Data Visualization Analysis of the Washington Post Police Killings Dataset

CSC 465

Harsh Hareshkumar Shukla, Max Rodrigues, Manjula Shetty, Rama Mani Deepika Maram, and Vivek Bhavsar

TABLE OF CONTENTS

INTRODUCTION

Background

In an ongoing project by the Washington Post, data have been compiled on every killing committed by a police officer since January 2015. While the FBI and CDC log police killings as well, The Washington Post documents more than double the amount of shootings at the FBI, indicating this is likely the most complete dataset available.

Variables

The data include a dozen details about each fatality. These variables can be split up into variables that are information about the person killed, variables describing the circumstance of the shooting, and geographic/time information

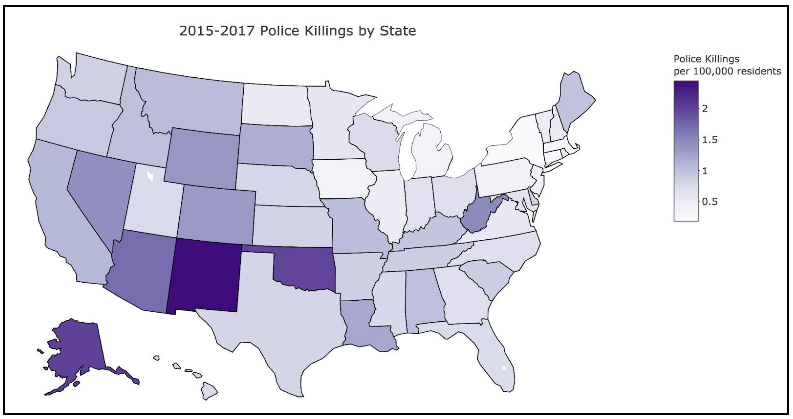
Variables About the Person Killed	Variables About Circumstance of the Killing	Geographical/Time Variables Included
Race	Manner of Death	City/State of incident
Age	Type of weapon was carrying (or unarmed)	Date of incident
Gender	Was the individual fleeing?	
Signs of Mental Illness (True/False)	Was the victim attacking?	
	Was the officer wearing a body camera?	

In addition to the variables included in the main table, other tables included demographic information about almost all of the cities and states in the USA including poverty rates, income levels, racial demographics, and high school graduation rates. In order to calculate per capita values, data from the census bureau were also obtained.

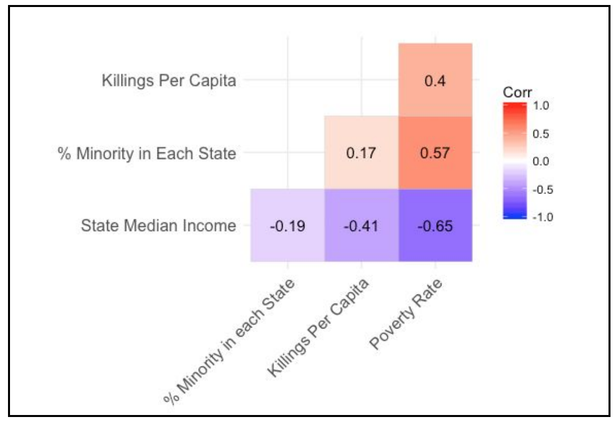
Key Variables for Visualization

We focused on identifying geographic and temporal patterns first and once a pattern was identified, we further inspected variables related to information about the individuals killed as well as patterns regarding the circumstance of the killings. Some of variables were similar (i.e. median income level and poverty rate), and had near identical correlations with police killings per capita. However, in order to further explore the relationship between economic status and police killings, we chose poverty rates because the correlation was positive rather than negative which was better suited for the visualization performed. The key variable among information about the individual killed was race since we found the most drastic differences in police killings to be between different races. While also considering race, we investigated the armed versus unarmed variable among variables about the circumstance of the killing because unarmed individuals being killed by the police was the main motivation for the creation of this data set.

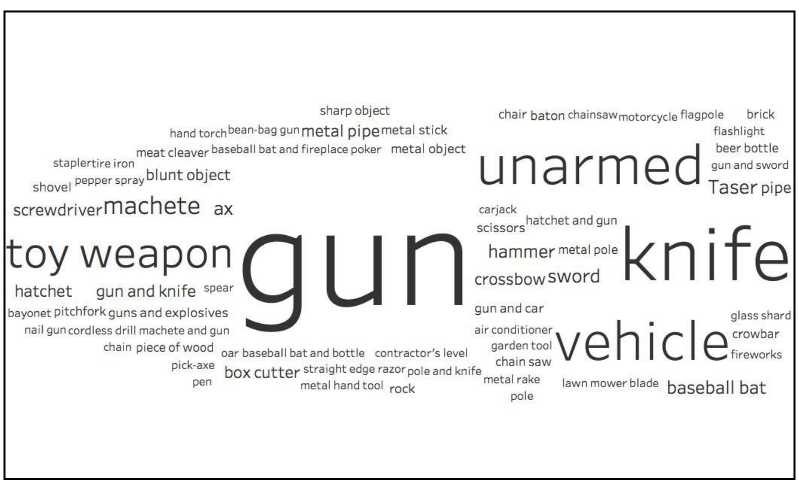
EXPLORATORY VISUALIZATIONS



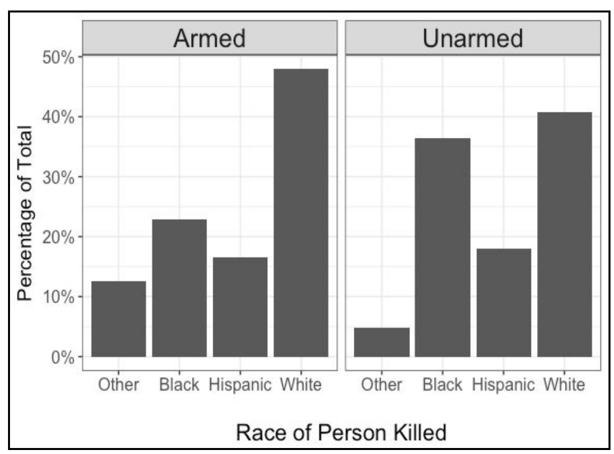
Exploratory Visualization 1:
Police Killings Per Capita Choropleth



Exploratory Visualization 2:
Correlation Matrix of Police Killings Per Capita (per state) and State Level Demographics



Exploratory Visualization 3:
Word Cloud of Objects Held by Each Individual while they were killed.



Exploratory Visualization 4:
Distribution of Race Between Armed (carrying any object) and Unarmed Individuals

Police Killings Per Capita Exploration: Identifying State and Regional Differences

Exploratory Visualization 1

One of the first visualizations we created to identify potential geographical patterns was a choropleth of per capita police killings by state. We used per capita values because otherwise, the choropleth would essentially be showing which states have the highest populations. This exploratory visualization indicates that there are regional differences, specifically that northeastern states seem to have the fewest number of police killings while southern and western states seem to have more with New Mexico experiencing the most police killings per capita.

State Per Capita Killings vs State Level Demographics

Exploratory Visualization 2

Killings per capita were calculated per state and correlations between this variable and other numeric variables for states were calculated with a correlation matrix. Variables concerning wealth (median income per state and poverty rate per state) had the strongest correlations with per capita police killings. The relationship between wealth and per capita police killings was determined suitable for further exploration.

Weapons Carried by Each Person Killed

Exploratory Visualization 3

With the word cloud, we wanted to see what types of weapon were used most among civilians who were killed by police. The word cloud shows that a vast majority of the individuals were armed with some sort of object. The word cloud shows that guns were the most common object carried among individuals who were shot and killed, while knives were the second most common object. Unarmed individuals were the third most common in the dataset.

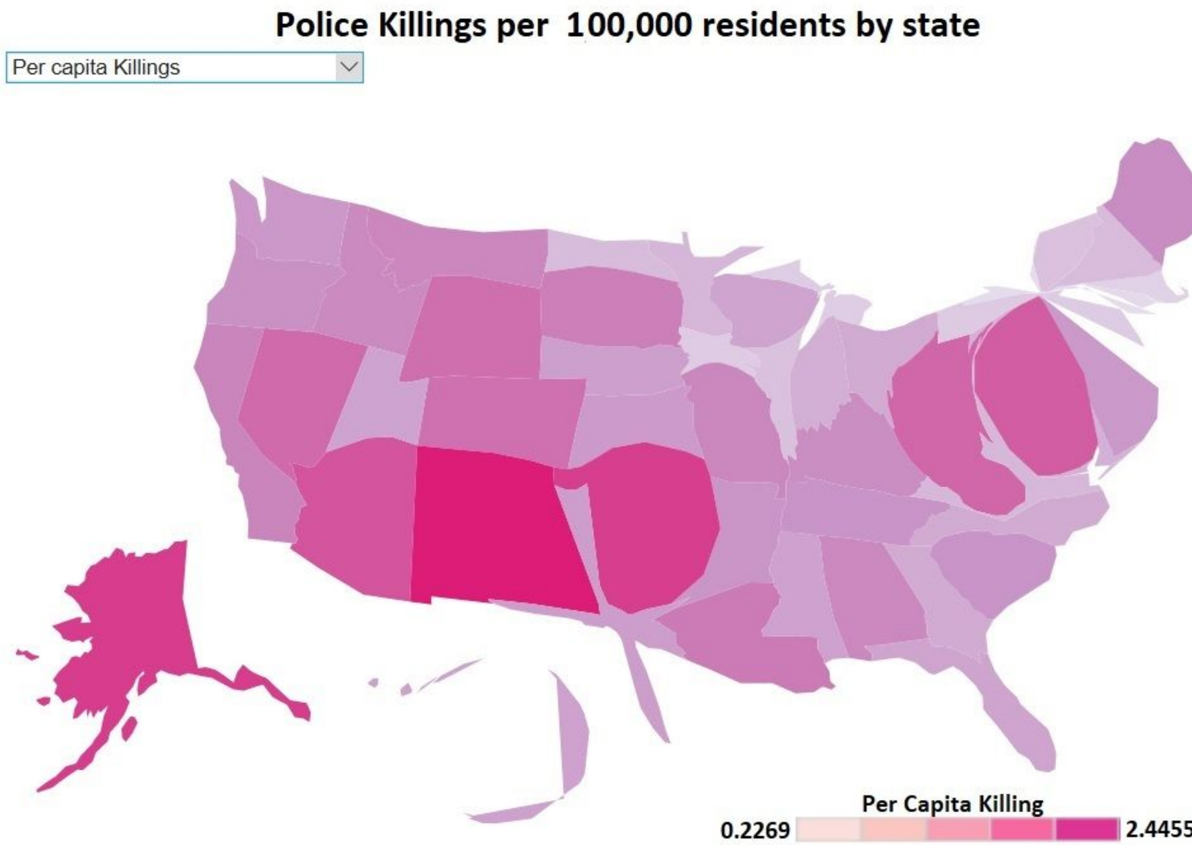
Race and Armed/Unarmed Victims

Exploratory Visualization 4

Given the creation of this dataset was motivated by the killing of Michael Brown in 2014 (an unarmed black man killed by a police officer), we found it imperative to explore both the distribution of race as well as the distribution of unarmed people who were killed. Unsurprisingly, white people made up the majority of the dataset (white people make up about 70% of the total United States population according to the Census Bureau). What we found interesting was that black people make up a significantly higher proportion of unarmed police killings than armed police killings. About 35% of police killings among unarmed victims were black, despite black americans only making up less than 15% of the entire United States population according to the U.S. Census Bureau.

