Hannah Yu*

My subtitle if needed

Hannah Yu

January 23, 2024

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

1 Introduction

You can and should cross-reference sections and sub-sections.

The remainder of this paper is structured as follows. Section 2….

2 Data

To investigate the homicide cases in Toronto, I obtained the dataset " Police Annual Statistical Report - Homicides (Services (2024)) from the Toronto Open Data Portal (Gelfand (2022)). Data was cleaned and analyzed using the open source statistical programming language R (R Core Team 2023), with additional support from 'tidyverse' (Wickham et al. 2019), 'ggplot2' (Wickham 2016), 'kableExtra' (Zhu 2021), Further details about data extraction and analyzation processes will be discussed in the following sections below.

This dataset includes a thorough list of alongside every homicide case in Toronto from the year 2004-2020. We cleaned the dataset and selected variables that are relevant to our analysis of homicide cases. Our variable of interest includes the year the offence occurred, police division where the offence occurred, type of homicide (shooting, stabbing, other), date of offence, and identifier of neighbourhood using Toronto's 158 neighbourhood structure. In addition to these variables, we also constructed three more variables: day of the week when the offence occurred (Monday-Sunday), weekend? season when the offence occurred (fall, winter, spring, summer), and count of homicide cases by year to further aid our data analysis (See summary statistics). This dataset was last refreshed on January 11, 2024.

*Code and data are available at: .

Table 1: Sample of Cleaned Homicide Data

Year	Division	Homicide Type	Date	Hood ID	Day of the Week	Season	Homicide Count
2004	D53	Other	2004-01-03	97	Saturday	Winter	64
2004	D42	Shooting	2004-01-08	137	Thursday	Winter	64
2004	D42	Shooting	2004-01-08	132	Thursday	Winter	64
2004	D13	Shooting	2004-01-25	93	Sunday	Winter	64
2004	D42	Shooting	2004-01-25	131	Sunday	Winter	64

We listed out the number of homicide cases each year in Table 2 and created the summary statistics table of the homicide count per year. We found the year 2011 has the least cases of homicide with only 51 cases while the year 2018 with the total of 97 cases has the most number of homicide cases.

```
Warning: Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
```

Table 2: Total Homicide Cases Each Year

Total Homicide Cases Each Year

occurrence_year	total_cases
2004	64
2005	80
2006	70
2007	86
2008	70
2009	62
2010	65
2011	51
2012	57
2013	57
2014	58
2015	59
2016	75
2017	65
2018	97
2019	79
2020	71

Of all the homicide cases, the cases caused by shooting is significantly higher than other type of homicides. In the years of our examination, shooting takes up more than 600 cases of homicides while stabbing takes over more than 250, and all other types together takes up around 300.

We then shift our attention to the homicide trend over time. Figure 2 portrays the trend of homicide cases throughout the span of almost 2 decades. With the obvious fluctuations, We do not observe a clear increase or decrease of cases across time. However, the homicide cases

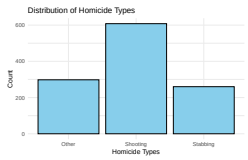


Figure 1: Distribution of Homicide Types

are decreasing after it peaked in 2018.

Following the grand picture across time, Figure 3 breaks down the homicide cases over time into further details based on the type of homicide. The trend of the three types of homicide cases generally follow the overall trend in figure 2. Shooting always prevail over other types of homicide cases even in its lowest year. We conclude that there is no clear increase or decrease of homicide across time.

