Title: Data Challenge #2

Authors: Brandon Wolff, Zachary Heinemann, and Stephanie Langeland

Due date: 3/22/2017

Review/load the data:

```
rm(list = ls(all = TRUE))   # cleans everything in the workspace

library(readr)          # easier reading of flat files
library(caret)          # classification and regresssion training package
```

```
## Loading required package: lattice
```

```
## Loading required package: ggplot2
```

```
# ::::::: SOME USEFUL DEFINITIONS ::::::::::::::::::::::::::::::::::::::::::::

# set the general path for the project at its root, specific files will define
# their own branches individually
# NOTE that specifying your path will be different in Windows

path <- "/Users/StephanieLangeland/Desktop/Columbia/Applied Data Science/Git/QMSS_G5069_
Applied_D_S/Data Challenges/datachallenge2" ## Stephanie's path

# path <- "C:\\Users\\Brandon\\Documents\\GitHub\\QMSS_G5069_Applied_D_S\\Data Challenge
s\\datachallenge2" ## Brandon's path

# path <- "/Users/zachheinemann/GitHub/QMSS_G5069_Applied_D_S/Data Challenges/datachalle
nge2" ## Zachary's Path

# define additional paths for files you will use. In each case, determine
# appropriate additions to the path

inFileName1   <- "data/processed/AllViolenceData_170216.csv"     # cleaned data on viole
nce
outFileName1  <- "graphs/RF_VarImportance.pdf"
outFileName2  <- "graphs/RF_MSE.pdf"

# ::::::: APPLY INITIAL DEFINITIONS ::::::::::::::::::::::::::::::::::::::::::

# set your path to that defined above, and confirm it
setwd(path)
getwd()
```

```
## [1] "/Users/StephanieLangeland/Desktop/Columbia/Applied Data Science/Git/QMSS_G5069_A
pplied_D_S/Data Challenges/datachallenge2"
```

```
# ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
# :::::::::::::::::::: LOADS DATA ::::::::::::::::::::::::::::::::::::::::::::


# ::::::: LOADING RAW DATA
# note that read_csv() guesses column types, so that date is read as a date
# very useful for plotting time series

AllData <- read_csv(inFileName1)
```

```
## Parsed with column specification:
## cols(
##   .default = col_double(),
##   date = col_date(format = ""),
##   state = col_character(),
##   state.abbr = col_character(),
##   municipality = col_character(),
##   source = col_character(),
##   category = col_character(),
##   global.id = col_integer()
## )
```

```
## See spec(...) for full column specifications.
```

```
# rough validations that data was correctly loaded
## names(AllData)
## nrow(AllData)
## summary(AllData)

## The authories involved in each event are noted in their own columns in the
## dataset.  For example when `AllData$army == 1` the army was involved.
## Guide to authority lookup:
auth_lookup <- "data/external/LookupAuthorityNames.csv"
auth_lookup_data <- read.csv(auth_lookup)
```

This is a team assignment. Please create a file on your team GitHub repo where you answer the challenge, including links to your code, and graphs.

Model #1 for question #1:

1.) Ask two (2) questions that might help you understand better the dynamics of violence contained on our dataset. Apply one algorithm per question and share your insights from each analysis. [50 pts] Remember: a non-finding is also a finding! It tells you whether a question is worth pursuing further or not.

- Which type of regression model (linear, quadratic, cubic, quartic) best predicts whether or not people will be detained using all weapons (defined as `clips.seized`, `cartridge.sezied`, `small.arms.seized`, and `long.guns.seized`) seized and wounded organized crime members as predictors? Will this model help the Mexican government formulate a strategy when the goal is to maximize the number of people detained?

