# Police Killings in the United States

2015-Present

By Christina Hartnett and Steven Geiser


## References

- https://www.washingtonpost.com/graphics/investigations/police-shootings-database/
- https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html
- https://data.world/devyl/u-s-state-populations
- https://simplemaps.com/data/us-cities
- https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/
- https://nces.ed.gov/pubs2010/2010015/tables/table_1a.asp
- https://www.businessinsider.com/united-states-regions-new-england-midwest-south-2018-4#the-east-north-central-division-of-the-midwest-begins-at-ohio-and-stretches-west--5


## Introduction

Since January 1st, 2015, Washington Post has maintained and updated comprehensive documentation regarding individual police shootings throughout the United States. Such an issue rose to prominence in the wake of the birth of the Black Lives Matter movement, bringing public-police relations into the spotlight. Each fatal police encounter brings with it considerable backlash from the public, especially in cases where the victim's threat level makes lethal force unwarranted. Given the nature of police intervention, we acknowledge that it may be difficult to accurately assess whether the amount of force used against possible perpetrators is justified. However, these data may allow us to discover more in-depth relationships between victims and their likelihood of experiencing a fatal outcome.

## Methods

### DATA AGGREGATION

- The data are expressed in terms of victim-specific data, such as the age and race of each victim.
- We chose to use locations as our data points instead of individuals where appropriate, as aggregation enables users to interpret larger amounts of data at once.

### LOCATION-BASED STATISTICS

- The population of each county is important to take into account due to the expectation that more populous counties will report more fatalities.
- Adjusting parameters based on demographics greatly reduces the bias effect of population and offers more accurate inter-county comparisons.

## Data Cleaning

The data was collected from many different sources including The Washington Post, The United States Census and The Department of Agriculture. The original police shooting dataset did not include an important attribute: "Fips". This value is the unique county code identifier across the United States. We merged our police shooting dataset with a dataset that included Fips and Cities. We were able to match the cities in our police shooting dataset with the cities in the Fips dataset to return the Fips values. Using this Fips information, we were now able to merge new attributes into our dataset including, median income, poverty level, percent high school dropout, population and the percentages of each race and ethnicity. We believed these attributes were important to know in order to get a deeper understanding of fatal police shooting activity across the United States. We also merged our dataset with information about what region different states are in to better analyze assumptions about underlying beliefs of these regions.

After our fused dataset was complete there were still missing values for race and age of some victims. We used mean imputation techniques across the state to fill the ages. To fill missing race values we used the highest occurring race in the state the victim was from. We felt this was more representative because states with a higher percentage of a certain race killed are more likely to have that race be a victim of a police shooting.

## Data Analysis

This dashboard was built to test four hypotheses:

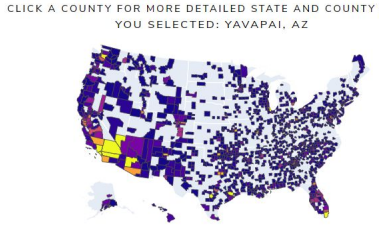
**Hypothesis 1:** Lower median income, lower graduation rate and higher population at poverty is correlated with more deaths at the county level.

**Hypothesis 2:** Minorities are killed by police at a much higher rate than that of nonminority races.

**Hypothesis 3:** Young Black males will be the highest age category killed by police across the United States

**Hypothesis 4**: Southern States where gun culture is more prominent will have a much higher percentage of deaths given their population.

## Choropleth map

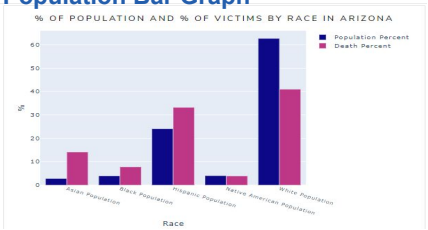



This map is shows the deaths by county across the United States. The user can click any county on this map and it will update all other plots on the dashboard. Warmer colors are higher deaths and cooler colors are places with lower deaths.

## Scatterplot of Regional Data


This plot is a scatterplot of regional information. There is a dropdown menu that has different attributes, including Deaths, Median Income, High School Graduation rate and Unemployment Rate. This creates an easy comparison between different regions throughout the country.

## Population Bar Graph




This bar chart compares the percentage of the population of a particular race in a county, state or the USA to the deaths percentage of each of these races in that area selected using the radio Buttons

