Investigating Bicycle Theft in Toronto: Trends and Patterns*

Taejun Um

September 27, 2024

In urban environments like Toronto, bicycle theft is as common as in any major city worldwide. This paper examines the trends and patterns of bicycle theft in Toronto from 2014 to 2023, focusing on the time of occurrence, the types of locations where the theft occurred, and the cost of bicycles. The following are the main findings from this data: the overall number of bicycle thefts has decreased since 2021, the an increasing trend of occurrences of incidents in both commercial districts and public transportation and there is a notable rise in high-end bicycles getting stolen. These findings can inform community efforts and assist Toronto's public authorities in developing more effective strategies to protect citizens' property and reduce bicycle theft.

1 Introduction

Toronto is the most populous and densely populated city in Canada. As a result, traffic congestion makes it difficult for people to navigate the city, and many citizens report this as a significant issue (Hrobsky 2024). This has led to an increase in the number of cyclists using bicycles for efficient point-to-point travel within the city. Currently, approximately 659 km of cycling routes are interwoven like a spider's web throughout Toronto ("Network Status - City of Toronto" 2022). Unfortunately, bicycle thefts remain a persistent problem for residents.

This paper examines reported bicycle thefts in Toronto from 2014 to 2023. The main objective is to observe and analyze the number of thefts, the locations of the incidents, and the price range of the stolen bicycles. Specifically, we will compare trends in the locations of incidents by year alongside changes in the price range of the targeted bicycles. Through this investigation, I confirmed that thefts steadily increased from 2014 until 2020 but thefts dropped since 2021. Additionally, there have been notable changes in the locations where thefts occurred and the

*Code and data are available at: https://github.com/tonytto/Investigating-Bicycle-Theft-in-Toronto

price range of the stolen bicycles over the years. Understanding these trends and patterns will help prevent future bicycle thefts, and inform new measures by Toronto's public authorities to enhance theft prevention.

The remainder of this paper is structured as follows: Section 2 discusses the data and process of data cleaning, and Section 3 discusses our findings, limitations of the study, and suggestions to reduce bicycle theft.

2 Data

2.1 Overview

The dataset used in the paper covers reported bicycle thefts to the Toronto Police System between 2013 and 2024. This dataset was updated in June 2024 and is considered reliable as it is based on the Toronto Police system. Since the police system only has reported data, cases that are not reported to the police are not on the data. This data has variables of occurred time, reported time, stolen location, type of bike, cost of a bike, and current status of the bike.

The dataset was downloaded from Toronto's Open Data Portal using opendatatoronto package (Gelfand 2022). Its setup was completed through R (R Core Team 2023) with the help of tidyverse package (Wickham et al. 2019), janitor package (Firke 2023), knitr package (Xie 2024), here package (Müller 2020), ggplot2 package (Wickham 2016), dplyr package (Wickham et al. 2023), tibble package (Müller and Wickham 2023), and readr package (Wickham, Hester, and Bryan 2023).

The original dataset contained 28 variables, but I used the following: ID, occurrence date, occurrence year, occurrence month, occurrence hour, premises type, and bike cost. Then, I created a new group called "cost group" based on the bike cost, categorizing the bicycles with an increment of $500 according to their price range. Since the dataset is based on the date the theft was reported, I organized the data to reflect events from 2014 to 2023, excluding incidents that occurred before 2014 and after 2023. Additionally, during the data cleaning phase, observations with missing information were excluded. As a result, the dataset used in this paper contains 29,217 observations from 34,962 observations. Table 1 shows the first 5 rows of the data used.

Table 1: First 5 rows of data

id	occ_date	occ_year	occ_month	occ_hour	premises_type	bike_cost	cost_group
8	2014-01-01	2014	January	12	Apartment	1019	$1000 < $1500
12	2014-01-03	2014	January	8	Other	560	$500 < $1000

Table 1: First 5 rows of data

id	occ_date	occ_year	occ_month	occ_hour	premises_type	bike_cost	cost_group
13	2014-01-04	2014	January	18	Apartment	200	< $500
15	2014-01-04	2014	January	20	House	900	$500 < $1000
19	2014-01-05	2014	January	23	Commercial	1800	$1500 < $2000

2.2 Number of Occurrences by Time

2.2.1 By Year

First, we will examine how many bicycle thefts occurred by year. In Figure 1, the line chart records thefts yearly and the table records the number of incidents in numbers. Bicycle thefts showed an increasing trend from 2014, peaking in 2018 with 3,418 recorded incidents, followed by a noticeable decrease starting in 2021. The decrements that happened in 2021 continued through 2023. According to Table 2, the number of incidents recorded in 2022 was 2,491 which was fewer than 2,511 incidents recorded in 2014. This indicates that the trend of bicycle theft has been decreasing over the last three years.

