Comparing Canada and USA during an era of rising police related deaths*

James Bao

22 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 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.

*The data can be found at Fatal Encounters and CBC

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 using R (R Core Team (2020)). 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 (Burghart (2020)) 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(Ross et al. (2020)) 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 form 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 (Marcouz (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, that there could be many names missing and are asking the public to provide more names where possible. However, both of these data sets are commonly used when analyzing statistics regarding police related deaths in the absence of official statistics.

Since these data sets are not government statistics, it feel reasonable to assume that these data sets will not purposefully miss information. Since both the organizations that collected the data sets had the goal to provide the public visibility towards police related deaths, it feels reasonable to assume that the collecting organizations are well motivated to collect every single name possible without leaving out any names. It is possible that the collection process purposefully left out names of certain demographics in order to make the police seem more or less discriminatory, however both data sets (especially the Fatal Encounters project) has gone through many critiques and has been used in various other statistics to check for validity. Another method bias could have been introduced is in the collection process themselves. Ignoring theories that police actively cover up police related deaths, we can assume information gathered through information requests by CBC and Fatal Encounters is unbiased and contain as much of the full picture as possible, however another method of information gathering is through news headlines, which has been known to be more biased. Between the data sets, CBC is much more likely to contain bias as the data set has been used less in other works, as well as the fact that the news network has to write an article using their own data set afterwards. It is much more appealing to talk about how bad the police is treating certain minorities rather than to write an article about how normal the police is. However their data set is available for public inspection and critique and seems unlikely a reputable news station will risk manipulating data when a well collected data set can easily yield an interesting headline.

While both of these data sets are assumed to be unbias and reasonably accurate, they are collect 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. Using Wickham et al. (2019) and Wickham et al. (2020) we graphed the variables below.

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

Table 2: Count of USA victims by gender

Gender	Count
Female	946
Male	17254
Transgender	15
Unknown	11

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:

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, inorder to compare our data, we needed the category naming to be consistent as well (ex Arab vs Middle Eastern).

From Figure **??**, 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 **??**). 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 Angelos, 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.

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 gun shot. 96% of all victims were males (figure **??**)

