US Distribution of Police Involved Fatalities

GEOG 469 Section AA

UW Fatal Force Research Group

Dan Castro, John Crinkey, and Wendy Liang

June 4, 2018

Executive Summary

Throughout history there has been an increasing phenomenon of police involved fatal force. Unfortunately, many of these cases go unnoticed and much of the data is not being collected nor updated. With a purpose similar to our client, the UW Fatal Force Research Group, we want to bring awareness to the public regarding police involved fatal force. With our research project, and perhaps future research based upon recommendations for further analysis, we want to drive public discussion to address these issues.

Our stakeholders include the general public, public agencies, and other key research priorities. With a collective motivation, we want to determine if there are any unusual patterns and change points in the locations and numbers of people killed by police, mainly from 2013 to 2016, and in regards to the policies that were passed within those years. This includes fatality trends at the regional, state, county and national levels. Once we have a focus on our scale, we will begin to further our analysis based off trends and change points found.

With the many reasons present as to why police involved fatal force is such a big issue, we intend to identify the patterns in fatalities across time and space. Especially now more than ever, with the constant headlines made across all the social media platforms of police brutality, our group is extremely interested in social justices regarding these events. Unfortunately, currently there is a gap in knowledge when it comes to archiving information on cases of fatal force in the US. In addition, many police departments do not publically offer this information. To fill in the gaps of knowledge, questions were created to answer issues related to variables such as time, space, age, gender, and race.

We ultimately want to produce a website, in addition to the ShinyApp created by the UW Fatal Force Research Group (found at https://ben-marwick.shinyapps.io/FFSG-ShinyApp/), that we can submit to WA state legislative committees with empirically-grounded recommendations on how to reduce the number of people killed by police in Washington state. This website, similar to the ShinyApp will include tables from the fatalencounter dataset gathered by the UW FFRG, graphs that depict the dataset, digital interactive maps of different locations and scales of analysis with the data and change point analytics made. In addition, we hope to provide written interpretations and discussions of results alongside each visualization to accurately drive our purpose and objectives for the research project across to our audience. This website can be found at https://wliang25.github.io/geog469/.

Table of Contents

1. Background and Problem Statement

1.1 Background

UW Fatal Force Research Group (FFRG) started at the University of Washington by Professor Martina Morris. Morris' background in sociology and statistics led her to creating this research group to fight injustice in police using fatal force. FFRG's mission is to bring justice and peace to communities most impacted by police fatal force.

This research group started at the beginning of 2017 with two students of Martina Morris. Jainul Vaghasia scraped, cleaned, and merged the two datasets called killedbypolice and US Police Shootings Database to form the dataset Fatal Encounters. Madeline Cummins is the undergraduate who created the ShinyApp to demonstrate visualizations of the data. Ben Marwick, one of the professors, studies Archeology but also has a background in statistics and social science. Because of Marwick's background he was approached by Morris to join the team.

1.2 Problem Statement

With the dataset provided by Jainul Vaghasia, we looked at some important locational questions to help us conclude our hypothesis around de-escalation policies and a correlation to declining per capita values. We analyzed different variables that included age, race, gender, and method of force. With a combination of our findings we related it back to the state of Washington to give some credibility to Washington's Initiative 940. I-940 was introduced to provide training on violence de-escalation.

Our group will gain experience using real-world, crowd-sourced data to scale contemporary issues of public safety and social justice. The data produced can give a client better knowledge about taking care of the fatalities. The client hopes to, with the final deliverable and policy research done, to support and back the passing of state policies that can decrease fatalities across the country.

2. System Requirements

Identifying the system requirements is important for any project as it helps recognize and balance the three constraints of scope, schedule, and budget in relation to the resources that will be used to answer the need to know questions. The main system that we conducted our analysis is from R Studio and the package Shiny App. We also used ArcMap to do some statistical clustering analysis, Leaflet for mapping visualization, and PostgreSQL to manipulate some of our data. GitHub was used for coding version control to communicate between our group and FFRG. Finally, Google Drive was used between our group to organize and maintain our documentation throughout the project.

Identifying the system requirements is important for any project as it helps recognize and balance the three constraints of scope, schedule, and budget in relation to the resources that will be used to answer the need to know questions. The main system that we will be conducting our analysis is from R Studio and the package Shiny App. The source of the data comes from two outside entities which are *killedbypolice* and *U.S. Police Shootings Database* where both sets of data are being combined by an algorithm for our use and analysis. The spatial types of data that we are dealing with are points and polygons to display the shapes of each state/county and the points to display the location of the encounter. This data has been an ongoing collection since 2013 until the most recent fatal force but the datasets have recently been combined in 2018.

The disk space needed for data storage to manipulate and analyze is about 2 GB of space disk because we included some digital interactive visualizations such as maps, charts, and graphs. The disk space calculated also includes pulling three GitHub repositories from GitHub. The size formats of output device of plots of maps, graphs and charts are about 1GB of space disk and the format exported an interactive session by using RStudio. The displays are about 500MB of space disk. We also used about 1GB of space disk for report pages.

