# Analysis of Traffic Offences Across Toronto neighbourhoods*

## The effect of neighbourhoods incomes

Gadiel David Flores

September 26, 2024

This paper examines the relationship between traffic offences and neighbourhood income in Toronto. Using police reports from 2014 to 2023 and income data from the 2016 Toronto neighbourhood census, we analyze how offences such as speeding, aggressive driving, and distracted driving vary across income levels. Through visualizations and linear regression models, we find that while a weak positive correlation exists between neighbourhood income and the frequency of citations, other factors likely contribute to traffic violations. Our findings raise questions about equitable enforcement practices and suggest further research into socioeconomic disparities in traffic offence distributions.

## 1 Introduction

Toronto has a diverse range of neighbourhoods, from busy urban centers to quieter residential areas. Even with public transportation, driving remains a popular option, and nothing is more frustrating than receiving a traffic ticket. By examining potential patterns in traffic offences, we may be able to identify traps set to catch drivers off guard or pinpoint areas in Toronto that are consistently dangerous and require attention. Analyzing this data could also help drivers avoid future speeding tickets.

We will explore how neighbourhood income affects police presence and citations overall. Another potential pattern could involve dangerous roads that are prone to human error, leading to a higher number of violations, which may be exploited by officers who are aware of these trends. We will use R Core Team (2023) for our analysis and Services (2024) for our primary data.

---

*Code and data are available at: https://github.com/DavidFJ207/Traffic_offences_Geography_Toronto

The remainder of this paper is structured as follows: Section 2 will explore our data from Services (2024) and Social Development (2024), highlighting any potential clusters and patterns in the number of offences, locations, and types of citations. We will then further analyze these findings in Section 3. Finally, we will discuss our conclusions in **?@sec-result**.

## 2 Data

In this section, we will specifically examine the types of offences and the number of citations issued from 2014 to 2023. Our focus will be on aggressive driving offences, which include tailgating, excessive lane changing, and road rage, as well as distracted driving, such as texting while driving—both of which pose risks to road safety. Other traffic offences covered are speeding, violations under the Compulsory Automobile Insurance Act (CAIA) related to vehicle insurance, and miscellaneous offences under the Highway Traffic Act (HTA). The frequency of each citation is visualized in (Figure 1) and was extracted from Services (2024). This dataset records the number of tickets issued by year, ticket type, offence, age group, division, and neighbourhood. In **?@sec-data-cleaning**, we discuss how this data was processed for our analysis. This dataset contains the most recent data available, as of August 2nd. Our analysis will focus on the total number of offences committed by neighbourhood, regardless of age.
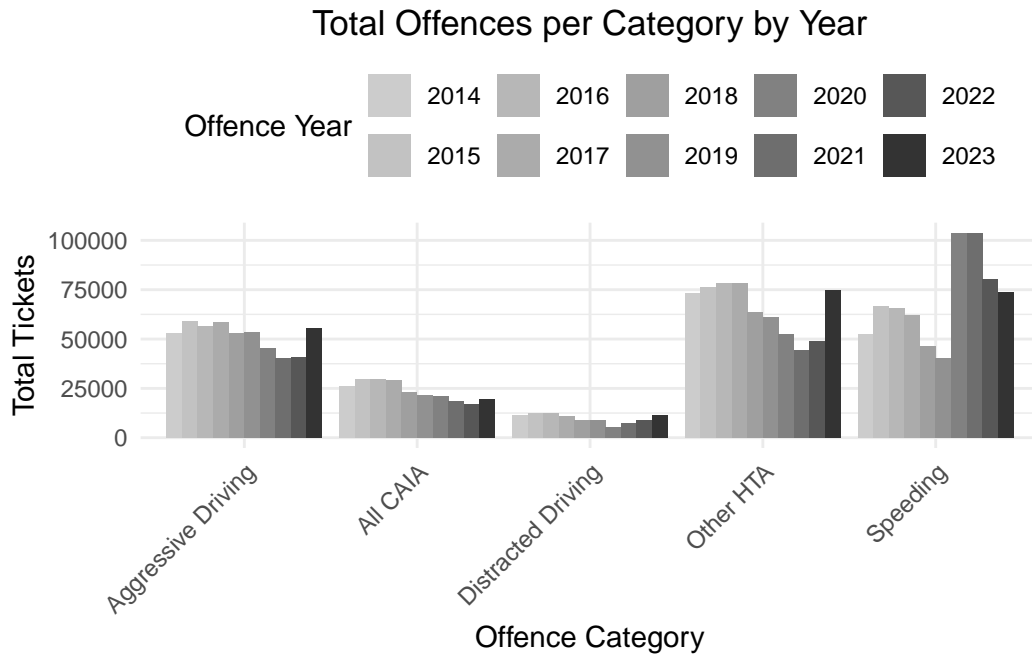


