

Analysis of Police Homicides in America (2015)

Summary of Research Questions:

1. Which state experienced the highest number of police killings as compared to other states in 2015?

This question will be addressed by computing killings per capita for each state and then comparing the resulting values. Since, state populations are variable, a per capita measurement will provide a more accurate representation of the situation. The answer to this question is important because it can help understand which state contributed the most to the total number of police killings in 2015.

2. From the data, can it be inferred that people who were armed were more likely to be victims of police homicide in 2015 when compared to people who were unarmed?

To answer this question, the likelihood of being a victim of police homicide will be calculated for both sections of the American population – armed and unarmed. This answer can help understand a police officer's decision to shoot and can play an important role while debating if an officer's decision to shoot was justified.

3. Were people from a particular section of the American population (race / ethnicity) more likely to be victims of police homicide?

To answer this question, police killings per million (adjusting for variability in population of each race) for each race / ethnicity will be calculated and compared. The answer will help understand if any racial bias was involved in the decision to shoot.

4. Is the current trend (2016) of police killings different from the trend in 2015?

To answer this question, the total number of police killings in 2016 by month will be compared with the total number of police killings in 2015 by month.

Motivation and Background

Police in the US has often been accused of brutality and abuse of authority. Rallies and protests have been carried out against police brutality. However, according to the Guardian, the US government has no comprehensive record of the number of people killed by law enforcement. This lack of basic data has been glaring amid the protests, riots and worldwide debate set in motion by the fatal police shooting of Michael Brown, an unarmed 18-year-old, in Ferguson, Missouri, in August 2014. Analysis of this data will help uncover trends which can be used as evidence and hence contribute to an informed public discussion about the use of force by the police. Analyzing and visualizing data for the highest number of police killings by state can uncover information about a particular state that has been contributing the most to the total number of police killings and thus, should take steps to prevent such incidents in the future. Police brutality is also associated with racial profiling. Some sections of the US population perceive victims of police homicide as relatively powerless groups such as racial or cultural minorities. If the analysis of the data shows that a particular section of the American population (race / ethnicity) is more likely to be a victim of police

homicide, then this data can be used by law enforcement agencies to take steps to protect this particular section of the US population. A look at the number of police killings for armed vs. unarmed can help address the issue of justified killing. If there is a low difference between the number of killings (armed vs. unarmed) per capita, then it can be questioned if the actions of the officers were justified. A comparison of the police homicide trend in 2015 and 2016 can help make conclusions about preventive and public safety measures taken based on the difference in the number of killings between 2015 and 2016.

Dataset

The program utilizes 3 data files:

1. Guardian's interactive database of *Americans killed by police – The Counted* (<http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/the-counted-police-killings-us-database>) - A .zip file containing two comma-separated values (CSV) files – 'the-counted-2015.csv' and 'the-counted-2016.csv' – can be downloaded from <http://www.theguardian.com/us-news/ng-interactive/2015/jun/01/about-the-counted>. However, I have made both these files available via Dropbox so that the program can access them directly:
 - a. the-counted-2015.csv: <https://www.dropbox.com/s/ohknei1ivvogq7z/the-counted-2015.csv?dl=0>
 - b. the-counted-2016.csv: <https://www.dropbox.com/s/h9kff40kamgut69/the-counted-2016.csv?dl=0>

The data set includes demographic information on those killed, such as age, sex and race, as well as address of where the incident occurred and cause of death.

2. National, State, and Puerto Rico Commonwealth Totals Datasets: April 1, 2010 to July 1, 2015 - I have obtained this data from the United States Census Bureau's website. The data can be accessed in CSV format from the following link - <http://www.census.gov/popest/data/national/totals/2015/files/NST-EST2015-alldata.csv>. I will be using this dataset to extract the individual state populations (2015) in the US. The dataset contains population, population change, and estimated components of population change by state and year – change in total resident population, net migration, etc.
3. Projected Population by Single Year of Age, Sex, Race, Hispanic Origin and Nativity for the United States: 2014 to 2060 – I have obtained this data from the United States Census Bureau's website. The CSV version of the data can be accessed from https://www.census.gov/population/projections/files/downloadables/NP2014_D5.csv. The dataset contains population projections by The data contains various variables – Origin, Nativity, Race, Sex, Year, Total Population and the population age. I will be using this data file to obtain the total population (2015) of individuals by race and origin in the US.

Methodology:

The program will access all 3 data files directly via the web. Data will be obtained and stored after performing data cleaning.

1. First question – To answer this question, the program will access Guardian’s ‘The Counted’ data file (the-counted-2015.csv) and the ‘National, State, and Puerto Rico Commonwealth Totals Dataset’ – state populations - data file.
 - a. The program will utilize ‘the-counted-2015.csv’ data file to obtain the total number of killings for each state. The total number of killings for a particular state will be mapped to that state.
 - b. The total population of each state will be mapped to that state and this data will be accessed from the state populations data file.
 - c. The program will compute the number of killings per capita for each state and then map the data to the respective state.
 - d. A histogram for state vs. killings per capita will be created and saved as ‘state_killings.png’ and the name of the state with the largest number of per capita killings will be printed.
2. Second Question: ‘the-counted-2015.csv’ data file will be required to help address this question.
 - a. The total number of killings will be mapped to the respective weapon or unarmed.
 - b. A pie chart will be created displaying the number of killings for each armed / weapon and the number of killings for unarmed. The pie chart will be saved as ‘armed_unarmed.png’
 - c. However, all weapons will be displayed in the same color (belong to armed section) and with the weapon name as the label. The unarmed section will be in a different color.
3. Third Question: The program will utilize ‘the-counted-2015.csv’ data file and ‘Projected Population by Single Year of Age, Sex, Race, Hispanic Origin and Nativity for the United States’ – race populations - data file to help answer this question.
 - a. The total number of killings for each race (obtained from ‘the-counted-2015.csv’) will be mapped to that particular race.
 - b. Population for each race will be obtained from the race populations data file and will be mapped to the respective race.
 - c. The number of killings per capita by race will be computed and mapped to the respective race.
 - d. A histogram for race vs. killings per capita will be created and saved as ‘race_killings.png’.
 - e. The race with the highest number of police killings per capita will be printed.
4. Fourth Question:
 - a. Required data files – ‘the-counted-2015.csv’ and ‘the-counted-2015.csv’.
 - b. Two Lists of total number of killings for each month will be created for each year – 2015 and 2016.
 - c. A month vs. number of police killings line graph with two trend lines (different colors) for each year will be created and saved as ‘trend_compare.png’

Work Plan