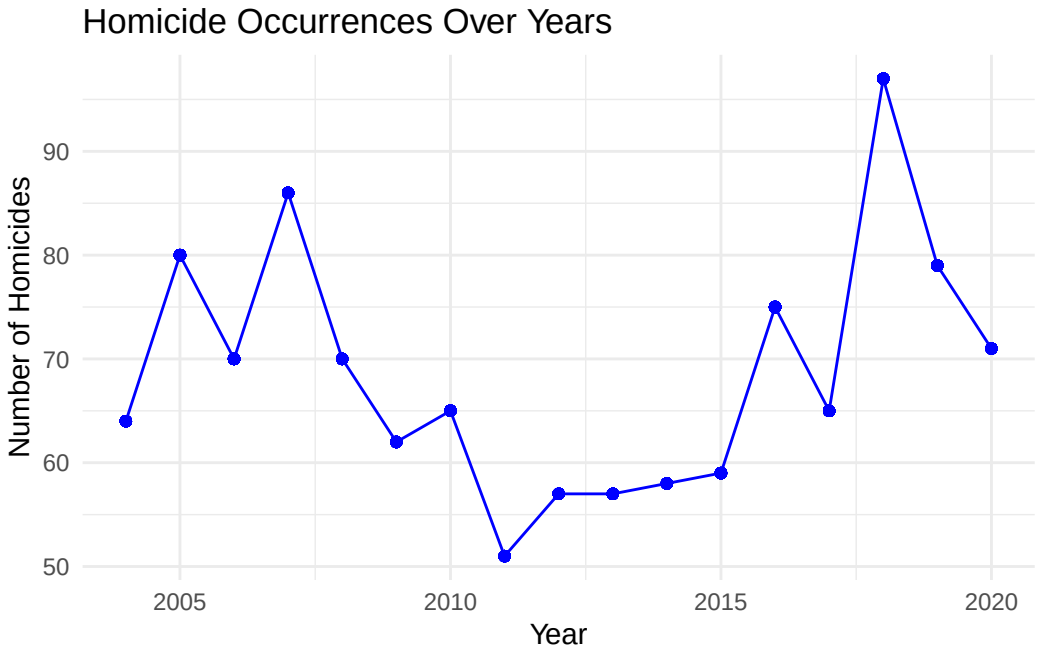


Figure 2: Homicide Occurrences Over Years

We want to determine whether if there are a specific time in the year when homicides are more likely to occurr. To investigate our interest, we created figure 4 that displays the homicide counts by every day of the week. Sunday with a total of 200 cases is the day with the most cases following by Saturday and Friday with 197 and 175 respectively. Thursday is the day with the least cases with 124 cases. We conclude that homicides are more likely to occur in weekends.

We perform similar analysis with the count of homicide cases by season and found that the

