Investigating Relationship Between Police Expenditure and Solved Crime Rate *

Randall Ni

February 1, 2022

Abstract

The debate of whether or not to cut police funding has gone on for ages. The propagators feared police brutalities due to the constant news portrayal, while others argued that police activity should be required to ensure people's safety. Recent news of violent crimes such as the stabbings in some of the TTC stations have highlighted the urgency of this issue; the topic of this debate is more relevant to us than we think. We can use data to settle this debate once in and for all, and they are critical for understanding the correlations between police expenditure and the number of solved reported crimes. My analysis has shown that more police expenses does not necessarily make their work more efficient. Furthermore, they are extremely bad at solving property related crimes. Therefore, the view held by proponents does have some merit, and the governing body should focus more of the expense on actual crime solving.

1. Introduction

The Toronto police force ensures the safety of around 2900000 civilians daily, and their presence can deter some degrees of blatant criminal activities (Guze 2022). However, being civil servants, the Toronto police department requires funding to continue their daily operations. In 2022, the total police budget went up to approximately 1.1 billion dollars, which takes up around 7% of the entire city's operating cost (Jones 2022). This is not only in Toronto, major cities such as Montreal, Vancouver, and Calgary all have police expenses that take up a large portion of their city's budget. The question of whether we should give the police that much resources is always looming, and the death of George Flyod made this issue a hot topic. This event has also sparked the movement of "Defund the police". Although it began in the United States of America, it slowly migrated to Canadian cities as well. Critics have questioned the enormous amount of resources that the police force was using and argued that the resource (money) should be placed elsewhere such as community initiatives. One of the proponents mentioned that they "don't have a good sense of what the police do with the money we give them", and they questioned the efficiency of police work (Jones 2022). Although the debate surrounding this issue is heated, a lot of people are basing their opinions purely on what they saw in the news or on social media platforms. Therefore, it is imperative that we use data to support certain claims and prevent possible misleading information shown on our screens. Data sets regarding police expenditure and reported crimes rate are extremely important in order to understand the potential correlations between the two elements. There are some very interesting factors that we can examine to determine if some of the issues raised in the "Defund the police" movement are justified. Since reported crimes get documented regardless of the police solving them, we can easily get the percentage of solved crimes; the percentage indicates that the police force allocated enough expense and manpower and gave the victims a satisfying result. The analysis of these elements will surely settle or support some of the concerns raised by proponents of the "Defund the police" movement. In this paper, I will try to analyse the connections between police expenditure and the percentage of crimes cleared based on different types of offense in order to answer

---

*Code and data are available at: https://github.com/randalln1140/Examining-Police-Expenditure-and-Rate-of-Major-Crime

the question of whether the police should get less funding or not. More specifically, I will first investigate the trend of police expenditure and how they fluctuate from 2014-2019. Next, I will take a look at overall reported crimes within that time frame in order to determine if there are any outside influences. Lastly, I would compare the percentage and expenditure data separated by unique division numbers to address the efficiency questions of police divisions. This analysis will be performed in R (R Core Team 2022), using the dplyr (Wickham et al. 2022), knitr (Xie 2023), janitor (Firke 2021), scales (Wickham and Seidel 2022) and tidyverse (Wickham et al. 2019) packages. All figures in the report are generated using ggplot2 (Wickham 2016) and tables are created with knitr (Xie 2023).

2. Data

2.1 Data Collection Method and Possible Bias/ Flaws

This analysis report will be using two data sets: "police annual statistical report reported crimes" and "police annual statistical report gross expenditures by division", and they were all obtained from opendatatoronto's open data portal. I accessed the raw data sets using the opendatatoronto package (Gelfand 2022) and saved them in the input/data file. Both data sets are a part of the police annual statistical (provided by the Toronto Police Department) report from 2014 to 2019, and they were last refreshed on November 18th, 2022. There are some factors that those two data sets cannot address. For example, due to the Municipal Freedom of Information and Protection of Privacy Act, the police cannot disclose locations of the crime and personal information about the crime. Therefore, using the reported crimes data won't answer questions surrounding police brutality on minorities. Furthermore, the police expenditure data set also suffers from a lack of information. We can only obtain the overall expenditure of divisions separated by year; however, that does not tell us how much percentage of the expenditure actually went into crime solving. Although the above evidences mean that the attributes given does not translate into a clear answer that solves the argument completely, the remaining data on those two data sets could still produce a relatively sufficient graph and gain us more insight for our proposed correlation.

