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

Zhai

2024-01-21

Abstract

This study contains well organized sections to explore the complex topic of the "Police Annual Statistical Report - Arrested and Charged Persons" dataset from Open Data Toronto. The visualization of the complex patterns of arrests, demographics, and criminal classifications, reveals undiscovered aspects of Toronto's law enforcement landscape. The data is sourced from Open Data Toronto database on the latest datasets section. The data is properly explored by visualizing the demographic distribution, crime conjuration analysis, and division-wise analysis. The results of the visualization gives undiscovered insights, on the complex distribution of arrests and set the stage for better discussions on the mystery of public safety in Toronto.

Introduction

Law enforcement organizations play a key role in coordinating social order so as to ensure there is enough public safety. In order to come up with a well-informed policy making, one must visualize and explore very well the specific patterns of arrests and charges (Bell, 2021). The patterns will then be used in coming up with meaningful conclusions for the our topic. Using the "Police Annual Statistical Report" dataset as our topic of study to explore the distribution of arrests in Toronto, we put on our exploratory metrics for this investigation. The main aim of this study is to uncover the complex patterns that are not easily determined by putting our emphasis towards divisions, communities, demographics, and crime categories. The study is well organized into sections which explore specific aspect of the dataset with the aim of uncovering new undiscovered properties of the dataset. Section 1 explores the data set, its composition and the types of variables. Section 2 reveals the data set's origins and explores the ethical composition of the data set. The rest of the paper that is the last section is similarly is very well constructed and organized because it is the major part of the study which explores hidden components of the data. A combination of graphs and tables are used in this section with the main aim of revealing a detailed picture of the data though visualizations, analyses and interpretations. A well arranged interpretation of the results, Section 4, sheds light on the possible implications of our discoveries by integrating the data within the body of current research. In the conclusion, the paper comes to a close with Section 5, a final section of complexity that summarizes the main discoveries and points out future directions for further investigation in this field.

Data

The Data Sources

The dataset, which extracted from the large database of the Open Data Toronto, shows the number of different set of people who have been prosecuted and detained (Open Data, 2023). Filtered by the arrest year and age, which is a variable derived from the date of occurrence, the dataset, which is very big, sections the information on divisions, neighborhoods, sex, age cohorts, and crime categories. An additional variable of dataset is added as a “No Specified Address” category as the last column of the dataset, which suggests a location outside of Toronto or one that cannot be verified. The data transformed into unknown records in the real world of ethics, in order to maintain full privacy for all the secretive data in the dataset. The gathering and analysis of this large data was directed by ethical considerations, to make sure that no confidential data was visible to the general public.

Data Expedition

The data exploration process starts by exploring hat dataset deeply including its variables. The dataset is made up of 11 variables. This makes it 11 columns , the variables of the dataset, which consists of the following: _id, ARREST_YEAR, DIVISION, NEIGHBOURHOOD_158, SEX, AGE_COHORT, AGE_GROUP, CATEGORY, SUBTYPE, and ARREST_COUNT. The _id is the unique id for every arrest that takes place within the set time frame.The arrest year is the year in which the arrest occurs. The rest of variables are self explanatory, that is their names explains what they represent.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 becom
```

```r
# 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)
```

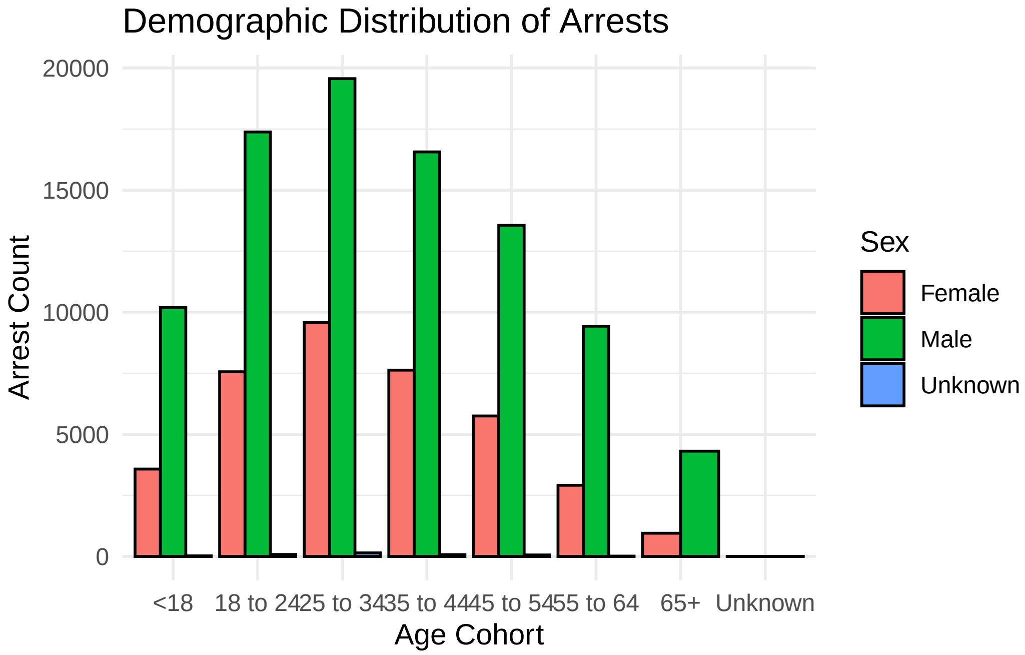
```
  X_id ARREST_YEAR DIVISION HOOD_158            NEIGHBOURHOOD_158    SEX
1    1        2019      D14       83          Dufferin Grove (83) Female
2    2        2022      D12       30     Brookhaven-Amesbury (30)   Male
3    3        2018      D14      165 Harbourfront-CityPlace (165)   Male
4    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

