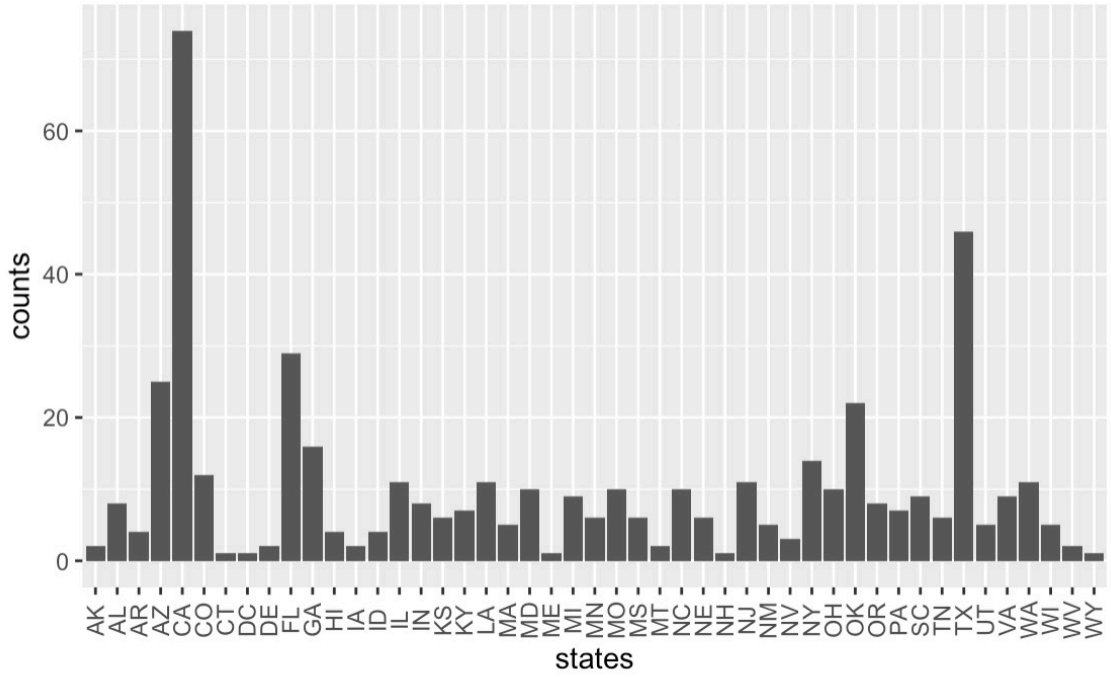
Introduction

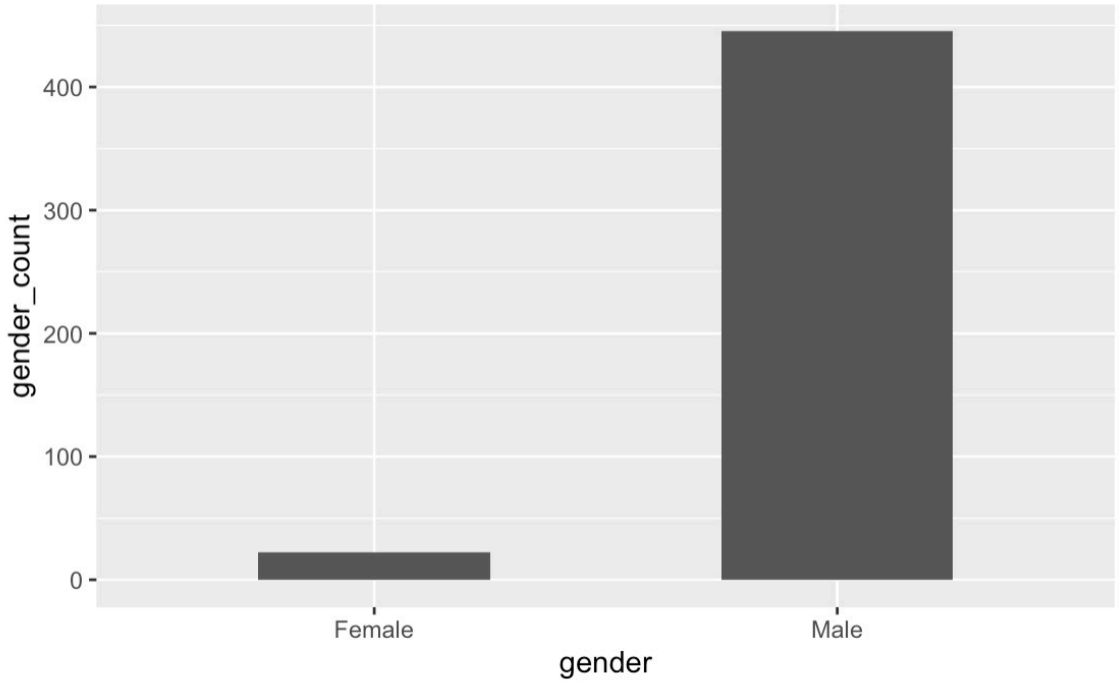
This project is mainly discussed about a very serious problem happening in the United States society, which is the police killing issue.



