

# Canada and USA facing an increase in police related deaths

James Bao

16 December 2020

## Abstract

We analyzed police use of lethal force from data sets collected by the Fatal Encounters Project and CBC in order to analyze the severity of police killing within Canada and the United States of America. We found a concerning trend in an annual increase of death of police cases every year, after taking the annual increase in population into account. Although Canada has less death per million and a slower rate of increase of death per million every year compared to the US. We also found that variables can alter the chances a person would die from gun shot versus other unintentional uses of lethal force, although a majority of victims are killed from gun shot in both countries. These findings can be important to analyzing Canada's police use of force reports and suggest that the Canadian police system are not immune to the faults American policing system is accused of.

**Keywords:** Police, Deadly-force, Canada, USA

The (currently incomplete) code generated in this markdown file can be found on my Github.

## 1 Introduction

On May 25 2020, the death of George Floyd became another name on a long list of people who has died at the hands of police. This incident was caught on camera and sparked riots all across the country, as people called for police reform throughout the US. 2 days later, Regis Korchinski-Paquet in Toronto died in a controversial incident after police became involved, and a similar protest sparked within the city of Toronto, with the Toronto City council voting for changes in policing policies (Boisvert (2020)). These are just two deaths among the many people of a visible minority who has died within America and Canada, which is just a portion of all the people who die during an encounter with law enforcement within America and Canada, most recently a police shooting resulted in a death of an infant in Canada. These deaths are a few among over 500 deaths in Canada and 18000 in USA over the past 20 years. This paper attempts to analyze and compare the data from both Canada and USA and model a trend in police related deaths in Canada and USA, as well as model how various external variables could effect the police's use of lethal on people.

Throughout the paper, we analyze the consistent increase of death per million in both Canada and USA, as well as the demographics of police related deaths in each country. We analyzed the statistical significance of our models, the p-values and their implications. Overall, we found a trend of increasing police related death in both the countries, with USA reaching almost 3 people per a million people dying and Canada going close to 1 death per million. Furthermore, we selected 4 explanatory variables and built a logistic regression model to investigate whether age, gender, race and location is a significant indicator on how a victim is killed, however we found no statistical significance overall. We further investigated demographics of each country and identify if any demographics are over represented within our data sets and we identified males and certain minorities have a disproportional representation within our data set.

In the following sections of this paper, we will discuss the two data sets that was used, how it was collected, and summarize key variables used from these data sets. We will then discuss the models this paper uses and how they were chosen, followed by a summary of the results from our charts and models. Finally we will discuss the significance and implications of our findings and conclude with our weaknesses and future steps.

## 2 Data

In order to complete our analysis and models, we used two data sets detailing police related deaths in both Canada and USA. For this, we used data collected by Fatal Encounters for US data and by CBC for Canadian data. We will discuss how both the data set was collected and highlight certain characteristics of both these data sets.

### 2.1 Data Collection

#### 2.1.1 USA Data set

The Fatal Encounters Project is a publicly funded organization started in 2012 that is recognized as a charity organization in the United States of America. The organization was started by D. Brian Burghart, after he realized that no official federal organization tracks deaths of people at the hands of police. At the time of writing this report, Fatal Encounters has collected 29192 names, locations, situations and more of people who has been killed by the police. The data was collected by D. Brian Burghart along with other paid researchers and volunteers through primarily 3 methods. They obtained information by gathering law enforcement data through the Freedom of Information Act requests and other public record requests, they crowd sourced internet and news paper searches, and finally they cross checked their data with other reputable data sets such as The Counted by the Guardian and the Fatal Force project by The Washington Post (this project won the Pulitzer Prize in 2016 (“2016 Pulitzer Prizes” 2016)) among many others (Finch et al. (2019)). The data starts from Jan 1, 2000 and is updated weekly in order to maintain the most update to information possible.

#### 2.1.2 Canada Data set

The Canadian data set containing information on police related deaths is collected by the Canadian Broadcasting Corporation (CBC), and according to the news station’s website, it is “the first country-wide database of every person who died or was killed during a police intervention (Marcoux and Nicholson (2018))”. Similar to the United States of America, Canada does not provide information on police related deaths on an official federal government level, and therefore other organizations such as CBC has collected information across the country. The project originally began in 2017 and lasted 6 months, concluding with a list of 461 names ranging from 2000 to 2017 (Marcoux and Nicholson (2018)), and this data was updated again in the summer of 2020 to include a total of 556 people. The team of researchers at CBC collected the information by going through independent investigator reports, coroner reports, court records, news reports and family interviews (Marcoux (2018)). In contrast to Fatal Encounters, this CBC data set only contains incidences where police purposefully used force, and as such incidences such as in-custody deaths, suicide/accident while evading police or accidental deaths were left out.

#### 2.1.3 Data Characteristics

As mentioned before, both the data sets used in this paper are collected through research and requests for information from official sources. However, these sources are not official statistics and are likely to not be a full representation of all the data across either countries. Some low profile deaths may have escaped the attention of any public resources and would be very difficult to locate through research. Due to the expense of collecting and processing the large amounts of data, these data sets could contain some mistakes. As stated by the founder of Fatal Encounters, their data set is not comprehensive and are missing data. The purpose of this project was to identify the names and information about the people who has died, rather than for academic research. Similarly, CBC has stated a similar sentiment regarding the comprehensiveness of their data set. However, both of these data sets are commonly used when analyzing statistics regarding police related deaths.

Table 1: Count of USA victims by race

Race	Count
Asian	324
Black	4523
Indigenous	233
Latin American	2817
Middle Eastern	38
Unknown	3745
White	6546

A noticeable strength of these data sets however is lack of bias. Since these projects are started with the mission objective of finding as many cases of police use of deadly force as possible, it is reasonable to assume that researchers will attempt to find every case possible in order to represent the entire picture. Since both data sets used are crowd sourced, the public can and are encouraged to provide information on missed data, and the responsible organization will use reliable sources to confirm new names to be added.