Figure 1: Frequency of Citations by Neighbourhood

As seen in this graph, speeding is the most frequently issued offence, while aggressive driving has shown little change over the years. Now that we have identified the most common offences, we will examine the neighbourhoods. In (Figure 2), we explore the relationship between neighbourhood incomes and the number of offences committed. This data comes from the 2016 Toronto neighbourhood Profiles, which is a census that collects information on households, including age, sex, education, and, for our purposes, income. Specifically, we will analyze the 2016 census data, focusing on _id 1030, "Average after-tax income of households in 2015 ($)," to approximate neighbourhood income. In **?@sec-data-cleaning**, we discuss the selection of neighbourhoods and the rationale behind these choices.

Each point in these scatterplots represents an individual neighbourhood. In these graphs (Figure 2), it is evident that there is a weak positive correlation between income and the number of citations, with offences like distracted driving and speeding showing the strongest correlations among all offence types. This is further illustrated in Section 3 and **?@sec-model-details**. Lastly, we note that the graphs display high variation and a wide spread, raising the question of whether there is issues with our correlation model between income and the number of citations issued. In the next section, we will explain how we created a correlation model to draw the red line, which helps us assess the strength of this correlation.

# 3 Model

The goal of our modeling strategy is twofold. First, to determine whether there is a relationship between neighbourhood income and the number of offences. Second, to assess how strong this relationship is.
We will use a simple linear regression model to examine the effect of income on the number of citations issued in each neighbourhood. Background details and diagnostics are included in **?@sec-model-details**.

## 3.1 Model Setup

Define $y\_i$ as the number of offences in a given neighbourhood, and $x\_i$ as the average income of that neighbourhood measured in thousands.
$E(Y|X) = \_0 + \_1(X)$