1a) perform the necessary transformations in your data - if any are needed, and explain why you did that

- The variables `clips.seized`, `cartridge.sezied`, `small.arms.seized`, and `long.guns.seized` will be summed into one variable named `weapons_seized`. This new variable will allow us to analyze the total number of weapons seized. We want to understand how seizing any type of weapon relates to whether people are detained rather than the number of each type of weapon seized. We are including `clips.seized` because clips are part of weapons.
- A new variable will be created named `detained_resp` to record whether people where detained (`1`) or not (`0`). We want to predict the response of whether people will be detained or not rather than the number of people detained.

```
AllData$weapons_seized <- transform(AllData$clips.seized +
                                    AllData$cartridge.sezied +
                                    AllData$small.arms.seized +
                                    AllData$long.guns.seized)

typeof(AllData$weapons_seized)
```

```
## [1] "list"
```

```
AllData$weapons_seized <- as.numeric(unlist((AllData$weapons_seized))) # convert
# from list to numeric for regression models

AllData$detained_resp <- ifelse(AllData$detained > 0, 1, 0)

typeof(AllData)
```

```
## [1] "list"
```

```
AllData <- as.data.frame(AllData) # convert from list to data.frame
```

1b) show the output from your analysis in a consumable form

- **Regression Models:**

```
set.seed(1234)  # for reproducibility
train <- sample(nrow(AllData), 2698) # randomly sample half of the data
training <- AllData[train, ] # create training set
testing <- AllData[-train, ] # create testing set

linear <- lm(detained_resp ~ weapons_seized + organized.crime.wounded,
             data = training) # linear regression

quadratic <- lm(detained_resp ~ poly(weapons_seized + organized.crime.wounded, degree =
2),
                data = training) # quadratic regression

cubic <- lm(detained_resp ~ poly(weapons_seized + organized.crime.wounded, degree = 3),
            data = training) # cubic regression

quartic <- lm(detained_resp ~ poly(weapons_seized + organized.crime.wounded, degree =
4),
              data = training) # quartic regression

p_linear <- predict(linear, newdata = testing, type = "response") # predict the repsonse

p_liner_table <- confusionMatrix(testing$detained_resp,
                                 as.integer(p_linear > 0.5)) # confusion matrix

p_quadratic <- predict(quadratic, newdata = testing, type = "response")
p_quadratic_table <- confusionMatrix(testing$detained_resp, as.integer(p_quadratic >
0.5))

p_cubic <- predict(cubic, newdata = testing, type = "response")
p_cubic_table <- confusionMatrix(testing$detained_resp, as.integer(p_cubic > 0.5))

p_quartic <- predict(quartic, newdata = testing, type = "response")
p_quartic_table <- confusionMatrix(testing$detained_resp, as.integer(p_quartic > 0.5))
```

- **Regression Results:** A confusion matrix was constructed for each of the models and ranked in order from worst to best prediction accuracy:

```
p_quadratic_table # quadratic regression confusion matrix
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 1889   22
##          1  760   27
##
##                Accuracy : 0.7102
##                  95% CI : (0.6926, 0.7272)
##     No Information Rate : 0.9818
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.0315
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.71310
##             Specificity : 0.55102
##          Pos Pred Value : 0.98849
##          Neg Pred Value : 0.03431
##              Prevalence : 0.98184
##          Detection Rate : 0.70015
##    Detection Prevalence : 0.70830
##       Balanced Accuracy : 0.63206
##
##        'Positive' Class : 0
##
```

```
p_liner_table # linear regression confusion matrix
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 1880   31
##          1  748   39
##
##                Accuracy : 0.7113
##                  95% CI : (0.6938, 0.7283)
##     No Information Rate : 0.9741
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.0455
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.71537
##             Specificity : 0.55714
##          Pos Pred Value : 0.98378
##          Neg Pred Value : 0.04956
##              Prevalence : 0.97405
##          Detection Rate : 0.69681
##    Detection Prevalence : 0.70830
##       Balanced Accuracy : 0.63626
##
##        'Positive' Class : 0
##
```

```
p_quartic_table # quartic regression confusion matrix
```

