A visualization of Toronto shooting data*

A look into how neighborhood and time affected shootings in Toronto from 2012 to 2021

Myra Li

3 February 2023

For Toronto residents, shootings are always one of the biggest concerns. After analysing data from 2012 to 2021, it is easy to see that the majority of shootings occur at evenings and nights. At the same time, neighbourhoods have a greater relationship with the probability of a shooting occurring. This paper attempts to visualise these findings.

Introduction

The recent spate of violence in Toronto TTC has led to more concerns about one's safety (https://toronto.ctvnews.ca/timeline-of-violent-incidents-on-the-ttc-1.6252945). Although most of the time, I walk to school. But with the frequency of violence, I have started to worry more about the safety of my environment. When walking in downtown Toronto, it is almost impossible for you not to encounter strange people. Sometimes they will suddenly yell and scream or look at you with a strange look. Worse still, they may carry out some hurtful incidents, such as violence, and shootings.

In this analysis, I have focused on the data from the shooting incidents in Toronto. This is described in more detail in the data section. In the analysis, I focused on the two factors of time and neighbourhood. Overall, we found a relationship between the number of shootings in Toronto and neighbourhoods between 2012 and 2021. Some neighbourhoods, such as Glenfield-Jane Heights (25), have a significantly higher number of shootings. Also, we find that shootings are usually more frequent at evenings and nights. In this paper, I will attempt to visualise these findings in order to present my findings visually. I will then discuss the implications of the findings, as well as some of the limitations of the findings.

*Code and data are available at: https://github.com/myraliym/INF312_Paper1_TorontoShooting_Final.git

Data

Data collection

All data used in this paper was retrieved from the City of Toronto Open Data Portal and can be retrieve from https://open.toronto.ca/dataset/shootings-firearm-discharges/. By using the R package opendatatoronto (Gelfand 2020), we downloaded the data in the R-script called download_and_clean. The data is uploaded and funded by the City of Toronto and it is refreshed annually. It contains all shooting-related occurrences reported to the Toronto Police Service. And it includes 14 columns, such as "_id","Event_Unique_ID""Occurrence_Date", "Occurrence_year", "Month", "Day_of_week", "Occurrence_Hour", "Time_Range", "Division", "Death", "Injuries", "Hood_ID", "Neighbourhood" and "geometry".

Since the last refreshed date is Oct 11, 2022, we decided to not to include 2022 as our data. Instead, we used 10 whole year data from 2012 to 2021 as our data.

Variables of interest

In cleaned_toronto_shooting, we picked up 6 columns we need from the dataset, and we renamed some columns to simplify. We renamed Occurrence_year to year and Time_Range to time. we only kept the variables of interest, they are "year", "time", "death", "injuries", "neighbourhood", "geometry". In addition to this, we have created a new variable, called "injury_level". This variable is used to indicate whether the event is an injury, death or both.

Data processing

We use some essential libraries to help us process the data. For example, opendatatoronto (Gelfand 2020), which is used to load data into R (R Core Team 2020). Lubridate (Grolemund and Wickham 2011) is used to help process date data. Also, we used knitr (Xie 2021),janitor (Firke 2021),tidyverse (Wickham et al. 2019) and dplyr (Wickham et al. 2021) to process other data and generate pdf. After the data was loaded, we saved it as a csv file named toronto_shootings.csv. Since we targeted our analysis to the decade from 2012 to 2021, we cleaned up the rest of the data. By filtering the years, we ended up cleaning up the data to keep only what happened during this decade. Considering the false alarms, we removed the cases of both injuries and death show NA from consideration in our analysis. Apart from that, we have replaced all the NA with zeros. Finally, we saved these cleaned data as a csv file called cleaned_toronto_shooting.csv.

Table 1: First ten rows of a dataset of cleaned shooting data

year	time	death	injuries	neighbourhood	geometry	injury_level
2012	Night	1	0	Clairlea-Birchmount (120)	c(-79.28272435, 43.72794544)	death
2012	Night	0	1	Clairlea-Birchmount (120)	c(-79.29142356, 43.72699591)	injuries
2012	Evening	0	1	Tam O'Shanter-Sullivan (118)	c(-79.29342081, 43.7821167)	injuries
2012	Evening	0	1	Church-Yonge Corridor (75)	c(-79.37827718, 43.65626242)	injuries
2012	Night	0	1	Regent Park (72)	c(-79.35864726, 43.66241529)	injuries
2012	Night	0	1	Weston-Pellam Park (91)	c(-79.45997962, 43.67419535)	injuries
2012	Evening	0	3	West Humber-Clairville (1)	c(-79.56497836, 43.69773424)	injuries
2012	Evening	0	1	Kennedy Park (124)	c(-79.26066651, 43.73366576)	injuries
2012	Evening	0	2	Willowridge-Martingrove-Richview (7)	c(-79.55757503, 43.69207108)	injuries
2012	Evening	0	1	Mount Olive-Silverstone-Jamestown (2)	c(-79.58637766, 43.73250887)	injuries

Now, let's have a quick look at some of the cleaned data we will be working with. They are illustrated above in table (Table 1).