Demographic distribution is the distribution of the population in question which includes the specific characteristics of the population, population activities, the region, administrative units, the general grid and other units according to the data variables. The demographic distribution of the variables in this data set will be visualized using a bar graph to give deep insights of the data and its variables. 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 clear visualization that offers insights into the patterns of the distribution of arrests by each gender. 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))
```

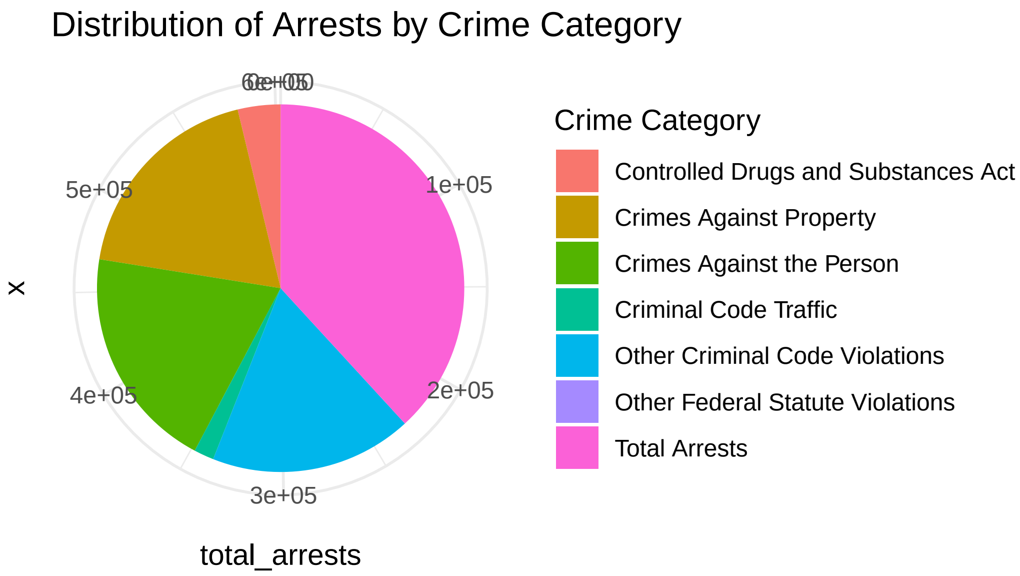
The above graph when analyses keenly gives very helpful insights on the distribution of arrests on the gender variable. The arrest age ranges from 18 to 65+ and male and female genders are compared on the arrest count. The above graph clearly shows that arrests are highly distribution on the male gender. This exploration opens door for further study on the male gender arrest counts.

Crime Conjuration Analysis

Delving further into the dataset, we investigate the distribution of arrests in relation to different criminal categories. The categories which will be explored and compared include controlled drugs and substances act, crimes against property, crimes against the person, criminal code violations, and other federal statute violations A 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")
```

