Police Fatatlity and Military Surplus Spending

Duke MIDS - IDS 702 Final Project

Christopher Oblak

11/20/2020

Summary

Police Brutality has led to an epic polarization of today's public opinion of police forces across the country. These stressful times have led to what I considered an identity crisis for police officers. Instead of taking a step back and reassessing a department's professional conduct, there seems to be a ramping up of militarization: both in equipment and rhetoric. This analysis investigates the potential association between the money a state law enforcement department spends on military surplus and the rates of deaths caused by those proprietors. This report's key findings show that there is a statistical significance that the use of body cameras and the smaller the ratio of military spending consumed by police departments directly decrease the odds of deaths being caused by Policing and Trooper (future reference in the report with only reference "police") Departments within a state. In addition, the more money a state government spends per person on military surplus, the odds of policing departments to cause deaths increase.

Introduction

There are images all across the nation where police officers carry highly customized military rifles and ride around in vehicles at home in news reports of Afghanistan and Iraq's front lines. This report seeks to look at the effects of police forces' physical militarization and how that affects its propensity for increase lethal force. In highlighting potential trends, I hope to shed light on the need for police departments to seek other approaches to keeping their city and themselves safe instead of instigating an issue by seeking militarization rather than better policing. To help highlight potential trends for increased death rates among police and trooper departments, I began this analysis with the following questions in mind.

1. Is there a relationship between the overall values of military surplus a state spends and the normalized police brutality rate to their population?
2. Are there specific equipment (weapons vs. officer equipment as an example) that provide better indicators of police-involved deaths?

By asking these questions, I hope to shed light on dangerous trends that could enlighten policing forces and state governments on how their spending habits influence their policing ideology and ultimately encourage less aggressive approaches to "keeping the peace" within our nation.

Variable	Description
Year	Year data was collected 2014-2019
State	State of Police Fatalities and Surplus Bought
TotalDeathCount	Total Police Fatality for the year and state
Population	Population estimate for each state by year
BodyCamUsageCat	Category of Body Cam Usage in Fatality Shooting ("0-40%", "40-55%", "55-70%", "70-85%", "85-100%")
TotalAcquisitionValueSum	Total Money Spent on Military Surplus by State and Year
NormalizedDeaths	Total Deaths normailized by the popualtion
NormalizedMiliSurSpending	Total Military Surplus Spending by each State
PoliceSpendingCat	Precentage of Police Dept Spending of the Total ("0-40%", "40-60%", "60-100%")
SheriffSpendingCat	Precentage of Sheriff Dept Spending of the Total ("0-40%", "40-60%", "60-100%")
DEMIL_Code_A	Number of Mis Equipment (tents, fridges, backpacks, ect)
DEMIL_Code_B	Number of Tools and Aircraft Equipment
DEMIL_Code_C	Number of Tactical Vehicles
DEMIL_Code_D	Number of Weapons
DEMIL_Code_E	Number of Uniforms / Personal Equipment
DEMIL_Code_F	Number of Weapons Scopes and Auxiliary
DEMIL_Code_Q	EUD Robots, Repair Parts, Tools, Non Tactical Equipment
PoliceAcquSum	Aggrigated amount spent by poilce deparmtents from the over all state spendings
PoliceCount	Number of fatalities cuased by police department personell
SheriffAcquSum	Aggrigated amount spent by sheriff/marshall deparmtents from the over all state spendings
SheriffCount	Number of fatalities caused by sheriff department personnel
PublicSafetyAcquSum	Aggrigated amount spent by highway safety/wildlife/fishing/mis deparmtents from the over all state spendings
PoliceTrooperDeath	Number of fatalities caused by highway safety/wildlife/fishing/mis department personnel

Figure 1: Data Variable Table

Data

When looking at how to answer the questions above, I acquired two data sources that would allow me to investigate potential trends of military surplus spending against police-induced deaths. The first was a comprehensive list of deaths from 2015 to 2019 (relatively present) consolidated by the Washington Post. This data included everything from the victim's name, the use of body cameras during the incident, the police department's title, and much more. I was also able to find the Defence Logistics Agency's 1033 report that contained information on surplus equipment being bought by all types of law enforcement agencies. The report has been annually published since the program's inception in 1997. The data set's key factors are the description of the items, acquisition cost, and DEMIL Codes. The DEMIL code categorizes the equipment by its "sensitivity," which is how the government distinguishes a weapon from repair parts for radios.

