Analysis of Crime and Police Data

Robert Hall

CPSC4180 Final Project

University of Tennessee: Chattanooga

Analysis of JohnM's Kaggle Data-set and Chattanooga's arrest charges to see trends among gender, racial groups, and armed status of individuals involved in crime and police brutality.

Abstract

After careful organizing, cleaning, this analysis of John M's Kaggle data and Chattanooga's Police data reveals trends among racial groups, gender groups, and armed status in regards to police brutality, citations, and overall crimes. The trends in regards to gender show that males have a higher chance and percentage than females to be the involved in police brutality. Trends among racial groups show that overall white Americans have higher occurrences and numbers compared to other racial groups, but in some instances other groups have higher percentages. Lastly Armed individuals show higher occurrences to be involved in crime and police brutality, but the unarmed percentages is surprisingly high as well.

Data Sources and Tools Used

I. *JohnM's Police Violence & Racial Equity Data Set[1]*

- JohnM's data set from Kaggle consists of multiple .csv files that contain information about citizens that have been killed by police, officers that have been killed in the line of duty, and death arrests. This set is used widely and is an important component to this analysis.

II. *Chattanooga Police Data[2]*

- This data that comes from Chattanooga's own police department can be described consisting of similar pieces just like JohnM's data. The Chattanooga police data on the other-hand has a data set that includes citations and relevant information about them. This set is used to see trends locally to the Chattanooga, TN area and see how they compare to the other sets.

III. *Jupyter Notebooks and Python 3[3]*

- All analytics for the respective data-sets was computed using the program Jupyter Notebooks, and Python 3. Using various methods such as cleaning, visualizations, and reading files were all done using the aforementioned tools.

---

1   https://www.kaggle.com/jpmiller/police-violence-in-the-us

2   https://www.chattadata.org/stories/s/26bg-ejs3

3   https://jupyter.org/

Introductory Methods

Reading and Cleaning the Data

In order to read and modify the information in the data-set files, the Python Data Analysis library pandas was used for these methods. The first data set (police_killings_MPV.csv) is described as each row being a person who has been killed by the police, and the columns being the respective information such as: race, city, state, name, age and race. Using subset techniques such as filtering by a value (whether it be filter by a state or a city), we were able to clean the data set to be used in future graphs and analytics.

Other data-sets from the sources such as: CPD_Arrest_Charges.csv , and fatal_encounters_dot_org.csv were both cleaned and filtered using the aforementioned techniques of filtering by values and data-frame slicing to ensure no erroneous values were involved in the analysis. Listed in Figure I is a sample code from the included Jupyter Notebook that shows basic sub-setting based on gender:

Figure I:
```
citizens_killed = pd.read_csv("police_killings_MPV.csv")
male_killed = citizens_killed[citizens_killed["Victim's gender"] == "Male"]
women_killed = citizens_killed[citizens_killed["Victim's gender"] == "Female"]
```

Handling Missing Values and Unknown Fields

In the data-sets used for this analysis, missing fields and values where handled using sub-setting techniques and filtering them into their own groups to be used in comparison to the correct values to show discrepancies and the total occurrences in the sets.

In the next section there are charts and figured included that show the high number of occurrences of missing values in the data-sets. In regards to the topic of police brutality, these missing values where either not reported due to political issues, or human error. The effect of the missing values in regards to the broader scope is not researched in this analysis. A sample of the police_killings_MPV.csv data-set with the mentioned missing values is labeled and shown in Figure II.

Victim's name	Victim's age	Victim's gender	Victim's race
Name withheld by police	Unknown	Male	Unknown race
Name withheld by police	Unknown	Male	Unknown race
Name withheld by police	61	Male	Unknown race

Figure II: police_killings_MPV.csv missing values and unknown fields sample.

