Exploring Major Crime Indicators*

following the data relating to reported MCIs in 2014

Shreya Sakura Noskor

September 22, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

*Code and data are available at: ShelterStory

1 Introduction

2 Data

2.1 Source

First was loading the initial raw data from Open Data Portal provided by the city of Toronto(Gelfand 2022). This data set is titled "Major Crime Indicators". Data was cleaned and analysed in R(R Core Team 2023) by various helpful packages like, knitr(Xie 2014), leaflet(Cheng et al. 2024), tidyverse(Wickham et al. 2019), dplyr(Wickham et al. 2023), and lubricate(Grolemund and Wickham 2011) .

2.2 Variables and Measurement

The initial data set was large as it had a total of 27 variables that were recorded. However out of them all we chose to investigate 5 of them: report date, occurrence date, police divisions of Toronto, MCI category and HOOD_158(). The reason for this is because the goal of this report is to try and investigate if there are trends associated with the number of major incident cases reported and where they took place both in terms of which division and which neighborhood. We also see how long it took for a case to be reported after it has occurred. There are of course many other analysis that can be done with all 27 variables but that is outside the scope of this paper and it will quite frankly be too long of a paper as well.

First to explain some of the pre-existing variables. Major Crime Indicators consists of 5 categories: Assault, Auto Theft, Break and Enter, Robbery and Theft Over. And HOOD_158 represent the new 158 neighborhoods present in the city of Toronto. There are extra columns through out this data set that we have included as well. We first added a date difference colomn which shows how long after the day the reported incident happened, was the incident reported. This variable is worth studying as this tells a story about whether the reported incident was not filed as a Major Incident the date it occurred or it was never reported by the victim/witnesses. Another variable that we added was total MCI in each neighborhood and division. This was to see if there are some areas more prone to a specific type of MCI. And finally we also added a column for counting the total number of charges for each MCI category in total, regardless of their location. This was to see if there is a specific MCI that is commonly commited.

The measurement of this data set is through open data Toronto meaning that they likely reported the values that were given to them by the Toronto Police Department. As for how they got the data, police are required to submit a report for every case that they handle, and it is no different in this case. The likely scenario is that they reported all the written documents that they have to submit to file a report/investigation.

There are similar data sets that could have been explored likely with more accurate data as this data set only contains information from a decade ago. However, these are trends worth studying as a lot of important events have taken place in the year 2014 that could influence the results in the graph and hence we can learn from it an apply to the present or future.

3 Results

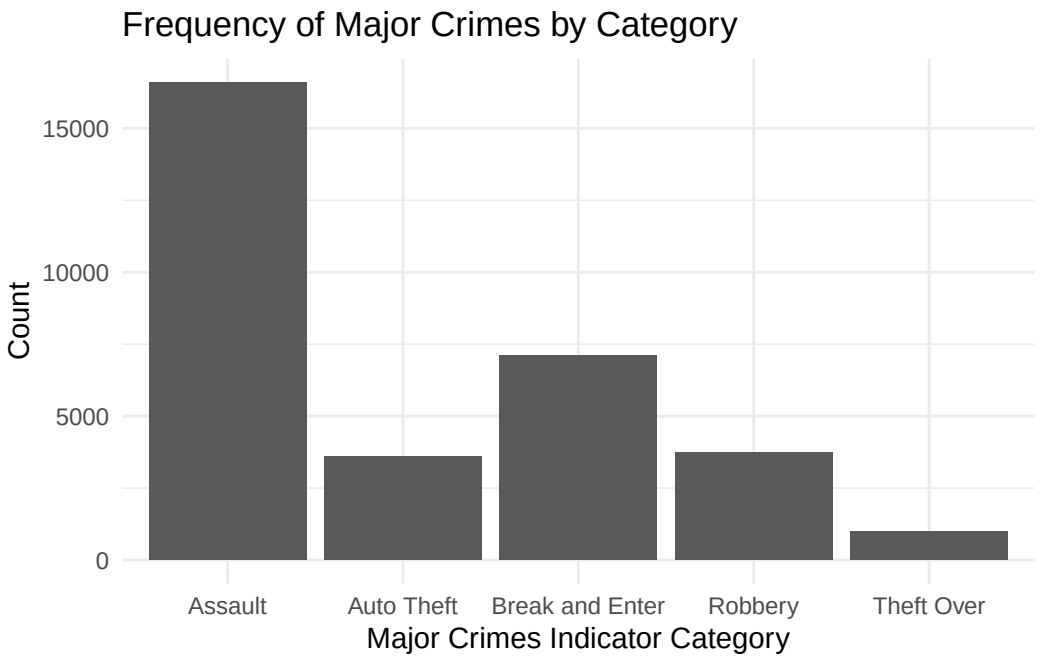


Figure 1: Examining the frequency of each MCI category: we see here that Assult is the most common to be reported.

