U.S. Police-caused Fatalities

2020-06-14

A group of people standing in front of a crowd posing for the camera

Description automatically generated*PC: Drew Angerer/Getty Images*

In this post, I will analyze the U.S. Policy Fatalities dataset included in the data folder of this repository. I will use a variety of exploratory and modeling techniques to answer the following questions:

* How many people have been killed by police over time?
* What is the race/age/gender of these people?
* What are the demographics of the offending police officers?
* What are the demographics of the general population in which the incident occurred? This includes the political affiliations of the mayor, city council, and the presidential candidate that the city voted for in 2016.

## Police-caused Fatalities Data

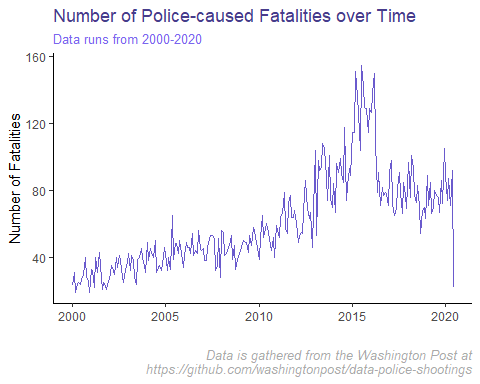
In this analysis, I’ll be working with a few datasets. Different datasets include different ranges of time for police fatalities, with some overlap. On top of that, I have some census, incident, and subject data. Our data runs from **2000-2020**, which makes it an extremely expansive dataset of police-caused fatalities/killings over the *past 20 years*.

## # A tibble: 14,925 x 18  
## id name age gender race date city stateCode manner\_of\_death  
## <int> <chr> <dbl> <chr> <chr> <date> <chr> <chr> <chr>   
## 1 1 John~ 45 Male Black 2000-01-05 Doth~ AL Shot   
## 2 2 Adri~ 31 Male <NA> 2000-01-06 Musc~ CA Shot   
## 3 3 Darr~ 23 Male <NA> 2000-01-06 Char~ NC Shot   
## 4 4 Sonn~ 24 Male <NA> 2000-01-09 More~ CA Shot   
## 5 5 Erin~ 26 Male Black 2000-01-10 Bala~ PA Shot   
## 6 6 Guid~ 17 Male <NA> 2000-01-11 Tampa FL Shot   
## 7 7 Jorg~ NA Male <NA> 2000-01-12 Garl~ TX Shot   
## 8 8 Dona~ 44 Male White 2000-01-13 Phoe~ AZ Shot   
## 9 9 Migu~ 41 Male Hisp~ 2000-01-14 Mode~ CA Shot   
## 10 10 Jose~ 25 Male Hisp~ 2000-01-15 Nort~ CA Shot   
## # ... with 14,915 more rows, and 9 more variables: armed <chr>,  
## # mental\_illness <lgl>, flee <chr>, threat\_level <chr>, body\_camera <lgl>,  
## # state <chr>, popEst2014 <dbl>, percent\_completed\_hs <dbl>,  
## # poverty\_rate <dbl>

After joining and binding together five disparate datasets, we get our final results. We can see that we’re working with **14925** rows of data that shows the individual who died at the hands of police, some demographic information about them and about the city/state where they’re from, as well as information about the killing. It’s worth pointing out that demographic data is pulled from the 2014 Census, which should serve solely as an *estimation* of the actual demographics of a city/state, mainly because some of the killings in this dataset occurred as early as 2000.

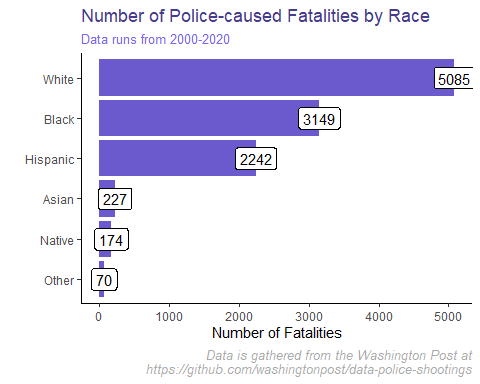
Now that we know what data we are working with, let’s start exploring it further.