Trends and Results from Analysis

I.   Trends and the Results Among Racial Groups

One of the most important things to look at when it comes to police violence is trends around certain ethnic groups compared to the other one. Using the matplotlib.pyplot package, we were able to show an overall view of the trends using Figure III.
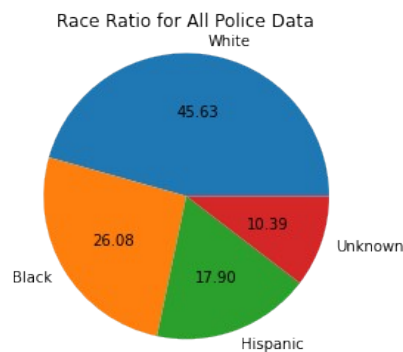

Figure III: Occurrences of each racial group in the police_killings_MPV file.

From looking at the figure, one would see that White Americans take up the majority with 45.63%, with African Americans, Hispanics, and the unknown values taking up 26.08%, 17.90% and 10.39% respectively. Just like previously mentioned in the introduction, the analytics shown here are just for the data-sets from Kaggle and are not the whole encompassing picture, rather just an analysis of this set.

Now that we see the overall occurrences of the racial groups in the set, let's take a look at specific location examples. Using the filter_by_race function from the included Jupyter notebook, we are able to see specific places, such as Tennessee in located in Figure IV.
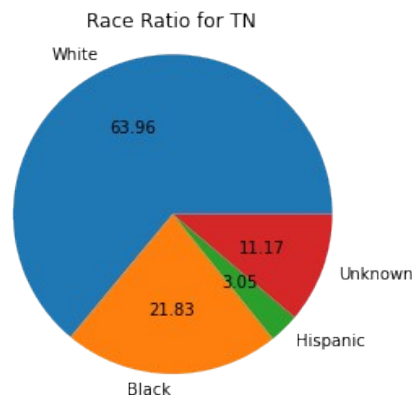

Figure IV: Occurrences of each racial group specifically in Tennessee

As we can see in Figure IV, in Tennessee the trend of White Americans having a higher occurrence in the police killing data set is higher than any other group. By looking at these few examples once can assume it's true across most places. From gathering and looking at population in regards to race, a few states such as in Figures V and Figure VI.
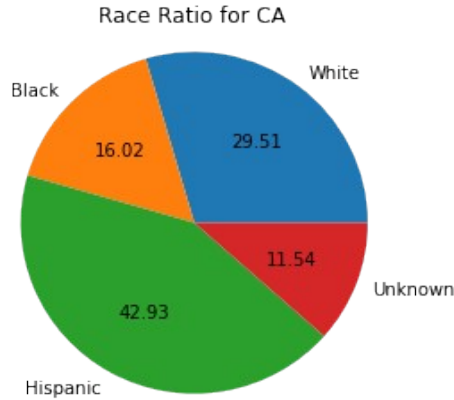

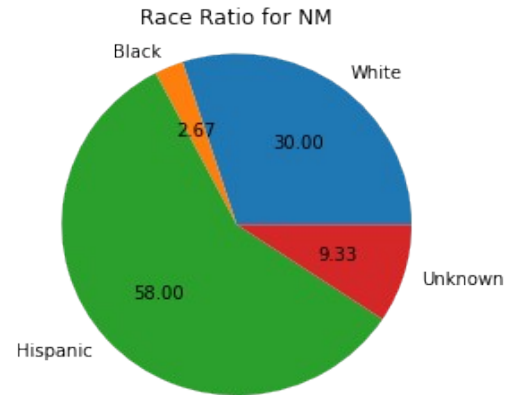
Figure V. Occurrences in California   Figure VI: Occurrences in New Mexico

From looking at these figures, once could see that the Hispanic group takes up over 42% and 58% of the occurrences in their respective states. We can also see that White Americans are both the second highest count in each state as well. While we can look at these pie-charts and conclude White Americans have a higher occurrence, the next few figures show a total numerical breakdown in bar-charts to show numbers rather than overall percentages. Figure VII shows the numerical break down of the same states used in the pie-charts.
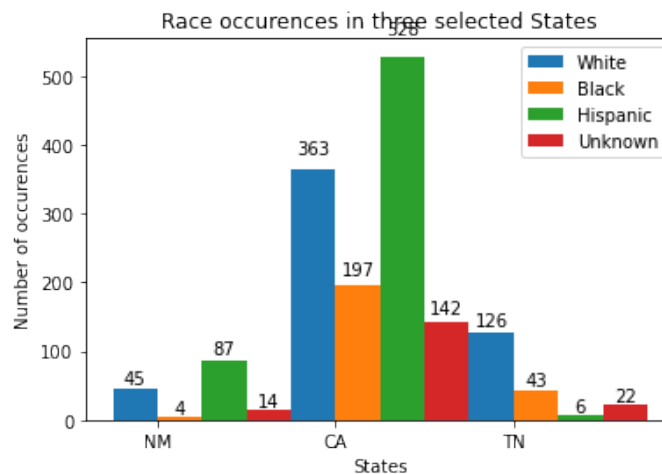

