

Data Analysis of Crime Statistics*

Toronto's Police Service Annual Statistical Report, 2014 - 2019

Youjing Li

31 January 2021

Abstract

This report examines reported and cleared crime offenses using open-source crime data provided by Toronto Police Services on Open Data Toronto. Using various methods of data analysis, we estimate emerging crime patterns and suggest limitations toward the current collection of data. Our results provide evidence that detailed collection and proper storage of criminal offense data may provide an alternative information source to improve crime investigations.

1 Introduction

Analysis on crime statistics allows a greater understanding of the dynamics of criminal activities and is important to governments, law enforcement agencies, and the residents themselves (Phillips and Lee 2011). Although law enforcement agencies have the primary responsibility to monitor and reduce the rate of criminal activities (Baculo et al. 2017), residents may increase their awareness of local crime activities using open-source data (Smith 2014). Therefore, data analysis on publicly available data provides an alternative information source to inform residents of local crime activities. This report examines data provided by Toronto's Police Service Annual Statistical Report and suggest important trends in law enforcement—the number of crimes reported are increasing while the number of crimes cleared are decreasing.

In the following sections, we will first explain the source of the data—what the dataset is, where the dataset came from, and how the dataset was obtained. Second, we will have a discussion around any biases and limitations that exist within the dataset. Lastly, we will compute statistics and generate visualizations using the dataset to show emerging trends.

2 Data

This paper uses the R statistical programming language (R Core Team 2020) to perform data analysis, and more specifically, the `opendatatoronto` package for data imports (Gelfand 2020), the `tidyverse` package for data manipulation (Wickham et al. 2019), the `kableExtra` package for table formats (Zhu 2020).

2.1 Data Source

This paper uses the R package `opendatatoronto` (Gelfand 2020) to obtain Toronto's Police Service Annual Statistical Report on reported crimes. The data includes all crime offenses reported to the Toronto Police Service between 2014 to 2019, and has been aggregated by year, category, subtype, and geographic division (Toronto Police Services 2020). For this paper, we will examine the number of crimes reported and the number of crimes cleared using filters like year, category, and subtype.

Description of key data features below:

- **ReportedYear:** Year crime was reported

*Code and data are available at https://github.com/lilydia/Toronto_Crime_Rates.

- **Category:** Crime category
- **Subtype:** Crime category subtype
- **Count_:** Total number of crimes
- **CountCleared:** Total number of crimes identified as cleared

2.2 Bias and Limitation

There are a few considerations when using this dataset to estimate crime rate patterns. First, the data includes all crimes reported to the Toronto Police Service, including, and not limited to, those that may have been deemed unfounded after investigation, those that have no verified location, and those that may have occurred outside the City of Toronto limits (Toronto Police Services 2020). While a detailed record containing all data points can provide a better picture of reported criminal offenses, these unfounded or not valid cases will skew crime rates and present an inaccurate picture of activities that constitute criminal offenses—behaviors that are prohibited by law and considered violate the moral standards of society (Schmidt 2020). One suggestion is to provide case specific details to the current dataset so that crime rate patterns can be better studied.

Another bias is that the variable count does not indicate the number of distinct crimes—if an offense fits under multiple categories, it would be included (Toronto Police Services 2020). This again adds a layer of bias because the same case can be counted multiple times and increase the overall crime rate.

Lastly, the geographic division label of No Specified Address also includes occurrences that have no verified locations, which is not a good representation of local criminal occurrences (Toronto Police Services 2020). In this paper, we are involving all data presented in The Annual Statistical Report to provide an estimate of reported cases processed by the Toronto Police Service.

2.3 Data Analysis

The tables and figures presented in this section are all created using R (R Core Team 2020). The **tidyverse** package (Wickham et al. 2019) is used and more specifically, **ggplot2** (Wickham 2016) for generating plots, **dplyr** (Wickham et al. 2020) for hiding system messages, **readr** (Wickham, Hester, and Francois 2018) for importing the dataset. Additional table formatting support is also provided by **kableExtra** (Zhu 2020).

In crime statistics, an offense is cleared if one person has been arrested, charged with the commission of the offense, or turned over to the court for prosecution (Common Sense for Drug Policy 2021). Table 1 outlines the cleared cases in relation to the reported cases to consider the progression of resolved cases as the number of reported cases increase. In comparison, we see that there is a steady increase in cases reported over time while the increase in cases cleared remains the same, causing a 10% drop in cleared crimes from 2014 to 2019.