This data and population data collected from the US Census Bureau were aggregated and merged to build a data set that provides variables death counts, the amounts of money spent (both normalized over the population), types of equipment, ratios of spending on various types of law enforcement, and total deaths (See above table for data descriptions). It is also important to note that the Poisson distribution assumption hold for this data that the events occur independently, at random, and the probability of an event occurring in a given interval does not vary with time.

Exploritory Data Analysis

After aggregating and merging all the data and checking that my aggregate variables were correct, I looked at the number of data points each year and state had that I would be able to use. In this case, I had aggregated the data from the year 2013 to 2019. The below tables break down the number of states I had for each year, and from this, I felt comfortable moving forward with the data I had and continued my EDA.

Table 1: Number of Variables by Year

2013	2014	2015	2016	2017	2018	2019
48	49	44	45	46	47	44

Interested in a potential trend relating directly to my original question, I started my EDA by plotting the total number of deaths and total spending by each state. The graph showed a positive correlation: the more money spent, the more likely a state would have more deaths. Unfortunately, the graph was also heavily skewed toward the bottom left, making me unsure if I could trust this trend. I then normalized these variables by dividing them by the population and graphing them again. The positive trend of spending and deaths still showed, but the massive group of points in the bottom left again made me question the trend's validity (see Fig 2: Death vs Spending Plots).

After looking at the significant interactions, I was interested in plotting states as boxplot against their respective normalized death counts to see if there were any large variance between states (See Fig. 2: Normalized Deaths by State). As you can see from the plot below, there is a considerable variation that I needed to take into account for model building. I eventually transitioned to a hierarchical model and grouped by state to encompass those variations.

It is also important to note that I looked at interactions between spending ratios categories and DEMIL codes to see if any trends would elicit them to be included in model selection. I took careful note to find and exclude highly correlated variables, like the variable for total spending on military surplus and subcategories of spending normalized spending and total spending by different types of departments.