While both of these data sets are assumed to be unbiased and reasonably accurate, they are collected with slightly different purposes in mind. The Fatal Encounters project attempts to collect data on any person who has died after a police officer has been involved, both on and off duty and regardless of intent to use lethal force. The CBC fatal police encounters data set is focused on police use of force, therefore the data collected are specifically cases where the police purposefully exerted force, both lethal and non-lethal on an individual who died as a result of the use of force. In order to compare the two data sets in a meaningful way, we removed all the cases where use of force was not intended from the Fatal Encounters project’s data set. As a result our two data sets now contain data that are more similar and the Fatal Encounters data set will not be bloated with extra data.

#### 2.1.4 Costs

Collecting data is expensive, and more so in this case as the process is tedious, slow and often depends on third parties to cooperate. A large part of both organizations’ collection process involved obtaining public records through various means, which involved filing requests and waiting. It is hard to estimate the cost of collecting a data set of this size as many steps of the process contain variable costs, and the size of both the teams are unknown. However, due to the rarity of strong data sets regarding this topic, it is safe to assume it is an extremely costly process.

## 2.2 Data Features and Visualization

The two data sets combine to almost 30000 data points, and share 35 unique variables across both data sets. Between the two data sets, we chose 5 variables: Date of death, Age, Gender, Race, Province or State, Police department involved, and the cause of death. To filter out the unwanted data, we used the “intended use of force field” to remove the names that were not caused by intentional use of force.

### 2.2.1 USA Data from Fatal Encounters

The Fatal Encounters project has collected a total of 29192 names at the time the data set was downloaded, and contains 27 variables. Among these 27 variables, 3 variables are not recommended to analysis (labeled internal use only) and one variable is in progress (labeled developing).

The variables we represent are as below:

Table 2: Count of USA victims by gender

Gender	Count
Female	946
Male	17254
Transgender	15
Unknown	11

Table 3: Count of cause of death in USA

Cause.of.Death	Count
Asphyxiated/Restrained	258
Beaten/Bludgeoned with instrument	172
Burned/Smoke inhalation	22
Chemical agent/Pepper spray	32
Drug overdose	41
Fell from a height	7
Gunshot	16587
Medical emergency	139
Other	12
Stabbed	15
Tasered	892
Undetermined	41
Vehicle	8

Table 4: Age summary of USA victims

Statistic	Value
Min.	0.25000
1st Qu.	25.00000
Median	34.00000
Mean	35.59942
3rd Qu.	44.00000
Max.	107.00000
NA's	319.00000

Table 5: Death counts by state

State	Count
CA	3224
TX	1583
FL	1254
AZ	686
GA	664
IL	576

Table 6: Death counts by police department

State	Count
Los Angeles Police Department	381
Los Angeles County Sheriff's Department	298
Chicago Police Department	282
City of New York Police Department	253
Phoenix Police Department	206
Houston Police Department	202