VISUALIZATIONS

Visualization 1: Diffusion Cartogram of Per Capita Killings



Techniques for the visualization 1

After exploratory analysis, we saw that per capita killings vary from state to state as well as region to region. To all the viewer to see the patterns of the choropleth on an even deeper level, we decided to morph the data into a diffusion Cartogram. The intensity of the color purple shows the amount of killing per 100,000 residents in each state. The states are morphed from their original shape and size based on the value of police killings per capita. This allows the user to see small states with high police killings that were difficult to see in the choropleth. It can be clearly seen that New Mexico has high amount of killing. There are also clear outliers in the North East (most states are morphed to be very small except Washington D.C., which is extremely large relative to its land mass). A possible explanation for these regional and state differences is shown in visualization 3.

Design Criteria

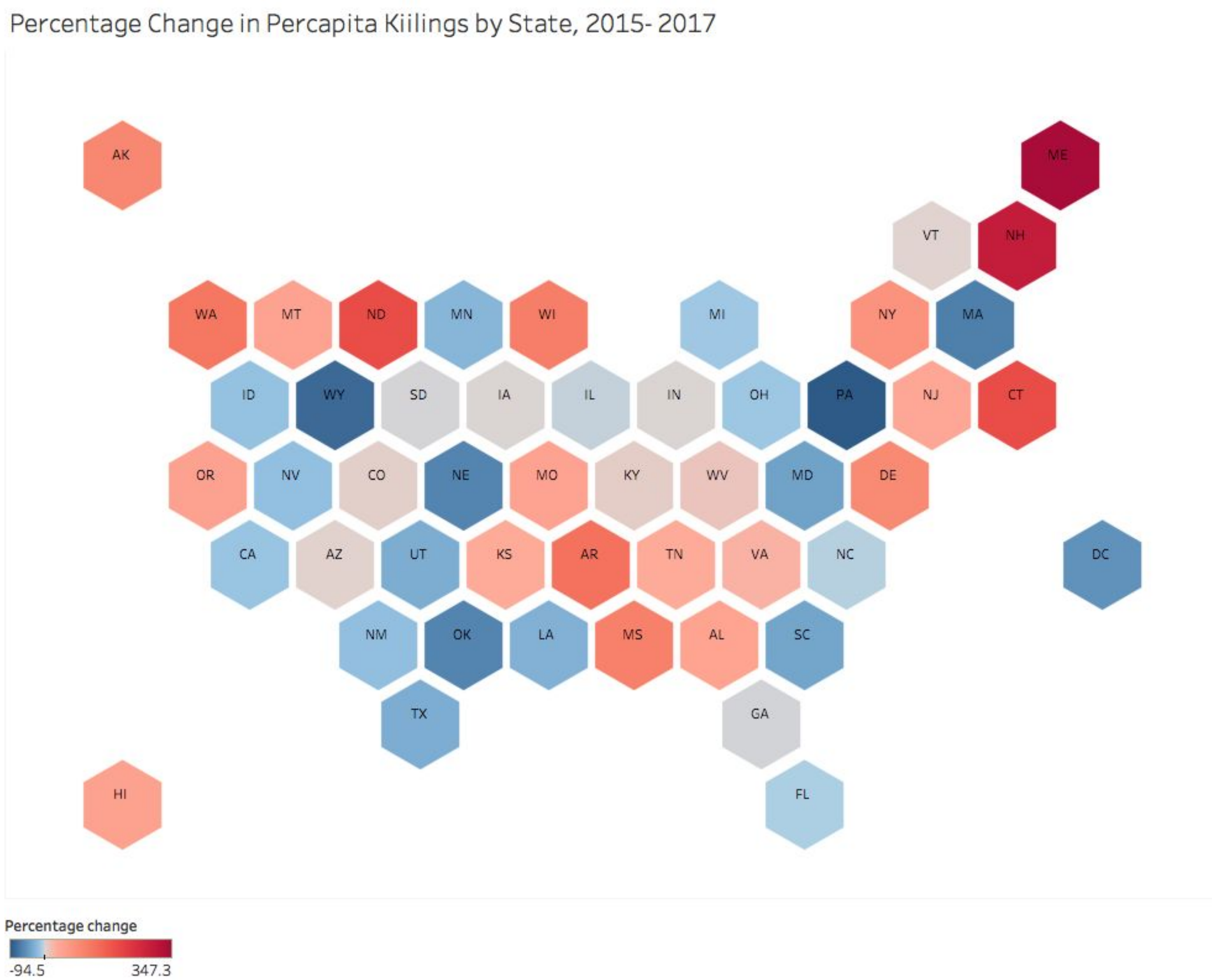
We used a color blind palette to show the increase per capita as well since it's diffusion cartograms states are morphed accordance with the increase and decrease in the per capita killing.

Interactivity

For the interactive version of visualization we have provided the following link where the user can browse through the variables and when the size of the states morphes accordance with the data user can simply visualize the different killings and their correlation with other variables. User can also hover above states and see their values based upon various selected variables from the drop down menu on top.

http://rpubs.com/harsh3838/394408

Visualization 2: Tile Plot showing trend in shooting.



Techniques for the visualization 2

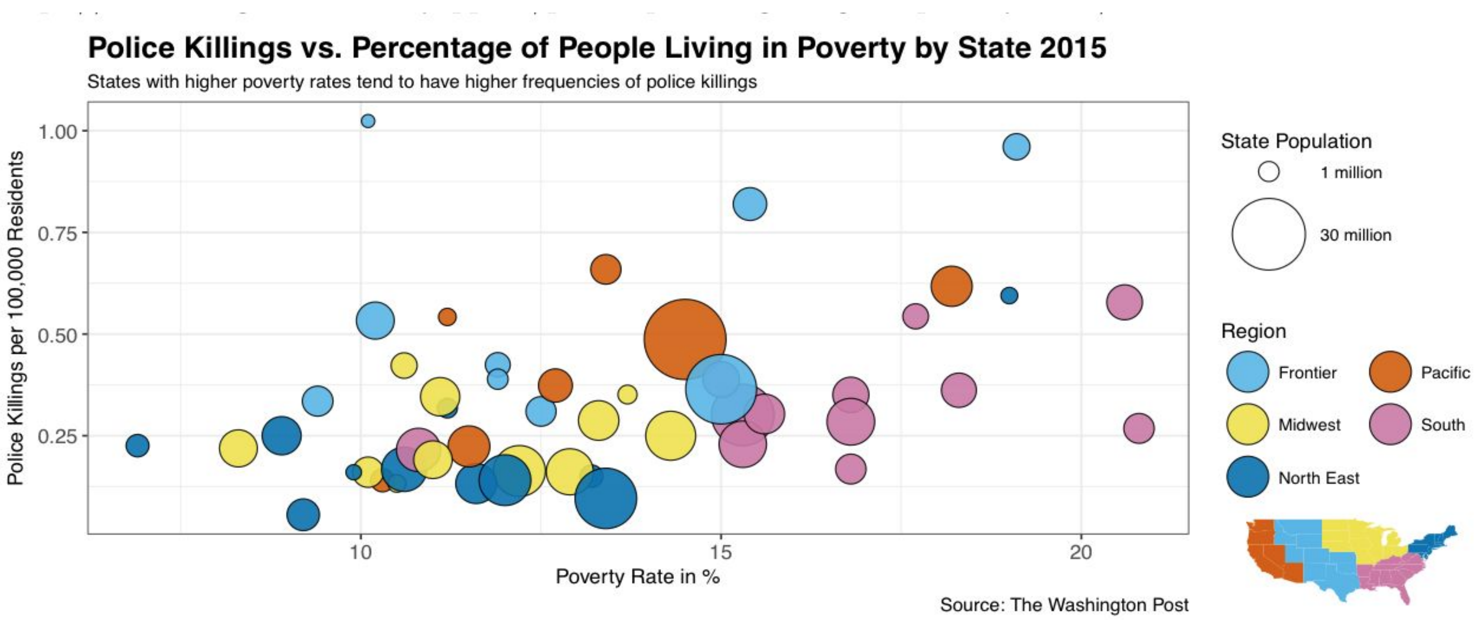
The hex- tile plot above shows, by state percentage change in per capita killings for 2015 to 2017.From exploratory analysis it appears that number of people killed from police shooting appears to have dropped since 2015 and there seems to be a strong correlation between poverty rate and per capita killings. So, we tried to future explore trends in killing across state and how they are changing over years.

Design Criteria

Each tile represents a state and the color is populated by percentage of change in per capita killings. The darker red shows positive increase in rate of percentage of killings for year 2017. And dark blue shows negative percentage of change in killings.

The graph shows an interesting angle for 2017, it shows high percentage change in killings for States that have low poverty rate and low minority and black population. This is unlike the visualization 1, where we see high number of people are shot at in State that have high poverty rate. The hex plot also shows that States that had high per capita killings in 2015 showed decrease in number of shooting for year 2017. This could be related to increase in number police front-end training in States like California, Chicago, Texas.

Visualization 3: Scatterplot of Per Capita Killings and Poverty Rate



Technique for Visualization 3

As was demonstrated by both the choropleth and diffusion cartogram, there are differences in police killings depending on what region of the United States you are looking at (with some outliers within regions). To better understand what is driving these differences, the above scatter plot compares state level police killings per capita against poverty rates for each state with color indicating what region of the United States the state is in. As the scatterplot shows, states with higher poverty rates tend to have higher frequencies of police killings per capita. Also, regions tend to cluster together, indicating homogeneity between states within certain regions. We can also see some states differ wildly from other states within the same region (for example, most Northeast states have low police killings per capita and low poverty rates, except for Washington D.C. which is the 4th poorest in the scatterplot.

The correlation between poverty rate and state per capita killings is statistically significant at alpha = .05 (n = 53, r = .40, non-directional p-value = 0.003) and poverty rates were chosen over median income because a positive correlation worked better as a visualization. A scatterplot was determined as the most appropriate visualization because scatterplots are best at showing the relationship between two numeric variables. Other techniques don't allow the viewer to see relationships as clearly, and a scatterplot allows mapping to both the size and the color of each dot, allowing the viewer to see regional differences.