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

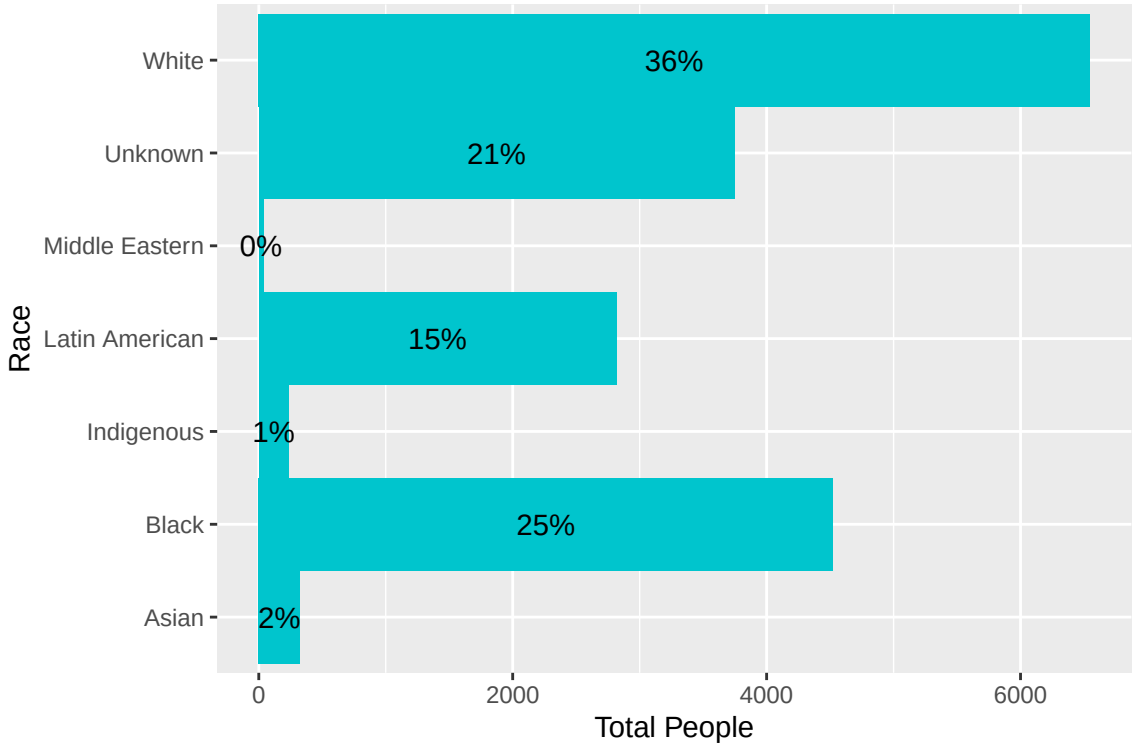

Figure 1: Distribution of the race of Americans who were killed by police

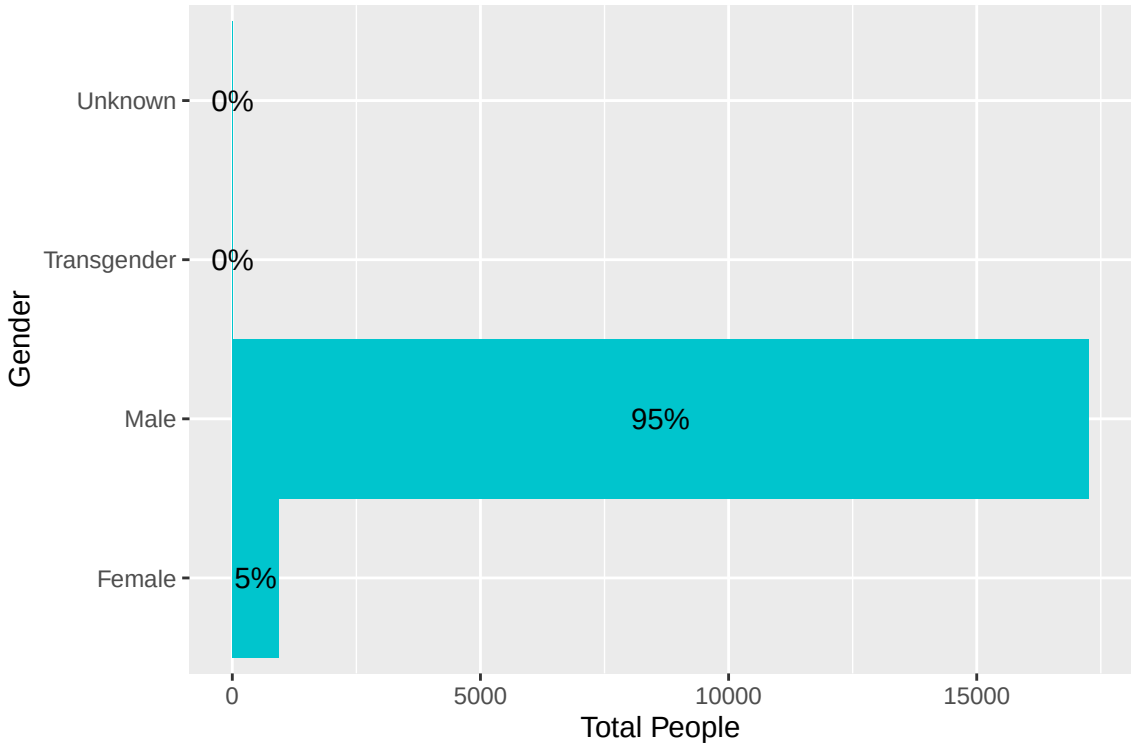

Figure 2: Distribution of the gender of Americans who were killed by police

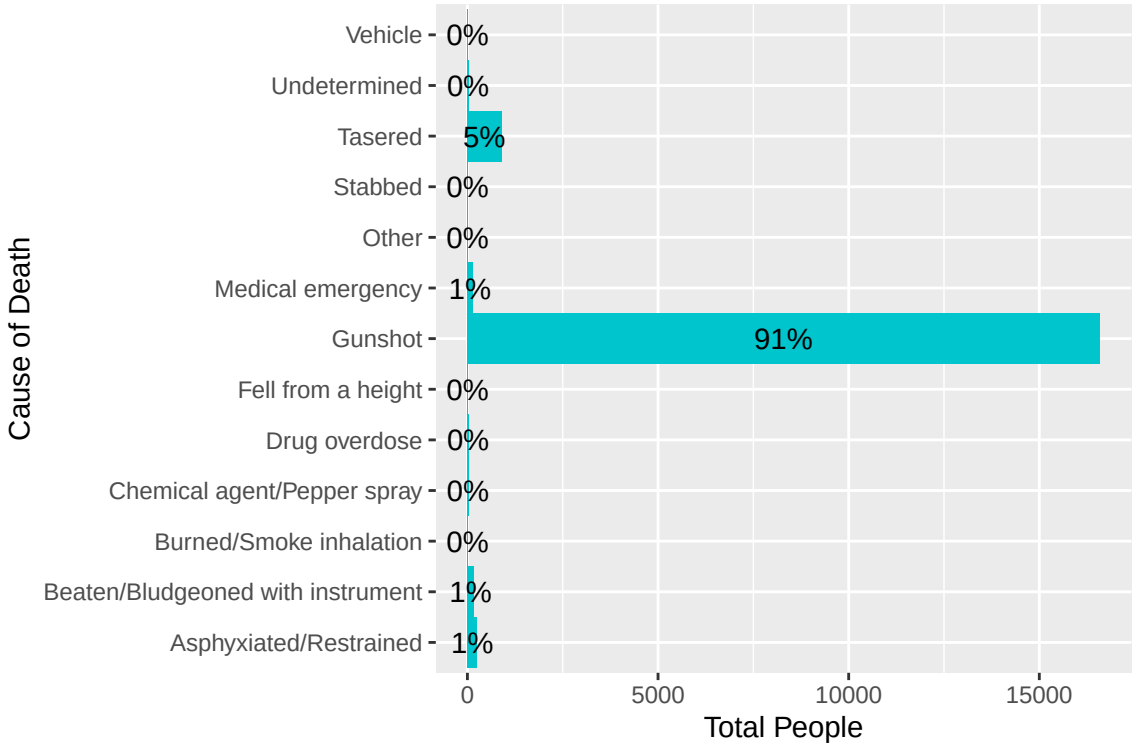