Table 2: Number of bicycle theft incidents by year

Year	Number of Occurrences
2014	2511
2015	2724
2016	3249
2017	3321
2018	3418
2019	3103
2020	3278
2021	2643
2022	2491
2023	2479

2.2.2 By Month

Figure 2 shows the number of incidents that occurred by month with each line representing a different year. This graph indicates there are clear differences by season. Each year, there

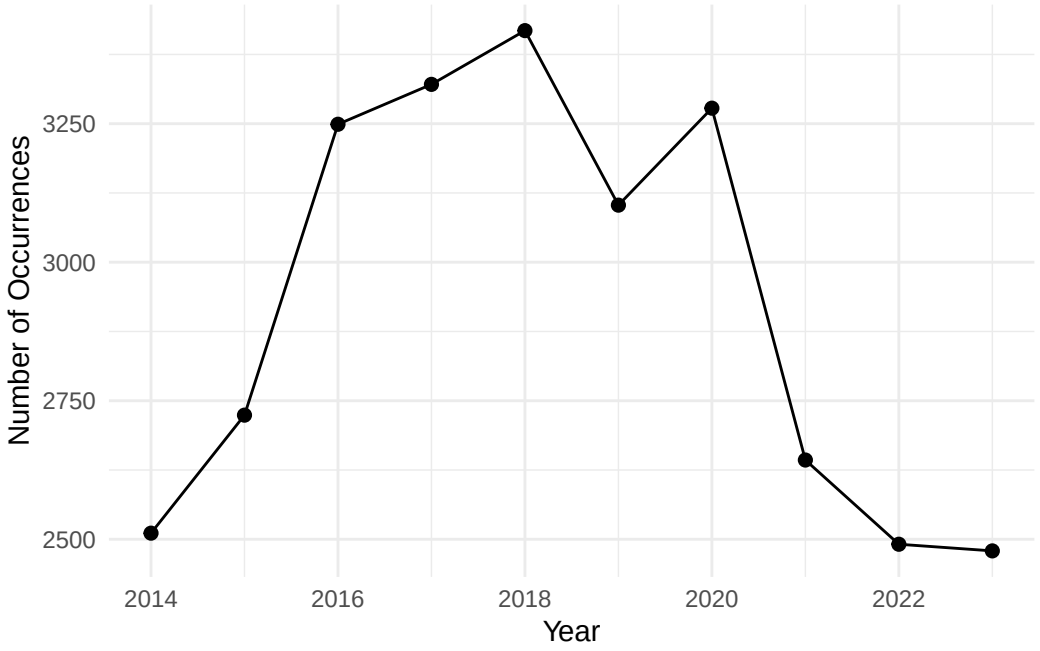


Figure 1: Number of bicycle theft incidents by year

was a sharp increase in thefts during the middle of the year, while the beginning and end of the year saw a sharp decline, showing a cyclical pattern. The number of incidents peaked in the summer months (June to August) and the incidents in January and December dropped to less than one-third compared to the summer months in all years.

