

# An Analysis of Motor Vehicle Theft in Toronto: Patterns and Trends (2020-2024)\*

Xinze Wu

September 24, 2024

This paper explores patterns of motor vehicle theft in Toronto between 2020 and 2024 using data from Open Data Toronto. The analysis reveals a significant correlation between location type and the frequency of thefts, with parking lots being the most common theft sites. This study shows how public data can inform targeted strategies to prevent motor vehicle theft.

## 1 Introduction

Motor vehicle theft is getting more and more serious across Great Toronto Area. This phenomenon damages the safety of the citizens' personal property, and it has increased the price of vehicle insurance to a certain extent and increased the financial pressure on residents. Therefore, understanding the patterns and trends behind such thefts can help law enforcement and policymakers implement preventive measures.

The availability of Open Data Toronto provides a valuable resource for analyzing crime trends. This paper uses data from Open Data Toronto on thefts from motor vehicles to identify key factors that contribute to these incidents.

The reminder of this paper is constructed as follows: @#sec-data: Describe data sources and simulation process @#sec-model: Visualize the data to have a better Readability @#sec-result: Provide a comprehensive description and summary of the data. @#sec-diss: Express personal thinking on the result

---

\*Code and data are available at: [https://github.com/ke3w/term\\_paper1](https://github.com/ke3w/term_paper1)

## 2 Data

The data used in this analysis was sourced from Open Data Toronto. Specifically, the dataset focuses on reported incidents of motor vehicle theft in Toronto from 2020 to 2024. As for the key variables in the dataset, `REPORT_DATE`: The date on which the theft was officially reported.

`OCC_DATE`: The date on which the theft occurred.

`LOCATION_TYPE`: Describes the type of location where the vehicle was stolen (e.g., “Parking Lot”, “Street”).

`PREMISES_TYPE`: Indicates whether the theft occurred inside or outside (e.g., “Outside”, “House”).

`OFFENCE`: A categorical variable that describes the type of offense (e.g., “Theft From Motor Vehicle Under”).

`DIVISION`: Represents the police division that handled the report.

`HOOD_158`: Represents the neighborhood where the incident occurred.

Shown as the diagram (Figure 1), the number of reported incidents of motor theft is relatively close every year. As for 2024, the total number is approximately half of the other year’s since the dataset used in this paper only up to June 2024. The relationship between the number of cases and year indicates that there is not a strong connection between these two variables. The number of incidents is likely not to be affected by any certain year.

One of the most critical independent variable in this dataset is location, the following diagram shows the number of incidents in the top 5 locations where car thefts are most likely to occur, shown as (Figure 2).

