doupi3*

sid

Mengze Zhang

2 February 2023

we consider data about and find that

1.Introduction

2.Data

2.1 Data Source

This report utilizes data on crime rates in Toronto neighborhoods obtained from the Toronto Police Service Annual Statistical Report (ASR) (Toronto Police Service 2022b). The ASR is a comprehensive overview of police related statistics which has been openly available to the public since 2019 (Toronto Police Service 2022a). The neighborhood crime rate dataset analyzed in this report was obtained in csv format from the City of Toronto Open Data Portal using the R package `opendatatoronto` (Gelfand 2020). The dataset was last updated on May 18th, 2022.

2.2 Data Collection

2.3 Data Characteristics

The Neighborhood Crime Rates dataset contains aggregated data of 8 types of crime in each Toronto neighborhood between the years 2014 and 2020. There are 16920 observations in the dataset and 7 attributes: index, object ID, hood name, hood ID, 2021 population projection, count of each crime type in each year, crime rate of each crime type in each year. The first 4 attributes are identifiers which were removed prior to analysis. An additional attribute to

*Code and data are available at github.com/MissyZhang/Crime_Data_by_Neighbourhood.

categorize neighborhoods based on population was created during analysis by dividing the population into quartiles and label each neighborhood accordingly (e.g. “small”, “medium”, “large”, “giant”). The crime counts are aggregated by neighborhood, year, and crime type, which I deleted because I only want to focus on crime rate out of such a big dataset. Finally, the crime rate is calculated as the crime count per 100,000 population per year, which I selected data from 2017 to 2021 to cut the data short and clean while also remains rigorousness for reasonable analysis.

2.4 Data Analysis

Analysis for this project uses the R statistical programming language (R Core Team 2022), as well as `tidyverse` (Wickham et al. 2019), `tidyr` (Wickham 2022) and `dplyr` (Wickham et al. 2021) programming packages. Because the data is managed using R Projects, `here` is used to reference file locations (Müller 2020). Figures and tables are created with `ggplot2` (Wickham 2016) and `kableextra` (Zhu 2020), and new data frame is created with `tibble` (Müller and Wickham 2022). The package `knitr` (Xie 2021) is used to generate the PDF report. The package `palettetown` (Lucas 2016) is used to create color palette for figures, and the packages `patchwork` (Pedersen 2022) and `ggpubr` (Kassambara 2022) are used to patch graphs together.

The first aspect of the data I investigated is the year-on-year changes in crime rates for different crime categories. Figure 1 shows changes in average crime rates for 8 different crime types from 2017 to 2021. It is clear that assault has the highest crime rate in each year, fluctuating around 600, while homicide has the lowest and remains stable around 10. Both the motorcycle theft rate and break and enter rate have a trend of rising first and then falling, with turning points in 2020 and 2019 respectively. The robbery rate decreases year by year from 2017 to 2021, while auto theft rate sees an increase and even exceeded break and enter rate in 2021. Similar to homicide rate, both theft rate and shooting rate remain relatively low and stable over the time period, with theft rate (around 50) slightly higher than shooting rate (around 30) each year.

The next aspect of the data I analyzed is the relationship between population and crime rate.

- Small neighborhood: Has a population fewer than 10,000
- Medium neighborhood: Has a population between 10,001 and 15,000
- Large neighborhood: Has a population between 15,001 and 2,0000
- Giant neighborhood: Has a population more than 20,000