Table 1: Total Number of Crimes Over Time (2014-2019)

Year	Cases Committed	Cases Cleared	Percent Cleared
2014	113277	59803	0.53
2015	117290	59465	0.51
2016	122572	59889	0.49
2017	129931	60780	0.47
2018	143280	59975	0.42
2019	144570	60064	0.42

To look at the potential causes of these unsolved (uncleared) crimes, we aggregate this target, the percent cleared, by different categories of offenses. Table 2 shows the average annual cases by 6 main categories of crimes. While majority of the categories indicate a high clearance percentage, crimes that are labeled “Crimes Against Property” shows a low clearance percentage at 30%—only 30% of the property crimes are resolved (cleared).

Table 2: Average Annual Cases by Crime Type (2014-2019)

Category	Cases Committed	Cases Cleared	Percent Cleared
Controlled Drugs and Substances Act	3652	3456	0.95
Crimes Against Property	80012	23742	0.30
Crimes Against the Person	27594	17029	0.62
Criminal Code Traffic	1863	1803	0.97
Other Criminal Code Violations	15051	13686	0.91
Other Federal Statute Violations	314	280	0.89

Furthermore, Figure 1 looks at the progression of each crime category over the years. It is shown that property crimes not only have the highest count of cases, but also have the greatest rise in numbers over the years.

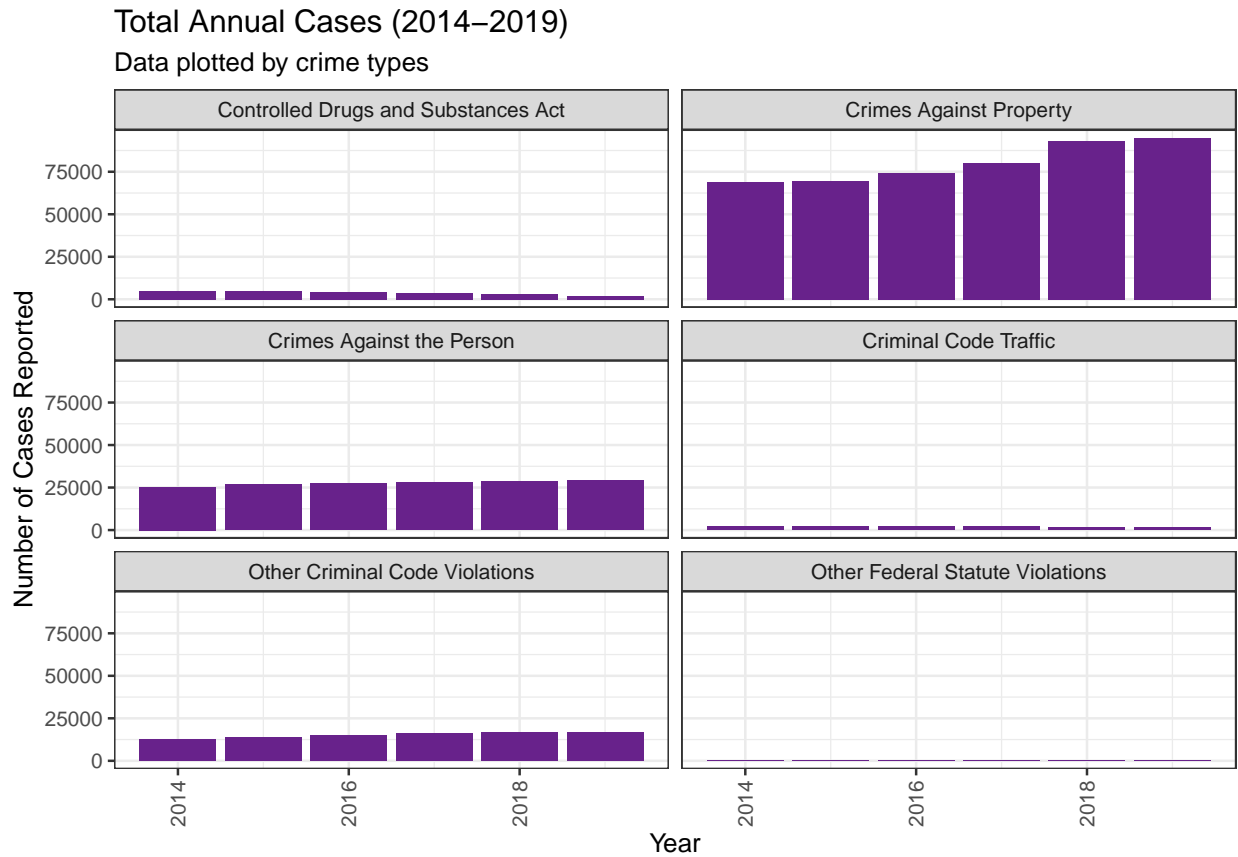


Figure 1: Total Annual Cases Plotted by Crime Categories (2014-2019)