2.2 Police Expenditure Data

To begin with, I will first take a look at the police expenditure data. This raw data set contains 128 observations and seven variables. Due to the ease of analysis, I have also used the mutate function in the dplyr package (Wickham et al. 2022) to create a new column for the variable "gross_expenditure_final" called "expense_amount". I got rid of the dollar sign and commas present in the previous column for ease of graphing and made the values in the column switch from "character" to "numeric". Extract of the cleaned data set is shown below (Table 1).

Table 1: Extracting First Six Rows form the Edited Version of Police Expenditure Data

ID	Object ID	Reported Year	Category	Division Responsible	Command	Gross Expenditure Final	Expense Amount
1	1	2014	Expenditures	Division 11	Central	$25,008,526.00	25008526
2	2	2014	Expenditures	Division 12	Central	$24,858,615.00	24858615
3	3	2014	Expenditures	Division 13	Central	$21,740,037.00	21740037
4	4	2014	Expenditures	Division 14	Central	$31,984,099.00	31984099
5	5	2014	Expenditures	Division 51	Central	$31,724,579.00	31724579
6	6	2014	Expenditures	Division 52	Central	$29,684,845.00	29684845

The three most valuable variables shown with the above table are year, division (where the police station was stationed), and expense amount. These variable will yield interesting insight regarding this debate, since we can see the total amount of each division's expenses and the fluctuation of them from 2014 to 2019. I

am interested to see the general trend of police expenses throughout the years, and we can examine specific divisions' efficiency using this data.
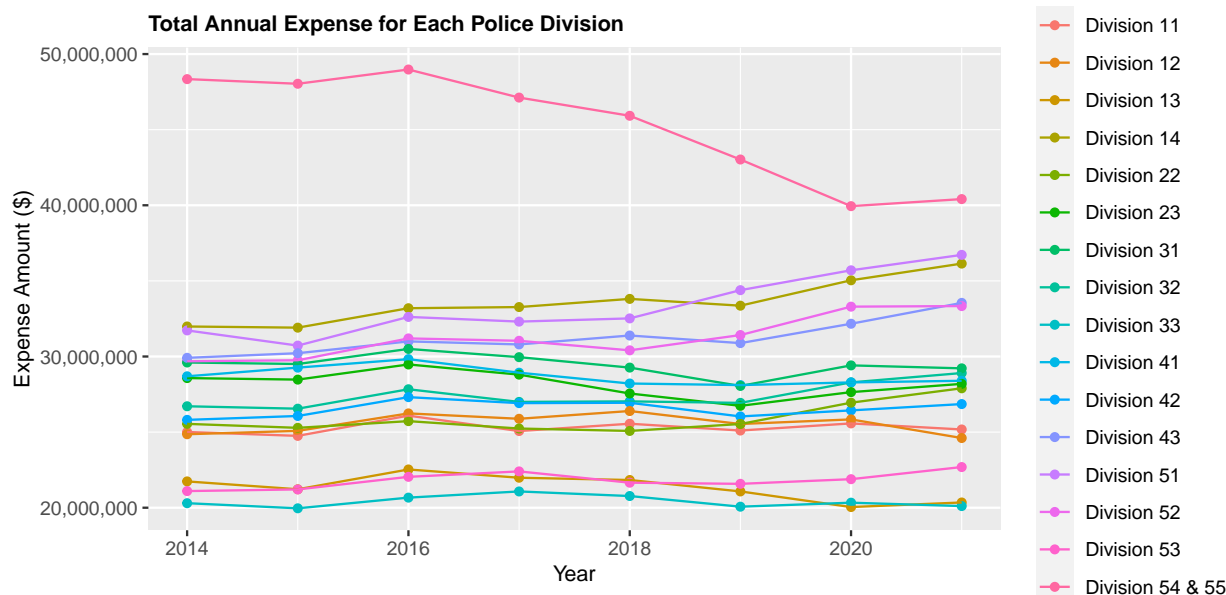

