

Crime Patterns Across Toronto Neighbourhoods

Daniel Biel, Dhruv Modi, Lawson Hung, Saket Sharma, Yasha Gupta

2025-11-21

Introduction

Crime Patterns Across Toronto Neighbourhoods uses the Neighborhood Crime Rates dataset published by the City of Toronto's Open Data Portal. The dataset contains 158 neighborhood-level observations and 203 variables covering a 10 year period from 2014 to 2024. Each neighborhood has annual crime counts (# unique) as well as rate values (per 100,000 population) for Assault, Auto Theft, Break and Enter, Robbery, Theft Over, Homicide and Shooting & Firearm Discharges.

Data was originally published on September 18, 2020 by Toronto Police Services, refreshed and refreshed annually. The dataset was last refreshed on June 30, 2025. The dataset includes:

- **Geographic Unit:** 158 Officially defined Toronto neighborhoods
- **Temporal Coverage:** 2014-2024
- **Measurement:** Crime rates per 100,000 residents
- **Method of Access:**
 1. Downloaded from Open Toronto and download the CSV file
 2. Imported into Rstudio with tidyverse
 3. Irrelevant variables removed (focused on Assault, Robbery, Auto Theft)

Variable	Description	Type	Recoding/Cleaning
AREA_NAME	Toronto Neighborhood name	Character	Trimmed whitespace; treated as categorical
ASSAULT_2014 ... ASSAULT_2024	Assault rate per 100,000 residents by year	Numeric	Converted year suffix to numeric for long-format analysis
ROBBERY_2014 ... ROBBERY_2024	Robbery rate per 100,000 residents by year	Numeric	Converted year suffix to numeric for long-format analysis
AUTOTHEFT_2014 ... AU-TOTHEFT_2024	Auto Theft rate per 100,000 residents by yaer	Numeric	Converted year suffix to numeric for long-format analysis

Transformations Introduced for Analysis 1. Reshaping Crime data into time-series format: Reorganized data so that each neighborhood has multiple rows; one row per year, instead of one column per year. This made it easier to plot changes across years and run regression models 2. Handling Missing Values: Any missing values in the crime rates were kept null rather than being filled in or estimated. Additionally, no zeroes were replaced or modified, a zero is treated as a real observed value.

1. How do assault, robbery, and auto theft rates vary across Toronto neighbourhoods in 2023?

To examine how assault, robbery, and auto theft rates vary across Toronto neighbourhoods in 2023, we created a subset of the dataset containing only the 2023 crime variables. This allowed us to focus specifically on neighbourhood-level variation for these three crime types. The 2023 data reveal substantial variation across neighbourhoods for all crimes studied.

