Are some women more susceptible to crime than others in Toronto?*

A visualisation of crime rates over the years, and the ages they affect most

Sehar Bajwa

January 24, 2024

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

Table of contents

*Code and data are available at: https://github.com/SEHB2012/crime-victims-toronto

1 Introduction

You can and should cross-reference sections and sub-sections.

The remainder of this paper is structured as follows. Section 2….

2 Data

The first step is to construct a relevant dataset derived from the Police Annual Statistical report on crime victims. This section delves into the characteristics of note and outlines the process undertaken to create a ready-to-analyse dataset, including addressing missing values and standardizing variable names. ## Data characteristics Published by the Toronto Police service on the Toronto Open Data Portal, the original dataset covers all crimes committed against the person, including those deemed unfounded post investigation. Updated annually, the entries are filtered by the reported year, and each year is associated with the following features : type of crime, age cohort of victim, gender, and counts of the same crime. Note that the dataset is not impervious to missing/incomplete data across several categories, and while any incomplete datapoint has been discounted, this causes serious implications for the analysis and is examined later. The original dataset features 1111 datapoints across 9 variables. Age_cohort is a defining feature of this dataset, and refers to one of 8 age bins that are as follows: “>12”, “12 to 17”, “18 to 24”, “25 to 34”,“35 to 44”,“45 to 54”,“55 to 64”,“65+”. Accordingly, efforts were made to attribute recognisable labels to these cohorts, but there was admittedly difficulty distinguishing between the middle aged adults that arguable spanned three of the 9 age cohorts. Therefore, the original identifiers were retained. ## Data cleanup and processing R (R Core Team 2021) was the language and environment used for the bulk of this analysis, alongside the dplyr (Wickham et al. 2021),tidyverse (Wickham et al. 2019), janitor (Firke et al. 2021), heatmap(cite).

I began with cleaning up the data after downloading it using the opendatatoronto package. Following are the steps involved in the primary cleanup stage: Cleaning and standardising column names Selecting relevant categories, these being report_year,subtype,sex, age_cohort,count Removing datapoints with unknown/missing data Filtering to only female victims.

Below are the first ten rows of the cleaned dataset, generated using the package kable.

```
# A tibble: 6 x 3
  report_year age_cohort sum_count
        <dbl> <chr>          <dbl>
1        2014 <12              642
2        2015 <12              602
3        2016 <12              533
4        2017 <12              589
5        2018 <12              608
6        2019 <12              617
```

3 Visualisation

3.1 Changing crime trends in Toronto over the years

3.1.1

3.1.2 Discussion

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

We run the model in R (R Core Team 2022) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.3 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in **?@tbl-modelresults**.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

In **?@fig-ppcheckandposteriorvsprior-1** we implement a posterior predictive check. This shows…

In **?@fig-ppcheckandposteriorvsprior-2** we compare the posterior with the prior. This shows…

Examining how the model fits, and is affected by, the data

Figure 1: **?(caption)**

B.2 Diagnostics

?@fig-stanareyouokay-1 is a trace plot. It shows… This suggests…

?@fig-stanareyouokay-2 is a Rhat plot. It shows… This suggests…

Checking the convergence of the MCMC algorithm

Figure 2: **?(caption)**

References

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

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