3. Benefit-Cost

Benefit and cost analysis is an important aspect that we had to consider in order to implement a GIS into our project. To perform benefit and cost analysis effectively, we considered what is included and what is not included because we determine how to minimize the costs and maximize the benefits. We assumed we would spend high cost for the project. We found out that benefit exceeds the cost at the end of the project. The cost and benefit analysis made our project goal and objective more obvious, so we could have planned well for performing the project with low cost and more benefits.

For Type 1 Benefits the redundancy in data collection was improved as an algorithm to scrape and match data was created. As the algorithm continues to be improved to account for all the data points, manual computations will be reduced and the labor can be pointed elsewhere for improved practices in regards to data collection. With the algorithms improvement, the setup for RStudio to run the ShinyApp requires multiple package installations and library loads. To reduce the amount of time and effort put into rerunning and reloading each package and library, a loop was created at the very beginning of the application to have it automatically identify and load the appropriate packages and libraries for the users. This reduces the first 10-15 minutes to run the ShinyApp each time.

For Type 2 Benefits added functionalities that were too costly to perform in the past includes a detailed change point analysis and graph. In the past, not all the data frames were readily outputted in R so it required more manual analysis. But now that most of the data are ready to be aggregated, change point analysis functions can be easily loaded and performed to determine where the change points are in regards to each state.

Although the cost are close, the benefits exceed the cost towards the end of the project. Possibly looking into other types such as type 4 benefits to account for intangible benefits and see if that adds extra value to our project or if it ends up costing more. From our perspective accounting for intangible benefits would increase to total benefit because of increased production and efficiency.

4. Data Acquisition

Data acquisition is an important part for our project because we had to produce interactive maps and visuals by transforming the data into the information. The outcomes of the project would be different based on the data we use. The data schema specification helped to organize our datasets in a better manner. Determining the data acquisition is an important part for us to get a sense of what path of a project how we look out the project to proceed forward. There are main two parts of data acquisitions which are data collecting and data manipulating.

Our data acquisition is based on information needs analysis. The data acquisition was separated into several steps. The first step was to organize what data we need to obtain based on our research question. The second step was to figure out the source and database where we need to gather the data from. The third step was to figure out how to scrape and merge the data, and figure out how we will convert the data ready to operation to information. We could have decided the scope for what data is included and what data is not included. We were working with UW Fatal Force Research Group and the research group provided most data that we needed to perform data as information. We had found the data on our own if the group did not have the data we needed. The data was stored in ffsg GitHub Repository and the spatial data type of our project was choropleth maps and cartograms maps to present a distribution of patterns in occurring fatalities with police force occurrence throughout the U.S. The UW Fatal Force Research Group has used Shiny App in RStudio to display web page with interactive maps and analysis.

4.1 Database Schema Specification

4.1.1 Data Category Name

Filename: statnet/ffsg GitHub Repository

Source of the data: UW Fatal Force Research Group, Martina Morris

Spatial object type: polygons are represented on choropleth maps and cartogram maps.

The data in RStudio are created and stored as below:

Attribute Field Name	Data Type	Size
Choroplethmapfunc	R FIle	1KB
forPBS	R File	6 KB
Permillgraphfunc	R File	2KB
Shiny App	R File	3KB
Merger	R File	5KB

The good thing in regards to our data acquisition is that there is no significant cost to our project since our data is mostly given from the FFRG and the data was from open sources that are free. Our group kept contacting with the FFRG and Professor Marwick

to make sure that we are on the right track of acquiring the data. We kept thinking of our purpose for the project that is given from the client because our data acquisition is based on our project's purpose and the client's favor.

5. Data Analysis Strategy

Data analysis is an significant step of the project. The data analysis is how you transform the data into the information and analyze what the information means in the project. For the data analysis, the scope of the project should be considered to make sure the project is ongoing well with the goal. Since we got the information after we transformed the data, we had to analyze how the information matches to our topic. Our project's goal is to find patterns in distribution of police killings including civilian fatalities because excessive use of police enforcement has become a controversial problem throughout the U.S. Our objective was to investigate if there are any unusual patterns in the numbers of fatalities by police over the last five years. Our group had to consider how we would digitize the visuals with the given data from the Fatal Force Research Group and what analysis we need to get for to need to know questions that we considered.

The need to know questions include, what are the demographics of victims, when did these events occur, where did these events occur, what happened, why did the event escalate as such, and were there any biases involved with the events. Each of need to know questions include different variables, so we had to plan to analyze the data of different variables in different ways.

For the demographics of victims, there is a 'Descriptive Statistics' page on the Shiny App that will show demographic filters by 'age', 'race', and 'gender'. The demographics of each fatality are displaying on the bar graph in Shiny App. The bar graph helped us to analyze more details in the trends and distribution of a pattern. Roughly 75% of all incidents of people killed by police using a gun as the method of force. In the case of female, there is a higher percentage rate from 2013-2017 over male and transgender of use of a vehicle. Although there is a small amount of data collected from transgender, almost all cases resulted in using a gun as the method of force.

For the events occurred, a choropleth mapping of each state, scaled at victims per capita per state. This choropleth map will have the ability to filter by state and by years the user wants to see. There will also be a cartogram option to display the victims per state capita information as well. The analysis for events occurred was manipulated through putting the data of fatal encounters into the mean operation and the outcome was fatal encounters split by mean for victims per capita per state. The choropleth map indicates that the states surrounding Washington DC, Texas and New York are top three states that seems most fatalities occurred in the U.S.

