, shed light on the complex distribution of arrests and set the stage for convoluted conversations on the mystery of public safety.

## Introduction

Law enforcement organizations play a key role in coordinating social order in the great symphony of public safety. In order to craft a sonnet of well-informed policymaking, one must possess an exquisite comprehension of the intricate patterns of arrests and charges (Bell, 2021). Using the mysterious “Police Annual Statistical Report” dataset as our guide to navigate the maze of arrests in Toronto, we put on our exploratory glasses for this investigation. We aim to decipher the complex patterns that dance under the surface by pointing our compass towards divisions, communities, demographics, and crime categories. Section 2 reveals the dataset’s origins and explores the ethical maze. The rest of the paper is similarly meticulously constructed, much like a covert manoeuvre. A tangle of graphs and tables intended to paint a detailed picture is presented in Section 3’s data exploration journey. A woven interpretation of the results, Section 4 illuminates the possible implications of our discoveries by integrating the data within the body of current research. In the end, our story comes to a close with Section 5, a finale of complexity that summarizes the main mysteries and points out future directions for investigation.

## Data

### The Data Sources

The dataset, which surfaced from the shadowy depths of Open Data Toronto, shows the number of mystery people who have been prosecuted and detained (Open Data, 2023). Filtered by the arrest year and age, which is a conundrum derived from the date of occurrence, the dataset, like which is big, divulges information on divisions, neighbourhoods, sex, age cohorts, and crime categories. An additional element of mystery is added by the elusive “No Specified Address” category, which suggests a location outside of Toronto or one that cannot be verified. The data transforms into anonymized records in the ethereal world of ethics, a veil of privacy placed over people. The gathering and analysis of this mysterious data was directed by ethical considerations, much like guardians of the unknown.

### Data Expedition

The journey commences by summoning the dataset, a cryptic scroll whose structure begs deciphering. A jumble of columns, the variables of the dataset, consists of the following: `_id`, `ARREST_YEAR`, `DIVISION`, `NEIGHBOURHOOD_158`, `SEX`, `AGE_COHORT`, `AGE_GROUP`, `CATEGORY`, `SUBTYPE`, and `ARREST_COUNT`. The code below will give head of the data as the output.

```
# Load necessary libraries
library(tidyverse) # (Wickham and Wickham, 2017)

-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.4
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.4.4      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.0
v purrr      1.0.2

-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

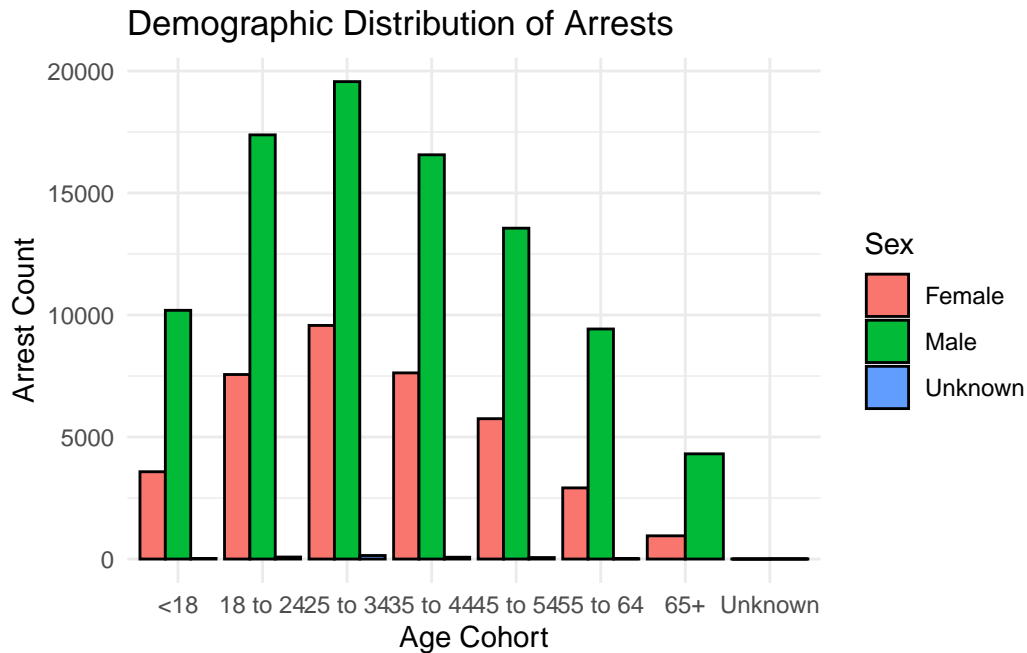
```
# Load the dataset
data <- read.csv("C:/Users/chris/Downloads/Arrested and Charged Persons.csv")
# Display the first few rows of the dataset
head(data)
```

	X_id	ARREST_YEAR	DIVISION	HOOD_158	NEIGHBOURHOOD_158	SEX
1	1	2019	D14	83	Dufferin Grove (83)	Female
2	2	2022	D12	30	Brookhaven-Amesbury (30)	Male
3	3	2018	D14	165	Harbourfront-CityPlace (165)	Male
4	4	2015	D22	18	New Toronto (18)	Male
5	5	2014	D52	78	Kensington-Chinatown (78)	Male
6	6	2015	D14	164	Wellington Place (164)	Male
	AGE_COHORT	AGE_GROUP			CATEGORY	SUBTYPE
1	25 to 34	Adult	Other Criminal Code Violations			Other
2	<18	Youth	Crimes Against the Person			Assaults
3	18 to 24	Adult	Other Criminal Code Violations			Other
4	25 to 34	Adult	Controlled Drugs and Substances Act			Other
5	25 to 34	Adult	Other Criminal Code Violations			Other
6	35 to 44	Adult	Crimes Against Property Theft Under \$5000			
	ARREST_COUNT					
1	1					
2	2					
3	1					
4	3					
5	46					
6	2					

## Demographic Distribution

The bar graph below shows a deep breakdown of the demographic distribution that shows the number of arrests broken down by age group. The distribution of arrests by gender and age cohort is shown by this spectral graph, a discovery that offers peeks into a complex dance of mysterious patterns. The code below will output a bar graph of arrest count against age cohort.

