STAT 443: Assignment 2

Kevin Liu (94200474)

March 14, 2025

For this project, as suggested on Piazza, I will look at Bayesian classification methods on the crime category. Alternatively, if that does not work out, I will use time series to find potential trends/seasonal effects on crime statistics (potentially splitting by neighborhood). Data is from here

```
crime_data = read.csv("crime_data/crime_data.csv")
head(crime_data)
```

```
##                         TYPE YEAR MONTH DAY HOUR MINUTE   HUNDRED_BLOCK
## 1 Break and Enter Commercial 2012    12  14    8     52
## 2 Break and Enter Commercial 2019     3   7    2      6   10XX SITKA SQ
## 3 Break and Enter Commercial 2019     8  27    4     12 10XX ALBERNI ST
## 4 Break and Enter Commercial 2021     4  26    4     44 10XX ALBERNI ST
## 5 Break and Enter Commercial 2014     8   8    5     13 10XX ALBERNI ST
## 6 Break and Enter Commercial 2020     7  28   19     12 10XX ALBERNI ST
##   NEIGHBOURHOOD        X       Y
## 1      Oakridge 491285.0 5453433
## 2      Fairview 490613.0 5457110
## 3      West End 491004.8 5459177
## 4      West End 491007.8 5459174
## 5      West End 491015.9 5459166
## 6      West End 491015.9 5459166
```

A second data set that I could potentially use is from the WHO, with road traffic accidents for people aged 15-24. I could also use time series on this data set, or potentially apply a simple model and improved model. Data is from here

```
headers = read.csv("WHO.csv", skip = 6, header = F, nrows = 1, as.is = T)
WHO_data = read.csv("WHO.csv", skip = 7, header = F)
colnames(WHO_data)= headers
head(WHO_data)
```

```
##   Region Code                     Region Name Country Code Country Name Year
## 1         NAC North America and the Caribbean          CAN       Canada 2000
## 2         NAC North America and the Caribbean          CAN       Canada 2000
## 3         NAC North America and the Caribbean          CAN       Canada 2000
## 4         NAC North America and the Caribbean          CAN       Canada 2000
## 5         NAC North America and the Caribbean          CAN       Canada 2001
## 6         NAC North America and the Caribbean          CAN       Canada 2001
##    Sex Age group code Age Group Number
## 1  All       Age15_19   [15-19]    354
```

```
## 2  All       Age20_24   [20-24]    348
## 3 Male       Age15_19   [15-19]    228
## 4 Male       Age20_24   [20-24]    259
## 5 Male       Age15_19   [15-19]    233
## 6 Male       Age20_24   [20-24]    266
##   Percentage of cause-specific deaths out of total deaths
## 1                                                33.74643
## 2                                                26.40364
## 3                                                32.02247
## 4                                                27.03549
## 5                                                32.00549
## 6                                                28.20785
##   Age-standardized death rate per 100 000 standard population
## 1                                                           NA
## 2                                                           NA
## 3                                                           NA
## 4                                                           NA
## 5                                                           NA
## 6                                                           NA
##   Death rate per 100 000 population NA
## 1                          17.07010 NA
## 2                          16.66438 NA
## 3                          21.42471 NA
## 4                          24.27537 NA
## 5                          21.40232 NA
## 6                          24.64398 NA
```