Analysis of Police Related Death From 2015 to 2016

G.Zhu

6/3/2020

R Markdown Introduction

To better understand the topics of police related death in the US. We collected the datasets from kaggle (data science/ML training and file exchange website) to analyze the detail statistics of police related death between the year of 2015 and 2016. All analysis are subjective to the data we received from Kaggle. Each visualization and analysis will be accompanied by short descriptions.

1. Read in 2015 and 2016 police related death data files. Data files are stored in .csv format and no modification were made when downloaded from Kaggle.

```
infile_2015 <- read.csv('/Users/gezhu/Desktop/Police_Related_Death/datasets_617_1199_2015.csv')
infile_2016 <- read.csv('/Users/gezhu/Desktop/Police_Related_Death/datasets_617_1199_2016.csv')
```

Victim profile - Average age of the victim

We would like to know the average, range, and standard deviation of the victim's age between 2015 and 2016. In order to understand the in depth information about the victim age, we have to eliminate the rows with 'unknown' age data.

1. Mean (Average of 2015 victim age) = 37.12073

```
library("stringr")
age_2015 <- infile_2015[,3]
clean_2015 <- str_remove(age_2015, "Unknown")
num_2015 <- as.numeric(clean_2015)
filtered_2015 <- na.omit(num_2015)
mean_2015 <- mean(filtered_2015)
mean_2015
```

```
## [1] 37.12073
```

2. Range (Youngest to oldest victim of 2015) Oldest victim: 87 years old Youngest victim: 6 years old Range: (6 - 87)

```
min_2015 <- min(filtered_2015)
max_2015 <- max(filtered_2015)
range_2015 <- range(filtered_2015)
range_2015
```

```
## [1]  6 87
```

3. Standard Deviation Standard deviation of 2015 victim age = 13.25565

```
sd_2015 <- sd(filtered_2015)
sd_2015
```

```
## [1] 13.25565
```

4. Repeat the above protocols for 2016 data Average age of 2016 victim: 36.72071 Range of the age: Youngest 10 and oldest 87 Standard deviation of 2016 age: 13.00465

```
library("stringr")
age_2016 <- infile_2016[,3]
clean_2016 <- str_remove(age_2016, "Unknown")
num_2016 <- as.numeric(clean_2016)
```

```
## Warning: NAs introduced by coercion
```

```
filtered_2016 <- na.omit(num_2016)
mean_2016 <- mean(filtered_2016)

range_2016 <- range(filtered_2016)

sd_2016 <- sd(filtered_2016)

mean_2016
```

```
## [1] 36.72071
```

```
range_2016
```

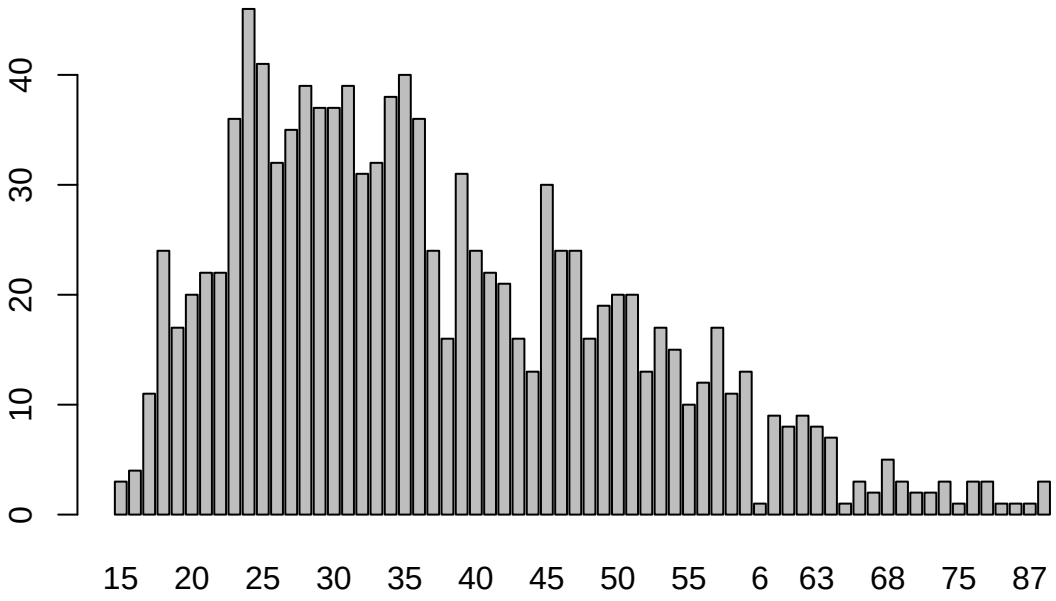
```
## [1] 10 87
```

```
sd_2016
```

```
## [1] 13.00465
```