### 3.1.1 Model Justification

A linear regression model was chosen because it allows us to analyze the effect of income on the number of citations. As neighbourhood income increases or decreases, do offences decrease or increase? This will help us evaluate the strength of the relationship between citations and average neighbourhood income. While we could add more variables for a deeper analysis, that
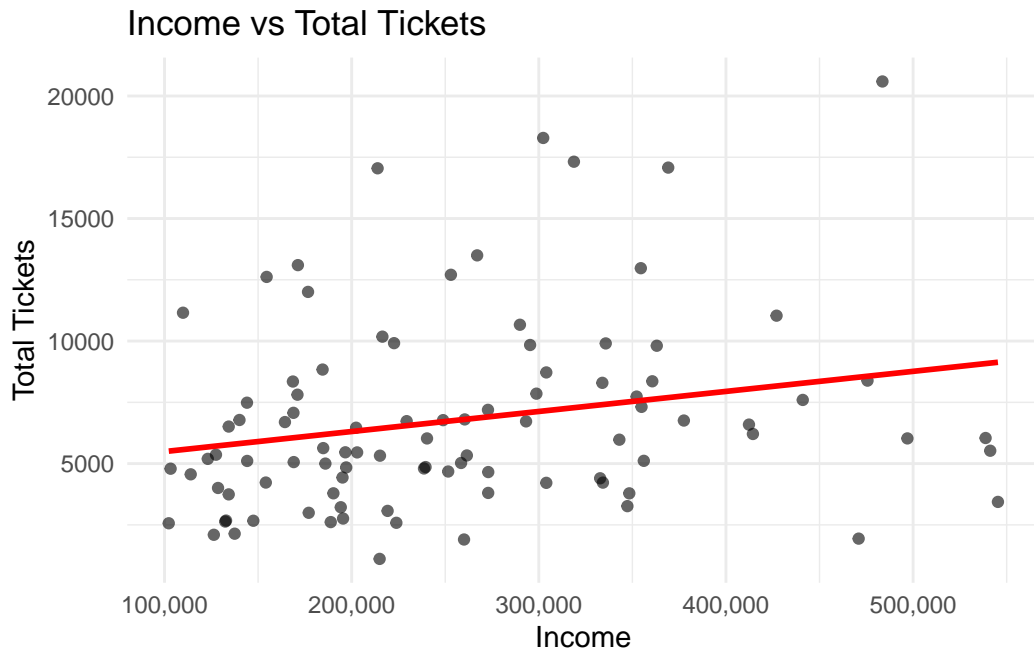
(a) Total Offences by Neighbourhood
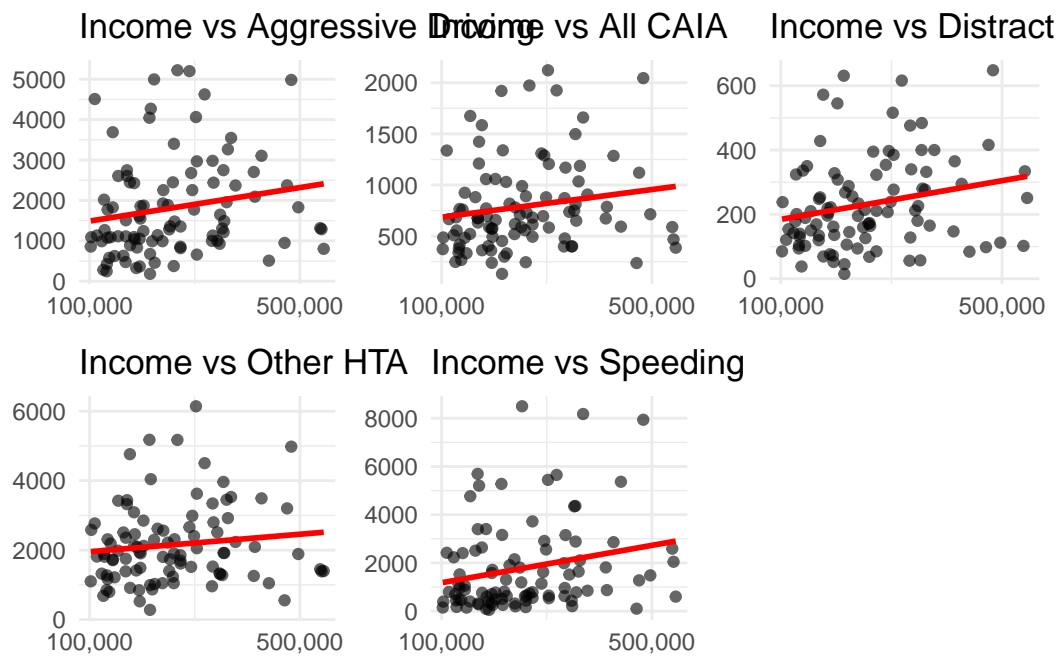


(b) Specific offence Category

Figure 2: Toronto Neighbourhoods Income vs. Offences

would fall outside the scope of this paper, which is focused solely on determining if a correlation exists between income and offences. While other models may give us a more accurate picture of the correlation, for the purpose of this analysis, we are just interested to see if there is a correlation or if other factors affect the amount of ciations given out. So for that purpose, a linear regression serves our purpose.