### Police-caused Fatalities by Date

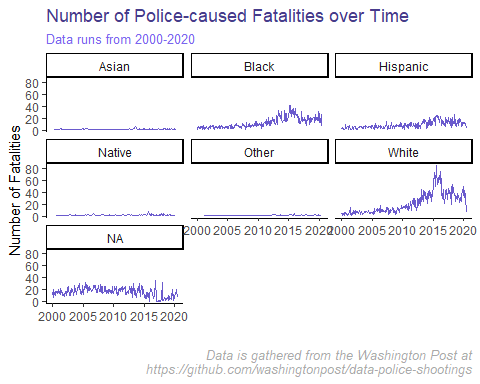


We can see that the number of fatalities caused by police has grown significantly over time, with some peaks in 2015. This is alarming and, to me, shows a change in approach by police in the way that they handle some of their most difficult situations.

### Police-caused Fatalities by Race

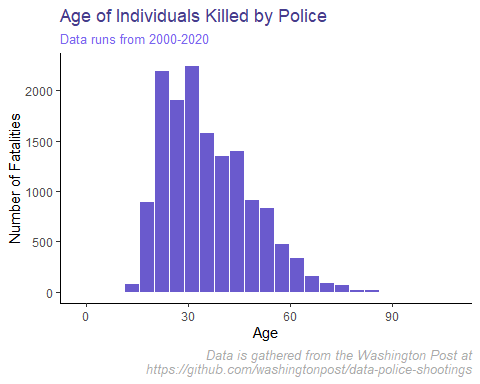


Interestingly enough, more people of a white ethnicity have been killed by police than black people, at least according to the dataset. This seems surprising to me, especially with everything going on in the news lately. One reason that this may be the case is that white people make up a larger proportion of the communities they live in, and this dataset may look different when compared against a race’s proportion of the population. At least for now, I’ll have to check on this to see if this data holds water.



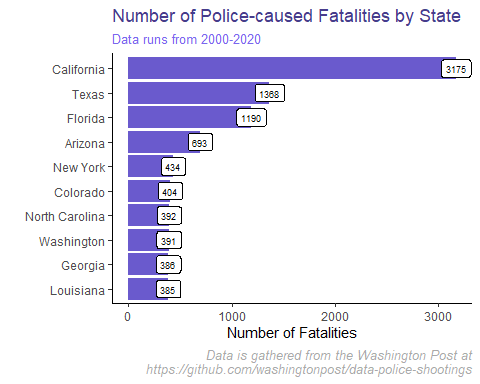
One thing I’ve noticed here is the spike in killings of white individuals in 2015. As a note, I merged two disparate datasets, one that runs from 2000-2015 and the other that runs from 2015-2020. It’s possible that something’s amiss with the latter. The other facet that may be skewing the data is that 31% of the data on race has not been collected (i.e. is null). If most of these individuals were black, that would constitute another 3800 individuals.

### Police-caused Fatalities by Age

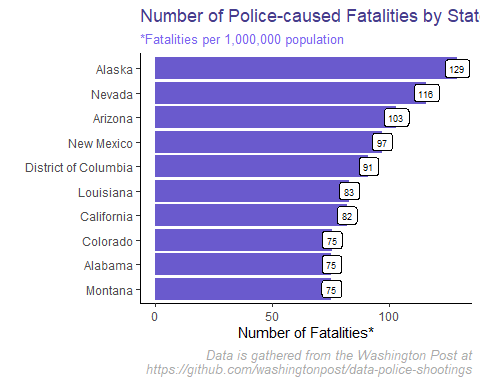


It looks like a pretty normal distribution by age, with possibly a right skew in the data. Most people who are killed are on the younger side, with the median age in our dataset being **33**.

### Police-caused Fatalities by State



This isn’t helpful since the populations are so different. For example, some of our biggest states by population, like California and Texas, are showing up here. I have a hunch that if we normalize by the state population using some of our census data from 2014, we’ll get a better sense of fatalities.