Figure 1 shows the total number of reported incidents for each MCI category. It shows that the highest number is a little over 16000 (16601 to be exact) reported cases of Assault in the GTA in 2014. And the lowest is 988 reported cases of Theft Over (a certain amount of money). The graphs shows from the most to least reported MCI case is Assault, Break and Enter, Robbery, Auto Theft and then Theft Over.

Figure 2 shows a more detailed version of Figure 1, where we are able to once again see that Assault by far is the most reported out of all the MCI categories and Theft Over is the least. However we are able to see that while the number of reports for Theft Over is similar through out the months of 2014 (except for the little increase in November). That is not the case for Assault, as we see that the number of reported cases increase significantly in May and June and then decrease only to increase back in November. The rate of increase is also much higher

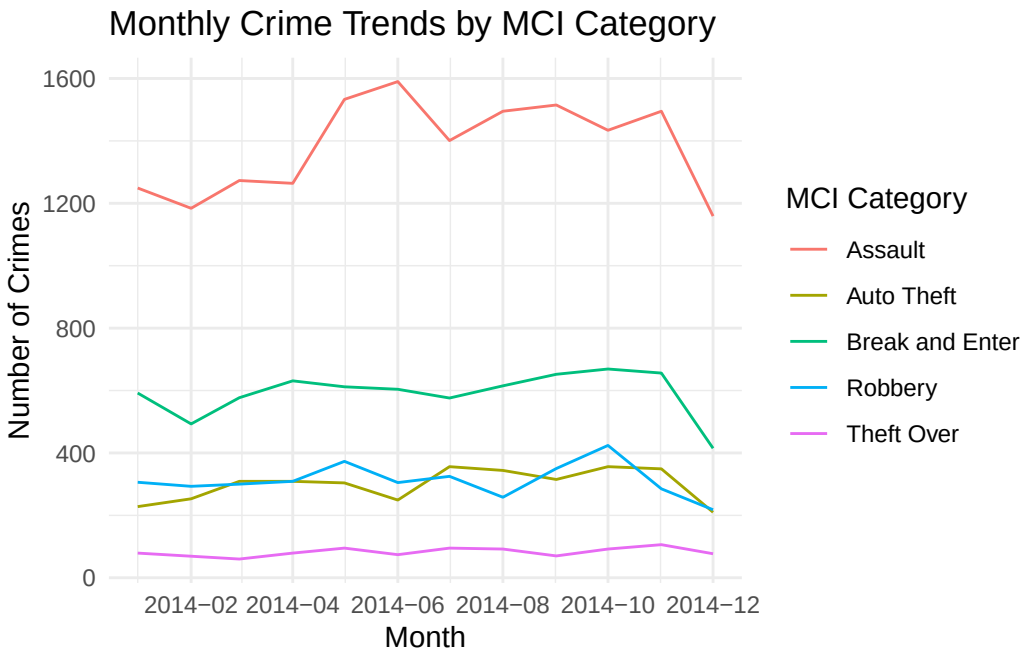
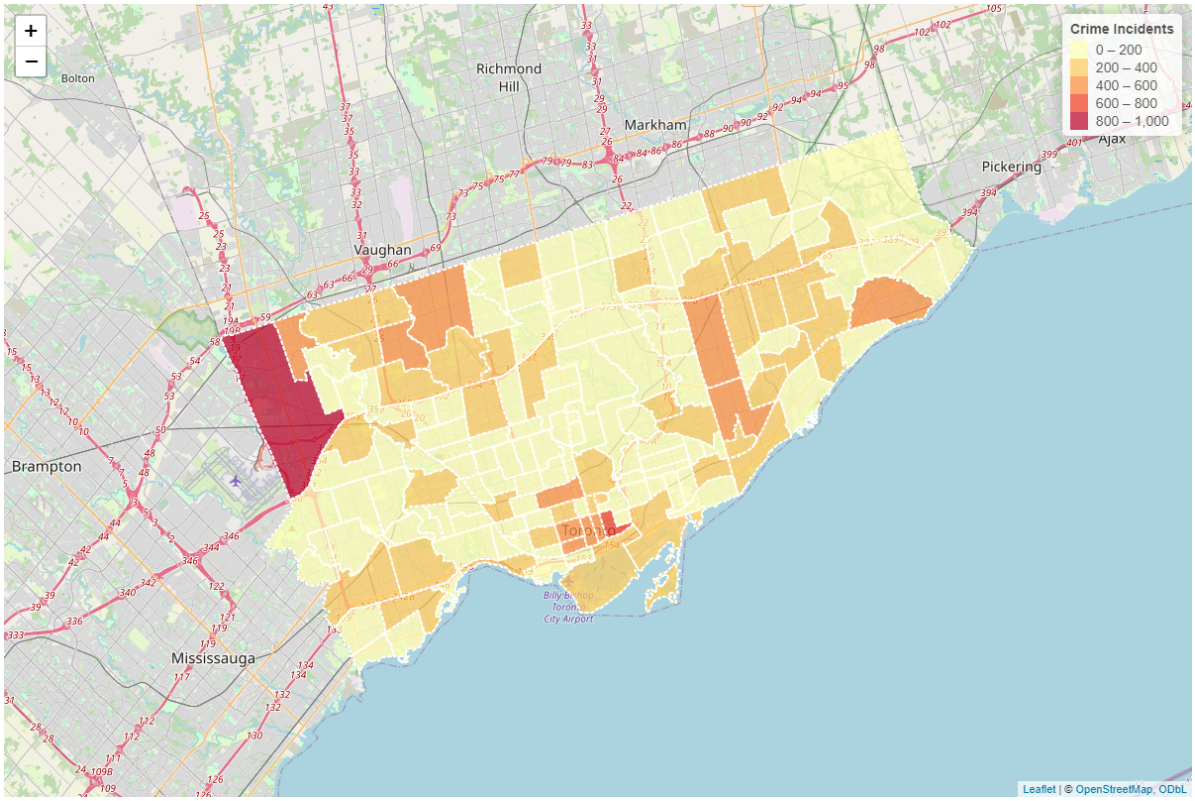
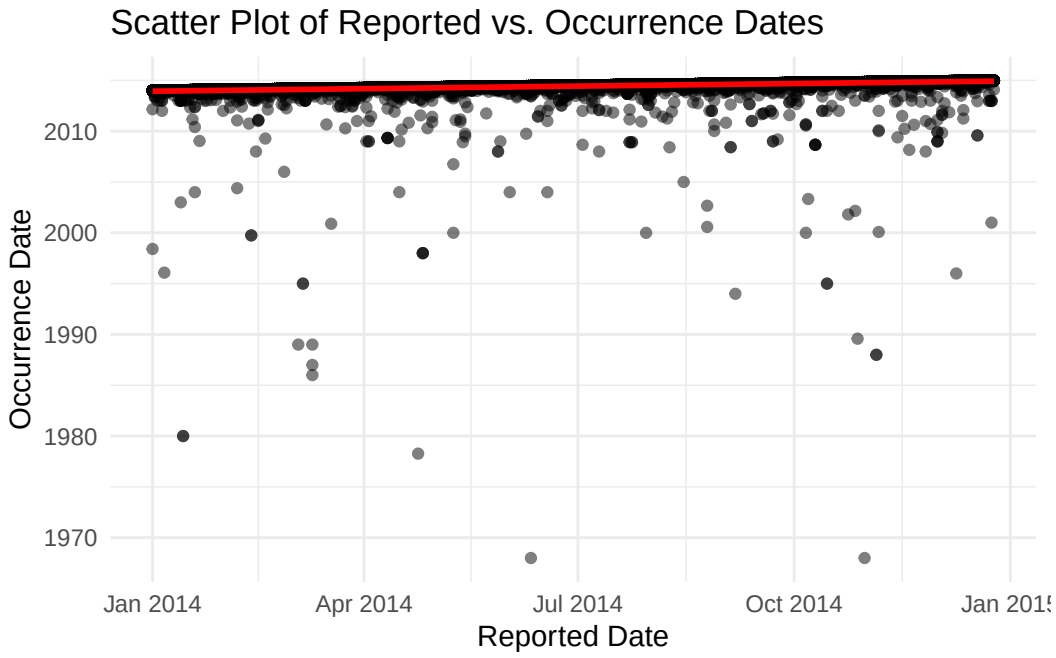


Figure 2: Examining the the number for each MCI reported against time.

for Assault then any other MCI category. Auto Theft and Robbery seem to have inverse connections where the months were Robbery increases Auto Theft decreases and then vice versa for other months. Break and Enter seems to have the same number of reported cases except for the 2 downward peaks in February and December.