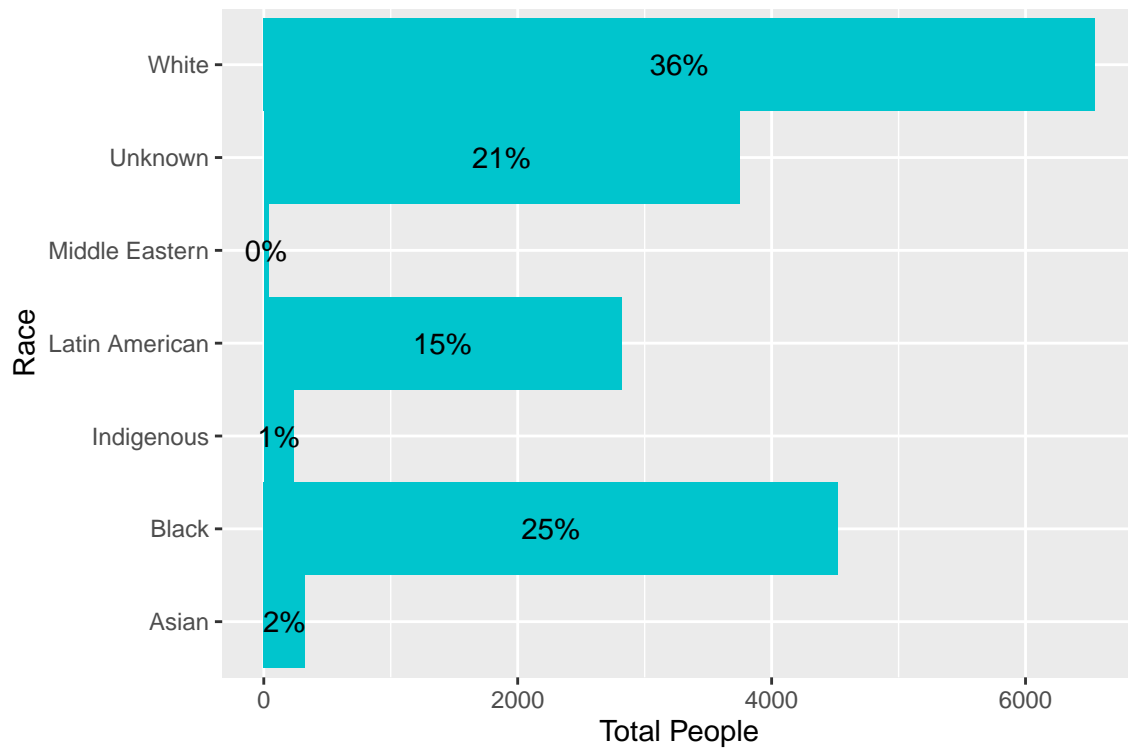


Figure 1: Distribution of the race of Americans who were killed by police

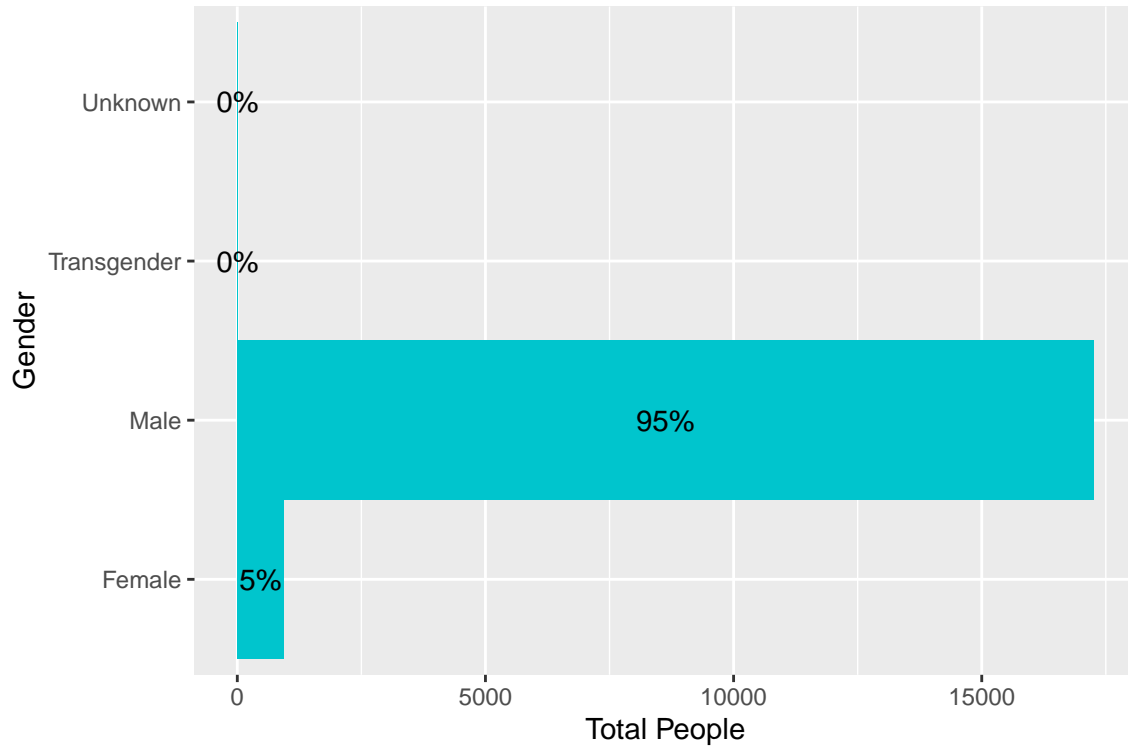


Figure 2: Distribution of the gender of Americans who were killed by police

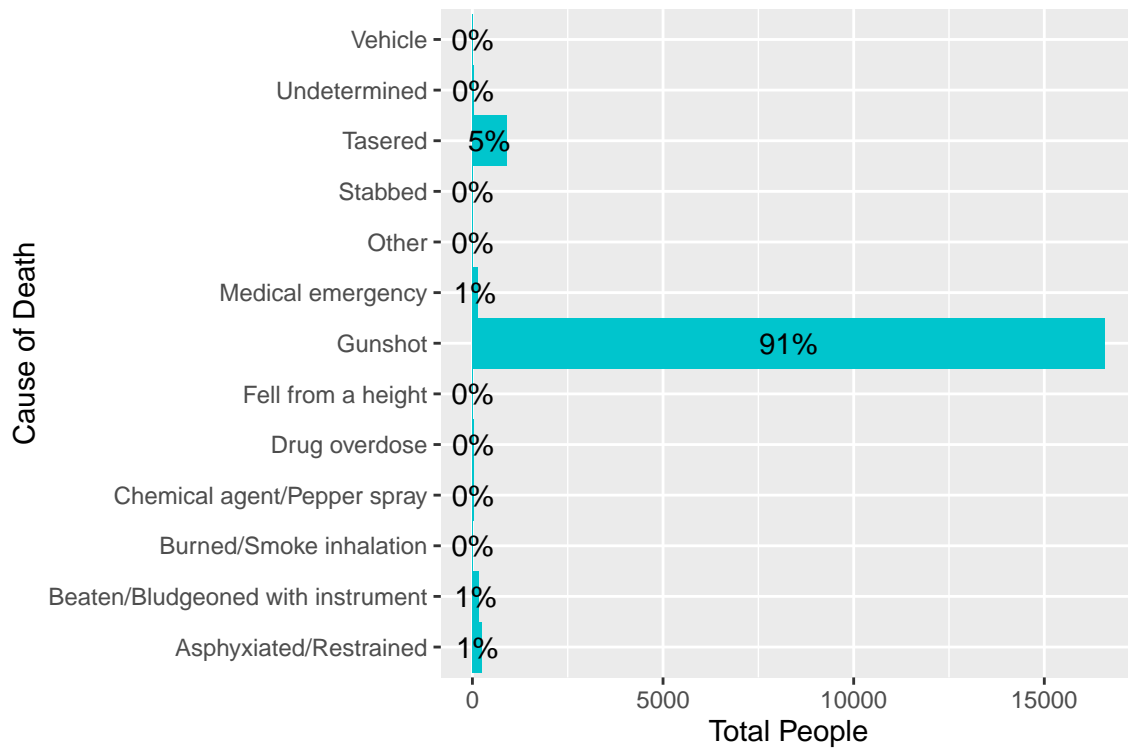


Figure 3: Distribution of the Cause of death for US

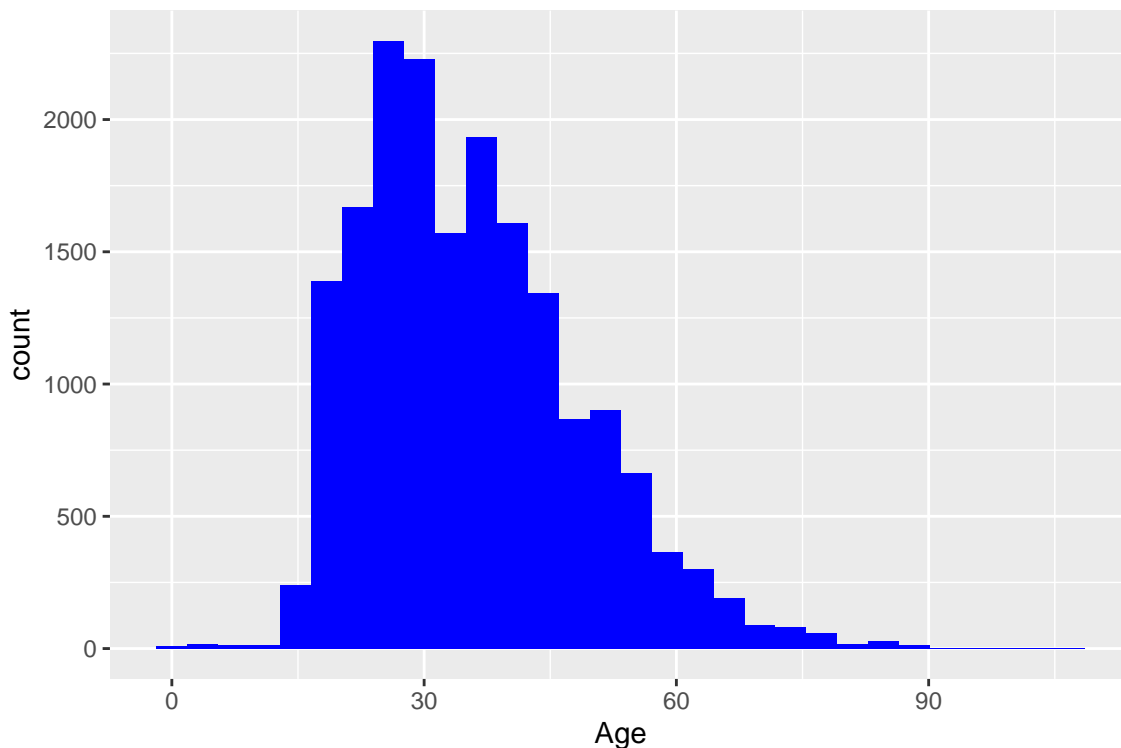


Figure 4: Distribution of the age of Canadians who were killed by police

When we were processing the data set from Fatal Encounters, we had to make both variable names as well as category types. We decided to use the Canadian data set as our standards and updated variable naming accordingly. Furthermore, in order to compare our data, we needed the category naming to be consistent as well (ex Arab vs Middle Eastern).

From Figure 1, we can see that approximately 36% of the victims were white, and 25% were black, while 21% of the victims race was unknown. About 95% of the victims were male, and 5% were females, and 0% were transgender or other (Figure 2). The oldest victim was 107 years old, and the youngest was less than 1 years old, with the mean age of victims being 35.5 years old, with over 300 unknown ages (Table 5). The states with the most people killed were California