Exploring Major Crime Indicators*

following the data relating to reported MCIs in 2014

Shreya Sakura Noskor

September 26, 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), kableExtra (Zhu 2024) 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 that 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 had occurred. There are of course many other analyses 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 consist of 5 categories: Assault, Auto Theft, Break and Enter, Robbery, and Theft Over. And HOOD_158 represents the new 158 neighborhoods present in the city of Toronto. There are extra columns throughout this data set that we have included as well. We first added a date difference column 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 the total MCI in each neighborhood and division. This was to see if there are some areas more prone to a specific type of MCI. 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 committed.

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 report 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 took place in the year 2014 that could influence the results in the graph and hence we can learn from them and apply them to the present or future.

3 Results

3.1 Summary Analysis

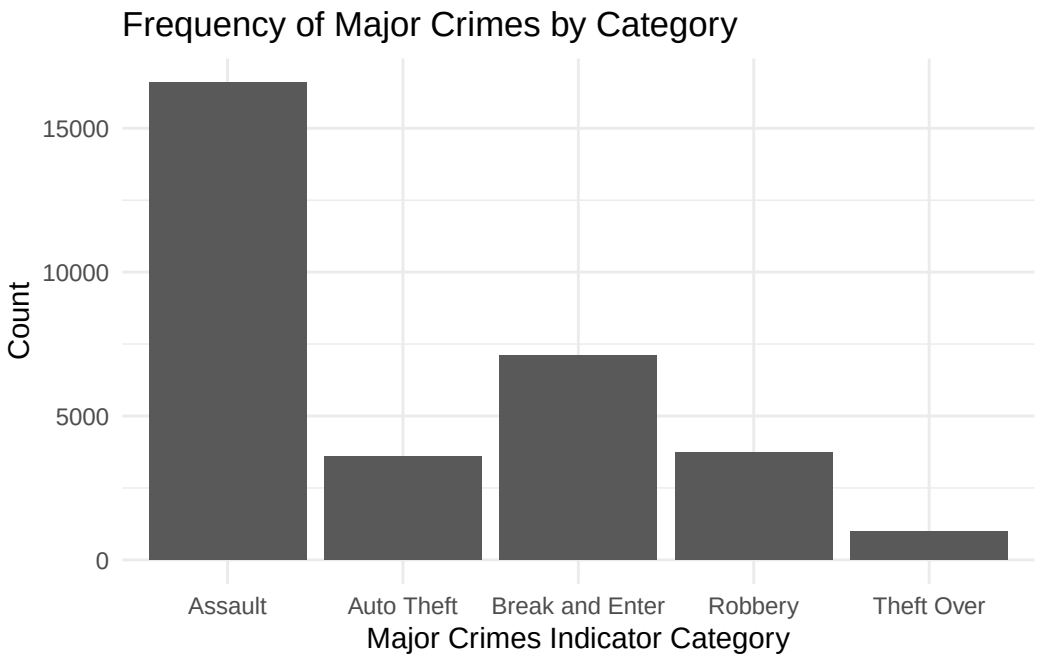


Figure 1: Examining the frequency of each MCI category: we see here that Assult is the most common to be reported.