A deeper dive into crimes against property is presented in Figure 2 and Figure 3, where we mapped the number of cases reported and the number of cases cleared according to specific subtypes. As shown in Figure 2, subtypes like theft under \$500 and fraud have a positive slope in crimes reported; however, similar increases are not observed in Figure 3 when we look at the cleared cases. The reported-cleared relationship is clear from these two plots where the crimes reported are continuing to increase in recent years while the number of cleared cases remain the same.

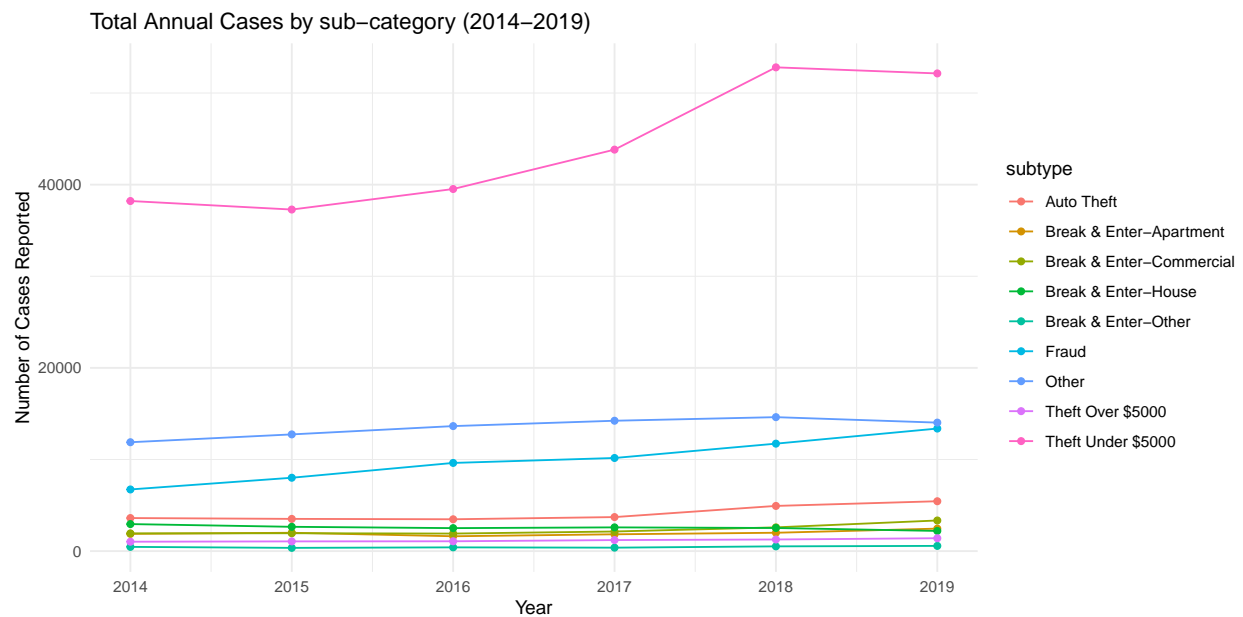


Figure 2: Total Reported Crime Cases (2014-2019)