Figure VII: Numerical breakdown of the occurrences of racial groups in the police_killings_MPV file.

As shown in the figure above, the pie-charts calculate percentages of instances, while Figure VII shows us the numerical breakdown. From this figure, we can see that the number of occurrences is not the same across the states. New Mexico has 150 instances from the set, California has 1230 instances from the set, and finally Tennessee has 197 instances from the set. From looking at the figure you can see that White Americans have a higher numerical count than others in the set reinforcing our figures.

Further trends in cities and states follows a similar trend to the aforementioned states and have not been included in this report.

Chattanooga Analysis and Trends Among Race

Now that we have looked at the overall trends from the police killings data-set, let's look at some local examples here in Chattanooga. The data-set being used here is the CPD_Arrest Charges[4] set that is instances of arrests and the information about them (race, gender, etc.) Figure VIII shows the ratio of racial groups broken down by percentages to show the trends:
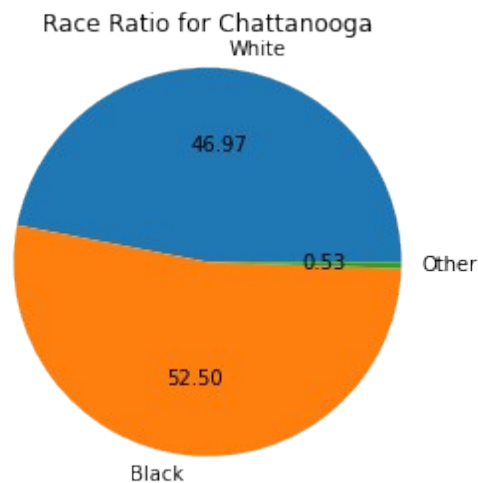

Figure VIII: Ratio of Racial Groups by Percentages in the set.

The most interesting thing about Figure VIII is that there are only White Americans, and African Americans available in the set, excluding the "other" category. After looking through manually at the set, it is confirmed that there are only white Americans and African Americans in this set. In this specific set, it would appear as if African Americans make up the majority with over 52% of occurrences in the set.

---

4   https://www.chattadata.org/Public-Safety/Public-CPD-Arrest-Charges/v9y9-uavb

II.   Trends and the Results Among Gender

Another interesting trend to look at with police violence is gender. Similar to above the code for the functions used to calculate and subset based on gender is in the included Jupyter Notebook. Figure IX shows the overall percentages based on gender in the data-set.
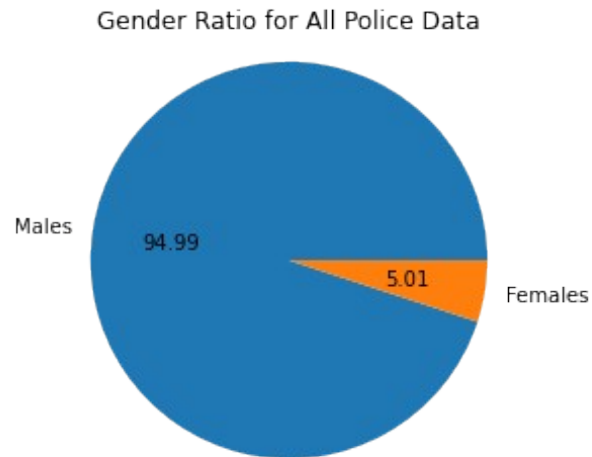

Figure IX: Overall percentages based on gender in police_killings_MPV[5] data-set file.