5. Visualization of Victim Age The barplot shows the frequency of 2015 victim age. Majority of the victims are in their mid 20s to mid 30s.

```
library(MASS)
dataViz = infile_2015$age
dataViz.freq = table(dataViz)
barplot(dataViz.freq)
```



Victim Demographics Now we can analyze the demographics of the victim from 2015 and 2016 police brutality data. 1. Basic demographics breakdown

```
race_data_15 <- infile_2015[,5]
race_data_16 <- infile_2016[,5]
table(race_data_15)
```

```
## race_data_15
##          Arab-American Asian/Pacific Islander                  Black
##                      4                     24                    307
##        Hispanic/Latino        Native American                  Other
##                    195                     13                      1
##                Unknown                  White
##                     18                    584
```

```
table(race_data_16)
```

```
## race_data_16
##          Arab-American Asian/Pacific Islander                  Black
##                      3                     21                    261
##        Hispanic/Latino        Native American                Unknown
##                    183                     21                     25
##                  White
##                    566
```

```
race_df_15 <- as.data.frame(table(race_data_15))
race_df_16 <- as.data.frame(table(race_data_16))
```

2. Visualization of demographics The piechart shows the demographics data of police brutality victims between the year of 2015 and 2016.

- 2015 demographics pie chart

```
library(ggplot2)
bp_15<- ggplot(race_df_15, aes(x="", y=Freq, fill=race_data_15))+
geom_bar(width = 1, stat = "identity")
pie_2015 <- bp_15 + coord_polar("y", start=0)
pie_15 <- pie + scale_fill_manual(values=c("#999999", "#E69F00", "#56B4E9", "#7a425c", "#61137d", "#cf4
pie_15
```

```
## NULL
```

- 2016 demographics pie chart

```
bp_16 <- ggplot(race_df_16, aes(x="", y=Freq, fill=race_data_16))+
geom_bar(width = 1, stat = "identity")
pie_2016 <- bp_16 + coord_polar("y", start=0)
pie_16 <- pie + scale_fill_manual(values=c("#999999", "#E69F00", "#56B4E9", "#7a425c", "#61137d", "#cf4
pie_16
```

```
## NULL
```

3. Gender of the victims 2015 victim gender data

```
gender_data_2015 <- infile_2015[,4]
gender_df_2015 <- as.data.frame(table(gender_data_2015))
gender_df_2015
```

```
##   gender_data_2015 Freq
## 1           Female   52
## 2             Male 1093
## 3   Non-conforming    1
```

2016 victim gender data

```
gender_data_2016 <- infile_2016[,4]
gender_df_2016 <- as.data.frame(table(gender_data_2016))
gender_df_2016
```

```
##   gender_data_2016 Freq
## 1           Female   58
## 2             Male 1022
```

Location of the incidents

Now we can start to look into the location of the incidents. -Incidents location data from 2015

```
location_2015 <- infile_2015[,12]
state_data_2015 <- as.data.frame(table(location_2015))
top5_state_2015 <- as.data.frame(head(state_data_2015[order(-state_data_2015$Freq),],5))
top5_state_2015
```

```
##    location_2015 Freq
## 5             CA  211
## 44            TX  112
## 10            FL   71
## 4             AZ   44
## 11            GA   39
```

-Incidents data from 2016 We select the top five states with highest police shooting incidents for visualizations.

```
location_2016 <- infile_2016[,12]
state_data_2016 <- as.data.frame(table(location_2016))
top5_state_2016 <- as.data.frame(head(state_data_2016[order(-state_data_2016$Freq),],5))
top5_state_2016
```

```
##    location_2016 Freq
## 5             CA  160
## 44            TX   92
## 10            FL   70
## 4             AZ   49
## 28            NC   36
```