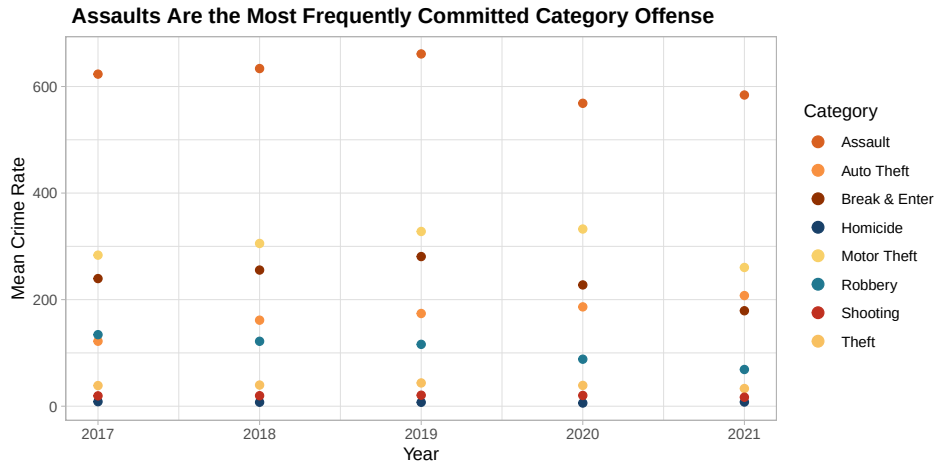

Figure 1: Crime rate in Toronto by category and year

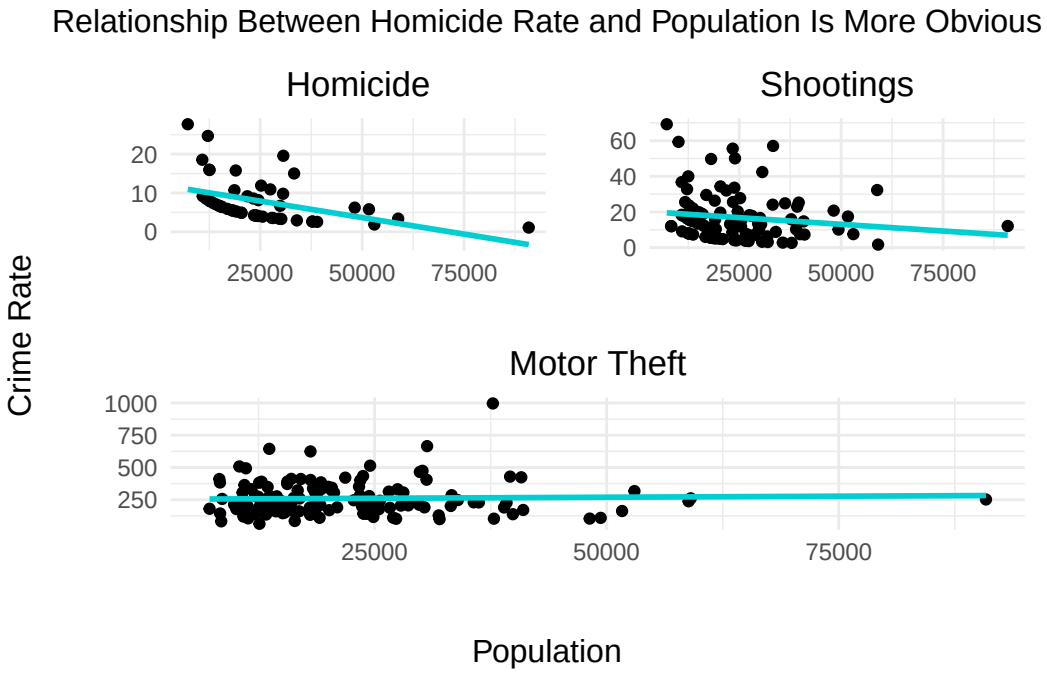

Figure 2: Relationship between crime rate and population

Table 1: Mean crime rate by neighborhood size in 2021

Neighborhood Size	Small	Medium	Large	Giant
Crime Category				
Assault	641.53	499.48	514.11	666.62
Auto Theft	136.01	224.89	201.09	208.64
Break & Enter	223.34	163.58	173.26	186.77
Robbery	119.00	49.31	65.14	77.99
Theft	31.37	30.84	30.87	35.77
Homicide	27.69	11.80	7.35	5.91
Shooting	31.09	20.30	13.69	16.11
Motor Theft	238.48	247.92	264.42	268.32

3. Discussion

Reference

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal*. https://CRAN.R-project.org/package=opendatatoronto.

Kassambara, Alboukadel. 2022. *Ggpubr: 'Ggplot2' Based Publication Ready Plots*. https://CRAN.R-project.org/package=ggpubr.

Lucas, Tim. 2016. *Palettetown: Use Pokemon Inspired Colour Palettes*. https://CRAN.R-project.org/package=palettetown.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Müller, Kirill, and Hadley Wickham. 2022. *Tibble: Simple Data Frames*. https://CRAN.R-project.org/package=tibble.

Pedersen, Thomas Lin. 2022. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Toronto Police Service. 2022a. "Annual Statistical Report." https://data.torontopolice.on.ca/pages/annualstatisticalreport.

———. 2022b. *Toronto Police Service Public Safety Data Portal*. https://data.torontopolice.on.ca/pages/annualstatisticalreport.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2022. *Tidyr: Tidy Messy Data*. https://tidyr.tidyverse.org.

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.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2021. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2020. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.