Visualizing the Data and The Implications

Visual 1 Toronto shooting data by year

What interests me in this analysis is the number of Toronto SHOTS per year. In this section, we show the number of reported shooting injuries in Toronto over a ten-year period from 2012 to 2021, which includes the number of all injuries and deaths. In the following table (Table 2), we can see the number of shooting injuries per year, as well as the specific number of people at different injury levels (injuries/death).

Table 2: Number of shootings in Toronto from 2012 to 2021

Year	Total shootings	Injuries	Death
2012	148	114	34
2013	141	119	22
2014	103	76	27
2015	152	126	26
2016	193	152	41
2017	187	148	39
2018	229	178	51
2019	284	240	44
2020	217	178	39
2021	209	163	46

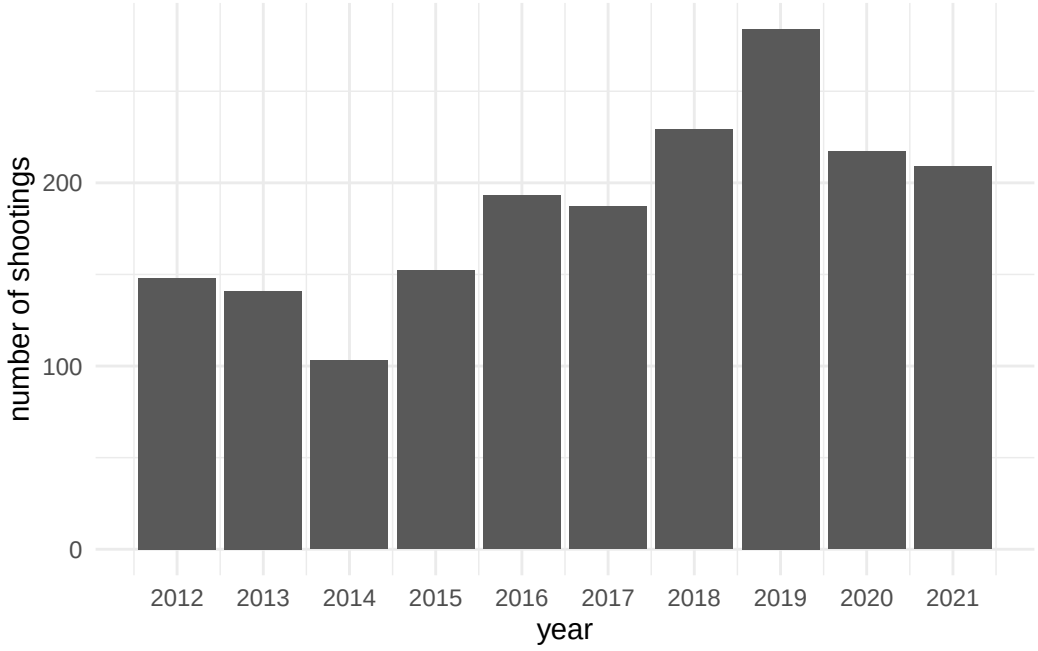

Figure 1: Shooting in Toronto from 2012 to 2021

Also, we have generated graphs of the data. In the chart (Figure 1), we can see that between 2012 and 2021, the most shootings occurred in 2019. the number of shootings in 2019 was more than twice the number in 2014. Since 2019, the number of Toronto shootings are again on a gradual downward trend.

Implication 1 Toronto shooting data by year

I think we need to take this downward trend with a watching skepticism. the decline in 2020 and 2021 may be largely due to the quarantine policy introduced after COVID-19, so people have less opportunity to go out. Since, the number of shootings per year is influenced by many factors, we do not intend to pick a specific year for analysis. Next, we will look more closely at the time when shootings occurred.

Visual 2 Shootings data by time

Moving on, lets see what implications we can draw when we filled the date by time.

