

# Is denser population area has more crime than others?

Thu Dong

This is my abstract.

```
#install.packages("kableExtra")
#install.packages("psych")
```

```
#### Workspace setup ####
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2     3.4.4      v tibble     3.2.1
v lubridate  1.9.3      v tidyr      1.3.0
v purrr       1.0.2

-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
library(ggplot2)
library(knitr)
library(kableExtra)
```

Attaching package: 'kableExtra'

The following object is masked from 'package:dplyr':

```
group_rows
```

```
library(psych)
```

Attaching package: 'psych'

The following objects are masked from 'package:ggplot2':

```
%+%, alpha
```

```
neighborhood_crime_clean <-  
  read_csv(  
    "neighborhood_crime_2023_clean.csv",  
    show_col_types = FALSE  
  )
```

New names:

```
* `` -> `...1`
```

## Table of Content

### 1. Introduction

Toronto is considered one of the safest cities in Canada and North America with homicide rates of approximately 1.95 per 100,000 people in 2020 (Armstrong and Jaffray 2021). However, it was reported that there was a 14% increase in the overall crime rate in 2022 and it will continue to increase in 2023 (Rosenbaum 2023). It is believed that more populated areas tend to have higher crime incidents. This paper attempts to investigate whether areas with higher populations experience higher levels of crime occurrences.

There have been many studies in criminology that discuss the relationship between population size and crime level. One of the first works in criminology is research conducted across France in 1825 by Balbi and Guerry which reveals that the area with a higher population has more crime occur than others. However, it was later discussed that these 2 relationships are not linear and might be a correlation-based conclusion. (Friendly 2007). On the other hand, some other research in criminology disproved the theory that crime increases with population. (Chamlin and Cochran 2004). This research argued that there is no direct association between

population and crime rate and other factors that are related to population growth such as quality of life have a stronger correlation to the volume of crime than population. Most of the research has used crime rates to measure the intensity of crime in an area, standardize a measure, and manage risk assessments. This paper will be using crime volume or the number of crimes to assess the magnitude of crime level. Using crime volume also helps policymakers to understand the safety state of each area to give efficient resource allocation.

To discuss the relationship between population density and crime volume, this paper will present in the following section: Data, Results, Discussion, and Conclusion. In the Data section, I will discuss the nature of the data obtained through the City of Toronto's OpenDataToronto Library and the steps I took to clean the data. The Result section will highlight some trends found in the data. Additionally, the Discussion Section will explore the relationship between the level of crime occurrence and population density. Finally, The conclusion section will summarize the main findings of the paper.

## **2. Data**

The data used in this paper was obtained through the City of Toronto's OpenDataToronto Library. The data set is called 'Neighborhood Crime Rate' ("Open Data Dataset" 2024). Data was collected and analyzed using the statistical programming software R (R Core Team 2023), with additional support from tidyverse (Wickham et al. 2019), ggplot2 (Wickham 2016), dplyr (Wickham et al. 2023), readr (Wickham, Hester, and Bryan 2023), ( **add all package used**)

### **2.1.Crime volume in 2023.**

The original "Neighborhood Crime Rate" Data set includes \_\_\_\_\_ Variables and \_\_\_\_\_ observations. It records the number of Assaults, Auto Theft, Break and Entering, Robbery, Theft Over, Homicide, and Shooting from the year 2014 to the year 2023, as well as the crime rate of each crime in each year.

To get the data for crime volume in 2023, I selected 8 columns: Assaults, Auto Theft, Breaking and Entering, Robbery, Theft Over, Homicide, and Shooting in 2023 since it is the closest data recorded. Additionally, all the "NA" values were set to 0 to allow for mathematical operation on the data. Then, I summed the number of each crime to get the total crime of each area in the year 2023. Finally, I renamed all the columns needed to allow for clarity.