Figure 1: Expenses for Each Division Seperated by Year

I have used the scale package (Wickham and Seidel 2022) to clean up my y-axis, so it shows numbers instead of staying in a log format; this allows us to more easily identify trends and learn how much actual money police stations are spending. In figure 1 above, we can easily see that police divisions' annual expenses are astronomical from the year of 2014 to 2021. However, there is no uniformed trend amongst those divisions; around half (eight) of the divisions experienced an expense increase, while the rest either have no significant expense changes or have decreased their spendings. Division 54 & 55 is an outlier in this case, since the data set originally combined their expenses; after division 54 joined 55, the spending has reduced dramatically. For the convenience of the analysis, I will be examining the police divisions that have increased their spending in order to see whether more spending equates to a more efficient rate of solving crimes. In this case, division 14, 51, 52, and 43 all showed a steady increase in their expenses from 2014 to 2021; I will filter their data out in the next section for more specific analysis.

2.3 Reported Crimes Data

In order to make the observations above into analysis, we must also take into the account of the efficiency of the police. While a lot of data sets can show the number of incidents occurred during specific years, the data set I will be using shows the number of cases cleared (solved) as well. This allows us to gain more insight in the overall ability of the police. It is reasonable to argue that having only the amount of cases reported or having the amount of cases cleared is useless, since we cannot determine how well the police do their job and protect their people from malicious acts. Examining the total amount of reported crimes can especially be misleading. Police force is not God, therefore they cannot totally prevent the crimes from happening. The best thing they can do is to bring the perpetrator down with the force of law after crimes have been committed. Furthermore, by only examining the total number of cases, they can be easily influenced by outside forces that are unpredictable such as Covid-19 pandemic. This raw data set contains 2701 observations and eight variables, and the extract of the cleaned original reported crimes data set is shown below (Table 2).

Table 2: Extracting First Six Rows form the Original Version of Reported Crimes Data

ID	Object ID	Reported Year	Division Responsible	Category	Subtype	Cases Count	Solved Cases
1	1	2014	D11	Controlled Drugs and Substances Act	Other	201	195
2	2	2014	D11	Crimes Against Property	Auto Theft	124	43
3	3	2014	D11	Crimes Against Property	Break & Enter-Apartment	85	37
4	4	2014	D11	Crimes Against Property	Break & Enter-Commercial	58	18
5	5	2014	D11	Crimes Against Property	Break & Enter-House	89	34
6	6	2014	D11	Crimes Against Property	Break & Enter-Other	23	7

Table 2's data is very descriptive; however, we won't be needing the specific subtypes of crime categories and they are out of scope. Therefore, I created a new table that adds up each individual subtype within their main category. Furthermore, I also calculated the percentage of each division's success rate in solving crimes in each category. Extract of the cleaned, edited version of the reported crimes data set is shown below (Table 3).

Table 3: Extracting First Six Rows form the Edited Version of Reported Crimes Data

Category of Crime	Reported Year	Division Responsible	Cases Count	Solved Cases	Solved Percentage
Controlled Drugs and Substances Act	2014	D11	201	195	97.01493
Controlled Drugs and Substances Act	2014	D12	333	321	96.39640
Controlled Drugs and Substances Act	2014	D13	169	162	95.85799
Controlled Drugs and Substances Act	2014	D14	559	541	96.77996
Controlled Drugs and Substances Act	2014	D22	224	216	96.42857
Controlled Drugs and Substances Act	2014	D23	251	235	93.62550

