**Analysis of Police Related Fatalities from 2000 to 2020**

Stat 3355.001

Group 17

Luke Robbins, Guy Fonkouop, Aatifa Khan

October 19, 2020

**1 Introduction**

This report is an analysis of a dataset tracking police related deaths in the United States. In order to analyze the data we are using R to clean, organize, and support meaningful conclusions found in the dataset. The data being used has been obtained through a website called Kaggle after being posted for the free use of anyone who is interested in using it. It should be noted that the dataset originally comes from an organization called Fatal Encounters that attempts to track and update a comprehensive and searchable national database of people killed during interactions with the police.

The dataset contains observations of 28,335 people who have died from police interactions, and 29 variables. These variables are as follows:

|  |  |
| --- | --- |
| Subject’s name | Name of deceased subject |
| Subject’s age | Age of deceased subject |
| Subject’s gender | Gender of deceased subject |
| Subject’s race | Race of deceased subject |
| Subject’s race with imputations | Imputed race if necessary |
| Imputation probability | Accuracy score of imputation |
| URL of image of deceased | Image of deceased subject |
| Date of injury resulting in death | Month, day, and year of injury |
| Location of injury (address) | Exact address of event |
| Location of death (city) | City the subject died in |
| Location of death (state) | State the subject died in |
| Location of death (zip code) | Zip code the subject died in |
| Location of death (county) | County the subject died in |
| Full address | Unique address of subject’s home |
| Latitude | Latitude of interaction with police |
| Longitude | Longitude of interaction with police |
| Agency responsible for death | Police department involved in event |
| Cause of death | Category of cause of death |
| A brief description of circumstances | Description of event causing death |
| Dispositions/Exclusions | For internal use of Fatal Encounters |
| Intentional Use of Force | Description of force used if intentional |
| Link to news article | URL to news article covering event |
| Symptoms of mental illness | Whether or not mental illness was reported |
| Video | Video of event |
| Date & Description | Date and description of event |
| Unique ID formula | Unique ID formula |
| Unique Identifier (Redundant) | Unique ID |
| Date (year) | Year of event |

Some of the provided variables have only null values or are intended for use by only the organization forming the database.

In light of the recent spotlight on both police brutality and police funding across America, our group aims to provide a clear look at the true causes of death by police and how it correlates with the current narrative across the country.

**2 Data Cleaning**

As the data needed a lot of cleaning, we started by deciding which variables were more important and necessary. We removed the variables that we weren’t going to focus on and renamed the important ones to more simple and easy variable names. The original variables “Death Cause”, “Age”, “Race”, “Date of Death”, “City”, “County”, “State”, “ZIP”, “Street”, “Gender”, “Name”, “Year”, “Agency” were changed to “Cause of Death”, “Subject’s Age”, “Subject’s Race”, “Date of Injury Resulting in Death Month Day Year”, “Location of Death City”, “Location of Death County”, “Location of Death State”, “Location of Death Zip Code”, “Location of Injury Address”, “Subject’s Gender”, “Subject’s Name”, “Date Year”,and “Agency responsible for Death” respectively and the other variables we kept had the variable name as is. We also changed all black spaces in the data to “NA”.

We started by filtering the values in the Cause of Death variable. As there were multiple values for cause of death that we could remove and focus mainly on others, we deleted the values that were not of our focus by using the mutate method. We then modified the other values by renaming them to more simple names. We took the original value names “Beaten/Bludgeoned with instruments”, “Asphyxiated/Restrained”, and “Vehicle” and changed them to “Beaten”, “Asphyxiated”, and “Pursuit” respectively, while keeping “Gunshot”, “Stabbed”, and “Chemical Agent/Pepper Spray” as is. By filtering the values, we were able to concentrate on values that were more repetitive.

We applied the same process we used to clean the data for the Cause of Death variable to the Race variable. We replaced the values to 6 values as there were similar values that we could combine into broader categories by using the mutate method. We combined “HIspanic/Latino” and “Hispanic/Latino” to “Hispanic” as a way to fix the typo, “Race Unspecified” and “NA” to “Unknown”, “Native American/Alaskan” and “European-American/White” to “White”, “Other Race” and “Middle Eastern” to “Other”, “Asian/Pacific Islander” to “Asian” and “African-American/Black” to “Black”. This way, the values are more easily identifiable and organized. Then we converted the variable to a factor variable.

We then worked on cleaning the Gender variable. As it did not need much cleaning, we just added another category for all the blank spaces by using mutate so the values were divided to “Female”, “Male”, “Transgender”, and “Unknown” and changed it to a factor variable. We then cleaned the Agency variable by modifying and renaming the values by using mutate into 4 different values. We renamed “Sheriff”, “Police”, and “Federal Bureau” to “Sheriff’s Office”, “Police Department”, and “FBI” respectively while combining the remaining agencies into one value named as “Other” and then changed the variable to a factor variable. We then worked on the Age variable by first changing it to a numeric variable and deleting the category of ages > 10 as they would not have much of a difference on the overall statistic by using mutate. We then categorized the remaining ages by using bucket in groups of 10’s from ages 10-100. The complete police\_fatalities dataset is summarized as below:

'data.frame': 26430 obs. of 23 variables:

$ Unique.ID : chr "25746" "25747" "25748" "25749" ...