# 4 Results

The linear regression model was applied to examine the relationship between neighbourhood income and the number of traffic offences. The model's equation is as follows:

$E(Y|X) = \beta_0 + \beta_1 \cdot \text{average\_income}$

The summary of the linear regression model is provided below in (**ressidual_plots.These?**) plots tells us that theres other factors affecting the amount of offences committed than just the neighbourhoods income. This means that our model isnt fully cpaturing all the factors that influence the neighbourhoods offences. Although, Figure 6 does show us that aggressive driving has the strongest relationship between the number of offences and a neighbourhoods income. Overall, while there is a relationship between neighbourhoods income and their offences, there is other factors that are missing that we should consider.

# 5 Discussion

## 5.1 Exploring the Relationship Between Income and Offences

The primary focus of this paper is to determine whether there is a relationship between the number of offences and the income of a neighbourhood. The results indicate that this relationship is weak but present, with aggressive driving and speeding showing the strongest connections to income. This raises questions about why wealthier neighbourhoods are penalized more than lower-income ones. Perhaps wealthier neighbourhoods have more speed cameras, or maybe speed cameras in lower-income areas are more frequently vandalized. It is also possible that police are targeting wealthier individuals, or that wealthier individuals are more willing to pay speeding tickets, while those in lower-income neighbourhoods are more cautious about speeding. These questions lead to broader inquiries, such as whether fines should be income-adjusted.

## 5.2 Scope of Analysis

Our data also suggest that we do not have the full picture. Other variables should be considered when analyzing the correlation between police activity and specific neighbourhoods. For instance, busier neighbourhoods might receive more offences simply due to higher traffic volume compared to quieter areas. Some neighbourhoods have school zones, which may lead to an increase in offences. The time of offences could also be another factor to take into account. Lastly, considering what Lima (2022) discusses in 'Making Forest Data Fair and Open' about how data is collected, it raises questions about whether traffic data has been collected without consideration of the socioeconomic background of neighbourhoods, potentially reinforcing pre-existing disparities and affecting how offences are recorded and managed. There are risks related to who collects traffic offence data, under what conditions, and what biases they may have. Accounting for these factors could provide a clearer understanding.

## 5.3 The Broader Implications and Ethical Considerations

We have raised more questions than answers, but that was the purpose of this paper. If we can find a relationship between income and traffic offences, even if weak, there might be other factors affecting offences across Toronto. Identifying all these variables can help us make Toronto fairer and safer. At the end of the day, it is crucial to have driving laws that are enforced equally and fairly across the city. This paper should serve as a starting point to delve deeper into ensuring that police practices are just and equitable. While it might be tempting to use this data for personal benefit to avoid penalties, we have an ethical responsibility to use public data in the best interest of society. As stated in D'Ignazio and Klein (2020), 'What gets counted counts.' The data presented here can shape perceptions of traffic enforcement and reveal broader patterns of unequal enforcement across different income levels.

## 5.4 Weaknesses and Next Steps

There may be bias in the data, as the income of a neighbourhood could influence the level of police activity in that area. This could result in more citations being issued in lower-income neighbourhoods due to higher police presence. Additionally, the linear regression model we used makes certain assumptions, which are further discussed in **?@sec-model-details**, regarding how well the model fits the data. As described in D'Ignazio and Klein (2020), 'Rethinking Binaries' is crucial to capture the bigger picture, as focusing on only two variables can lead to overly simplified analyses and insufficient conclusions.

The next steps for this research would involve incorporating additional variables to provide a more comprehensive view of the factors influencing the number of citations a neighbourhood receives. Furthermore, using more sophisticated models could lead to a more accurate understanding of the relationship between income and traffic offences. Lastly, the data does

not account for citations that were dismissed or dropped after review, which could affect the
accuracy of the results.