For the motivational biases involved with the fatalities, the change point analysis is used with the variables of age of death, year, and races. Three most common races involved with cases of fatal force are Black, Latino, and White. As displayed above, in 2013, 2014, and 2017 young Black victims had the highest percentage of incidents resulting in fatal force. In 2015 and 2016, Latinos that were a bit older had a higher percentage of cases over White and Black. This was created through data management functions through data entry done during data collection.

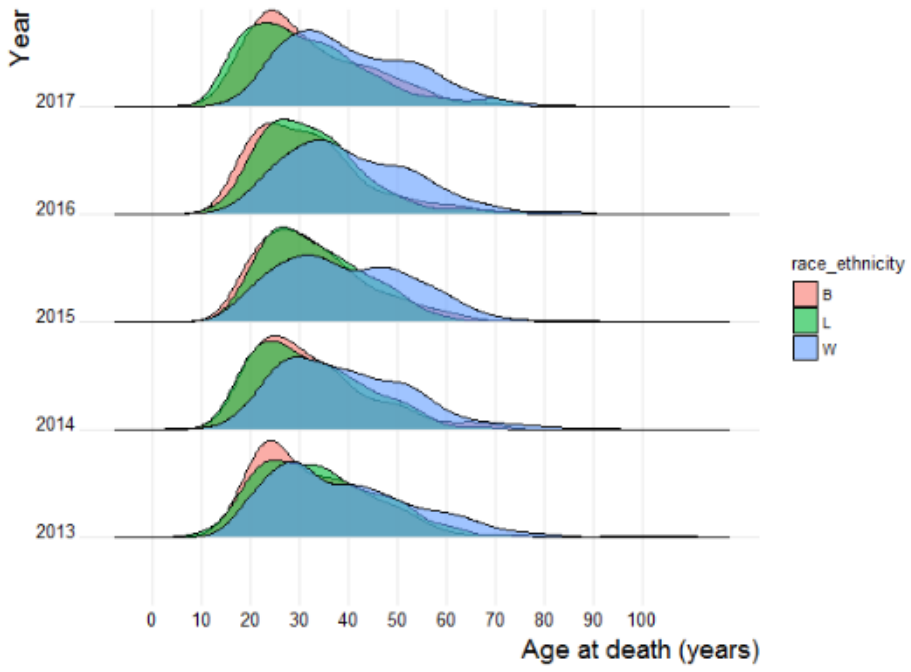
We had focused on the goal of our project while we were analyzing the data and information. We wanted to analyze and explain the data we gathered from the data acquisition relating to our project goal and how these data are important as a part of our project since the data analysis is an important part for the outcome of the project that meets the client's favor. The data and information for the need to know questions helped us to analyze the patterns of fatalities occurred by the police in the United States.

6. Discussion of Results

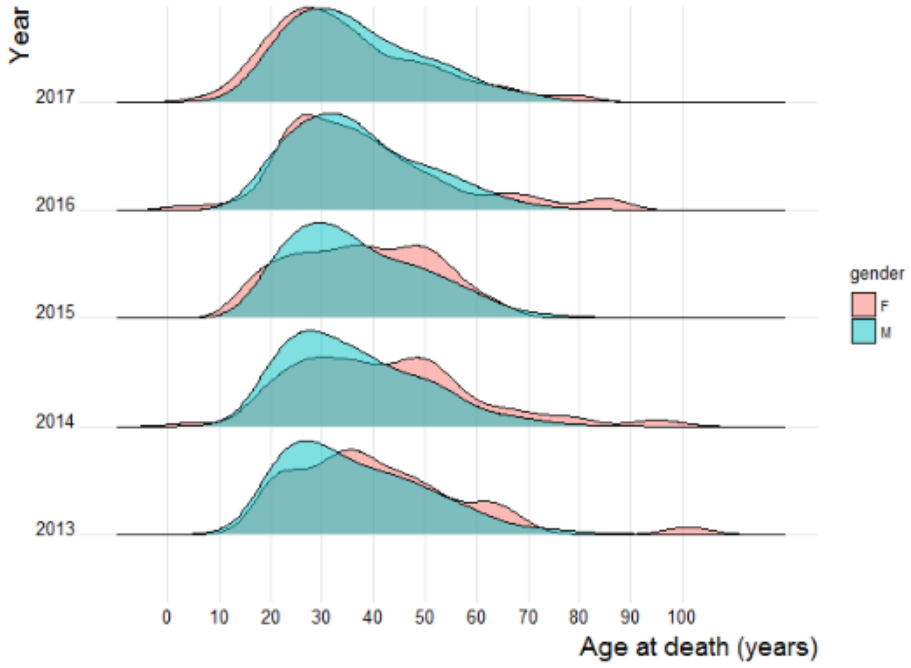
Having a discussion of results is important for any research project to start the conversation towards answering the questions posed in the beginning. Discussing the results will help bring all the research and information gathered together. For our project, we had five main need to know questions: “What are the demographics of the victims?”, “When did these cases happen?”, “Where are the cases of fatal force happening?”, “What are the methods of force between male, female, and transgender used in the events that resulted in fatality?”, and “Which states do we see a decrease in declining fatality?”. Dynamic and static maps were produced to answer these questions, generated by both the UW Fatal Force Research Group and some by our team.