- Which police department has the highest number of police shooting? We can count the incident frequency by going through the data from both 2015 and 2016. We display the top ten highest fatality in police shooting incidents here.

```
pd_2015 <- infile_2015[,16]
pd_data_2015 <- as.data.frame(table(pd_2015))
top10_2015 <- as.data.frame(head(pd_data_2015[order(-pd_data_2015$Freq),],10))
top10_2015
```

```
##                                       pd_2015 Freq
## 386              Los Angeles Police Department   22
## 384     Los Angeles County Sheriff's Department   16
## 297                   Houston Police Department   11
## 115                   Chicago Police Department   10
## 473                  New York Police Department   10
## 309 Indianapolis Metropolitan Police Department    9
## 363    Las Vegas Metropolitan Police Department    9
## 30                     Austin Police Department    8
```

```
## 491             Oklahoma City Police Department    8
## 532                   Phoenix Police Department    8
```

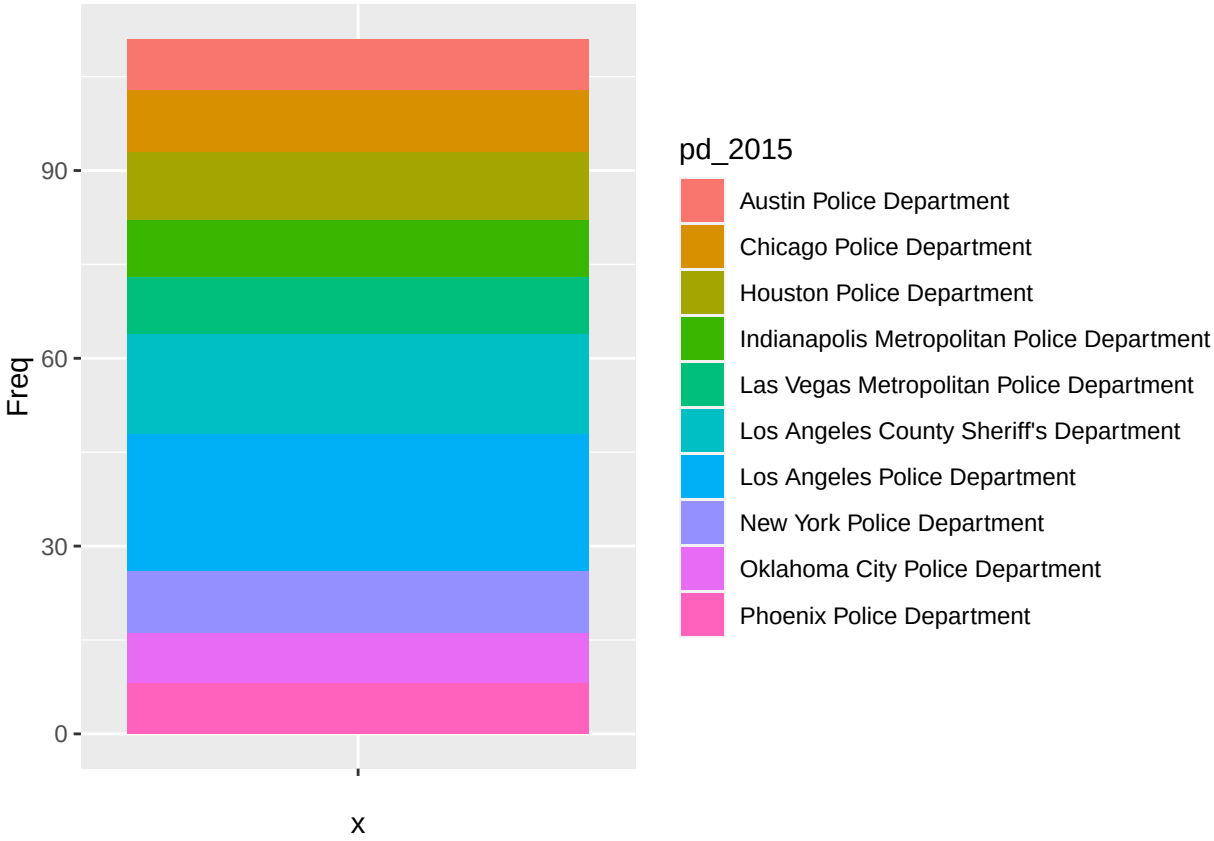
2016 data:

```
pd_2016 <- infile_2016[,16]
pd_data_2016 <- as.data.frame(table(pd_2016))
top10_2016 <- as.data.frame(head(pd_data_2016[order(-pd_data_2016$Freq),],10))
top10_2016
```

```
##                                    pd_2016 Freq
## 368           Los Angeles Police Department   19
## 367 Los Angeles County Sheriff's Department   17
## 496                Phoenix Police Department   16
## 651                                  Unknown   16
## 115                Chicago Police Department   11
## 444               New York Police Department   11
## 560            San Antonio Police Department    9
## 643                  Tulsa Police Department    8
## 649          United States Marshals Service    8
## 95                 California Highway Patrol    7
```