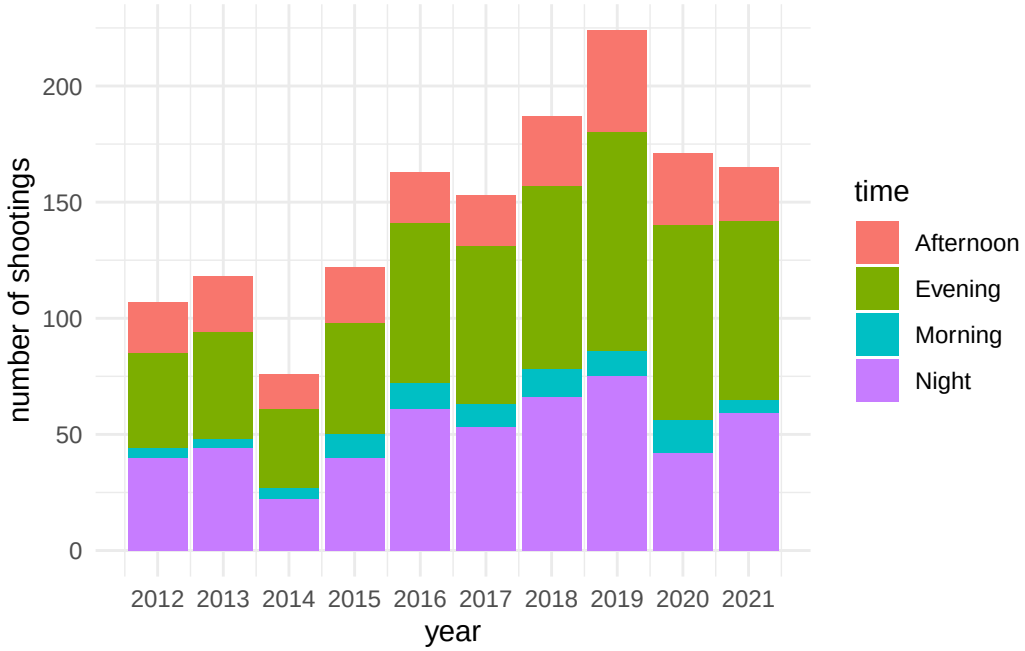

Figure 2: Shootings in Toronto at different times from 2019 to 2021

In the chart (Figure 2), we can see that between 2012 and 2021, shootings occur mostly at evenings and nights, with very few occurring at mornings, and the highest number of shootings occurring at evenings.

Implication 2 Shootings data by time

Although shootings are usually irregular and unpredictable, I think from this result we can protect ourselves by avoiding going out at night as much as possible, thus reducing the prob-

ability of encountering shootings. Next, we will analyze different neighbourhoods in more details.

Visual 3 Shootings data by neighbourhood

We have used neighbourhood to classify the data to find out whether neighbourhood is related to shooting. Since the data covers 140 neighbourhoods, we decided to focus our analysis on the top 10 most frequent neighbourhoods where shootings occur. These neighbourhoods are Glenfield-Jane Heights (25), Mount Olive- Silverstone-Jamestown (2), Black Creek (24), Waterfront Communities-The Island (77), Yorkdale-Glen Park (31), Downsview-Roding-CFB (26), West Humber-Clairville (1), West Hill (136), Malvern (132) and York University Heights (27). More detailed values are in the table (Table 3) below.

Table 3: TOP10 neighbourhood where shooting occurs most often from 2019 to 2021

Neighbourhood	Total shootings	Injuries	Death
Glenfield-Jane Heights (25)	85	73	12
Mount Olive-Silverstone-Jamestown (2)	69	55	14
Black Creek (24)	62	52	10
Waterfront Communities-The Island (77)	55	43	12
Yorkdale-Glen Park (31)	54	43	11
Downsview-Roding-CFB (26)	53	44	9
West Humber-Clairville (1)	50	41	9
West Hill (136)	46	41	5
Malvern (132)	44	33	11
York University Heights (27)	41	33	8

We then decided to conduct a closer analysis of these ten communities where shootings most often occur. We analyzed when shootings occurred in each neighborhood and presented them in the graphs (Figure 3, Figure 4, Figure 5, Figure 6, Figure 7, Figure 8, Figure 9, Figure 10, Figure 11, Figure 12) below.

We can find that except for Waterfront Communities-The Island (77), Yorkdale-Glen Park (31) and West Humber-Clairville (1), the highest number of people injured time range is evening. In York University Heights (27), the number of people who were injured during the evening and night hours was equal. We also confirm that the number of people shot and injured during the morning hours is the lowest.