Figure 3: Distribution of the Cause of death for US

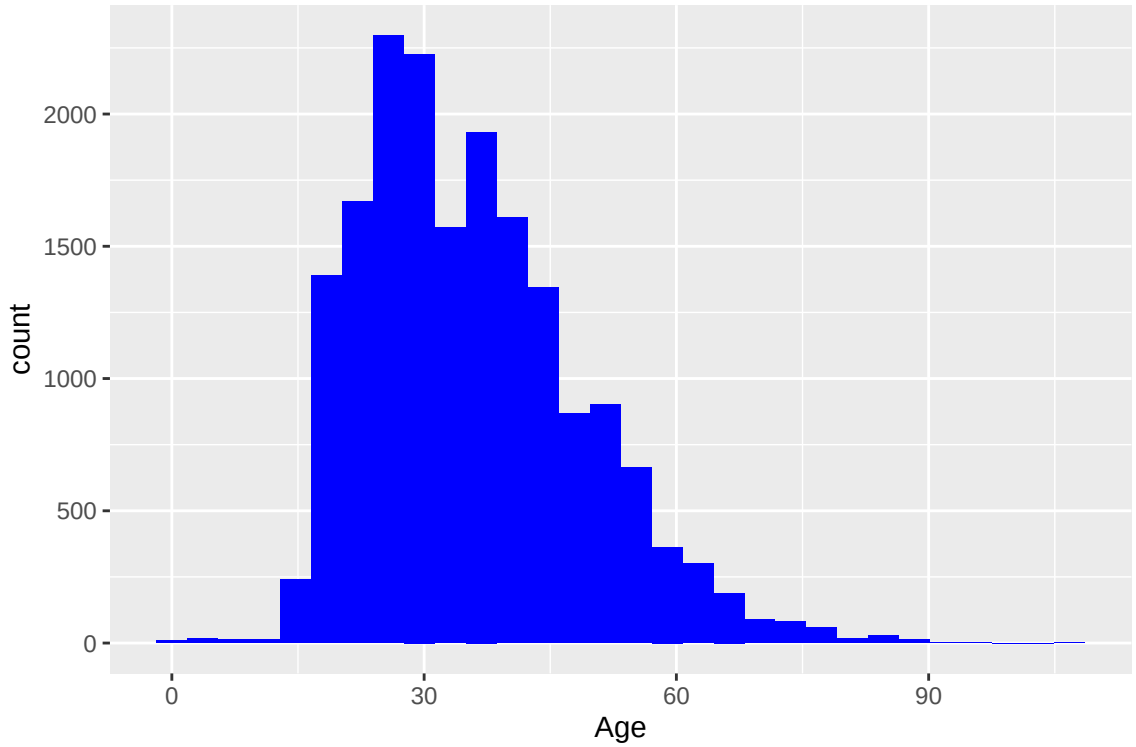

Figure 4: Distribution of the age of Canadians who were killed by police

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 MontrÃ©al	33
SÃ»retÃ© du QuÃ©bec	31
Edmonton Police Service	27
Calgary Police Service	26

