

Understanding Urban Crimes: An Analysis of Police Tickets in Toronto*

Yiyan Zhang

January 22, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

We will be delving into a fascinating dataset in this essay—the Police Annual Statistical Report on Tickets Issued in Toronto. The objective is to go through the data and determine what it says about how individuals are breaching the law in various regions of the city and whether there is a pattern in who is receiving tickets. It’s like taking a closer look at the city to see who is and isn’t abiding by the rules. It’s very relevant since it can provide us a better understanding of the state of law enforcement and public behavior in the city.

I’ll start by investigating the different kinds of offenses that are taking place. Are people breaching some rules more often than others? And does it differ depending on where in the city you are? This is essential to comprehending not only what’s wrong but also where our efforts should be directed in order to improve things. Seeing any surprising trends in the data, such as if a particular rule is broken significantly more frequently in one neighbourhood than the rest of Toronto, is also fairly interesting.

Finally, I will investigate who is really receiving these tickets. Are the majority of them young? Or is it dispersed among several age groups? This portion of the study may provide valuable insights into which individuals are more likely to violate the law, or it may just indicate that they are being detected more frequently. Furthermore, knowing how offenses are distributed throughout different age groups can give politicians and city planners important information. With this information, they can devise more effective strategies to maintain city safety and ensure that everyone abides by the regulations.

The remainder of this paper is structured as follows. Section 2,

*Code and data are available at: <https://github.com/manfred1201/Term-Paper1>

2 Data

The dataset breaks down the tickets handed out by the Toronto Police from Open Data Toronto (<https://open.toronto.ca/dataset/police-annual-statistical-report-tickets-issued/>). It's sorted by different stuff like the year the ticket was given, what kind of ticket it was, what the offence was, the age group of the person who got the ticket, which police division was involved, and the neighbourhood where it happened. This data covers several years, which lets us see how the patterns in ticket giving have changed over time. The data is loaded directly from open data toronto. (Gelfand 2022)

