Final Report

Hening Cui, Jialiang Hua, Congyang Xie, Wenhan Bao, Qihang Wu

Dec 9 2021

Motivation

In the wake of George Floyd's murder, a number of protests arisen across the nation. This motivates us to investigate innocent deaths under police violence and the association of confounders with excessive police force. Although the majority encounters that meet death with law enforcement officials present a lethal threat to officers or society, there are still people that involved in the police action meets death with a quantity not to be ignored. This project analyses the data on people involved in a police operation and results in fatalities. Despite the assessment of overuse of police force across geographical factors, this project also focuses on socioeconomic factors, political factors and periodical patterns.

Initial Questions

- What is the trend of the innocent killing by police over time in the U.S.
- Which state have higher rates of gun ownership, crime rates, and unemployment rates than others.

- Whether the factors above have potential correlations with the number of innocent killings in the U.S.

Data

The main dataset for our project is from Kaggle and to make a more comprehensive analysis, we also combined other datasets from sources like the World Population Review, Statista, and the U.S. Bureau of Labor Statistics, etc. To make this report have both relaistic social signifiance and far-reaching historical significance, we only analyze the cases between 2010 and 2020, even though the main dataset contains a wide time range from 2000 to August 2021.

Exploratory Data Analysis

US Total

The total amount of innocent killings across 50 states by year in the U.S. from 2010 to 2020 is shown as following, which illustrates that the number of cases has increased year by year and stabilized after 2013.
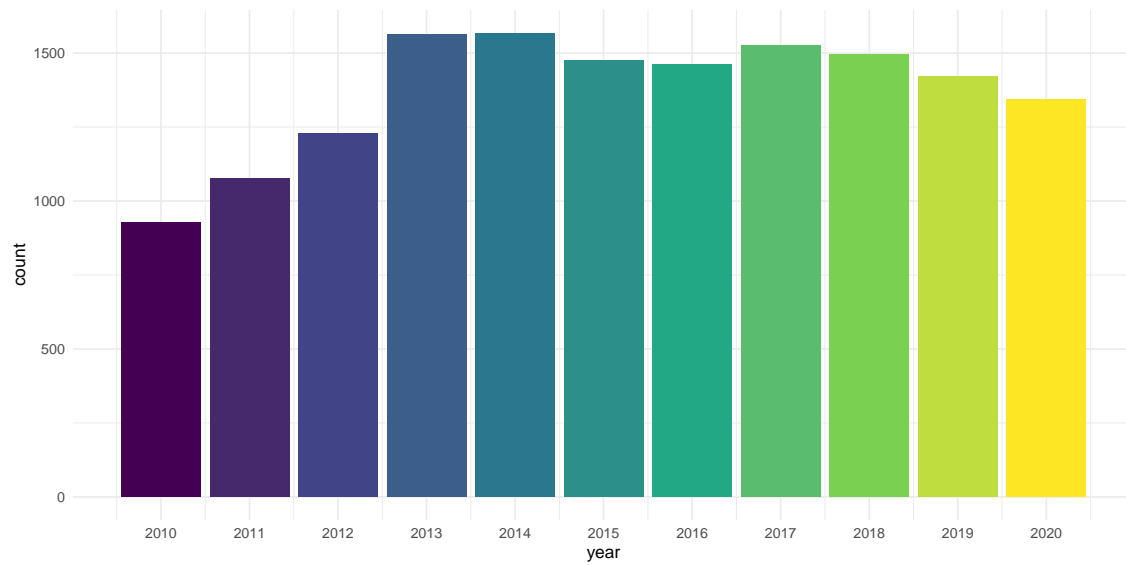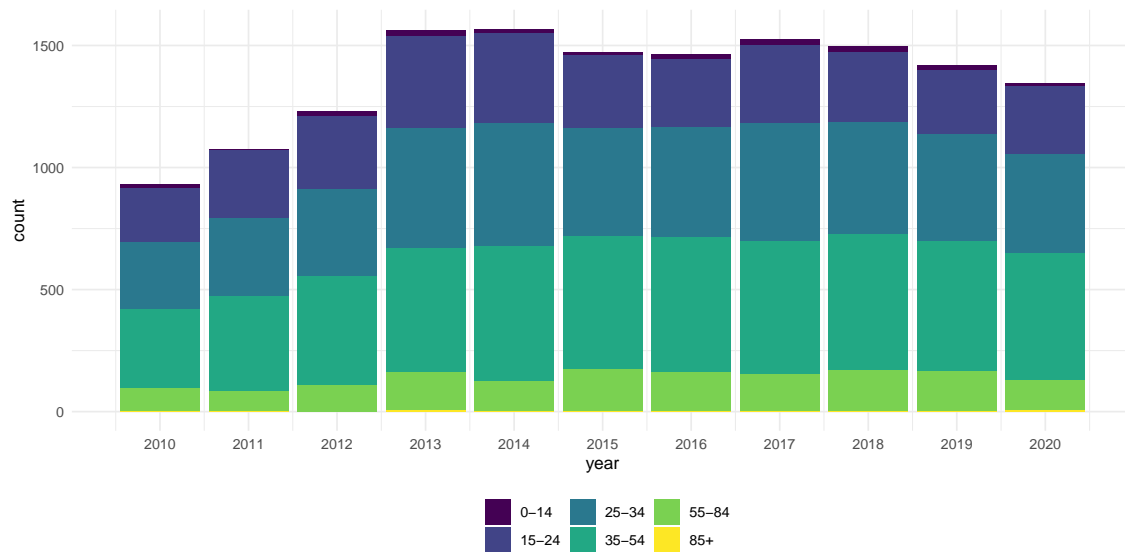
Sex