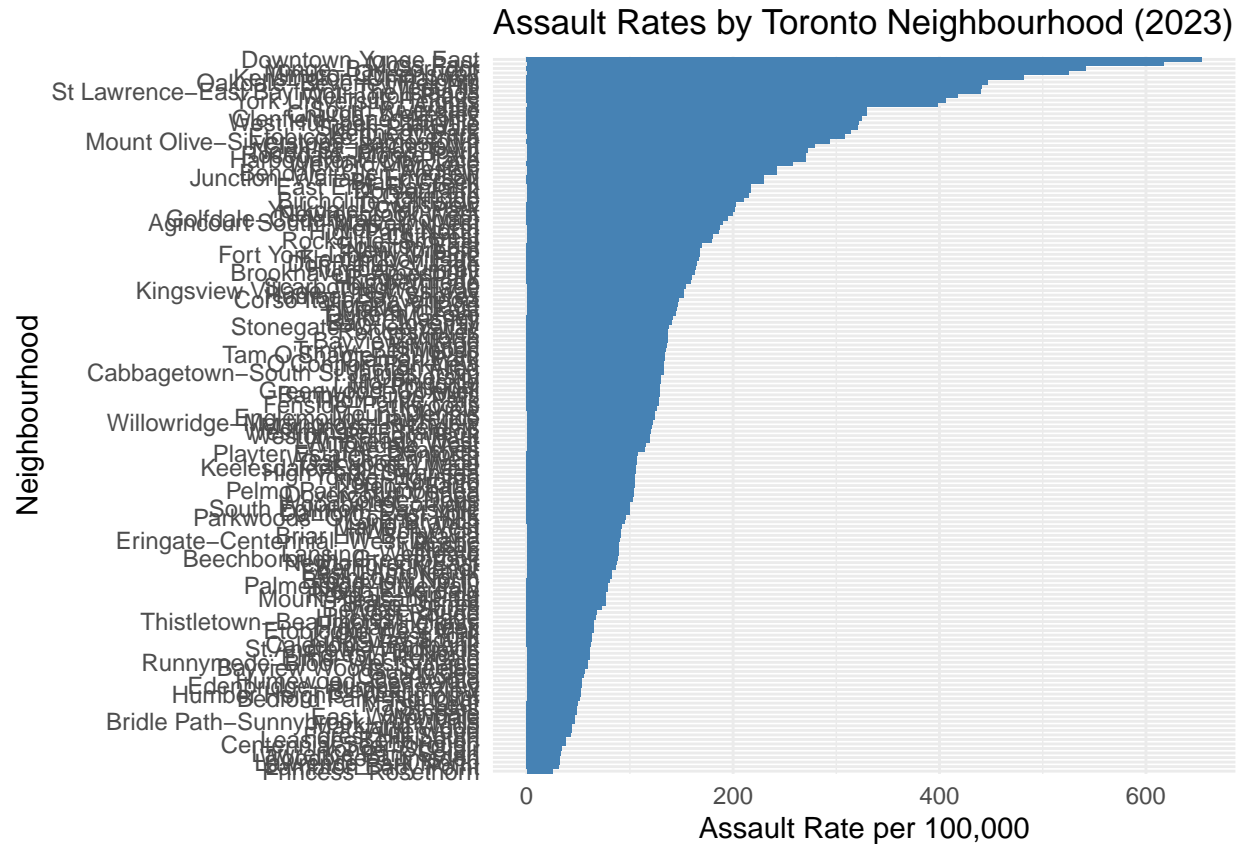


Figure 1: Assault Rate by Toronto Neighbourhood (2023)

Figure 1 shows the distribution of assault rates across Toronto’s 156 neighbourhoods. The bar chart ranks neighbourhoods from lowest to highest assault rates, clearly showing that some neighbourhoods experience much higher levels of assault than others. The long right tail of the graph indicates a small group of neighbourhoods with particularly high assault rates, while many others cluster near the lower and middle ranges.