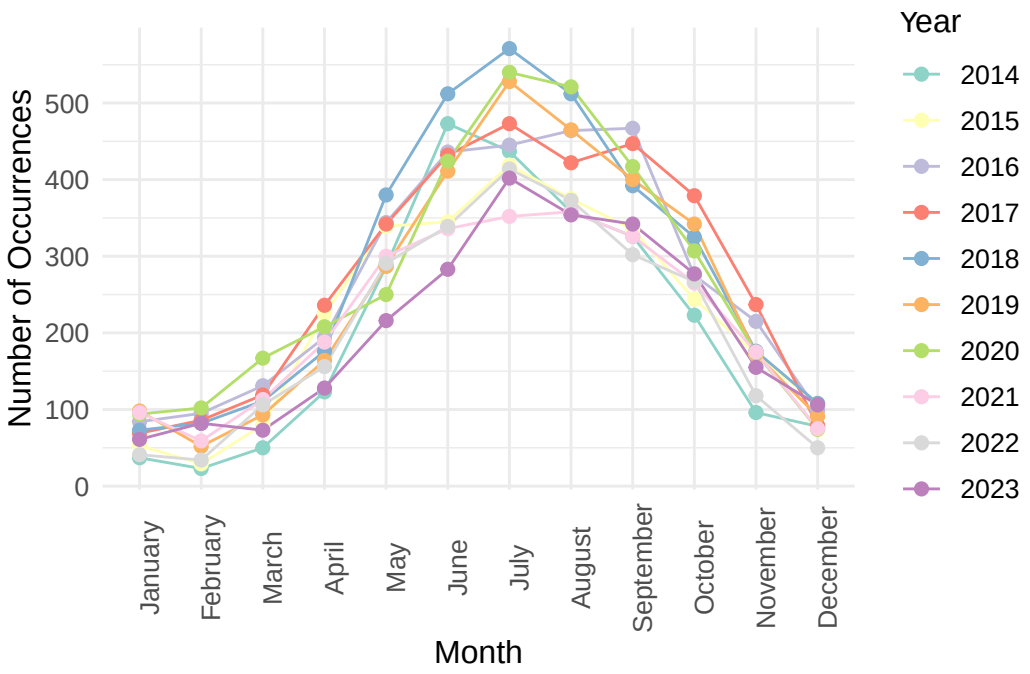
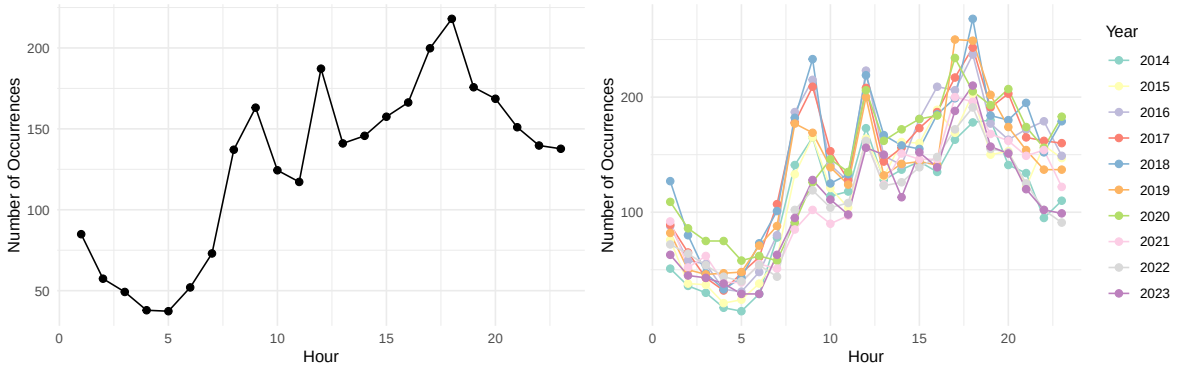


Figure 2: Number of bicycle theft incidents by month

2.2.3 By Hour

Figure 3 represents the number of bicycle theft incidents over an hour. Figure 3a shows the average number of incidents across all years, while the Figure 3b illustrates the number of incidents for each year over an hour. Figure 3a shows that there is a steady increase in thefts from 5 AM to 6 PM, and after that, the numbers begin to decline. Notably, there are three distinct peaks at 9 AM, 12 PM, and 6 PM, corresponding to periods of high social activity and movement. On average over the past decade, the lowest number of incidents occurred at 5 AM, with 37.3 thefts reported annually, whereas 6 PM saw the highest, with 218 thefts per year. From 1 AM to 7 AM, fewer than 100 incidents were recorded each year, but beyond that, the number consistently exceeded 100.



(a) Average by Year

(b) Separated by Each Year

Figure 3: Number of bicycle theft incidents by hour

2.3 Types of Crime Scenes

In this dataset, crime scenes have been categorized into seven types: Apartment, Commercial, Educational, House, Outside, Transit, and Others. As seen in Figure 4, the number of thefts shows a different pattern depending on the crime location. In the case of apartments, incidents surged from 315 in 2014 to 1,101 in 2020, marking a 249.5% increase over seven years. However, residential areas like apartments and houses peaked in 2020 and then declined. Apartment thefts dropped by 45.1% in just three years, with 604 incidents recorded in 2023, while house thefts saw a 42.3% decrease. 'Outside' sector also experienced a peak in 2018, followed by a downward trend until 2023, with a 43.7% decline. 'Commercial' buildings have shown a sinusoidal pattern since 2014. Although they hit a low point in 2021, thefts increased by 25.7% by 2023, indicating a recent and consistent upward trend. The 'Transit' sector, while having relatively fewer incidents compared to other sectors, has seen a rising trend since 2020. Starting with 40 incidents in 2020, the number nearly doubled to 78 by 2023, marking a 95% increase over three years. The 'Academic' and 'Other' sectors showed little variation in incident numbers by year. However, both sectors experienced declines in 2020 and 2022 respectively, followed by gradual increases.

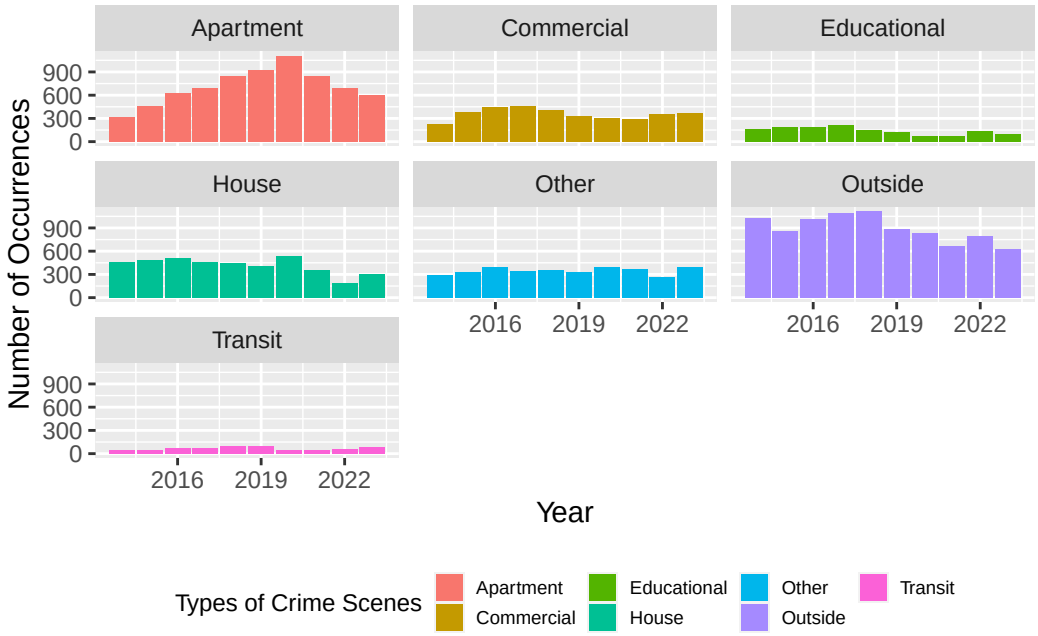


Figure 4: Number of occurrences categorized by crime scene

2.4 Cost of Stolen Bicycle

The original dataset includes a variable for the price of stolen bicycles, but it does not contain a variable for grouped prices. Therefore, I created a new variable called cost_group by grouping the prices in increments of $500 and it's divided into 5 groups. Figure 5 shows a bar graph with number of bike stolen by year categorized by cost group. The patterns of occurrence depend on the price range of the stolen bicycles. Bicycles under $1,000 have been on a downward trend. Bicycles under $500 and between $500 and $1,000 decreased 52.2 percent and 42.2 percent, respectively, in 2023 compared to 2017 and 2018. The number of stolen bicycles under $1,000 fell by about half from their peak, while high-end bicycles saw an increase in the number of occurrences. Bicycles priced between $1,000 and $1,500 saw a 64.4% increase in theft compared to 2014, while those priced above $2,000 saw a 119.4% increase compared to 2014. Bicycles ranging from $1500 to $2000 have recently seen a 34.8% increase from 2021 to 2023, although the number of thefts has not increased as much as other $1,000+ bikes.