at over 3000 people between 2000 and 2019, with Texas and Florida following behind at over 1000 deaths each. Most police departments had relatively low death counts, with the exception of police departments from major cities such as Los Angeles, Chicago and New York. 91% of the victims died from gunshot.

## 2.3 Canada Data from CBC

CBC has collected 526 data points with 27 variables. Listed below are more details regarding the variables that were chosen for this paper.

Table 7: Count of Canadian victims by race

Race	Count
Asian	27
Black	46
Indigenous	85
Latin American	3
Middle Eastern	5
Unknown	127
White	233

Table 8: Count of Canadian victims by gender

Gender	Count
Female	19
Male	507

Table 9:

State	Count
ON	171
BC	108
QC	92
AB	88
MB	24
SK	20

Table 10: Count of cause of death in Canada

Cause.of.Death	Count
Gunshot	379
Intermediat weapon	14
Other	33
Physical force	10
Restraint	73
Unknown	17

Table 11: Age summary of Canadian victim

Statistic	Value
Min.	15.00000
1st Qu.	27.00000
Median	35.00000
Mean	36.53786
3rd Qu.	44.00000
Max.	77.00000
NA's	11.00000



Table 12:

State	Count
RCMP	136
Toronto Police Service	54
Service de police de la Ville de Montr��al	33
S��ret�� du Qu��bec	31
Edmonton Police Service	27
Calgary Police Service	26