$ name : chr "Samuel H. Knapp" "Mark A. Horton" "Phillip A. Blurbridge" "Mark Ortiz" ...

$ age : chr "17" "21" "19" "23" ...

$ gender : Factor w/ 4 levels "Male","Female",..: 1 1 1 1 1 2 1 1 2 2 ...

$ Subject.s.race : chr "European-American/White" "African-American/Black" "African-American/Black" "Hispanic/Latino" ...

$ race : Factor w/ 6 levels "White","Black",..: 1 2 2 3 2 2 2 2 6 6 ...

$ Imputation.probability : chr "not imputed" "not imputed" "not imputed" "not imputed" ...

$ date\_of\_death : chr "01/01/2000" "01/01/2000" "01/01/2000" "01/01/2000" ...

$ street : chr "27898-27804 US-101" "Davison Freeway" "Davison Freeway" "600 W Cherry Ln" ...

$ city : chr "Willits" "Detroit" "Detroit" "Carlsbad" ...

$ state : chr "CA" "MI" "MI" "NM" ...

$ zip : int 95490 48203 48203 88220 30294 95823 28501 28501 21201 21201 ...

$ county : chr "Mendocino" "Wayne" "Wayne" "Eddy" ...

$ Full.Address : chr "27898-27804 US-101 Willits CA 95490 Mendocino" "Davison Freeway Detroit MI 48203 Wayne" "Davison Freeway Detroit MI 48203 Wayne" "600 W Cherry Ln Carlsbad NM 88220 Eddy" ...

$ Latitude : num 39.5 42.4 42.4 32.5 33.6 ...

$ Longitude : num -123.4 -83.1 -83.1 -104.2 -84.2 ...

$ agency : Factor w/ 4 levels "Sheriff Office",..: 1 4 4 1 1 2 2 2 2 2 ...

$ death\_cause : chr "Pursuit" "Pursuit" "Pursuit" "Pursuit" ...

$ Intentional.Use.of.Force..Developing. : chr "Vehicle/Pursuit" "Vehicle/Pursuit" "Vehicle/Pursuit" "Vehicle/Pursuit" ...

$ Symptoms.of.mental.illness..INTERNAL.USE..NOT.FOR.ANALYSIS: chr "No" "No" "No" "No" ...

$ year : int 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 ...

$ ages : num 17 21 19 23 53 24 30 29 42 41 ...

$ age\_grps : Factor w/ 9 levels "[10,20)","[20,30)",..: 1 2 1 2 5 2 3 2 4 4 ...

'data.frame': 26430 obs. of 14 variables:

$ name : chr "Samuel H. Knapp" "Mark A. Horton" "Phillip A. Blurbridge" "Mark Ortiz" ...

$ age : chr "17" "21" "19" "23" ...

$ age\_grps : Factor w/ 9 levels "[10,20)","[20,30)",..: 1 2 1 2 5 2 3 2 4 4 ...

$ race : Factor w/ 6 levels "White","Black",..: 1 2 2 3 2 2 2 2 6 6 ...

$ gender : Factor w/ 4 levels "Male","Female",..: 1 1 1 1 1 2 1 1 2 2 ...

$ street : chr "27898-27804 US-101" "Davison Freeway" "Davison Freeway" "600 W Cherry Ln" ...

$ city : chr "Willits" "Detroit" "Detroit" "Carlsbad" ...

$ state : chr "CA" "MI" "MI" "NM" ...

$ zip : int 95490 48203 48203 88220 30294 95823 28501 28501 21201 21201 ...

$ county : chr "Mendocino" "Wayne" "Wayne" "Eddy" ...

$ date\_of\_death: chr "01/01/2000" "01/01/2000" "01/01/2000" "01/01/2000" ...

$ death\_cause : chr "Pursuit" "Pursuit" "Pursuit" "Pursuit" ...

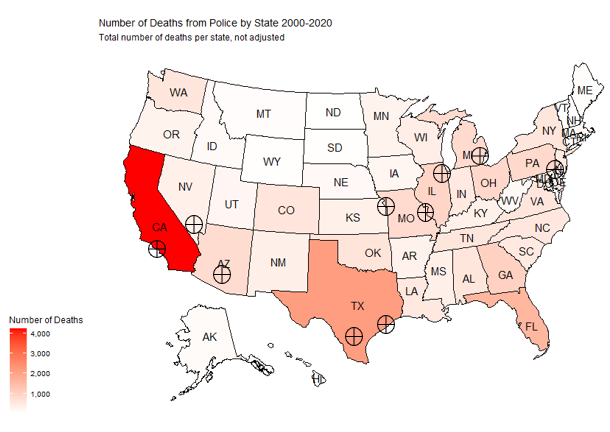
$ year : int 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 ...