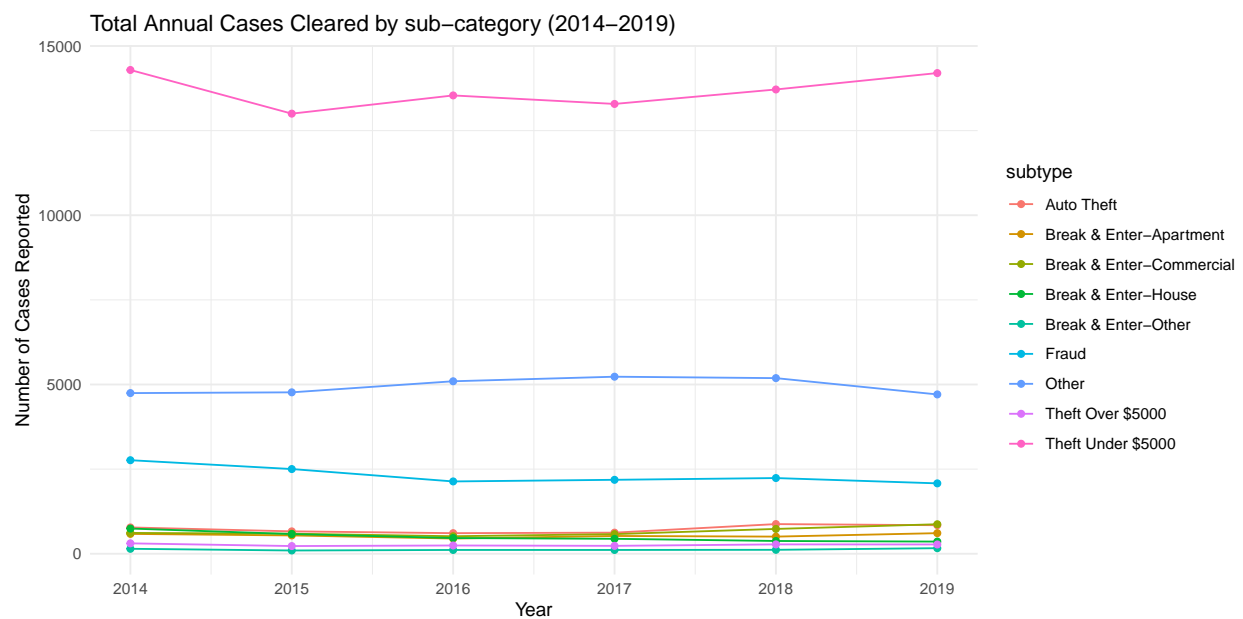


Figure 3: Total Cleared Crime Cases (2014-2019)

Perhaps what this specific dataset tells us is that the Toronto police capacity is not meeting the rise in demand—more and more cases are being reported over the years but the number of cases cleared stay consistent (see Figure 2 and Figure 3). Theft under 5000 dollars appears to be the most commonly reported crimes amongst all subtypes within the property crime category and resolution of such crime might not be prioritized in relation to other types of offenses. Moreover, the property crime category contributes more than 62% of the the total crimes reported, which suggests that the drop in percentage cleared that we witnessed in Table 1 is mainly caused by unresolved crimes reported under this category. However, limitations exist such that we cannot detect distinct crimes. As discussed in Section 2.2 Bias and Limitation, a single crime can be reported multiple times as long as they fit the category description. The great number of cases under property crimes might be caused due to the broad inclusion of activities (Schmidt 2020). Furthermore, the inclusion of unfounded cases might also increase the number of property crimes that are of little significance (eg. Theft under \$5000).

References

- Baculo, M. J. C., C. S. Marzan, R. de Dios Bulos, and C. Ruiz. 2017. *Geospatial-Temporal Analysis and Classification of Criminal Data in Manila*. <https://doi.org/10.1109/CIAPP.2017.8167050>.
- Common Sense for Drug Policy. 2021. *Definition of 'Clearance' in Crime Statistics*. Drug War Facts. <https://www.drugwarfacts.org/node/2641>.
- Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- Phillips, Peter, and Ickjai Lee. 2011. *Crime Analysis Through Spatial Areal Aggregated Density Patterns. GeoInformatica*. Vol. 15. <https://doi.org/10.1007/s10707-010-0116-1>.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Schmidt, Jonathan. 2020. *303 Legal: What Constitutes a Criminal Offense?* 303 Legal. <https://303.legal/what-constitutes-a-criminal-offense/#:~:text=Therefore>.
- Smith, Amanda M. 2014. *Police.uk and Data.police.uk: Developing Open Crime and Justice Data for the UK. JeDEM - eJournal of eDemocracy and Open Government*. Vol. 6. <https://doi.org/10.29379/jedem.v6i1.326>.
- Toronto Police Services. 2020. *Open Data Toronto: Police Annual Statistical Report - Reported Crimes*. Open Data Toronto. <https://open.toronto.ca/dataset/police-annual-statistical-report-reported-crimes/>.
- 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, and Kirill Müller. 2020. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, Jim Hester, and Romain Francois. 2018. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.
- Zhu, Hao. 2020. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>.