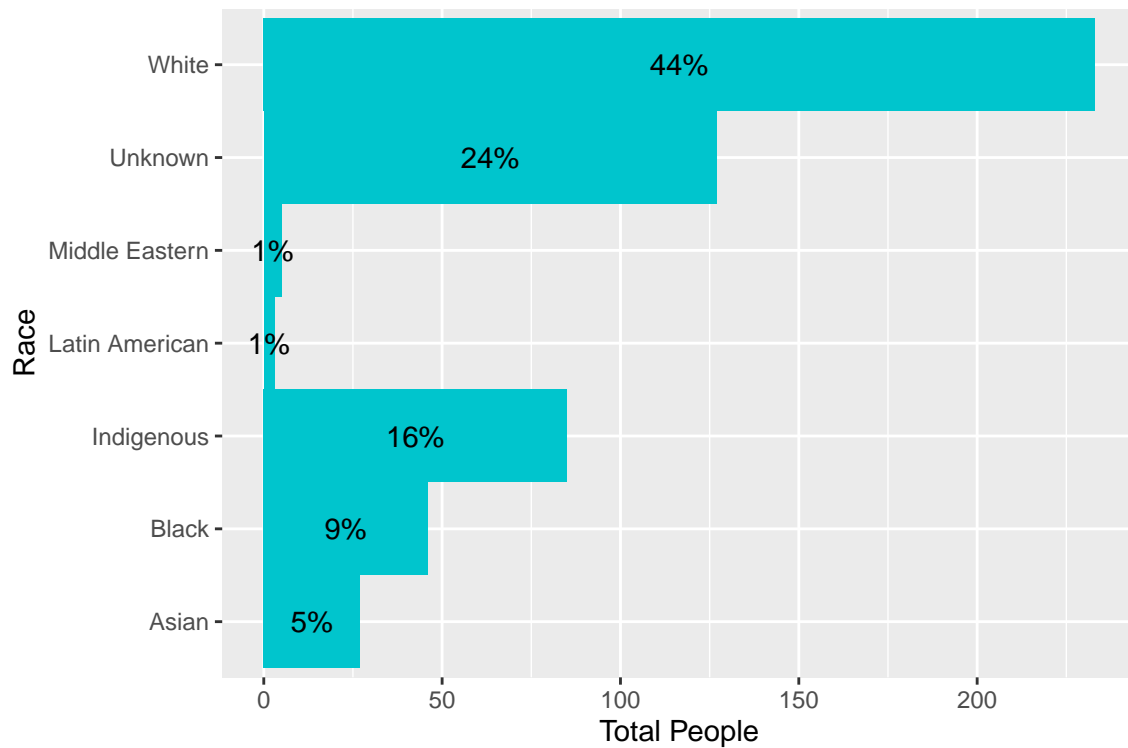
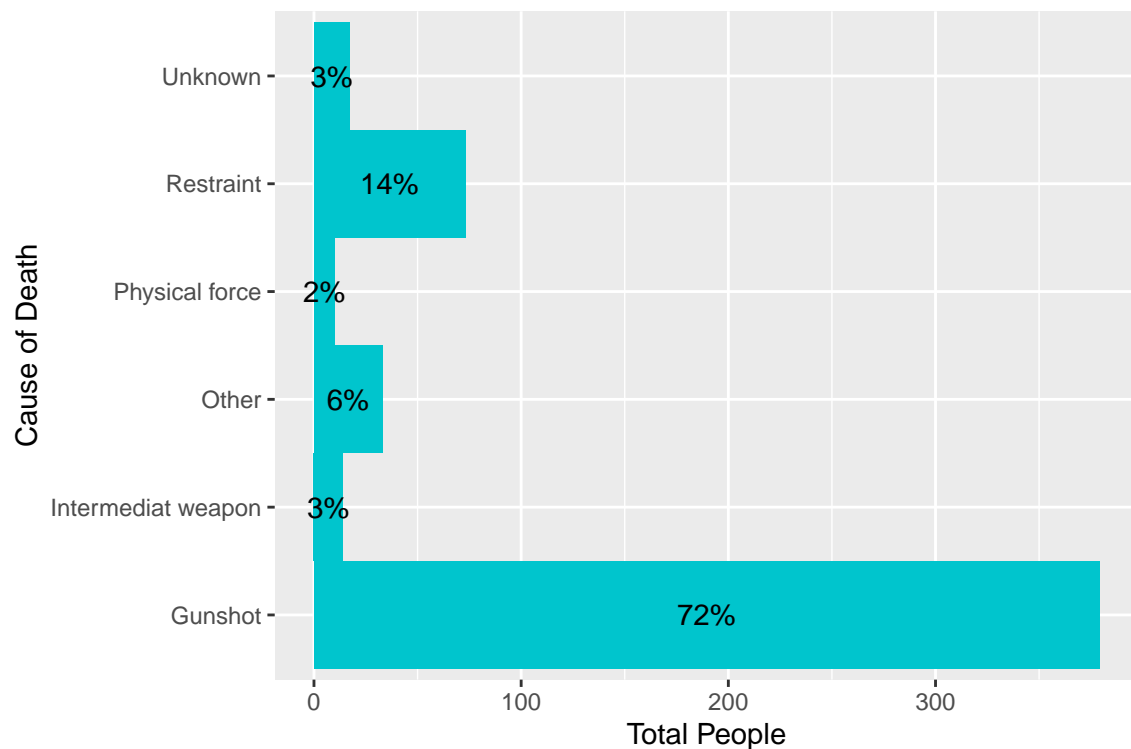
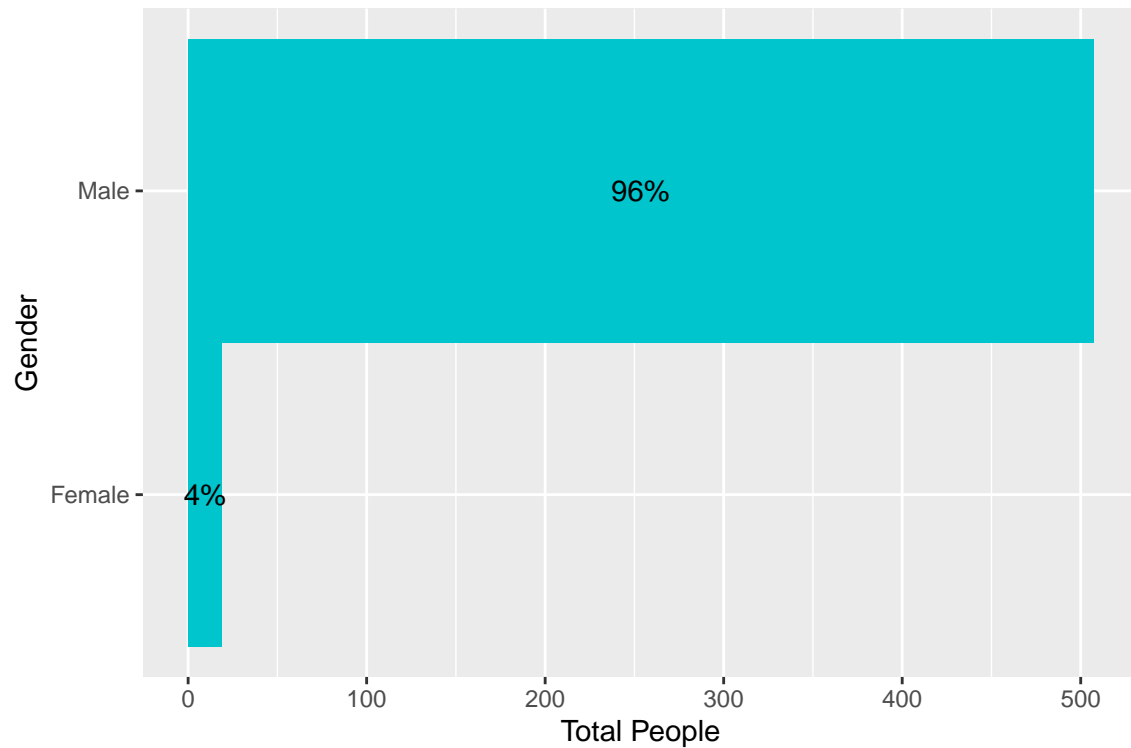