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 1850   61
##          1  713   74
##
##                Accuracy : 0.7131
##                  95% CI : (0.6956, 0.7301)
##     No Information Rate : 0.95
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.0821
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.72181
##             Specificity : 0.54815
##          Pos Pred Value : 0.96808
##          Neg Pred Value : 0.09403
##              Prevalence : 0.94996
##          Detection Rate : 0.68569
##    Detection Prevalence : 0.70830
##       Balanced Accuracy : 0.63498
##
##        'Positive' Class : 0
##
```

```
p_cubic_table # cubic regression confusion matrix
```

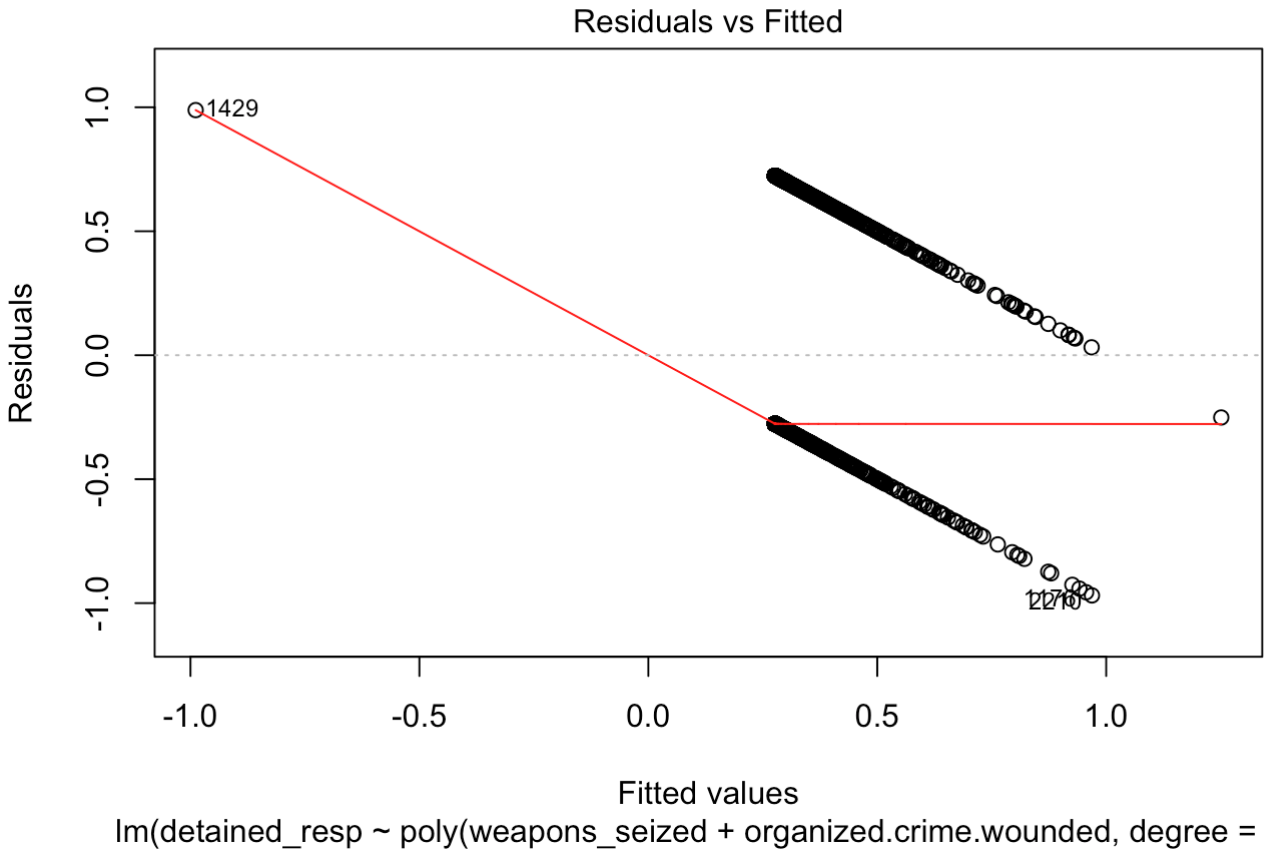