The graph above is showing the total number of people from each states get killed by the police in 2015. The data that I use to build the graph and the model is from Kaggle. As we can see, the highest number of people get killed by the police is in California, and the second one will be Texas, so the population and the area of the states is somehow effecting the result of police killing.

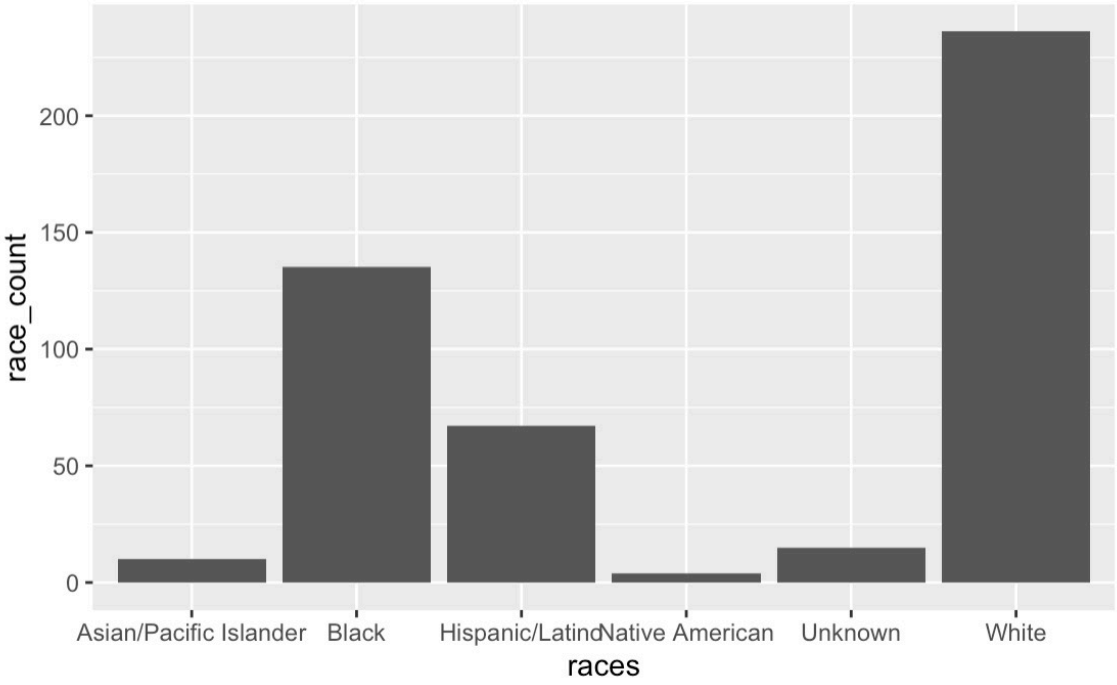
Problem statement

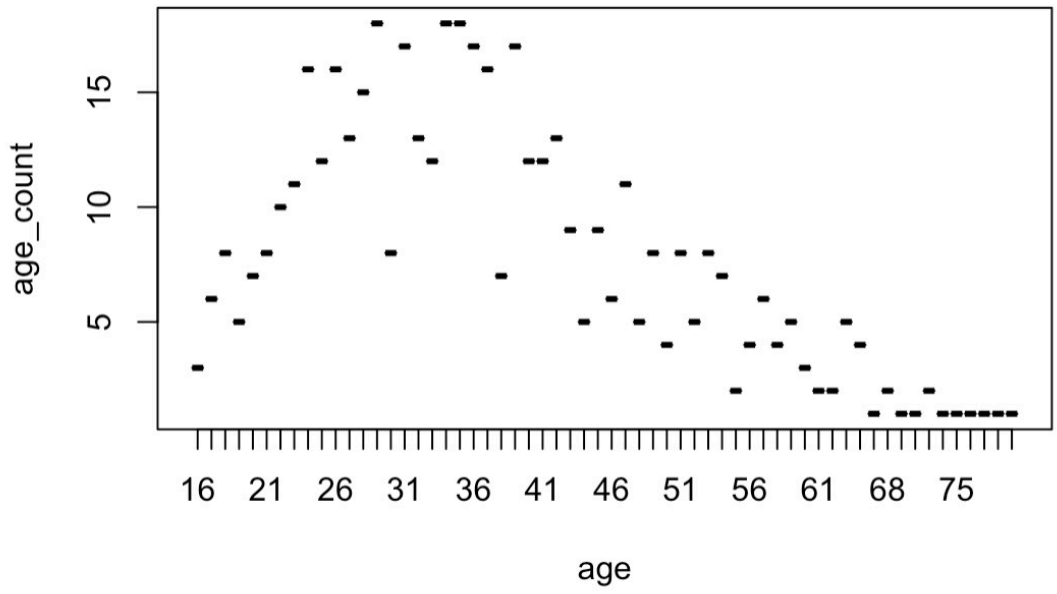
The question that I am trying to find out is to see are people of different gender killed by the police correlated to their age, race and the Area that they are living?