Figure 5: Distribution of the race of Canadians who were killed by police



The CBC data set showed 44% of all victims were white, 24% were unknown, and 16% were indigenous. The average age was 36.5 years old, with the oldest victim being 77 years old and the youngest being 15 years old, and 11 victims had an unknown age. 96% of the victims were male, and 4% were females. Ontario and British Columbia both had over 100 deaths, at 171 and 108 respectively. The Royal Mounted Canadian police is responsible for the most police related deaths in Canada at 168 deaths. 72% of deaths were a result of a

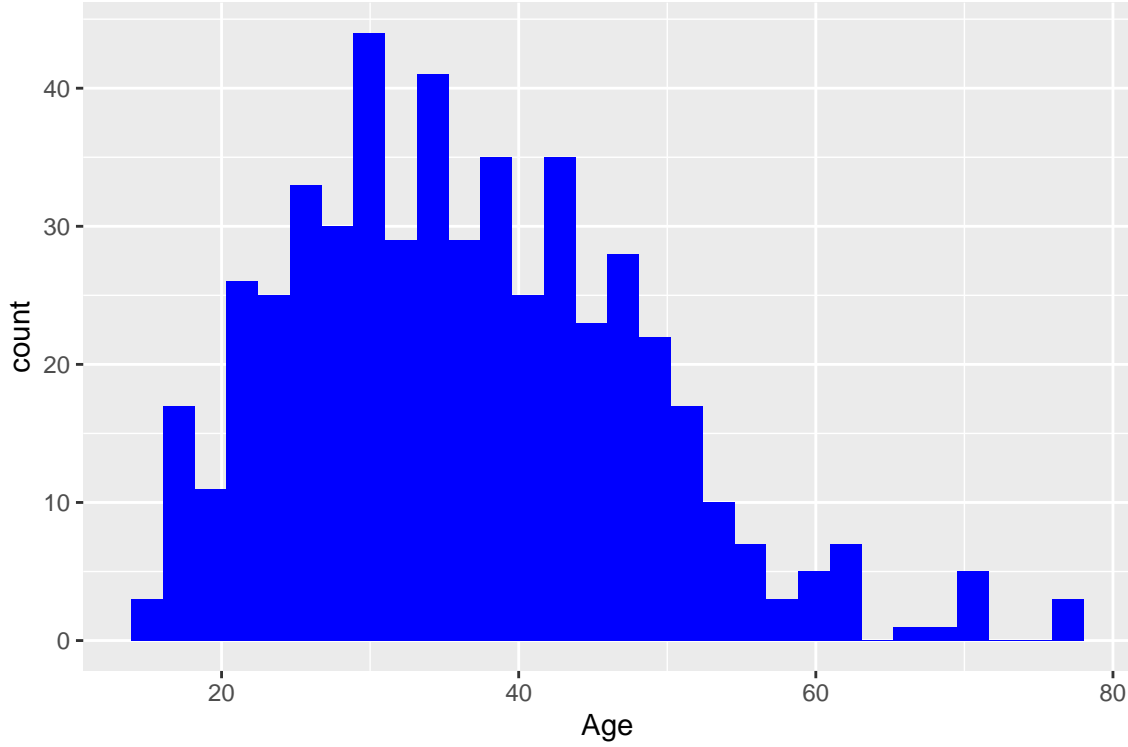


Figure 6: Distribution of the age of Canadians who were killed by police

gun shot.

### 3 Model

In this paper, we used 2 models on both our data sets to gather information for analysis. First we used a simple linear regression in order to model and predict death per million in the year 2020 using data from the last 19 years. Our second model is a multivariate logistic regression to model the affect of race, age, state/province and police department on the most likely type of forced used that lead to loss of life.

#### 3.1 Simple Linear Regression

One of the most basic ways to get information from our data sets is to determine if police across both countries are killing more or less people every year. A simple linear regression is a great way to model the trend of the death per million in each country as each year passes. By building a model to estimate the trend in the death per million for each country, we can determine if the death per million is progressively getting better or worse or the same. The model takes the equation as follows:

Equation 1:

$$\hat{y} = \beta_0 + \beta_1 x \quad (1)$$

Where  $\beta_0$  is the intercept of a linear equation and  $\beta_1$  represents the rate of change with respect to the year and  $x$  is the year. The  $\hat{y}$  is the predictor, and represents an estimate of the death per million given a year. This model was generated by calculating the death per million for each country using information from our data set and the population in each year from the year 2000 gained from each countries online census

information ((???), (???)). To get the death per million for each year, we divided the death total for that year by the population of that year and multiplied the result by a million. By using the death per million statistic, our model would be able to predict the trend and take into account the growing population every year. Since an increase of population would logically cause an increase of police related deaths, this statistic would take the increase of population into account.