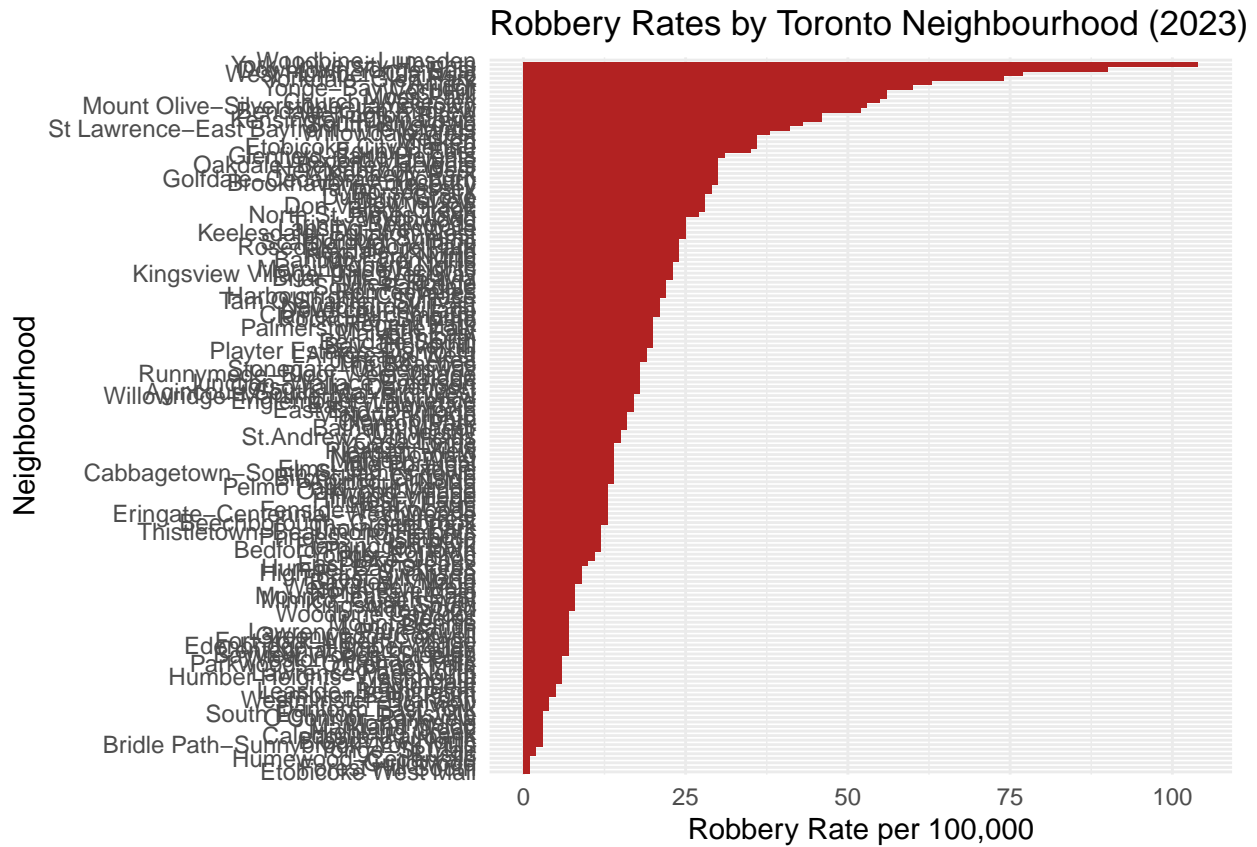


Figure 2 displays a similar ranking for robbery rates in 2023. Compared to assault, robbery rates are lower overall but still show noticeable variation between neighbourhoods. A few neighbourhoods stand out with distinctly higher robbery levels, indicating localized hotspots, while most neighbourhoods fall within a moderate or low range. The pattern suggests that robbery tends to be more concentrated geographically than assault.