The graph above shows that almost 99% people who get killed by the police are male, and there only 1% are female. And I also create a graph below that shows the number of different race of people get killed by the police. We can see there are three races were mainly killed by police, white people, black people and the Latino, white people has the highest number of people that killed by police, and

then second high is black people, third will be Latino.





The graph above shows different age of people get killed by the police, and we can see the age range from 25 to 40 has the highest

chance that killed by police.

Related work

I did search online that there is a related work show clear relationship between police killing and the race, gender, age and the region of the deceased by using the same dataset. In this related work, it breaking down the number of killings over time by race, then it shows there are a notable number of Black, White, and Hispanic killings. Then it use 2010 census data pulled from Wikipedia to show which racial groups are most frequently victims. Based on the data, it shows that any given Black male in 2015 was more than 3 times more likely to be killed by police than a given White male. And this data also shows the vast gender discrepancy across all racial groups; males are much more likely to be victims of police killings. Moving on to which states have the most police killings, it create a map that show which region has the high rate of police killings per number of state citizens. Next, it show how the average age of victims compares with the average life expectancy. And the result it get is the gap between the average age of death of these victims and the average life expectancy is significant. In this related

work, it also do the analyst to show the police killing correlated with other variable such as income and armed status. It concluded that Police violence that results in death undermines the American criminal justice system by executing potential criminals before they have access to a fair trial.

My Model

In my model, first I did pick the useful variable that related to my question from the dataset. Because my question is seeing how the police killing correlation with the gender, race, age and the living area, so I create a new dataset that only contain those variables. Then I convert gender with number, that male is 1, female is 0, after that I also sort the states into 4 regions: northeast, Midwest, south, west. And I sort the age into 2 categories, if the age lower than 21 then I put it in teenage category, else is in adult category. After the new dataset was created, I use linear regression model to see how the different gender people killed by police correlation with their race, age and the living area. Then use summary() function to get the summary of the result, the result show in below.

```
Call:
lm(formula = sex ~ raceethnicity + age_range + region, data = new_data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.97145  0.02855  0.03860  0.06931  0.12124

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                    0.92278    0.07072  13.048   <2e-16 ***
raceethnicityBlack             0.03471    0.06927   0.501    0.617
raceethnicityHispanic/Latino   0.08670    0.07103   1.221    0.223
raceethnicityNative American   0.11406    0.12381   0.921    0.357
raceethnicityUnknown           0.03374    0.08807   0.383    0.702
raceethnicityWhite             0.04475    0.06769   0.661    0.509
age_rangeteenage              -0.04092    0.03635  -1.126    0.261
regionNortheast               -0.01460    0.04180  -0.349    0.727
regionSouth                    0.00391    0.02896   0.135    0.893
regionWest                    -0.03684    0.03097  -1.190    0.235
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2087 on 453 degrees of freedom
Multiple R-squared:  0.01548,   Adjusted R-squared:  -0.004083
F-statistic: 0.7913 on 9 and 453 DF,  p-value: 0.6246
```

Results

According to the result above, we can see that all those variables have some correlation between each other, the race has the strongest relationship with the police killing and the gender and the age didn’t affect too much on the result.

Conclusions

According to the model, gender didn’t affect too much on the result

of the police killing. The race is the main variable that effect the result of the police killing. In today society, racism is one of the biggest issue in our life, it can affect the fairness of police enforcement, so I hope recruiting police should raise the bar and reduce the police killing.

References:

https://www.kaggle.com/stevechadwick/police-killings-analysis#Analysis: