Exploring the 2020 Homicide Rate in Toronto

Julia Kim

```
#### Preamble ####
# Purpose: Read in data from the 2022 Australian Elections and make a
 ↪  graph
# of the number of seats each party won.
# Author: Julia Kim
# Email: juliaym.kim@mail.utoronto.ca
# Date: 12/02/24
# Prerequisites: Know where to get the Canadian elections data.
```

The city of Toronto is considered a relatively safe city to live in and travel to, with homicide rates generally declining over the past decade (Statistics Canada 2020). In this document, we are interested in the number of homicides committed per month in Toronto in 2020. We investigate this question using a workflow and code similar to Example 2.3. of Chapter 2 of *Telling Stories with Data* (Alexander 2023).

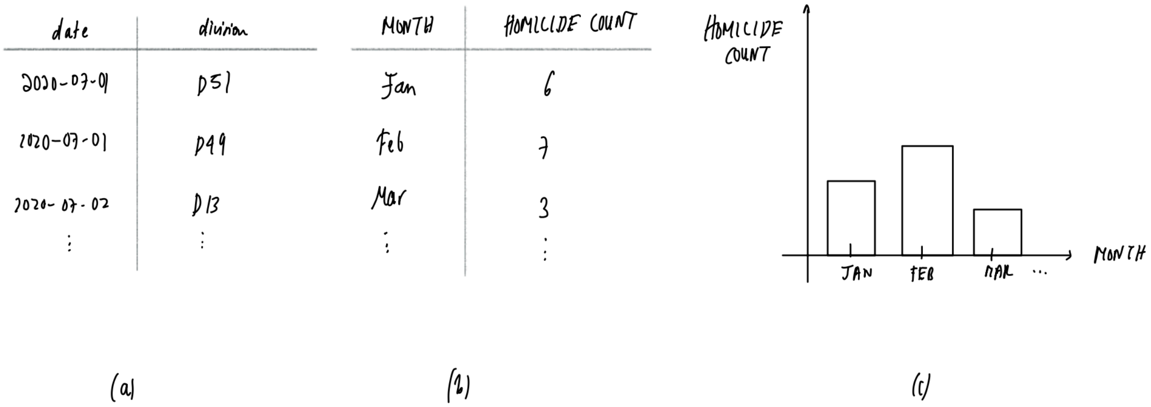
1. Plan

Figure 1 on p.2 shows a quick sketch of how the dataset, table and final graph should need to look like. The dataset that we are interested in would have the date and the location (or division) where the homicide occurred (Figure 1(a)). The table would have the number of homicides occurring per month (Figure 1(b)) and the final graph would display this data as a bar chart (Figure 1(c)).

2. Simulate

We first set up the workspace by downloading necessary packages in R, as follows.

```
#### Workspace setup ####
install.packages('tidyverse')
```



date	division
2020-07-01	D51
2020-07-01	D49
2020-07-02	D13
⋮	⋮

MONTH	HOMICIDE COUNT
Jan	6
Feb	7
Mar	3
⋮	⋮

Figure 1: Sketches of a dataset, table and graph related homicide counts in Toronto

```
install.packages('janitor')
install.packages('formatR')
install.packages('opendatatoronto')
library(tidyverse)
library(janitor)
library(formatR)
library(opendatatoronto)
knitr::opts_chunk$set(tidy.opts=list(width.cutoff=80), tidy=TRUE) #
 ↪  prevent source code from running off PDF page
```

To simulate the data, we create a dataset of two variables: `month` and `homicide_count`. Reasonable values for `month` would be the 12 months of the year; reasonable values for `homicide_count` would be randomly sampled from a Poisson distribution with mean, say, 6, though this is just an arbitrary choice.

```
#### Simulate ####
set.seed(853)  # for reproducibility
simulated_homicide_data <- tibble(month = factor(month.abb, levels =
 ↪  month.abb),
    homicide_count = rpois(n = 12, lambda = 6))

head(simulated_homicide_data)
```

```
# A tibble: 6 x 2
  month homicide_count
  <fct>          <int>
```

```
1 Jan                5
2 Feb                3
3 Mar                6
4 Apr                5
5 May                2
6 Jun                5
```