Design Criteria

We used a colorblind palette to color each circle so circles could be easily differentiated for everyone. A set range of circle size values was also used so population could be more easily interpreted. In the shiny interactive plot, a slider allows the user to move between different years of data.

Interactivity

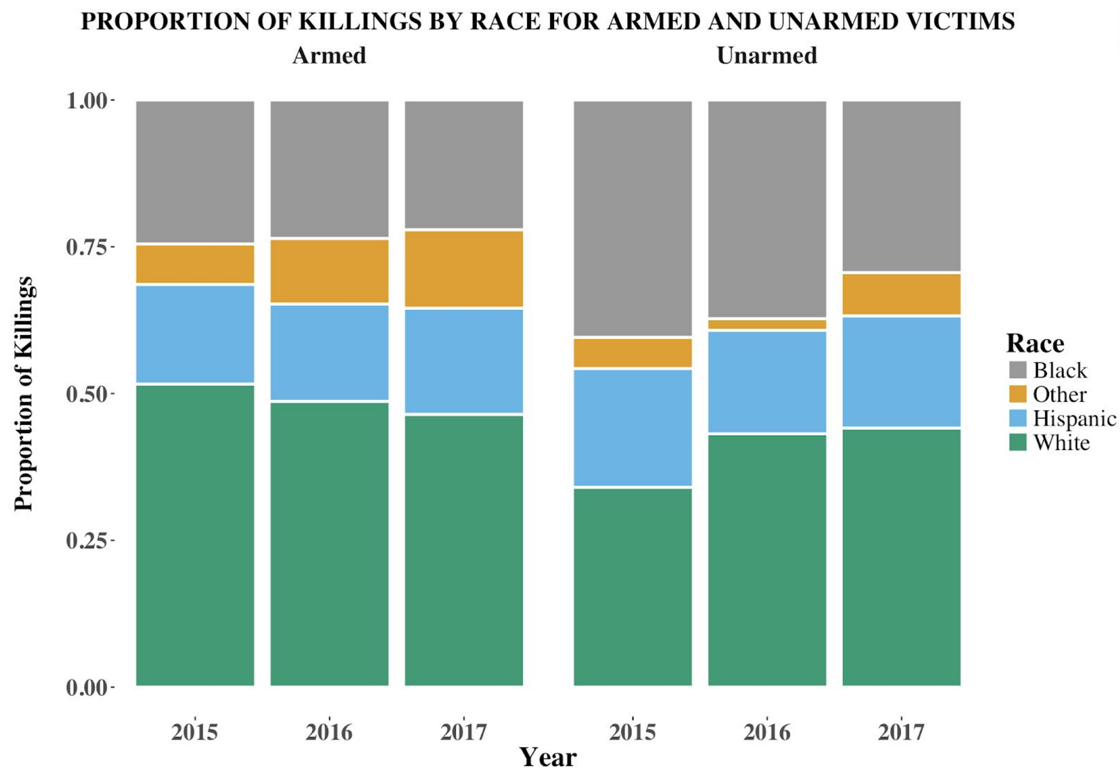
An interactive version of this visualization where the user can select the year is available at the following url. Users can see that regardless of the year, there is a positive association between poverty and police killings per capita.

https://maxrodrigues5591.shinyapps.io/per_cap_killings_region_poverty_rate/

Visualization 4: Proportional Stacked Bar Chart of Police Killings of Each Race Between Armed and Unarmed Victims Over Time

Technique for Visualization 4

The data in the dataset was collected after the shooting in which a police officer, in the line of duty, shot and killed a civilian — the circumstances that most closely parallel the 2014 killing of Michael Brown in Ferguson, Missouri, which began the protest movement culminating in Black Lives Matter and an increased focus on police accountability nationwide. Given this information, we investigated proportions of police killings by armed and unarmed individuals to see if there have been noticeable patterns since the events in Ferguson.



We wanted to compare levels of race variable over the years considering the armed and unarmed victims. So, the choice of Proportional stacked bar graph was a better choice as it shows a clear variation between the both over the years.

So, we clearly from the above graph we see a pattern that the proportion of unarmed black killings have decreased over the years. The killing of Michael Brown might be a reason as once the killing was recorded, the focus might have been turned towards the Black killings. This might be one of the reasons for which we see a decline in the killings of Black people over the years. Also, we observe that the Proportion killing whites was considerably high throughout.

Design Criteria:

The story of Michael Brown has lead us to observe this perspective of the data set and we can see the proportion of killings among black individuals has decreased over time. For this graph, we used R studio and we consolidated date into years, initially we explored using quarters of the year, but we didn't find any interesting pattern so we had to stick with year gave us more clear pattern.

Also since there were many races we choose White, Black, Hispanic and combined Non-Hispanic, Asian and other to a level called other as their proportions were very low compared to the Black, White and Hispanic.

We made sure that the Races Black and White are on the top and bottom levels of the stack which make the comparison easier, since they were the most deviating ones and surely hold a pattern to observe.

DISCUSSION

How Techniques Relate to Each Other (The Data Story)

For the main visualizations, we started with the diffusion cartogram that gives the viewer an understanding of state differences of police killings. This cartogram showed that not only are there regional differences in police killings, but that there are clear outliers within regions such as Washington D.C. in the Northeast. The Northeast generally sees fewer police killings per capita than most other parts of the United States but interestingly, the Northeast has also experienced some of the largest percentage increases in police killings per capita over the last 3 years (according to the hex plot of percentage change in police killings over time).

To better understand what could be influencing these regional differences as well as outliers within regions, a scatterplot between per capita killings per state and state poverty rates showed that poverty has a relationship with the number of police killings. Despite finding significant percentage change differences between years for certain states, the relationship of poverty and police killings is pervasive in the interactive version of the scatterplot.

We were interested in identifying any other patterns that occurred over time within the data, since the data started being compiled around the time when the Black Lives Matter movement was gaining more media attention. Our exploratory analysis found differences between armed and unarmed killing rates between different races, specifically that unarmed black people are killed at a disproportionate rate when compared to the actual proportion of black people who live in the United States. We wanted to see if a temporal pattern occurred within this pattern as well, and as the bar chart over time between unarmed and armed victims by race shows, unarmed black killings have been decreasing as a proportion of the total police killings since 2015. The tree map further explored the relationship between armed and unarmed individuals and showed state differences within regions between armed and unarmed victims.

Possible Statistical Techniques

There are several statistical techniques that could be applied to the data. For one, it would be interesting to know if the pattern of decreasing rates of unarmed black men is statistically significant or if the pattern is just an aberration. Building a classification model and then predicting the killings by state might be a future step towards this dataset, which might let us find interesting predictions on killings by place and race.

Other Directions

Our geographical visualizations looked at larger picture patterns at the state and regional level. With more time, it would have been interesting to take similar methodologies but apply them to the county level within states. While we believe we identified the most interesting pattern with race and armed/unarmed victims, we did find other other relationships that could be explored further. For example, we found that white people have a much higher rate of mental illness within the dataset than black people which may be indicating that a situation has to be more extreme for a white person than a black person in order for them to be shot and killed by the police.

Conclusions

After our initial exploration, we found that there were some patterns that could be explored on a deeper level within the datasets. We found that poverty is one of the strongest predictors of police killings on a state level and that a big reason we see regional differences in the data is that states within these regions tend to have similar rates of wealth. While our scatterplot did not show any dramatic differences per year, we did see a pattern in the stacked bar chart for proportions of killings by race between unarmed and armed individuals. A possible explanation for the pattern is that police officers may have started becoming more cautious in situations where black people were unarmed.

APPENDICES

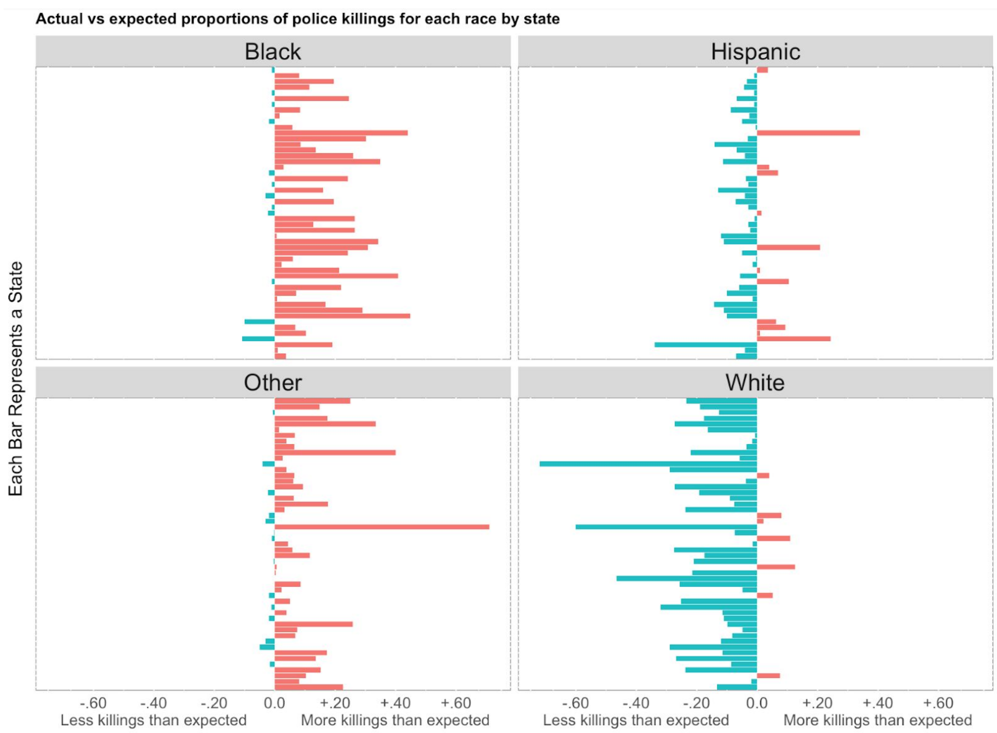
Max Rodrigues Individual Report:

Main Visualization Contributed:

My main contribution to the group was creating the scatterplot of state poverty rates vs state per capita police killings. I selected poverty rate over median state income despite median state income having a negligibly higher correlation with killings per capita because poverty rate had a positive association with killings per capita while income had a negative correlation, and it seemed like a positive correlation worked as a better visualization. I created the entire scatterplot within R including a map legend that shows the geographic locations of regions created. I then made this visualization interactive by allowing the user to select which of the four years of data to view in Shiny.