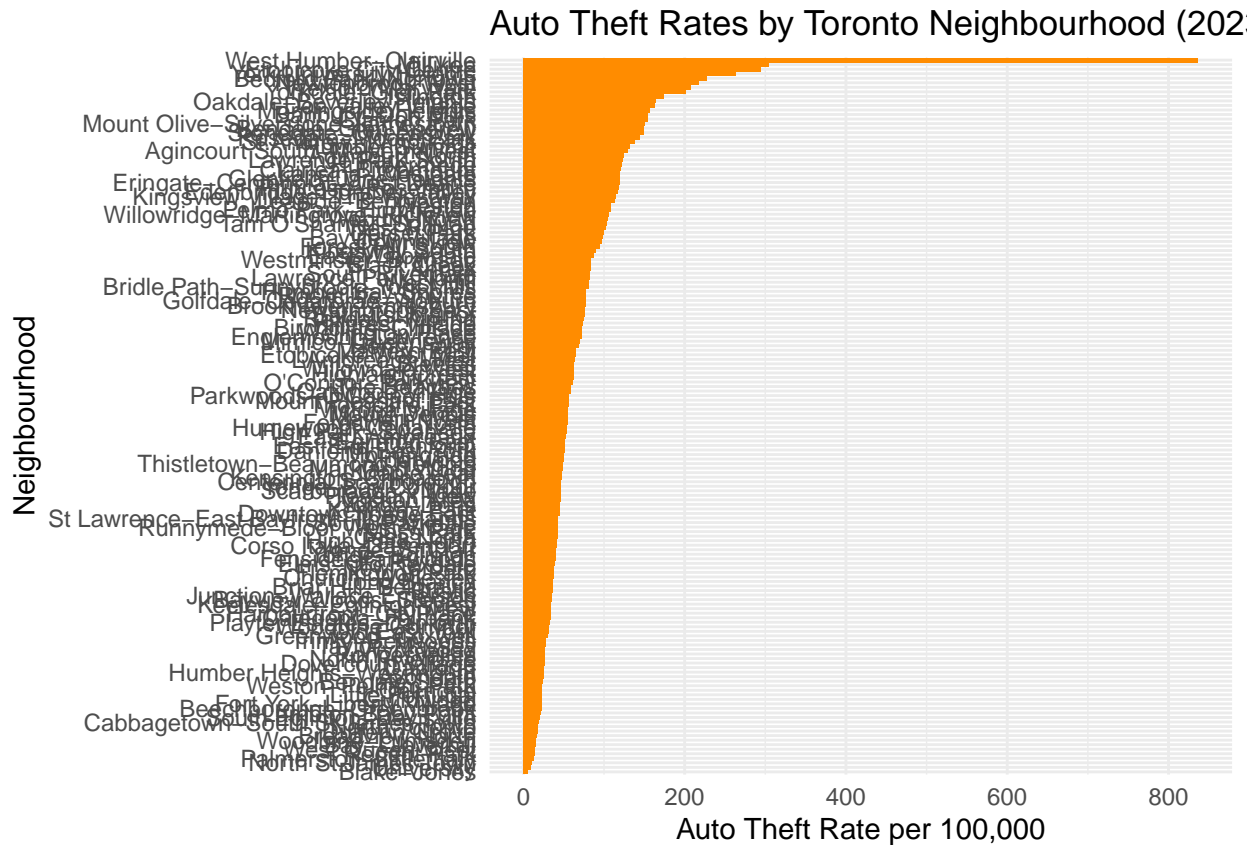


Figure 3 highlights auto theft rates by neighbourhood and shows the largest spread among the three crime types. Auto theft rates vary dramatically, with some neighbourhoods showing extremely high levels relative to the rest of the city. The long upper-tail of the distribution suggests that auto theft is heavily concentrated in certain areas. Compared to assault and robbery, auto theft exhibits the most uneven spatial distribution across Toronto.

To complement these graphs, we examined the summary statistics for each crime type, as shown in figure 4. Assault rates ranged widely, robbery rates were generally lower with a smaller spread, and auto theft rates showed the highest maximum values and the widest distribution. Together, the bar charts and summary statistics demonstrate that crime levels in Toronto are not evenly distributed. Some neighbourhoods consistently appear near the top across multiple crime categories, suggesting localized areas with higher criminal activity. Overall, 2023 data show clear differences in assault, robbery, and auto theft patterns, with each crime type having distinct levels of concentration and variability across neighbourhoods.

2. How have these crime types changed over time from 2014 to 2024?

To examine how crime has changed over time, we reshaped the dataset into long format and plotted annual trends for assault, robbery, and auto theft across Toronto neighbourhoods from 2014 to 2024.

```
## # A tibble: 6 x 3
##   AREA_NAME      Year AssaultRate
##   <chr>         <dbl>     <dbl>
## 1 South Eglinton-Davisville 2014         55
## 2 South Eglinton-Davisville 2015         56
## 3 South Eglinton-Davisville 2016         66
## 4 South Eglinton-Davisville 2017         73
## 5 South Eglinton-Davisville 2018         74
## 6 South Eglinton-Davisville 2019         62
```

