Public Security Concern: Bicycle Theft in Toronto Shows a Seasonal and Geographical Pattern*

Yufei Liu

November 5, 2023

Bicycle theft has been a serious issue for city safety in Toronto. The analysis uses data on opendatatoronto portal including occurrence date and details about Bicycle Thefts occurrences reported to Toronto Police Service. We find a decreasing trend and low recovery rates in overall offences in recent years even people tend to report immediately, a seasonal pattern that more offences occurred in summer, and a geographical pattern that outdoor bicycle thefts decreased. It implies possible ineffectiveness of tracking and security management, and further work could look at influencing factors contributing to the patterns.

1 Introduction

Cycling is one of the most popular transportation methods and relaxing activities in Toronto and contributes to reducing traffic congestion. The City of Toronto supports cycling by building infrastructures, such as bike lanes and bike parking facilities, and create a cycling-friendly environment (City of Toronto 2023). However, together with the prevalence of cycling, bicycle theft became a major concern for cyclists and the city safety in Toronto.

In this paper, we use R (R Core Team 2022) to analyze `bicycle thefts` dataset (Toronto Police Service 2023a) in `opendatatoronto` package (Gelfand 2022) to discover any potential time and geographical patterns in bicycle theft occurrence in Toronto from 2014 to 2022. We explore time and location variables as well as recovery rate and report habit through bar plots, scatter plots, and tables. We find that the overall occurrence of bicycle theft in Toronto decreased in recent years with a potential seasonal pattern that more thefts occurred in summer, and a geographical pattern that outdoor thefts decreased, along with relatively

*Code and data are available at: https://github.com/Florence-Liu/bicycle-theft

law recovery rates and immediate report habit. Biases exist in the dataset and may influence some analysis. Further work could explore possible reasons behind the patterns combining with other dataset.

The remainder of this paper is structured as follows: Section 2 discusses the data with Section 2.1 includes basic information about data collection and cleaning results, and Section 2.2 includes graphs and tables representing relationships between variables and some discussions.

2 Data

2.1 Data introduction

The dataset used in this analysis was obtained from Toronto's Open Data Portal and can be accessed from the package `opendatatoronto` (Gelfand 2022). The dataset `Bicycle Thefts` includes all bicycle theft occurrences reported to the Toronto Police Service from 2014 to 2022 (Toronto Police Service 2023a). Since the occurrence of bicycle theft was self-reported, there may exist reporting bias that the reporter may choose to hide or inaccurately report some information (Toronto Police Service 2023b). Also, the location information is blurred due to privacy consideration. We use the latest `csv` version refreshed on Apr 4, 2023 in this analysis.

We will use `R` to do the analysis in this report (R Core Team 2022). After acquiring the dataset from package `opendatatoronto` (Gelfand 2022), we used package `tidyverse` (Wickham et al. 2019) and package `janitor` (Firke 2021) to clean the data and select variables for further analysis.

We selected variables from the raw data (Toronto Police Service 2023a):

- `occ_year`: Year offence occurred
- `occ_month`: Month offence occurred
- `occ_date`: Date offence occurred
- `report_date`: Date offence reported
- `premises_type`: Premises type of offence
- `status`: Whether the bicycle is recovered, stolen, or unknown

To analyze the relationship between occurrence date and report date of offence, we constructed a new numerical variable `difference` to represent the difference in days between the report date and occurrence date, and filter any offences whose difference is larger than 365, that means we will not analyze offence that is not reported within one year from the occurrence date. The number of offence that reports long time later is quite small and may not be representative due to possible inaccurate report information.

After changing the type and name of some variables, filtering offence occurred between 2014 and 2022, and removing missing values, we selected our final variable for analysis:

- `occ_year`: Year offence occurred
- `occ_month`: Month offence occurred
- `difference`: Difference in days between occurrence date and report date
- `location`: Premises type of offence
- `bicycle_status`: Whether the bicycle is recovered, stolen, or unknown

2.2 Data visualization

In the analysis, we will use package `knitr` (Xie 2014), package `here` (Müller 2020), and package `formattable` (Ren and Russell 2021) to load the cleaned data and make figures and tables to explore relationship between variables.