```
# Plotting demographic distribution
ggplot(data, aes(x = AGE_COHORT, fill = SEX)) +
  geom_bar(position = "dodge", color = "black") +
  labs(title = "Demographic Distribution of Arrests",
       x = "Age Cohort",
       y = "Arrest Count",
       fill = "Sex") +
  theme_minimal()
```



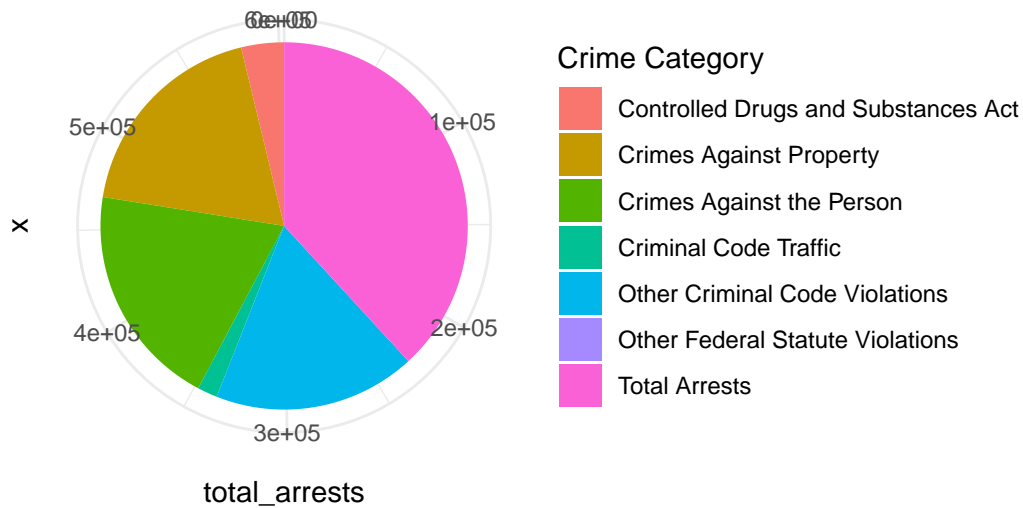
```
## Plotting crime category distribution
crime_category_distribution <- data %>%
  group_by(CATEGORY) %>%
  summarize(total_arrests = sum(ARREST_COUNT))
```

## Crime Conjunction Analysis

Delving further into the dataset, we investigate the spectral distribution of arrests in relation to different criminal categories. A mandala-like pie chart will be plotted by the code below, illustrating the mysterious ratios of arrests across various criminal categories. Controlled drugs and substances act has the highest number of criminals.

