My title*

My subtitle

Siming Shan

06 February 2022

Abstract

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

2 Data

2.1 Raw Dataset Introduction

I utilized an Annual Victims of crime report("Police Annual Statistical Report - Victims of Crime" 2021) from Toronto Open Data portal, using the `opendatatoronto` package(Gelfand 2020) to have a better understanding of the relationship between peace officers and the communities in Toronto. This data was published by Toronto Police Services, and the last time the data was refreshed on Aug 12, 2021. There are 854 observations in the raw data, each observation has 12 variables - id, index, ReportedYear, Category, Subtype, AssaultSubtype, Sex, AgeGroup, AgeCohort, Count, ObjectId, geometry. This dataset collects victim of crime which crime was reported between 2014 and 2020, and the victim of the crimes are classified as peace officers, other, and unknown. Subtypes of assault against peace officers include Aggravated Peace Officer, Assault Peace Officer, Assault Peace Officer Weapon/Bodily Harm, and Assault Resist Arrest. Since I am only interested in crimes against peace officers, I cleaned the data and extracted essential observations by using `R`(R Core Team 2020), `tidyverse`(Wickham et al. 2019), `dplyr`(Wickham 2021), and `janitor`(Sam Firke 2021).

2.2 Cleaning Process And Cleaned Dataset

The columns "index" and "geometry" of the dataset are all `NA` and `Null`, and the column "ObjectId" is redundant, thus I removed these columns as they did not provide useful information for analysis. Then I modified each column name to make them more organized. Finally, I extracted the observation that only peace officers were victims.

The cleaned dataset has 168 observations, each observations has 9 variables:

- id: Unique row identifier
- reported_year: Year crime was reported
- category: Crime category
- subtype: Crime category subtype
- assault_subtype: Breakdown of assault subtypes
- sex: Sex of identified victim
- age_group: Age group of identified victim, adult or youth
- age_cohort: Age cohort of identified victim
- count: Count of identified victims

*Code and data are available at: https://github.com/SimingShan/STA304-Project-1

3 Conclusion

Reference

Gelfand, Sharla. 2020. *Opendatatoronto: Access the City of Toronto Open Data Portal.* https://cran.r-project.org/web/packages/opendatatoronto/index.html.

"Police Annual Statistical Report - Victims of Crime." 2021. Toronto Police Services. https://open.toronto.ca/dataset/police-annual-statistical-report-victims-of-crime/.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sam Firke, Chris Haid, Bill Denney. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://cran.r-project.org/web/packages/janitor/index.html.

Wickham, Hadley. 2021. *Dplyr: A Grammar of Data Manipulation.* https://cran.r-project.org/web/packages/dplyr/index.html.

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.