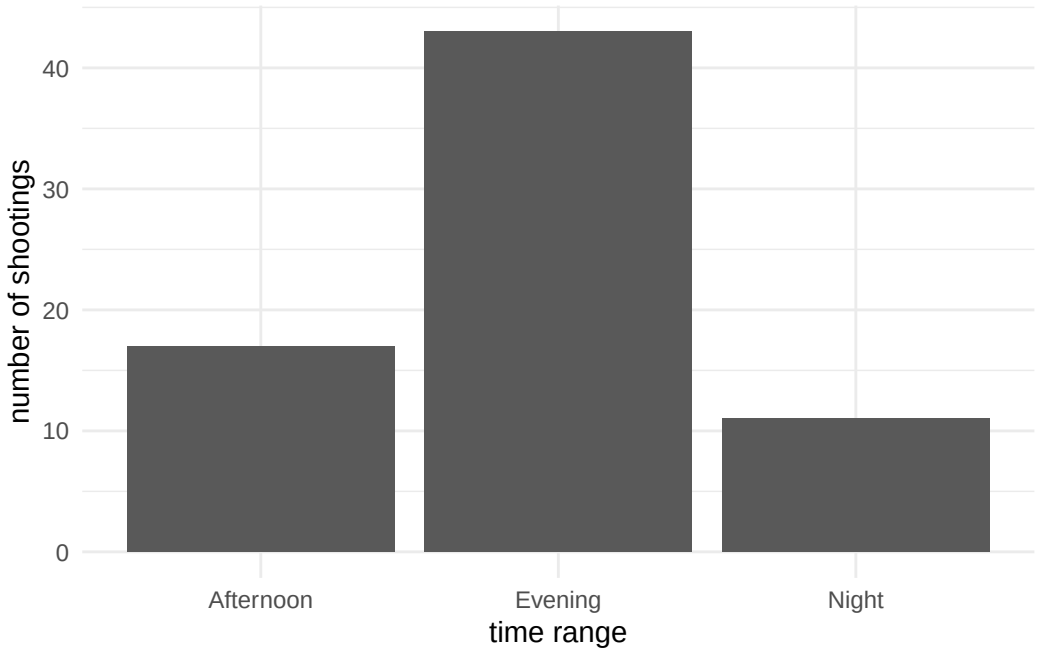

Figure 3: Shootings in Glenfield-Jane Heights (25) at different times from 2012 to 2021

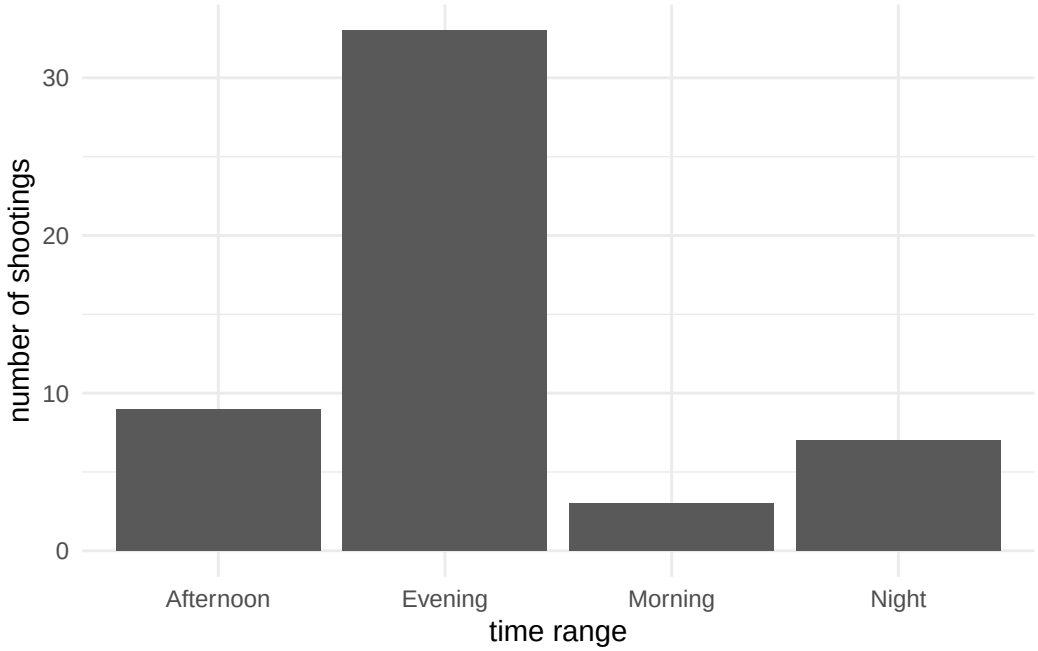

Figure 4: Shootings in Mount Olive-Silverstone-Jamestown (2) at different times from 2012 to 2021

