Investigating the motor vehicle collisions in Toronto*

Installing more red light cameras do not stop traffic accidents from happening

Alicia Yang

04 February 2022

Abstract

Vehicle collisions occur constantly in big citys. They sometimes could be severe issues when resulting in damage in public and personal properties, injuries, or even death. This report aims to investigate the motor vehicle collisions in Toronto that result in injuries or deaths from 2006 to 2020. By exploring the trend of these traffic collisions in respect of number and location, I find that the total number of accidents has a descending trend. However, the vehicle collisions taking place at intersections has not been reduced despite that more red light cameras are installed. The data used in this report is obtained from the City of Toronto Open Data Portal.

Contents

1 Introduction **2**

2 Data **2**
2.1 Motor Vehicle Collisions involving Killed or Seriously Injured Persons 2

3 Model **6**

4 Results **8**

5 Discussion **8**
5.1 First discussion point 8
5.2 Second discussion point 8
5.3 Third discussion point 8
5.4 Weaknesses and next steps 8

Appendix **9**

A Additional details **9**

B References **10**

*Code and data are available at: https://github.com/Alicia-y/Telling-stories-with-data.

1 Introduction

As one of the busiest and most modern city in Canada, Toronto constantly faces a large number of citizens, and, inevitably, heavy traffic. The road condition in Toronto is busy most of the time. The vast majority of roads and buildings under construction that take place almost everywhere narrow the roadway of vehicles and turn off traffic lights. Heavy traffic is likely to be one of the main reasons behind car accidents, because it increases the road rage and encourages the vehicle drivers to make dangerous moves to save time. Vehicle collisions have always brought large financial costs to the involving parties, and even more, they even brought injuries and deaths. In just 2019, traffic accidents resulted in more than 1700 deaths and more than 140000 injuries.

The Motor Vehicle Collisions data of Toronto provides a record of the past severe vehicle collisions in Toronto which result in injuries or deaths, and with the help of statistical data analysis, it can possibly offer insights on the ways to reduce this type of accident. This report investigates the motor vehicle collisions in Toronto from 2006 to 2020, with a focus on the ones result in injuries or deaths. In Section 2, the Motor Vehicle Collisions dataset is introduced and organized to visualize the change in the annual number of accidents, as well as comparing the frequency of the accidents by district, accident type, and level of injury. Then, I discussed about the possible implications of these visualizations.

However, numbers of annual motor vehicle collisions demonstrate a generally descending trend from 2006 to 2020.

You can and should cross-reference sections and sub-sections. For instance, Section 2. R Markdown automatically makes the sections lower case and adds a dash to spaces to generate labels, for instance, Section 5.1.

2 Data

2.1 Motor Vehicle Collisions involving Killed or Seriously Injured Persons

To investigate the motor vehicle collisions in Toronto, I obtained the dataset "Motor Vehicle Collisions involving Killed or Seriously Injured Persons" from the Toronto Open Data Portal. Provided by the Toronto Police Service, this dataset includes all vehicle collisions occurred in Toronto between 2006 and 2020 that result in injuries or deaths. It contains 16840 observations of collisions and 54 variables with all kinds of information of each collision. This report focuses on six of these variables which are: year, accident location, district, street1(of the collision occurred), street2(of the collision occurred), and classification of accident. By using R packages "tidyverse" (Wickham et al. 2019a), "janitor" (Firke 2021), "kableExtra" (Zhu 2021) and "knitr" (Xie 2021), an extract of the cleaned dataset (Table 1) is shown below.

Table 1: Extracting the first ten rows from the Motor Vehicle Collisions data

Year	Street1	Street2	Collision location	District	Classification of Accident
2006	BLOOR ST W	DUNDAS ST W	At Intersection	Toronto and East York	Fatal
2006	BLOOR ST W	DUNDAS ST W	At Intersection	Toronto and East York	Fatal
2006	MORNINGSIDE AVE	SHEPPARD AVE E	At Intersection	Scarborough	Fatal
2006	MORNINGSIDE AVE	SHEPPARD AVE E	At Intersection	Scarborough	Fatal
2006	EGLINTON AVE E	COMMONWEALTH AVE	NA	Scarborough	Fatal
2006	EGLINTON AVE E	COMMONWEALTH AVE	NA	Scarborough	Fatal
2006	EGLINTON AVE E	COMMONWEALTH AVE	NA	Scarborough	Fatal
2006	OSSINGTON AVE	DUNDAS ST W	At Intersection	Toronto and East York	Non-Fatal Injury
2006	OSSINGTON AVE	DUNDAS ST W	At Intersection	Toronto and East York	Non-Fatal Injury
2006	ISLINGTON AVE	DIXON RD	At Intersection	Etobicoke York	Fatal