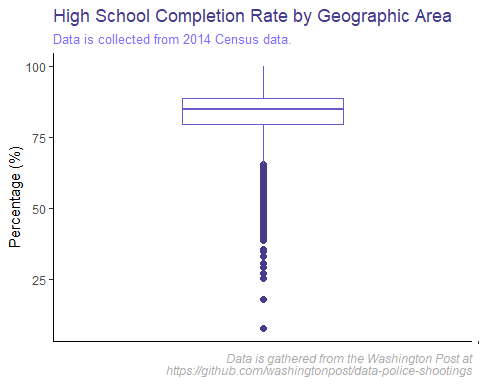


Super interesting! We now have a lot of states in the new visualization, showing that just because a state, like Texas, has a lot of police-caused fatalities, it doesn’t mean that it’s as high as other states proportionately to state population.

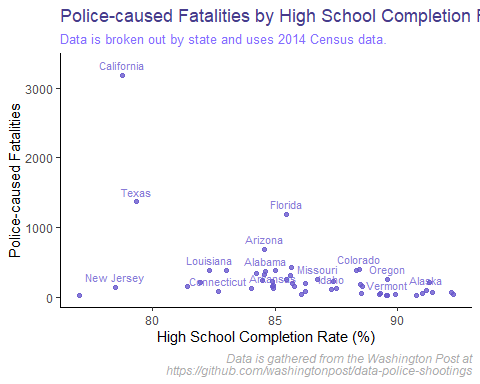
### High School Completion Rate Analysis

I’d like to analyze the number of shootings against the average city population which completed high school. My hypothesis here is that there will be more shootings in cities with low high school completion rates.

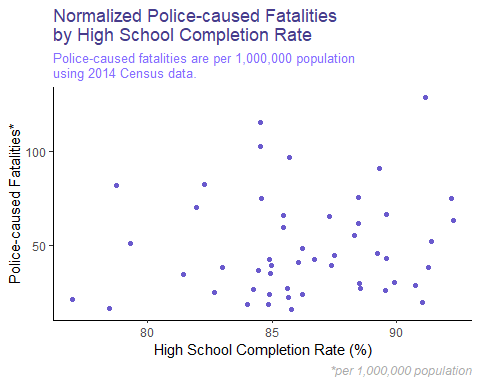
The first thing to do is to fill out our data a bit more. There is a lot of geographic data that wasn’t brought in earlier. To fill out this data, I plan to *impute* it by finding the average rate for each state and using that.



It looks like most of the areas have pretty high high school completion rates, with a median of **84.9** and an average of **83.6614947**.



Again, this doesn’t account for normalizing our data by population. Let’s see how that changes things.



Between both of these graphs, I can’t see any correlation in the data. It’s reassuring to know that my original hypothesis, that police-caused fatalities would drastically increase in areas with lower high school completion rates, was **wrong**. To add some more statistical rigor to our analysis, let’s quickly look at the correlation between high school completion rate and number of police-caused fatalities.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | fatalities | | hs\_completionavg | | popEst2014 |
| **fatalities** | 1 | | -0.4358 | | 0.8822 |
| **hs\_completionavg** | -0.4358 | | 1 | | -0.5087 |
| **popEst2014** | 0.8822 | | -0.5087 | | 1 |
| **fatalities\_normalized** | 0.2479 | | 0.1243 | | -0.0749 |
|  | | fatalities\_normalized | |
| **fatalities** | | 0.2479 | |
| **hs\_completionavg** | | 0.1243 | |
| **popEst2014** | | -0.0749 | |
| **fatalities\_normalized** | | 1 | |

From this we can see that the correlation is highest with population, which makes sense. Once normalizing for population, the correlation drops to -.025, which is highly uncorrelated. Overall, I would say that there is no correlation between fatalities and high school completion rate.

