Vancouver Crime Data Analysis and Visualization

AUTHOR
Eric Cheung

```
library(tidyverse)
```

2023 Vancouver Crime Data

```
crimedata <- read_csv("crimedata_csv_AllNeighbourhoods_2023.csv")
```

```
Rows: 26543 Columns: 10
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (3): TYPE, HUNDRED_BLOCK, NEIGHBOURHOOD
dbl (7): YEAR, MONTH, DAY, HOUR, MINUTE, X, Y

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.
```

```
crimedata
```

```
# A tibble: 26,543 × 10
   TYPE          YEAR MONTH   DAY  HOUR MINUTE HUNDRED_BLOCK NEIGHBOURHOOD      X
   <chr>        <dbl> <dbl> <dbl> <dbl>  <dbl> <chr>         <chr>          <dbl>
 1 Break and …   2023     4     1     4      7 10XX BEACH A… West End      4.90e5
 2 Break and …   2023     4     3     0     50 10XX BEACH A… Central Busi… 4.90e5
 3 Break and …   2023     5    11    18      0 10XX BEACH A… Central Busi… 4.90e5
 4 Break and …   2023     8     9     4     31 10XX BEACH A… Central Busi… 4.90e5
 5 Break and …   2023     9     2     4     24 10XX BIDWELL… West End      4.90e5
 6 Break and …   2023     3    16     4      0 10XX BROUGHT… West End      4.90e5
 7 Break and …   2023     8    12     1      3 10XX BUTE ST  West End      4.90e5
 8 Break and …   2023     3    17    15     23 10XX BUTE ST  West End      4.90e5
 9 Break and …   2023     7     9    21      0 10XX CAMBIE … Central Busi… 4.91e5
10 Break and …   2023     2     4     2     57 10XX CANADA … Central Busi… 4.91e5
# ℹ 26,533 more rows
# ℹ 1 more variable: Y <dbl>
```

This data describes all the crimes that were reported to the Vancouver Police Department in 2023. Each row is a crime reported and contains the time, location and type of crime which means that there were 26,543 crimes documented in the above data frame.