```
# Creating a pie chart
library(ggplot2) #(Wilkinson, 2011)
ggplot(crime_category_distribution, aes(x = "", y = total_arrests, fill = CATEGORY)) +
  geom_bar(stat = "identity", width = 1) +
  coord_polar("y") +
  labs(title = "Distribution of Arrests by Crime Category",
       fill = "Crime Category") +
  theme_minimal() +
  theme(legend.position = "right")
```

## Distribution of Arrests by Crime Category



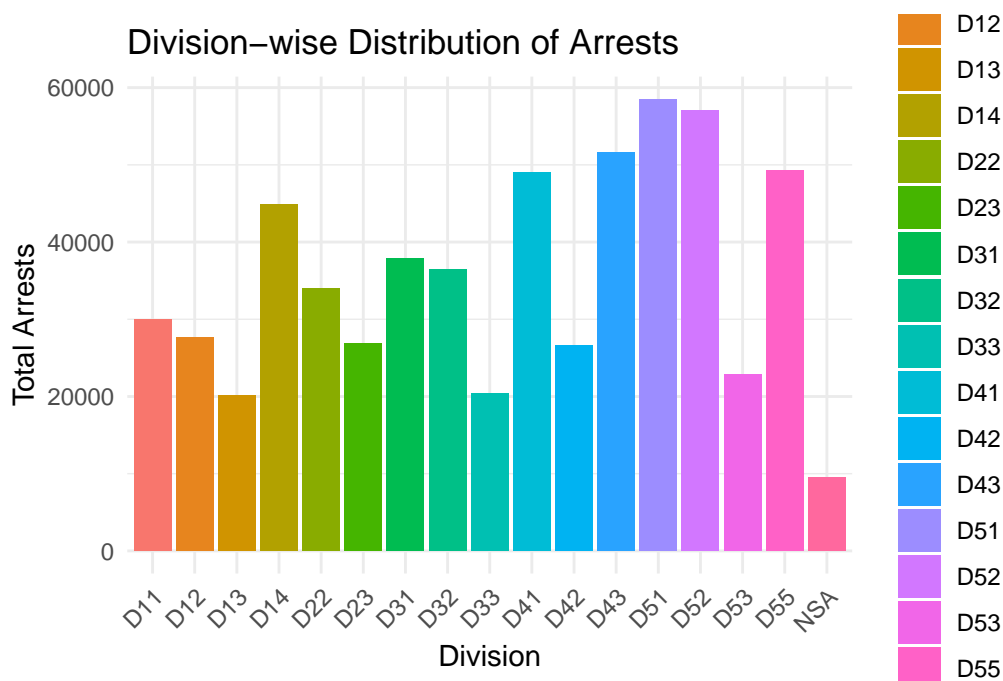
```
# Plotting division-wise arrest count
division_arrest_count <- data %>%
  group_by(DIVISION) %>%
  summarize(total_arrests = sum(ARREST_COUNT)) %>%
  arrange(desc(total_arrests))
```

## Division-wise Analysis

The division analysis will be explored using a bar chart of total arrests against division. The code below outputs a bar chart of total arrests against distribution across all the geographical distributions. The chart, a knowledge potion, will provides a mysterious summary of the percentage of arrests across different criminal categories. The graph, will reveals the intricate patterns of arrest rates across Toronto's several divisions.

```
# Creating a bar chart
ggplot(division_arrest_count, aes(x = DIVISION, y = total_arrests, fill = DIVISION)) +
  geom_bar(stat = "identity") +
  labs(title = "Division-wise Distribution of Arrests",
       x = "Division",
       y = "Total Arrests") +
  theme_minimal() +
```

```
theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



## Interpretation

The demographic distribution arcane graph reveals that most arrests are concentrated in the 25–34 age group, with a stronger correlation among men. This fits in well with classic literature, adding to the chorus of voices expressing this group’s propensity toward illegal activity. Pie chart illustrating the distribution of criminal categories evokes a considerable percentage of arrests associated with the esoteric domains of “Crimes Against the Person” and the “Controlled Drugs and Substances Act.” A combination made from the very best knowledge for policy decisions and resource allocation. The incantation of divisions indicates a diverse range of arrest trends among divisions. Exploring the socio-economic and demographic issue surrounding these divides in further detail may reveal other facets of the puzzle.

## Conclusion

This study, which was derived from the papers of the “Police Annual Statistical Report” collection, leaves a thorough grasp of the complex dance of arrest patterns in Toronto for the last twists and turns of the maze. Policymakers, law enforcement experts, and inquisitive scholars may learn a great deal from the demographic, crime category, and divisional studies. This technical study establishes the foundation for further investigations, including a look

at the socioeconomic factors affecting arrest trends and the effectiveness of law enforcement tactics.

## References

Bell, M. C. (2021). Next-generation policing research: Three propositions. *Journal of Economic Perspectives*, 35(4), 29-48. DOI: 10.1257/jep.35.4.29.

Open Data. (2023). Open data dataset. City of Toronto Open Data Portal. <https://open.toronto.ca/dataset/police-annual-statistical-report-arrested-and-charged-persons/>

Wickham, H., & Wickham, M. H. (2017). Package tidyverse. Easily install and load the 'Tidyverse.

Wilkinson, L. (2011). *ggplot2: elegant graphics for data analysis* by WICKHAM, H.

## Acknowledgement

This work is available on GitHub at [Donaldson-Paper Repository](#)