## 3.2 Logistic Regression

A logistic regression builds a logistic model, and measures the probability of a certain event happening given the values of its independent variable. The logistic regression has a equation for of:

Equation 2:

$$Pr(Y_i \in \{Gunshot, Other\}) = \text{logit}^{-1}(\alpha^{\text{age}} + \alpha_{g[i]}^{\text{gender}} + \alpha_{p[i]}^{\text{province/state}} + \alpha_{r[i]}^{\text{race}}) \quad (2)$$

In this equation, the  $Y_i$  represents the probability a individual is likely to be killed by a gunshot vs being killed to other forms of force.  $\alpha^{\text{age}}$  represents the age of a victim,  $\alpha_{g[i]}^{\text{gender}}$  represents the gender,  $\alpha_{p[i]}^{\text{province/state}}$  is the province or the state and  $\alpha_{r[i]}^{\text{race}}$  is the race. All of these variables are categorical with the exception of age with is numerical. We model the categorical data numerically in the equation by one-hot encoding the variables. In other words, if a variable  $V$  which has 3 categories  $A$ ,  $B$  and  $C$ , then the equation would represent the variables as  $W_A V_A$ ,  $W_B V_B$  and  $W_C V_C$ , where  $W_n$  is the weight and  $V_n$  is either 0 or 1, and  $V_A + V_B + V_C = 1$ .

## 4 Results

In this section, we will go over the results from our simple linear regression models and binomial logistic regression models in both data sets.

### 4.1 Model Selection

There were two primary goals this paper attempted to achieve. The first goal was to determine if the overall situation in Canada was as dire as it is in America. A simple linear regression model was chosen as it creates a linear model and can measure the change of death per million with respect to our independent variable, Year. With this model we can estimate the change in death per million from year to year. This model was chosen for its simplicity and ability to model current relationships as well as give an overall trend. The second goal was to examine the effects of certain variables on how an individual is treated in terms of fatal force. Since in North America, use of force guidelines are strict and fairly uniform, it was determined to be a reasonable standard to measure all of our results regardless of Country, State/Province and City. Specifically, in Canada and USA, the police follow a shoot to kill policy, therefore every time a gun is fired by the police, there is intent to kill a suspect. All other methods of force are strictly to detain and apprehend a suspect and does not carry the intention to kill. As such, we processed the data in our binomial regression model so that the cause of death are either gunshot, or other. Since a majority of deaths in each data set (72% in Canada, 91% in USA) were gunshots, our model almost always predicts gunshot as the cause of death. However, this model is able to provide some insight to what variables play a stronger role in an individuals cause of death. This may be important in identifying key factors that effect how an individual is treated in a confrontation with the police, since all non gun shot deaths can be assumed to be unintentional. I chose a logistic regression model as I was interested in the chances of an event happening over predicting the event that will happen and a logistic regression provides this information exactly. I chose a binomial predictor as one of the predictor is an overwhelming favorite and the remaining causes of death are spread thinly among the remaining percents. As a result, I chose to do a binomial logistic regression in order to model the effect various variables has on the overall outcome.