6.1 What are the demographics of the victims?

The graph for this need to know question, generated by the UW FFRG, was split into the demographics of race, age, and gender. While doing a comparison across time, a comparison is first done through the subjects’ race and age. The race shown is split into blacks, latinos, and whites. From the visualization, we can see that whites, compared to latinos and blacks, has fatalities more evenly spread across age, and the distribution increases with time. Blacks and latinos on the other hand, has the most fatalities during the ages of 20-30, peaking at the mid-20’s. Further analysis has to be done at a per fatality level in order to accurately determine the reason why, but the general can infer that there must be some biases involved as the distribution of white victims increasingly spreads out by age, yet blacks and latinos, people of color, seems to keep peaking at the fatal ages of their 20’s.

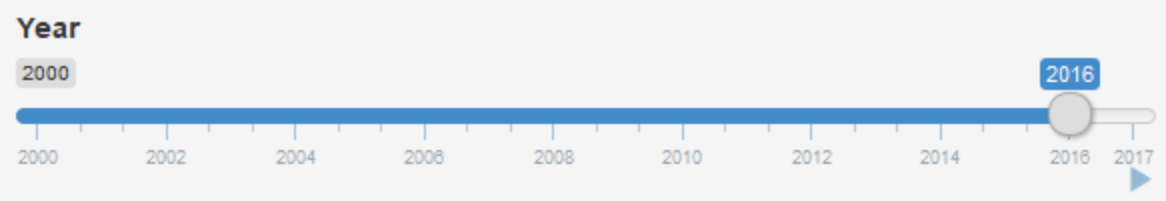


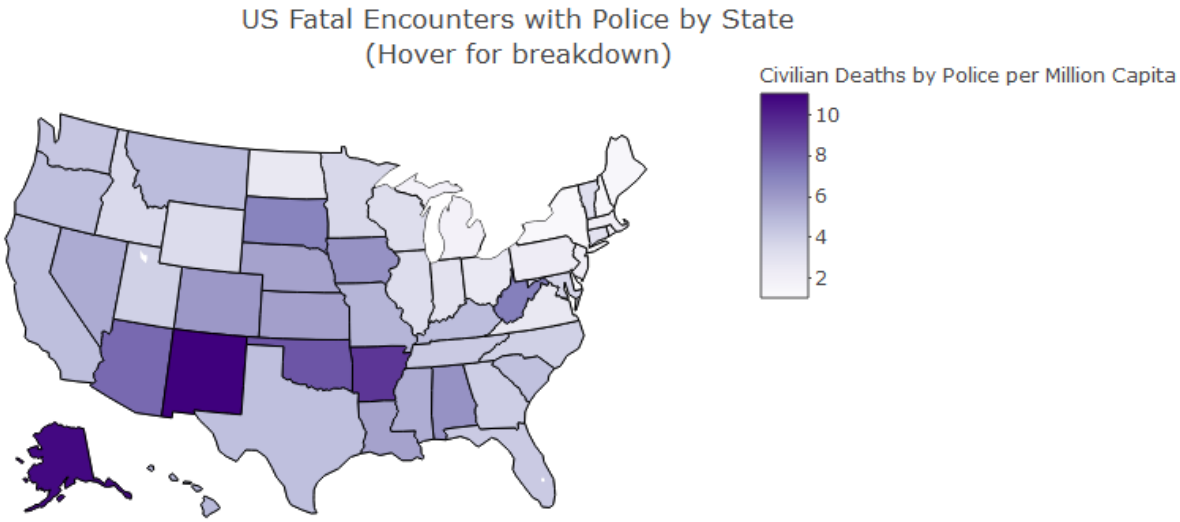
The second graph generated by the UW FFRG is a comparison across time, split between subjects' gender and age. Through the graph, a difference in distribution can be seen between the years of 2013 compared to 2014-2015 and 2016-2017. In recent years, the correlation between female and males seems to be catching up to each other as the blue and pink colors more closely overlap each other. In comparison, the years 2014 and 2015 has more distribution between females across time compared to their counterparts. Fatality for females had a drastic decrease in distribution from 2014 to 2017. In 2014, female fatality developed a peak into their 50's, but that peak is barely visible in 2017. Deeper analysis should be done to more accurately determine the reason for this shift in distribution. Overall, the peak age of fatality for males seems to be increasing, from mid-20's in 2013 to late-20's in 2017.



