Crime Rates in Toronto Neighborhoods By Year and Type*

Yuxuan Yang

04 February 2022

Abstract

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

1 Introduction

First paragraph is going to be motivational and broad.

Second paragraph is about what was done and what was found.

Third paragraph is about implications.

The remainder of this paper is: Section 2 explains the data. Section 4 covers results. . . . . .

2 Data

2.1 Data Source And Collection Methods

This paper uses data on neighborhood crime rates published by the Toronto Police Services on the City of Toronto Open Data Portal; the data was last refreshed on May 6, 2021, and I imported the data as csv file directly using the R package opendatatoronto (Gelfand, 2020). This dataset contains information on numbers of different types of crimes in each of the 140 Toronto neighborhoods from the year 2014 to 2020, together with crime rates of different types of crimes in each neighborhood from 2014 to 2020, where the crime rate is calculated as the crime count per 100,000 population* per year in each neighborhood (cite opendatatoronto here).

The Toronto Police Service did not specify how exactly the crime cases in this dataset were processed, but it mentions that the crime rates were calculated using "the standard definition by StatsCan" (cite here). In addition, crimes can be reported to the Police and collected through many channels. For example, citizens can voluntarily report crimes they have experienced or observed. It is also possible that the Police will discover crimes on site through daily patrols. Nevertheless, the fact that crimes can be self-reported indicates the possibility of crimes being under-reported, which could result in potential bias of number of crime cases on record.

2.2 Population and Samples

The population of this dataset is the total population in each Toronto neighborhood in each year from 2014 to 2020. Note that the total population in each neighborhood in this dataset is only a projection estimated by Environics Analytics instead of the true value. The sampling frame is all the crime cases reported and recorded in this dataset.

*Code and data are available at: https://github.com/Yuxuan-Yang-Maggie/starter_folder.

2.3 How Respondents Were Found

Respondents in our Neighborhood Crime Rates dataset are the reporters of crime cases in each neighborhood. Since crime cases are mainly collected through either self-reporting or police discoveries, our respondents here are discovered by the Toronto Police Service if evidences of crime are observed. In terms of crime data, the action of non-response will cause crime cases in Toronto to be under-reported and under-represented, leading to significant bias in the results of our crime data analysis. ## Major Strengths And Weaknesses of The Data Source

Our data is of penguins (Figure 1).

```
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.
```