Table 3 above gives us a clearer idea of the data set and weeded out unnecessary variables such as "Object ID" and "Subtype". The four specific variables that I wanted are "Category of Crime", "Reported Year", "Division Responsible", and "Solved Percentage". Before we dive into the specific division's efficiency by comparing their expenses and "Solved Percentage", we must first take a look at the overall amount of crime reported and see if there are any anomalies or outliers.
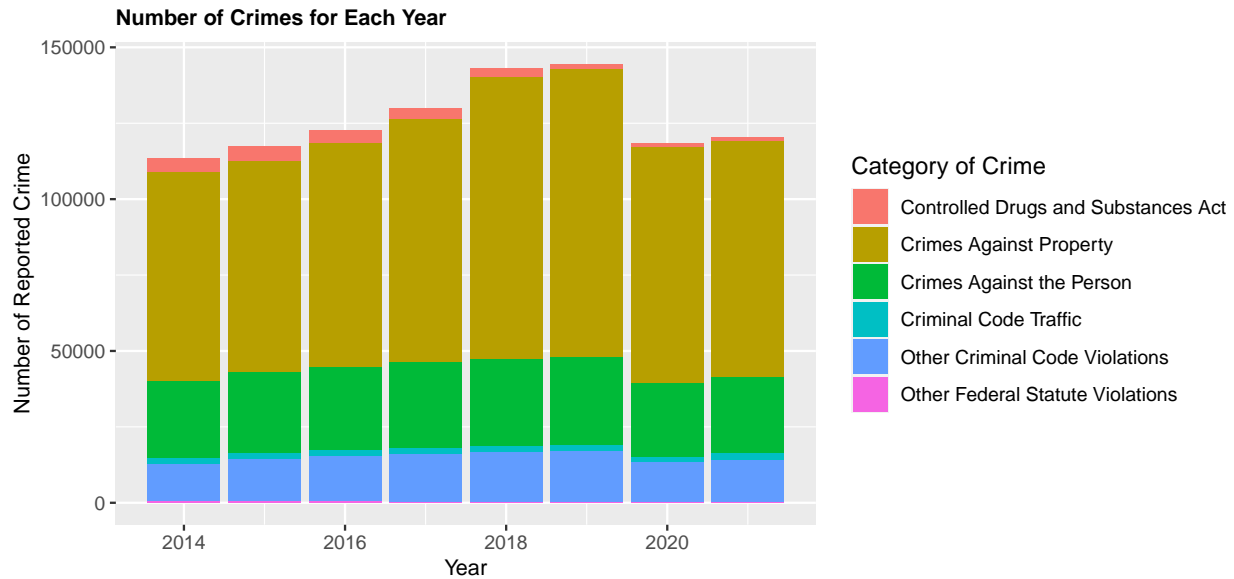
Figure 2: Number of Crimes Seperated by Categories from 2014 to 2021

Figure 2 provides us with a clear trend of each category of crime from 2014 to 2021. Interestingly, the total number of reported crimes has steadily gone up from 2014 to 2019, but it suddenly drop significantly after 2019. However, this does not mean that we should chalk up the significant difference due to statistical error. One of the most contributing factors towards the total number of crimes is "Crimes Against Property", and it is possible that the pandenmic caused a sudden increase in the crimes against property. Using the original data set, we can also see that the main subtype of "Crimes Against Property" is auto theft; this is possible due to the fact that it is the easist to commit and the pandenmic and lockdown mode created a perfect enviornment for this kind of crime to occur. When Toronto opened back up begining in 2020, all kinds of crimes generally decreased; this might be the possibility that a lot more police are going back to work and people having more awareness after lockdown for almost an entire year. Seeing that there are no glaring outliers or flaws in the number of crimes reported, we can move on to analyse specific divisons' efficiency. The four divisions (14, 51, 52, and 43) that I used for analysis all have a notable in crease in their police expenses, and they are relatively stable and do not fluctuate too much.
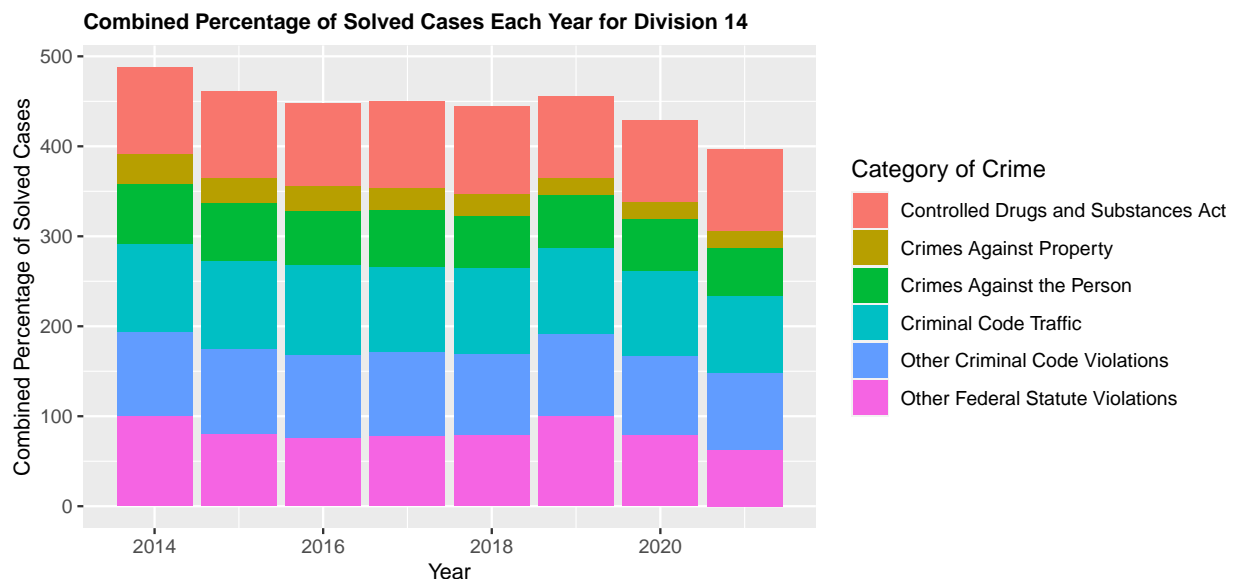