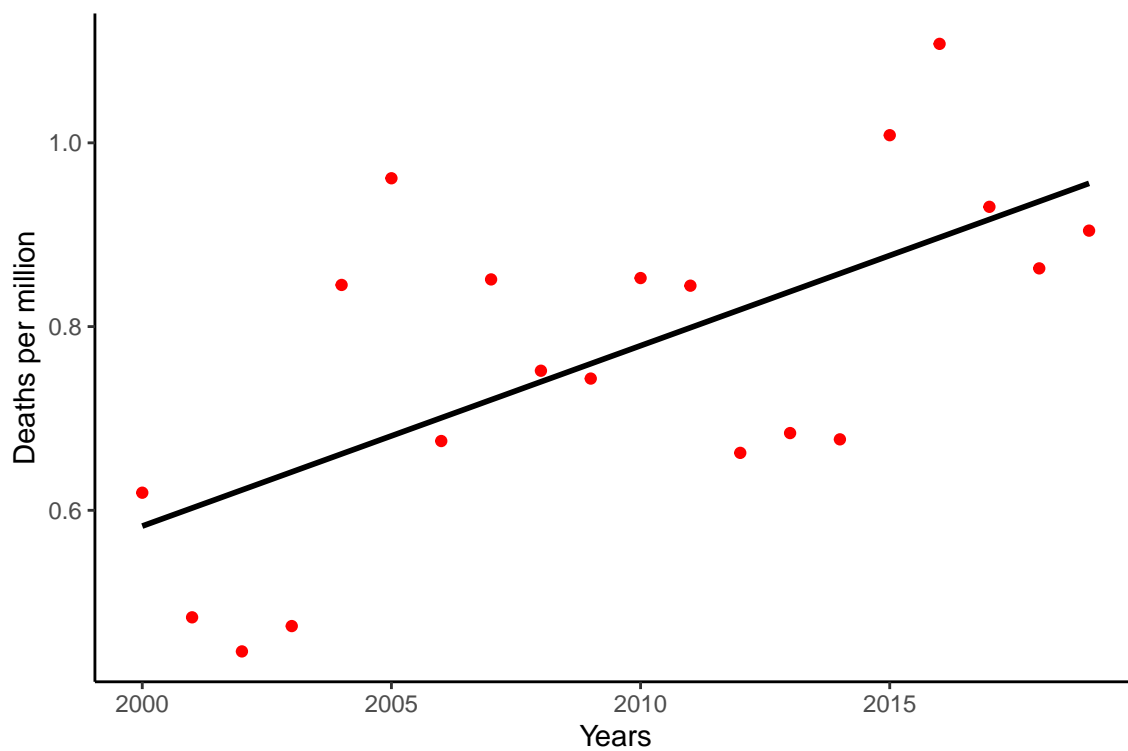


Figure 7: Linear regression model of death per million by year in Canada

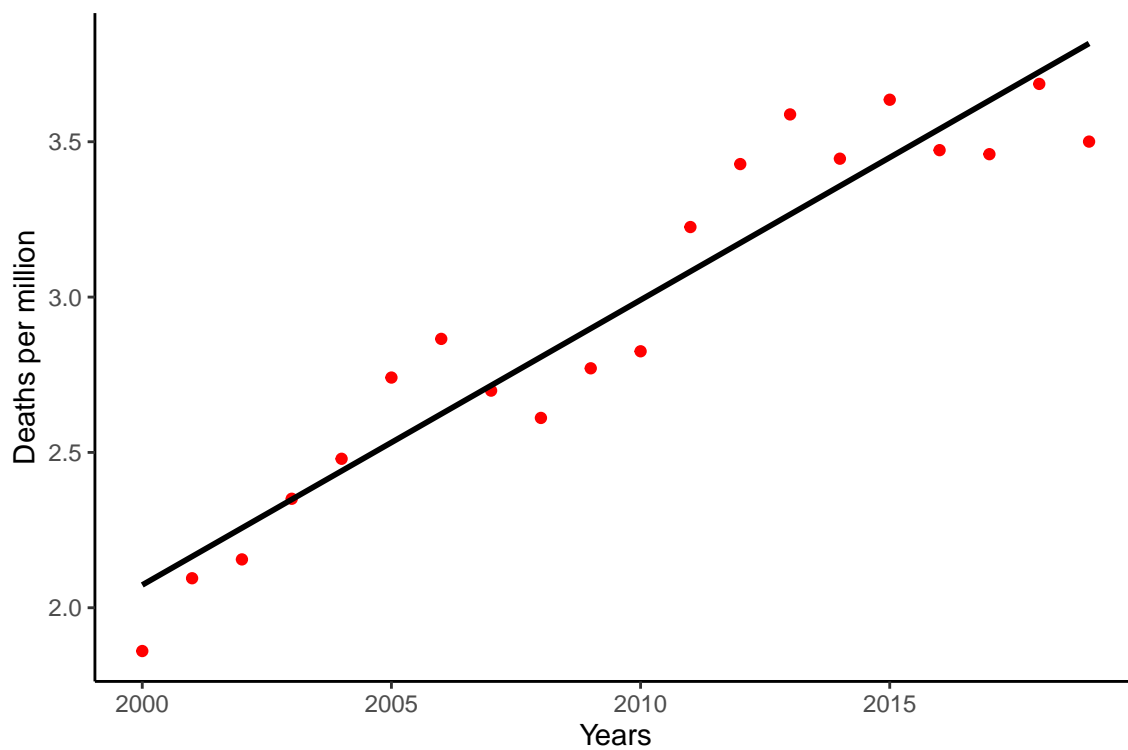


Figure 8: Linear regression model of death per million by year in USA

## 5 Dicussion

TBA

## 6 Appendix

TBA

## References

- “2016 Pulitzer Prizes.” 2016. *The Pulitzer Prizes*. <https://www.pulitzer.org/prize-winners-by-year/2016>.
- “Amy Coney Barrett: The Supreme Court Nominee on Abortion, Healthcare and Her Faith.” 2020. *BBC News*. BBC. <https://www.bbc.com/news/election-us-2020-54512678>.
- Boisvert, Nick. 2020. “Toronto City Council Votes Against Cut to Police Budget, for Body-Worn Cameras by 2021 | Cbc News.” *CBCnews*. CBC/Radio Canada. <https://www.cbc.ca/news/canada/toronto/toronto-city-council-defund-police-1.5630983>.
- Bürkner, Paul-Christian. 2018. “Advanced Bayesian Multilevel Modeling with the R Package brms.” *The R Journal* 10 (1): 395–411. <https://doi.org/10.32614/RJ-2018-017>.
- Finch, Brian Karl, Audrey Beck, D. Brian Burghart, Richard Johnson, David Klinger, and Kyla Thomas. 2019. ““Using Crowd-Sourced Data to Explore Police-Related-Deaths in the United States (2000–2017): The Case of Fatal Encounters”” *Open Health Data* 6 (1): 1. <https://doi.org/10.5334/ohd.30>.
- Johnston, Rene. 2020. “Doug Ford Says Canada Doesn’t Have Same ‘Systemic, Deep Roots’ of Racism as U.s. | Cbc News.” *CBCnews*. The Canadian Press. <https://www.cbc.ca/news/canada/toronto/ont-ford-racism-1.5595647>.
- Marcoux, Jacques, and Katie Nicholson. 2018. “Deadly Force.” *CBCnews*. CBC/Radio Canada. <https://newsinteractives.cbc.ca/longform-custom/deadly-force>.
- Marcouz, Jacques. 2018. “Deadly Force: How Cbc Analyzed Details of Hundreds of Fatal Encounters Between Canadians, Police | Cbc News.” *CBCnews*. CBC/Radio Canada. <https://www.cbc.ca/news/canada/manitoba/iteam/deadly-force-abc-analysis-1.4603696>.
- 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 Grolemond, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. <https://doi.org/10.21105/joss.01686>.
- Xie, Yihui. 2020. *Bookdown: Authoring Books and Technical Documents with R Markdown*. <https://github.com/rstudio/bookdown>.