6.2 When did these cases happen?

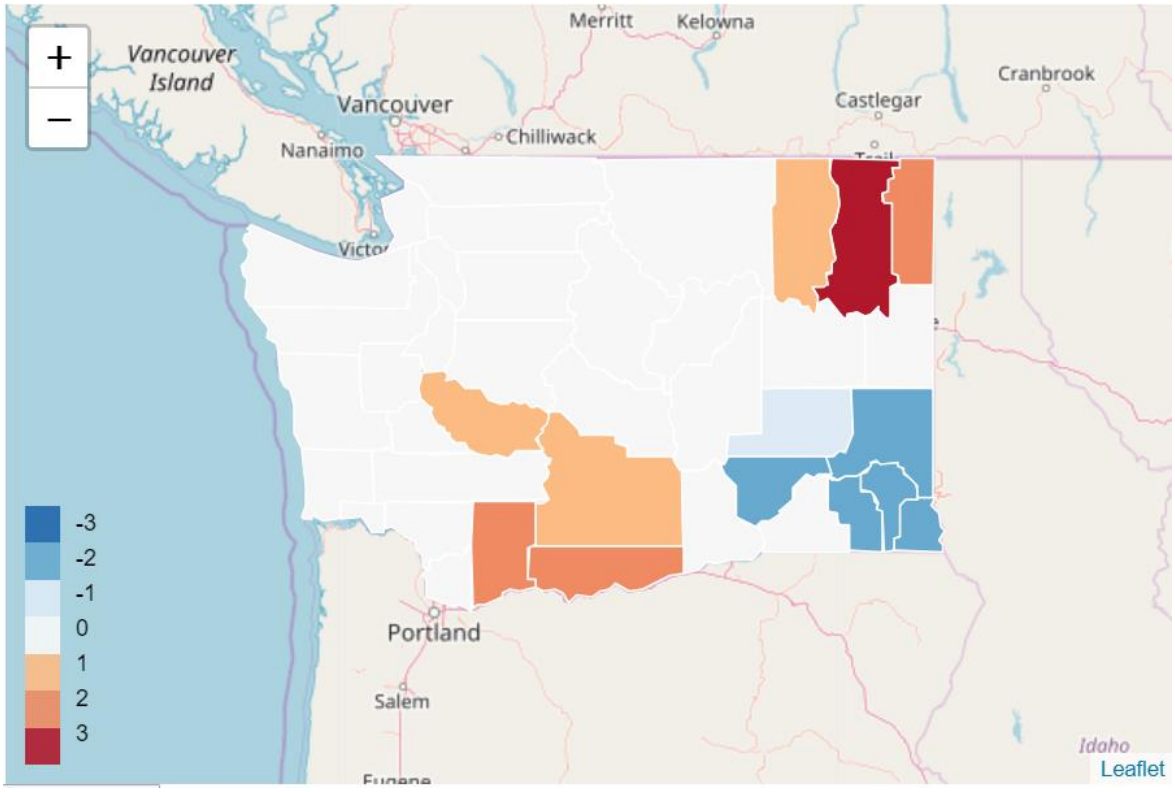
From the ShinyApp created by the UW FFRG, we can use the interactive choropleth map to answer our question of when. The choropleth depicts the fatal encounters by state at a level of civilian deaths by police per million capita across time. Using this map, a selection of year can be done to show the density of fatalities for the US. Using this, for the year 2016, we can see that there is high intensity for the states of New Mexico and Alaska. With this knowledge, a deeper analysis can be done to determine current events for those states in 2016 and tying them into policies passed or not passed that resulted in the concentration of fatalities.