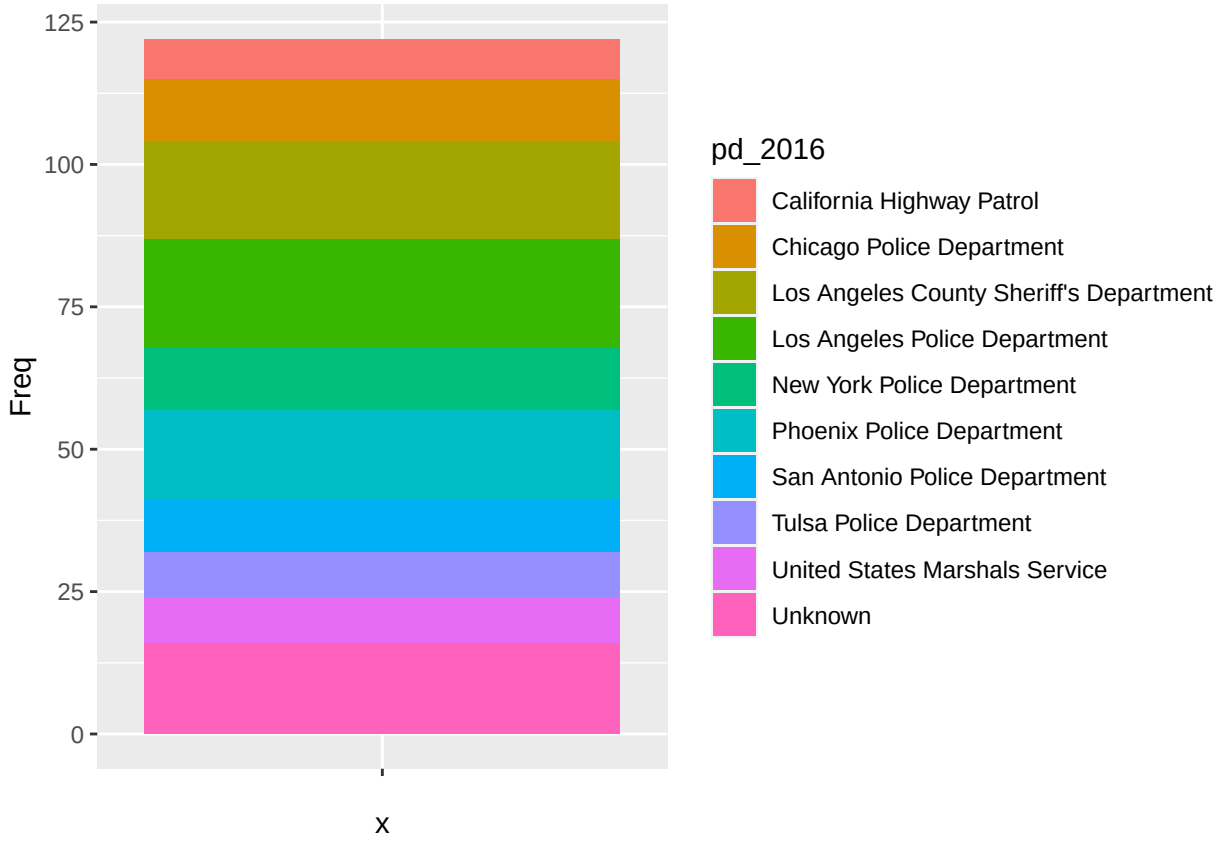
Visualization of the top ten shooting incident police department 2015:

```
library(ggplot2)
bp_pd_2015<- ggplot(top10_2015, aes(x="", y=Freq, fill=pd_2015))+
geom_bar(width = 1, stat = "identity")
bp_pd_2015
```



2016:

```
library(ggplot2)
bp_pd_2016<- ggplot(top10_2016, aes(x="", y=Freq, fill=pd_2016))+
geom_bar(width = 1, stat = "identity")
bp_pd_2016
```



In summary:

The basic data are listed below. Detailed demographics and incidents locations data are attached next to the visualizations above.

-2015 Data Mean age = 37.12073

Range of the age: Youngest 6 and oldest 87

Standard deviation of 2015 victim age = 13.25565

-2016 Data Mean age = 36.72071

Range of the age: Youngest 10 and oldest 87

Standard deviation of 2016 victim age = 13.00465