Figure 1 shows monthly number of offence for each year from 2014 to 2022. The bar plot provides a clear demonstration for not only the overall trend of bicycle thefts through years, but also the potential seasonal pattern that could be further explored. It shows that the total number of offence in a year increased before 2018 and began decrease since 2018. The change is relatively not significant within 1000, and the number of offence in 2022 is similar to that in 2014. Possible reasons include improvement of security equipment of the bike, the market behind bicycle theft influenced by pandemic, and people's usage habit of bike. These implications need more data to analyze and could be our further interest.

Figure 1 also illustrates potential seasonal pattern that there are much fewer bicycle thefts in winter, eg. December, January, and February. The offence increases a lot in other seasons especially in summer, eg. June, July, and August. This may be related to the weather in Toronto that many people will not ride bikes in winter due to extremely freezing weather and frequent snowfall.

Other than seasonal pattern, Figure 2 also demonstrates the relationship between the location type and number of bicycle theft for each year. It clearly shows that bicycles were mostly stolen when parked outside, and this aligns with common sense that outdoor parking is more vulnerable to theft. However, it is surprising that the second most common location for bicycle theft is apartment instead of house. Residents' transportation habits may be a possible reason that since apartments are usually located in the town center and close to public transportation stations, bicycles is a good choice for short trips. For residents in house, they may rather choose vehicles that are faster such as cars to commute.

Figure 2 also illustrates that the number of offence happened outside decreased while thefts happened indoor especially in residence increased in recent years. Factors contributing to the phenomenon could be further explored combining with other data.

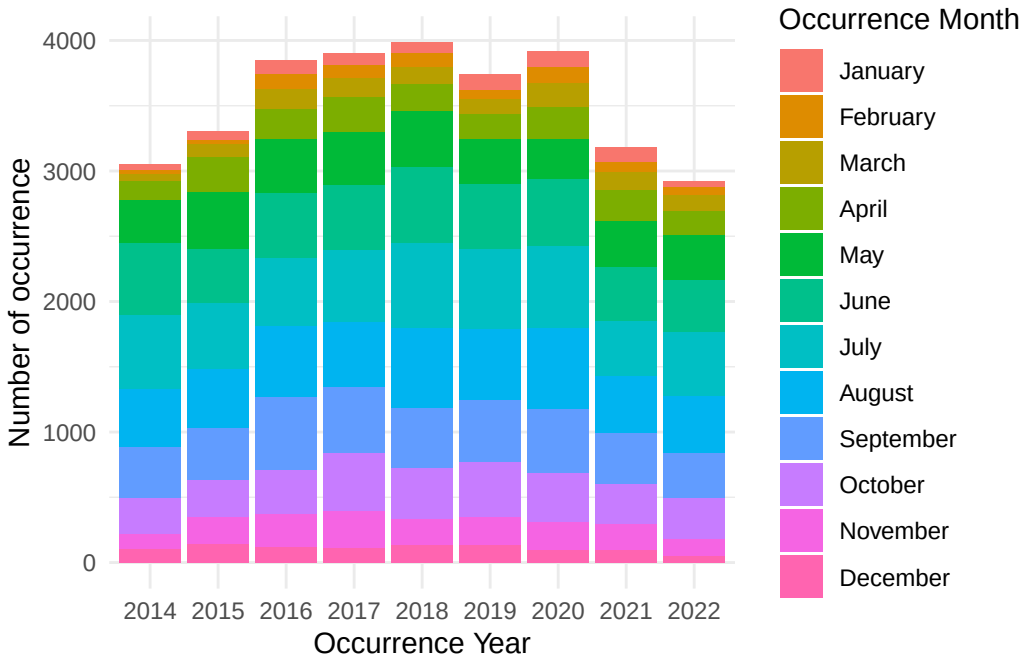

Figure 1: Monthly number of occurrence of bicycle theft in Toronto for each year from 2014 to 2022

It is noticeable that there are many reports with location type `Other`. Since it is not specified and the amount of `Other` type location cannot be neglected, how we treat the data may influence the analysis, and there may exists possible reporting bias.

Other than external pattern, we are also interested how internal variables associated with bicycle thefts. Figure 3 shows the monthly difference in days between occurrence date and report date for each year from 2014 to 2022. It illustrates that people tend to report the offence within the same day of occurrence since there are more scatters near 0. Combined with Figure 1, there are much fewer thefts happened in December, however, the interval between occurrence and report date is much shorter in December since there are much more dots representing December near 0.

Table 1 and Table 2 contains recovery rate of bicycle thefts in each year and month respectively. The recovery rate is calculated as the percentage of stolen bicycle that could be successfully returned to the owner. It does not have a clear trend of recovery rate through years, however, 2022 has the lowest recovery rate, which is much less than that in other years (See Table 1). This is not a good signal that the three lowest recovery rates come from 2022, 2019, 2021, although the highest recovery rate comes from 2020. Such unusual pattern is worth investigating.

Table 2 shows that the recovery rate in winter is lower than that in summer in general. Consider potential seasonal pattern for bicycle thefts in a year, this may also be related to the weather in

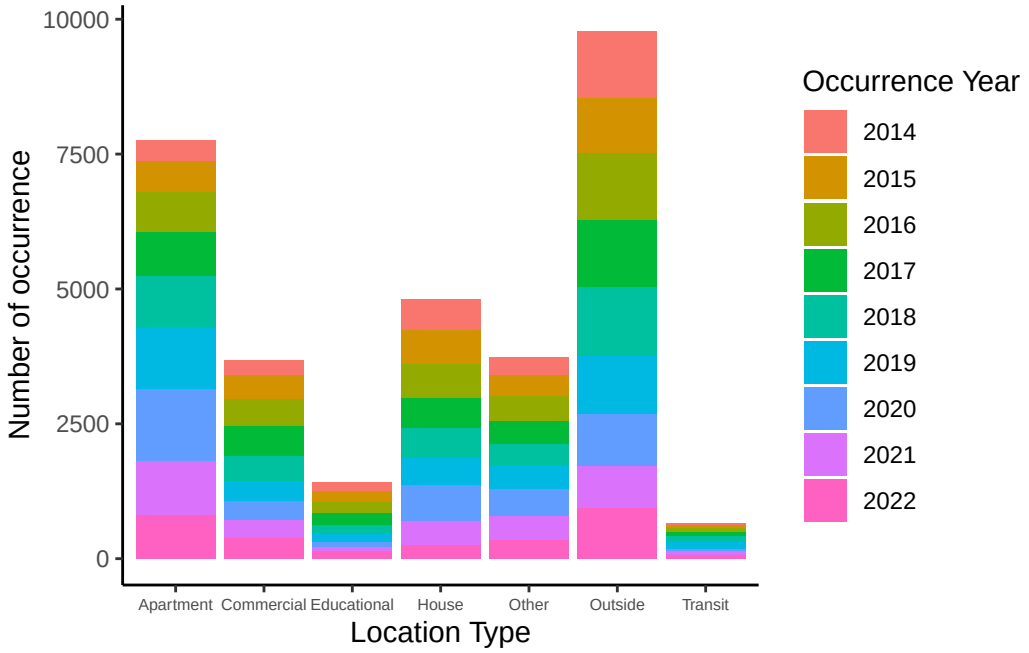

Figure 2: Number of bicycle theft in different type of location in Toronto for each year from 2014 to 2022.