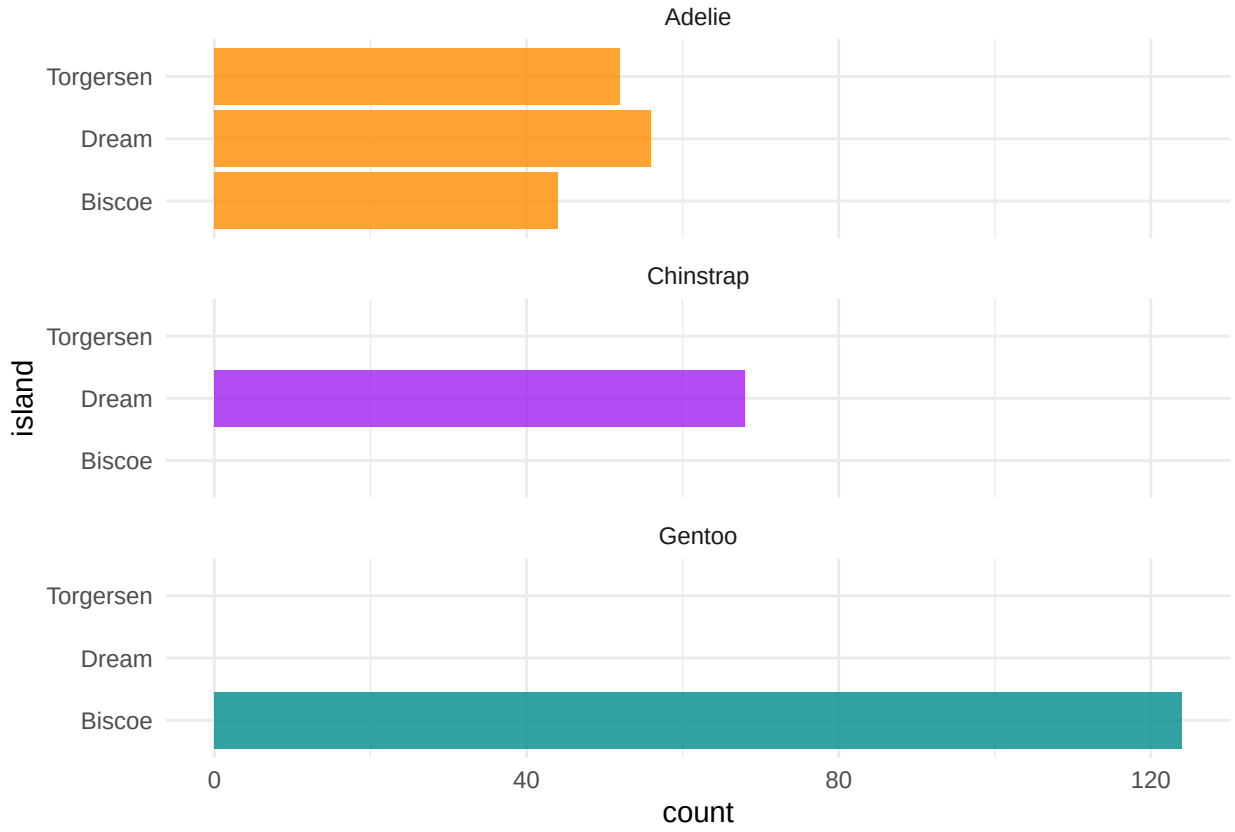

Figure 1: Bills of penguins

Talk more about it.

Also bills and their average (Figure 2). (Notice how you can change the height and width so they don't take the whole page?)

```
## Warning: It is deprecated to specify `guide = FALSE` to remove a guide. Please
## use `guide = "none"` instead.
```

Talk way more about it.

3 Model

$$Pr(\theta|y) = \frac{Pr(y|\theta)Pr(\theta)}{Pr(y)} \tag{1}$$

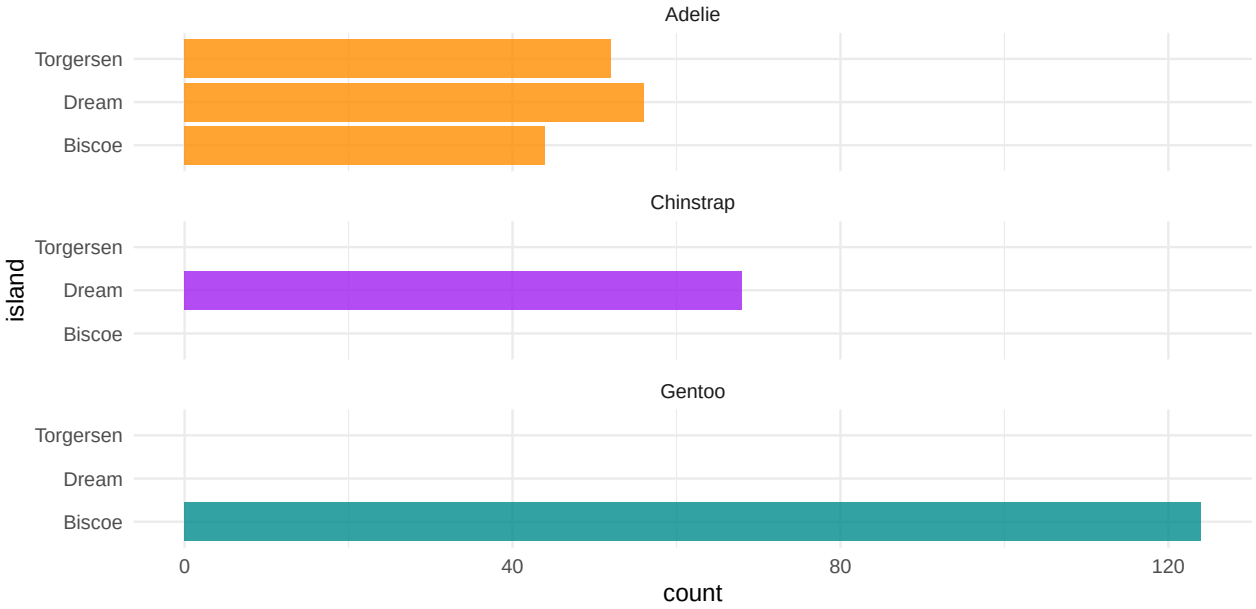

Figure 2: More bills of penguins

Equation (1) seems useful, eh?

Here's a dumb example of how to use some references: In paper we run our analysis in `R` (R Core Team 2020). We also use the `tidyverse` which was written by Wickham et al. (2019) If we were interested in baseball data then Friendly et al. (2020) could be useful.

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

4 Results

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 details

References

Friendly, Michael, Chris Dalzell, Martin Monkman, and Dennis Murphy. 2020. *Lahman: Sean 'Lahman' Baseball Database*. https://CRAN.R-project.org/package=Lahman.

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