```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction    0    1
##          0 1875   36
##          1  737   50
##
##                Accuracy : 0.7135
##                  95% CI : (0.696, 0.7305)
##     No Information Rate : 0.9681
##     P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.0606
##  Mcnemar's Test P-Value : <2e-16
##
##             Sensitivity : 0.71784
##             Specificity : 0.58140
##          Pos Pred Value : 0.98116
##          Neg Pred Value : 0.06353
##              Prevalence : 0.96812
##          Detection Rate : 0.69496
##    Detection Prevalence : 0.70830
##       Balanced Accuracy : 0.64962
##
##        'Positive' Class : 0
##
```

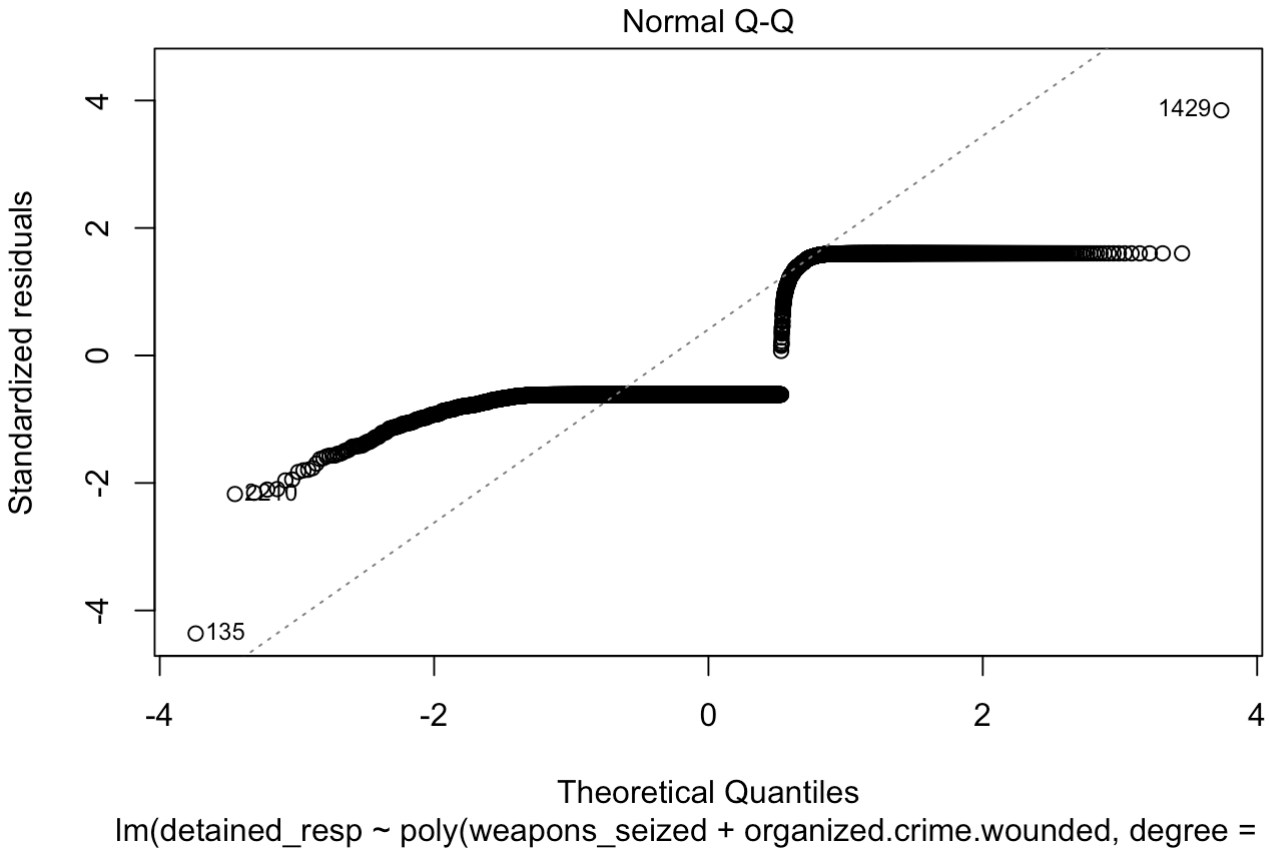
- **Analysis of the output:** The cubic regression model had the highest prediction accuracy of 0.7135 or 71.35% in using wounded organized crime members and all weapons seized to predict whether people would be detained or not. The cubic model's confusion matrix shows that the model correctly predicted that no one would be detained 1,875 times and people would be detained 50 times. Although the model has the highest prediction accuracy, it is mainly driven by correct predictions of no one being detained. Overall, it does not successfully predict when people will be detained. Therefore, this model will not help Mexican authorities understand drivers of detention rates, since it