Graph below shows that sex for each year. Innnocent killings generally occur in the male group, while the proportion of male and female remains roughly the same from 2010 to 2020.
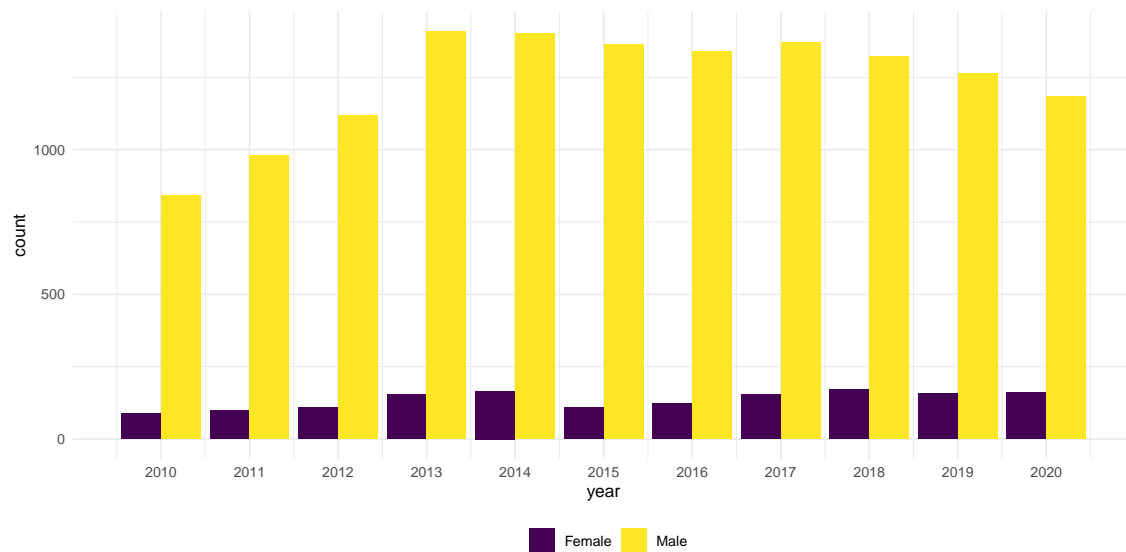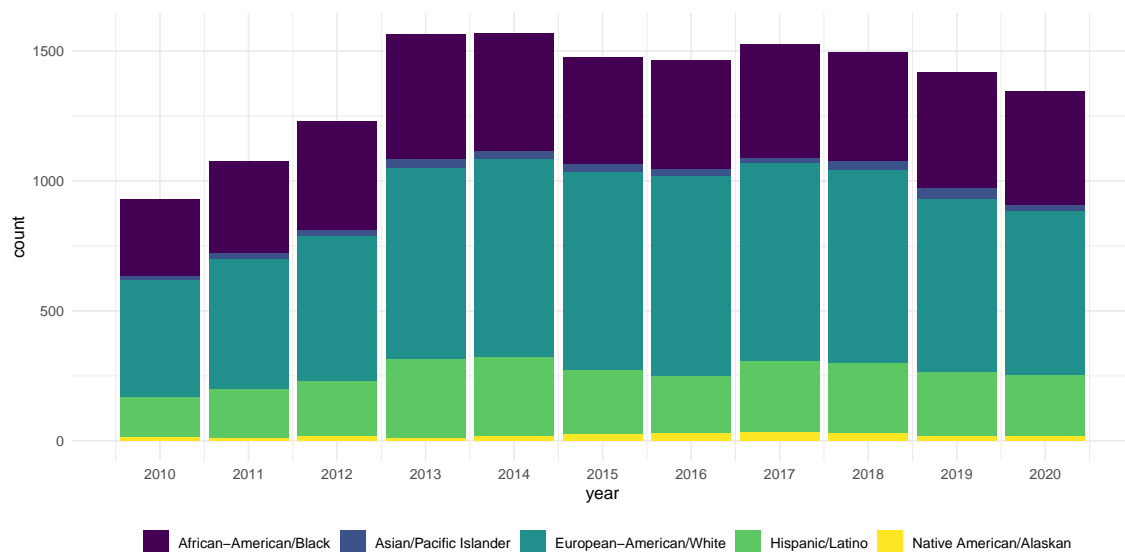