Table 1 shows the first ten rows of the Motor Vehicle Collisions data. Variable "Year" indicates the year in which the accident occurs. Variables "Street1" and "Street2" provides the name of the street where the accident takes place. Variable "District" indicates its district. Variable "Collision location" includes 8 types:

at intersection, intersection related, at/near private drive, laneway, non intersection, overpass or bridge, trail, and underpass or tunnel. Variable "Classification of Accident" involves fatal and non-fatal injury.

I'm interested in finding the trend of motor vehicle collisions from 2006 to 2020 and the types of injuries they result in. It reveals the traffic control in Toronto, and possibly the quality and knowledge of vehicle owners throughout these years.

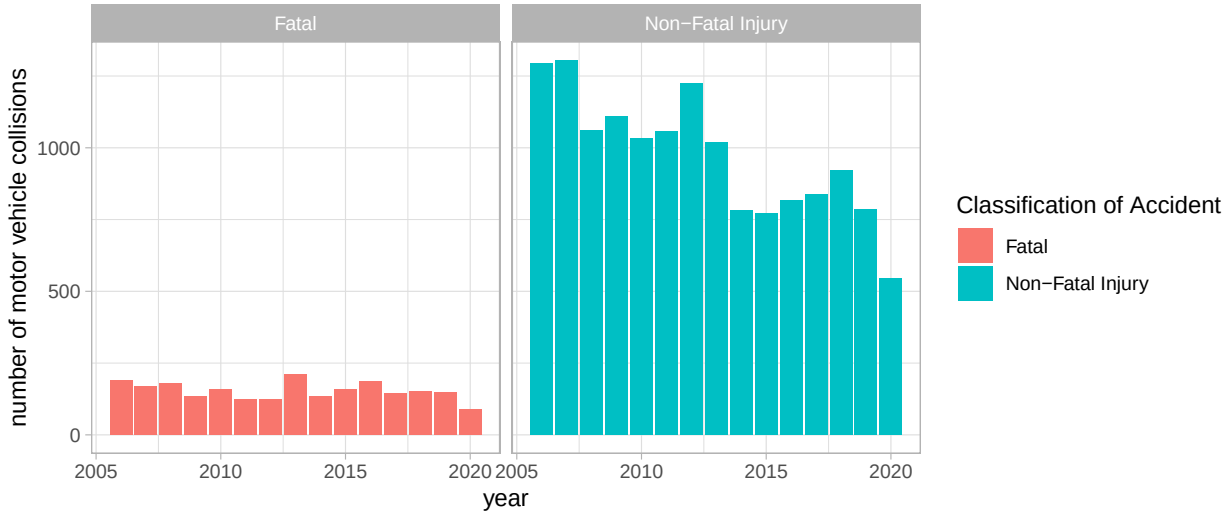


Figure 1: Comparing the number of traffic collisions each year from 2006 to 2020

Figure 1 shows a descending pattern for the number of car accidents involving non-fatal injuries from year 2006 to 2020. The number of fatal accidents does not increase nor decrease much. In general, the number vehicle collisions in Toronto has largely decrease, indicating that when combining this two factors(number and classification of accident), Toronto has done a better job on traffic control in general. I'm also interested in the trend of motor vehicle collisions if other factors are taken into considerations.