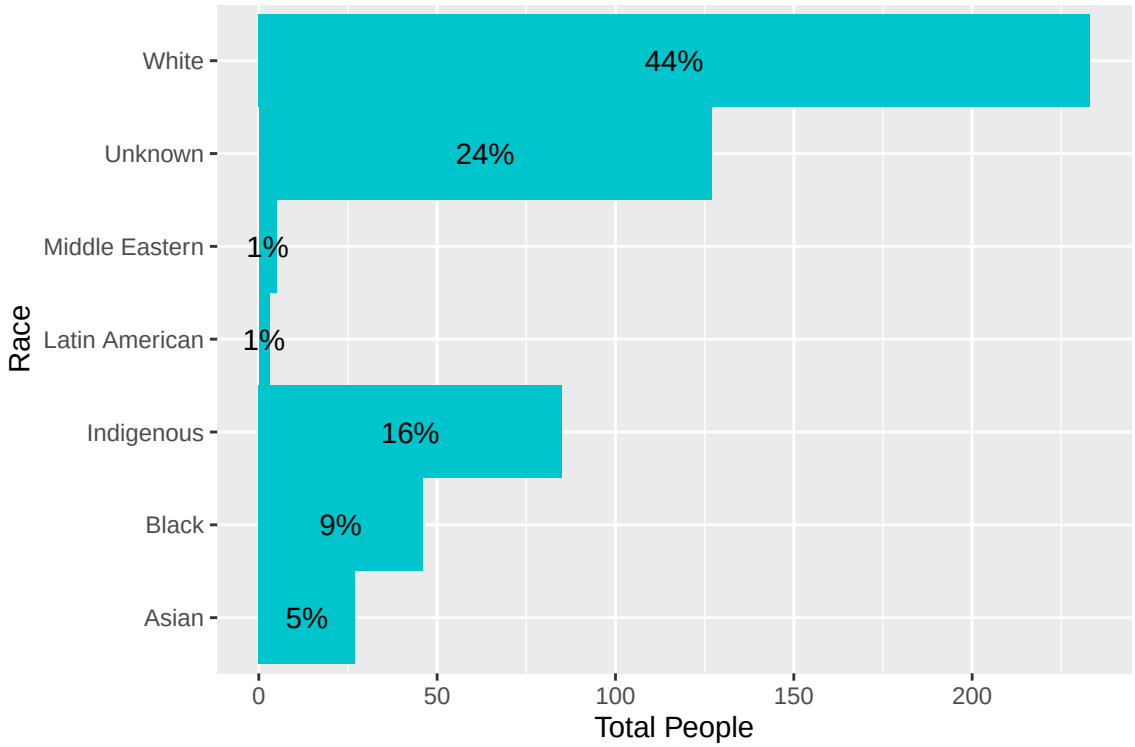

Figure 5: Distribution of the race of Canadians who were killed by police

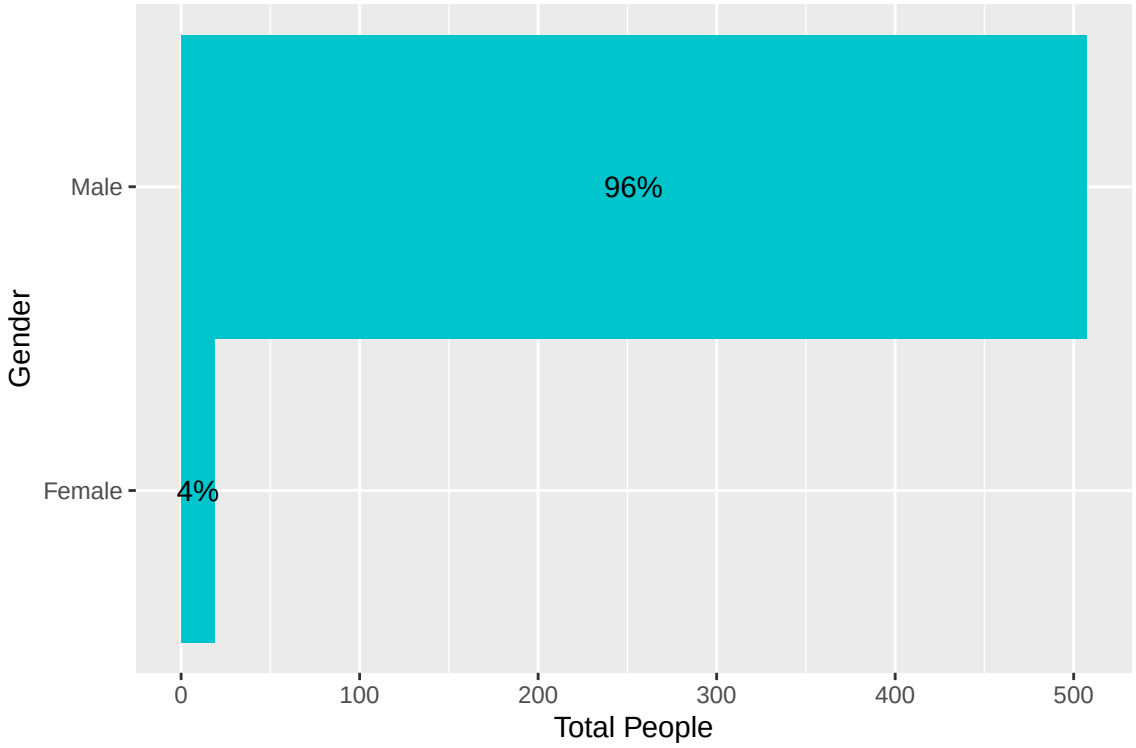

Figure 6: Distribution of the gender of Canadians who were killed by police

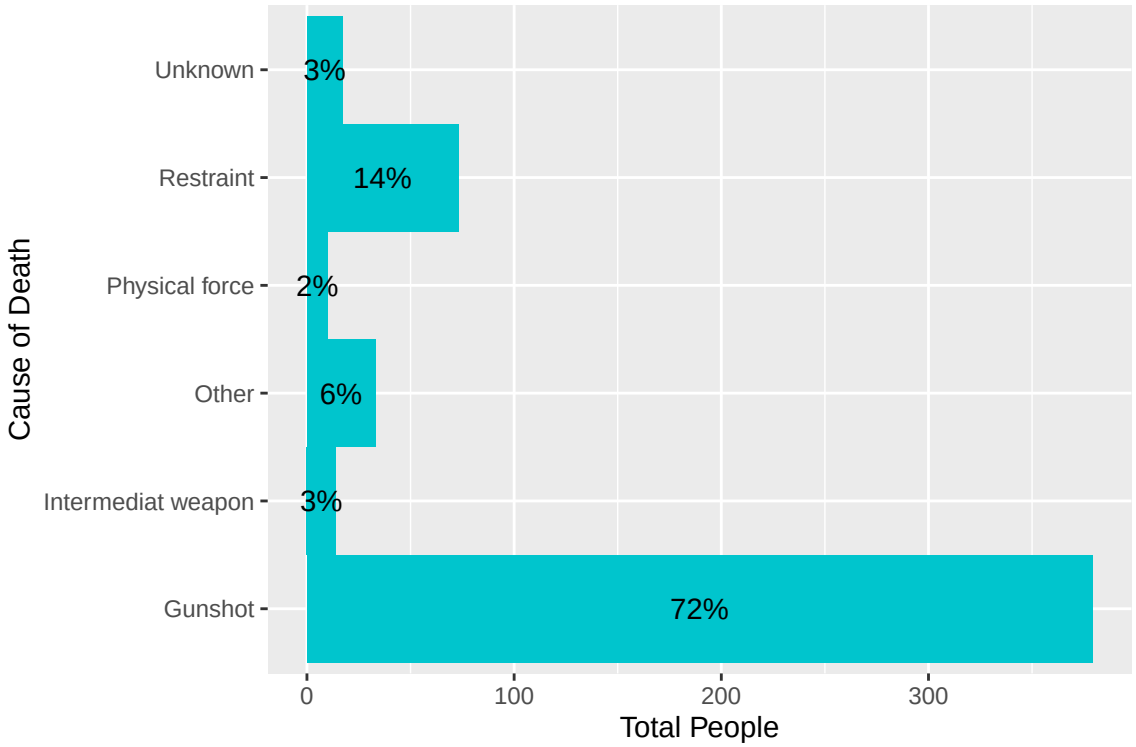

Figure 7: Distribution of the gender of Canadians who were killed by police

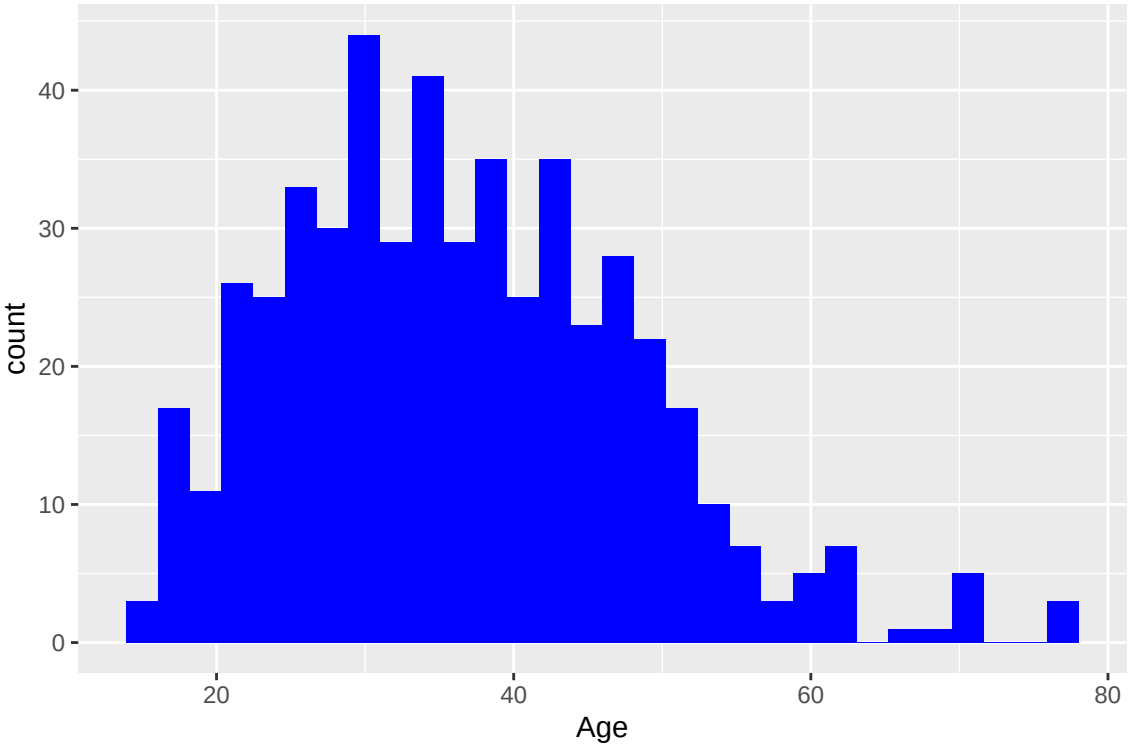

Figure 8: Distribution of the age of Canadians who were killed by police

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:

$$\hat{y} = \beta_0 + \beta_1 x \tag{1}$$

Where β_0 is the intercept of a linear equation and β_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 ((“USA Population” 2019), Government of Canada (2020)). 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:

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

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. α^{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 which is a numerical variable. 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 using the NNet library(Venables and Ripley (2002)). 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.