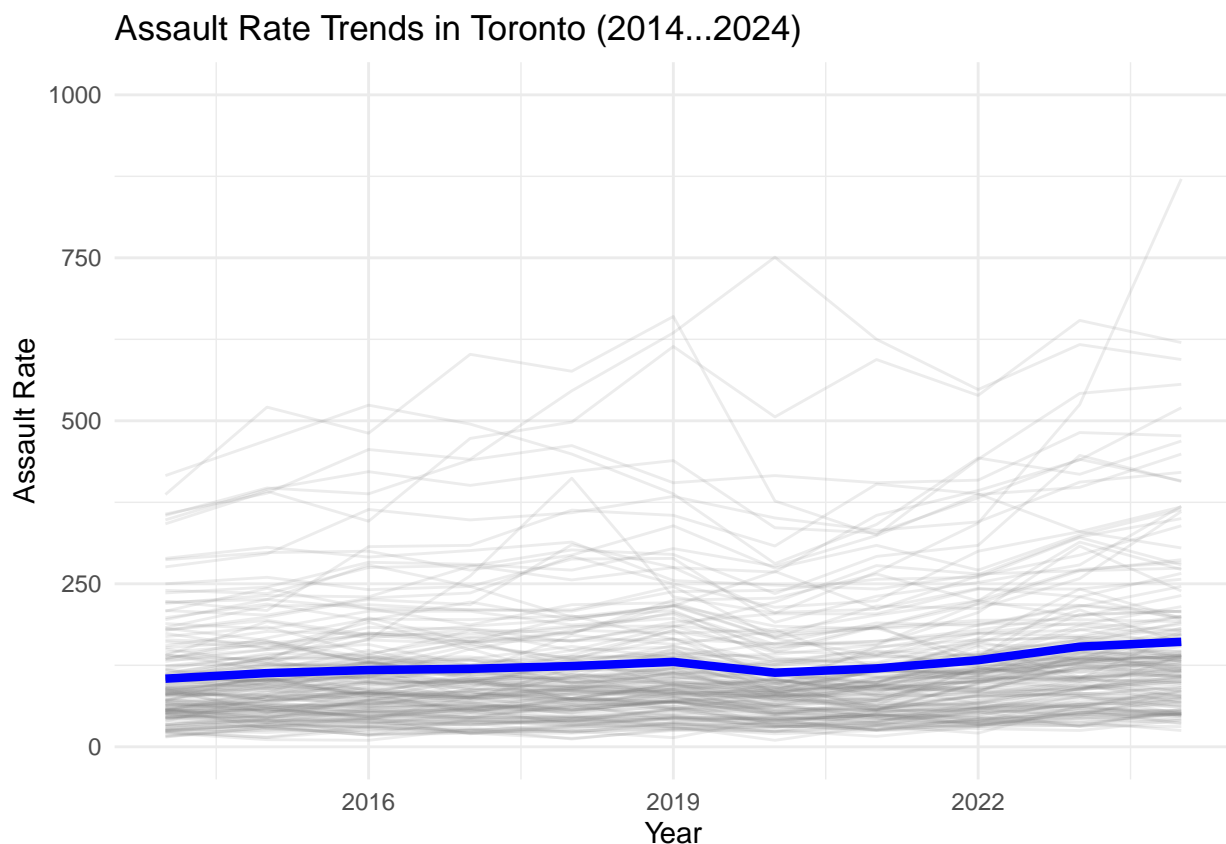


Figure 1 shows that assault rates have remained relatively stable over the ten-year period. Most neighbourhoods fall within a low-to-moderate range, and the city-wide average line is nearly flat, indicating no major long-term increase or decrease.