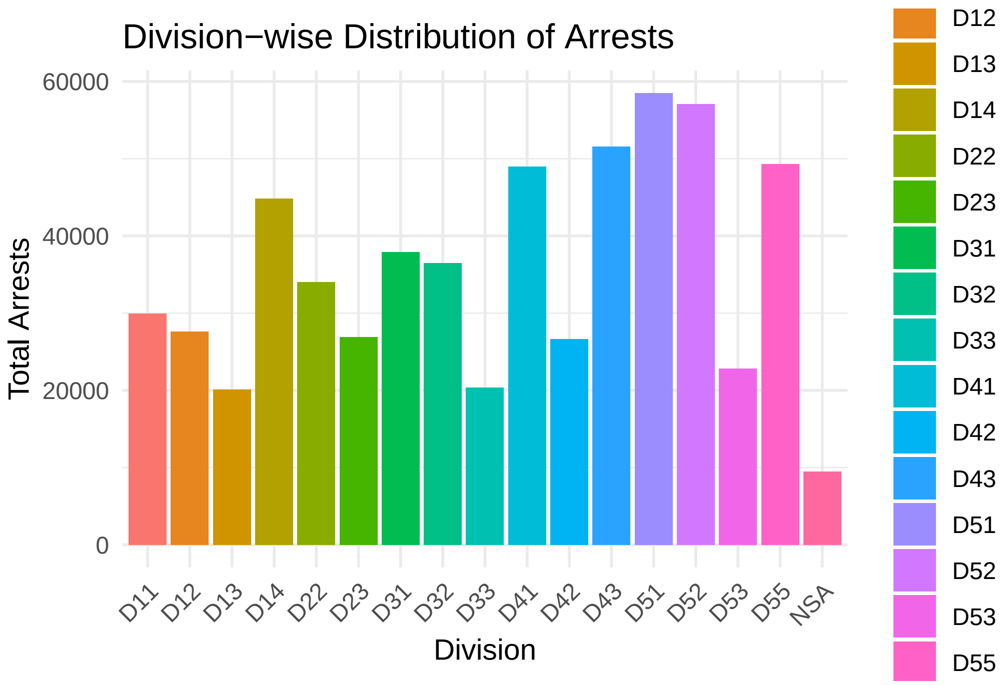


```r
# Plotting division-wise arrest count
division_arrest_count <- data %>%
  group_by(DIVISION) %>%
  summarize(total_arrests = sum(ARREST_COUNT)) %>%
  arrange(desc(total_arrests))
```

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, provides a mysterious summary of the percentage of arrests across different criminal categories. The graph, reveals the intricate patterns of arrest rates across Toronto's several divisions.

```r
# 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))
```



Time Trend Analysis for the Number of Arrests

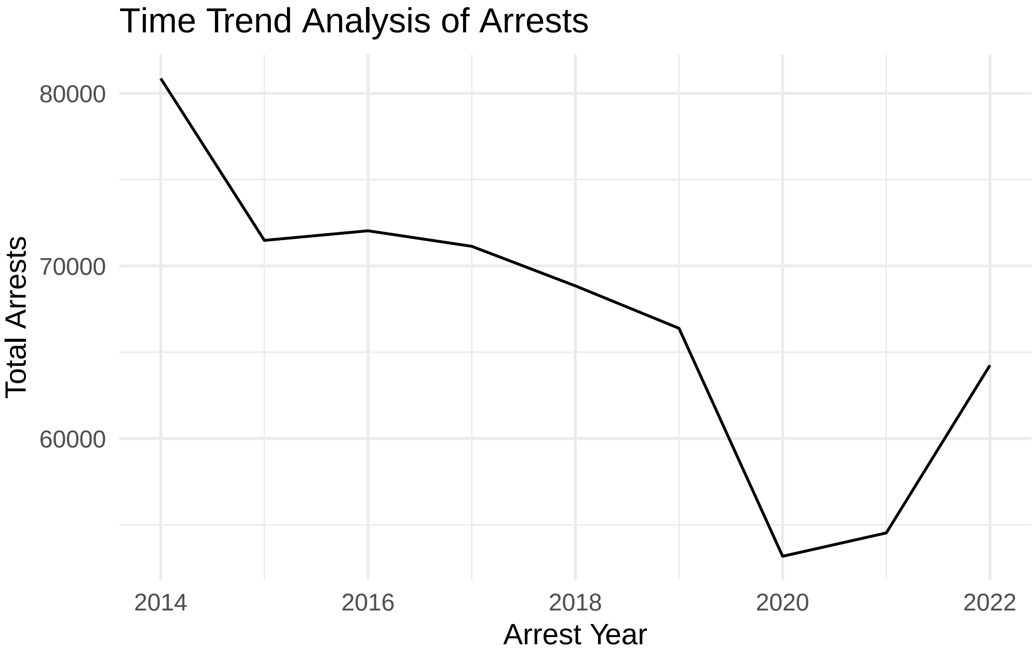
Next is exploration of arrests over the year in order to identify any specific patterns or trends in the number of arrests over the years. The time trend analysis is done using a line graph of total arrests against the year of arrest. The arrest year runs from 2014 to 2022. The graph shows a decreasing trend generally on the number of total arrests over the years.

```
# Load necessary libraries
library(tidyverse)
library(dplyr)

# Load the dataset
data <- read.csv("C:/Users/chris/Downloads/Arrested and Charged Persons.csv")

# Time Trend Analysis
time_trend_plot <- data %>%
  group_by(ARREST_YEAR) %>%
  summarize(total_arrests = sum(ARREST_COUNT))
```