As we can see in the figure above, males take up the majority with 94.99% of all members killed by police, and females with their 5.01% occurrence in the set. With large percentages such as this, looking at a location-by-location basis reinforces the fact that these percentages make sense.

Using the same states as above in the included figures, we can see that the overall male occurrences are higher than that of the female groups.

Gender Ratio for NM

Males 93.46

6.54 Females
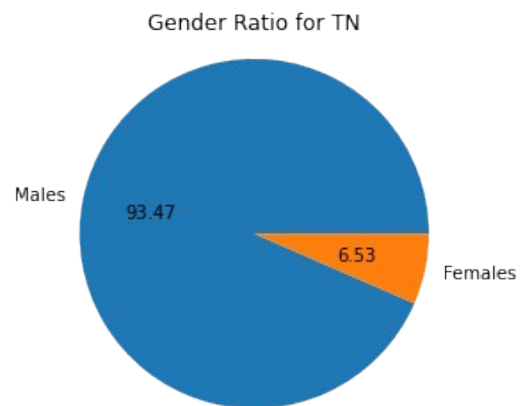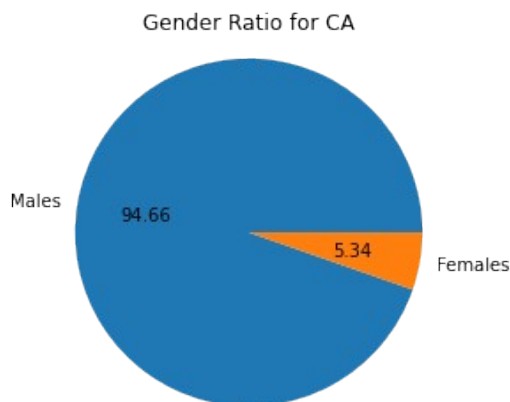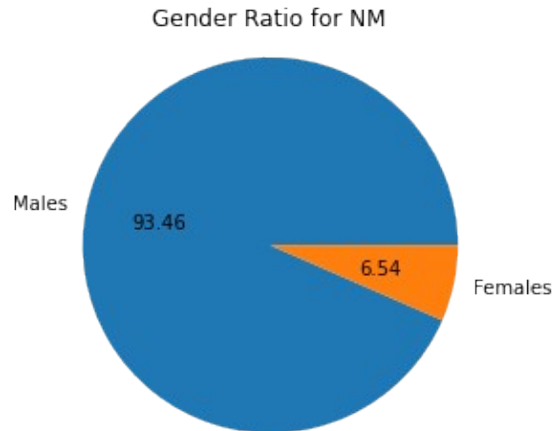
Figures X – XII : Each state's occurrence of gender in the police_killings_MPV file.

As we can see in the figures above, in all states males make up the majority when it comes to police killings. This trend is clear across the data-set for the whole and each individual state as well leading to conclude that males have a higher chance than females to be the victim of police brutality in respect to this data-set.

Chattanooga Analysis and Trends Among Gender

Using the same data-set as the one in the previous example (Chattanooga Arrest Charges) the trends are similar to the police killings set, but at different percentages, Figure XIII shows the percentages of males to females in regards to arrest charges:

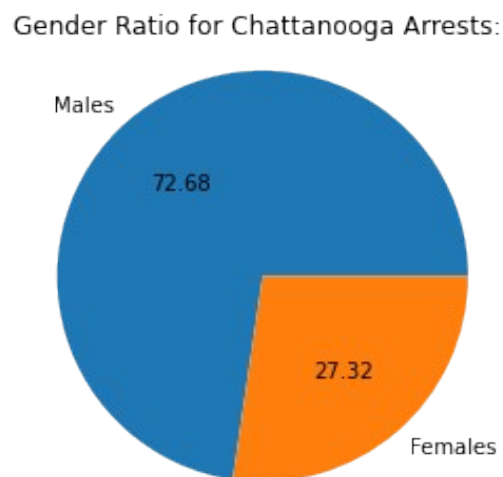Gender Ratio for Chattanooga Arrests:

Males

72.68

27.32

Females

Figure XIII: Percentages of males to females in Chattanooga Arrest Charges

One can see from the figure above that the trend that males have higher occurrences than females in regards to crime are higher for both this Chattanooga data-set and the police brutality set. Factoring in the number of males compared to females in the U.S, this trend can be validated by other analytics. For this analysis, it seems as if males do indeed have a higher occurrence than females when it comes to arrests.