Other Visualizations Contributed:

I also created several other visualizations (all within R) that were either exploratory or did not make it to the final report. These include the interactive choropleth shown in the exploratory visualizations (interactive link here: http://rpubs.com/mrod1791/389195), a visualization exploring the distribution of race between armed and unarmed individuals (see exploratory figures), the correlation matrix in the exploratory figures, and I also worked with Deepika on a visualization exploring race differences in police killings comparing the observed proportions of each race within our dataset to actual state-level race proportions obtained from to actual state racial demographic proportions.



This last visualization was created by obtaining state level race proportions from the census bureau and then these values were subtracted from the proportion of each race observed in the dataset. After performing this calculated field, a bar chart was created with more positive values indicating the state experienced more killings than expected for that particular race and values that were more negative experienced less killings than expected for that particular race.

Other Contributions:
I helped help create the stacked bar chart and consistently gave feedback to group members on how to improve visualizations. I wrote the introduction to the main report and explained exploratory visualizations as well as the scatterplot and met with the group on three separate 2 hour occasions (as well as numerous other meetings with individual members of the group).

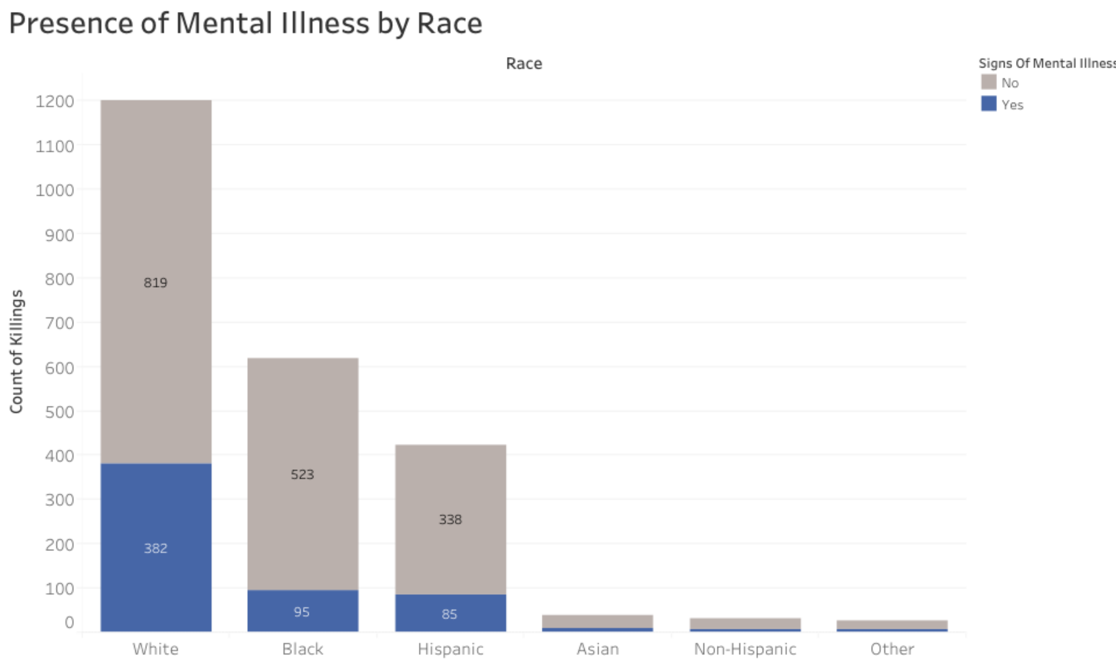
Summary of Takeaways:
My biggest takeaway from this project is a better understanding of identifying the most appropriate visualization and visual metaphors for the message that is being communicated. For example, with scatterplot that compares state level police killings with poverty rates, the first idea that jumped into my head was either a cartogram or choropleth. However, given that the best visualization tool for comparing the relationship between two numeric variables is a scatterplot, that is actually the more appropriate visualization (something I wouldn't have thought of without being exposed to the guidelines taught in this course). Although a cartogram could be created that morphs the state size by either poverty or police killings and the shades states by poverty or police killings, this visualization would be poor at showing the relationship between the two variables.

Rama Mani Deepika Maram Individual Report:

Main Visualization Contribution:
My main contributions to the group were the Stacked bar plot showing the Proportion of killings by race for Armed versus Unarmed which shows a clear pattern of reduced Unarmed killings of victims in USA over the years. I have used both RStudio and Tableau. Also explored d3.js for the mosaic plot.

Other Visualizations Contributed:
I have explored actual versus predicted from exploring the state level proportions from the US census bureau versus the proportions at state level in our dataset. The two-sided bar chart shows the if the state expected more or less killings than expected for a particular race.

I have explored another perspective of this dataset being if the victim had signs of Mental Illness. The interesting observation I found here is in the races Black and White. We see that the victims having no mental illness were killed more than the ones that had no mental illness.



This seem to be a very good observation but was limiting in direction for which I choose to give more focus on Armed and Unarmed killings. Based on my analysis, I am assuming that the races like Black have a higher chance of being shot. Also, when we look at percentage of unarmed victims on both “White” and “Black” races respectively, we see that black has a higher chance of killings than white when unarmed. Therefore, a small injustice of perspective when it comes to other races might be possible here.

Other Contributions:

I also created mosaic plot for the killings but choose a stacked bar over it. Also, I created a bar graph and choropleth for the signs of mental illness by state. I met the team for three two-hour session and individual group members when needed and was responsive to the emails.

Summary of Takeaways:

My takeaways from this project are figuring out the best visualizations based on the data we had. When Exploring the proportion of killings by race based on armed and unarmed victims I have tried both mosaic and proportional stacked bar graph. Selecting one over the other made sense and gave me a better picture of the visualizations. Exploring different variables and coming up with one right one relevant to the story was a bit challenging for me since my initial focus was more on the signs of the mental illness, but later we decided not to include it as wasn’t relevant to our story.

Harsh Hareshkumar Shukla Individual Report:

Main Visualization Contribution:

My contribution to group was creating the diffusion cartogram for multiple variables to see how does it varies from not only one perspective and morphed state sizes were representing to show how does per capita killing is varying. I have also provided an interactive link to make it simple differentiate and hover-over on various states and see how it is affecting.

Other Contribution:

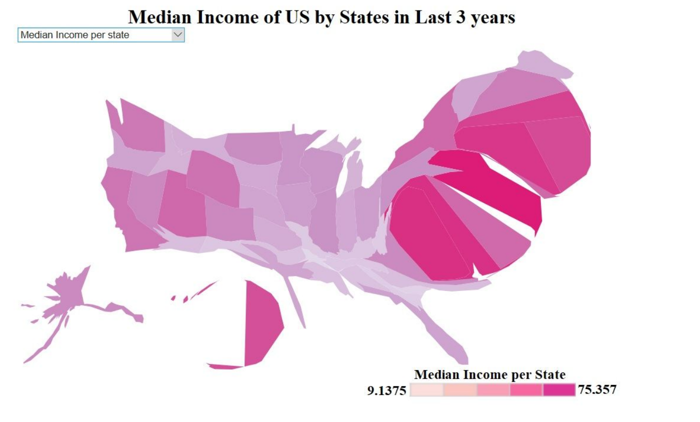
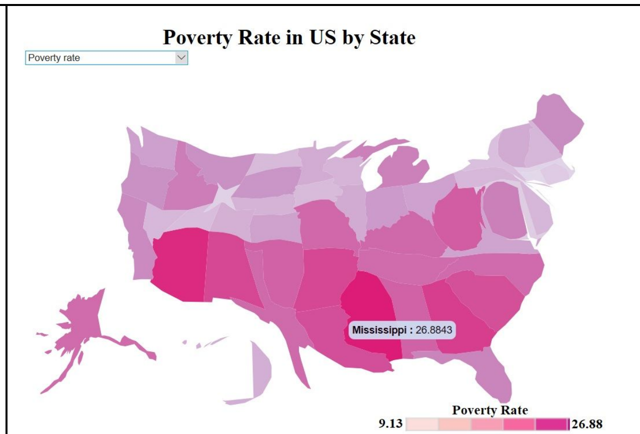
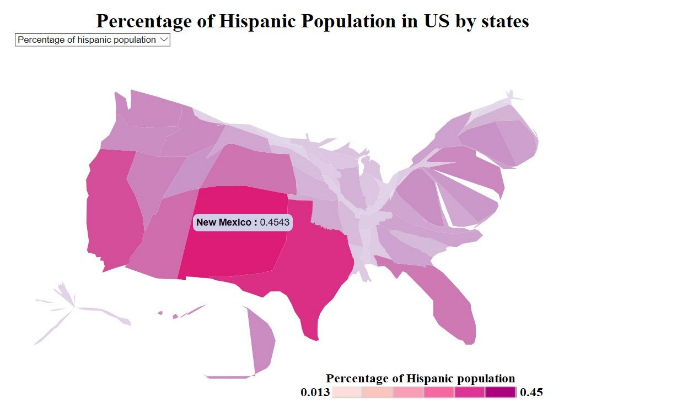
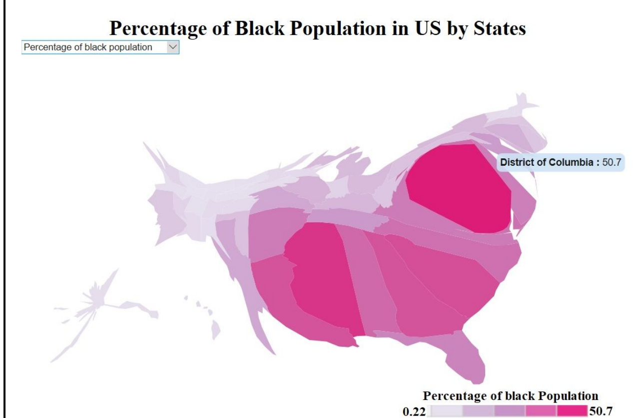
I have played part of team liaison and always helpful to my teammates. I usually try to divide work with everyone's compatibility and their availability. I have also helped in finding the data set and organizing meetings.

Other visualization contributions:

I have tried to visualize different story line where how median income, population estimates of black and hispanic, and poverty rate by states. So as can be observed in the first visualization the median income varies a lot from region to region where northeast part of us has higher median income per state however when it comes to poverty rate it is least in northeast however a lot high in southwest and southeast. And when it comes to black population they are living more in southeast and northeast while hispanic majority can be noticed in southwest since it is near from mexico. This all visualizations determines a small pattern where you can clearly see the poverty rate and proportion of different races is obviously making a biased effect on over all killings.

I have also tried to visualize does the lack of education in states has anything to do with killings. Since, it might be possible that lack of education might lead to not finding appropriate job and less amount of money consequenting in luring generation to lure into criminal activities to visualize that I used choropleth other than that I have ran various different story lines and possible alternative to current story such as how does Police wearing body cameras can increase or decrease in killings. however, none of them resulted as effective as current one.