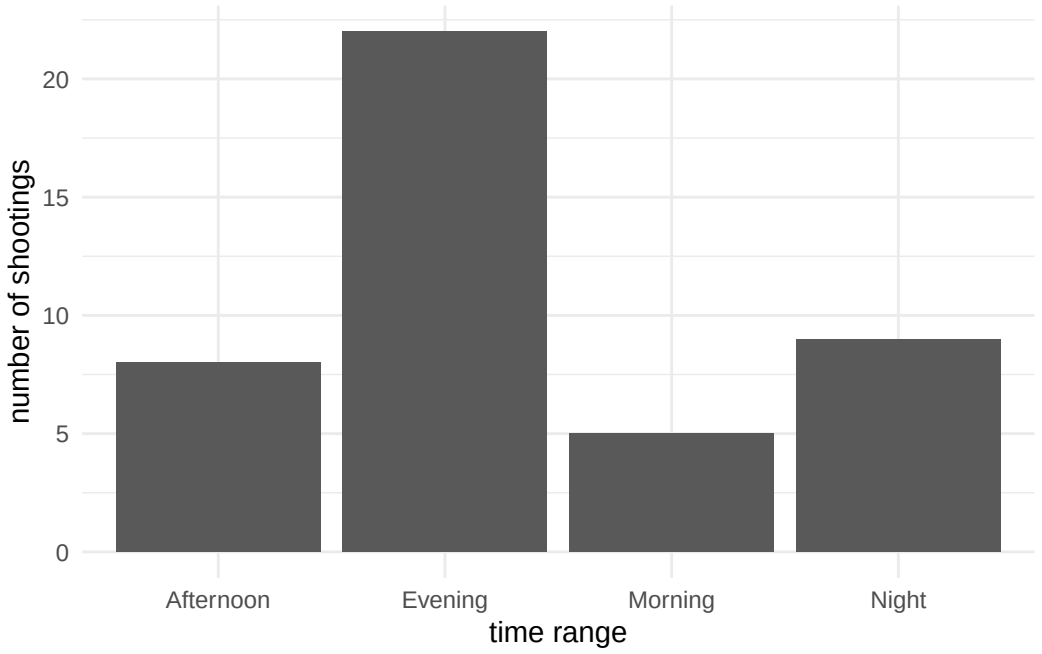

Figure 5: Shootings in Black Creek (24) at different times from 2012 to 2021

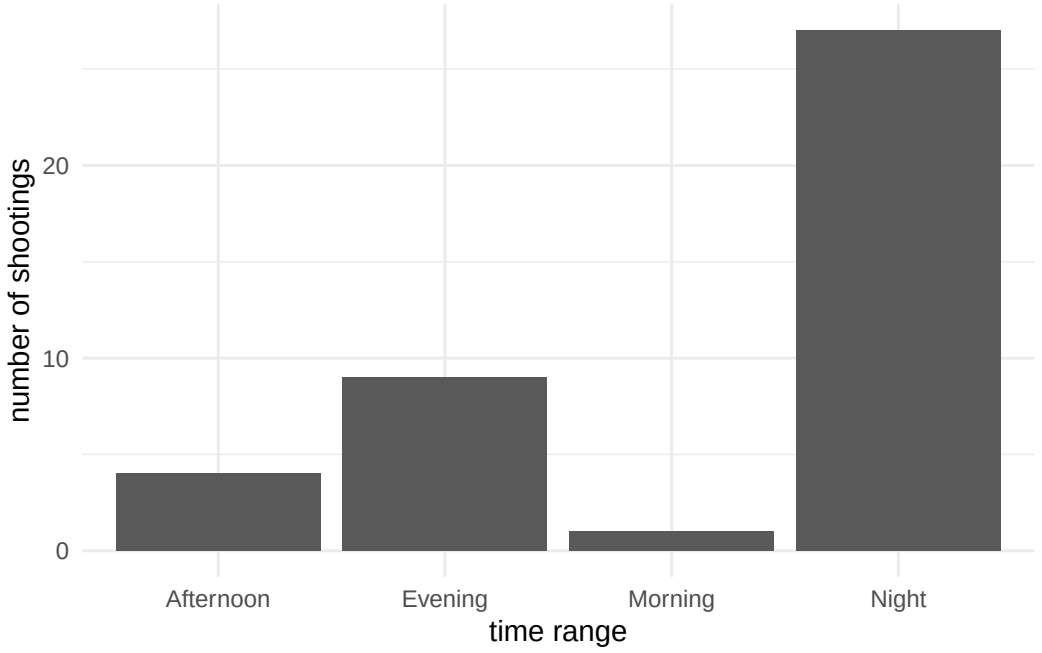

Figure 6: Shootings in Waterfront Communities-The Island (77) at different times from 2012 to 2021