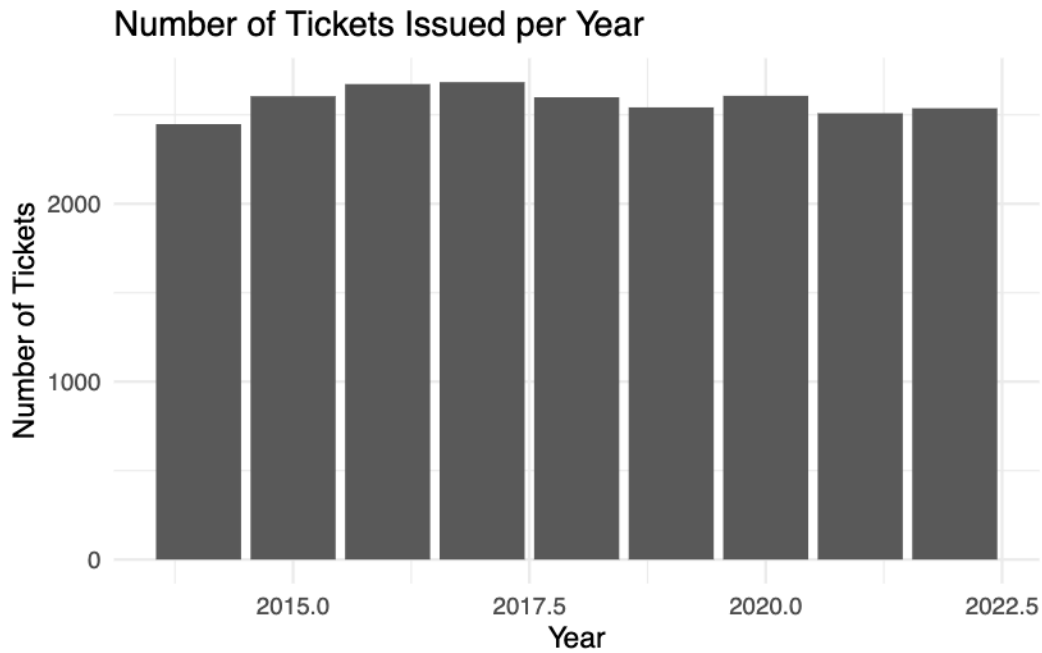


Figure 1: Bills of penguins

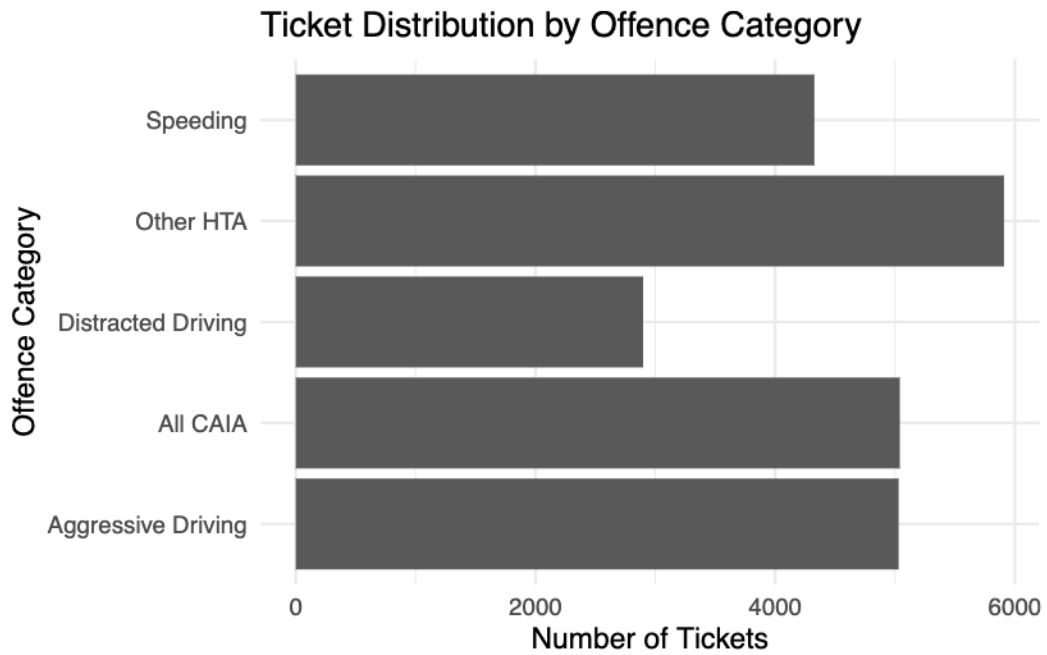


Figure 2: Bills of penguins

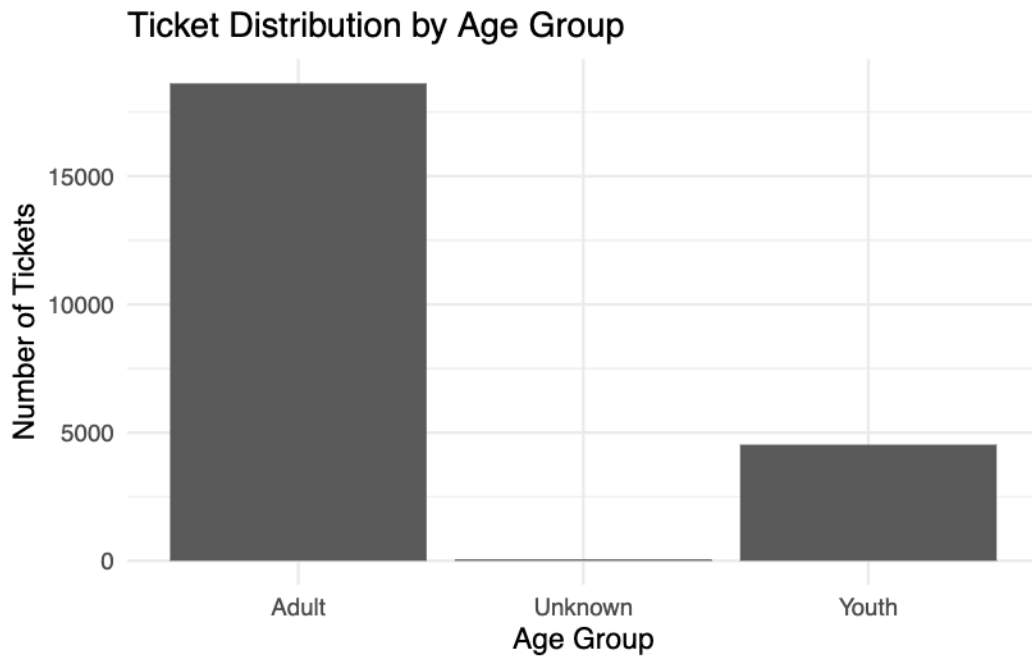


Figure 3: Bills of penguins

Based on the visualization results (**vis?**), the overall tickets throughout the years are steady. HTA, CAIA and aggressive driving are the main categories of tickets, indicating that people are more often having the highway traffic violations and insurance issues. With regarding to age, adult group has more violations than youths, which is normal because there are more adults.