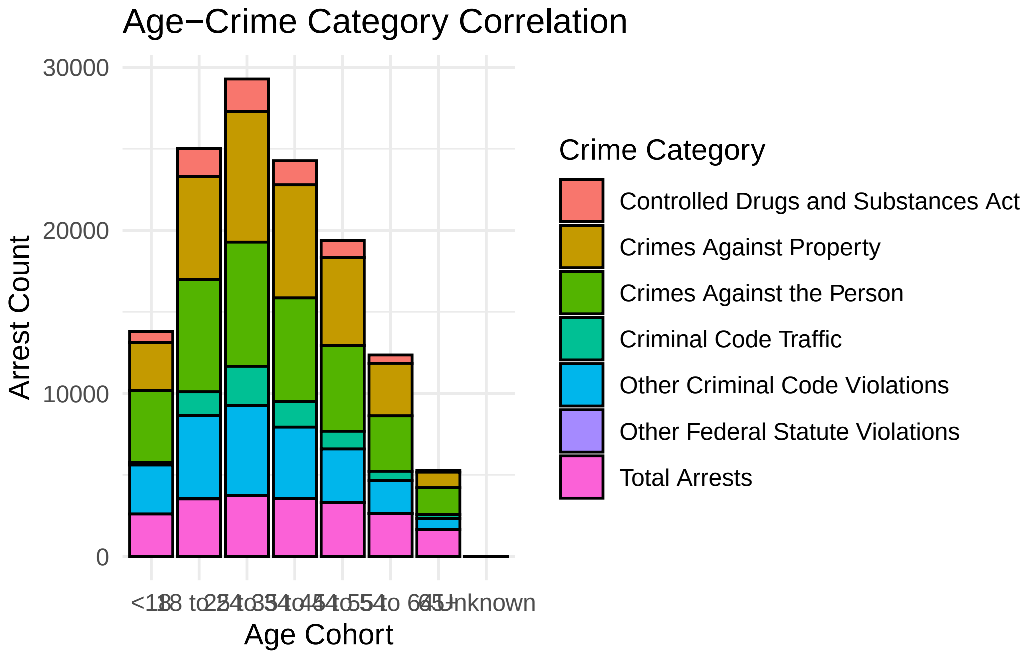
```
# Creating a line plot
ggplot(time_trend_plot, aes(x = ARREST_YEAR, y = total_arrests)) +
  geom_line() +
  labs(title = "Time Trend Analysis of Arrests",
       x = "Arrest Year",
       y = "Total Arrests") +
  theme_minimal()
```



The Correlation among Age-Crime Category

The correlation between age cohorts and crime categories can be visualized using stacked bar chart as shown below. This is a plot of arrest count against age cohort. The graph will show how different age cohorts contribute to arrest count and categories. The highest contributor to the arrest count is between age 25 to 25

```
# Creating a stacked bar plot for age-crime category correlation
ggplot(data, aes(x = AGE_COHORT, fill = CATEGORY)) +
  geom_bar(position = "stack", color = "black") +
  labs(title = "Age-Crime Category Correlation",
       x = "Age Cohort",
       y = "Arrest Count",
       fill = "Crime Category") +
  theme_minimal()
```



Interpretation

The demographic distribution bar 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 existing studies which have stressed in expressing this group's propensity toward illegal activity. Pie chart illustrating the distribution of criminal categories also shows the highest percentage of arrests associated are associated with the division category of "Crimes Against the Person" and the "Controlled Drugs and Substances Act." This knowledge can be helpful to the case managers or prison managers by giving them the best knowledge for policy decisions and resource allocation. The exploration of divisions indicates a diverse range of arrest trends among divisions. Exploring the socio-economic and demographic issue surrounding these divisions in further detail may reveal other factors which are captured and are most likely to affect the range of arrest trends.

Conclusion

This study, which was derived from the papers of the "Police Annual Statistical Report" collection, leaves a thorough exploration of the complex form of the arrest patterns in Toronto.

The dataset, which extracted from the large database of the Open Data Toronto, shows the number of different set of people who have been prosecuted and detained. The distribution of arrests by gender and age cohort has been shown by the bar graph, a clear visualization that gives insights into the patterns of the distribution of arrests for each gender. A bar chart has been used to visualize the distribution of arrests in various divisions. Controlled drugs and substances act has the highest number of criminals. 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/pol
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