$ agency : Factor w/ 4 levels "Sheriff Office",..: 1 4 4 1 1 2 2 2 2 2 …

**3 Analysis & Questions**

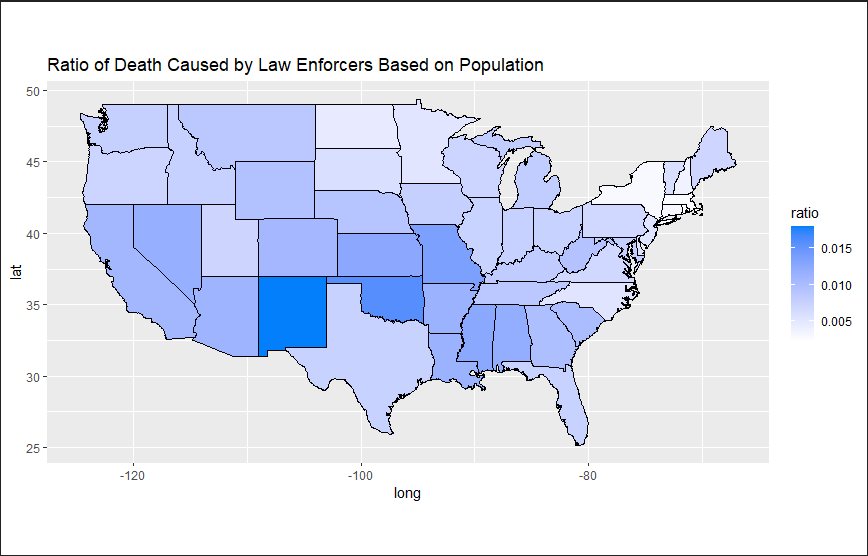
In this section we will present the questions that pushed our interest to evaluate this dataset and report our findings through charts that attempt to answer said questions. Overall, we attempt to compare our findings to the modern narrative of police related fatalities in the United States. Using the dataset we were able to focus on questions that provided a deeper look into the state of police violence in America. The main questions in this section deal with the geographic distribution of deaths, the distribution among race and gender, and the number of deaths per year.

**3.1 What is the geographic distribution of Law enforcement brutalities in the US?**

****

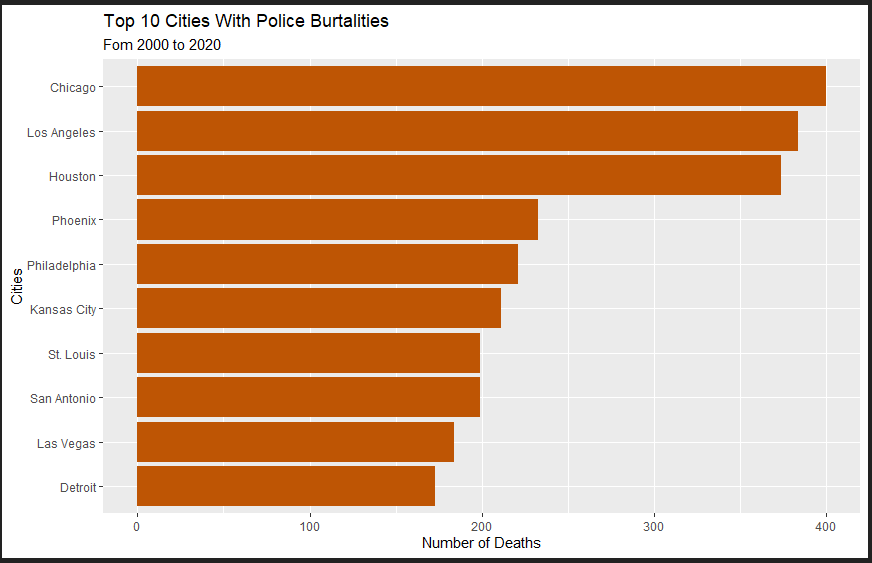
*Figure 1.1: US Map of Law Enforcers Fatalities.*

Figure 1.1 exhibits a map indicating the propagation of death across the US territory inflicted by law enforces officers. From the map we can see that the states with more intense red have the highest number of deaths and the states with less intense red registered fewer deaths. From this analysis we will be tempted to conclude that police brutalities were more fatal in California, Texas, Florida. But we know that those are the most populated states which can lead to higher numbers. Let’s now analyse if the ratio is the same by comparing the number of deaths based on each state's population. Since our dataset doesn’t have a variable listing the US population by states, we had to look for another dataset (*US population by States.csv*) from Kaggle and join it with our study dataset by region to obtain the map below.