Age

The ratio between different age groups are consistent, and people in the age group (25-84) are more likely to be victims.
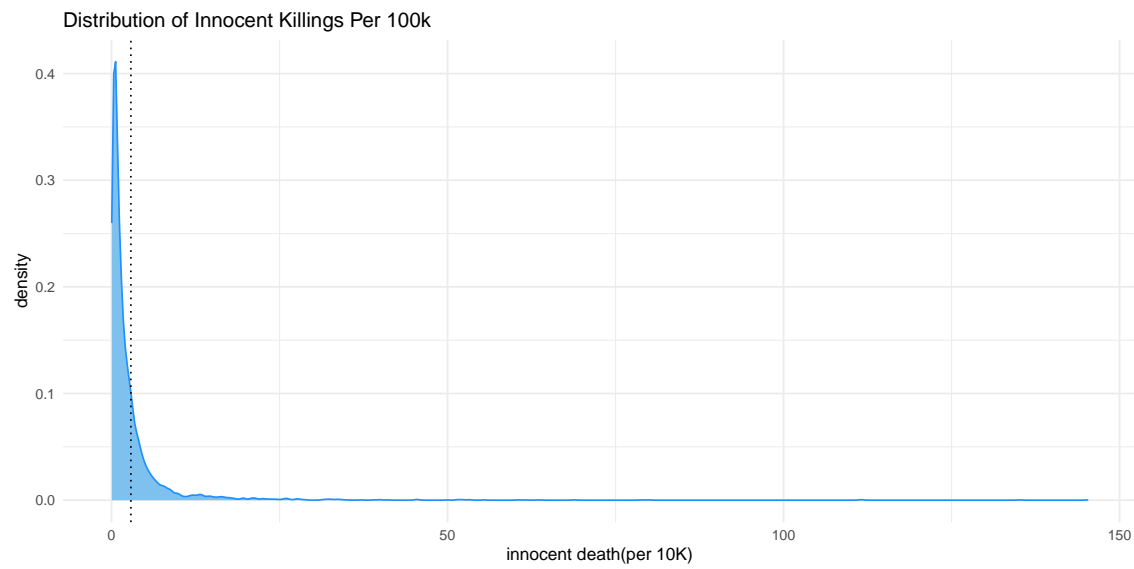
Race and ethnicity

Visualization

Statistical Analysis

For statistical analysis, we come up with two regression models. The first one contains variables including state, age, gender, and race based on the data from year 2010 to 2020, while the second one considers not only the above predictors but also the other three: gun ownership, crime rate, and unemployment rate in 2020. Since we believe that all these three factors may somehow connect with the rate of innocent killings by the police in the U.S. For instance, the police might be armed with more advanced weapons in the states where the crime rate is high, which will lead to a higher rate of innocent killings in these regions.

The following graph is the distribution of the target variable `innocent_kill_per100k`, we can see that the distribution has nothing to do with the normal, therefore we need to perform a transformation.

Distribution of Innocent Killings Per 100k

Now, it looks way more better.

Distribution of Innocent Killings Per 100k after transformation
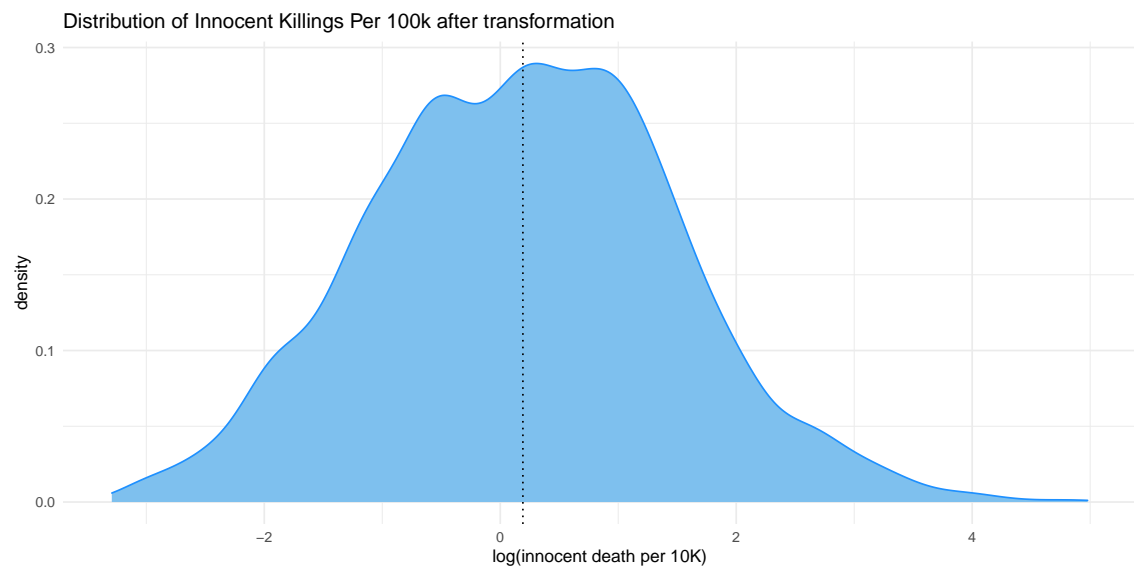
Results

Conclusion

Discussion

Contribution