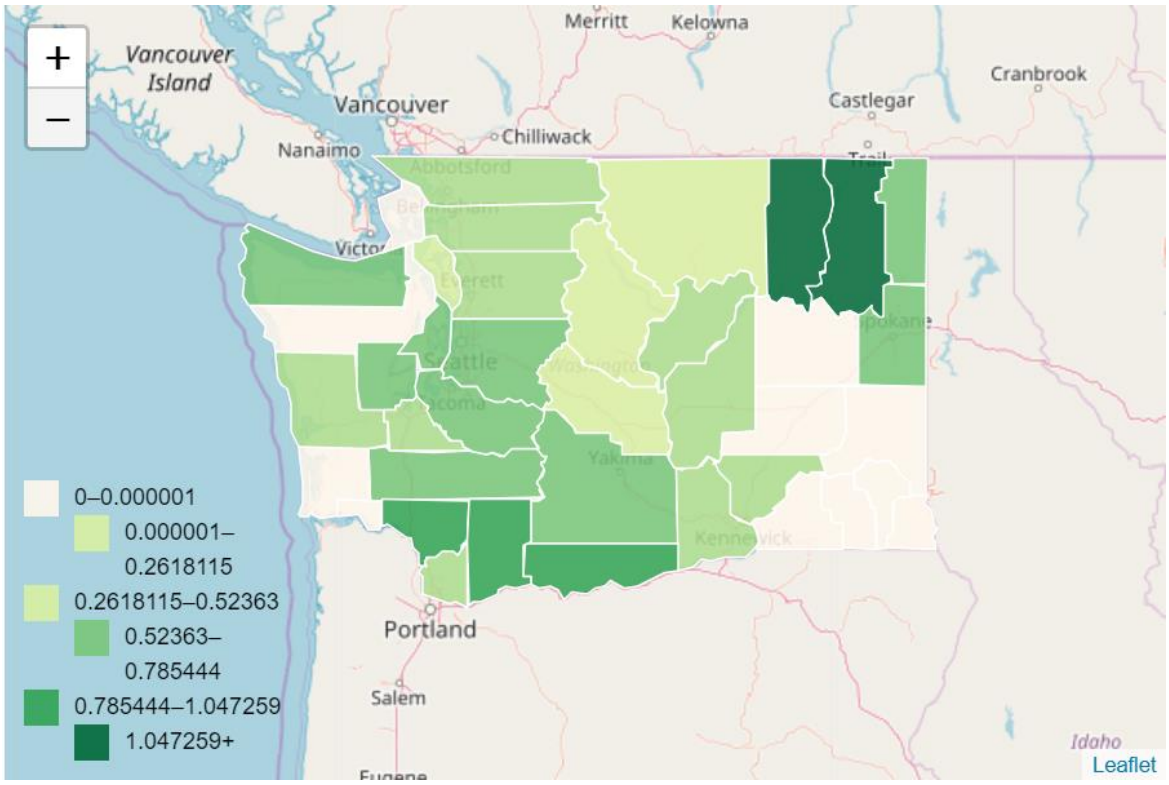




6.3 Where are the cases of fatal force happening?

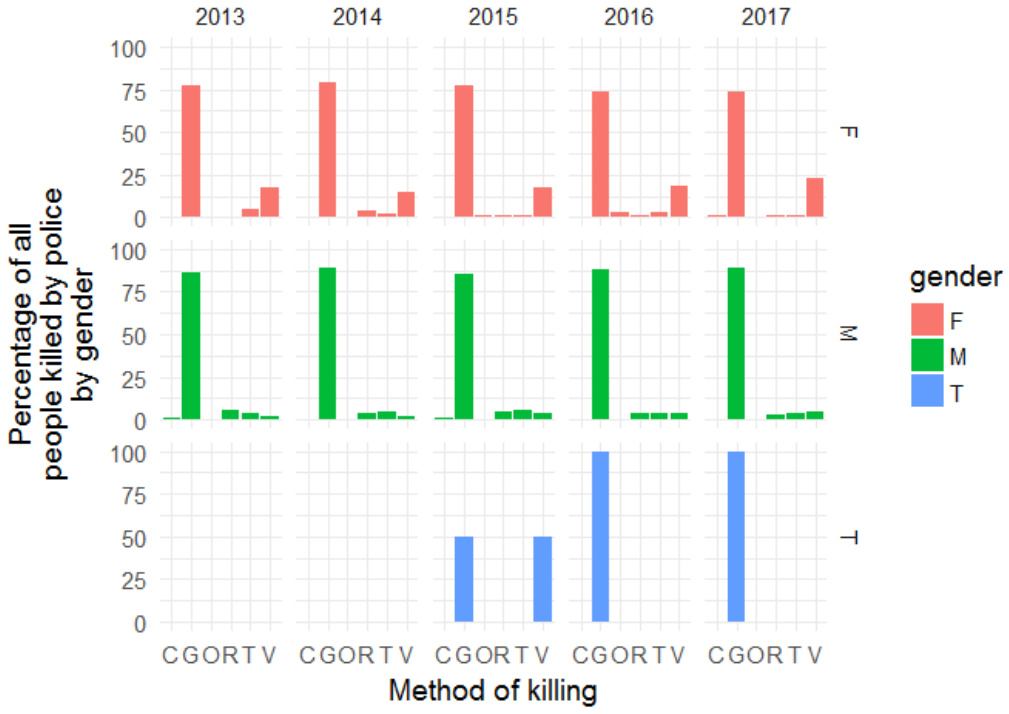
Using ArcMap and Leaflet, our group used the Getis-Ord Gi*, a mapping cluster analysis tool also known as hot spot analysis, to determine the highest values of cases per capita and lowest values of cases per capita. At first glance, much of the northeast counties of Washington state show some high values and the southeast counties show some low values. After running the hot spot analysis, Stevens County showed a 99% confidence for highest values of cases per capita. Although Stevens county has a smaller amount of fatal force cases, the county also has a small population creating a high per capita value. This analysis was only done for the state of Washington and at county level.





6.4 What are the methods of force between male, female, and transgender used in the events that resulted in fatality?

The bar graphs for this need to know question, generated by the UW FFRG, was split into the variables of male, female and transgender in addition to the type of force used. The types of force are split into the categories of chemical, gun, restraint, taser, vehicle, and other. Although the data for the transgender variable was not apparent until 2015, we can see an apparent increase from the years of 2015 to 2017. At first glance, the most common method of force resulting in fatality is gun, even when it comes to transgender. As such, we can conclude that vast changes in policies needs to be done for gun reforms, especially when it comes to police brutality. A discussion can be opened in regards to police usage of guns when it comes to dealing with victims. This includes opening past cases and asking questions such as “Was the gun used only in dire and extreme circumstances?” or “Were guns used as a police routine, regardless of circumstances?”. Using the bar graphs, an inference can also be made when comparing females to males. Such as that females are more likely to be killed with a vehicle compared to their male counterparts - with the percentages increasing across time.