*Figure 1.2: US Map Ratio of Law Enforcers Fatalities.*

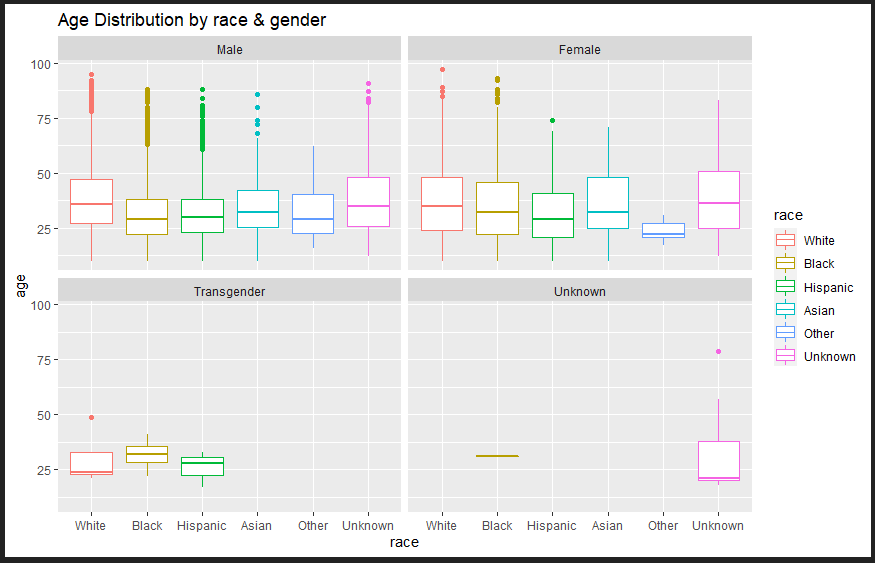
From Figure 1.2 we can see that the percentage of victims is more concentrated in Minnesota, Oklahoma, and New Mexico. Which makes us understand that although California, Texas, and Florida might be the top 3 states with higher numbers of death, the ratio of fatalities in those 3 states is much lower compared to Minnesota, Oklahoma, and Arizona. Now let’s gear our analysis toward cities.



*Figure 1.2b Highest deaths in cities bar plot*

It is somewhat obvious that we see a few of the cities with the highest number of deaths from police violence in the states that had the brightest red color in figure 1.1. However, it is surprising that places like Philadelphia or Michigan because their states have a relatively low death rate per capita. This shows that datasets breaking down specific states would be more useful for analysis, and this dataset is more useful for a national scope.

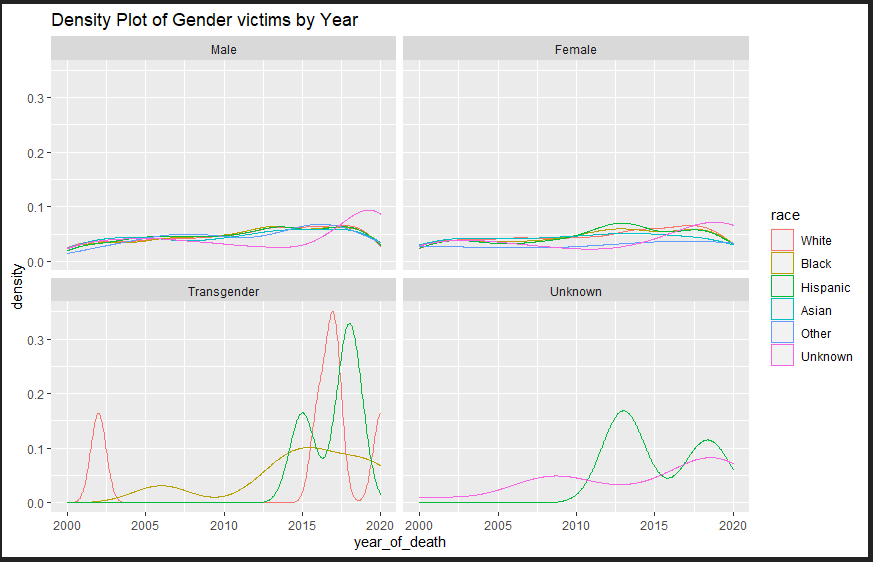
**3.2 What is the age distribution of Law enforcement victims by race and gender ?**

****

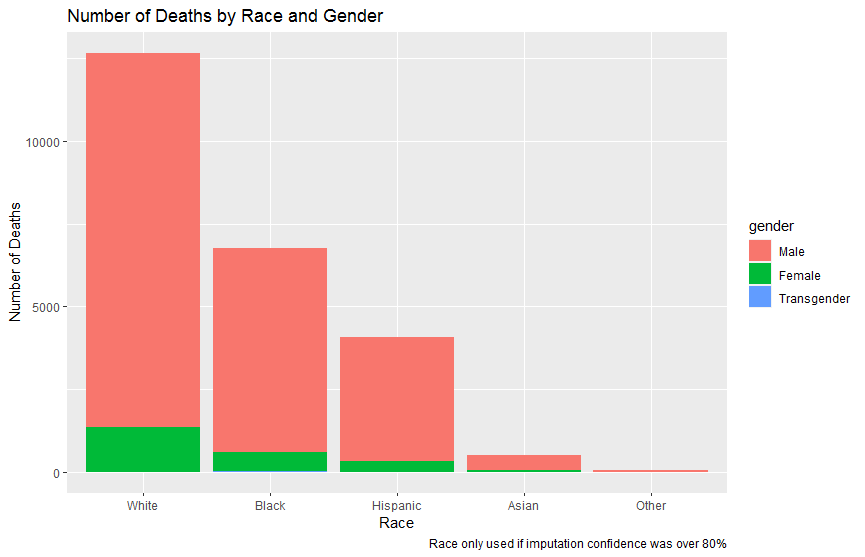
*Figure 1.3 Boxplot of Racial Gender affected by police brutality*

After analysing the fatalities geographically we wanted to see the racial impact and compare it by gender to see which age groups are mostly affected. It is important to note that we use imputed race, but assigned all imputed values with a confidence rating less than 80% to unknown. Figure 1.3 exhibits a boxplot of the races that were affected by their gender and ages. The most obvious implication for this visual representation is that the most affected age group is between 25 - 38 for each of the race and genders. Looking closer, we see that black and hispanic victims have a lower age on average than white victims.