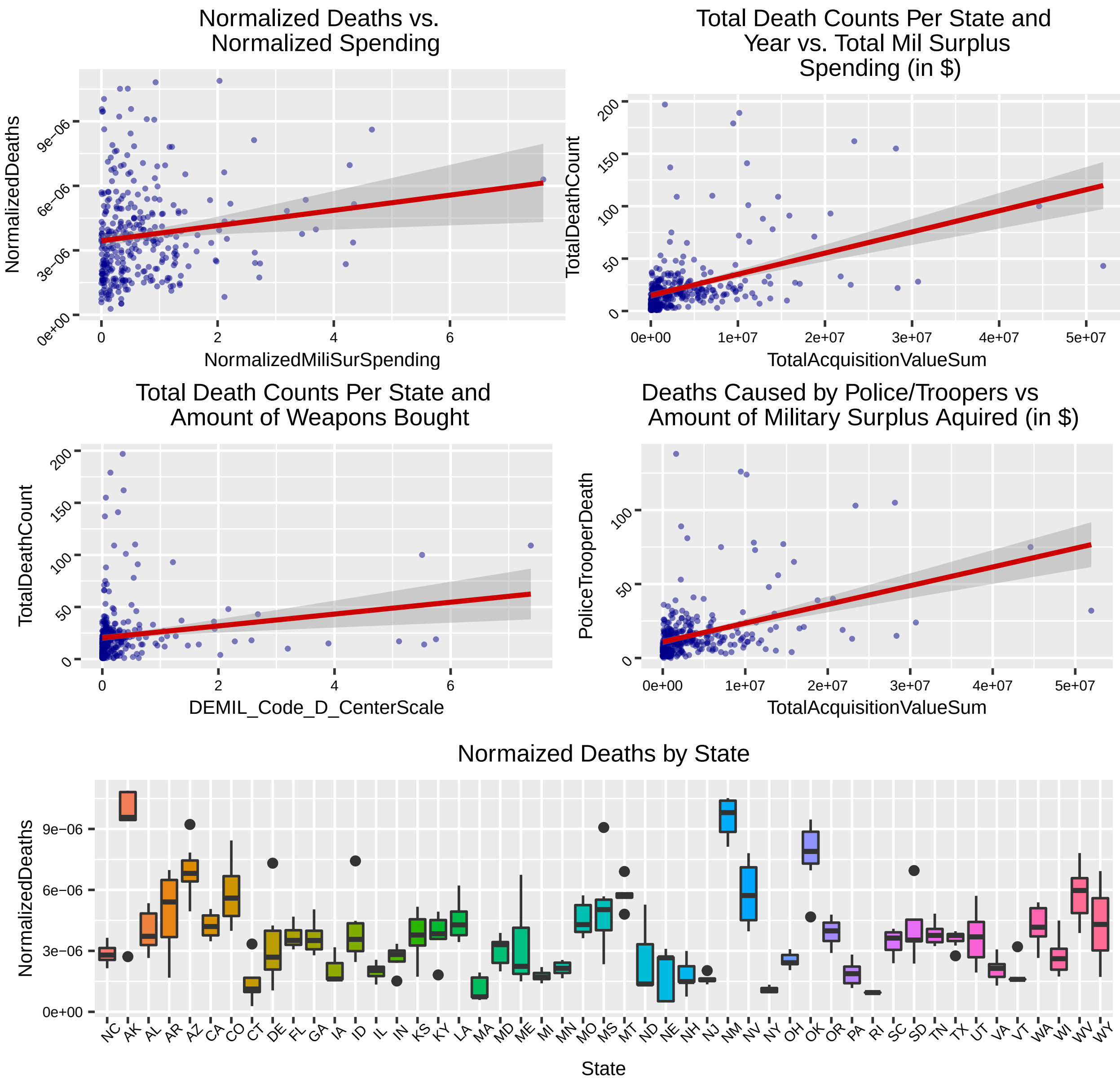

Figure 2: EDA Plots

Model

After grasping the data and some of the critical factors I believed could help answer the original questions, I began modeling the Total Death Counts and all significant variables from my EDA. Initially, there was no variable other than individual states that were statistically significant. This drove the decision to change the model to a hierarchical one that I could then group by state, specifically allowing for a varying intercept. The other important factor about this model is that I chose to use a Poisson model in terms of rate. By incorporating an offset for the response variable by the respective state and year's estimated population from the US Census Bureau, I was able to normalize the deaths and compare counts between states like Delaware and Texas on a practical scale.

From here, I conducted an AIC selection process that stepped in both directions to find variables that would be more significant than trying to force a full model of all my variables. Unfortunately, the total deaths' response variable was not returning relationships with the independent variables that made sense. I shifted my focus to only deaths that occurred by police departments. After conducting AIC selection again with the new count response variable, I began to see models that made more sense. At this point, the model included the variables of body camera use, state spending ratio on police departments, the overall normalized spending, and DEMIL Codes Q, F, and E. The DEMIL Code for Weapons (DEMIL Code D) was not selected, which was the direct variable needed to answer my second question. I was still curious if the number of weapons purchased would have an impact on the model. I tested this by comparing the models with an ANOVA test; resulting in the DEMIL Code D addition not being sufficiently significant. The results of the test illustrated worse AIC and BIC scores and a P-value of 0.077 for the model with the included variable.

$$log(PoliceDeathCount) = \beta_0 + \gamma_0 State + \beta_{2:5} BodyCamUsageCat+$$
$$\beta_6 TotalAcquisitionValue(centered/scaled) + \beta_{7:8} PoliceSpendingCat+$$
$$\beta_9 DEMIL\_Code\_Q(centered/scaled) + \beta_{10} DEMIL\_Code\_C(centered/scaled)+$$
$$log(Population)$$

After settling on the model above, I validated it by checking the normality with a QQ plot and the independence through a residual plot (see Fig 3). Both plots gave me confidence that the model was appropriate to interpret the relationships between the response variable and independent variables.

	PoliceCount		
Predictors	*Incidence Rate Ratios*	*CI*	*p*
(Intercept)	0.00	0.00 – 0.00	**<0.001**
BodyCamUsageCat [40-55%]	0.72	0.56 – 0.93	**0.011**
BodyCamUsageCat [55-70%]	0.64	0.57 – 0.72	**<0.001**
BodyCamUsageCat [70-85%]	0.78	0.72 – 0.84	**<0.001**
BodyCamUsageCat [85-100%]	0.68	0.62 – 0.75	**<0.001**
NormalizedMiliSurSpending	1.26	1.21 – 1.31	**<0.001**
PoliceSpendingCat [40-60%]	1.22	1.11 – 1.33	**<0.001**
PoliceSpendingCat [60-100%]	1.37	1.24 – 1.51	**<0.001**
DEMIL_Code_Q_CenterScale	1.05	1.00 – 1.10	0.059
DEMIL_Code_F_CenterScale	0.93	0.88 – 0.98	**0.003**
DEMIL_Code_E_CenterScale	1.03	1.01 – 1.05	**0.017**
Random Effects			
σ^2	13.22		
$\tau_{00\ State}$	0.36		
ICC	0.03		
N_{State}	49		
Observations	323		
Marginal R^2 / Conditional R^2	0.007 / 0.033		

Figure 3: Final Model Summary Table

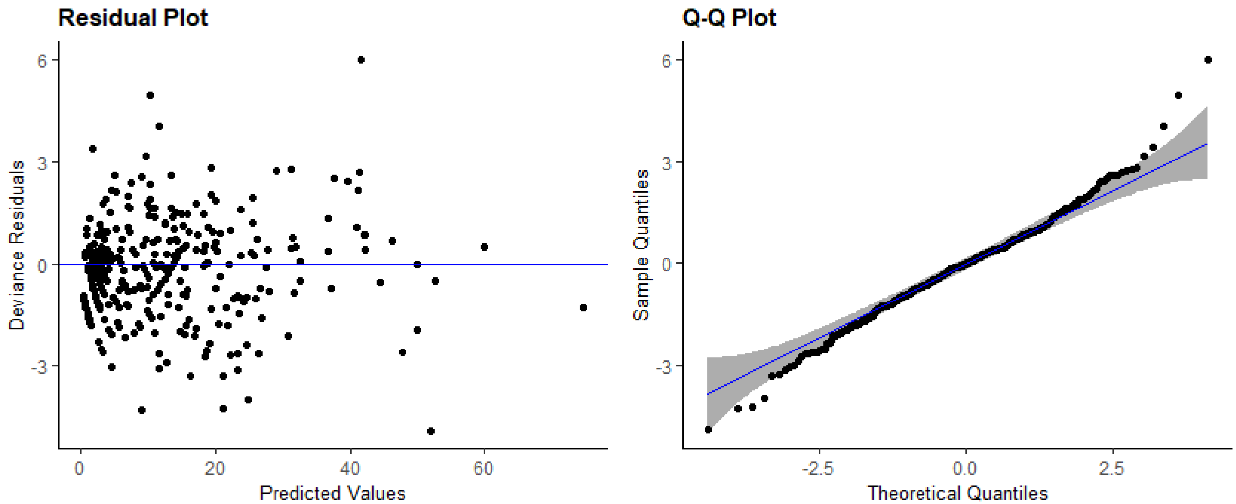

Figure 3: Final Model Summary Table

Significant findings can be interpreted as the following:

1. Random Effect: Intercepts vary by 0.36 between states and has a calculated variance with groups as 13.22. Showing that each state differs little compared to each other in terms of police caused death, but tent to have more considerable variation with in-state by year.
2. As body camera usage increases from the range of 0-40% (while all else holds the same) , we see police/trooper department related death fatalities decrease with a high level of statistical significance: (40-55% usage): 28% decrease, (55-70% usage): 36% decrease, (70-85% usage): 22% decrease, and (85-100% usage): 32% decrease.
3. As the ratio of the money spent on military surplus going to police/trooper departments within a state increase, the rate in which death also occurs increases: (40-60% usage): 22% (60-100% usage): 37%
4. As the total number of DEMIL-ed equipment under the "F" (Number of Weapons Scopes and Auxiliary) category increases on the mean, standard scale, and while everything else holds the same, death counts decreases at 7%.

Conclusion

The final model validates most of my assumptions about the effects of spending money on military surplus and its association with deaths caused. Interestingly enough, the number of weapons a state allows bought has little association with this metric. However, there are limitations to this analysis that I hope to address in the future. First, to conduct a true study of spending habits and its likely hood to show association with death cause, I would need access to law enforcement's full budgetary breakdown. Many military specific items that are not surplus and are being purchased by law enforcement agencies could highlight even more robust associations between police fatalities and the desire to mimic military forces. A second limitation is finding data before enacting the 1033 program and comparing death cases overtime when the military surplus was not readily available. Comparing the budget before and after and adding a military surplus component could be very insightful.

Arguably, and based on this analysis, when policing agencies feel the need to spend money on equipment generally designed for war, their willingness to use those equipment items becomes higher. I don't pretend to understand situations in which law enforcement personnel are in on a daily bases, and death will also be associated with policing, but I hope this analysis shows that spending money on "military equipment" can shed light on the message an agency has for dealing its own violent ends.

R Appendix

```r
knitr::opts_chunk$set(echo = TRUE, fig.pos = "h")

library(ggplot2)
library(readxl)
library(tibble)
library(tidyverse)
library(dbplyr)
library(rio)
library(tidyr)
library(readr)
library(jtools)
library(reshape)
library(sjPlot)
library(lme4)
library(pander)
library(stargazer)
library(knitr)
library(gt)
library(gridExtra)
library(grid)
library(qqplotr)
library(ggResidpanel)
setwd("C:/Users/Chris/Desktop/MIDS/Fall 2020/IDS 702 Modeling and Rep of Data/FinalProject")
policebrut_df = read.csv("data/MPVDatasetDownload.csv")[ ,
                  c('Victim.s.gender', 'Date.of.Incident..month.day.year.',
                  'City', 'State', 'County',
                  'Agency.responsible.for.death', 'Cause.of.death',
                  'Criminal.Charges.', 'Unarmed.Did.Not.Have.an.Actual.Weapon',
                  'Alleged.Threat.Level..Source..WaPo.',
          'Alleged.Weapon..Source..WaPo.and.Review.of.Cases.Not.Included.in.WaPo.Database.'
                  ,'Body.Camera..Source..WaPo.')]
#inspect data set
head(policebrut_df)

#redefine data set columns to more readable data
colnames(policebrut_df)[1]  <- "VictimGender"
colnames(policebrut_df)[2]  <- "DateofDeath"
colnames(policebrut_df)[6]  <- "PoliceAgencyName"
colnames(policebrut_df)[7]  <- "CauseOfDeath"
colnames(policebrut_df)[8]  <- "CriminalCharges"
colnames(policebrut_df)[9]  <- "Unarmed"
colnames(policebrut_df)[10] <- "AllegedTheatLevel"
colnames(policebrut_df)[11] <- "AllegedWeapon"
colnames(policebrut_df)[12] <- "BodyCamera"

#inspect and reassign the columns to appropriate data types

typeof(policebrut_df$DateofDeath)
as.Date(policebrut_df$DateofDeath, "%m/%d/%Y")

policebrut_df$DateofDeath <- as.Date(policebrut_df$DateofDeath, "%m/%d/%Y")
```

```
policebrut_df$YearofDeath <- format(as.Date(policebrut_df$DateofDeath), "%Y")


premerge_Police_df <- policebrut_df %>% group_by(State, County,
                                                  PoliceAgencyName, YearofDeath) %>%
  summarise(DeathCount = n_distinct(DateofDeath),
            GunShotDeathCount = sum(CauseOfDeath=="Gunshot"),
            TazerDeathCount = sum(CauseOfDeath=="Gunshot, Taser"),