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

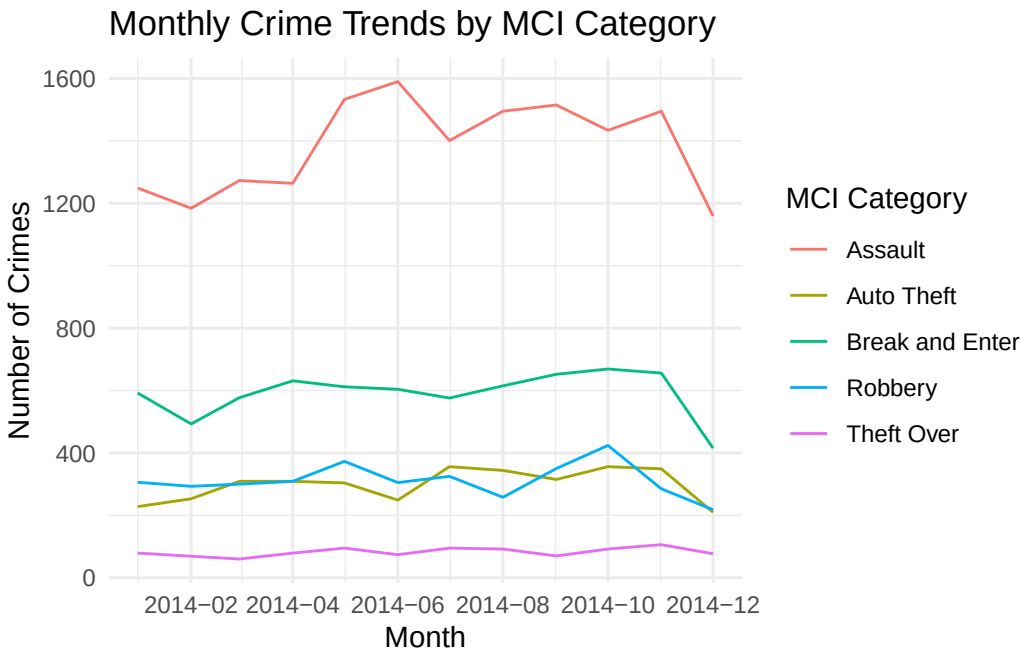


Figure 2: Examining the the number for each MCI reported against time.

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 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.