### **2.2. Population in 2023.**

Data for area in Toronto and population in 2023 was available in the " Neighborhood Crime rate" dataset. Therefore, the name of the area and the population in 2023 were added to the clean data set. Table 1 shows the first 5 rows of the data set with area name, population, and the number of crime occurrences in 2023.

AREA_NAME	POPULATION_2023	TOTAL_CRIME_2023
South Eglinton-Davisville	21987	253
North Toronto	15077	238
Dovercourt Village	13837	265
Junction-Wallace Emerson	26240	486
Yonge-Bay Corridor	14731	1117

```
neighborhood_crime <- select(neighborhood_crime_clean, AREA_NAME, POPULATION_2023, TOTAL_CRIME_2023)

write_csv(neighborhood_crime, "neighborhood_crime_analysis.csv")

kable(head(neighborhood_crime,5)) %>%
  kable_styling(full_width = FALSE)
```

### 3. Result

#### 3.1. Toronto Area and the population in 2023.

There are 158 areas divided in Toronto. Out of the 158 areas in Toronto, Beechborough-Greenbrook is the part with the lowest population of 7,057 and West Humber-Clairville is the area with the highest population at 36,388. Table 2 shows the 10 most populated areas in Toronto. The second most populated area is Mount Olive-Silverstone-Jamestown with 35,386 people but the third highest populated area is the Annex with 34502 people. Other area that are highly populated are Glenfield-Jane Heights, Harbourfront-CityPlace, Agincourt North, etc...

```
#Summary table
summary_result <- describe(neighborhood_crime$POPULATION_2023)
#summary_result

#area with lowest population
min_populated_area <- neighborhood_crime$AREA_NAME[which.min(neighborhood_crime$POPULATION_2023)]

sorted_population_table <- neighborhood_crime_clean %>%
  arrange(desc(POPULATION_2023)) %>%
  slice_head(n = 10)
```

```
sorted_population_table |>
  summarise(population = POPULATION_2023,
            .by = AREA_NAME) |>
  kable()
```

AREA_NAME	population
West Humber-Clairville	36388
Mount Olive-Silverstone-Jamestown	35386
Annex	34502
Glenfield-Jane Heights	32251
Harbourfront-CityPlace	31477
Agincourt North	31260
Banbury-Don Mills	31047
St Lawrence-East Bayfront-The Islands	31003
South Riverdale	30135
York University Heights	29897

### 3.2. Toronto Area and crime volume.

Table 3 shows 10 areas with the highest level of criminal incidents in Toronto. Out of the 158 areas in Toronto, West Humber-Clairville has the most crime occurrence in 2023 with 1856 crime cases and also has the most population. The area with the second highest level of crime occurrence is Downtown Young East with 1259 cases. Additionally, the area that came third in terms of highest crime occurrence is York University Height with 1236 cases. The fourth area where the most crime occurred is the Young-Bay Corridor with 1117 cases.

```
# 10 neighborhood with the most crime

sorted_table <- neighborhood_crime_clean %>%
  arrange(desc(TOTAL_CRIME_2023)) %>%
  slice_head(n = 10)

sorted_table |>
  summarise(total_crime = TOTAL_CRIME_2023,
            .by = AREA_NAME) |>
  kable()
```