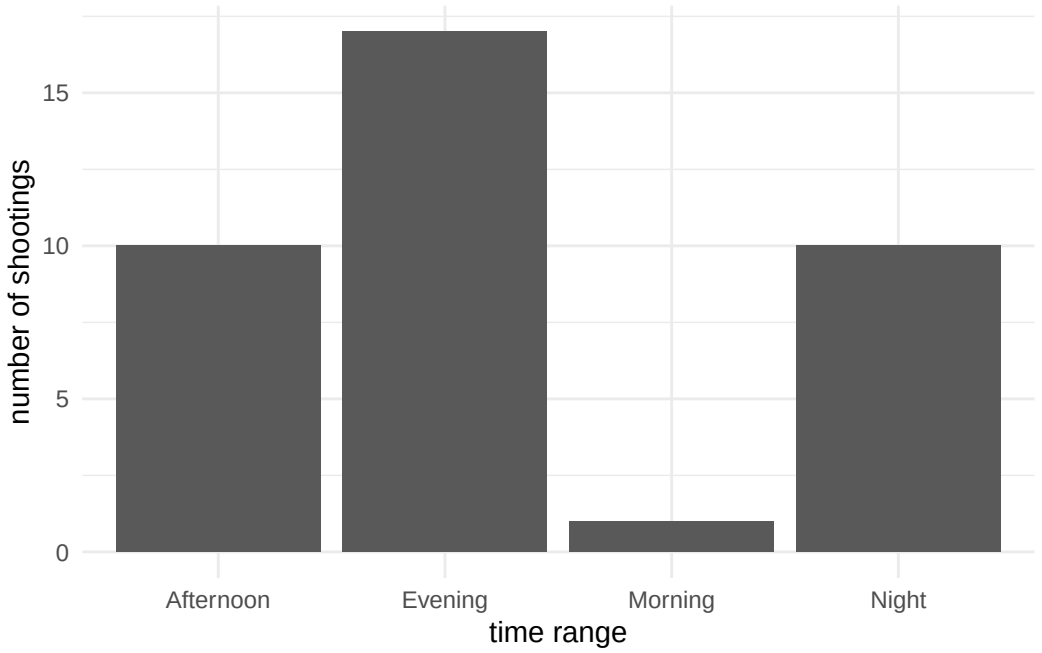

Figure 7: Shootings in Yorkdale-Glen Park (31) at different times from 2012 to 2021

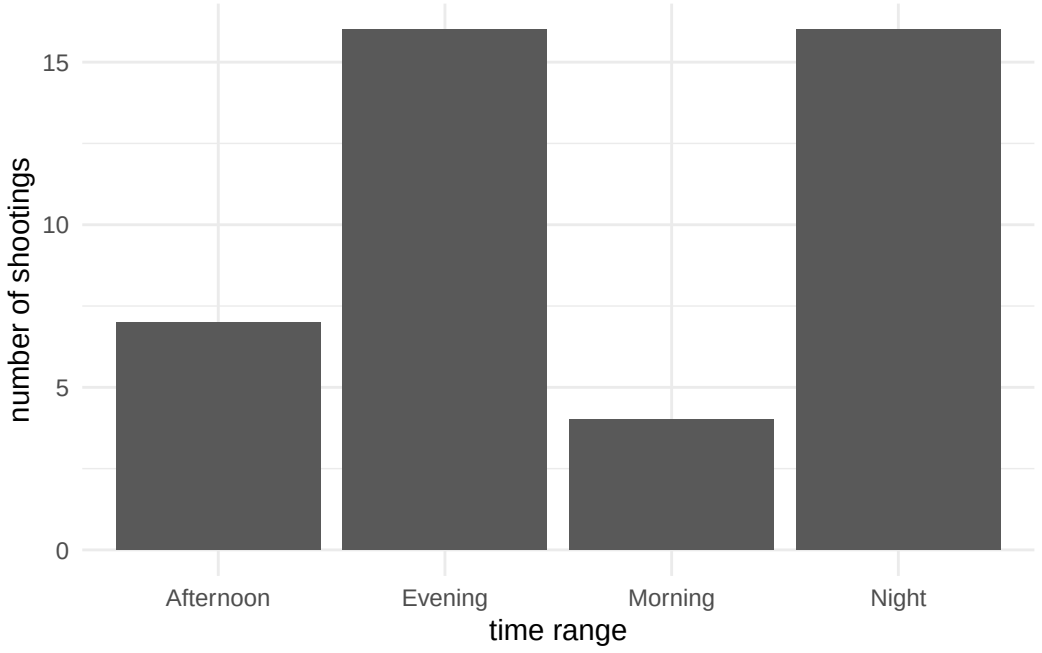

Figure 8: Shootings in Downsview-Roding-CFB (26) at different times from 2012 to 2021