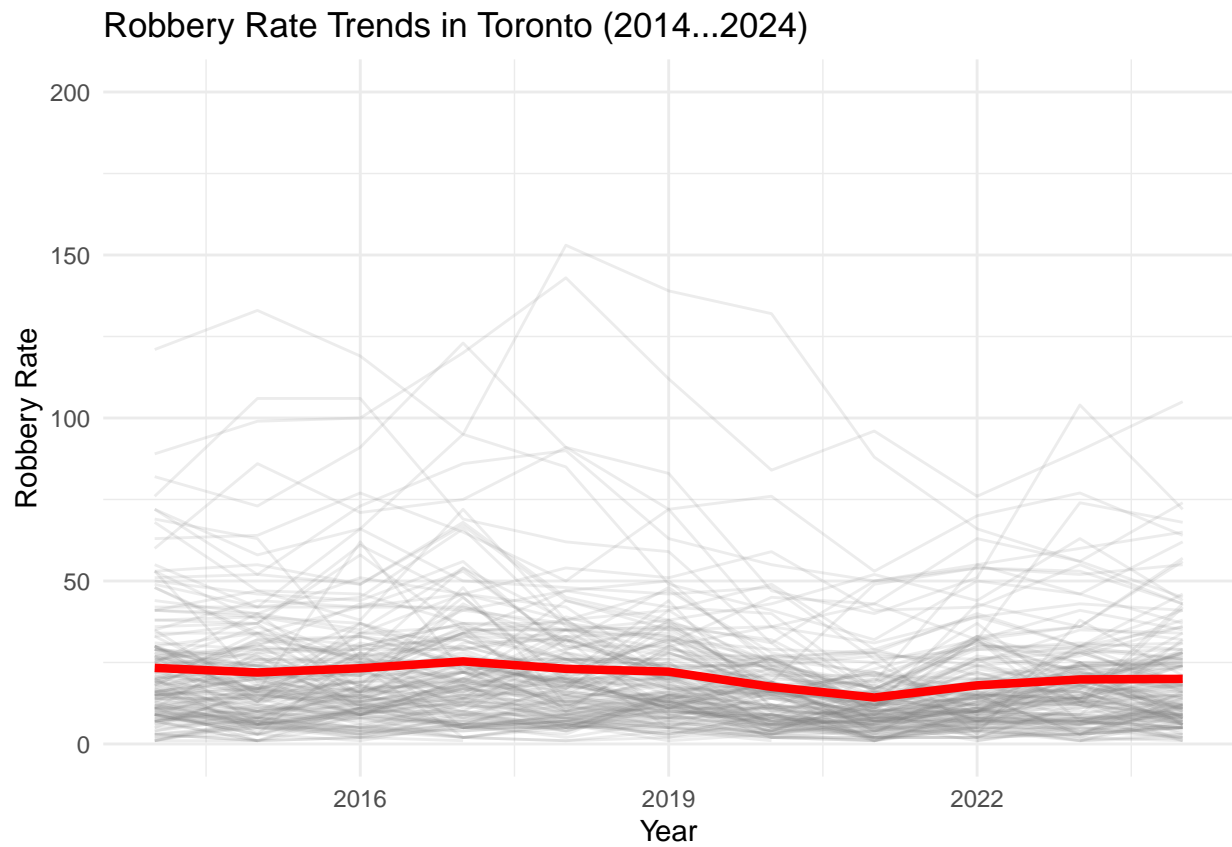


Figure 2 reveals that robbery rates are consistently lower than assault and show a gradual decline from 2014 to roughly 2020, followed by a slight rise after 2021. Most neighbourhoods remain below 50 robberies per 100,000, suggesting robbery became less common overall but may be increasing again in recent years.