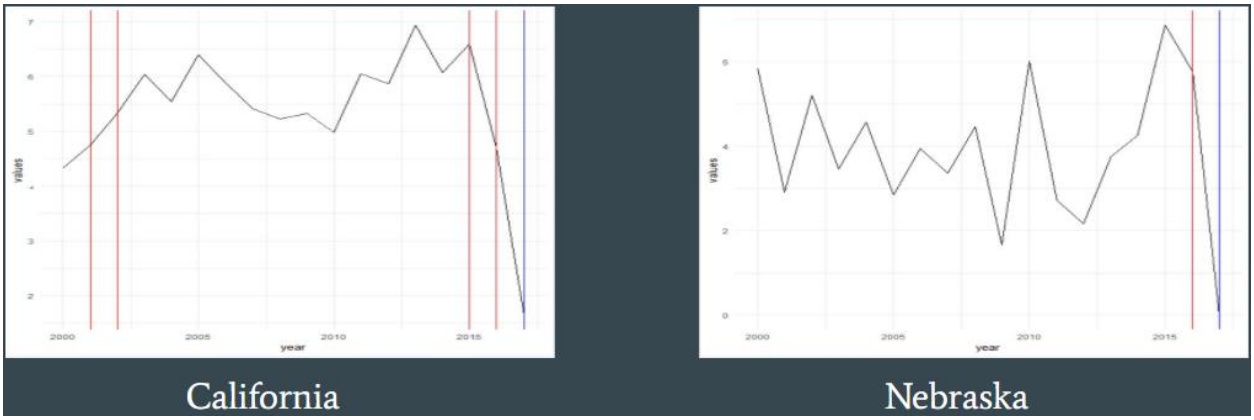


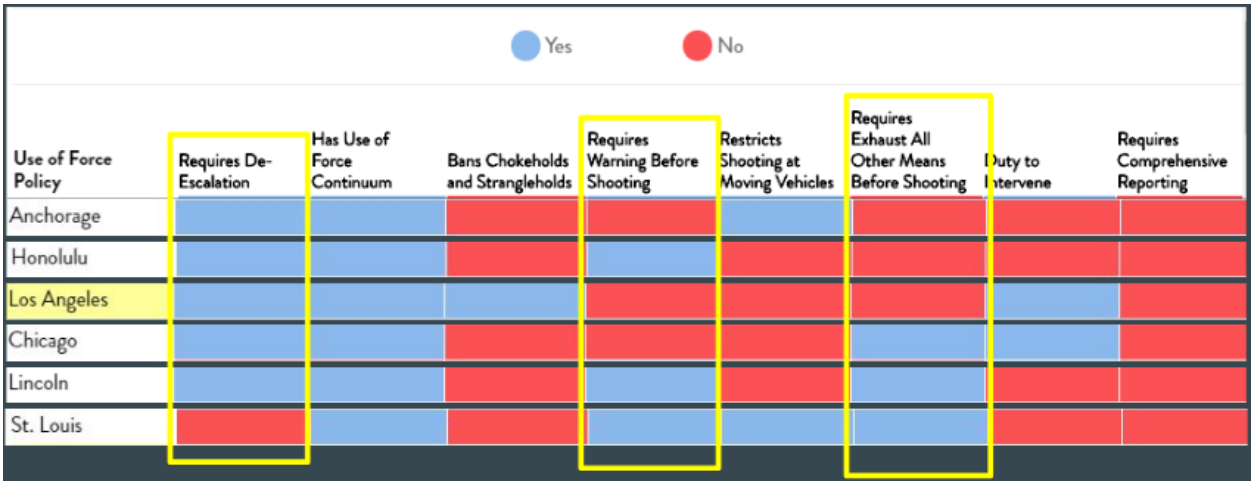
Legend	
C	Chemical
G	Gun
O	Other
R	Restraint
T	Taser
V	Vehicle

6.5 Which states do we see a decrease in fatality?

Professor Marwick introduced the method of change point analysis to determine the years in which significant changes are seen for each state to answer this question. We approached this question by first determining which states had a general decrease in police involved fatalities by applying a simple negative slope formula for each state. We concluded with the states of Hawaii, Illinois, California, Missouri and Nebraska to have a decline in police involved fatality. Using the resulting states, we picked large populated cities of each state to do research on the policies that were enacted in regards to police force with guns. For the cities of Anchorage, Honolulu, Los Angeles, Chicago, Lincoln, and St. Louis, we found three policies that each city had in common. 5 out of the 6 cities

require de-escalation before use of fatal force. 3 out of the 6 requires warning before shooting. And 3 out of the 6 requires police to exhaust all other means before shooting. With the change point analysis and decrease in fatality evidence, these policies can be used as evidence for other states to enact for their cities.





7. Summary, Conclusions, and Recommendations

As we complete our project, this section will work to summarize our research, draw any conclusions regarding our research as well as provide recommendations to anyone looking to continue this project. A summary of our research is helpful when tying together what our intended deliverable and target goals were to our results and finding. This paves the way for the conclusions of our findings and next steps for project. Identifying recommendations for this project will provide suggestions if further investigation were to be done in the future.

The objective for the project was to investigate if there were any unusual patterns in the number of fatalities by police over the last five years and if possible, further back. With this in mind, we were able to determine five important need-to-know questions to guide our research. We sought to identify the relationships of police involved fatalities between time and space prior to correlating our findings to state de-escalation policies. The UW FFRG created an online application that provided most of the graphs and maps. Using the dataset and interactive maps they provided we were able to further investigate and analyze them as appropriate and follow up with policies enacted for those findings.

Tying it back home, to Washington, we have also started to look more into passing policies related to police fatal force. Initiative 940, De-Escalate Washington was passed in March 2018 in conjunction with House Bill 3003. This initiative requires our police force to undergo training on approaching mental illness, de-escalation in regards to violence, and first-aid. Although Washington, with Seattle as one of its densest cities, can be regarded as a very liberal state, we did not come up in the results of the declining states. With the values that we hold ourselves to of inclusion, we hope that this initiative will put our state off of the statistics that adds onto police involved fatalities.