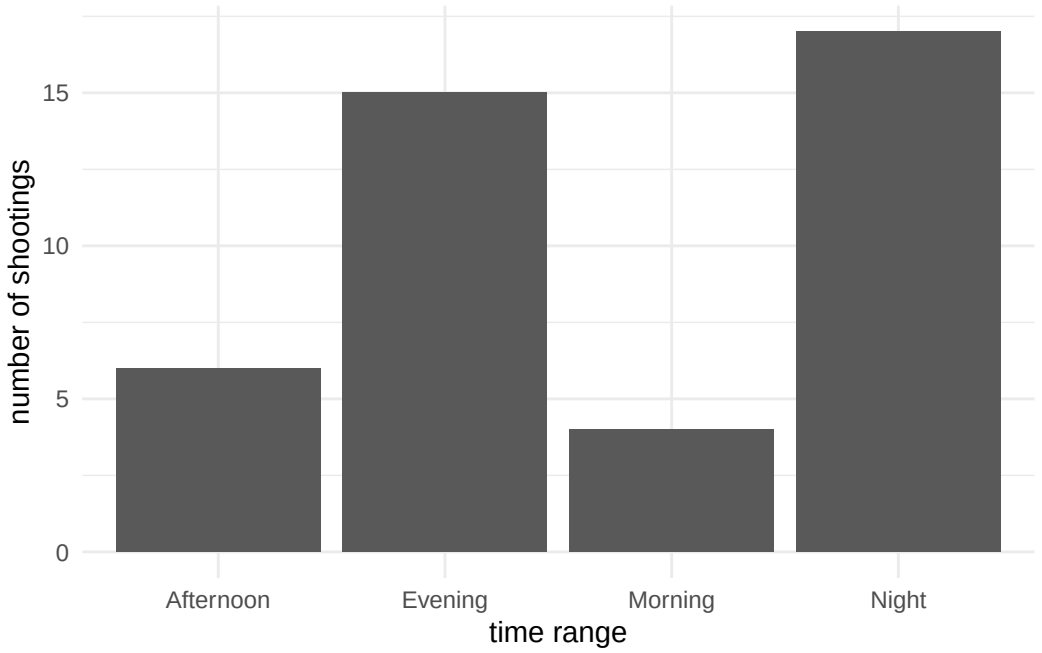

Figure 9: Shootings in West Humber-Clairville (1) at different times from 2012 to 2021