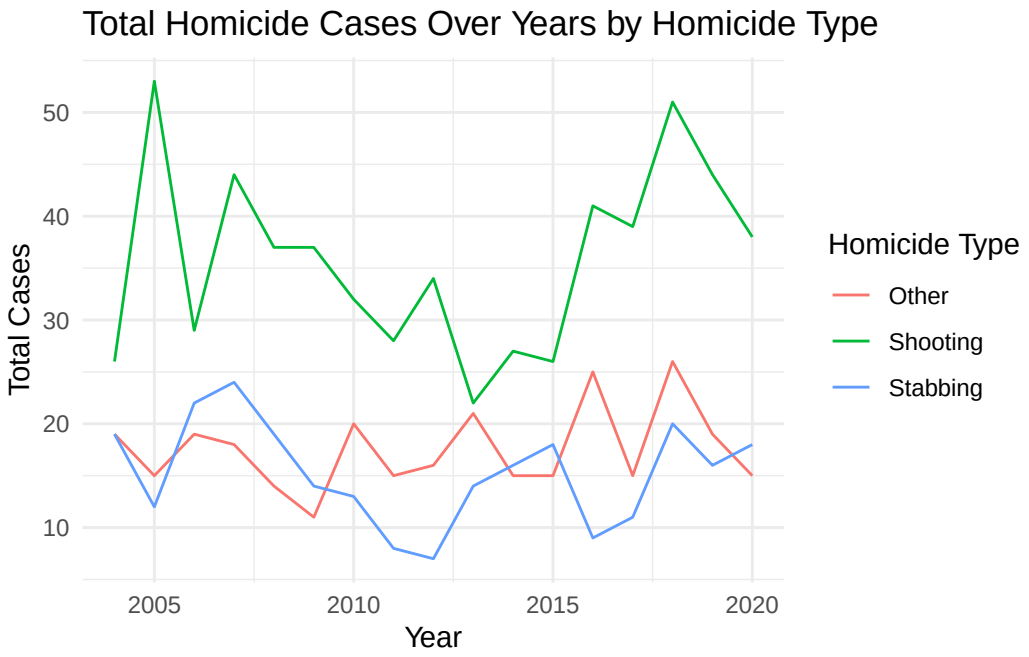


Figure 3: Total Homicide Cases Over Years by Homicide Type

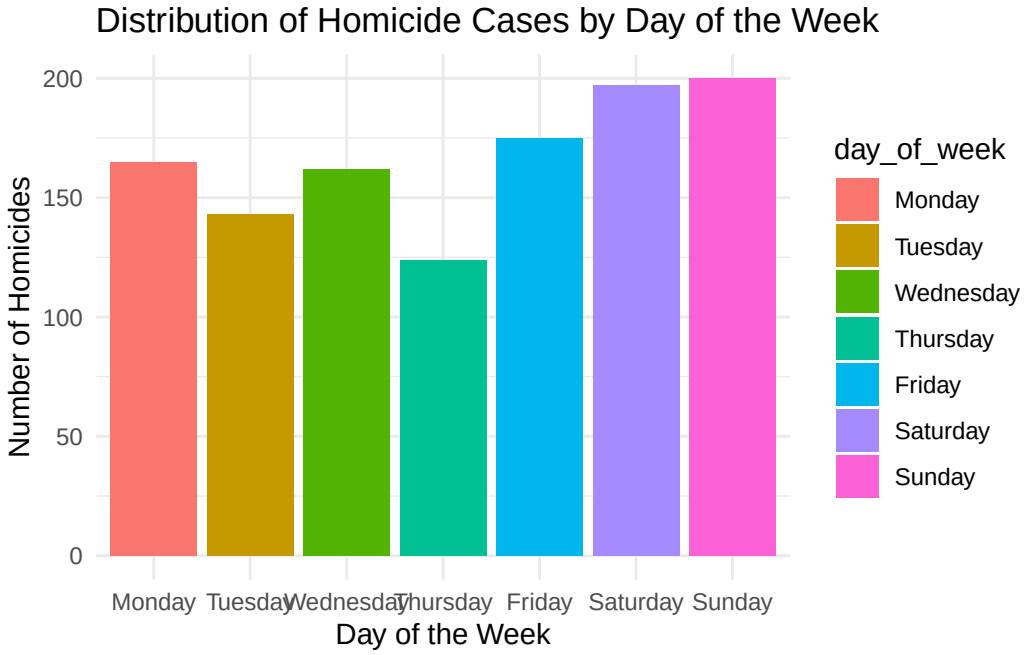
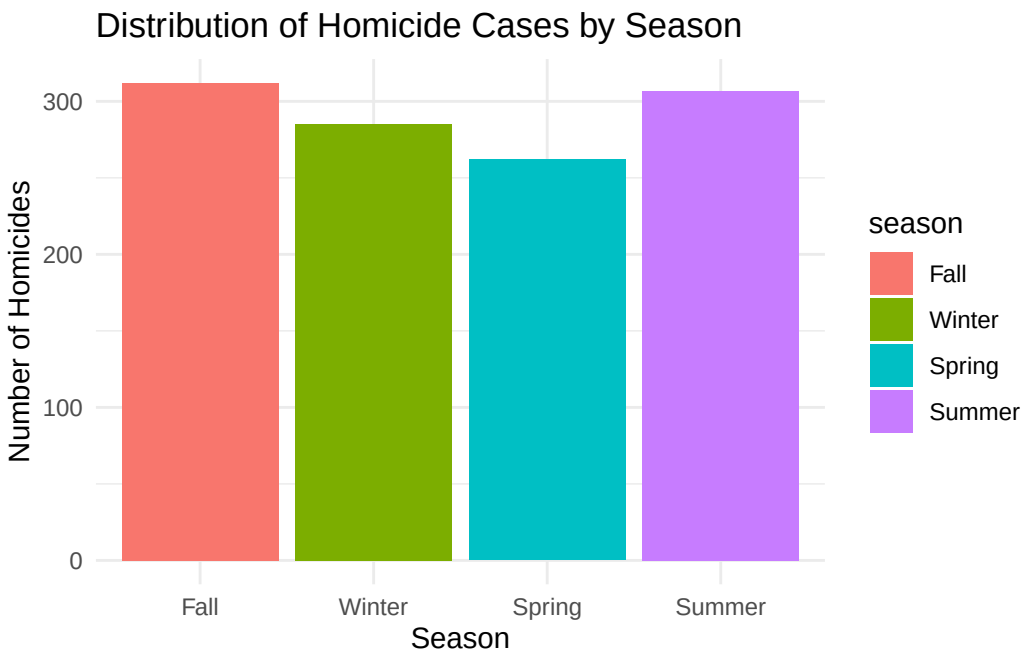


Figure 4: Distribution of Homicide Cases by Day of the Week

Table 3: **?(caption)**



seasonal data are distributed somewhat evenly with fall having slightly more cases than other weather but the differnece is to small and not worth noting.