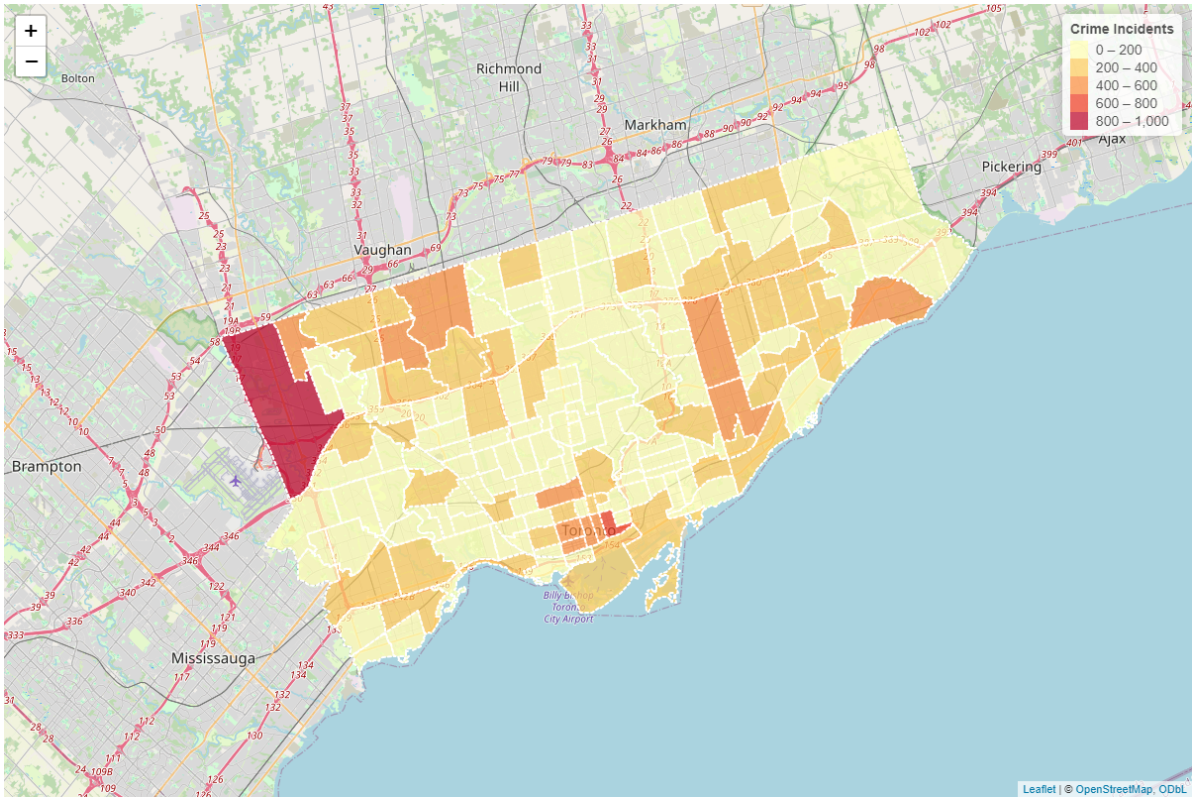
3.2 Neighbourhood Analysis

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

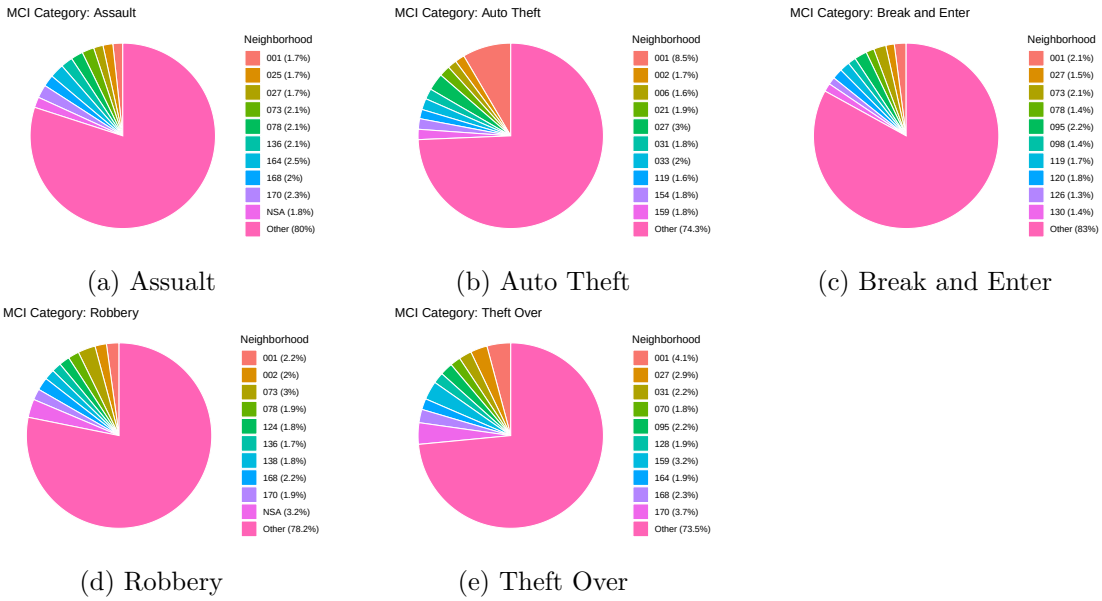
Figure 3: Table of the number of MCI cases reported in the neighbourhoods; break down by MCI category

From the above graphs we are only able to see trends but we can't tell anything about the actual numbers related to each neighbourhoods. Therefore we have Figure 3 to help with that. This table only shows the top 10 neighbourhoods with the highest count of charges related MCI. For a full table of 158 neighbourhoods go to Figure 7. From this table itself we see that West Humber-Clairville has the highest number of MCI charges with 858 charges. We also see that the most commited crime there was Auto Theft. This is however the only unique case. In the rest of the data, (both in this shortend version and the full version in the appendix), we see that the most crime that was commited was Assault, with the numbers being more than 50% of the cases there.



Now this picture gives a visual representaion to Figure 3. The sectors highlighted in dark red is where the most number fo MCIs are reported while the lighter it goes the less are reported. As seen in the table, West Humber-Clairville (in the top left corner) has the darkest shade along with Moss Park (in the middle section of the picture) has the highest number of MCI cases. An additional thing we can see from this graph is that majority of th eneighbourhoods in Toronto has a lighter shade meaning that the number of MCI cases is below 200. This includes majority of the west and mid towns of Toronto.

