Interrogating Independently Compiled Police Shooting Data from 2000 to 2016

William M.

January 6, 2018

Interrogating Independently Compiled Police Shooting Data from 2000 to 2016

Introduction

Using a publicly available data set compiling information drawn from the news coverage of police fatalities in the United states from 2000 to 2016, we hope to gain some unique insights into the phenomenon.

Source: https://data.world/awram/us-police-involved-fatalities

The Structure of the Data Set

The compiled data has variables representing the major circumstances of interest in a case of a police-perpetrated fatality:
* Demographic Information for the victim
+Name, Age, Gender, Race
* Fatality location
+Date, City, State
* Details of the Encounter
+The manner of death, whether or not the victim was armed (and if so, with what),
+whether or not the victim was identified to have a mental illness, whether or not the victim was fleeing

The entire data set contains 12,491 events from the years 2000 to 2016.

Armed Encounters with the Police

An obvious question to be asked given the data on hand is what portion of these police encounters resulting in a civilian fatality were the events of an armed (and potentially threatening) civilian suspect.
While every fatal encounter should considered individually, we can reasonably say that a killing of an unarmed civilian is an undo use of force whereas a killing of an armed civilian is more likely to be necessary act of self-defense on the part of the responding police officer.
We aggregate the data set by State and Year finding the fraction of fatalities of civilians where the civilian in question was armed*.
The plot is shown with an "loess" line (local polynomial regression) summarizing the general trend for all ten states over the time period.

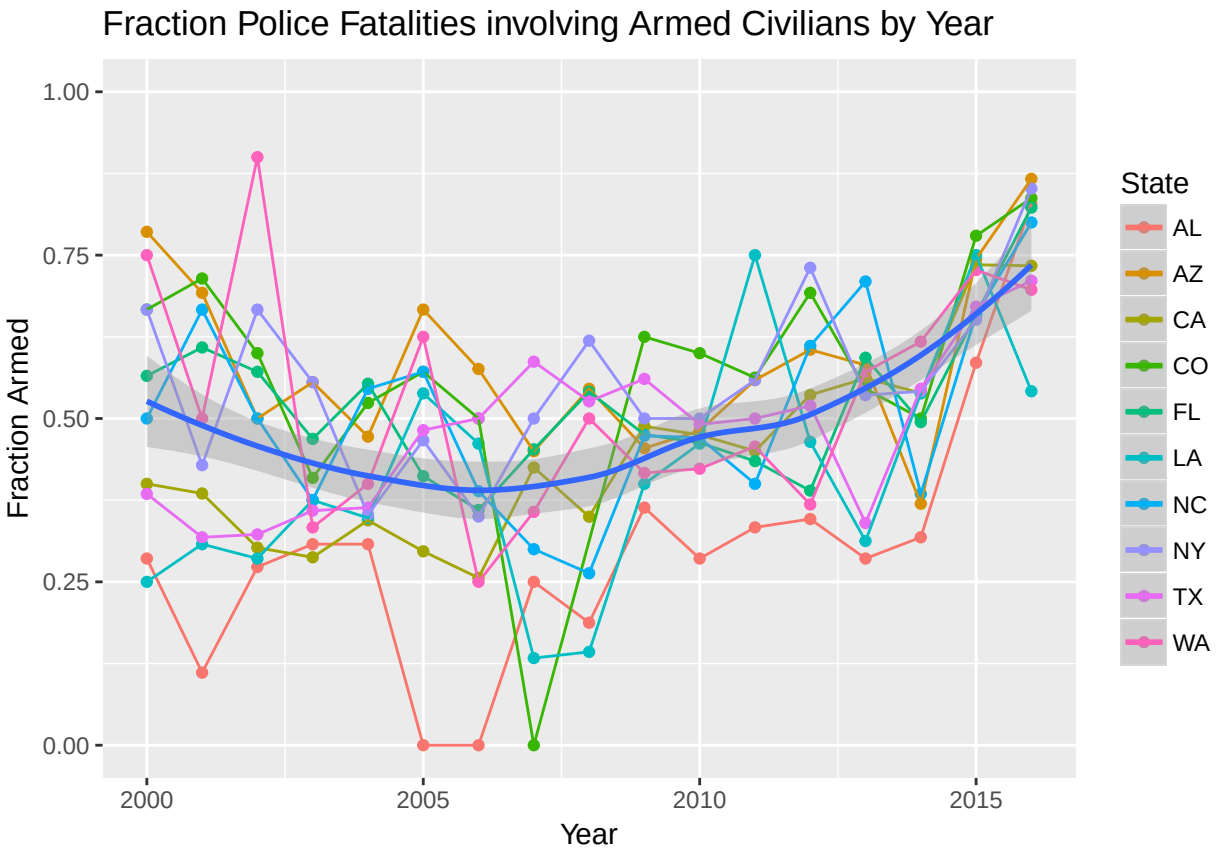
*Note: the list of weapons wielded by civilians in theses encounters is extensive and ranges from "Guns" and "Knives" to "Screwdriver" and "Metal Rake". We will not differentiate between them at this point.

```
Police_Fatalities$Year<-year(as.Date(Police_Fatalities$Date, "%m/%d/%Y"))
Police_Fatalities_Armed<-aggregate(Armed~State+Year, Police_Fatalities, FUN=function(a) sum(!is.na(a))/

#Order the states by total number of fatalities over entire period
```

```
numb_fatal<-aggregate(Armed~State, Police_Fatalities, FUN=length, na.action=NULL)
numb_fatal<-numb_fatal[order(-numb_fatal$Armed),]

#Plot the top ten states with the highest number of fatalities
ggplot(Police_Fatalities_Armed[Police_Fatalities_Armed$State %in% numb_fatal[1:10,1],], aes(x=Year, y=A
```