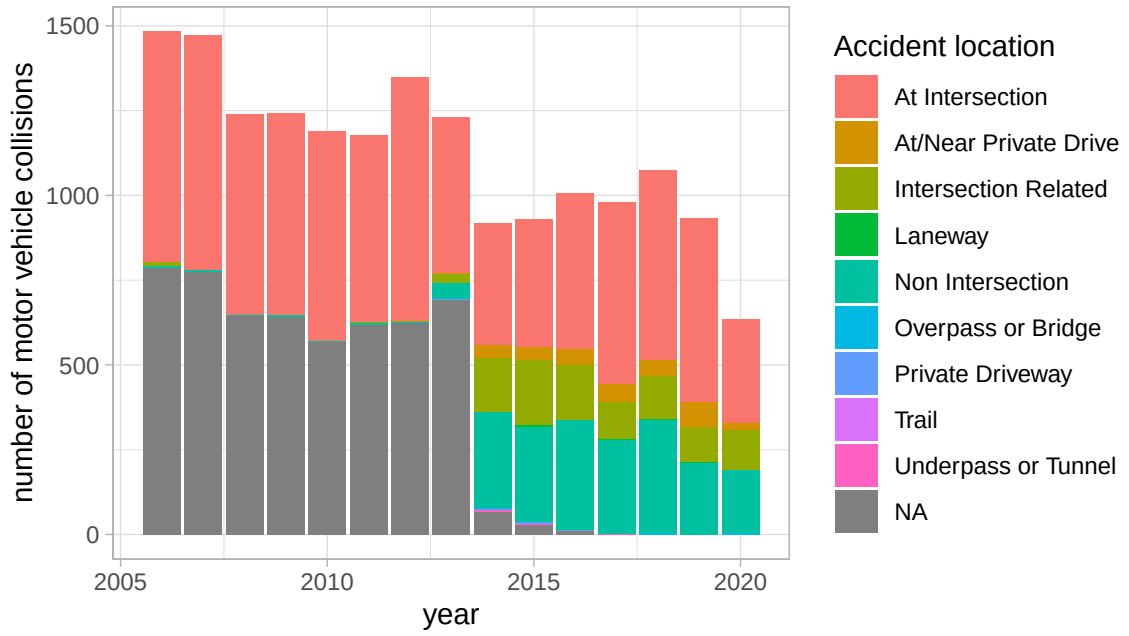
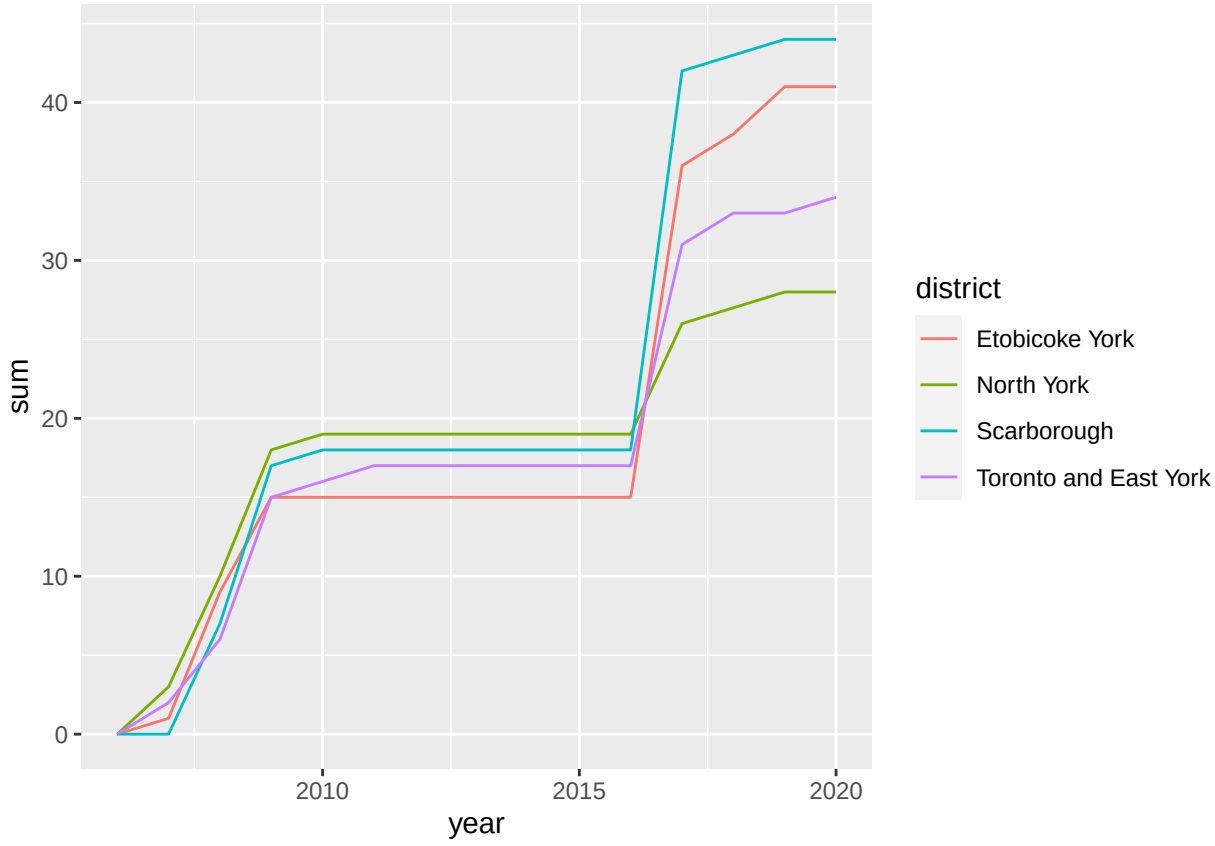