Figure 3: Division 14's total percentage of solved crime cases from 2014 to 2021
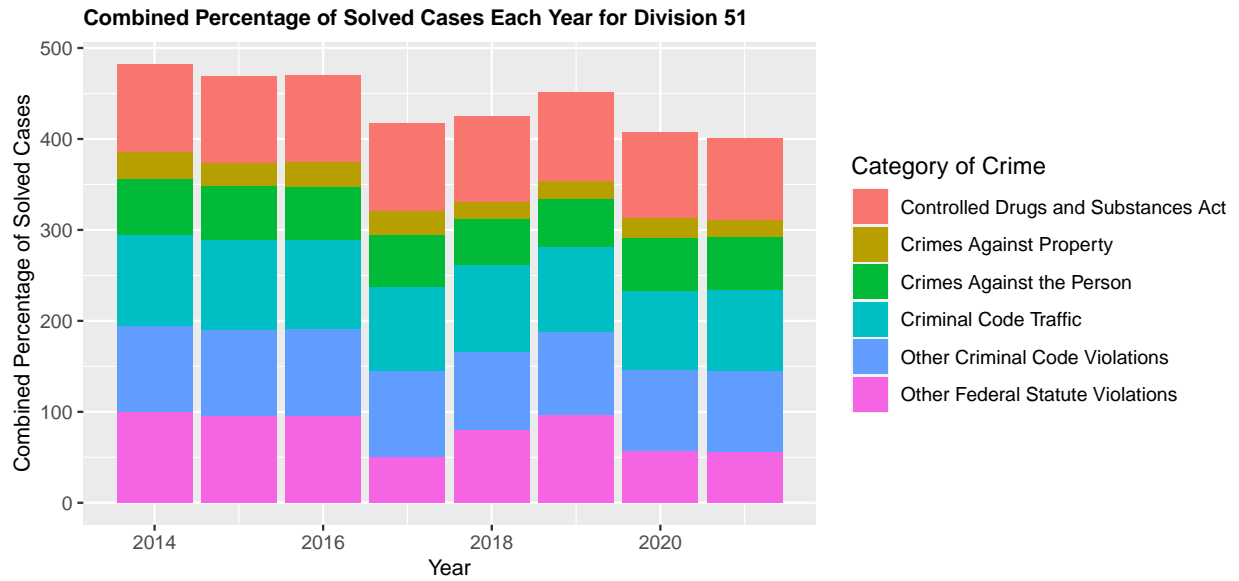
5