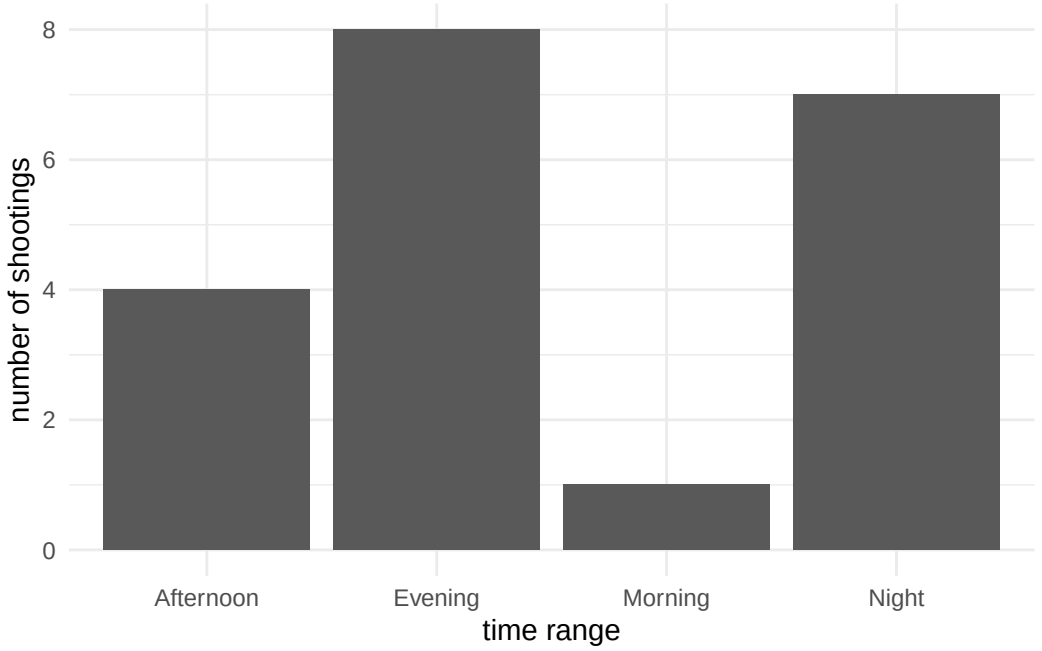

Figure 10: Shootings in West Hill (136) at different times from 2012 to 2021

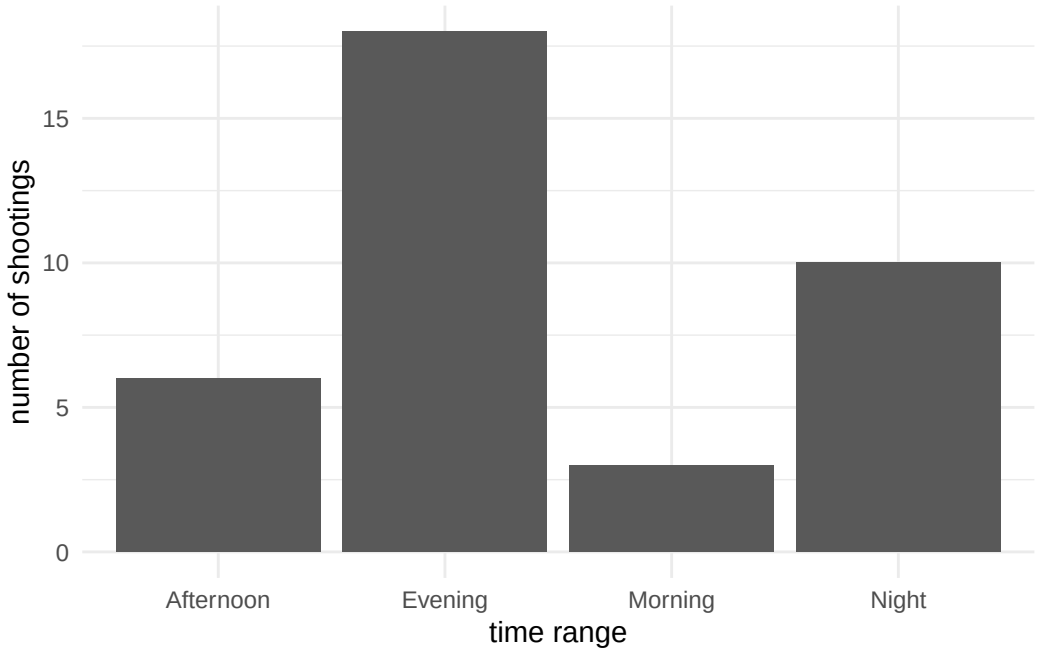

Figure 11: Shootings in Malvern (132) at different times from 2012 to 2021

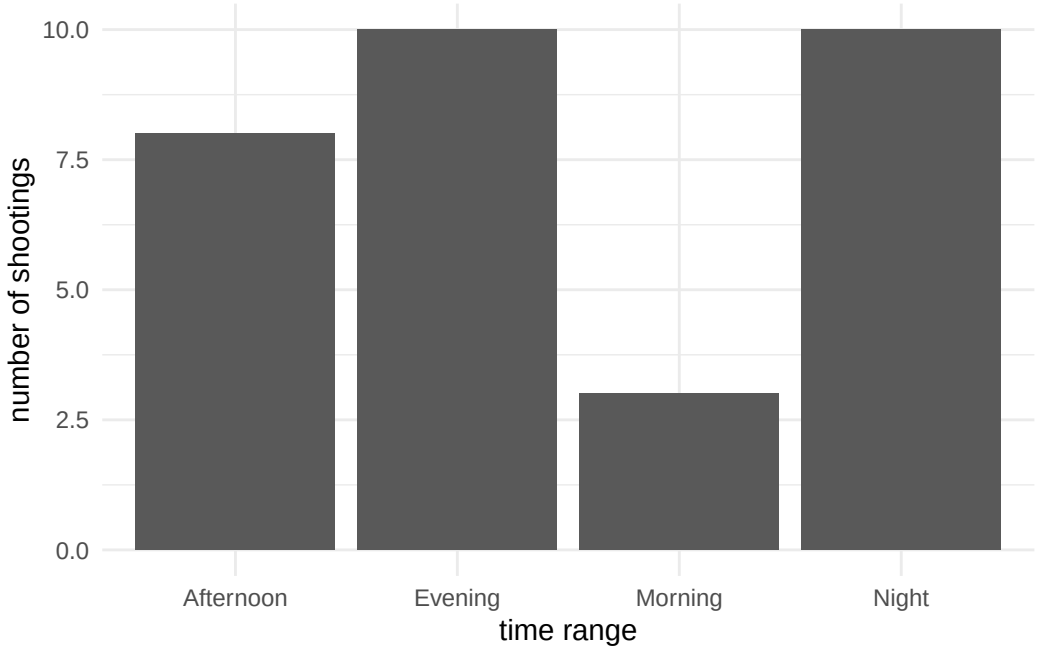

Figure 12: Shootings in York University Heights (27) at different times from 2012 to 2021

Implications 3 Shootings data by neighbourhood

We believe that there is some connection between the occurrence of shootings and regions. For example, Glenfield-Jane Heights (25) has several times as many occurrences as other communities.

According to the report (https://www.toronto.ca/ext/sdfa/Neighbourhood%20Profiles/pdf/2016/pdf1/cpa25.pd we found that the educational attainment and income levels in this community are much lower than the Toronto average. Also, we found that among these 10 neighbourhoods, Glenfield-Jane Heights (25), Downsview-Roding-CFB (26),and York University Heights (27) are three adjacent neighbourhoods. Although the shootings are unpredictable events, we recommend that you pay more attention to your safety when visiting these neighborhoods.

Limitations

In this analysis, we remove the data when both injuries and deaths appear in the data as NA. This could be a false alarm, or it could be that a shooting occurred but no one was killed or injured. Since our analysis focuses more on shootings that affect people's safety, we cleaned up all such data. Using neighborhood and time alone to analyze shootings is not sufficient. As I have said repeatedly, shootings are random, so there are many factors that can influence them. It's hard to tell when and where criminals will want to shoot. So we need more data to help us do as comprehensive an analysis as possible.

Next Step

As a next step, we can either obtain more data on factors that may affect the occurrence of shootings or conduct a more detailed analysis of existing communities so that our conclusions are not so vague. First, we can try to understand the education level, income level, etc. of the residents in different neighborhoods in Toronto. This would help us analyze in more detail the relationship between neighborhoods and the occurrence of shootings. Next, we can analyze the location and number of police stations in these neighborhoods to see if the police were able to arrive in time to solve the problem when the shooting occurred.

Reference

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*.

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. 2020. *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. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.