To investigate the relationship between geographical location and homicide cases, we constructed figure . This bar plot illustrates the distribution of homicide cases in Toronto's 140 neighborhoods. This visualization enables the identification of patterns or areas of concerns in the city. While most neighborhoods have fewer than 10 cases, we notice that several neighborhoods stand out with significantly higher homicide incidences compared to the rest.

```
#| message: false
#| echo: false
#| include: false
#| label: tbl-top_neighborhood_table
#| tbl-cap: Top Neighborhood Table

top_hood_table <- cleaned_homicide_data %>%
  group_by(hood_id) %>%
  summarise(total_homicides = n()) %>%
  arrange(desc(total_homicides)) %>%
  head(5) %>%
  inner_join(cleaned_homicide_data, by = "hood_id") %>%
```

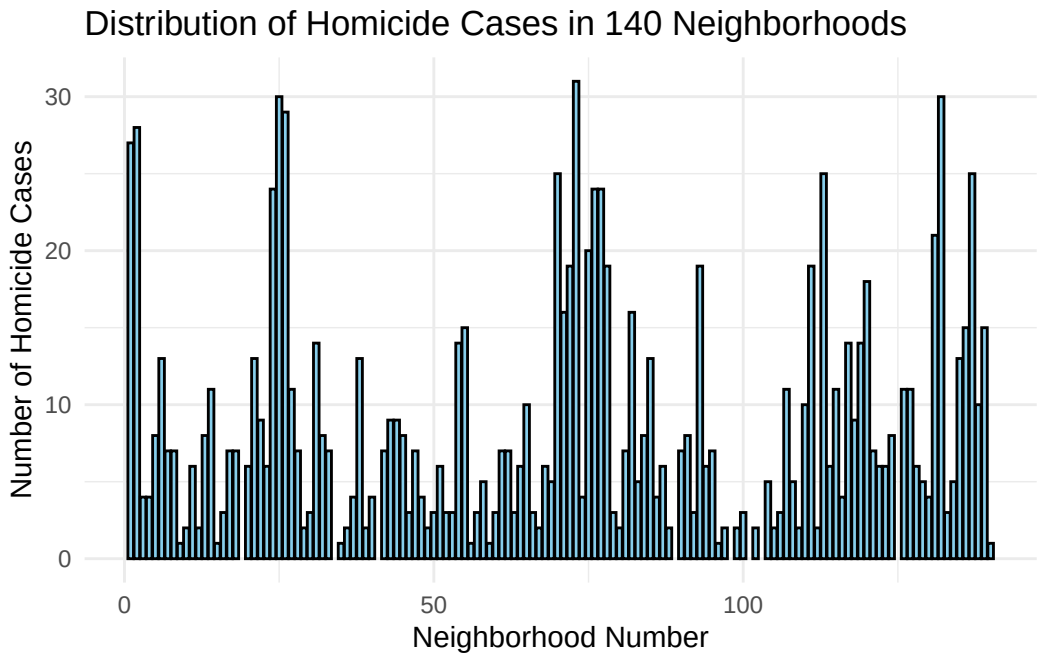


Figure 5: Distribution of Homicide in Neighborhoods

```
    select(hood_id, season, day_of_week, occurrence_year, homicide_type)

  # Print the resulting table
  print(top_hood_table)
```

```
# A tibble: 148 x 5
   hood_id season day_of_week occurrence_year homicide_type
     <int> <fct>  <fct>                 <int> <chr>
 1      73 Spring Thursday               2004 Other
 2      73 Summer Tuesday                2004 Other
 3      73 Summer Friday                 2004 Stabbing
 4      73 Winter Wednesday              2010 Stabbing
 5      73 Spring Wednesday              2010 Other
 6      73 Spring Thursday               2013 Shooting
 7      73 Summer Tuesday                2013 Other
 8      73 Summer Monday                 2016 Other
 9      73 Winter Saturday               2016 Stabbing
10      73 Winter Sunday                 2017 Shooting
# i 138 more rows
```

```
# #| message: false
# #| echo: false
# #| label: tbl-top5_season
# #| tbl-cap: Homicide Cases in Most Frequent Neighborhoods by Season
#
# # Find the top 5 neighborhoods
# top_neighborhoods <- cleaned_homicide_data |>
#   group_by(hood_id) |>
#   summarise(total_cases = n()) |>
#   arrange(desc(total_cases)) |>
#   top_n(5)
#
# # Filter data for the top 5 neighborhoods
# top_neighborhood_data <- cleaned_homicide_data |>
#   filter(hood_id %in% top_neighborhoods$hood_id)
#
# # Create a table with the relationship between season and the number of cases in the top
# season_relationship_table <- top_neighborhood_data |>
#   group_by(hood_id, season) |>
#   summarise(total_cases = n()) |>
#   arrange(hood_id, desc(total_cases))
#
# season_relationship_table$hood_id <- factor(season_relationship_table$hood_id,
#                                             levels = c("73", "25", "132", "26", "2"))
# # Transform the dataframe to the format required for printing as a table
# season_relationship_df <- season_relationship_table |>
#   select(hood_id, season, total_cases) |>
#   group_by(season, hood_id) |>
#   summarise(n = sum(total_cases)) |>
#   pivot_wider(names_from = hood_id, values_from = n) %>%
#   replace(is.na(.), 0) |>
#   ungroup() |>
#   arrange(season)
#
# # Use kbl for the entire table creation and styling
# kbl(season_relationship_df, booktabs = TRUE, linesep = "",
#     caption = "Homicide Cases in Most Frequent Neighborhoods by Season",
#     position = 'h', latex_options = "striped", font_size = 8, full_width = FALSE) |>
#   column_spec(1, width = "2cm") |>
#   add_header_above(c("", "Top 5 Neighborhoods" = 5), bold = TRUE)
```