3 Model

Our modeling strategy has two objectives. First, the future number of tickets can be predicted using a linear regression model. Making better patrol arrangements and projecting future police officer needs can both benefit greatly from this. Second, the association between the age group and the type of violation can be examined using a chi-square test.

The chi-square test and linear regression model that were utilized to look at are briefly described here. You can get diagnostics and background information in [Appendix -@sec-model-details].

3.1 Model set-up

For the regression model, we use number of tickets as the independent variable and year as the predictor.

We run the model in R (R Core Team 2022). We use the default OLS methods embedded with `lm()` function.

For the Chi-square model, we first create a contingency table between the ticket type and the age group, then utilize the `chisq.test()` function to test if there is a statistical relationship between two variables.

4 Results

The linear regression model is summarized in Table 1. From the result, we can tell that the overall tendency of the number of tickets in Toronto is decreasing. However, the parameter of the regression is not significant, which states that this is only a minor decrease.

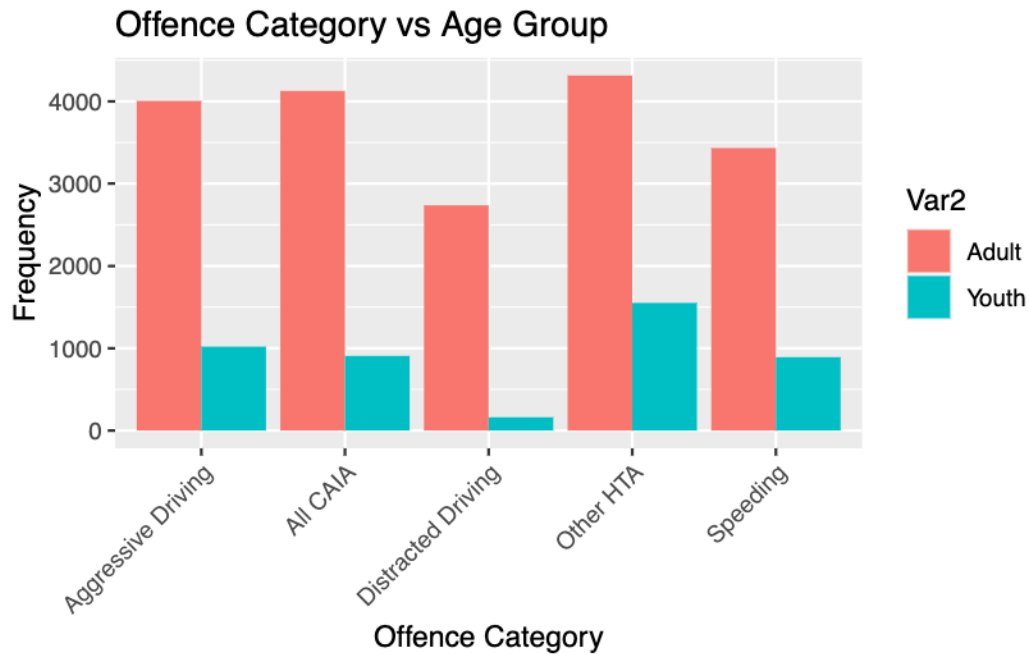
The Chi-Square result is summarized in (**chi-result?**). The p-value for the chi-square test is significant, which means that there is a relationship between age group and ticket type. As the figure (**chi-result?**) shows, the younger group will have more tickets on speeding/aggressive driving problems.

Table 1: Explanatory models of flight time based on wing width and wing length

Linear Regression model	
(Intercept)	8 842 912.60 (5 230 253.40)
OFFENCE_YEAR	−4274.20 (2591.80)
Num.Obs.	9
R2	0.280
R2 Adj.	0.177
AIC	207.6
BIC	208.2
Log.Lik.	−100.805
F	2.720
RMSE	17 705.35

Pearson's Chi-squared test

data: contingency_table
X-squared = 550.08, df = 4, p-value < 2.2e-16



5 Conclusion

Our initial research indicates notable differences in the issue of tickets between various years and neighborhoods. There seems to be a relationship between particular age groups and particular kinds of offenses. According to spatial data, certain parts of Toronto are hotspots for specific types of offenses. The findings also demonstrate temporal trends in the data, with some offense categories being more or less prevalent over time.

Our statistical analyses verify that these differences are meaningful and not merely anomalies. An intelligible and clear explanation of these patterns and trends is made possible by the data visualization.

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/>.