# Police Killings in the United States

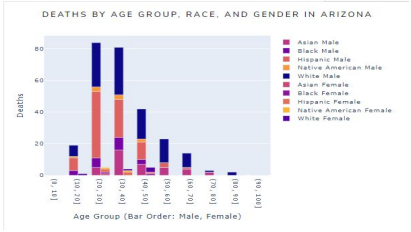
2015-Present

By Christina Hartnett and Steven Geiser


## References

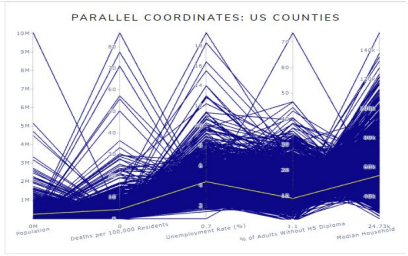
- https://www.washingtonpost.com/graphics/investigations/police-shootings-database/
- https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html
- https://data.world/devl/u-s-state-populations
- https://simplemaps.com/data/us-cities
- https://www.ers.usda.gov/data-products/county-level-data-sets/download-data/
- https://nces.ed.gov/pubs2010/2010015/tables/table_1a.asp
- https://www.businessinsider.com/united-states-regions-new-england-midwest-south-2018-4#the-east-north-central-division-of-the-midwest-begins-at-ohio-and-stretches-west--5


## Age and Race Bar Graph




This bar chart shows age, race, and gender in a stacked and grouped layout. This makes it easy to compare different attributes of victims

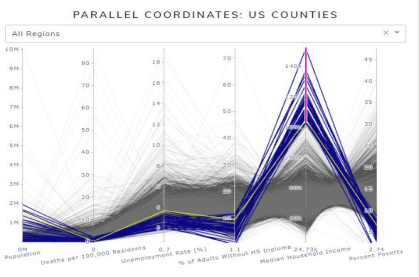
## Parallel Coordinates




This nonstandard plot shows how our different county attributes are correlated with each other. We can easily use brushing to determine important correlations.The parallel coordinates plot shows population, deaths, unemployment percent, percent of adults without a high school diploma, median income, and percent of those in poverty. This also has a dropdown menu where the user can highlight the different regions to see how a specific region differs from the country as a whole.

**Hypothesis 1:** Lower median income, lower graduation rate and higher population at poverty is correlated with more deaths at the county level.

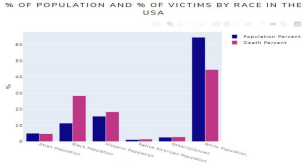
When we only highlight over the deaths axis, we don't necessarily get a clear picture of what's going on with the other variables. Highlighting the counties with the most deaths suggests a possible weak association with median income, but too many counties reported 0 deaths for us to see a clear pattern when we highlight the counties with fewer deaths. So, we went with a different approach: maybe there was an underlying causal model between one of these variables and the level of police shootings. When we highlight the highest incomes we see a very clear association between high income and low deaths, poverty, unemployment, and dropout rates. This should make sense intuitively and there are many possible explanations. What we have is that high incomes equals low deaths, but not necessarily the other way around.




**Brushing of high income counties**

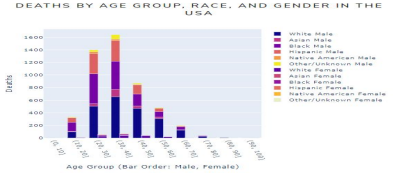
**Hypothesis 2:** Minorities are killed by police at a much higher rate than that of nonminority races.

To test this hypothesis, we created a bar chart of population percents. The blue bars represent the actual proportion of each race in the population and the ones to their right show the proportion among the people killed by police. By comparing the bar size, we can see that African-Americans are indeed killed at a disproportionately high rate, whereas Hispanic and Native American groups show this too but to a lesser extent. This is the whole picture, though, so we're bound to get different results based on the state or county. In general, the pattern for minorities is the rule rather than the exception.




**Hypothesis 3:** Young Black males will be the highest age category killed by police across the United States

To test this hypothesis we need to look at our deaths by age group, gender and race. Even though we see that white people have more deaths in most of the age groups, the bar chart we just talked about really puts these numbers in perspective. Even with a fraction of the total white population, black males in the 20-30 and the 30-40 age groups have almost the same number of deaths as white people do in those groups. Also, even though more people died in the 30-40 group, more African Americans died in the 20-30 group than in that one.




**Hypothesis 4**: Southern States where gun culture is more prominent will have a much higher percentage of deaths given their population.

To test this hypothesis we used our scatterplot of regional information. To our surprise, the West actually has more deaths per capita. It's not 100% clear why this is, but it could have to do with a mix between cultural similarities to the South and having a much more urban makeup.


## Results

Through our visualizations, we saw that several of our hypotheses are more grounded in fact than theory. There appears to be a weak but significant association between deaths and social welfare indicators made visible by the parallel coordinates plot. The bar charts demonstrate that more black males are killed by police relative to population size, with most fatalities occurring within the (20, 40] age bracket. Surprisingly, Western states displayed the highest percentage of deaths given their population.

## Conclusion

Over the past 5 and a half years more than 5,000 people have been killed due to police intervention. This has been a topic of interest for millions of Americans in recent years and has sparked the biggest human rights movement in centuries. The hypotheses tested and the information brought to light through this analysis, shows there is merit in this movement. The disproportionate rate of minority deaths to non-minority deaths speaks volumes on the underlying issues in our society. While we can not be certain these killings were unwarranted, we feel the data speaks for itself.