Right now, the analyses done in relation to tying the de-escalation policies back to our research has all been manual. We started out with the fatalencounters dataset, determined all the states that had a decrease in police involved fatalities and then did a change point analysis on those states. This led us to comparing and contrasting the de-escalation policies that those states had enacted that could lead to the decrease in fatalities. For the future, we would like to do some research on potential sites that has an inventory of all policies passed and pending for all the states. Finding such a site can reduce labor costs and increase efficiencies. We can use this site to create an automatic lookup of policies for any state that stakeholders would like to see. Creating this can lead to many other possibilities as well. This includes creating options for uses to be able to compare and contrast various states' policies, like we did for California and Nebraska, themselves.

8. Technical Appendices

8.1 Appendix A

All of our data and charts were collected and created by UW Fatal Force Research Group unless labeled otherwise

- Jainul Vaghasia - scrapped and merged all datasets to create main dataset, *Fatalencounters*
- Madeline Cummins - coded and created ShinyApp for interactive and informational data
- Ben Marwick - coded and created informational charts and graphs which we displayed for our analysis

8.2 Appendix B

Show 10 entries Search:

	race	n
1	African-American/black	1
2	African-American/Black	3652
3	Asian	1
4	Asian/Pacific Islander	270
5	European-American/White	4971
6	Hispanic/Latino	2409
7	Middle Eastern	31
8	Native American/Alaskan	179
9	Race unspecified	6635

Showing 1 to 9 of 9 entries Previous 1 Next

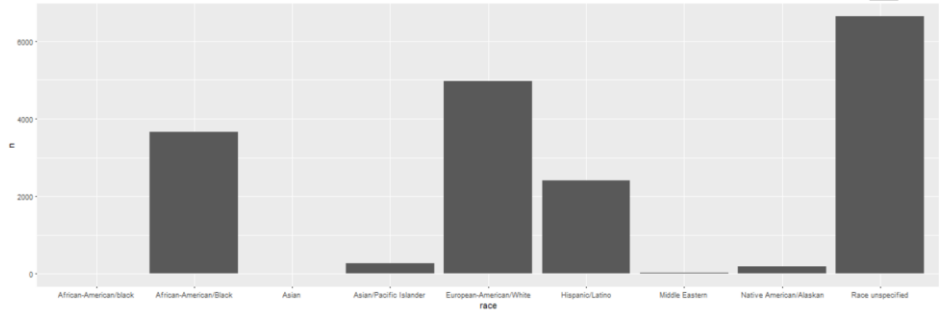


Table 1 - Race/Ethnicity of Victims

Show 10 entries Search:

	sex	n
1	Female	1719
2	Male	16391
3	Transgender	8
4	Unknown	2
5		29

Showing 1 to 5 of 5 entries Previous 1 Next

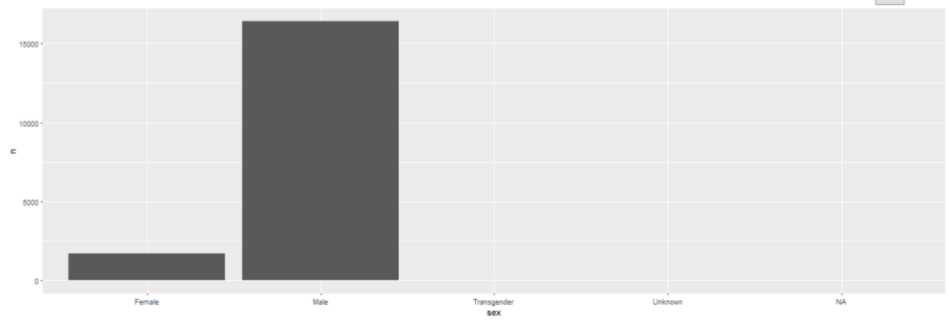


Table 2 - Gender of Victims

Show 10 entries Search:

	age	n
1	1	13
2	2	12
3	3	14
4	4	10
5	5	14
6	6	13
7	7	13
8	8	7
9	9	11
10	10	13

Showing 1 to 10 of 98 entries Previous 1 2 3 4 5 … 10 Next

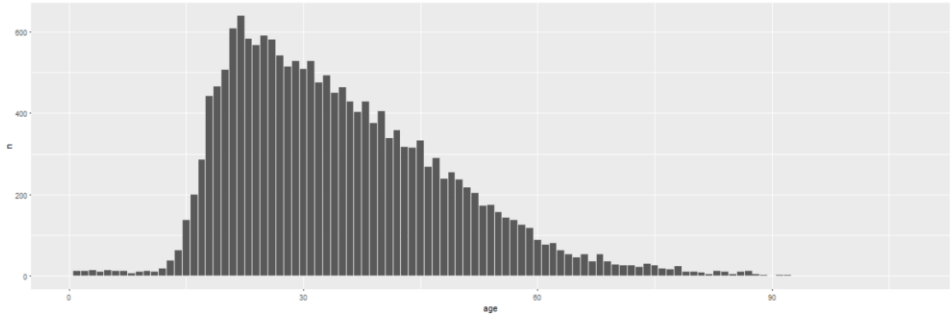


Table 3 - Age of Victims