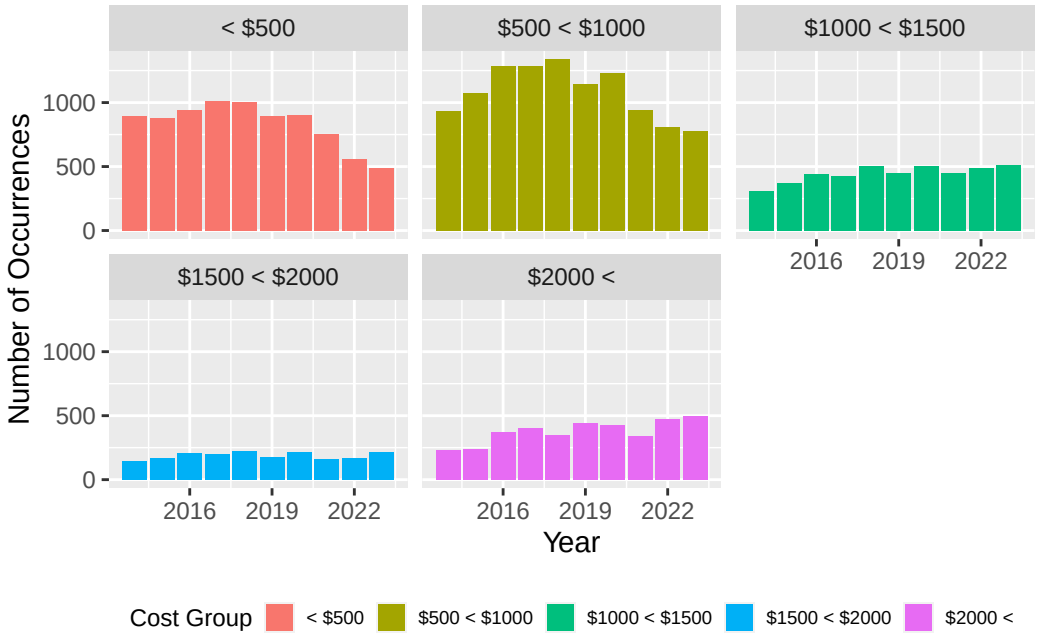


Figure 5: Number of occurrences categorized by price of bicycle

3 Discussion

3.1 Overall

The analysis of bicycle theft in Toronto has revealed trends and patterns over the ten years of theft incidents. Most of the trends shifted starting in 2021, which appears to be a result of the lockdowns due to the pandemic. First, the overall number of thefts began to decrease from 2021, indicating a significant reduction in outdoor activities. Bicycle owners were riding their bikes less frequently, and thieves were also unable to engage in outdoor activities due to the lockdowns.

One of the most notable trends is that the monthly theft graph exhibits a cyclical pattern. This indicates that most crimes occur in the summer, while very few happen in the winter. The winter in Toronto, with its snow and cold weather, makes cycling difficult for both owners and thieves, which likely explains the lack of activity in both groups during this season.

When examining thefts by location, we observed significant changes in residential area thefts. With the increase in remote work and online learning, more people have been at home, leading to a substantial decrease in thefts within residential spaces.

There were also intriguing observations regarding the price range of stolen bicycles. Year after year, the number of thefts of more expensive bicycles has been on the rise. This can be interpreted in two ways: either bicycle thieves have started to specifically target high-end bicycles, or the prices of bicycles available on the market have increased. What is clear is that the rising theft rates of expensive bicycles have led to more severe property losses for citizens. Bicycle theft may be considered a minor issue, but it is a clear crime that deprives someone of their valuable property or an essential means of transportation for their life.

3.2 Implication

According to this data, the Toronto Police should strengthen patrols and response efforts during the summer to prevent bicycle thefts. As of the data up to 2022, thefts in residential areas have significantly decreased; however, with the end of the lockdown, there is a possibility that these incidents may rise again. Therefore, it would be beneficial to conduct preventive campaigns targeting residential areas.

Regardless of their price, all bicycles are important, but there has been a recent trend of increasing thefts of high-end bicycles. To protect citizens' property, it is essential to deliver messages of caution to the community.

3.3 Weaknesses and next steps

This dataset has a significant amount of missing data, and unfortunately, I had to drop some parts of the dataset. Additionally, only theft incidents that were reported are included, meaning that incidents that were not reported for various reasons are not accounted for. These missing pieces of information may have influenced the trends or patterns identified in this paper.

References

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Hrobsky, Martin. 2024. "The Majority of Residents Say There Is a Traffic and Congestion Crisis in the Greater Toronto and Hamilton Area." https://www.ipsos.com/en-ca/majority-residents-say-there-is-traffic-and-congestion-crisis-greater-toronto-and-hamilton-area.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Müller, Kirill, and Hadley Wickham. 2023. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

"Network Status - City of Toronto." 2022. https://www.toronto.ca/services-payments/streets-parking-transportation/cycling-in-toronto/cycling-pedestrian-projects/network-status/.

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

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

Xie, Yihui. 2024. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.