

Assessing the Rates of Crimes Against the Person in Toronto

Sofia Sellitto

February 3, 2023

Abstract

This data was free for use, courtesy of Gelfand (2022). We are using R Core Team (2022) for analyzing this data.

Loading Packages and Data

We import tidyverse by Wickham et al. (2019) for convenience because it contains many other packages which will be useful for plotting and data manipulations. We also use Firke (2021) for cleaning and Wickham et al. (2022) for manipulating data.

```
## loading appropriate packages
```

```
library(tidyverse)
```

```
-- Attaching packages ----- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr  1.0.1
v tibble  3.1.8      v dplyr  1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
library(ggplot2)
library(knitr)
library(janitor)
```

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

chisq.test, fisher.test

```
library(dplyr)
```

```
## Calling on cleaned data
cleaned_data_dropped <- read_csv(here::here("inputs/data/cleaned_data.csv"))
```

Rows: 868 Columns: 4

-- Column specification -----

Delimiter: ","

chr (2): subtype, sex

dbl (2): reported_year, count

i Use `spec()` to retrieve the full column specification for this data.

i Specify the column types or set `show_col_types = FALSE` to quiet this message.

```
head(cleaned_data_dropped)
```

A tibble: 6 x 4

	reported_year	subtype	sex	count
	<dbl>	<chr>	<chr>	<dbl>
1	2014	Assault	M	2
2	2014	Assault	M	1
3	2014	Assault	F	1
4	2014	Assault	F	5
5	2014	Assault	F	3
6	2014	Assault	F	1

```
## summing count of crimes and making them its own group
```

```
cleaned_data_dropped_grouped <- cleaned_data_dropped |>  
  group_by(subtype, reported_year) |>  
  summarise(across(count, sum))
```

`summarise()` has grouped output by 'subtype'. You can override using the
`.groups` argument.

```
cleaned_data_dropped_grouped
```

```
# A tibble: 32 x 3  
# Groups:   subtype [4]  
  subtype reported_year count  
  <chr>         <dbl> <dbl>  
1 Assault      2014 15141  
2 Assault      2015 15897  
3 Assault      2016 16550  
4 Assault      2017 17089  
5 Assault      2018 17799  
6 Assault      2019 18819  
7 Assault      2020 16352  
8 Assault      2021 17222  
9 Other        2014  4726  
10 Other       2015  5605  
# ... with 22 more rows
```

```
### making a data frame that includes the variable sex so that I can call on it for a graph
```

```
sex_data <- cleaned_data_dropped |>  
  group_by(subtype, sex, reported_year) |>  
  summarise(across(count, sum))
```

`summarise()` has grouped output by 'subtype', 'sex'. You can override using
the `.groups` argument.

```
sex_data
```

```
# A tibble: 64 x 4
# Groups:   subtype, sex [8]
  subtype sex   reported_year count
  <chr>   <chr>         <dbl> <dbl>
1 Assault F             2014  7290
2 Assault F             2015  7502
3 Assault F             2016  7714
4 Assault F             2017  7880
5 Assault F             2018  8221
6 Assault F             2019  8679
7 Assault F             2020  7755
8 Assault F             2021  8097
9 Assault M             2014  7851
10 Assault M            2015  8395
# ... with 54 more rows
```

```
## calculating the mean number of crimes for each subcategory across 2014-2021
```

```
mean_data <- cleaned_data_dropped_grouped |>
  summarise(mean_count = mean(count))
```

```
mean_data
```

```
# A tibble: 4 x 2
  subtype      mean_count
  <chr>         <dbl>
1 Assault    16859.
2 Other       4522.
3 Robbery     3059.
4 Sexual Violation 2379.
```

```
## Creating a table to visually to display the mean number of crimes for each subcategory
```

```
knitr::kable(mean_data,"simple", caption = "Average Number of victims of crimes in Toronto",
  col.names=c("Category of Crime", "Average Number of Victims"),
  align=c('l', 'c'))
```

Table 1: Average Number of victims of crimes in Toronto across 2014-2021, grouped by category of crime

Category of Crime	Average Number of Victims
Assault	16858.625
Other	4521.500
Robbery	3059.250
Sexual Violation	2379.125

```
knitr::kable(head(cleaned_data_dropped_grouped), "simple", caption = "The first 6 rows of
col.names=c("Sub-Category of Crime", "Year", "Number of Victims"),
align=c('l', 'c', 'c'))
```

Table 2: The first 6 rows of the data frame used to represent the number of victims according to category of crime in Toronto Across 2014 - 2021

Sub-Category of Crime	Year	Number of Victims
Assault	2014	15141
Assault	2015	15897
Assault	2016	16550
Assault	2017	17089
Assault	2018	17799
Assault	2019	18819

```
knitr::kable(head(sex_data), "simple", caption = "The first 6 rows of the data frame used
align=c('l', 'c', 'c'))
```

Table 3: The first 6 rows of the data frame used to represent the gender of the number of victims according to category of crime in Toronto Across 2014 - 2021

Sub-Category of Crime	Sex	Year	Number of Victims
Assault	F	2014	7290
Assault	F	2015	7502
Assault	F	2016	7714
Assault	F	2017	7880
Assault	F	2018	8221
Assault	F	2019	8679

See (figure1?)

```

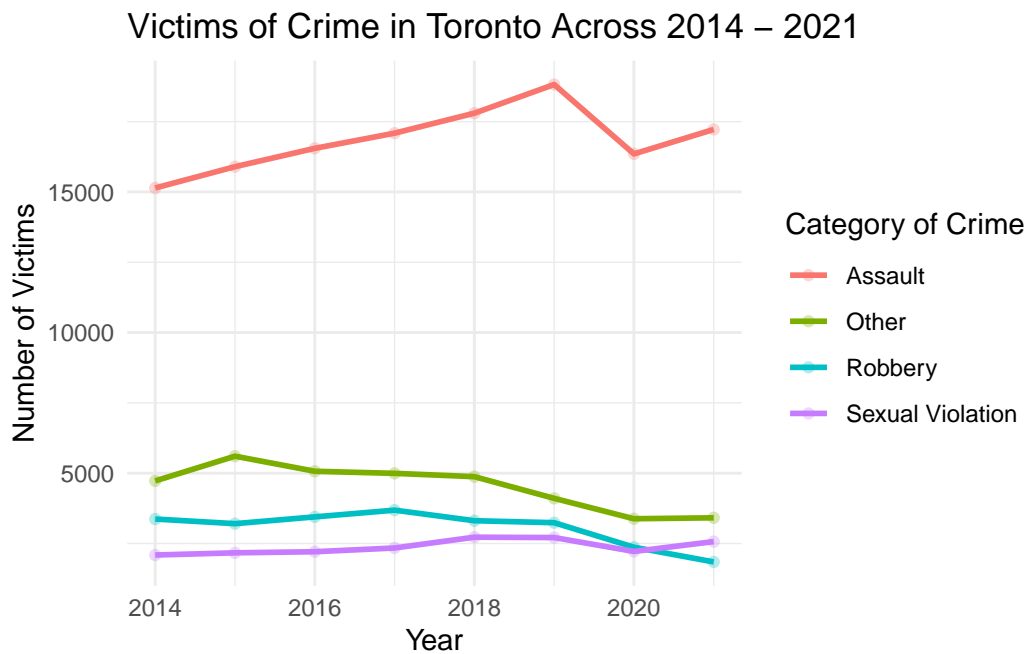
## creating a line graph that includes the number of victims according to category of crime

#| fig-cap: "Number of victims according to category of crime in Toronto Across 2014 - 2021"
#| label: figure1

cleaned_data_dropped_grouped |>
  ggplot(aes(x = reported_year, y = count, colour = subtype))+
  geom_point(linewidth = 5, alpha = 0.3)+
  geom_line(linewidth = 1)+
  theme_minimal()+
  labs(
    title = "Victims of Crime in Toronto Across 2014 - 2021",
    x = "Year",
    y = "Number of Victims",
    colour = "Category of Crime" )

```

Warning in geom_point(linewidth = 5, alpha = 0.3): Ignoring unknown parameters: ``linewidth``



See (figure2?)

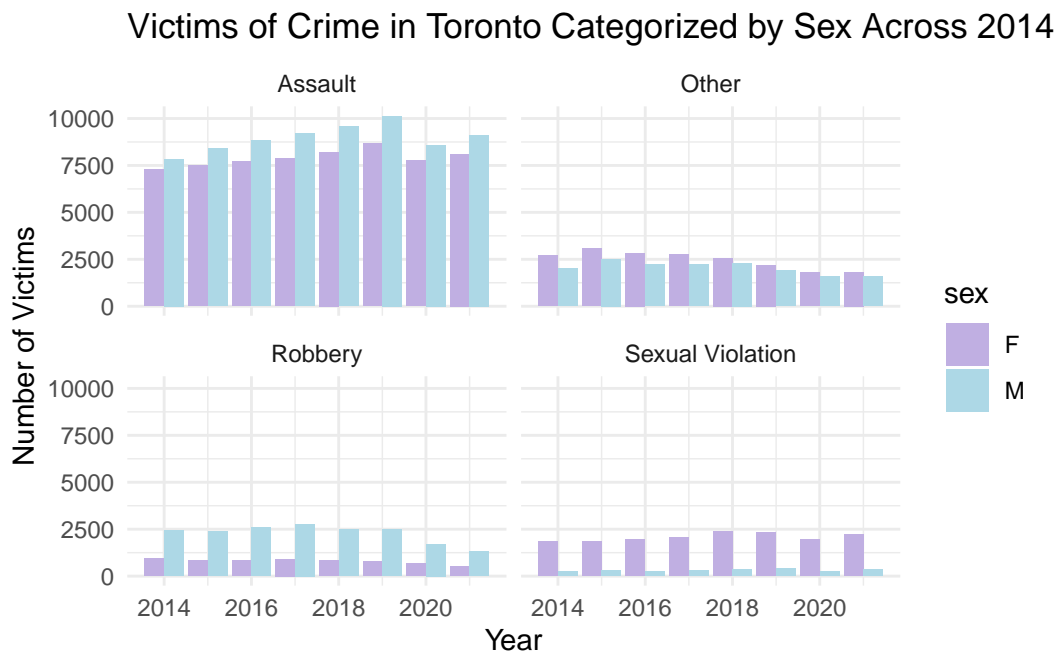
```

## calling on my data frame to create a bar plot of the sex of the victims across time

#| fig-cap: "How the victims of crimes varries across sex in Toronto from 2014-2021"
#| label: figure2

ggplot(sex_data, aes
      (x =reported_year,
       y =count,
       fill =sex))+
  geom_bar(position="dodge", stat="identity")+
  facet_wrap(~subtype)+
  scale_fill_manual(values=c("#C0AFE2", "#ADD8E6"))+
  theme_minimal()+
  labs(
    title = "Victims of Crime in Toronto Categorized by Sex Across 2014 - 2021",
    x = "Year",
    y = "Number of Victims",
    legend = "Sex")

```



```
library(ggmap)
```

i Google's Terms of Service: <<https://mapsplatform.google.com>>

i Please cite ggmap if you use it! Use ``citation("ggmap")`` for details.

References

- Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Gelfand, Sharla. 2022. *Opendatatoronto: Access the City of Toronto Open Data Portal*. <https://CRAN.R-project.org/package=opendatatoronto>.
- R Core Team. 2022. *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 Golemund, 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. 2022. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.