### Poverty Rate Analysis

I’d like to analyze the number of shootings against the average city poverty rate. My hypothesis here is that there will be more shootings in cities with high poverty rates, although my hypothesis earlier was debunked, so we will see!

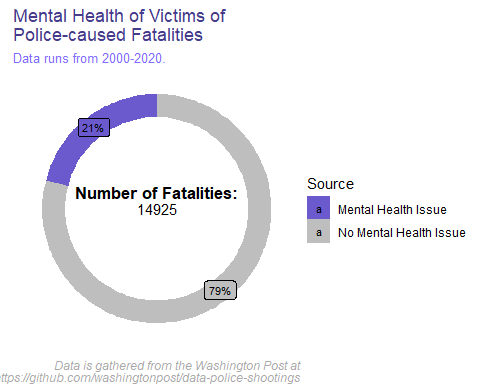
First, let’s see if there’s correlation between poverty rate and number of police-caused killings.

## fatalities avg\_poverty  
## fatalities 1.00000000 0.03892824  
## avg\_poverty 0.03892824 1.00000000

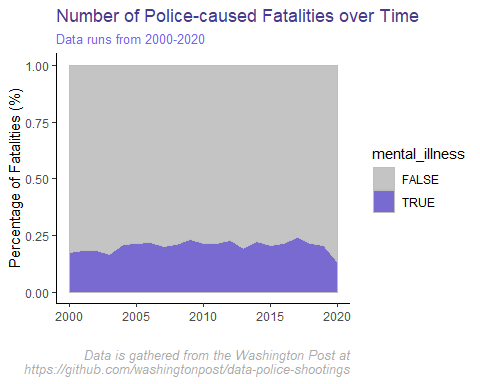
I really don’t see any correlation here, so let’s move on.

### Mental Health Analysis

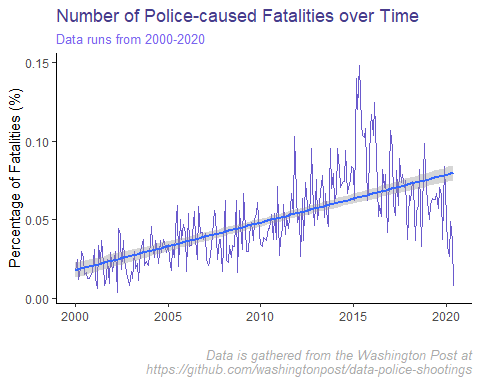
One thing that I often don’t hear covered in the media is the presence of mental health issues amongst victims of police-caused fatalities. Let’s take a look at what the data shows.



About 1 in 5 of the 14925 victims suffered from a mental health issue. Our data runs back to 2000 when, frankly, the conversation around mental health was not nearly as progressive as it is now. Thus, I would expect the proportion of victims who have mental health issues to **decrease** over time. Let’s check this out.



Interestingly enough, it seems that there is a slight upward trend in the percentage of victims of police-caused fatalities who were struggling with mental health at the time. This trend continues until about 2017 when it steadily declines. To me this says that there has been increased focus on training police officers on how to deal with individuals with mental health issues. Let’s look just at the purple section of the graph.



Above we can see just the percentage of fatalities caused by police officers of individuals struggling with mental health as a proportion of all those killed by police officers. It looks like there’s a strong case to say that every year, those dying at the hands of police officers with mental health issues is steadily increasing. By how much?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| **(Intercept)** | -0.05417 | 0.02572 | -2.106 | 0.04872 |
| **date** | 6.967e-06 | 1.741e-06 | 4.003 | 0.0007617 |

Fitting linear model: mental\_ill\_prop ~ date

|  |  |  |  |
| --- | --- | --- | --- |
| Observations | Residual Std. Error |  | Adjusted |
| 21 | 0.01764 | 0.4575 | 0.4289 |

Although it’s relatively small, we can see a strong trend between date and the proportion of victims who had mental health issues.

### Conclusion

### Additional Resources

Interested in learning more on the subject? Go to: Interested in seeing the Washington Post’s GitHub repository? Go to: Interested in seeing my original code? Go to my GitHub repository here: