

Analysis of Police killings in the US

Capstone Project - The Battle of Neighborhoods

Introduction

The 2014 killing of Michael Brown in Ferguson, Missouri, began the protest movement culminating in Black Lives Matter and an increased focus on police accountability nationwide.

Since Jan. 1, 2015, The Washington Post has been compiling a database of every fatal shooting in the US by a police officer in the line of duty. It's difficult to find reliable data from before this period, as police killings haven't been comprehensively documented, and the statistics on police brutality are much less available. As a result, a vast number of cases go unreported.

The Washington Post is tracking more than a dozen details about each killing - including the race, age and gender of the deceased, whether the person was armed, and whether the victim was experiencing a mental-health crisis. They have gathered this information from law enforcement websites, local news reports, social media, and by monitoring independent databases such as "Killed by police" and "Fatal Encounters". The Post has also conducted additional reporting in many cases.

Data

This project aims to determine the relationship between poverty rate, high school graduation rate, median household income, racial demographics and police killings in the US.

Therefore, I've collected on Kaggle a complete collection of the following datasets :

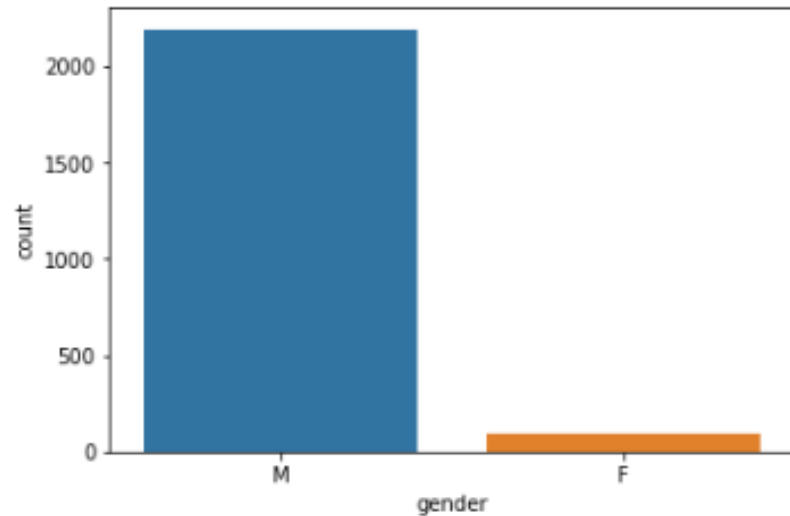
- Police killings data
- Poverty rate data
- High School graduation rate data
- Median household income data
- Racial demographics data

Data has been collected between 02/01/2015 and 31/07/2017.

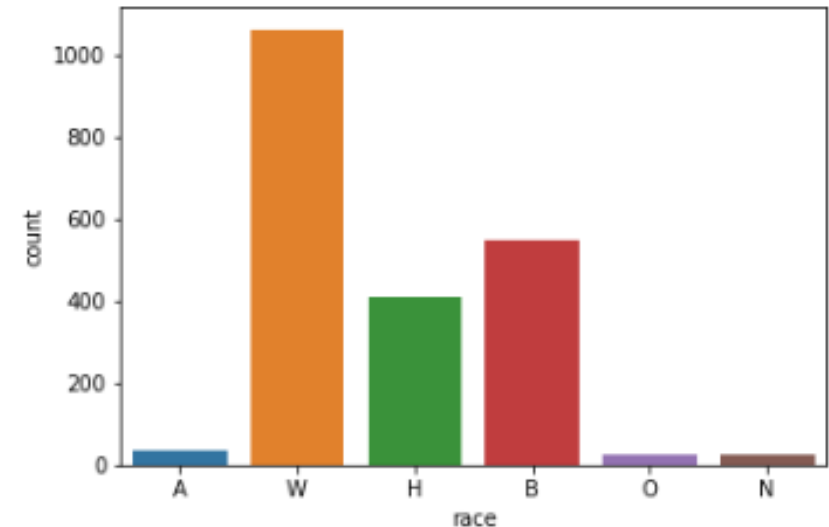
Methodology

- First step was to import datasets on IBM Watson Studio and connecting them to a Jupyter notebook. Police killings data didn't have the same location details as the other data sets. Indeed, the following texts were added after each city names in the other datasets : "city", "village", "town", "CDP". Therefore, I had to harmonize all the locations in each dataset to be able to go further in the analysis.
- Second step was to merge datasets using multiple left joiner to get a final dataset with all data required. After merging datasets, I had to change format of numerous columns from object to float.
- Third step was to visualize categorical data using seaborn to get an overview on the case.
- Final step was to visualize numerical data using matplotlib to capture trends.

Results (1/3)

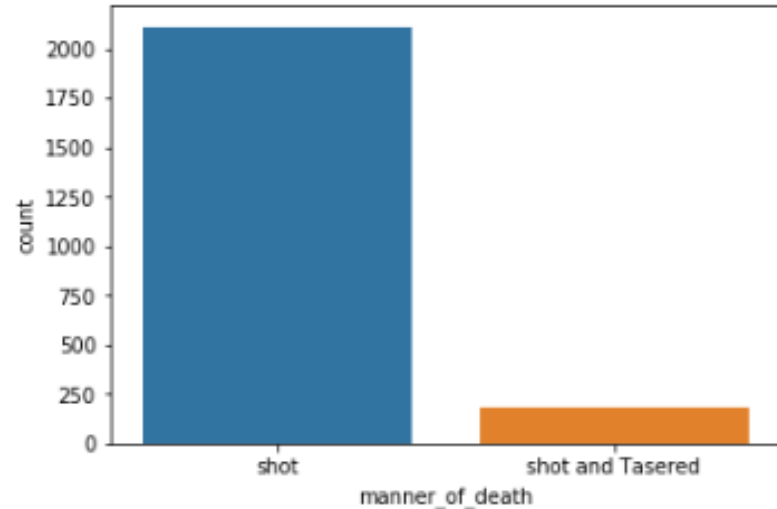


Mens are 23 times more likely to be killed by a police officer than womens.

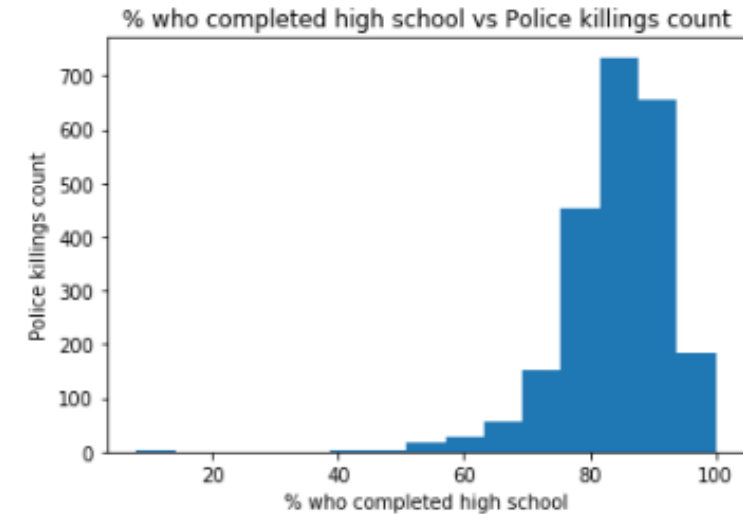


White people are almost 2 times more likely to be killed by a police officer than black people.

Results (2/3)

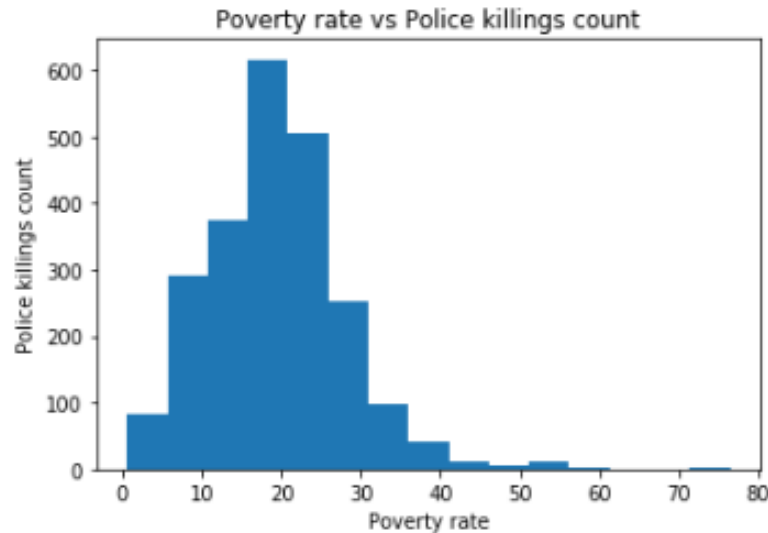


In only 8% of cases, the suspect has been tasered before being shot.

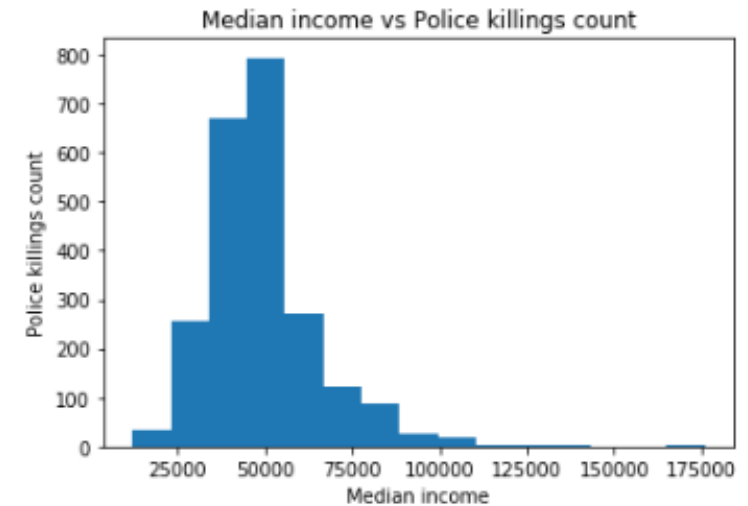


Most of suspect being killed by a police officer lived in an overall well educated area. Mean is at 84% of people who completed high school.

Results (3/3)



Most of suspect being killed by a police officer lived in a modestly poor area. Mean poverty rate is at 19.2%.



Most of suspect being killed by a police officer lived in an average area. Mean median income is at 49.313\$/year.

Conclusion

Contrary to popular belief, according to the data I have been able to analyze, most of the suspects killed by the police do not come from disadvantaged neighborhoods where access to education is complicated.

In addition, statistics show that whites are almost two times more likely to get killed by a police officer than black people.