3.3 Visual Graph of Neighbourhood Data



(a) Assualt (b) Auto Theft (c) Break and Enter

(d) Robbery (e) Theft Over

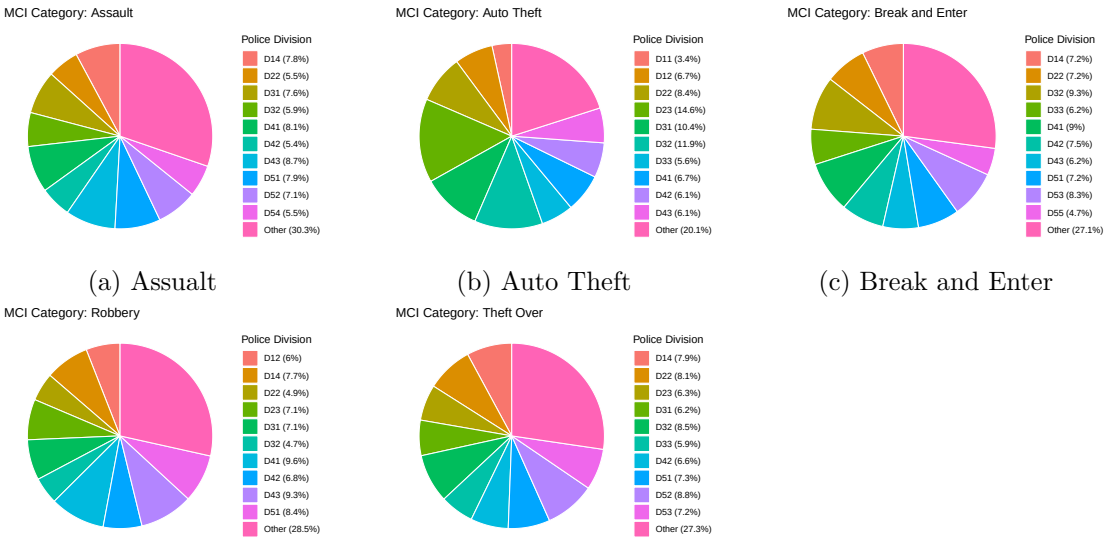
Figure 4: The top 10 neighbourhoods with the most MCI case and the rest grouped as other: This is to see if there is a obvious outliar

As per the caption Figure 4 looks at the top 10 neighbour hoods with the most cases in each MCI catagory. With all of the MCI catagory we see that neighbourhood 001 (Or Westville from the table) has the highest percentage of the graph meaning they have the highest number of cases. Take notice that the top 10 neighbourhoods occupy 20% to 25% of all the pie charts (or $\frac{1}{4}$ to $\frac{1}{5}$ of the pie chart).

As per the caption Figure 5 looks at the top 10 police divisions with the most cases in each MCI catagory. With MCI catagory Break & Enter and Robbery we see that the most perentage is made by Division 41 (9% and 9.6% respectivly). With Auto Theft we notice that Division 32 is the highest with 11.9%. Assualt has the highest percentage of 8.7% by Division 43. And lastly Theft Over MCI catagory has Division 52 holding 8.8% of the circle. Notice that unlike the first pie chart, the top 10 division holds anywhere from 60% to 80% of the chart (or $\frac{2}{3}$ to $\frac{4}{5}$ of the pie chart).

3.4 Occurence Date Vs Reported Dates

It is important to notice that the number of MCIs in each neighbourhood is not the only data that can be found from this data set so we look at the reported date vs the occurence date as well. In Figure 6 we notice that most of the scatters are in the top of the graph near y = 2014. This means that report date an dthe ocurence date is relativly close to each other.