AREA_NAME	total_crime
West Humber-Clairville	1856
Downtown Yonge East	1259
York University Heights	1236
Yonge-Bay Corridor	1117
Moss Park	1113
Kensington-Chinatown	999
Etobicoke City Centre	984
Annex	933
Wellington Place	898
St Lawrence-East Bayfront-The Islands	878

```

area <- neighborhood_crime_clean$AREA_NAME
max_assault <- which.max(neighborhood_crime_clean$ASSAULT_2023)
max_auto_theft <- which.max(neighborhood_crime_clean$AUTOTHEFT_2023)
max_bike_theft <- which.max(neighborhood_crime_clean$BIKETHEFT_2023)
max_break_enter <- which.max(neighborhood_crime_clean$BREAKENTER_2023)
max_homicide <- which.max(neighborhood_crime_clean$HOMICIDE_2023)
max_shooting <- which.max(neighborhood_crime_clean$SHOOTING_2023)
max_robbery <- which.max(neighborhood_crime_clean$ROBBERY_2023)
max_theft_mv <- which.max(neighborhood_crime_clean$THEFTFROMMV_2023)
max_theft_over <- which.max(neighborhood_crime_clean$THEFTOVER_2023)

type_of_crime <- data.frame(
  Type_of_Crime= c("Assault", "Auto Theft", "Bike Theft", "Breaking and Entering",

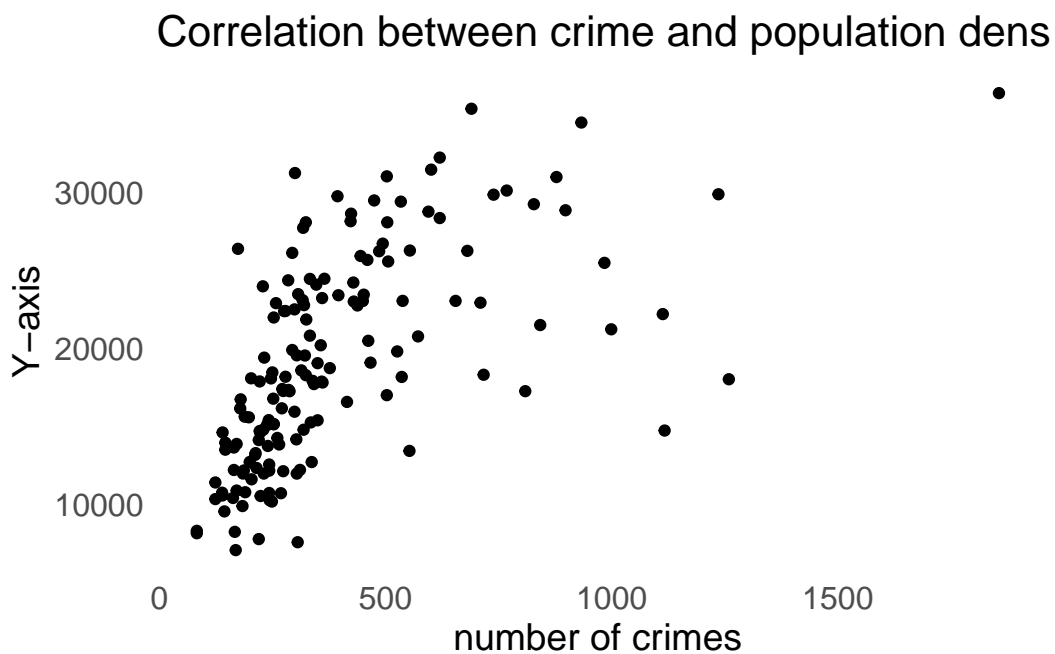
kable(type_of_crime)

```

Type_of_Crime	Area_with_max
Assault	Downtown Yonge East
Auto Theft	West Humber-Clairville
Bike Theft	Yonge-Bay Corridor
Breaking and Entering	York University Heights
Robbery	Moss Park
Homicide	Mount Olive-Silverstone-Jamestown
Shooting	York University Heights
Theft from moving vehicles	West Humber-Clairville
Theft Over	West Humber-Clairville

#### 4. Discussions.

```
#scatterplot
ggplot(neighborhood_crime_clean, aes(x = TOTAL_CRIME_2023, y = POPULATION_2023)) +
  geom_point() +
  labs(title = "Correlation between crime and population density", x = "number of crimes",
        theme_minimal() + # Minimal theme
  theme(
    plot.title = element_text(size = 16), # Title size and bold
    axis.title = element_text(size = 14), # Axis title size
    axis.text = element_text(size = 12), # Axis text size
    legend.position = "none", # Remove legend
    panel.grid.major = element_blank(), # Remove major grid lines
    panel.grid.minor = element_blank() # Remove minor grid lines
  )
```



#### 5. Conclusion.