does not consistently predict how the number of organized crime members wounded and weapons seized relate to whether people are detained or not. Furhermore, the model incorrectly predicted that people would be detained in 737 events when no one was detained. It also incorrectly predicted that people would not be detained when people were detained in 36 events.

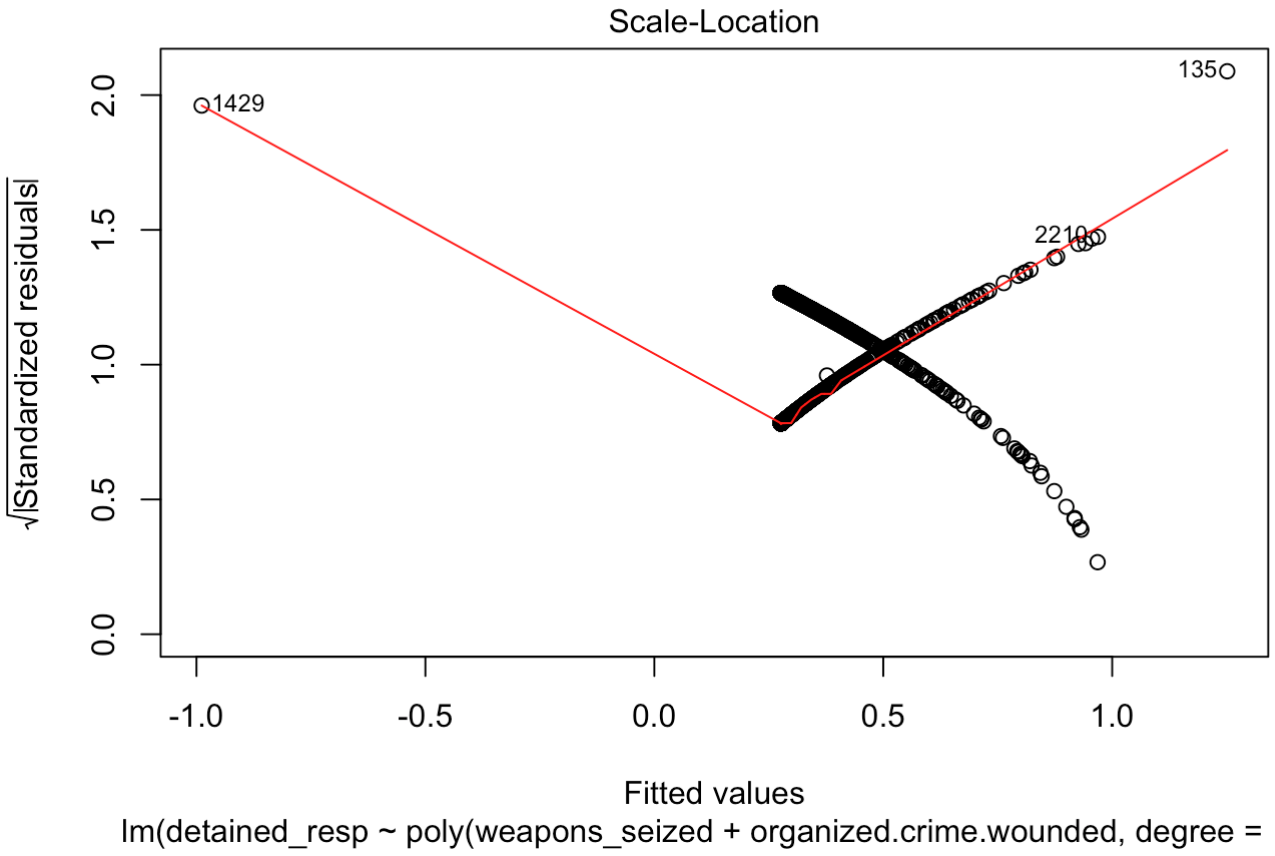
1c) be explicit about the limitations of your anaylisis, due to estimation or to the data itself

- The limitations of the cubic regression model are shown through various diagnostic plots using the entire dataset. Refer to the analysis below the plots.

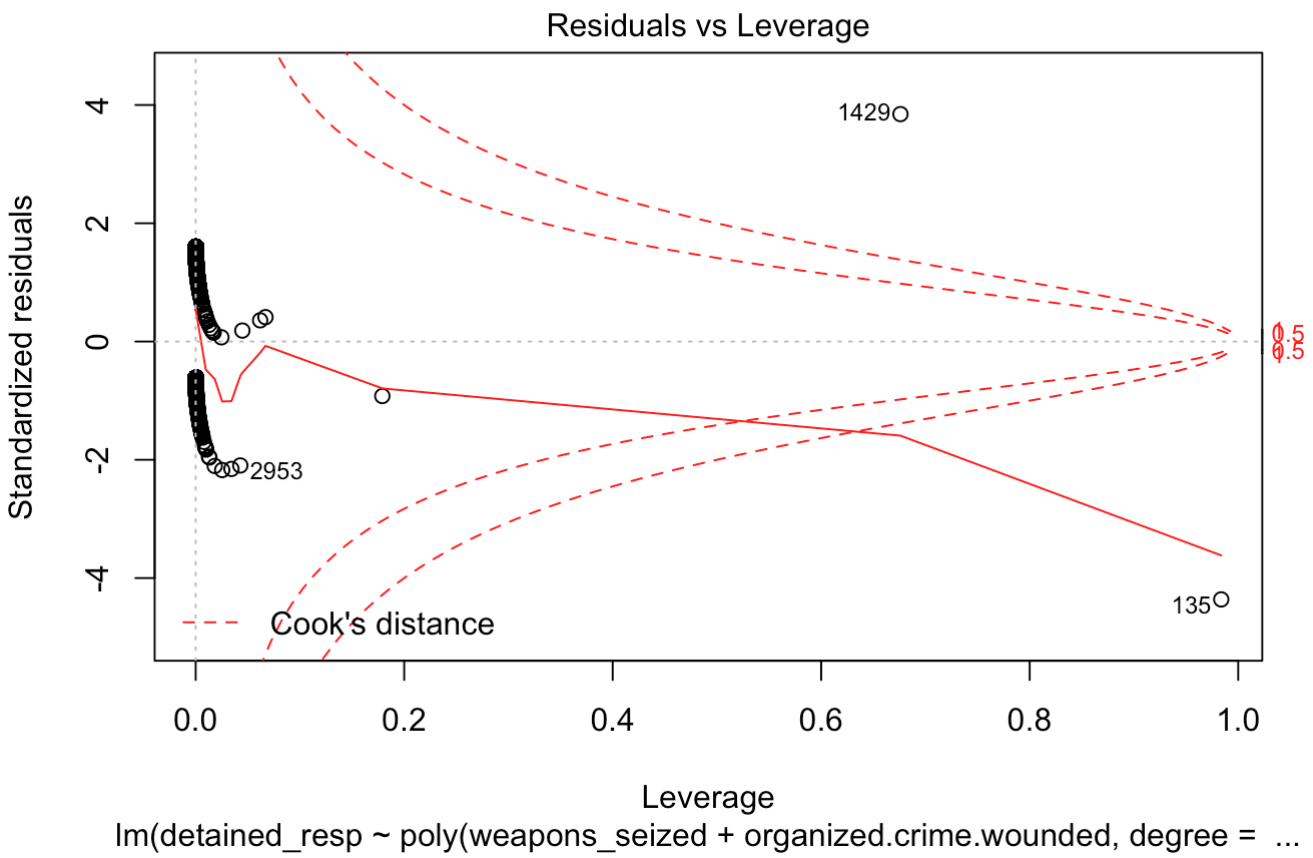
```