III.   Trends and the Results Among Weapon Status

One of the last things to look at when it comes to police violence and brutality is the armed status of the offender. In recent days there has been movements for police reform and overall change due to the occurrences of unarmed police brutality, the included figures later on show us numbers and percentages to see if there is a higher occurrence of unarmed vs armed police violence.

In the included Figure XV it shows the overall percentages of whether the individual was armed, unarmed, or was using a vehicle whenever the instance occurred:
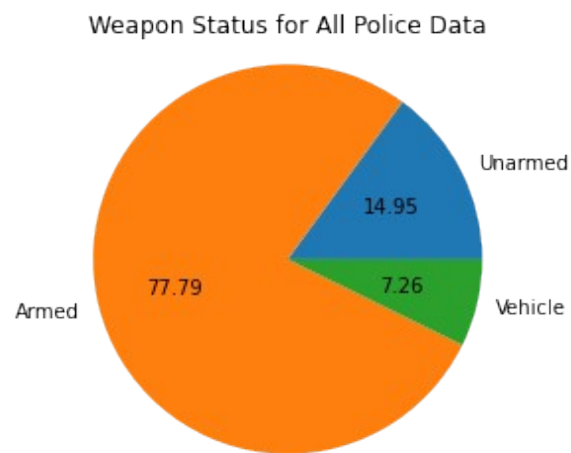

Figure XV: Percentages of Weapon Status for police_killing_MPV file.

From looking at the above figure, one could see that Armed occurrences of police brutality has a majority of 77.79, and 14.95 and 7.26% for unarmed and vehicle respectively. Constituting together Armed and Vehicle occurrences would have a 85.05% majority compared to the unarmed instances. Seeing how Armed percentages in the set are larger than the others, it is still important to note how high the amount of unarmed instances are in the set.

To further look at specific states and cities to see if this trend holds true for all or some leads us to compute Figures XVI-XVIII using the similar sub-setting techniques used for previous calculations based on location basis. The included figures show the breakdown of percentages for armed status using the previously used states of California, New Mexico and Tennessee respectively:
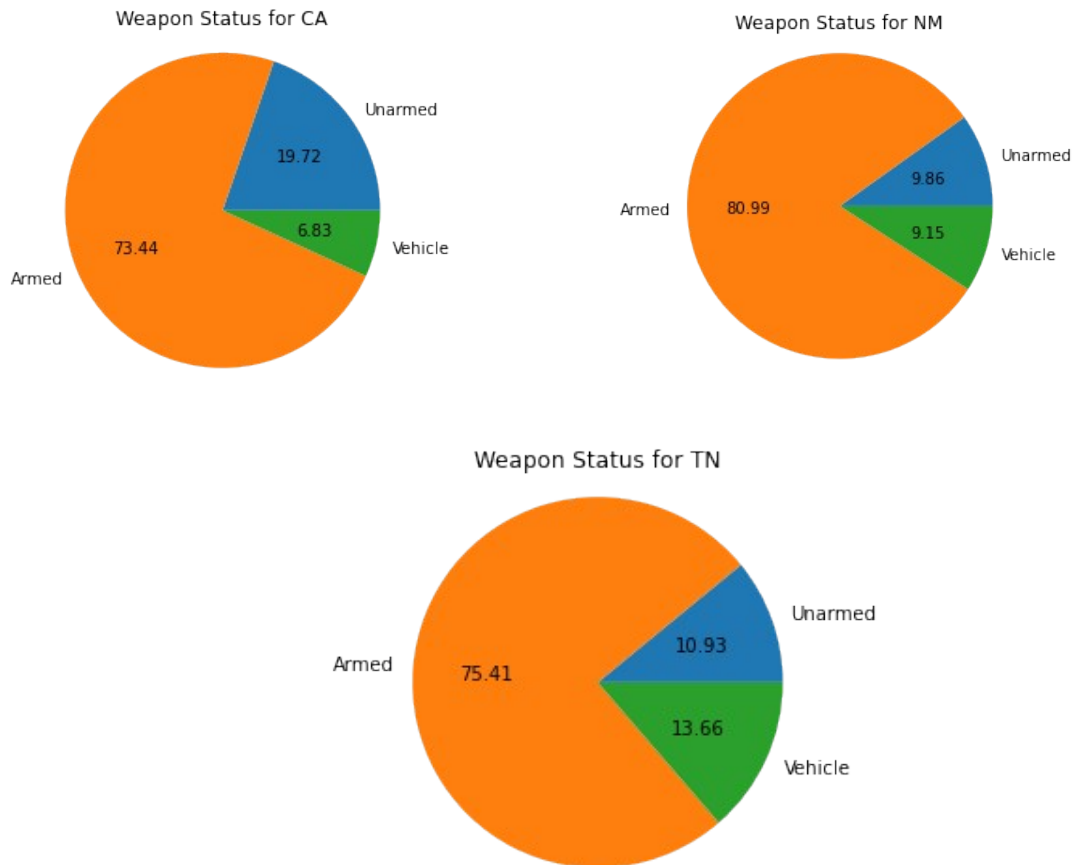
Figures XVI-XVIII: California, New Mexico, and Tennessee percent occurrences of armed vs unarmed statuses in the police_killings_MPV file.

Other states were computed and to have found similar trends and were not included in this report but were used in overall analysis and conclusion. It would appear that Armed individuals indeed take up a majority in the police_killings_MPV file, and this statement holds true to the computed figures for the states. It is also important to note how high the unarmed ratios are, in all computed figures it constitutes between 10-15% overall occurrences in the set.

Chattanooga's open data used in this analysis did not have columns or values relating to armed status of the individual for the interactions, so it was not included in this section.

Conclusions and Closing Remarks

Overall the data-sets used in this analysis shows interesting trends when it is broken down into certain groups such as gender, racial groups, and armed status. After computing and calculating there shows a clear trend that males have higher occurrences than females to be the victim of police brutality. Using other locations to reinforce this, the selected states and locations all shown the same trend that males have a higher majority. In the Chattanooga data analysis this statement that males have a higher occurrence than females stands true as well.

In regards to being armed compared to unarmed occurrences in the data-sets, Armed individuals make up the majority compared to the other groups such as unarmed and vehicle instances. Further analysis shows that the unarmed populations in the police_killings_MPV file show between 10-15% of each set whether it was broken down by state or overall analysis of the set. This number was staggeringly high and is an interesting point in this analysis.

Lastly, in the computations and figures provided show that White Americans have a higher percentage to be killed by police and take up a decent majority for the Chattanooga instances as well. Further analysis in a state-by-state breakdown shows that in some states such as California or New Mexico, Hispanic Americans take up a higher majority than the other racial groups. This analysis was more in-depth due to the number of instances of locations that have higher occurrences for one group compared to the hypothesis that White Americans are more likely to be victims of police brutality. In the Chattanooga data analysis, it shows that African Americans have a higher percent in terms of arrest charges than White Americans. It is also important to note for the Chattanooga analysis for race, the data-set only has instances of White and African Americans compared to the other set which included Hispanic Americans and others.

Overall purpose of this analysis is to show trends for individuals in regards to police violence and arrests for their respective data-sets. The data-sets used for this analysis were gathered from their respected sources and can serve as a good base-line for overall trends for all police violence and charges but this analysis does not show the general trend, but rather the trend for these specific instances in the sets.

References

John M's and JP Miller's Police Violence & Racial Equity

- https://www.kaggle.com/jpmiller/police-violence-in-the-us

Chattanooga's Open Data Policing and Racial Equity

- https://www.chattadata.org/stories/s/26bg-ejs3

Jupyter Notebook

- https://jupyter.org/

Matplotlib

- https://matplotlib.org/