# Appendix

# A Data Cleaning (sec-data-cleaning)

For data cleaning, our interests was all neighbourhoods that had data on their police offences and average income. The dates found offences ranged from 2014 to 2023, whilsts the the onyl data available for neighbourhoods incomes were from 2015. We used two data sets Social Development (2024) and Services (2024) and merged the information based on the neighbourhoods. It was also important to take a total of all the offences commited by neighbourhoods so it was important to add up the individual offence categories into one sum per neighbourhood. After data is merged, any data that was no available for the neighbourhoods, such as NA, were dropped. Then lastly, outliers were removed follwing an interquartile process. # Model details {#sec-model-details}

## A.1 Linear Regression

We implement a Simple Linear Regression Model.
This shows that every $1000 increase in neighbourhood income:
- Total Tickets: 8.18 additional tickets will be given.
- Aggressive Driving: 2.07 additional tickets will be given.
- All CAIA: 0.68 additional tickets will be given.
- Distracted Driving: 0.3 additional tickets will be given.
- Other HTA: 1.26 additional tickets will be given.
- Speeding: 3.87 additional tickets will be given.

## A.2 Diagnostics

Figure 3 is a Residual vs. Fitted Values Plot. It shows a pattern in the spread of residuals, suggesting potential heteroscedasticity.

Figure 4 is a Normal Q-Q Plot.It shows deviations from a straight line, suggesting that the residuals may not be normally distributed.

Figure 5 is a Residuals vs. Predictors Plot. It shows potential non-random patterns, suggesting that the predictors may not fully account for all variability in the response variable.
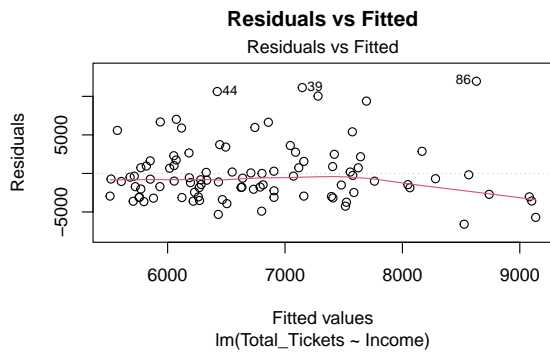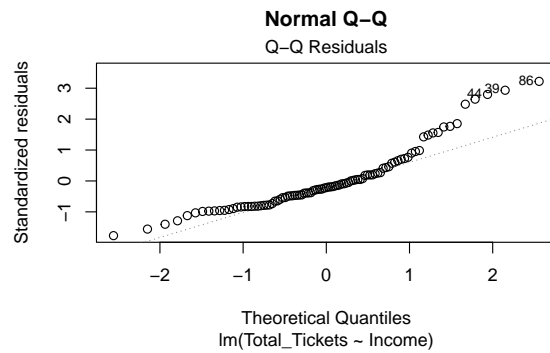
Figure 6

Figure 3: Residual vs. Fitted
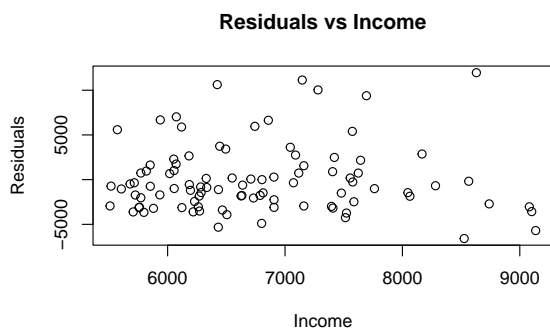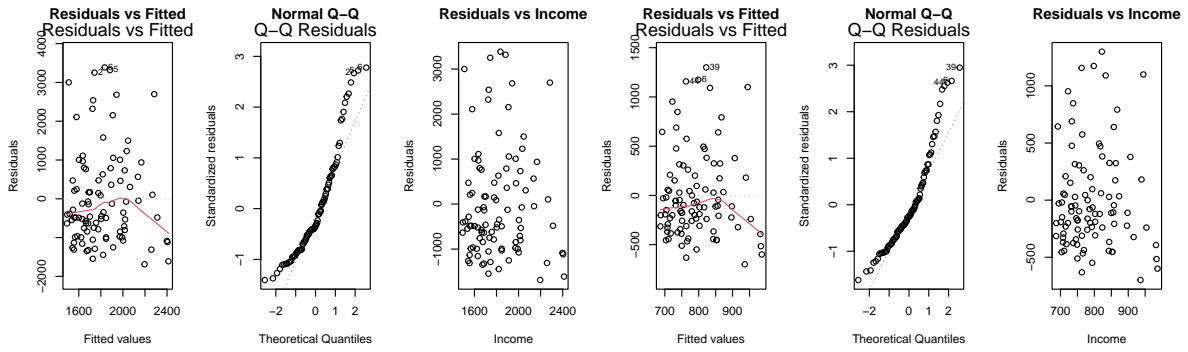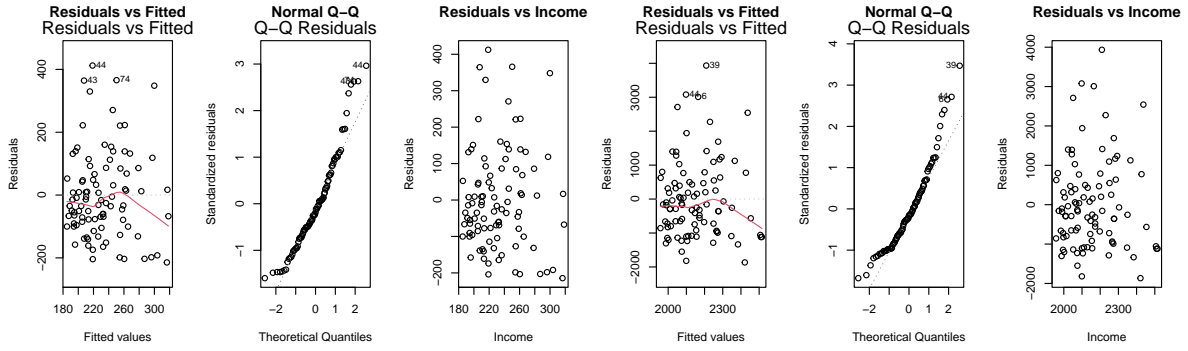


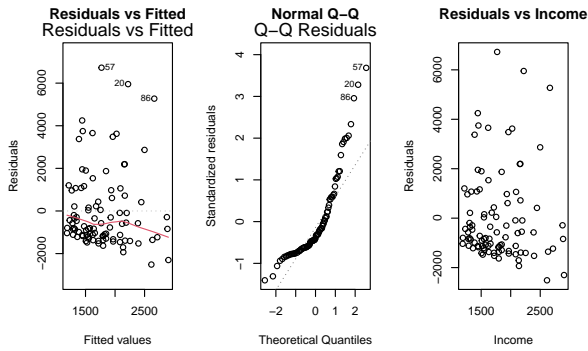Figure 4: Normal Q-Q Plot



Figure 5: Residuals vs. Predictors

(a) Aggressive Driving

(b) All CAIA

(c) Distracted Driving

(d) OtherHTA

(e) Speeding

Figure 6: Plots to Assess Residuals.

## References

D'Ignazio, Catherine, and Lauren F. Klein. 2020. *Data Feminism.* Cambridge, MA: MIT Press. https://data-feminism.mitpress.mit.edu.

Lima, Renato de de, Lima. 2022. "Making Forest Data Fair and Open." *Nature Ecology & Evolution.* https://doi.org/10.1038/s41559-022-01738-7.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Services, Toronto Police. 2024. "Police Annual Statistical Report - Tickets Issued." https://open.toronto.ca/dataset/police-annual-statistical-report-tickets-issued/.

Social Development, Finance & Administration. 2024. "Neighbourhood Profiles." https://open.toronto.ca/dataset/neighbourhood-profiles/.