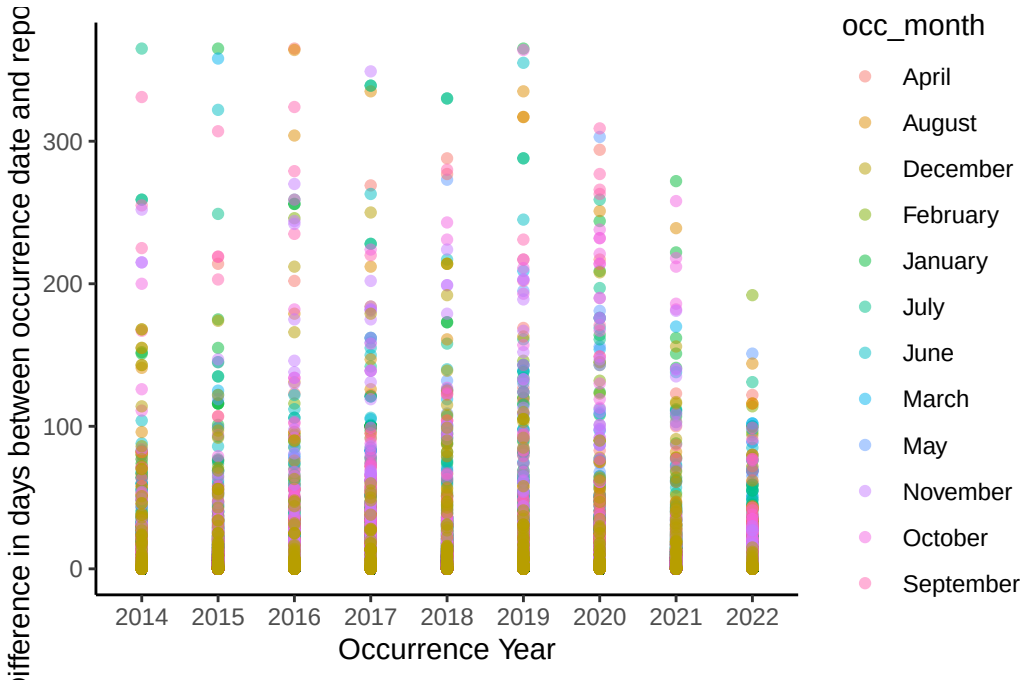

Figure 3: Scatter plot of monthly difference in days between occurrence date and report date for bicycle thefts in Toronto for each year from 2014 to 2022.

Toronto that severe and freezing weather in winter constrains people's activity in both positive and negative way. Even recovery rate is higher in summer, it is still not a high value that the largest recovery rate for each year is 17.58% and for each month is 15.70%.

In general, the data shows the overall trend of bicycle theft in Toronto decreased in recent year with potential seasonal pattern that more offence occurred in summer than in winter, and location pattern that outdoor theft decreased but indoor theft increased in recent year with Figure 1 and Figure 2. Some implications about people's reporting habit and the overall recovery rate are delivered by Figure 3, Table 1, and Table 2. Although the occurrence of bicycle theft decreased recently, bicycle theft is still common in Toronto and it is hard to get the stolen bikes successfully returned due to many factors such as lack of surveillance and ineffectiveness of tracking measures.

Table 1: Yearly recovery rate of bicycle theft in Toronto from 2014 to 2022.

Year	Recovery Rate
2022	2.20%
2019	6.91%
2021	7.85%
2015	13.50%
2017	13.81%
2018	15.38%
2014	15.70%
2016	16.01%
2020	17.58%

Table 2: Monthly recovery rate of bicycle theft in Toronto from 2014 to 2022

Month	Recovery Rate
January	1.88%
February	2.20%
December	2.51%
March	5.65%
November	7.54%
October	10.67%
April	11.30%
September	12.56%
August	12.87%
July	12.87%
May	13.19%
June	15.70%

References

City of Toronto. 2023. *Cycling in Toronto*. https://www.toronto.ca/services-payments/streets-parking-transportation/cycling-in-toronto/#:~:text=Cycling%20is%20one%20of%20the,environment%2C%20and%20promotes%20physical%20activity.

Firke, Sam. 2021. *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.

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

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ren, Kun, and Kenton Russell. 2021. *Formattable: Create 'Formattable' Data Structures*. https://CRAN.R-project.org/package=formattable.

Toronto Police Service. 2023a. *Bicycle Thefts*. https://open.toronto.ca/dataset/bicycle-thefts/.

———. 2023b. *Online Reporting*. https://www.tps.ca/services/online-reporting/.

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.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.