(a) Assualt (b) Auto Theft (c) Break and Enter

(d) Robbery (e) Theft Over

Figure 5: The top 10 Police Division with the most MCI case and the rest grouped as other: This is to see if there is a obvious outliar

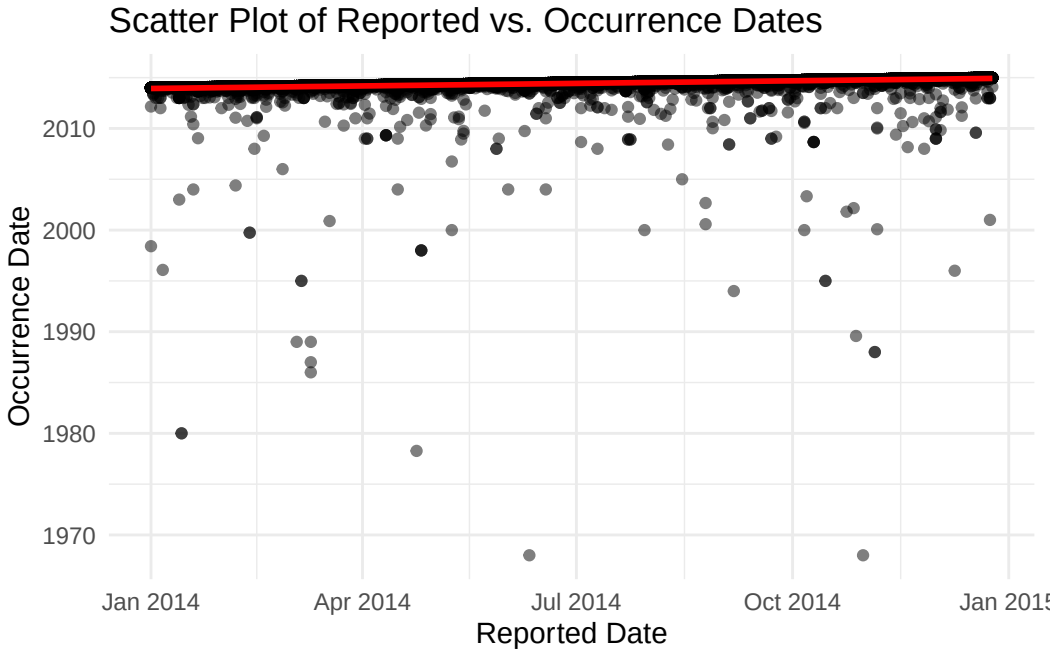


Figure 6: Scatter plot of Occurence date vs Reported Date

Looking at the cleaned data we see the same thing. However there is a handful of points is no where near the read line. This means that those points are where the occurence date is a lot earlier than the reported date. There are many MCI cases that have been reported more than 10 years after the occurence date, and even more that has been reported 5 years later.