            BodyCameraNo = sum(BodyCamera=="No", na.rm = TRUE))

colnames(premerge_Police_df)[4]  <- "Year"

grouped_police_df <- premerge_Police_df %>%
  group_by(State, Year) %>%
  summarise(TotalDeathCount = sum(DeathCount),
            GunShotDeathCount = sum(GunShotDeathCount),
            TazerDeathCount = sum(TazerDeathCount),
            BodyCameraNo = sum(BodyCameraNo),
            #PoliceDeath = sum(DeathCount[str_count(PoliceAgencyName, "Police|PD|Troopers")]),
            PoliceTrooperDeath = sum(DeathCount[grepl("Police|Troopers",
                                                      PoliceAgencyName)]),
            #SheriffTrooperDeath = sum(DeathCount[str_count(PoliceAgencyName, "Sheriff")]),
            SheriffMarshPatrolDeath = sum(DeathCount[grepl("Sheriff|Marshals|Patrol",
                                                           PoliceAgencyName)]),
            #SafetyOtherDeath = sum(DeathCount[str_count(PoliceAgencyName, "Safety|Game")])
            SafetyOtherDeath = sum(DeathCount[grepl("Safety|Game|Border|Forest",
                                                    PoliceAgencyName)])
            )

#Input and clean military surplus Dataset in preparation for merging

surplus_df <- excel_sheets("data/DISP_AllStatesAndTerritories_09302020.xlsx") %>%
  map_df(~read_xlsx("data/DISP_AllStatesAndTerritories_09302020.xlsx",.))

colnames(surplus_df)[2]  <- "RecivedStation"
colnames(surplus_df)[4]  <- "ItemName"
colnames(surplus_df)[7]  <- "AcquisitionValue"
colnames(surplus_df)[8]  <- "DEMIL_Code"
colnames(surplus_df)[9]  <- "DEMIL_IC"
colnames(surplus_df)[10] <- "ShipYear"
colnames(surplus_df)[11] <- "StationType"

surplus_df$ShipYear <- format(as.Date(surplus_df$ShipYear), "%Y")

truckstr = c("TRUCK", "VEHICLE")
Safety_df = c("SAFETY", "GAME")

premerge_surplus_df <- surplus_df %>% group_by(State, ShipYear) %>%
  summarise(TotalAcquisitionValueSum = sum(AcquisitionValue),
            PoliceAcquSum = sum(AcquisitionValue[grepl("POLICE|PD|TROOPERS",
                                                       RecivedStation)]),
            PoliceCount = sum(str_count(unique(RecivedStation),"POLICE|PD|TROOPERS")),
            SheriffAcquSum = sum(AcquisitionValue[grepl("SHERIFF|MARSHALS|PATROL",
```

```
                                                        RecivedStation)]),
            SheriffCount = sum(str_count(unique(RecivedStation), "SHERIFF|MARSHALS|PATROL")),
            PublicSafetyAcquSum = sum(AcquisitionValue[grepl("SAFETY|GAME|BORDER|FOREST",
                                                             RecivedStation)]),
            PublicSafety = sum(str_count(unique(RecivedStation),
                                         "SAFETY|GAME|BORDER|FOREST")),
            DEMIL_Code_A = sum(Quantity[str_count(DEMIL_Code, "A")]),
              #Mis Equipment (tents, fridge, assault packs, ect)
            DEMIL_Code_B = sum(Quantity[str_count(DEMIL_Code, "B")]),
              #tools and aircraft equipment
            DEMIL_Code_C = sum(Quantity[str_count(DEMIL_Code, "C")]),
              #Vehicles
            DEMIL_Code_D = sum(Quantity[str_count(DEMIL_Code, "D")]),
              #Weapons
            DEMIL_Code_E = sum(Quantity[str_count(DEMIL_Code, "E")]),
              #Uniforms
            DEMIL_Code_F = sum(Quantity[str_count(DEMIL_Code, "F")]),
              #Sights/Scopes
            DEMIL_Code_G = sum(Quantity[str_count(DEMIL_Code, "G")]),
            DEMIL_Code_P = sum(Quantity[str_count(DEMIL_Code, "P")]),
            DEMIL_Code_Q = sum(Quantity[str_count(DEMIL_Code, "Q")])
            )

#Merge our two datasets

colnames(premerge_surplus_df)[2]  <- "Year"