Figure 1.4 below gives a representation of the density of victims of each race by year. We can see from this visualization that female victims are more dense than male victims, and transgender victims are very dense from the year 2010 - 2020 for the white, hispanic and black race. The analysis below also shows that there is a high density of unknown gender victims most likely due to the quality of the report to the nonprofit who hosts the dataset.



*Figure 1.4 Racial Density Plot of Victims by year*



*Figure 1.3b bar plot showing race vs. number of deaths, separated by gender.*

Figure 1.3b shows that white males have a higher number of deaths overall than any other race. Although this seems to indicate that white people are almost twice as likely as black people to die from police brutality, it is important to contextualize by percentage of population. When we take the total number of a specific race of people in the United States and divide by the amount of deaths shown in this data set, we get a more reliable indication of likelihood of death. The respective numbers are listed below:

White - .005%

Black - .015%

Hispanic - .007%

Asian - .0026%

This shows that blacks are almost 3 times as likely to die from police violence than whites, and significantly more likely than any other race.

**3.3 What is the age distribution of all Law enforcement victims by cause of death ?**

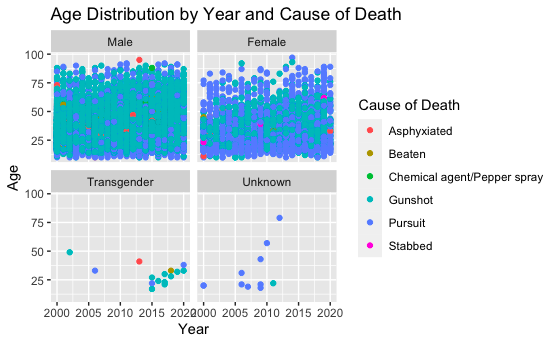
*Figure 1.5 Dot plot of Cause of Deaths and Age of victims per Year*

Figure 1.5 depicts a dot plot indicating the correlation between Age and Year of Death by Cause of Death. As visually shown, most deaths have occured in males with the cause of death having been by gunshots. While female victims had fewer deaths with the cause of death as pursuit. This can be due to the fact that males are seen more as a threat than females and the fact that police shootings have been rising recently so deaths caused by deaths are more common. There have been a couple of deaths for transgender and unknown with a variety of causes of deaths. Gunshots and pursuits are the most common form of death, although there are some cases including asphyxiation and beatings.

**3.4 What effect does agency have on the other details of the victim from the incident?**

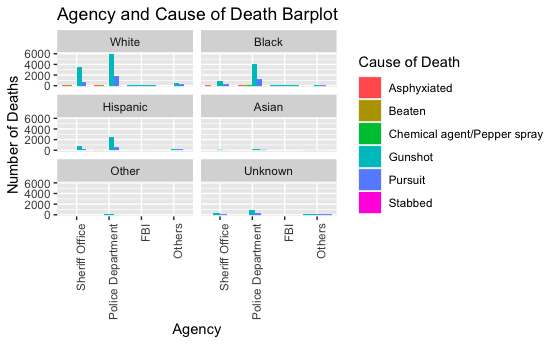
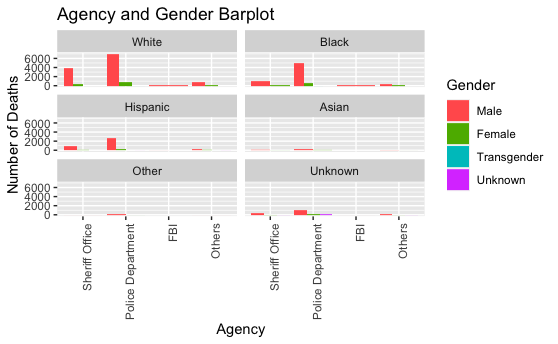
*****Figure 1.6 Bar plot of Relationship Between Agency and Cause of Death*

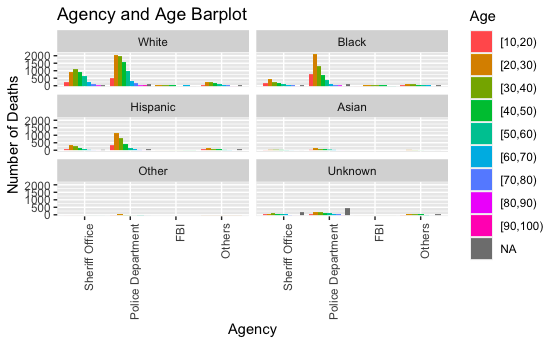
Figure 1.6 depicts a bar plot showing the correlation between law enforcement agency, race, and cause of death. As shown in the above bar plot, the agency with the most deaths is the police department for all types of race. Also, the most common cause of death is from gunshots as mentioned earlier. White people have had the most deaths with the least deaths being from unknown and other. The reason there are some unknown victims could be due to the fact that the victims could have been extremely injured thus being unidentifiable. An interesting thing that this chart reveals is that, with respect to portion of total population, a greater percentage of black people are shot than white people.

**

*Figure 1.7 Bar plot of Relationship Between Agency and Gender*

Figure 1.7 displays any correlation between agency and gender regarding race. As shown above and mentioned earlier, more deaths occur in men than women. Most deaths have occurred in white males while least deaths occurred in Asian males. In almost all types of gender and race, deaths caused by police departments have been the most common. The least deaths have been caused by the FBI and other smaller agencies.

Figure 1.8 shows the correlation between agency and age groups regarding race. There have been many deaths caused by the police department and sheriff's office in white people, but the most deaths have been caused by the police department in black people ranging from the age 20-30. This can be due to the many cases against black people recently. There have been very minimal incidents in other races but there have been a few in races that are unknown, which can be due to the fact that the victims were unrecognizable. As the victims grow older, we have noticed that there have been less incidents as well. From looking at these three bar plots, we realize that the most deaths have occured in white people by the police department regardless of the age, gender, and cause of death.

* Figure 1.8 Bar plot of Relationship Between Agency and Age*

**4 Conclusion**

After using the data contained in the Kaggle dataset we have presented evidence of a few conclusions. The first conclusion is that it is reasonable to suggest that the number of police related fatalities has risen since the year of 2000. The second conclusion is that this has disproportionately affected black people as shown by the amount of death with respect to percent of population. This is interesting because it affirms the modern narrative that police fatalities are found more within minority communities, and can be presented as a data supported reason to believe this. A third conclusion we found was that police departments around the country are the main agency responsible for death related to brutality by an officer. Although this may be considered well known, it is important to offer an empirical explanation of the breakdown of responsibility in authority in the US. Lastly, we have found that the median age for victims is younger and around the 30 year-old mark. It was important to know this because there is a political narrative that people who find themselves victims of police violence are “young and dumb”. Finding that the average age is much higher than people would think helps support the idea that there may be more to it than just being young.

This dataset has provided valuable insights on the current status of police violence in our country, but doesn’t fully explain the situation. One huge weakness of our dataset was the inability to make meaningful connections with the reasons victims were stopped or made to interact with the police. This was because the variables were very long descriptions housed in strings and contained little of the same keywords. Moving forward it would be important to apply these findings to larger datasets that come from government organizations that have more rigorous data collection methods. We could also use machine learning dictionaries to pull key ideas out of the descriptions provided within this dataset. Ultimately, this analysis can be the starting point for a much larger conversation supported with empirical data to further cause positive change in our country.

**Code**

**Initial Data Cleaning Code**

library(dplyr)

*library(ggplot2)*

*library(stringr)*

*library(maps)*

*library(sf)*

*p\_data <- read.csv("police\_fatalities.csv")*

*table(p\_data$Subject.s.age)*

*table(p\_data$Cause.of.death)*

*str(p\_data)*

#Data cleaning

#replace subject.age missing values with NA, change ages less than 15 to NA

#create gender variable and assign subject.race with imputation values and change it as factor

#filter data to cause of death by: Gunshot, Stabbed, chemical agent/pepper spay, Tasered, Asphyxiated/Restrained,

#Beaten/Bludgeoned with instrument, vehicle

#Removing unused columns

*drop <- c("URL.of.image.of.deceased",*

*"A.brief.description.of.the.circumstances.surrounding.the.death",*

*"Dispositions.Exclusions.INTERNAL.USE..NOT.FOR.ANALYSIS",*

*"Supporting.document.link",*

*"Link.to.news.article.or.photo.of.official.document",*

*"Unique.ID.formula",*

*"Unique.identifier..redundant.",*

*"Video",*

*"Date.Description")*

*p\_data <- p\_data[, !names(p\_data) %in% drop]*

*cause\_list <- c("Gunshot","Pursuit","Asphyxiated/Restrained","Beaten/Bludgeoned with instrument","Vehicle",*

*"Chemical agent/Pepper spray","Gunshot","Stabbed")*

*myData <- filter(p\_data, p\_data$Cause.of.death %in% cause\_list)*

#Changing empty cells to NA

*na\_if(myData, " ")*

#change values

*myData <- myData %>% mutate(Cause.of.death = replace(Cause.of.death, which(Cause.of.death == "Beaten/Bludgeoned with instrument"), "Beaten"))*

*myData <- myData %>% mutate(Cause.of.death = replace(Cause.of.death, which(Cause.of.death == "Vehicle"), "Pursuit"))*

*myData <- myData %>% mutate(Cause.of.death = replace(Cause.of.death, which(Cause.of.death == "Asphyxiated/Restrained"), "Asphyxiated"))*

#rename columns

*myData = dplyr::rename(myData, death\_cause = Cause.of.death)*

*myData = dplyr::rename(myData, age = Subject.s.age)*

*myData = dplyr::rename(myData, race = Subject.s.race.with.imputations)*

*myData = dplyr::rename(myData, date\_of\_death = Date.of.injury.resulting.in.death..month.day.year.)*

*myData = dplyr::rename(myData, city = Location.of.death..city.)*

*myData = dplyr::rename(myData, county = Location.of.death..county.)*

*myData = dplyr::rename(myData, state = Location.of.death..state.)*

*myData = dplyr::rename(myData, zip = Location.of.death..zip.code.)*

*myData = dplyr::rename(myData, street = Location.of.injury..address.)*

*myData = dplyr::rename(myData, gender = Subject.s.gender)*

*myData = dplyr::rename(myData, name = Subject.s.name)*

*myData = dplyr::rename(myData, year = Date..Year.)*

*myData = dplyr::rename(myData, agency = Agency.responsible.for.death)*

*myData = dplyr::rename(myData, lat = Latitude)*

*myData = dplyr::rename(myData, long = Longitude)*

# cleaning values for race

*myData <- myData %>% mutate(race = replace(race, which(race %in% c("HIspanic/Latino","Hispanic/Latino") ), "Hispanic"))*

*myData <- myData %>% mutate(race = replace(race, which(race %in% c("Race unspecified",NA) ), "Unknown"))*

*myData <- myData %>% mutate(race = replace(race, which(race %in% c("Native American/Alaskan","European-American/White") ), "White"))*

*myData <- myData %>% mutate(race = replace(race, which(race %in% c("Other Race","Middle Eastern") ), "Other"))*

*myData <- myData %>% mutate(race = replace(race, which(race == "Asian/Pacific Islander" ), "Asian"))*

*myData <- myData %>% mutate(race = replace(race, which(race == "African-American/Black" ), "Black"))*

#change race as factor

*myData$race = factor(myData$race, levels = c("White","Black","Hispanic","Asian","Other","Unknown"))*

#cleaning gender values

*myData <- myData %>% mutate(gender = replace(gender, which(gender == "" ), "Unknown"))*

*myData$gender = factor(myData$gender, levels = c("Male","Female","Transgender","Unknown"))*

#categorize agency

*myData <- myData %>% mutate(agency = replace(agency, str\_detect(agency, "Sheriff"), "Sheriff Office")) %>%*

*mutate(agency = replace(agency, str\_detect(agency, "Police"), "Police Department")) %>%*

*mutate(agency = replace(agency, str\_detect(agency, "Federal Bureau"), "FBI"))*

*main\_agency <- c("Sheriff Office","Police Department","FBI")*

#Create a not in function

*`%!in%` = Negate(`%in%`)*

*# change remaining agencies to Others*

*myData <- myData %>% mutate(agency = replace(agency, which(agency %!in% main\_agency), "Others"))*

#change agency as factor

*myData$agency = factor(myData$agency, levels = c("Sheriff Office","Police Department","FBI","Others"))*

#cleaning age

*myData$age = substr(myData$age,1,2)*

*myData$age = as.numeric(myData$age)*

*myData <- myData %>% mutate(age = replace(age, which(age < 10 ), NA))*

#Bucket ages in groups of 10s

*myData$age\_grps <- cut(myData$age, breaks=c(10, 20, 30, 40, 50, 60, 70, 80, 90, 100), right = FALSE)*

#Create variable for year of death

*myData$year\_of\_death = as.numeric(substr(myData$date\_of\_death,7,length(myData$date\_of\_death)))*

#Final data set

*df <- select(myData,name, age,age\_grps,race,gender,street,city,state,zip,county,date\_of\_death,death\_cause,year\_of\_death,agency)*

*str(myData)*

*str(df)*

*unique(myData$year\_of\_death)*

*unique(police\_data$Cause.of.death)*

**Code for Plots form Section 3**

**Figure 1.1**

# **Creating table of state freq for death**

*state\_count <- as.data.frame(table(myData$state))*

*state\_count <- dplyr::rename(state\_count, state = Var1)*

*# Plotting with state\_count df*

*plot\_usmap(data = state\_count, values = "Freq",*

*labels = TRUE, label\_color = "black", color = "black") +*

*scale\_fill\_continuous(low = "white", high = "red",*

*name = "Number of Deaths",*

*label = scales::comma) +*

*theme(legend.position = "left") +*

*labs(title = "Number of Deaths from Police by State 2000-2020",*

*subtitle = "Total number of deaths per state, not adjusted*")