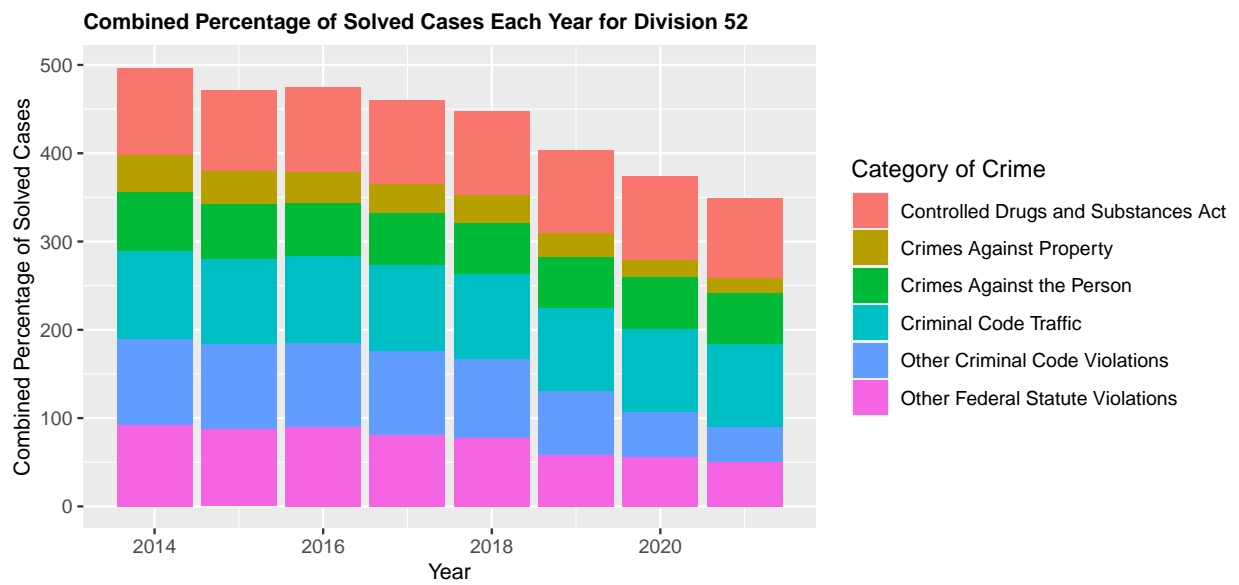Figure 4: Division 51's total percentage of solved crime cases from 2014 to 2021

Figure 5: Division 52's total percentage of solved crime cases from 2014 to 2021

Combined Percentage of Solved Cases Each Year for Division 43
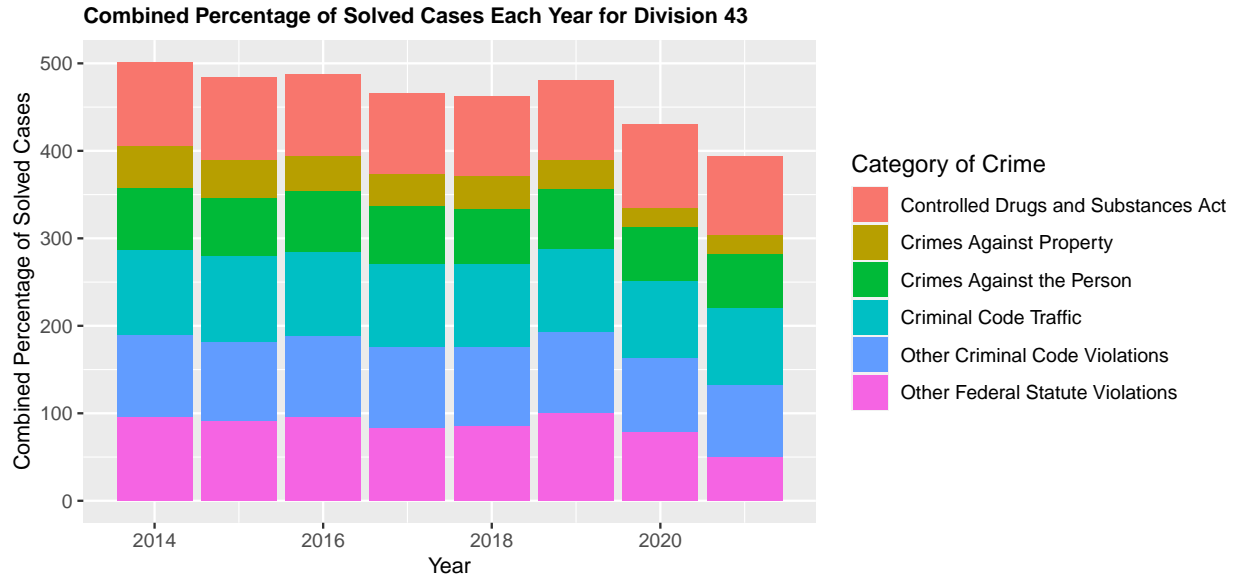