merged_df <- merge(grouped_police_df, premerge_surplus_df, by = c("State",
                                                                  "Year"))

#Pull in population datasets

pop_df <- read_xlsx("data/nst-est2019-01.xlsx", skip = 3)
colnames(pop_df)[1]  <- "State"
pop_df$State <- str_replace(pop_df$State, ".", "" )
pop_df <- pop_df[6:56,]
pop_df <- pop_df[-c(9), ]
pop_df <- pop_df[,-c(2,3,4,5,6)]

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

pop_df <- pop_df %>%
  pivot_longer(!State, names_to = "Year", values_to = "Population")

#Merge population data and the rest of the dat

final_merged_df <- merge(merged_df, pop_df, by = c("State", "Year"))


final_merged_df$NormalizedDeaths =
  final_merged_df$TotalDeathCount / final_merged_df$Population
final_merged_df$NormalizedMiliSurSpending =
  final_merged_df$TotalAcquisitionValueSum / final_merged_df$Population
```

```r
final_merged_df$State   = factor(final_merged_df$State)
final_merged_df$Year    = as.integer(final_merged_df$Year)
final_merged_df$Year14  = final_merged_df$Year - min(final_merged_df$Year)
final_merged_df <- within(final_merged_df, State <- relevel(State, ref = "NC"))


# Ratio Variables

final_merged_df$PoliceSpentRatio =
  (final_merged_df$PoliceAcquSum / final_merged_df$TotalAcquisitionValueSum)
final_merged_df$SheriffSpentRatio =
  (final_merged_df$SheriffAcquSum / final_merged_df$TotalAcquisitionValueSum)

# Categorical Police/Sheriff Ratios

final_merged_df$PoliceSpendingCat =
  cut(final_merged_df$PoliceSpentRatio, breaks = c(-Inf,.4,.6, 1),
                                        labels = c("0-40%", "40-60%", "60-100%"))

final_merged_df$SheriffSpendingCat =
  cut(final_merged_df$SheriffSpentRatio, breaks = c(-Inf,.4,.6, 1),
                                        labels = c("0-40%", "40-60%", "60-100%"))

final_merged_df$BodyCamUsage =  (final_merged_df$BodyCameraNo / final_merged_df$TotalDeathCount)
final_merged_df$BodyCamUsageCat = cut(final_merged_df$BodyCamUsage,
                                      breaks = c(-Inf,.4, .55 ,.7, .85, 1),
                                      labels = c("0-40%", "40-55%",
                                                 "55-70%", "70-85%", "85-100%"))


final_merged_df = final_merged_df[, c("Year", "State", "TotalDeathCount",
                                      "Population","BodyCamUsageCat", "TotalAcquisitionValueSum",
                                      "NormalizedDeaths",  "NormalizedMiliSurSpending",
                                      "PoliceSpendingCat", "SheriffSpendingCat",
                                      "DEMIL_Code_A", "DEMIL_Code_B", "DEMIL_Code_C",
                                      "DEMIL_Code_D", "DEMIL_Code_E", "DEMIL_Code_F",
                                      "DEMIL_Code_Q", "PoliceAcquSum", "PoliceCount",
                                      "SheriffAcquSum", "SheriffCount", "PublicSafetyAcquSum",
                                      "PoliceTrooperDeath", "SheriffMarshPatrolDeath",
                                      "SafetyOtherDeath", "Year14")]

#scaling varialbes

final_merged_df$TotalAcquisitionValueSum_CenterScale =
  scale(final_merged_df$TotalAcquisitionValueSum, center = TRUE, scale = TRUE)
final_merged_df$DEMIL_Code_C_CenterScale = scale(final_merged_df$DEMIL_Code_C,
                                                 center = FALSE, scale = TRUE)
final_merged_df$DEMIL_Code_A_CenterScale = scale(final_merged_df$DEMIL_Code_A,
                                                 center = FALSE, scale = TRUE)
final_merged_df$DEMIL_Code_D_CenterScale = scale(final_merged_df$DEMIL_Code_D,
                                                 center = FALSE, scale = TRUE)