Figure 2: Comparing the location of traffic collisions each year from 2006 to 2020

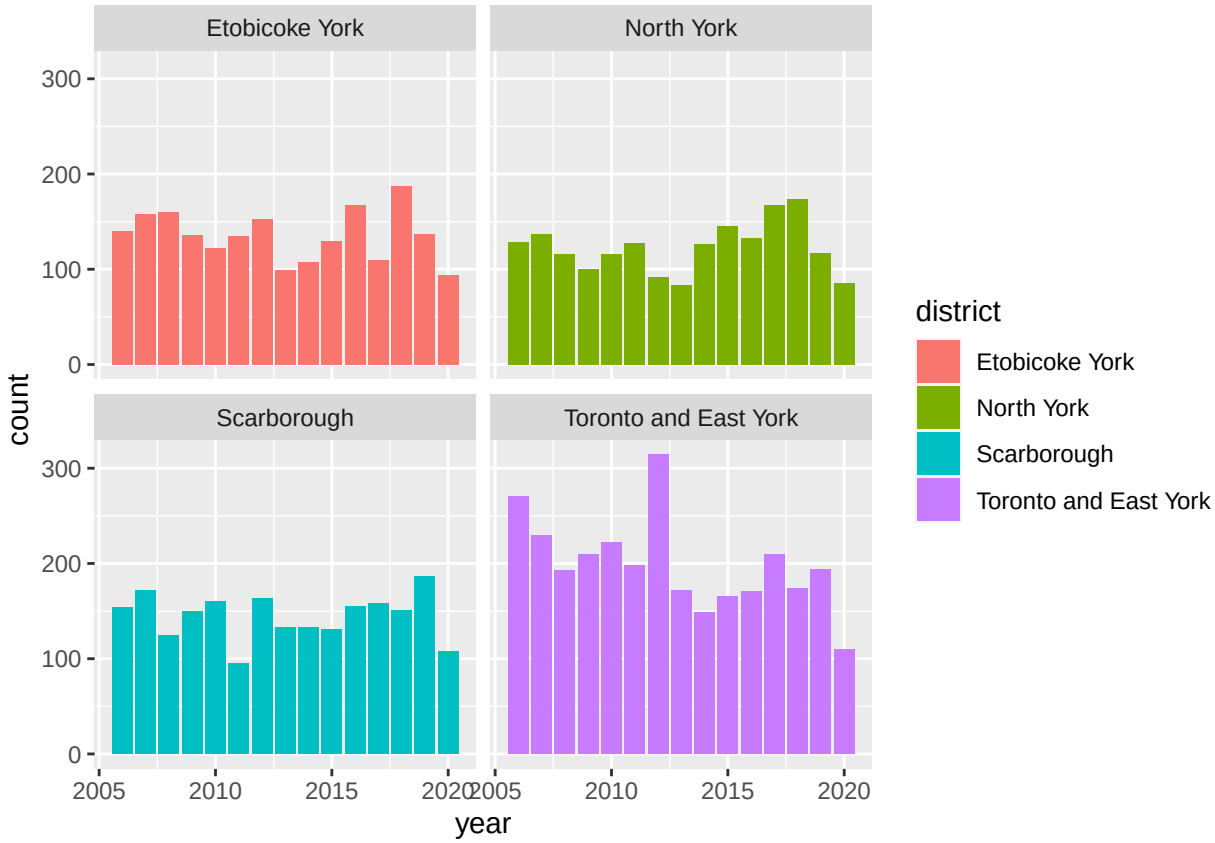
Figure 2 illustrates the where these vehicle collisions usually take place from 2006 to 2020. Before 2013, half of the data doesn't come with accident location, which is understandable because back in the days there wasn't an easy approach to record and save the data, so they often got lost. A large proportion of the accidents take place either at intersections or is intersection related. The other seven accident locations types add up to less than half of all vehicle collisions. At intersections, a lot more rule breaking behaviors can be committed by drivers such as ignoring stop signs or traffic lights, some drivers even claimed that they rarely yielded. These behaviors would result in collisions with other vehicles and the pedestrians.