Figure 6: Division 43's total percentage of solved crime cases from 2014 to 2021

Figure 3 to Figure 6 above showcase the four divisions that I mentioned at the ending part of section 2.2. They all have increased their expenditure in recent years compared to 2014, and there is a general trend of steady increase in expenses. Unfortunately, the figures above underscore a general trend of decline in the percentage of solved crime. Those four police division samples are particularly inept at solving property related crimes, while those crimes are most prevalent in the total number of reported crimes. Furthermore, those four divisions all had an average of roughly 83% of solving all types of crimes in 2014, but that percentage has dropped significantly in 2021 to around 64%; that is a barely passing grade for the police divisions that continue to increase their expenses. This trend is particularly noticeable in Figure 5 above, which represents Division 52; it had an approximately 58% case solving rate across all categories in 2021, and it saw a staggering 25% decrease in its efficiency since 2014. Some of the figures above show a slight bump in percentage of solved crimes in 2019, but that can be negligible since the general trend is that the efficiency is decreasing.

3. Discussion

The decreasing trend of percentage of solved crimes is alarming, and the evidence clearly shows that more expenses does not mean a higher efficiency of solving crimes. Criminals do not suddenly get smarter within the span of eight years, and technologies do not suddenly make criminals harder to trace. It is reasonable to argue that police divisions do not allocate enough of the increase in their expenses to crime investigations. The lower number of total crimes in 2020 and 2021 (Figure 2) further highlights the incompetence of the Toronto Police Department. They somehow performed worse even with a lesser workload and higher expenses. Therefore, the proponents of the "Defund the police" movement do have a valid point when they argue that the police are eating up way too much public resources (Jones 2022). I suggest that the Toronto police divisions do a complete reorganization of their expenses; they should cut unnecessary costs and focus more resources on crime solving. The government should consider more strictly when divisions are proposing more budget increases, and the decision should closely tie with the specific division's performance. In addition, the abysmal solving rate for "Crimes Against Property" is shocking. The success rate never succeeded 50% for all divisions, which means that the police department failed at solving this type of crime more than half of the time. Even more shocking is the floor of some specific divisions performances. Division 33 only had a 8% solving rate for "Crimes Against Property" in 2021; out of 3842 cases, they have only closed 325 of them. To put it more in the perspective of regular people reading this paper, Division 33 has around the same amount of odds solving a regular property related case and you being accepted into an Ivy League school (Jotwani 2022). Jokes aside, with the analysis above, I sense under-powered performances regarding

property related crimes, and I think that there might be some underlying issues that haven't been solved yet. Judging from the already horrible rate in 2014, the success rate for this type of crime has almost never been high; Toronto Police Department should spend more resources on figuring out the reason behind the issue and solve it. In conclusion, my analysis has yielded fruitful results. While I can confidently say that the Toronto Police Department increasing their budget is relatively unjustified, I can not guarantee that putting the expenses in community initiatives will yield better results. On the other hand, by examining the reported crimes data, some of the success rates are alarmingly low, and I seriously hope the provincial government or the head of the police department would take notice and revamp their investigation procedures. Although I also can not promise a better result after revamping their system, more awareness of this pressing issue is always welcomed by the civilians of Toronto.

References

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.

Guze, John. 2022. *More Cops, Less Crime.* https://www.johnlocke.org/more-cops-less-crime-2/#:~:text=They%20found%20that%20%E2%80%9Can%20increase,crimes%20of%20auto%20theft%20and.

Jones, Ryan Patrick. 2022. *The Police Budget Is One of Toronto's Largest Expenses. Here's What You Need to Know about It.* https://www.cbc.ca/news/canada/toronto/toronto-police-spending-1.6623747.

Jotwani, Rohan. 2022. *Ivy League Acceptance Rates Vs. Other Schools.* https://www.quadeducationgroup.com/blog/ivy-league-acceptance-rates-vs-other-schools.

R Core Team. 2022. *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, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Dana Seidel. 2022. *Scales: Scale Functions for Visualization.*

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r.* https://yihui.org/knitr/.