Other visualizations:

 Harsh's individual visualization 1: about median Income by the states in last 3 years	 Harsh's individual visualization 2: about Poverty rate by the states in last 3 years
 Harsh's individual visualization 3: about Hispanic Population by the states in US	 Harsh's individual visualization 4: about Black Population by the states in US

Summary of Takeaways

I have learned a lot from this project like using one dataset or relying only on one website is not sufficient and it is better to see multiple perspectives than one. Most importantly, I have learned which visualization is appropriate to convey a certain message to audience even if it is a non-technical audience. A selection of wrong visualization or wrong color pallets may lead to unexpected misleading delivery of a message so it is quite important be sure to use appropriate techniques and proper design. For instance, in my visualizations I have used a color pallet that is sequential since my data was continuous and never had neutral point moreover I have used size and color pallet to determine the same thing which makes easier for audience to differentiate from both perspective.

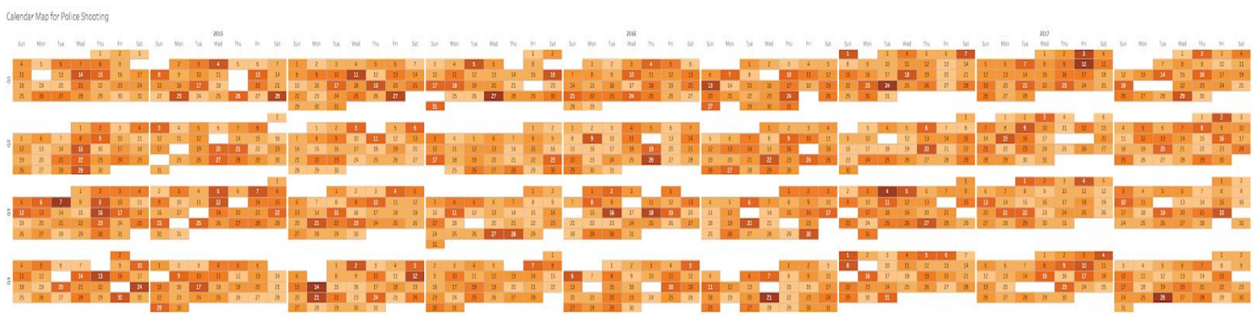
Manjula Shetty Individual Report:

Main Visualization :

My main visualization was *Hex- Tile Plot* which was showing the Change in Police Shooting for three Years. Each tile represents a state and the color is populated by percentage of change in per capita killings. The darker red shows positive increase in rate of percentage of killings for year 2017. And dark blue shows negative change in killings for States.The graph shows an interesting angle for killings, 2017 shows high percentage change in killings for States that have low poverty rate and low minority and black population. And the States that shows high per capita killings in 2015 showed decrease in shooting for year 2017.

Other Visualizations:

Calendar Map Showing Police Shooting for 2015- 2017



Design Criteria: The calendar map shows the shooting for year 2015, 2016 and 2017.

The row and column are both generated by date .Color represents the number of killings for in a day. Rows and Columns are generated by Calculated Field to get Calendar look and divided years into 4 Quarters.

The calendar map analysis matches the rest of project data analysis. Some of the trends seen were-

- 2016 shows less shooting compared to 2015 and 2016.
- Quarter 4 show less number of shooting than Quarter 1, Quarter 2, Quarter 3.
- Weekends show less number of shootings than weekdays.

Police Shooting for 2015, 2016, 2017

Summary of Takeaway:

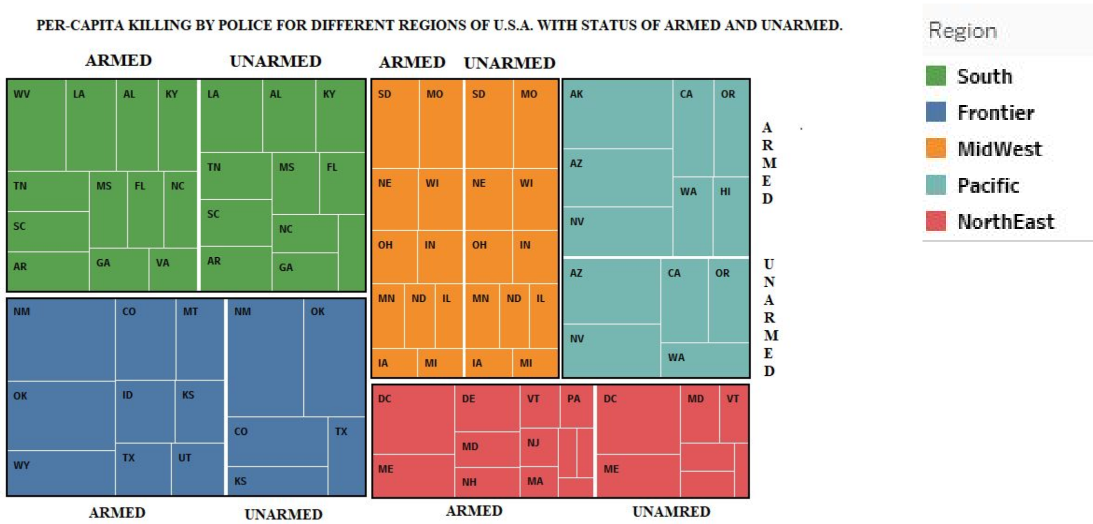
My biggest takeaway from this course has been, looking at the data and taking time to figure out data and appropriate visualizations that is best to communicate the message to the audience . I have more understanding of colors, shapes, angle, length etc now, which i wouldn't focus on much before this course.

Vivek Bhavshar Individual Report:

Main Visualization :

My main visualization is tree map which depicts information of per capita killings for states by different regions. Moreover, the regions are encoded with different colors also explained by the legend in the visualization, the size of each region in tree map depicts the amount of people killed in that region. Furthermore, the size of each state in particular region depicts killing rate for that state. In addition to that there are two main categories for victims that whether they were armed or unarmed.

In nutshell, we concluded that the south region has more killings where as the northeast region has the less killings when compared to other regions for both categories. Moreover, in the tree map we start from the top left corner having high values(i.e. high killing) to right bottom (i.e. less killing). For our data set we have different state name noted in each region component stating the per-capita killing in each state for two different categories.



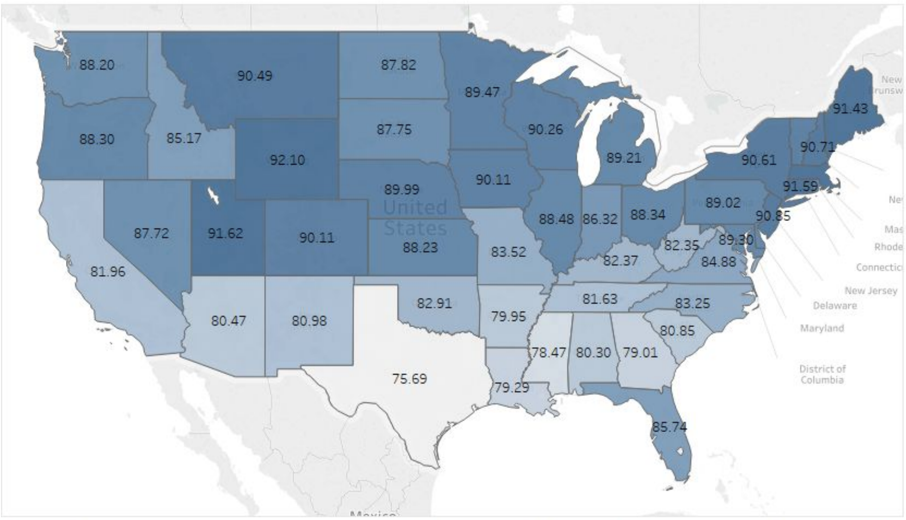
Other Visualizations contributions:

As a part of the other visualization, Choropleth of high school education for different states was made for education level in each state. The percentage in each state depicts the amount of people above the age of 25 who atleast have passed high school. Moreover, as we can see from the choropleth that northeast region has high education level while the south region has less when compared to the northeast.

Furthermore, if we correlate the education level and per capita killings, we can conclude that lesser the education in region higher the killing rate for that region. As we can see in the tree map as well that south region has more killings and northeast has less killings.

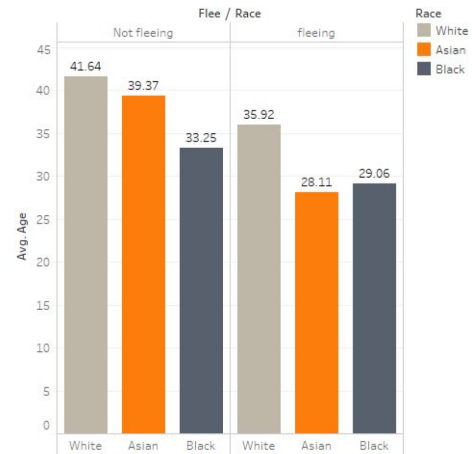
Another visualization depicts the fleeing status of victim whether fled or not, by different races.

Percent of People Completed Highschool for Different States.



Map based on Longitude (generated) and Latitude (generated). Color shows average of Percent Completed Hs. The marks are labeled by average of Percent Completed Hs. Details are shown for State. The view is filtered on State, which keeps 51 of 51 members.

Fled Status by Different Race.



Other Contributions:

Other than that I played part of suggesting and helping rest of the teammates with their visualization. Moreover, I also tried looking into different aspects of the dataset also made visualization but continuing the storey I kept the relevant visualizations. I have also helped to review literature and tried to bring unique perspective to the group also met team members on group meetings as well as individual when needed for any amendment.

Summary of Takeaways:

My takeaways from this course is that I learned how to represent the data into effective visualization which clearly conveys the message underlying the statics and to make the visualization considering the audience to convey clear message without any barrier.