#**death map**

*us\_map <- map\_data("state")*

*region\_key <- data.frame(state.abb, state.name)*

*#****Rename columns***

*region\_key = dplyr::rename(region\_key, state = state.abb)*

*region\_key = dplyr::rename(region\_key, region = state.name)*

*region\_key$region <- tolower(region\_key$region)*

**Figure 1.2**

# **LOAD STATE POP**

*state\_pop <- read.csv("State Populations.csv")*

*state\_pop$State <- tolower(state\_pop$State)*

*state\_pop <- dplyr::rename(state\_pop, pop = X2018.Population )*

*state\_pop <- dplyr::rename(state\_pop, region = State )*

*df <- left\_join(df, region\_key, by = "state")*

*murder <- data.frame(table(df$region))*

*murder <- dplyr::rename(murder, region = Var1 )*

# **join murder with state pop**

*murder <- left\_join(murder, state\_pop, by = "region")*

*us\_map <- left\_join(us\_map,murder, by = "region")*

*us\_map$ratio <- (us\_map$Freq/us\_map$pop)\*100*

*ggplot(us\_map, aes(x = long, y=lat, group = group))+*

*geom\_polygon(aes(fill=Freq),color = "black")+*

*coord\_quickmap() +*

*scale\_fill\_gradient(low = "white", high = "brown")*

# **map by pop ratio**

*ggplot(us\_map, aes(x = long, y=lat, group = group))+*

*geom\_polygon(aes(fill=ratio),color = "black")+*

*coord\_quickmap() +*

*scale\_fill\_gradient(low = "white", high = "#037ffc")+*

*labs(title="Ratio of Death Caused by Law Enforcers Based on Population")*

**Figure 1.2b**

*city\_count <- as.data.frame(table(myData$city))*

*city\_count <- dplyr::rename(city\_count, city = Var1)*

*city\_count <- arrange(city\_count, -Freq)*

*city\_count <- city\_count[1:10, ]*

*ggplot(data = city\_count) +*

*geom\_bar(aes(x = Freq, y = reorder(city, Freq)),*

*stat = "identity", fill = "#F15533", color = "black") +*

*labs(title = "Top 10 Cities With Police Brutality Death",*

*subtitle = "From 2000 to 2020", x = "Number of Deaths", y = "Cities") +*

*theme(axis.title.y = element\_text(margin = margin(t = 0, r = 15, b = 0, l = 0)),*

*axis.title.x = element\_text(margin = margin(t = 10, r = 0, b = 0, l = 0)))*

**Figure 1.3**

*ggplot(df, aes(x=race, y=age, col= race))+*

*geom\_boxplot() +*

*facet\_wrap(~gender, ncol = 2 )+*

*labs(title = "Age Distribution by race & gender")*

**Figure 1.3b**

*df\_death <- df*

*df\_death <- na.omit(na\_if(df\_death, "Unknown"))*

*ggplot(df\_death, aes(x = race)) +*

*geom\_bar(aes(fill = gender)) +*

*labs(x = "Race", y = "Number of Deaths",*

*title = "Number of Deaths by Race and Gender",*

*caption = "Race only used if imputation confidence was over 80%") +*

*geom\_text(stat = 'count',aes(label =..count.., vjust = -.6), color = "#1A6D0F") +*

*guides(fill = guide\_legend(title = "Gender"))*

**Figure 1.4**

*ggplot(df, aes(x=year\_of\_death, color=race)) +*

*geom\_density(alpha=0.5)+*

*labs(title = "Density Plot of Gender victims by Year")+*

*facet\_wrap(~gender, ncol = 2 )*

**Figure 1.5**

#**Age Distribution by Year and Cause of Death**

*ggplot(df, aes(x= year\_of\_death, y=age, col= death\_cause)) + geom\_point() + facet\_wrap(~gender, ncol = 2) + labs(x = "Year", y = "Age", col = "Cause of Death", title = "Age Distribution by Year and Cause of Death")*

**Figure 1.6**

#**Correlation of Agency and Cause of Death regarding race**

*ggplot(df, aes(x= agency, fill=death\_cause)) + geom\_bar(position = 'dodge') + labs(title = "Agency and Cause of Death Barplot", x = "Agency", y = "Number of Deaths", fill = "Cause of Death") + facet\_wrap(~race, ncol = 2) + theme(axis.text.x = element\_text(angle = 90, hjust = 1))*

**Figure 1.7**

#**Correlation of Agency and Gender regarding race**

*ggplot(df, aes(x = agency, fill = gender)) + geom\_bar(position = 'dodge') + labs(title = "Agency and Gender Barplot", x = "Agency", y = "Number of Deaths", fill = "Gender") + facet\_wrap(~race, ncol = 2) + theme(axis.text.x = element\_text(angle = 90, hjust = 1))*

**Figure 1.8**

#**Correlation of Agency and Cage regarding race**

*ggplot(df, aes(x = agency, fill = age\_grps)) + geom\_bar(position = 'dodge') + labs(title = "Agency and Age Barplot", x = "Agency", y = "Number of Deaths", fill = "Age") + facet\_wrap(~race, ncol = 2) + theme(axis.text.x = element\_text(angle = 90, hjust = 1))*

Reference

**Kaggle data set original source**

Burghart, Brian. “Home.” *Fatal Encounters*, 2020, fatalencounters.org/.