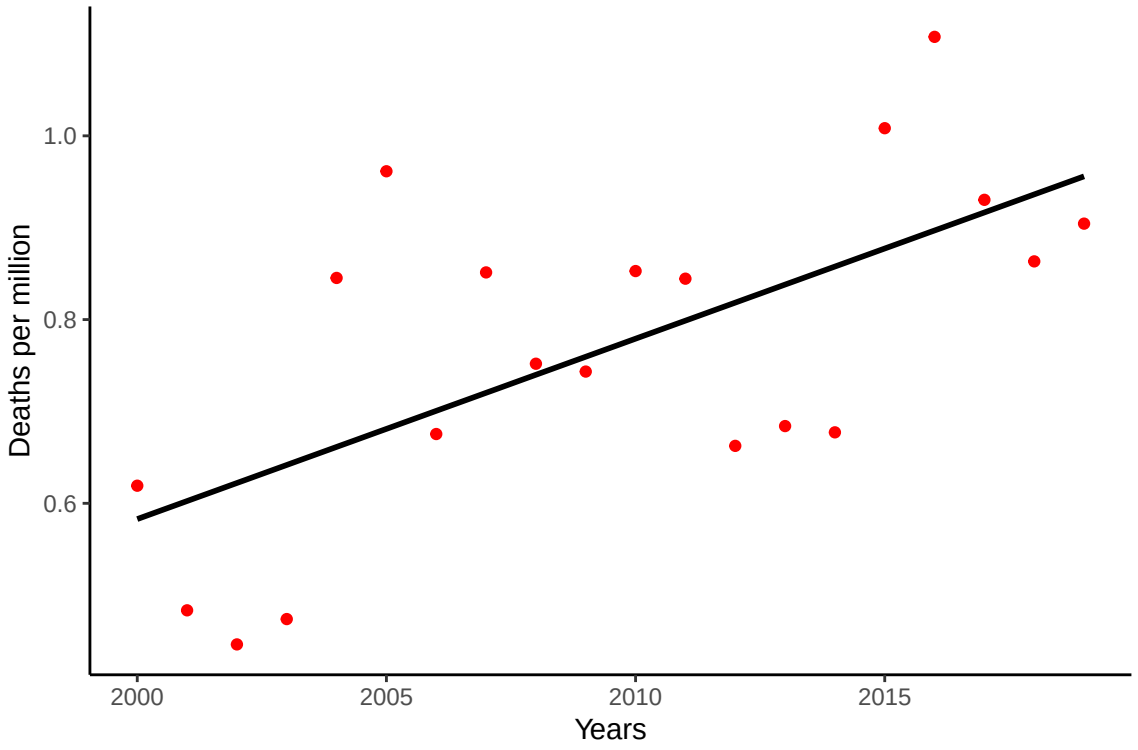

Figure 9: Linear regression model of death per million by year in Canada

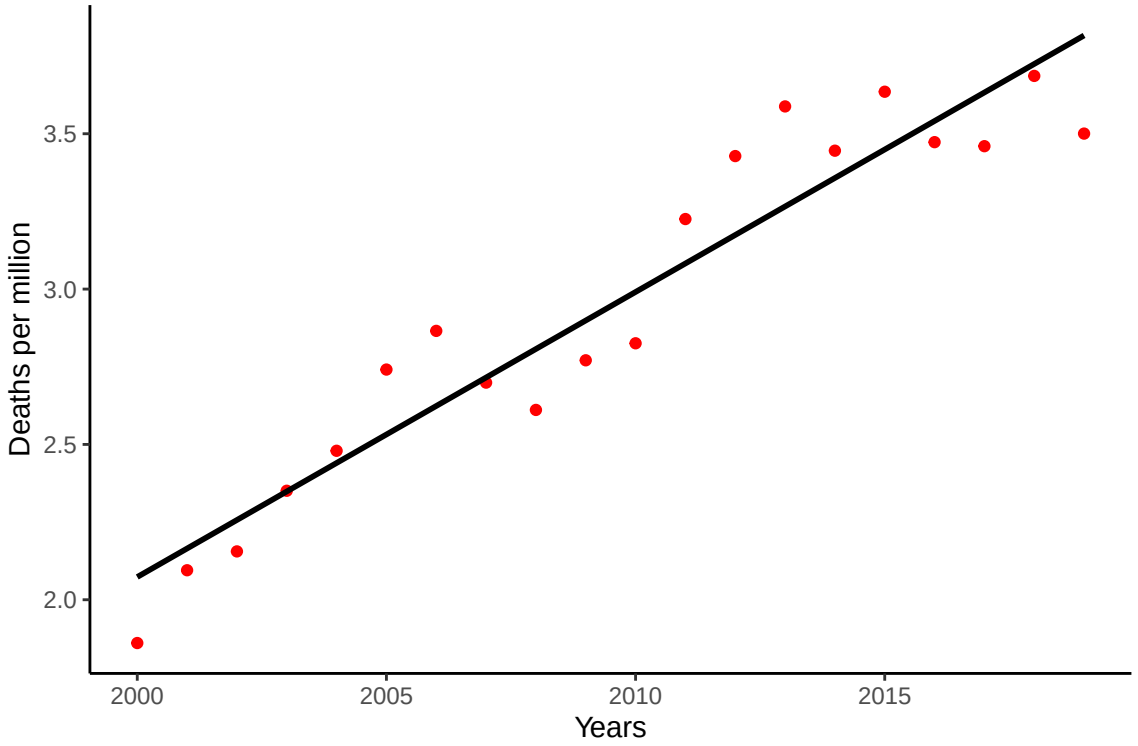

Figure 10: Linear regression model of death per million by year in USA

Table 13: Estimate values of Linear Regression on Canadian Dataset

	Estimate	Std. Error	T value	P value
(Intercept)	-38.6838874	11.0656270	-3.495860	0.0025803
year	0.0196334	0.0055066	3.565403	0.0022109

Table 14: Estimate values of Linear Regression on America Dataset

	Estimate	Std. Error	T value	P value
(Intercept)	-181.4116506	14.3833092	-12.61265	0
year	0.0917425	0.0071576	12.81744	0

4.2 Linear Regression Model

From figure **??** we see that Canada is facing a rise of death per million per year, at the rate of 0.02 (**??**). These estimates has a significant result as the p value is much less than 0.05. In the American data set, we observe a similar trend in the death per million per year however it has a steeper rate of 0.09, rising almost 5 times as fast. The American data set yields a p-value of 0, meaning it is also statistically significant. A p value is defined as the probability that our findings are a result of pure luck. In this case, a p-value would represent the statistical likelihood that our correlations are a matter of luck rather than an actual correlation. The lower the value is, the less likely luck is involved and the stronger our correlation is. Since we have low p-values in both models, we can claim these results are statistically significant.

From the tables above, we see that different variables have varying effects on the chances of an individual to be shot. However, the errors with these estimates are fairly high and since a majority of police related deaths in both Canada and USA, the model well overwhelmingly predict the cause of the death to be gunshot. However the estimates can act as an indicator to certain predictions as well be discussed in the following section.

5 Dicussion

The topic of police related deaths has been a commonly debated and discussed topic in 2020, with the events in the summer drawing attention to how serious the situation is in both Canada and USA. As indicated by figures **??** and **??**, both Canada and USA are facing an increase in their police related death per million. In 2019, USA had approximately 3.5 deaths per million people while Canada had approximately 0.9 deaths per million related to the police. Although this is a large and significant difference, there could be many reasons for this. The most likely reason behind this difference in police related deaths between the countries is gun control within both these countries. USA has per state regulations for gun control and are known to be very loose compared to many countries across the world. Estimates by BBC ((“America’s Gun Culture in Charts” 2019)) show that America has by far the most guns per population, being the only country who on average has more than one gun per person (120.5guns per 100 people). Canada has much stricter gun restrictions is 6th on the list of estimates with 34.7 guns per 100 people. Other countries that make the list include Yemen with 52.8 guns per 100 people and Finland with 32.4 guns per 100 people. These countries has mandatory military service for males above the age of 18 so the higher gun ownership statistics could be a result of these factors. The sheer amount of guns in America could contribute to the fact that police are more likely to use lethal in situations due to fear of life for themselves or others, which is the only situation police are supposed to use lethal force. However gun control is not the only contributing factor as crime statistics as well as police training can be a major contribution. Within Canada, police departments such as Toronto Police and the RCMP have Cadet training programs taking 24-26 weeks (6 months) to complete and Vancouver police offers a 3 block training program taking as long as 42 weeks to finish according to their websites. In America, police academies takes anywhere between 18 to 30 weeks, with Los Angeles police training taking

Table 15: Estimates of Coefficients for the Logistic Regression on Canadian Data set

Coefficients	Estimates
(Intercept)	-1.3076386
RaceBlack	-0.0545142
RaceIndigenous	-0.4932026
RaceLatin American	1.8998558
RaceMiddle Eastern	-17.7878397
RaceUnknown	0.5704698
RaceWhite	-0.1259627
Age	0.0087189
GenderMale	-0.0847875
Prov.StateBC	0.6154495
Prov.StateMB	-0.1607783
Prov.StateNB	0.5386535
Prov.StateNL	-18.1376110
Prov.StateNS	-0.1586050
Prov.StateNT	-11.8968970
Prov.StateNU	-17.1337537
Prov.StateON	0.0524817
Prov.StatePE	-12.9813228
Prov.StateQC	-0.3189684
Prov.StateSK	-0.0331582
Prov.StateYT	23.2811882

6months and Chicago requiring 1000 hours of instruction according to their respective websites. Most police departments offer further optional training, and require to new police to work alongside senior officers as part of their training process. In comparison, Germany requires officers to spend 2 and a half to 4 years of training before becoming an police officer (Stute (2014)). In comparison, Germany has 0.13 deaths per million between 2017-2018. Taking all this into account, both Canada and USA has a noticeably high rate of police related death, and although Canada does face less deaths, it does not mean that Canadian police should be safe from criticism.