4 Discusion

4.1 Conclusion

4.2 weakness and limitation

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

4.4 Future Research

5 Appendix

Table 2: Incidents by Neighborhood and MCI Category

NEIGHBOURHOOD_158	Assault	Auto Theft	Break and Enter	Robbery	Theft Over	total
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
Downtown Yonge East (168)	335	26	84	84	23	552
NSA	305	55	32	120	13	525
West Hill (136)	345	22	80	62	7	516
Annex (95)	232	13	155	51	22	473
Clairlea-Birchmount (120)	244	43	129	49	4	469
Wexford/Maryvale (119)	190	58	124	53	17	442
Glenfield-Jane Heights (25)	286	44	35	40	7	412
Mount Olive-Silverstone-Jamestown (2)	239	60	24	75	5	403
Oakdale-Beverley Heights (154)	217	66	55	41	10	389
South Riverdale (70)	206	19	85	55	18	383
Kennedy Park (124)	205	26	48	69	6	354
Church-Wellesley (167)	219	12	60	40	10	341
Black Creek (24)	215	42	26	49	9	341
Eglinton East (138)	188	18	40	68	8	322
Golfdale-Cedarbrae-Woburn (141)	197	22	55	41	3	318
Etobicoke City Centre (159)	140	64	55	20	32	311
Dorset Park (126)	132	49	91	33	6	311
East End-Danforth (62)	177	19	77	31	4	308
St Lawrence-East Bayfront-The Islands	177	20	56	32	16	301
Weston (113)	170	55	33	35	5	298
Oakridge (121)	167	21	60	47	2	297
Bendale-Glen Andrew (156)	133	35	71	52	6	297
Woburn North (142)	154	24	64	44	8	294
O'Connor-Parkview (54)	182	21	50	29	11	293
Don Valley Village (47)	109	47	80	46	9	291
North St.James Town (74)	152	23	71	31	7	284
Newtonbrook West (36)	139	49	52	27	7	274
University (79)	131	11	71	51	7	271
Scarborough Village (139)	161	20	49	30	3	263
Milliken (130)	71	38	102	37	15	263
Yorkdale-Glen Park (31)	104	66	47	23	22	262
South Parkdale (85)	181	9	45	20	5	260
Rockcliffe-Smythe (111)	133	45	36	38	7	259
Agincourt South-Malvern West (128)	103	33	73	26	19	254
Humber Summit (21)	88	68	54	26	14	250
Junction-Wallace Emerson (171)	162	21	35	26	5	249
Birchcliffe-Cliffside (122)	139	13	73	16	2	243
Trinity-Bellwoods (81)	116	11	67	27	12	233
Mimico-Queensway (160)	148	28	42	9	3	230
Stonegate-Queensway (16)	90	38	70	15	8	221
Flemingdon Park (44)	124	8	49	27	9	217
Malvern East (146)	138	17	29	24	6	214
Clanton Park (33)	72	70	60	7	5	214
Downsview (155)	111	35	30	29	4	209
Humbermede (22)	105	37	35	30	1	208
Kingsview Village-The Westway (6)	90	57	37	18	5	207
Rosedale-Moore Park (98)	70	9	99	19	7	204
Tam O'Shanter-Sullivan (118)	103	23	48	25	4	203
Bedford Park-Nortown (39)	56	40	83	13	9	201
Willowridge-Martingrove-Richview (7)	85	44	42	22	4	197
Greenwood-Coxwell (65)	112	16	48	16	4	196
Morningside Heights (144)	101	31	39	15	7	193
Morningside (135)	121	21	30	17	1	190
Taylor-Massey (61)	118	12	31	25	4	190
Palmerston-Little Italy (80)	87	15	55	26	5	188
West Queen West (162)	123	10	34	17	3	187
Islington (158)	86	27	51	19	2	185
Westminster-Branson (35)	81	29	60	12	3	185
Englemount-Lawrence (32)	79	29	42	30	4	184
Roncesvalles (86)	97	16	38	28	2	181
Harbourfront-CityPlace (165)	119	14	25	10	12	180
Cabbagetown-South St.James Town (71)	87	8	71	11	3	180
Fenside-Parkwoods (150)	106	19	45	5	4	179
Junction Area (90)	100	25	36	14	4	179
Oakwood Village (107)	100	18	25	33	2	178
Brookhaven-Amesbury (30)	81	39	23	30	5	178
St.Andrew-Windfields (40)	56	18	81	10	10	175
Cliffcrest (123)	80	19	42	29	4	174
Briar Hill-Belgravia (108)	79	18	42	25	9	173
Regent Park (72)	102	9	31	22	8	172
Victoria Village (43)	102	18	38	7	5	170
Mount Dennis (115)	95	26	18	28	3	170
Agincourt North (129)	67	17	46	36	1	167
Ionview (125)	102	5	33	24	1	165
High Park-Swansea (87)	79	15	45	19	7	165
West Rouge (143)	84	30	35	12	3	164
Bay-Cloverhill (169)	104	6	26	12	15	163
Danforth (66)	92	11	38	18	3	162
Bayview Village (52)	80	25	46	3	8	162

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.

Zhu, Hao. 2024. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.