R CODE

Correlation Matrix (Exploratory)

```
library(dplyr)
library(ggplot2)
library(gdata)
library(scales)
library(stringr)
library(ggcorrplot)

police_killings <- read.csv("fatal-police-shootings-data.csv")
median_household_income <- read.csv("MedianHouseholdIncome2015.csv")
percent_over_25_complete_hs <- read.csv("PercentOver25CompletedHighSchool.csv")
share_race_by_city <- read.csv("ShareRaceByCity.csv")
population_data <- read.csv("populationData.csv")
poverty_rates <- read.csv("poverty_rate.csv")



police_killings$date <- as.character(police_killings$date)
police_killings$date <- str_sub(police_killings$date, -2,-1)
police_killings$date <- as.factor(police_killings$date)

police_killings_grouped <- police_killings %>%
  group_by(date, state) %>%
  summarise(killings = n())

police_killings_grouped <- left_join(police_killings_grouped, population_data,
                                     by = c("state" = "Abbreviations"))

median_household_income$Median.Income <-
as.character(median_household_income$Median.Income)
median_household_income$Median.Income <-
as.numeric(median_household_income$Median.Income)

median_household_income_state <- median_household_income %>%
  group_by(Geographic.Area) %>%
  summarise(Median.Income = median(Median.Income, na.rm = T))
```

```
police_killings_grouped <- left_join(police_killings_grouped, median_household_income_state,
                   by = c("state" = "Geographic.Area"))

share_race_by_city$share_black <- as.character(share_race_by_city$share_black)
share_race_by_city$share_hispanic <- as.character(share_race_by_city$share_hispanic)

share_race_by_city$share_black <- as.numeric(share_race_by_city$share_black)
share_race_by_city$share_hispanic <- as.numeric(share_race_by_city$share_hispanic)


share_race_by_state <- share_race_by_city %>%
  group_by(Geographic.area) %>%
    summarise(perc_black = mean(share_black, na.rm = T), perc_hisp = mean(share_hispanic,
na.rm = T))

police_killings_grouped <- left_join(police_killings_grouped, share_race_by_state,
                   by = c("state" = "Geographic.area"))

police_killings_grouped$perc_minority <- police_killings_grouped$perc_black +
police_killings_grouped$perc_hisp

police_killings_grouped$perc_minority <- police_killings_grouped$perc_minority/100


poverty_rates$State <- state.abb[match(poverty_rates$State,state.name)]


police_killings_grouped <- left_join(police_killings_grouped, poverty_rates,
                   by = c("state" = "State"))


police_killings_grouped$per_capita_killings <- police_killings_grouped$killings
police_killings_grouped$per_capita_killings[police_killings_grouped$date == "15"] <-
(police_killings_grouped$killings[police_killings_grouped$date ==
"15"]/police_killings_grouped$POPESTIMATE2015[police_killings_grouped$date == "15"])
*100000
police_killings_grouped$per_capita_killings[police_killings_grouped$date == "16"] <-
(police_killings_grouped$killings[police_killings_grouped$date ==
```

```
"16"]/police_killings_grouped$POPESTIMATE2015[police_killings_grouped$date == "16"])
*100000
police_killings_grouped$per_capita_killings[police_killings_grouped$date == "17"] <-
(police_killings_grouped$killings[police_killings_grouped$date ==
"17"]/police_killings_grouped$POPESTIMATE2015[police_killings_grouped$date == "17"])
*100000
police_killings_grouped$per_capita_killings[police_killings_grouped$date == "18"] <-
(police_killings_grouped$killings[police_killings_grouped$date ==
"18"]/police_killings_grouped$POPESTIMATE2015[police_killings_grouped$date == "18"])
*100000

Pacific <- c("CA", "HI", "AK", "WA", "OR", "NV", "AZ")
Frontier <- c("ID", "MT", "UT", "OK", "TX", "KS", "NM", "CO","WY")
Midwest <- c("ND", "SD", "NE", "IA", "IL", "IN", "WI", "MI", "OH", "MN", "MO")
South <- c("KY", "TN", "WV", "FL", "GA", "MS", "AL", "LA", "AR", "VA", "NC", "SC")
NorthEast <- c("NY", "DC", "CT", "DE", "ME", "MD", "MA", "NH", "NJ", "PA", "RI", "VT")


police_killings_grouped$Region[police_killings_grouped$state %in% Pacific] <- 'Pacific'
police_killings_grouped$Region[police_killings_grouped$state %in% Frontier] <- 'Frontier'
police_killings_grouped$Region[police_killings_grouped$state %in% Midwest] <- 'Midwest'
police_killings_grouped$Region[police_killings_grouped$state %in% South] <- 'South'
police_killings_grouped$Region[police_killings_grouped$state %in% NorthEast] <- 'North East'
police_killings_grouped$Region <- as.factor(police_killings_grouped$Region)

all_states <- map_data("state")

MidwestNames <- tolower(state.name[match(Midwest,state.abb)])
SouthNames <- tolower(state.name[match(South,state.abb)])
FrontierNames <- tolower(state.name[match(Frontier,state.abb)])
PacificNames <- tolower(state.name[match(Pacific,state.abb)])cm
NorthEastNames <- tolower(state.name[match(NorthEast,state.abb)])

levels(police_killings_grouped$date)[1] <- "2015"
levels(police_killings_grouped$date)[2] <- "2016"
levels(police_killings_grouped$date)[3] <- "2017"
levels(police_killings_grouped$date)[4] <- "2018"

grouped_by_state <- police_killings_grouped %>%
```

```
  group_by(state) %>%
  summarise(State.Median.Income = mean(Median.Income),
          Percentage.Minority = mean(perc_minority),
          Killings.Per.Capita = mean(per_capita_killings),
          Poverty.Rate = mean(PovertyRate))

corr <- round(cor(grouped_by_state[-8,2:5]),2)
ggcorrplot(corr, lab = TRUE, type = "lower") +
  ggtitle("") +
  scale_x_discrete(labels = c("% Minority in each State", "Killings Per Capita", "Poverty Rate")) +
   scale_y_discrete(labels = c("State Median Income", "% Minority in Each State", "Killings Per
Capita"))
```

Armed vs Unarmed Bar Plot (Exploratory)

```
library(dplyr)
library(ggplot2)
library(reshape2)
library(scales)
police_killings_usa <- read.csv("fatal-police-shootings-data.csv")


levels(police_killings_usa$race)[levels(police_killings_usa$race)=="W"] <- "White"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="H"] <- "Hispanic"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="B"] <- "Black"
levels(police_killings_usa$race)[levels(police_killings_usa$race)==""] <- "Other"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="A"] <- "Other"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="N"] <- "Other"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="O"] <- "Other"

levels(police_killings_usa$armed)[levels(police_killings_usa$armed)=="unarmed"]          <-
"Unarmed"
levels(police_killings_usa$armed)[levels(police_killings_usa$armed)!="Unarmed"] <- "Armed"

police_killings_by_race <- police_killings_usa %>%
     filter(race == "White" | race == "Black" | race =="Hispanic" | race == "Other") %>%
     group_by(armed, race) %>%
     summarise(killings = n())
```

```
armed <- police_killings_by_race[police_killings_by_race$armed == "Armed", ]
unarmed <- police_killings_by_race[police_killings_by_race$armed == "Unarmed", ]

armed_total <- sum(armed$killings)

armed$total <- armed_total

armed$proportion <- armed$killings/armed$total

unarmed_total <- sum(unarmed$killings)

unarmed$total <- unarmed_total

unarmed$proportion <- unarmed$killings/unarmed$total

grouped_data <- rbind(armed, unarmed)

ggplot(data = grouped_data, aes(x = race, y = proportion)) +
    geom_bar(stat = "identity") +
    facet_wrap(~armed) +
    ylab("Percentage of Total") +
    xlab("\nRace of Person Killed") +
    scale_y_continuous(labels = percent) +
    theme_bw() +
    theme(strip.text.x = element_text(size = 18),
        axis.text.x = element_text( size = 12 ),
        axis.text.y = element_text( size = 12 ),
        axis.title.x = element_text(size = 16),
        axis.title.y = element_text(size = 14))
```

Choropleth of Killings Per Capita

```
library(dplyr)
library(ggplot2)
library(gdata)
library(scales)
library(plotly)
library(maps)
```

```
library(binomen)

options(scipen=999)
police_killings <- read.csv("PoliceKillingsUS.csv")
median_household_income <- read.csv("MedianHouseholdIncome2015.csv")
percentage_below_poverty <- read.csv("PercentagePeopleBelowPovertyLevel.csv")
percent_over_25_complete_hs <- read.csv("PercentOver25CompletedHighSchool.csv")
share_race_by_city <- read.csv("ShareRaceByCity.csv")
population_data <- read.csv("populationData.csv")

police_killings$date <- as.Date(police_killings$date, '%d/%m/%y' )

police_killings_grouped <- police_killings %>%
    group_by(state) %>%
    summarise(killings = n())

police_killings_grouped <- left_join(police_killings_grouped, population_data,
                    by = c("state" = "Abbreviations"))

median_household_income$Median.Income                                          <-
as.character(median_household_income$Median.Income)
median_household_income$Median.Income                                          <-
as.numeric(median_household_income$Median.Income)

median_household_income_state <- median_household_income %>%
    group_by(Geographic.Area) %>%
    summarise(Median.Income = median(Median.Income, na.rm = T))

police_killings_grouped <- left_join(police_killings_grouped, median_household_income_state,
                      by = c("state" = "Geographic.Area"))

share_race_by_city$share_black <- as.character(share_race_by_city$share_black)
share_race_by_city$share_hispanic <- as.character(share_race_by_city$share_hispanic)

share_race_by_city$share_black <- as.numeric(share_race_by_city$share_black)
share_race_by_city$share_hispanic <- as.numeric(share_race_by_city$share_hispanic)


share_race_by_state <- share_race_by_city %>%
```

```
    group_by(Geographic.area) %>%
     summarise(perc_black = mean(share_black, na.rm = T), perc_hisp = mean(share_hispanic,
na.rm = T))

police_killings_grouped <- left_join(police_killings_grouped, share_race_by_state,
                                     by = c("state" = "Geographic.area"))

police_killings_grouped$perc_minority <- police_killings_grouped$perc_black +
police_killings_grouped$perc_hisp

police_killings_grouped$perc_minority <- police_killings_grouped$perc_minority/100

police_killings_grouped$per_capita_killings <-
(police_killings_grouped$killings/police_killings_grouped$POPESTIMATE2016) *100000


library(maps)

percentage_below_poverty$poverty_rate <-
as.character(percentage_below_poverty$poverty_rate)
percentage_below_poverty$poverty_rate <-
as.numeric(percentage_below_poverty$poverty_rate)


percentage_by_state <- percentage_below_poverty %>%
    group_by(Geographic.Area) %>%
    summarise(perc_poverty = mean(poverty_rate, na.rm = T))

police_killings_grouped <- left_join(police_killings_grouped, percentage_by_state,
                                     by = c("state" = "Geographic.Area"))

library(plotly)

police_killings_grouped$perc_minority <- police_killings_grouped$perc_minority*100

police_killings_grouped$hover <- with(police_killings_grouped, paste(state, '<br>',
                                          "Median income", paste("$", format(Median.Income,big.mark =
",", sep = "")), "<br>",
```

```
                                        "Percent in poverty", paste(round(perc_poverty, 2),"%",sep =
""),"<br>",
                                            "Percent Hispanic or Black", paste(round(perc_minority,
2),"%",sep = "")))

l <- list(color = toRGB("white"), width = 2)

g <- list(
  scope = 'usa',
  projection = list(type = 'albers usa'),
  showlakes = TRUE,
  lakecolor = toRGB('white')
)

p <- plot_geo(police_killings_grouped, locationmode = 'USA-states') %>%
  add_trace(
    z = ~round(per_capita_killings, 2), text = ~hover, locations = ~state,
    color = ~per_capita_killings, colors = 'Purples'
  ) %>%
  colorbar(title = "Police Killings\nper 100,000 residents") %>%
  layout(
    title = '2015-2017 Police Killings by State',
    geo = g
  )

p
```

Scatterplot of Per Capita Killings vs Poverty Rates (RMD Code)

```
---
title: "Police Killings By Year"
output: flexdashboard::flex_dashboard
runtime: shiny

---

```{r global, include=FALSE}
library(dplyr)
library(ggplot2)
```
```

```
library(gdata)
library(scales)
library(stringr)
library(gridExtra)
library(maps)
options(scipen=999)
police_killings <- read.csv("fatal-police-shootings-data.csv")
median_household_income <- read.csv("MedianHouseholdIncome2015.csv")
percentage_below_poverty <- read.csv("PercentagePeopleBelowPovertyLevel.csv")
percent_over_25_complete_hs <- read.csv("PercentOver25CompletedHighSchool.csv")
share_race_by_city <- read.csv("ShareRaceByCity.csv")
population_data <- read.csv("populationData.csv")
poverty_rates <- read.csv("poverty_rate.csv")

police_killings$date <- as.character(police_killings$date)
police_killings$date <- str_sub(police_killings$date, -2,-1)
police_killings$date <- as.factor(police_killings$date)

police_killings_grouped <- police_killings %>%
  group_by(date, state) %>%
  summarise(killings = n())

police_killings_grouped <- left_join(police_killings_grouped, population_data,
                                     by = c("state" = "Abbreviations"))

median_household_income$Median.Income                                         <-
as.character(median_household_income$Median.Income)
median_household_income$Median.Income                                         <-
as.numeric(median_household_income$Median.Income)

median_household_income_state <- median_household_income %>%
  group_by(Geographic.Area) %>%
  summarise(Median.Income = median(Median.Income, na.rm = T))

police_killings_grouped <- left_join(police_killings_grouped, median_household_income_state,
                                     by = c("state" = "Geographic.Area"))

share_race_by_city$share_black <- as.character(share_race_by_city$share_black)
share_race_by_city$share_hispanic <- as.character(share_race_by_city$share_hispanic)
```

```
share_race_by_city$share_black <- as.numeric(share_race_by_city$share_black)
share_race_by_city$share_hispanic <- as.numeric(share_race_by_city$share_hispanic)


share_race_by_state <- share_race_by_city %>%
  group_by(Geographic.area) %>%
    summarise(perc_black = mean(share_black, na.rm = T), perc_hisp = mean(share_hispanic,
na.rm = T))

police_killings_grouped <- left_join(police_killings_grouped, share_race_by_state,
                                     by = c("state" = "Geographic.area"))

police_killings_grouped$perc_minority <- police_killings_grouped$perc_black +
police_killings_grouped$perc_hisp

police_killings_grouped$perc_minority <- police_killings_grouped$perc_minority/100

percentage_below_poverty$poverty_rate <-
as.character(percentage_below_poverty$poverty_rate)

percentage_below_poverty$poverty_rate <-
as.double(percentage_below_poverty$poverty_rate)

percentage_below_poverty <- percentage_below_poverty %>%
        group_by(Geographic.Area) %>%
        summarise(poverty_rate = mean(poverty_rate, na.rm = T))

police_killings_grouped <- left_join(police_killings_grouped, percentage_below_poverty,
                                     by = c("state" = "Geographic.Area"))




police_killings_grouped$per_capita_killings <- police_killings_grouped$killings
police_killings_grouped$per_capita_killings[police_killings_grouped$date == "15"] <-
(police_killings_grouped$killings[police_killings_grouped$date ==
"15"]/police_killings_grouped$POPESTIMATE2015[police_killings_grouped$date == "15"])
*100000
```

```
police_killings_grouped$per_capita_killings[police_killings_grouped$date == "16"] <- (police_killings_grouped$killings[police_killings_grouped$date == "16"]/police_killings_grouped$POPESTIMATE2015[police_killings_grouped$date == "16"]) *100000
police_killings_grouped$per_capita_killings[police_killings_grouped$date == "17"] <- (police_killings_grouped$killings[police_killings_grouped$date == "17"]/police_killings_grouped$POPESTIMATE2015[police_killings_grouped$date == "17"]) *100000
police_killings_grouped$per_capita_killings[police_killings_grouped$date == "18"] <- (police_killings_grouped$killings[police_killings_grouped$date == "18"]/police_killings_grouped$POPESTIMATE2015[police_killings_grouped$date == "18"]) *100000

Pacific <- c("CA", "HI", "AK", "WA", "OR", "NV", "AZ")
Frontier <- c("ID", "MT", "UT", "OK", "TX", "KS", "NM", "CO","WY")
Midwest <- c("ND", "SD", "NE", "IA", "IL", "IN", "WI", "MI", "OH", "MN", "MO")
South <- c("KY", "TN", "WV", "FL", "GA", "MS", "AL", "LA", "AR", "VA", "NC", "SC")
NorthEast <- c("NY", "DC", "CT", "DE", "ME", "MD", "MA", "NH", "NJ", "PA", "RI", "VT")


police_killings_grouped$Region[police_killings_grouped$state %in% Pacific] <- 'Pacific'
police_killings_grouped$Region[police_killings_grouped$state %in% Frontier] <- 'Frontier'
police_killings_grouped$Region[police_killings_grouped$state %in% Midwest] <- 'Midwest'
police_killings_grouped$Region[police_killings_grouped$state %in% South] <- 'South'
police_killings_grouped$Region[police_killings_grouped$state %in% NorthEast] <- 'North East'
police_killings_grouped$Region <- as.factor(police_killings_grouped$Region)

all_states <- map_data("state")

MidwestNames <- tolower(state.name[match(Midwest,state.abb)])
SouthNames <- tolower(state.name[match(South,state.abb)])
FrontierNames <- tolower(state.name[match(Frontier,state.abb)])
PacificNames <- tolower(state.name[match(Pacific,state.abb)])
NorthEastNames <- tolower(state.name[match(NorthEast,state.abb)])

levels(police_killings_grouped$date)[1] <- "2015"
levels(police_killings_grouped$date)[2] <- "2016"
levels(police_killings_grouped$date)[3] <- "2017"
levels(police_killings_grouped$date)[4] <- "2018"
```

````
poverty_rates$State <- state.abb[match(poverty_rates$State,state.name)]

police_killings_grouped <- left_join(police_killings_grouped, poverty_rates,
                            by = c("state" = "State"))

police_killings_grouped$PovertyRate[police_killings_grouped$state == "DC"] <- 19.0

police_killings_grouped$date <- as.character(police_killings_grouped$date)

police_killings_grouped$date <- as.numeric(police_killings_grouped$date)

```

Inputs {.sidebar}
-----------------------------------------------------------------------

Each dot represents a state. Larger dots represent larger states and colors represent what region of the United States the state is in. Use the slider below to view the scatterplot by year.

```{r}
sliderInput("year", label = "Year:",
        min = 2015, max = 2018, value = 2015, ticks = F)

```


Column
-----------------------------------------------------------------------


```{r flexdashboard = T}
library(ggplot2)
renderPlot({

#xleft = 1.63
#xright = 0.57
#ybottom = -.07
#ytop = .2
````

```
xleft = 1.62
xright = 0.57
ybottom = -.07
ytop = .2


plot1 <- ggplot(data = police_killings_grouped[police_killings_grouped$date == input$year,],
aes(x = PovertyRate, y = per_capita_killings, size = POPESTIMATE2016, fill = Region)) +
  geom_point(alpha = .9, pch=21, color = "black") +
  #ylab("Police Killings per 100,000 Residents") +
  #xlab("Median State Income in US Dollars") +
  #scale_size_continuous(guide = 'none', range = c(3,20)) +
  scale_size_continuous((name = "State Population"),
                        breaks = c(1000000,30000000),
                        labels = c("1 million", "30 million"), range = c(3,20)) +
  theme_bw() +
  guides(fill = guide_legend(override.aes = list(size=10), nrow = 3)) +
  labs(title = "Police Killings vs. Percentage of People Living in Poverty by State",
        subtitle = "States with higher poverty rates tend to have higher frequencies of police
killings",
      caption = "Source: The Washington Post",
      x = "Poverty Rate in %", y = "Police Killings per 100,000 Residents") +
    theme(axis.text = element_text(size = 11),
      plot.title = element_text(size = 16, face = "bold")) +
  scale_fill_manual(values = c("#56B4E9", "#F0E442", "#0072B2", "#D55E00", "#CC79A7"))

plot2 <- ggplot(all_states, aes(x=long, y=lat, group = group)) +
  geom_polygon(fill="grey", colour = "white") +
  coord_fixed(xlim = c(-130, -65), ylim = c(18, 55)) +
  geom_polygon(fill="#D55E00", data = filter(all_states, region %in% PacificNames)) +
  geom_polygon(fill="#56B4E9", data = filter(all_states, region %in% FrontierNames)) +
  geom_polygon(fill="#F0E442", data = filter(all_states, region %in% MidwestNames)) +
  geom_polygon(fill="#CC79A7", data = filter(all_states, region %in% SouthNames)) +
  geom_polygon(fill="#0072B2", data = filter(all_states, region %in% NorthEastNames)) +
  # scale_fill_continuous((name  ="Killings per\n100,000 residents")) +
  theme_void()

l1 = ggplot_build(plot1)
```

```
x1 = l1$layout$panel_ranges[[1]]$x.range[1]
x2 = l1$layout$panel_ranges[[1]]$x.range[2]
y1 = l1$layout$panel_ranges[[1]]$y.range[1]
y2 = l1$layout$panel_ranges[[1]]$y.range[2]
xdif = x2-x1
ydif = y2-y1
xmin  = x1 + (xleft*xdif)
xmax  = x1 + (xright*xdif)
ymin  = y1 + (ybottom*ydif)
ymax  = y1 + (ytop*ydif)


g2 = ggplotGrob(plot2)
plot3 = plot1 + annotation_custom(grob = g2, xmin=xmin, xmax=xmax, ymin=ymin, ymax=ymax)

plot3
#}, width = 800, height = 650)
}, width = 1020, height = 460)
```

Visualization 3: Proportion of Killings by Race for Armed and Unarmed Victims

```
library(dplyr)
library(ggplot2)
library(reshape2)
library(zoo)
library(stringr)
library(ggthemes)


police_killings_usa <- read.csv("fatal-police-shootings-data.csv")

police_killings_usa$date <- as.Date(police_killings_usa$date, '%m/%d/%y' )

#filter out 2018
police_killings_usa$year <- as.character(police_killings_usa$date)
police_killings_usa$year <- str_sub(police_killings_usa$year, 1, 4)
police_killings_usa <- police_killings_usa[police_killings_usa$year != "2018",]
```

```
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="W"] <- "White"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="H"] <- "Hispanic"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="B"] <- "Black"
levels(police_killings_usa$race)[levels(police_killings_usa$race)==""] <- "Other"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="A"] <- "Other"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="N"] <- "Other"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="O"] <- "Other"

levels(police_killings_usa$armed)[levels(police_killings_usa$armed)=="unarmed"] <-
"Unarmed"
levels(police_killings_usa$armed)[levels(police_killings_usa$armed)!="Unarmed"] <- "Armed"

police_killings_usa$race <- factor(police_killings_usa$race, levels = c("Black", "Other",
"Hispanic", "White"))


police_killings_by_race <- police_killings_usa %>%
    filter(race == "White" | race == "Black" | race =="Hispanic" | race == "Other") %>%
    group_by(armed, year, race) %>%
    summarise(killings = n())

armed <- police_killings_by_race[police_killings_by_race$armed == "Armed", ]
unarmed <- police_killings_by_race[police_killings_by_race$armed == "Unarmed", ]

#armed_total <- sum(armed$killings)
#armed$total <- armed_total
#armed$proportion <- armed$killings/armed$total
#unarmed_total <- sum(unarmed$killings)
#unarmed$total <- unarmed_total
#unarmed$proportion <- unarmed$killings/unarmed$total
#grouped_data <- rbind(armed, unarmed)

#2015 armed
armed_2015 <- armed[armed$year == "2015", ]
armed_total_2015 <- sum(armed_2015$killings)
```

```
armed_2015$total <- armed_total_2015
armed_2015$proportion <- armed_2015$killings/armed_2015$total
#2016 armed
armed_2016 <- armed[armed$year == "2016", ]
armed_total_2016 <- sum(armed_2016$killings)
armed_2016$total <- armed_total_2016
armed_2016$proportion <- armed_2016$killings/armed_2016$total
#2017 armed
armed_2017 <- armed[armed$year == "2017", ]
armed_total_2017 <- sum(armed_2017$killings)
armed_2017$total <- armed_total_2017
armed_2017$proportion <- armed_2017$killings/armed_2017$total

#2015 unarmed
unarmed_2015 <- unarmed[unarmed$year == "2015", ]
unarmed_total_2015 <- sum(unarmed_2015$killings)
unarmed_2015$total <- unarmed_total_2015
unarmed_2015$proportion <- unarmed_2015$killings/unarmed_2015$total
#2016 unarmed
unarmed_2016 <- unarmed[unarmed$year == "2016", ]
unarmed_total_2016 <- sum(unarmed_2016$killings)
unarmed_2016$total <- unarmed_total_2016
unarmed_2016$proportion <- unarmed_2016$killings/unarmed_2016$total
#2017 unarmed
unarmed_2017 <- unarmed[unarmed$year == "2017", ]
unarmed_total_2017 <- sum(unarmed_2017$killings)
unarmed_2017$total <- unarmed_total_2017
unarmed_2017$proportion <- unarmed_2017$killings/unarmed_2017$total


grouped_data <- rbind(armed_2015, armed_2016, armed_2017, unarmed_2015, unarmed_2016, unarmed_2017)


ggplot(grouped_data, aes(x = year, y = proportion, fill = race)) +
  geom_bar(stat="identity", position = "fill", colour = "white", lwd = 1) +
  facet_wrap(~armed) +
  scale_fill_manual(values = c("#999999", "#E69F00", "#56B4E9", "#009E73")) +
  theme_tufte() +
```

```
theme(legend.title=element_text(size=20, face = "bold"),
      legend.text=element_text(size=17),
      axis.title.x = element_text(size = 20, face = "bold"),
      strip.text.x = element_text(size = 18, face = "bold"),
      axis.text.x = element_text(size = 17, face = "bold"),
      axis.title.y = element_text(size = 18, face = "bold"),
      axis.text.y = element_text(size = 17, face = "bold"),
      plot.title=element_text(face="bold", size=18, hjust = 0.5)) +
guides(fill = guide_legend(override.aes = list(size = 10))) +
guides(fill=guide_legend(title="Race")) +
xlab("Year\n ") +
ylab("Proportion of Killings\n ")+
ggtitle("PROPORTION OF KILLINGS BY RACE FOR ARMED AND UNARMED VICTIMS")
```

Actual vs Predicted Police Killings by Race per State (Max Rodrigues and Deepika Individual Report)

```
library(dplyr)
library(ggplot2)
library(reshape2)

setwd("/Users/maxrodrigues/Desktop/DePaul  Grad  School/Data  Visualization/Final/predicted vs actual")
police_killings <- read.csv("police_killings_grouped_by_state.csv")
share_by_race <- read.csv("ShareRaceByCity.csv")
race_by_state <- read.csv("RaceByState.csv")

police_killings_usa <- read.csv("PoliceKillingsUS.csv")

#police_killings$POPESTIMATE2015 <- police_killings$POPESTIMATE2015/1000000
#police_killings <- police_killings[,c(5,9,12,13,14)]
#chart.Correlation(police_killings, histogram = T, method = pearson)

race_by_state <- race_by_state[,2:6]
race_by_state <- melt(race_by_state, "Abbrev")

levels(police_killings_usa$race)[levels(police_killings_usa$race)=="W"] <- "White"
```

```
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="H"] <- "Hispanic"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="B"] <- "Black"
levels(police_killings_usa$race)[levels(police_killings_usa$race)==""] <- "Other"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="A"] <- "Other"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="N"] <- "Other"
levels(police_killings_usa$race)[levels(police_killings_usa$race)=="O"] <- "Other"

levels(police_killings$armed)[levels(police_killings$armed)=="gun"] <- "Gun"
levels(police_killings$armed)[levels(police_killings$armed)=="gun and knife"] <- "Gun"
levels(police_killings$armed)[levels(police_killings$armed)=="guns and explosives"] <- "Gun"
levels(police_killings$armed)[levels(police_killings$armed)=="knife"] <- "Knife"

police_killings_by_state_race <- police_killings_usa %>%
    filter(race == "White" | race == "Black" | race =="Hispanic" | race == "Other") %>%
    group_by(state, race) %>%
    summarise(killings = n())

police_killings_by_state_race <- dcast(police_killings_by_state_race, state ~ race )

police_killings_by_state_race[,2:5][is.na(police_killings_by_state_race[,2:5])] <- 0

police_killings_by_state_race$total <- police_killings_by_state_race$Other +
                    police_killings_by_state_race$Black +
                    police_killings_by_state_race$Hispanic +
                    police_killings_by_state_race$White
police_killings_by_state_race$White                                    <-
police_killings_by_state_race$White/police_killings_by_state_race$total
police_killings_by_state_race$Other                                    <-
police_killings_by_state_race$Other/police_killings_by_state_race$total
police_killings_by_state_race$Hispanic                                 <-
police_killings_by_state_race$Hispanic/police_killings_by_state_race$total
police_killings_by_state_race$Black                                    <-
police_killings_by_state_race$Black/police_killings_by_state_race$total


police_killings_by_state_race <- melt(police_killings_by_state_race[,1:5], "state")

colnames(police_killings_by_state_race) <- c("state", "race", "actual_proportion")
```

```
police_killings <- left_join(police_killings_by_state_race, race_by_state,
                    by = c("state" = "Abbrev", "race" = "variable"))

#more positive means more than expected
police_killings$Difference = police_killings$actual_proportion - police_killings$value
police_killings$Diff_format <- as.factor(police_killings$Difference < 0)




ggplot(data=police_killings, aes(x = state, y = Difference, fill = Diff_format)) +
    geom_bar(stat ="identity") +
    coord_flip() +
    facet_wrap(~race) +
    theme(axis.text.y = element_blank(),
        axis.ticks = element_blank(),
        panel.background = element_rect(fill = "white", colour = "grey50")) +
    scale_y_continuous(breaks = c(-.75,-.50, -.25, 0, .25, .50, .75),
                       labels = c("-.75","-.50\nLess than expected","-.25", "0.0","+.25","+.50\nMore
than expected", "+.75")) +
    ylab("") +
    guides(fill=FALSE) +
    xlab("State") +
    ggtitle("Actual vs expected proportions of police killings for each race by state")
```

R code for Diffusion Cartogram:

```
if (!require("devtools")) install.packages("devtools")
devtools::install_github("pvictor/topogRam")
library("topogRam")
library("RColorBrewer")
PKUS<-read.csv(file.choose())
head(PKUS)
key_var <- list(
  list(name = "killings", key = "killings"),
  list(name = "Population Estimate for year 2015", key = "POPESTIMATE2015"),
  list(name = "Population Estimate for year 2016", key = "POPESTIMATE2016"),
  list(name = "Population Estimate for year 2017", key = "POPESTIMATE2017"),
  list(name = "Median Income per state", key = "Median.Income"),
```

```
  list(name = "Percentage of black population", key = "perc_black"),
  list(name = "Percentage of hispanic population", key = "perc_hisp"),
  list(name = "Percentage of minority population", key = "perc_minority"),
  list(name = "Per capita Killings", key = "per_capita_killings"),
  list(name = "Poverty rate", key = "perc_poverty")
)
key_var <- lapply(
  X = key_var,
  FUN = function(x) {
    x$key <- sprintf(x$key)
    x$lab <- ""
    x
  }
)

topogRam(
  data = PKUS,
  key_var = key_var,
  shape = "usa-states",
  geo_lab = "region",
  col = brewer.pal("PuRd", n = 3),
  width = 800, height = 500
)
```

REFERENCES

[1] The dataset was taken from: https://www.kaggle.com/washingtonpost/police-shootings

[2] US Census Demographics: https://www.census.gov/quickfacts/fact/table/US/PST045217

[3] US Census Demographics by Race: https://www.census.gov/quickfacts/fact/table/US/PST045216

[4] The Washington Post is compiling a database of every fatal shooting in the United States by a police officer in the line of duty in 2015 and 2016. (2017). The Washington Post. Retrieved from https://github.com/washingtonpost/data-police-shootings (Original work published June 30, 2015)

[4] Kappeler, V. E. (2014, January 7). A Brief History of Slavery and the Origins of American Policing |http://plsonline.eku.edu/insidelook/brief-history-slavery-and-origins-american-policing

[5] The Counted: tracking people killed by police in the United States | US News. (n.d.). Retrieved June 15, 2017, from http://www.theguardian.com/us-news/series/counted-us-police-killings

[6]Jagger, T. (2017, July 04). How to Optimize Charts For Color Blind Readers. Retrieved May 26, 2018, from https://venngage.com/blog/color-blind-friendly-palette/