Motor vehicle collisions are most likely to take place at intersections. If we could take measures to restrict the driver's rule breaking behaviors, the numbers of car accidents would drop a lot in general.

```
red_light_camera <- read_csv(here::here("inputs/data/red_light_cameras.csv")) %>%
  clean_names() %>%
  select(activation_date, district) %>%
  drop_na(activation_date) %>%
  drop_na(district)
red_light_camera$year <- vector("numeric", length = 199)
for (i in 1:199){
  red_light_camera$year[i]=format(as.Date(red_light_camera$activation_date[i], format="%Y/%m/%d"),"%Y")
}
red_light_camera <- red_light_camera %>%
  filter(year != 2021)
cumulative <-tibble(
  year=c(2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2006
  sum=c(0, 2, 6, 15, 16, 17, 17, 17, 17, 17, 17, 31, 33, 33, 34, 0, 0, 7, 17, 18, 18, 18, 18, 18, 18, 18
  district=c("Toronto and East York", "Toronto and East York", "Toronto and East York", "Toronto and Eas
)
ggplot( aes(x=year, y=sum, group=district, color=district), data=cumulative) +
    geom_line()
```




```
co <- vehicle_collision %>%
  filter(accloc == "At Intersection" | accloc == "Intersection Related") %>%
  drop_na(district)
for(i in 1:9039){
  if(co$district[i] == "Toronto East York"){
    co$district[i] = "Toronto and East York"
  }
}
ggplot(data=co,aes(x=year, fill=district))+geom_bar()+facet_wrap(~district)
```




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

Talk more about it.

Also bills and their average (Figure 4). (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

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

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. (2019b) If we were interested in baseball data then (`???`) could be useful.