```
Warning: Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
```

Table 4: Homicide Cases in Most Frequent Neighborhoods by Season

Homicide Cases in Most Frequent Neighborhoods by Season

	Top 5 Neighborhoods				
season	73	25	132	26	2
Fall	4	9	5	9	10
Winter	7	5	15	9	9
Spring	5	6	4	4	5
Summer	15	10	6	7	4

```
Warning: Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
```

Table 5: homicide cases in most frequent neighborhoods by day of the week

Homicide Cases in Most Frequent Neighborhoods by Season

	Top 5 Neighborhoods				
day_of_week	73	25	132	26	2
Monday	8	7	3	2	4

Tuesday	3	7	6	2	5
Wednesday	5	1	2	6	2
Thursday	4	3	4	6	3
Friday	4	6	6	5	1
Saturday	3	1	3	6	7
Sunday	4	5	6	2	6

```
Warning: Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
Since gt v0.3.0, `columns = vars(...)` has been deprecated.
* Please use `columns = c(...)` instead.
```

Table 6: Homicide Cases in Most Frequent Neighborhoods by Type

Homicide Cases in Most Frequent Neighborhoods by Season

	Top 5 Neighborhoods				
homicide_type	73	25	132	26	2
Other	10	2	10	3	3
Shooting	13	21	20	17	21
Stabbing	8	7	0	9	4

```
ggplot(cleaned_homicide_data, aes(x = division,  y = occurrence_year, fill = homicide_count))
  geom_tile() +
  labs(title = "Correlation Between Division and Homicide Cases (Heat Map)",
       x = "Division",
       y = "Year",
       fill = "Homicide Cases") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

```
#
# Talk more about it.
#
# And also planes (@fig-planes). (You can change the height and width, but don't worry about
#
# ```{r}
# #| label: fig-planes
# #| fig-cap: Relationship between wing length and width
# #| echo: false
# #| warning: false
# #| message: false
#
# analysis_data <- read_csv(here::here("outputs/data/analysis_data.csv"))
#
# analysis_data |>
#   ggplot(aes(x = width, y = length)) +
#   geom_point(alpha = 0.8) +
#   theme_minimal() +
#   labs(x = "Wing width (mm)",
#        y = "Wing length (mm)")
```

3 Results

#Our results are summarized in **?@tbl-modelresults**.

4 Discussion

4.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

4.2 Second discussion point

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

#In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows...

#In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows...

B.2 Diagnostics

References

Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://sharlagelfand.github.io/opendatatoronto/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Services, Toronto Police. 2024. *Police Annual Statistical Report - Homicides*. https://open.toronto.ca/dataset/police-annual-statistical-report-homicide/.

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.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. http://haozhu233.github.io/kableExtra/.