Another statistic that is important to examine is the rate of change in the police related death per million of each country. The idea is more police related deaths despite accounting for the increase in population by calculating the death per million for each year is an unwanted result. There can be multiple reasons behind the increase of police related deaths per million in both countries, but the fact is, both countries are facing an increase in police related deaths, with Canada facing an increase of 0.02 deaths per million per year and America facing a growth of 0.09 deaths per million per year. Both countries' estimates are statistically significant meaning that the numbers are reliable to a certain degree and certainly point to a trend of the worsening scenario in Canada and USA. There has been many reasons suggested that may contribute to these rising numbers mostly surrounding an increase in crime. As in before, Canada is statistically doing better, as they have a lower rate of increase but it is still a significant growth per year, with 2020 projected to have 0.97 deaths per million in Canada and 3.9 deaths per million in the USA. Based on current population projections for 2020, that translates to 37 people in Canada and 1,290 people in America. According to numbers provided by Fatal Encounters and CBC for 2020 so far, Canada has 30 police related deaths as of June 2020, and USA has 1053 police related deaths as of November 27, 2020. Canada appears to be facing a worse than expected year, reach over 75% of its projected total in half the time, where as America appears to be doing similar to the estimated total (query on December 21 shows that USA currently has 1116 reported police related deaths due to intentional use of force within America in 2020). Another noticeable trend in figure **??** is how steady the overall trend is, and how closely it follow the estimated line, regardless of the countless reforms, dash cams, policy updates, training updates and educational programs provided by various governments and police departments over the years. Similarly in Canada, the grow is more sporadic yet still

Table 16: Estimates of Coefficients for the Logistic Regression on America Data set

Coefficients	Estimates		Coefficients	Estimates
(Intercept)	-4.9829653	31	Prov.StateMD	1.1896435
RaceBlack	0.7674502	32	Prov.StateME	0.2350663
RaceIndigenous	0.4167633	33	Prov.StateMI	1.3346825
RaceLatin American	0.3825139	34	Prov.StateMN	1.5405554
RaceMiddle Eastern	0.2375654	35	Prov.StateMO	1.3328949
RaceUnknown	0.5653943	36	Prov.StateMS	1.7653559
RaceWhite	0.2067476	37	Prov.StateMT	0.9021608
Age	0.0161254	38	Prov.StateNC	1.3420087
GenderMale	0.2963962	39	Prov.StateND	0.9296879
GenderTransgender	1.9952096	40	Prov.StateNE	1.8075328
GenderUnknown	2.3609464	41	Prov.StateNH	1.0634702
Prov.StateAL	1.4203671	42	Prov.StateNJ	1.3151212
Prov.StateAR	0.7148870	43	Prov.StateNM	0.5801088
Prov.StateAZ	0.7124670	44	Prov.StateNV	1.1226200
Prov.StateCA	1.4718530	45	Prov.StateNY	1.6948700
Prov.StateCO	1.0965327	46	Prov.StateOH	1.3258691
Prov.StateCT	2.4393895	47	Prov.StateOK	1.2652216
Prov.StateDC	0.4116973	48	Prov.StateOR	1.1658309
Prov.StateDE	1.0477509	49	Prov.StatePA	1.1121379
Prov.StateFL	1.3244872	50	Prov.StateRI	2.1197042
Prov.StateGA	1.0294383	51	Prov.StateSC	0.9553607
Prov.StateHI	1.8990546	52	Prov.StateSD	1.3712919
Prov.StateIA	1.2197821	53	Prov.StateTN	1.1403208
Prov.StateID	0.6061428	54	Prov.StateTX	1.7239237
Prov.StateIL	1.4308068	55	Prov.StateUT	0.7420657
Prov.StateIN	0.9717165	56	Prov.StateVA	0.9194792
Prov.StateKS	1.5628713	57	Prov.StateVT	1.5708805
Prov.StateKY	0.7000755	58	Prov.StateWA	1.4618950
Prov.StateLA	1.8796642	59	Prov.StateWI	0.3063954
Prov.StateMA	1.0630815	60	Prov.StateWV	0.2455830
		61	Prov.StateWY	0.8585335

faces that continuous climb despite multiple efforts to reduce police related deaths.

Police related deaths are generally followed by lengthy investigation that attempts to clarify a situation. Unfortunately in my cases such as in Toronto, the investigation is overseen by a police related organization and these reports are often accused of bias. As such it can be hard to determine whether or not a shooting was necessary or not. However, it should straightforward to analyze when a death was unintentional during an event of use of use of force. From figure **??** we see that in Canada 72 percent of police related deaths are a result of gunshot, compared to a 91% gun shot related cause of death in the USA (Figure **??**). This is concerning as in Canada 14% of police related deaths are due to restraint and 5 percent being physical force and intermediate weapons. In total almost 20% of deaths are a cause of police applying improper force and results in a death, compared to only 9 percent in the USA being unintentional. At the end of the day, these numbers arent entirely significant as how people are dying cannot really solve why people are dying or resolve the fact that they are dying. However in Canada's case, it feels like proper use of force training could potentially decrease the number of police related deaths. This is assuming that all the gunshot related deaths are intentional and justified, which is currently a heavily debated topics with arguments for both sides.

It is very hard to talk about police related deaths, gun shots and justifications behind these deaths without bringing race and other personal factors into the arguments. As such we attempted to find any correlation between race, gender, age and location and the chance of an individual dying from a gunshot rather than other causes. By going through Table **??** and **??** we can see that certain variables play a large role in the chances of an individual being lethally shot by the police. In table **??** we see that in white Americans has a higher chance of dying from a gunshot as compared to black Americans and Latin Americans, while Connecticut citizens has a lower chance of dying from gunshots as well as non-males. In table **??**, Latin Canadians has the lowest chance of being shot while many provinces has an extremely high chance of the cause of death being gunshot. In both models, age has a negligible effect. Overall there is nothing significant from these along, except for suggesting that people of different race, sex, and location are treated differently than police.