3. Acquire

We use the data made available by the City of Toronto as to homicide counts from 2004 to 2020. To access the data, we use `opendatatoronto`.

```
#### Acquire ####
raw_toronto_homicides_data <- read_csv(file = "homicides_4326.csv",
 ↪   show_col_types = FALSE)
# Save the data in case something happens to it.
write_csv(x = raw_toronto_homicides_data, file =
 ↪   "toronto_homicides.csv")
```

We quickly inspect the first six rows of the dataset by using `head()`.

```
head(raw_toronto_homicides_data)
```

```
# A tibble: 6 x 15
  `_id` EVENT_UNIQUE_ID      OCC_DATE OCC_YEAR OCC_MONTH OCC_DAY OCC_DOW OCC_DOY
  <dbl> <chr>                   <dbl>    <dbl> <chr>       <dbl> <chr>     <dbl>
1     1 GO-2004111878   1073106000000     2004 January         3 Saturd~       3
2     2 GO-2004125755   1073538000000     2004 January         8 Thursd~       8
3     3 GO-2004136086   1073538000000     2004 January         8 Thursd~       8
4     4 GO-2004148623   1075006800000     2004 January        25 Sunday       25
5     5 GO-2004148619   1075006800000     2004 January        25 Sunday       25
6     6 GO-2004152518   1077253200000     2004 February       20 Friday       51
# i 7 more variables: DIVISION <chr>, HOMICIDE_TYPE <chr>, HOOD_158 <chr>,
#   NEIGHBOURHOOD_158 <chr>, HOOD_140 <chr>, NEIGHBOURHOOD_140 <chr>,
#   geometry <chr>
```

To make this dataset similar to the one in which we are interested (Figure 1), we will need to change the column names using `clean_names()`, filter the data to the year 2020 using `filter()` and reduce the columns to only those that are relevant using `select()` .

```
toronto_homicides_clean <- clean_names(raw_toronto_homicides_data) |>
    filter(occ_year == 2020) |>
    select(occ_day, occ_month, occ_year, division)
head(toronto_homicides_clean)
```

```
# A tibble: 6 x 4
  occ_day occ_month occ_year division
    <dbl> <chr>        <dbl> <chr>
1       1 January       2020 D51
2      11 January       2020 D43
3      13 January       2020 D42
4      20 January       2020 D43
5      20 January       2020 D13
6      25 January       2020 D43
```

Having appropriately cleaned the dataset, we now save it.

```
write_csv(x = toronto_homicides_clean, file =
 ↪  "cleaned_toronto_homicides.csv")
```

4. Explore

We explore and visualise the dataset we just created by making a graph. First, we read in the dataset and obtain a count of the number of homicides per month using `group_by()` and `summarise()` from `dplyr` .

```
#### Read in the data ####
cleaned_toronto_homicides <- read_csv(file =
 ↪  "cleaned_toronto_homicides.csv", show_col_types = FALSE)
# obtain a count of the number of homicides per month
cleaned_toronto_homicides |>
    # arrange the months in chronological order from January to December
mutate(occ_month = factor(occ_month, levels = month.name)) |>
    group_by(occ_month) |>
    summarise(num_homicides = n())
```

```
# A tibble: 12 x 2
   occ_month num_homicides
   <fct>             <int>
 1 January               8
 2 February              4
 3 March                 5
 4 April                 7
 5 May                   7
 6 June                  3
 7 July                  4
 8 August                8
 9 September            11
10 October               4
11 November              4
12 December              6
```

In Figure 2, we use `ggplot2` of `tidyverse` to build a graph of the number of homicides committed per month in Toronto in 2020.

```
cleaned_toronto_homicides |>
  mutate(occ_month = factor(occ_month, levels = month.name, labels =
   ↪  month.abb)) |>
  ggplot(aes(x = occ_month)) +
  geom_bar() +
  theme_minimal() + # Make the theme neater
  labs(
    x = "Month",
    y = "Number of Homicides",
    caption = "Figure 2: Number of Homicides per month in Toronto in
     ↪  2020"
  ) +
  theme(plot.caption = element_text(hjust = 0.5, size = 10))
```

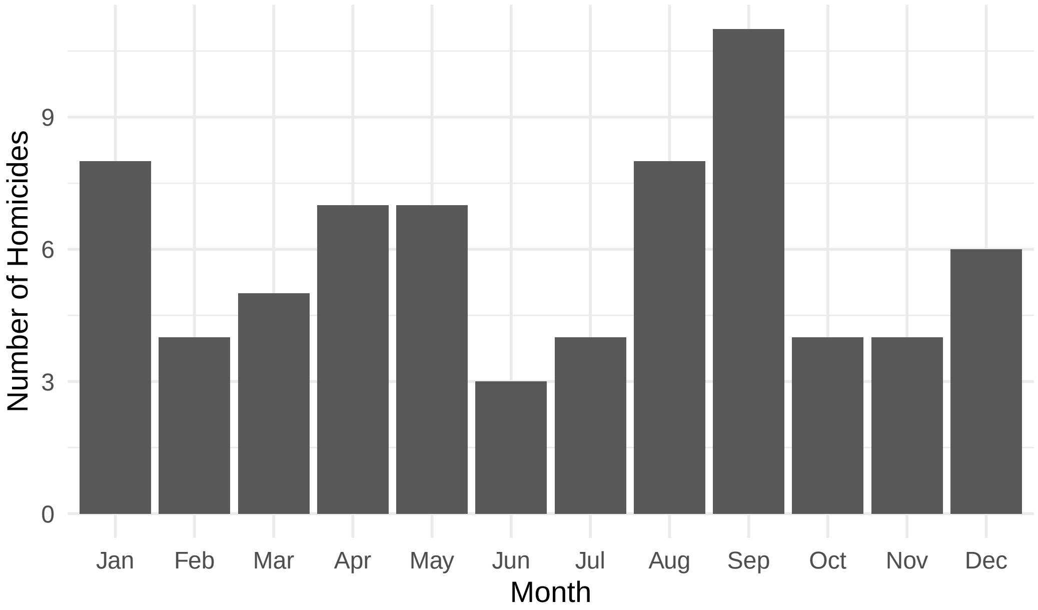


Figure 2: Number of Homicides per month in Toronto in 2020

5. Share

Toronto is considered a relatively safe city to inhabit, with homicide rides declining over the past decade (Statistics Canada 2020). We were interested in the number of homicides in Toronto per month in 2020.

We used data provided by the City of Toronto as to the homicide counts from 2004 to 2020. We cleaned, tidied, and analyzed the dataset using the statistical language R (R Core Team 2023) as well as the `tidyverse` (Wickham et al. 2019), `janitor` (Firke 2023), `opendatatoronto` (Gelfand 2022) and `formatR` (Xie 2023) packages. We then created a bar graph of the number of homicides committed per month in 2020.

We found that the number of homicides was highest in September, with 11 homicides, and lowest in June, with 3 homicides. It may be that the monthly increase or decrease in the number of homicides are due to various factors, including social activities, community dynamics and law enforcement strategies. In general, the causes for temporal fluctuations in the number of homicides per month or for longer time periods are of interest in future studies.

6. References

Alexander, Peter. 2023. *Telling Stories with Data with Applications in r*. Chapman & Hall. https://tellingstorieswithdata.com/.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

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." https://www.R-project.org/.

Statistics Canada. 2020. "Safe Cities Profile Series: Key Indicators by Census Metropolitan Area - Toronto, Ontario." https://www150.statcan.gc.ca/n1/pub/85-002-x/2020001/article/00001/toronto-eng.htm.

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.

Xie, Yihui. 2023. *formatR: Format r Code Automatically*. https://github.com/yihui/formatR.