Crime Incidents
0 – 200
200 – 400
400 – 600
600 – 800
800 – 1,000
Bolton
Richmond Hill
Markham
Ajax
Pickering
Vaughan
Brampton
Toronto
Mississauga
Leaflet | © OpenStreetMap, ODbL

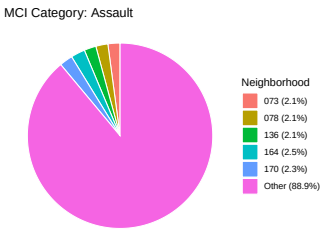
?@fig-map


Scatter Plot of Reported vs. Occurrence Dates
Occurrence Date
2010
2000
1990
1980
1970
Jan 2014
Apr 2014
Jul 2014
Oct 2014
Jan 201
Reported Date

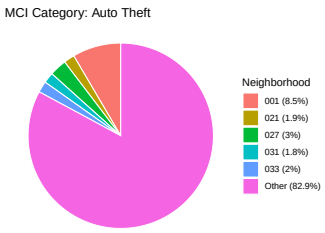
West Humber-Clairville (1) has one of the most and moss park following second.

Table 1: Incidents by Neighborhood and MCI Category

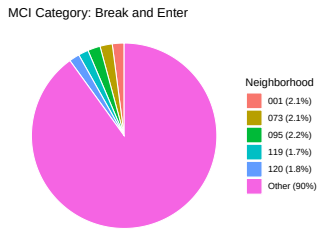
NEIGHBOURHOOD_158	Assault	Auto Theft	Break and Enter	Robbery	Theft Over	total
West Humber-Clairville (1)	286	303	146	82	41	858
Moss Park (73)	350	25	152	114	7	648
York University Heights (27)	274	107	105	60	29	575
Yonge-Bay Corridor (170)	382	12	66	72	37	569
Kensington-Chinatown (78)	350	27	101	72	17	567
Wellington Place (164)	409	21	77	26	19	552



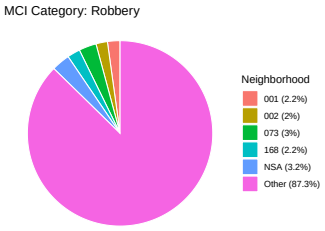
(a) Posterior prediction check



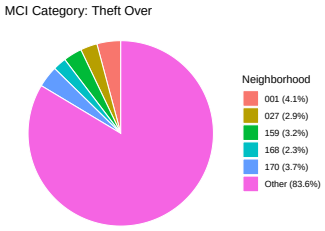
(b) Comparing the posterior with the prior



(c) 2

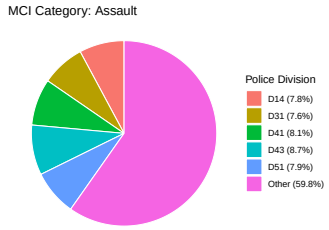


(d) 3

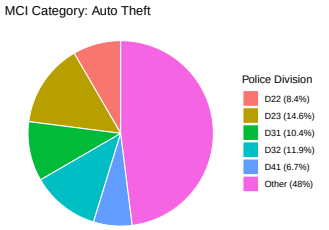


(e) 4

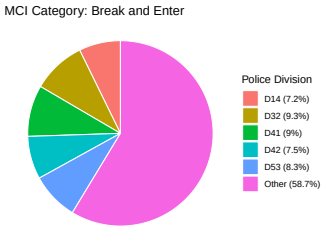
Figure 3: Examining how the model fits, and is affected by, the data



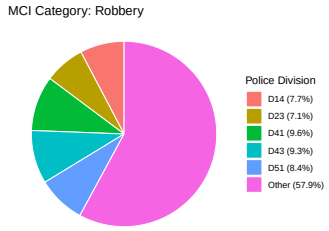
(a) Posterior prediction check



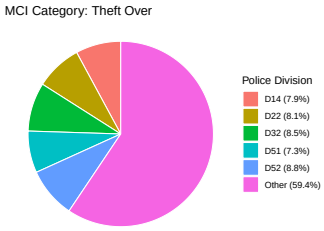
(b) Comparing the posterior with the prior



(c) 2



(d) 3



(e) 4

Figure 4: Examining how the model fits, and is affected by, the data

4 Discusion

4.1 Conclusion

4.2 weakness and limitation

This is a very old data but the point was to see the trends and learn from the history.

We think that the data set chosen for this paper is very strong as it is from a credible source and there were no NA values in the raw data file meaning that every data had a value for each attribute. This makes our data strong however, some of the limitations of this data set was the fact that this is a data set from 2014. It is a decade old data set so there is a possibility that the analysis doen here doesn't apply to our current date however, it is still good to take note of such trends. Another limitation could be that not all incidents get reported so, there is a strong chance that this data is very accurate depiction of all the Major Crime Indicators that happened in Toronto in 2014. There could be cases where the victims/witnesses didn't come forward or a incident was wrongly reported as MCI or not.

4.3 Real-world application

5 Appendix

References

Cheng, Joe, Barret Schloerke, Bhaskar Karambelkar, and Yihui Xie. 2024. *Leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. https://CRAN.R-project.org/package=leaflet.

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

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

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, 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. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

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