final_merged_df$DEMIL_Code_Q_CenterScale = scale(final_merged_df$DEMIL_Code_Q,
                                                 center = FALSE, scale = TRUE)
```

```r
final_merged_df$DEMIL_Code_B_CenterScale = scale(final_merged_df$DEMIL_Code_B,
                                                 center = FALSE, scale = TRUE)
final_merged_df$DEMIL_Code_F_CenterScale = scale(final_merged_df$DEMIL_Code_F,
                                                 center = FALSE, scale = TRUE)
final_merged_df$DEMIL_Code_E_CenterScale = scale(final_merged_df$DEMIL_Code_E,
                                                 center = FALSE, scale = TRUE)
final_merged_df = within(final_merged_df, BodyCamUsageCat <-
                           relevel(BodyCamUsageCat, ref = "0-40%"))
final_merged_df = within(final_merged_df, PoliceSpendingCat <-
                           relevel(PoliceSpendingCat, ref = "0-40%"))
pander(table(final_merged_df$Year), caption = "Number of Variables by Year")

p1 = ggplot(final_merged_df, aes(y = NormalizedDeaths, x = NormalizedMiliSurSpending)) +
  geom_point(alpha = .5, size = .5, colour="blue4") +
  geom_smooth(method = lm, color ="red3") +
  ggtitle("Normalized Deaths vs. \n Normalized Spending") +
  theme(text = element_text(size = 8), plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(color="black", size=6, angle=0),
        axis.text.y = element_text(color="black", size=6, angle=45),
        plot.margin=grid::unit(c(0,0,0,0), "mm"))

p2 = ggplot(final_merged_df, aes(y = TotalDeathCount, x = TotalAcquisitionValueSum)) +
  geom_point(alpha = .5, size = .5, colour="blue4") +
  geom_smooth(method = lm, color ="red3") +
  ggtitle("Total Death Counts Per State and \n Year vs. Total Mil Surplus \n Spending (in $)") +
  theme(text = element_text(size = 8), plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(color="black", size=6, angle=0),
        axis.text.y = element_text(color="black", size=6, angle=45),
        plot.margin=grid::unit(c(0,0,0,0), "mm"))

p3 = ggplot(final_merged_df[(final_merged_df$DEMIL_Code_D!=max(final_merged_df$DEMIL_Code_D)), ], aes(y
  geom_point(alpha = .5, size = .5, colour="blue4") +
  geom_smooth(method = lm, color ="red3") +
  ggtitle("Total Death Counts Per State and \n Amount of Weapons Bought") +
  theme(text = element_text(size = 8), plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(color="black", size=6, angle=0),
        axis.text.y = element_text(color="black", size=6, angle=45))

p4 = ggplot(final_merged_df, aes(y = PoliceTrooperDeath, x = TotalAcquisitionValueSum)) +
  geom_point(alpha = .5, size = .5, colour="blue4") +
  geom_smooth(method = lm, color ="red3") +
  ggtitle("Deaths Caused by Police/Troopers vs \n Amount of Military Surplus Aquired (in $)")+
  theme(text = element_text(size = 8),
        axis.text.x = element_text(color="black", size=6, angle=0),
        axis.text.y = element_text(color="black", size=6, angle=45),)

p5 = ggplot(final_merged_df, aes(x = State, y = NormalizedDeaths ,fill=State)) +
  geom_boxplot() +
  ggtitle("Normaized Deaths by State") +
  theme(legend.position="none",
        text = element_text(size = 8), plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(color="black", size=6, angle=45),
        axis.text.y = element_text(color="black", size=6, angle=0))
```

```
lay = rbind(c(1, 2),
            c(3, 4),
            c(5, 5))

grid.arrange(p1, p2, p3, p4, p5, layout_matrix= lay, bottom = "Figure 2: EDA Plots")
knitr::include_graphics("C:/Users/Chris/Desktop/MIDS/Fall 2020/IDS 702 Modeling and Rep of Data/FinalPro
knitr::include_graphics("C:/Users/Chris/Desktop/MIDS/Fall 2020/IDS 702 Modeling and Rep of Data/FinalPro
```