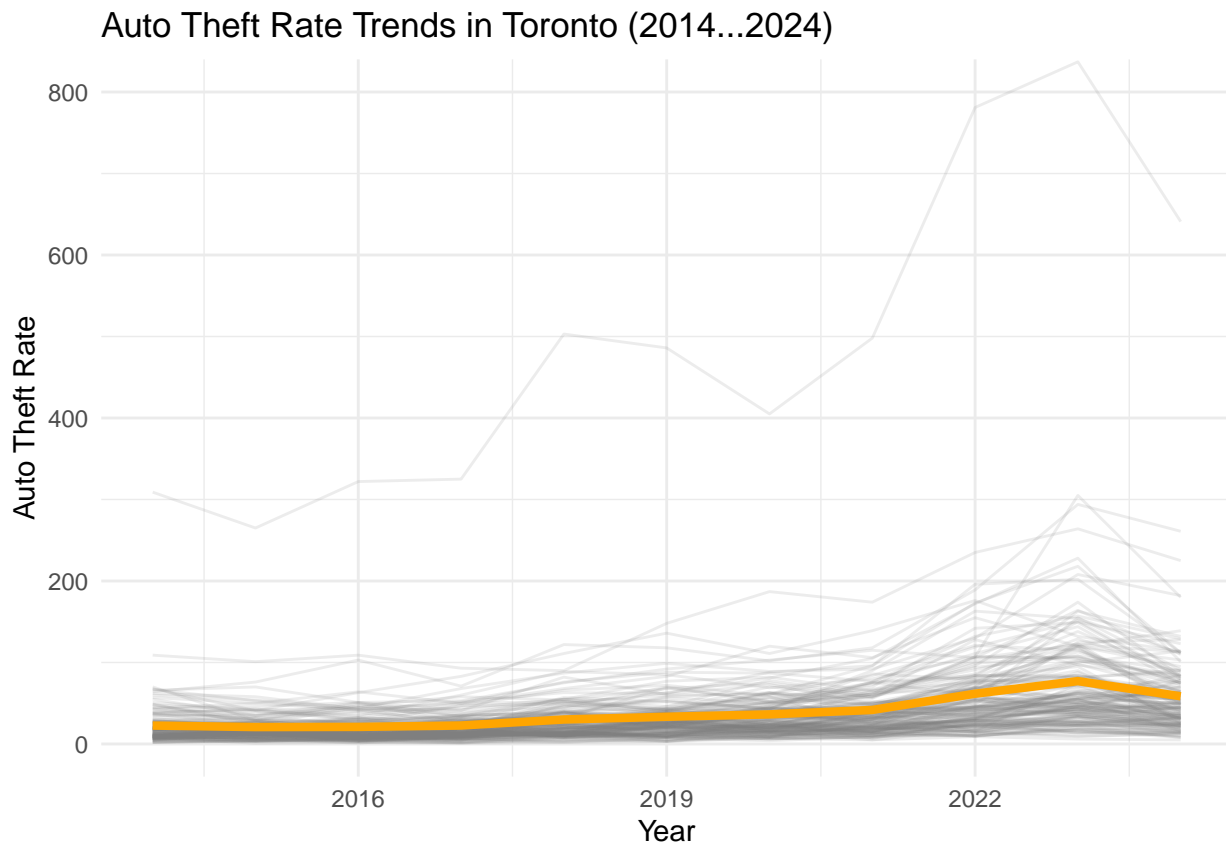


Figure 3 displays the strongest change among the three crimes. Auto theft rates begin rising noticeably around 2018, and the city-wide average increases sharply in the final years of the dataset. More neighbourhoods also show higher auto theft levels, indicating that auto theft has grown into a significant and expanding issue across Toronto. Collectively, the trends show that while assault has stayed stable and robbery has mildly declined, auto theft has increased substantially, especially after 2018, highlighting a major shift in crime patterns over the last decade.

3. Can we build a regression model to predict assault rates using time (year) or other crime types as predictors?

To investigate whether assault rates can be predicted using time or other crime types, we fitted two regression models. The first model examined whether assault rates changed over time by regressing assault rates on Year alone.

```
##
## Call:
## lm(formula = AssaultRate ~ RobberyRate + AutoTheftRate, data = combined)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2022.0  -192.8   -40.1   137.7  3713.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  296.866323   1.408502  210.77  <2e-16 ***
## RobberyRate    3.716887   0.008361  444.58  <2e-16 ***
## AutoTheftRate -0.193217   0.004234  -45.64  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 359.3 on 210935 degrees of freedom
## (1098 observations deleted due to missingness)
## Multiple R-squared:  0.484, Adjusted R-squared:  0.484
## F-statistic: 9.892e+04 on 2 and 210935 DF, p-value: < 2.2e-16
```

The output shows a small positive slope of approximately +4.37, indicating a slight upward trend over the 2014–2024 period. Although this trend is statistically significant due to the large number of observations, the model has a very low R^2 of about 0.018, meaning Year explains only 1.8% of the variation in assault rates. This demonstrates that time alone is a very weak predictor and that assault levels remain relatively stable over the decade.

```
##
## Call:
## lm(formula = AssaultRate ~ RobberyRate + AutoTheftRate, data = combined)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2022.0  -192.8   -40.1   137.7  3713.2
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  296.866323   1.408502  210.77  <2e-16 ***
## RobberyRate    3.716887   0.008361  444.58  <2e-16 ***
## AutoTheftRate -0.193217   0.004234  -45.64  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 359.3 on 210935 degrees of freedom
## (1098 observations deleted due to missingness)
## Multiple R-squared:  0.484, Adjusted R-squared:  0.484
## F-statistic: 9.892e+04 on 2 and 210935 DF, p-value: < 2.2e-16
```

The second model used robbery rates and auto theft rates as predictors of assault. The regression output

shows that robbery rate is strongly associated with assault rate: for every one-unit increase in robbery, assault increases by about 3.7 units, and this effect is highly significant. Auto theft shows a very small negative coefficient, which is statistically significant only because of the extremely large sample size and is not practically meaningful. This model has an R^2 of approximately 0.484, meaning robbery and auto theft together explain about 48% of the variation in assault rates—a much stronger model than the time-only model. The high F-statistic confirms that the crime-based model is highly significant overall.

Taken together, the results show that while assault rates do not meaningfully change over time, they are strongly linked to other violent crime levels in a neighbourhood. Robbery, in particular, is a strong predictor of assault, whereas auto theft adds very little explanatory power. Therefore, a regression model using other crime types, especially robbery, is much more effective at predicting assault rates than a model using time alone.