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

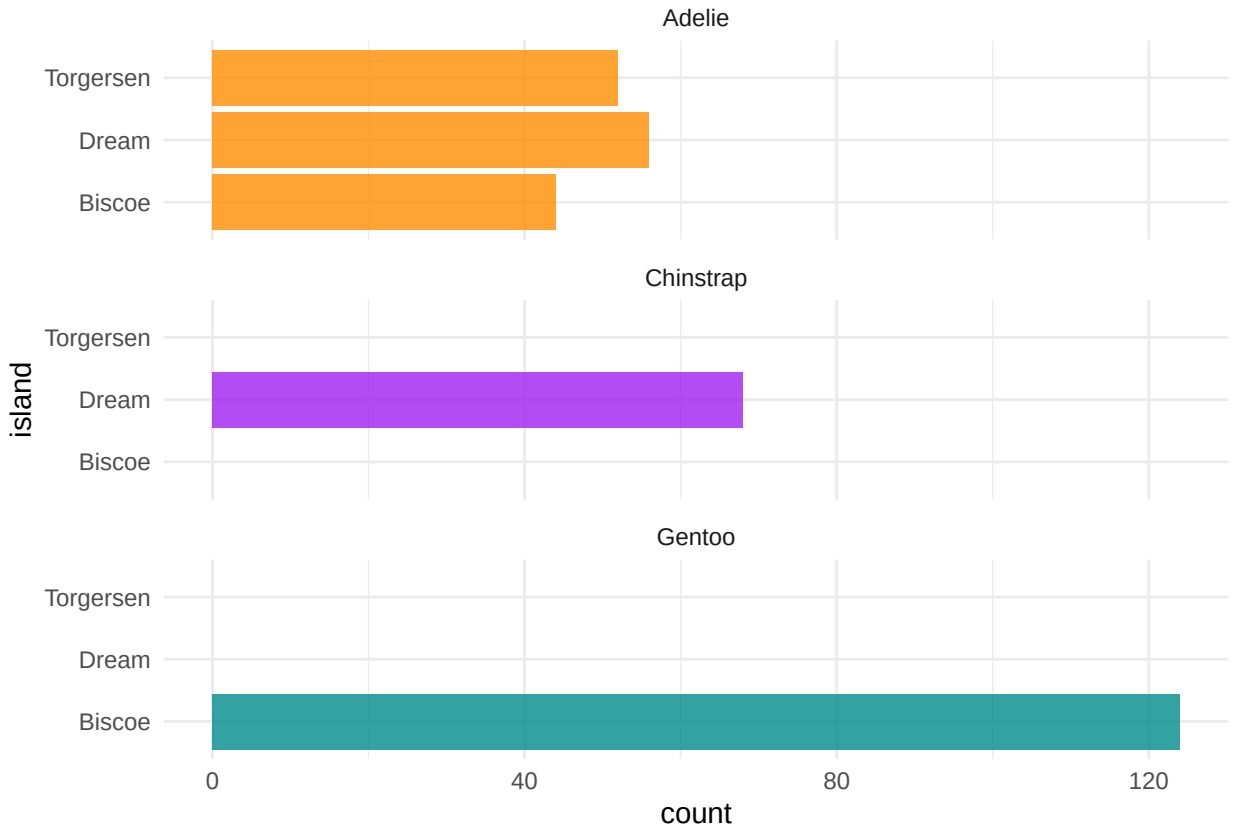


Figure 3: Bills of penguins

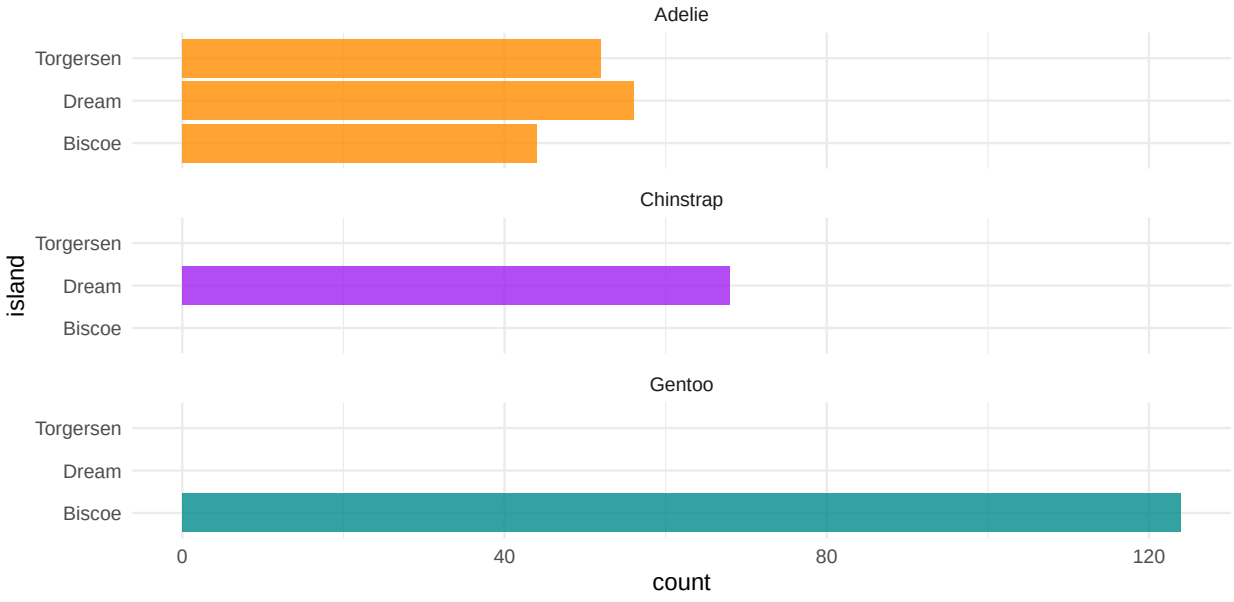


Figure 4: More bills of penguins

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

B References

https://tc.canada.ca/en/road-transportation/statistics-data/canadian-motor-vehicle-traffic-collision-statistics-2019

https://newcanadianlife.com/why-does-toronto-have-so-much-traffic/

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

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. 2019a. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

———. 2019b. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

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

Zhu, Hao. 2021. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*.