## 3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in Appendix B.

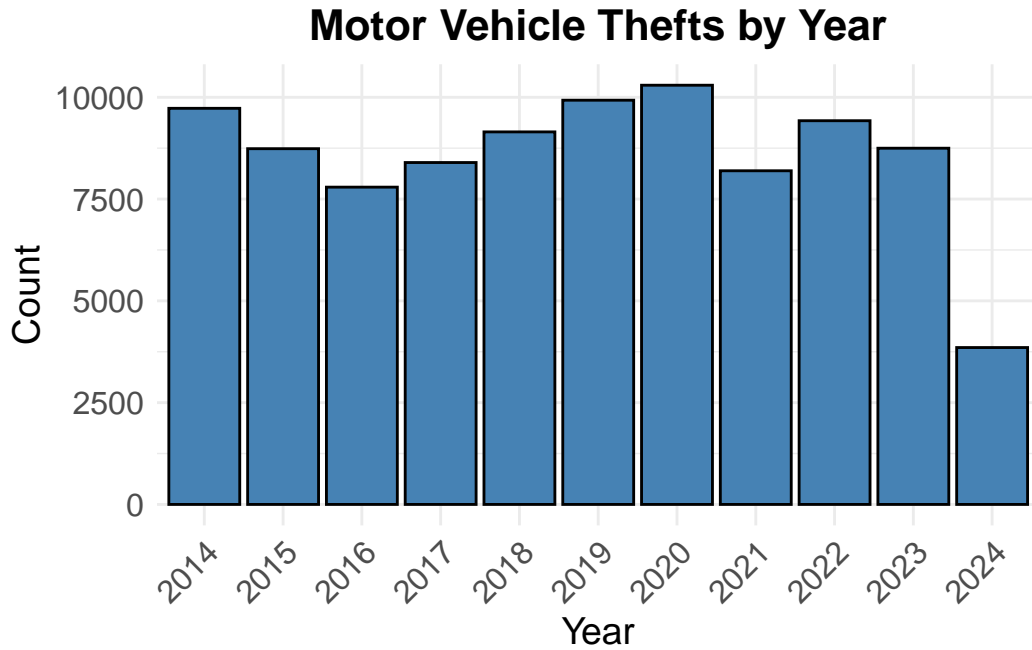


Figure 1: Motor Vehicle Thefts by Year

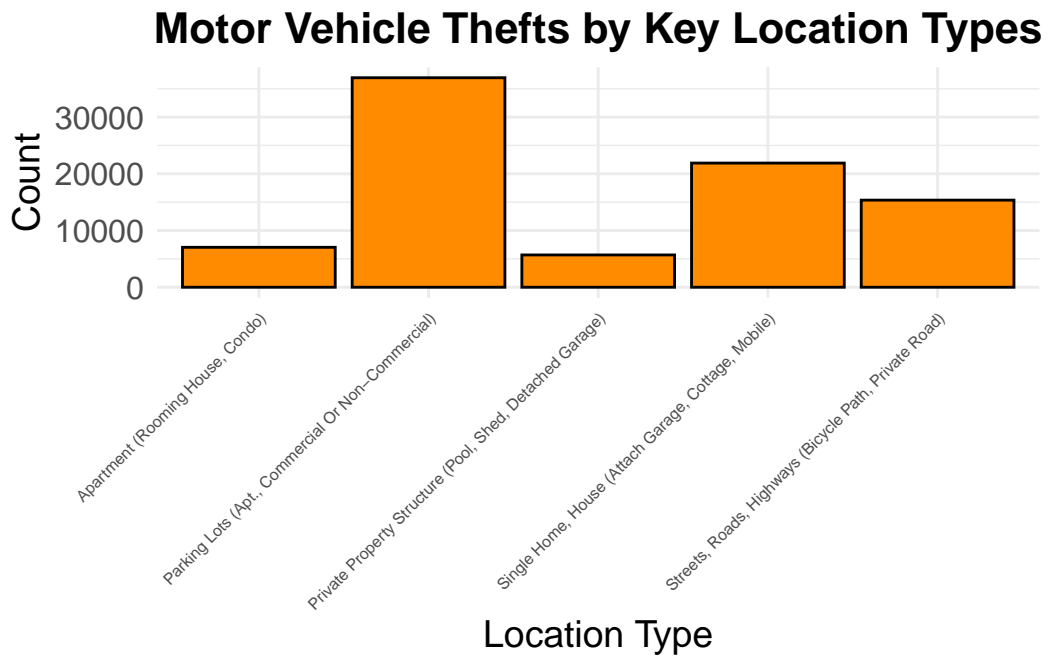


Figure 2: Motor Vehicle Thefts by Key Location Types(Top 5)

### 3.1 Model set-up

Define  $y_i$  as the number of seconds that the plane remained aloft. Then  $\beta_i$  is the wing width and  $\gamma_i$  is the wing length, both measured in millimeters.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$

$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$

$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$

$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$

$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

#### 3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance  $\theta$ .

## 4 Results

Our results are summarized in `?@tbl-modelresults`.

## 5 Discussion

### 5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

## **5.2 Second discussion point**

## **5.3 Third discussion point**

## **5.4 Weaknesses and next steps**

Weaknesses and next steps should also be included.

## Appendix

### A Additional data details

### B Model details

#### B.1 Posterior predictive check

In `?@fig-ppcheckandposteriorvsprior-1` we implement a posterior predictive check. This shows...

In `?@fig-ppcheckandposteriorvsprior-2` we compare the posterior with the prior. This shows...

#### B.2 Diagnostics

`?@fig-stanareyouokay-1` is a trace plot. It shows... This suggests...

`?@fig-stanareyouokay-2` is a Rhat plot. It shows... This suggests...

## References

- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “Rstanarm: Bayesian Applied Regression Modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.