To further investigate this discussion, we break down the demographics of the victims in each country. In both countries, a majority of victims are males (Figure **??** and **??**), and this is obviously an misrepresentation of the population however this may be because due to a variety of societal, cultural and evolutionary causes suggested by a study by (Kanazawa (2008)). Further more, younger people (15- 40) are much more likely to be shot, and as with gender, younger individuals feel more need to prove themselves as well as having a lower impulse control. However by far the most controversial topic is the discussion of race within these data sets. In both data sets, the majority of victims within the data set are white, with 36% of American victims being white and 44% of Canadian victims being white (Figure **??** and **??**). This observation lead to the president of the United States to claim "And so are white people. So are white people, what a terrible question to ask. So are white people. More white people by the way. More white people" (Walters (2020)) in response to a reporter questioning why black people are dying at the hands of police. However what was failed to be considered was the representation of the population, since most of Canada and USA are white (above 70%) this is a under-representation while other races face an over representation. For example 16% of victims in Canada are Indigenous, 25% of victims in USA are black however according to statistics Canada only 4.9% of Canadians are indigenous and the U.S Census Bureau states 13.4% of Americans are black. More over claiming more white people are being killed is the opposite of a solution and makes the solution worse overall. Seeing the numbers from this perspective, it is hard not to see why people across the country are angry at the police, viewing the police as potentially racist. The numbers do support the fact that minorities are killed disproportionately. However another contributing factor for the misrepresentation within our data is the fact that overall, people belonging to lower socio-economic status tend to commit more crime, and are therefore more likely to be involved in a violent confrontation with police (("8.3 Who Commits Crime?" 2016)), and studies has shown that racial minorities are much more likely to have a lower socio-economic status. However it is clear that racial minorities are being over represented within our data sets and regardless of whether or not each case was justified or not, there is a disparity of race within our data sets and the fact that over demographics are also victims of police related deaths further shows that both Canada and USA needs to reconsider how policing works in both countries.

At the end of the day, Canada does seem to have a better policing system in terms of death by police, however being better should not be enough in these cases, and as the situation gets worse every year, there is a

rising urgency for changes within both Canada and USA. The fact that the data sets are not government collected is a problem in itself. In 2016, The Washington Post did a major national collection of police related deaths called Fatal Force, winning a Pulitzer prize and the following year the Justice Department started a program that will aim to collect a comprehensive set of plice use of force statistics, however this program is still incomplete and are missing information from many states and I was not able to locate a satisfactory dataset. The FBI has also started collecting a data base of police related deaths, but so far only about half the states are participating. In Canada, there were no other data bases that was comprehensive and up to date besides the CBC database. Scrolling through the data base provided by CBC and Fatal Encounters' respective websites, it's hard not to notice how many faces there are listed, and its even harder not to notice how many names are faceless. Whether or not the police are biased against a certain race or gender, the fact is that there is an increasingly worsening situation in Canada and USA that is resulting in some of the highest death per million rates in the world.

5.1 Weaknesses and Future steps

Throughout this paper, I attempted to dicuss the issue of police related deaths from a purely numerical point of view, however it is not completely prone to bias. Since I assumed that America policing system needed improvement, and judged Candian statistics compared to American statistics, I attempted to find graphs and research that would be beneficial to discussion and focused on the negatives of each country. Furthermore the logistic regression couldve been more detailed and in the future I would like to extend the regression to model the likelihood that an individual is killed by the police instead by finding more research and data on this topic. Furthermore, since most cases of police use of force does not result in death, it would be interesting to see how the statistics of police use of force in general and how it compares to police related death statistics. The CBC and Fatal Encounters data set also provide variables such as details on inidividual situations, whether the victim was armed at the time, as well as mental health of the victim which would be interesting to explore in the future.

References

"2016 Pulitzer Prizes." 2016. *The Pulitzer Prizes*. https://www.pulitzer.org/prize-winners-by-year/2016.

"8.3 Who Commits Crime?" 2016. *Social Problems*. University of Minnesota Libraries Publishing edition, 2015. This edition adapted from a work originally produced in 2010 by a publisher who has requested that it not receive attribution. https://open.lib.umn.edu/socialproblems/chapter/8-3-who-commits-crime/.

"America's Gun Culture in Charts." 2019. https://www.bbc.com/news/world-us-canada-41488081.

"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.

Burghart, D Brian. 2020. "Home." *Fatal Encounters*. https://fatalencounters.org/.

Department, Published by Statista Research, and Nov 26. 2020. "Police Killings: Rate in Selected Countries 2019." *Statista*. https://www.statista.com/statistics/1124039/police-killings-rate-selected-countries/.

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.

Government of Canada, Statistics Canada. 2020. *Population Estimates on July 1st, by Age and Sex*. Government of Canada, Statistics Canada. https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000501.

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.

Kanazawa, Satoshi. 2008. “Why Are Almost All Criminals Men? Part I.” *Psychology Today*. Sussex Publishers. https://www.psychologytoday.com/ca/blog/the-scientific-fundamentalist/200807/why-are-almost-all-criminals-men-part-i.

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-cbc-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/.

Ross, Cathy, April Lavilla-Lount, Kate Zieman, and Diana Redegeld. 2020. “Deadly Force.” *CBCnews*. CBC/Radio Canada. https://newsinteractives.cbc.ca/fatalpoliceencounters/.

Stute, Dennis. 2014. “Why German Police Officers Rarely Reach for Their Guns: DW: 27.08.2014.” *DW.COM*. https://www.dw.com/en/why-german-police-officers-rarely-reach-for-their-guns/a-17884779.

“USA Population.” 2019. *Timelines Explorer - Data Commons*. https://datacommons.org/tools/timeline.

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. Fourth. New York: Springer. http://www.stats.ox.ac.uk/pub/MASS4.

Walters, Joanna. 2020. “Trump Twists Stats on Police Brutality: ’More White People’ Are Killed.” *The Guardian*. Guardian News; Media. https://www.theguardian.com/us-news/2020/jul/14/donald-trump-george-floyd-police-killings.

Wickham, Hadley. 2011. “The Split-Apply-Combine Strategy for Data Analysis.” *Journal of Statistical Software* 40 (1): 1–29. http://www.jstatsoft.org/v40/i01/.

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.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2020. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Xie, Yihui. 2020. *Bookdown: Authoring Books and Technical Documents with R Markdown*. https://github.com/rstudio/bookdown.