Conclusions

It should be noted that this analysis counts all cases with NA's for records of armed citizens as unarmed citizens. In some casese, these may just be armed civilians where the circumstances were not correctly reported. Therefore, the true fractions of armed civilian fatalities in each state are **at least** the fractions reported here.
This graphic highlights the trend in police departments across the country taking deliberate action to quell the incidents of extreme and unecessary uses of force after the public outcry against police violence following several high-profile cases in 2014.

Police Responses

Below is a table tabulating the manner of death of each civilian in this data set with the weapon they were reported to have been carrying. This gives us a concise picture of the responses to threatening situations that police officers have made in the broad context of this study.
The table is orderd by total frequency of fatal force applied per weapon and the twenty most frequently involved weapons are shown.

```
table_mannerdeath<-table(Police_Fatalities$Manner_of_death, Police_Fatalities$Armed)
table_mannerdeath<-table_mannerdeath[,order(-colSums(table_mannerdeath))]
kable(table_mannerdeath[,1:10])
```

	Gun	Knife	Unarmed	Vehicle	Toy Weapon	Toy weapon	Machete	Unknown Weapon	Swo
Other	2	4	1	0	0	0	0	0	
Shot	4633	1301	357	107	68	29	13	14	
Shot and Tasered	16	41	25	4	1	0	3	1	
Tasered	52	15	17	0	0	0	0	0	

```
kable(table_mannerdeath[,11:20])
```

	Hammer	Axe	Baseball Bat	Tasered	Hatchet	Metal Pipe	Crossbow	Gun And Knife	Guns
Other	0	0	0	0	0	0	0	0	
Shot	5	6	6	3	5	3	3	3	
Shot and Tasered	2	0	0	3	0	2	0	0	
Tasered	0	0	0	0	0	0	0	0	

Conclusions

Not unexpectedly, the use of lethal force was primarily made against civilians armed with guns or knives. The next most frequent use of force was made against unarmed civilians with a higher fraction of those fatalities involving tazers rather than shootings alone (p=0.117). Cases where the civilian was threatening lives in a vehicle were almost exclusively met with firearms.
Interestingly, the fifth most frequent police-civilian scenario leading to the use of lethal force involved reportedly "toy weapons". These cases are difficult to judge from the standpoint of data analytics alone.

Race and Lethal Force

The relationships between race and applied police force have been rigorously studied. One question that can be investigated using this data set is whether or not a fleeing civilian is more or less likely to be shot given their race. The assumption is that a civilian in the act of fleeing poses less of an immediate risk to the life of the police officer involved. While this is not immediately indicative of broader trends in race-based police violence, the answer may still be illuminating.
We tabulate the total number of police fatalities in this data set by indicated race and whether or not they were known to be fleeing. NA values were automatically omitted.

```
table_flee<-table(Police_Fatalities$Flee, Police_Fatalities$Race)
kable(table_flee)
```

	Asian	Black	Hispanic	Native	Other	White
FALSE	170	2371	1683	123	48	3609
TRUE	3	165	101	7	0	246

```
#Chi-Squared Test of Independence
chisq.test(table_flee[,c(2, 3, 6)])
```

```
##
```

```
##  Pearson's Chi-squared test
##
## data:  table_flee[, c(2, 3, 6)]
## X-squared = 1.4473, df = 2, p-value = 0.485
```

Conclusions

The table immediately indicates that race plays no part in the decision by a police officer to use lethal force against a fleeing civilian.
We can verify this intuition statistically using a Pearson's Chi-Squared Test which tests the given tabulated data against the null hypothesis that the two categories be tabulated are independent of each other. We had to only include the most populous races (White, Latino and Black) because the test is sensitive to disproportionate data caused by small sample sizes. The p-value (p=0.485) indicates a high likelihood that the two categories (race and whether or not the civilian was shot while fleeing the police) are in fact independent.

Technical Note

It should be clarified that, given the data we had available, we cannot interpret these ratios as the fraction of civilians fleeing from a police officer who where shot. The somewhat unintuitive interpretation of "the fraction of civilians who happened to be fleeing a police officer given that they were shot by the officer" is actually the correct one.
We could compare these two conditional statements using Baye's rule:

$$P(was\ fleeing|was\ shot) = \frac{P(was\ fleeing)P(was\ shot|was\ fleeing)}{P(was\ shot)}$$

However, this relation would require estimates for the global fractions of all civilian-police encounters that involved a fleeing civilian (P(was fleeing)) or fatality (P(was shot)). This is beyond the scope or capabilites of this study.