Crimes on January 25th, 2023

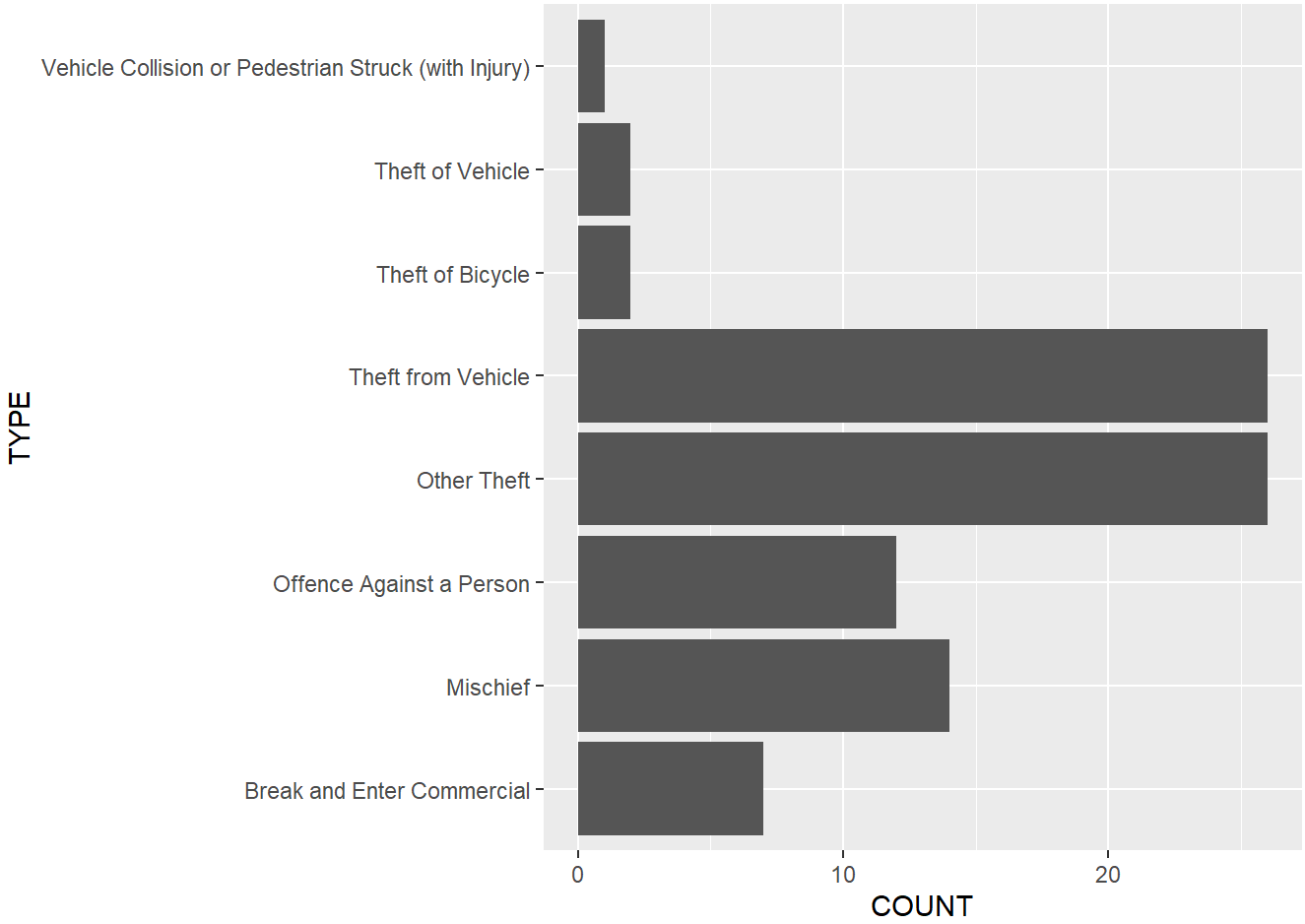
```
jan_25_crimes <- crimedata |>
  filter(YEAR == 2023, MONTH == 1, DAY == 25)

jan_25_crimes
```

```
# A tibble: 90 × 10
   TYPE          YEAR MONTH   DAY  HOUR MINUTE HUNDRED_BLOCK NEIGHBOURHOOD      X
   <chr>        <dbl> <dbl> <dbl> <dbl>  <dbl> <chr>         <chr>          <dbl>
 1 Break and …   2023     1    25     3     45 11XX COMMERC… Grandview-Wo… 4.95e5
 2 Break and …   2023     1    25     3     30 12XX BUTE ST  West End      4.90e5
 3 Break and …   2023     1    25     4     49 12XX GRANVIL… Central Busi… 4.91e5
 4 Break and …   2023     1    25    10     42 3XX SW MARIN… Marpole       4.92e5
 5 Break and …   2023     1    25     0     15 3XX WATER ST  Central Busi… 4.92e5
 6 Break and …   2023     1    25     7     58 5X W CORDOVA… Central Busi… 4.92e5
 7 Break and …   2023     1    25     2      0 7XX DAVIE ST  Central Busi… 4.91e5
 8 Mischief      2023     1    25    23      0 10XX W 52ND … Oakridge      4.91e5
 9 Mischief      2023     1    25     9      0 11XX ROBSON … West End      4.91e5
10 Mischief      2023     1    25    15     50 11XX UNION ST Strathcona    4.94e5
# ℹ 80 more rows
# ℹ 1 more variable: Y <dbl>
```

There were 90 crimes reported to the VPD on my birthday on January 25th, 2023. I can tell because I used the filter() function to only view observations if the year equaled 2023, month equaled 1, and day equaled 25. And the resulting data frame returned 90 rows with each row being a reported crime.

FIG 1:

```
jan_25_crimes |>
  group_by(TYPE) |>
  summarize(COUNT = n()) |>
  ggplot(aes(x = TYPE, y = COUNT)) +
  geom_col() +
  coord_flip()
```



The most common crime reported on my birthday was “Theft from Vehicle” and “Other Theft”. The least common was “Vehicle Collision or Pedestrian Struck (with injury)”. Theft could be most common due to it being comparatively low risk without much punishment if caught and low probability of being caught. And vehicle collision being the least common could be because people are generally good drivers.