Conclusion

This project examined crime patterns across Toronto neighbourhoods using data from the City of Toronto’s Neighbourhood Crime Rates dataset (2014-2024). Across our three research questions, we identified clear spatial and temporal trends in assault, robbery, and auto theft, and evaluated the extent to which these crimes can be predicted using regression models.

Our first research question focused on variation across neighbourhoods in 2023. The results show substantial differences in crime levels across the city. Assault, robbery, and auto theft rates were not evenly distributed; rather, certain neighbourhoods consistently exhibited higher crime levels across multiple categories. Auto theft displayed the widest spread, with a long right-tail indicating that crime was heavily concentrated in a small number of neighbourhoods. These findings demonstrate that crime in Toronto is geographically uneven and that certain areas experience disproportionate levels of criminal activity.

The second research question examined trends from 2014 to 2024. Assault rates remained relatively stable over the ten-year period, with only minor fluctuations. Robbery rates showed a gradual decline from 2014 until approximately 2020, followed by a slight increase after 2021. Auto theft exhibited the most notable change, rising sharply beginning around 2018 and increasing more dramatically in the final years of the dataset. These results highlight that crime types do not follow the same long-term trajectory; some remain steady while others experience significant growth or decline.

The third research question investigated whether assault rates could be predicted using time (Year) or other crime types. The time-only regression model showed a small but statistically significant upward trend in assault rates; however, the model’s explanatory power was extremely weak, with an R^2 value of approximately 0.018. In contrast, the model incorporating robbery and auto theft rates explained nearly half of the variation in assault levels. Robbery emerged as a strong and meaningful predictor of assault, while auto theft contributed very little practical explanatory value despite statistical significance. These findings suggest that temporal changes do not meaningfully predict assault, but cross-crime relationships within neighbourhoods provide a much stronger basis for understanding variation in assault rates.

Limitations

Although the dataset is rich and detailed, several limitations of the dataset and methodology should be considered when interpreting these results.

- **Neighbourhood-level aggregation:** Crime data is aggregated at the neighbourhood level, which may mask variation within neighbourhoods and limit the precision of the analysis.
- **Lack of demographic or socioeconomic variables:** The dataset contains crime rates only and does not include factors such as income, population density, policing presence, or housing characteristics. Without these contextual variables, our models cannot fully explain why crime varies across space and time.
- **Missing or irregular values:** Although missing values were left as-is rather than imputed, this may slightly reduce model accuracy. Additionally, zero crime entries may represent either genuinely low crime or reporting/measurement gaps.
- **Observational constraints:** The dataset provides correlational, not causal, insights. We cannot determine whether robbery causes higher assault rates or whether both arise from shared underlying conditions. Annual data granularity: Crime is measured yearly. Monthly or seasonal trends cannot be explored with this dataset.

Future Directions

There are several promising opportunities to extend this work in future analyses. Incorporating demographic and socioeconomic variables, such as income, education levels, unemployment rates, or housing characteristics, would provide additional explanatory power and help clarify the structural factors contributing to crime patterns. Spatial analysis methods, including mapping crime clusters or measuring spatial autocorrelation, could offer deeper insight into the geographic dynamics of crime across Toronto.

More advanced modelling approaches, such as random forests or gradient boosting, may capture nonlinear relationships and interactions that linear models cannot detect. Access to more granular temporal data (monthly or seasonal) would allow for the analysis of short-term fluctuations, post-pandemic changes, or seasonality in crime patterns. Additional crime categories from the dataset could also be explored to determine whether similar patterns of clustering or co-occurrence exist among other types of offences.

Finally, connecting the dataset with external policy or policing information, such as community safety initiatives, neighbourhood revitalization programs, or shifts in police resource allocation, could help contextualize observed trends and provide more meaningful interpretations for real-world applications.

Appendix

```
crime_2023 <- crime %>%
  select(AREA_NAME, ASSAULT_2023, ROBBERY_2023, AUTOTHEFT_2023)
ggplot(crime_2023, aes(x = reorder(AREA_NAME, ASSAULT_2023), y = ASSAULT_2023)) +
  geom_col(fill = "steelblue") +
  coord_flip() +
  theme_minimal() +
  labs(
    title = "Assault Rates by Toronto Neighbourhood (2023)",
    x = "Neighbourhood",
    y = "Assault Rate per 100,000"
  )
```