plot(lm(detained_resp ~ poly(weapons_seized + organized.crime.wounded, degree = 3),
        data = AllData)) # cubic regression
```


Residuals vs Fitted
Residuals
Fitted values
lm(detained_resp ~ poly(weapons_seized + organized.crime.wounded, degree = ...
1429
1.0
0.5
0.0
-0.5
-1.0

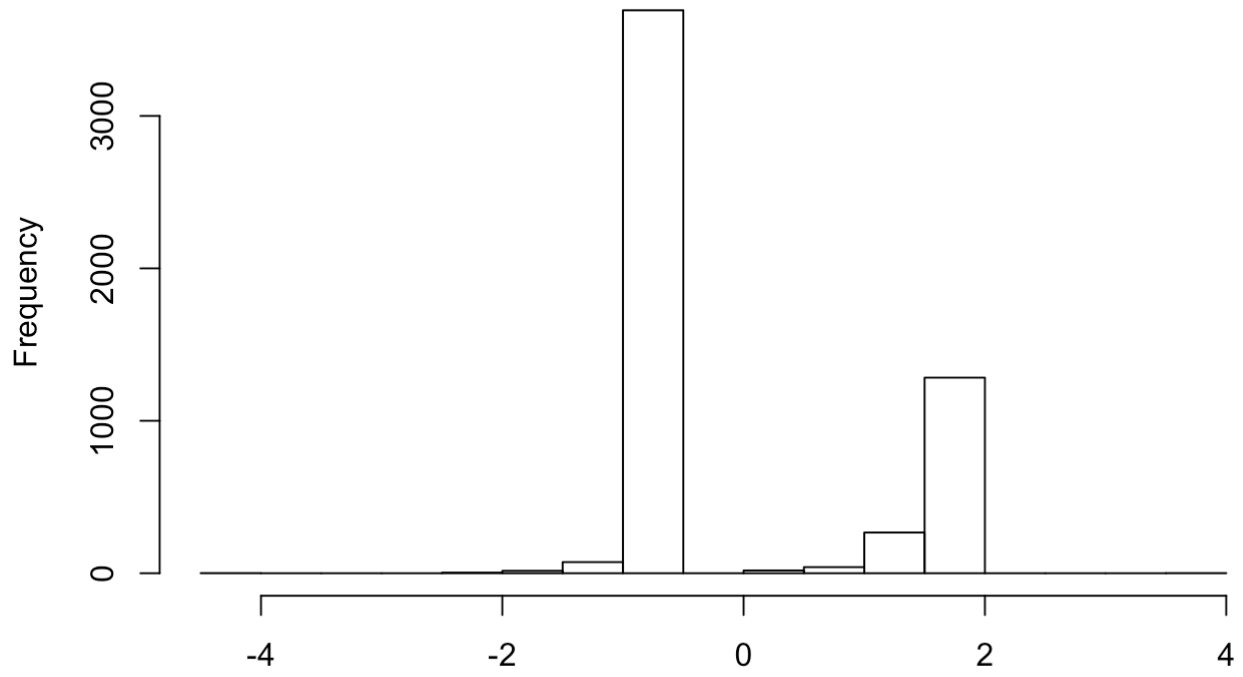

Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(detained_resp ~ poly(weapons_seized + organized.crime.wounded, degree = ...
1429
135
4
2
0
-2
-4




```
## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```




```
library(MASS)
hist(studres(lm(detained_resp ~ poly(weapons_seized + organized.crime.wounded, degree =
3),
                data = AllData)),
     main = "Distribution of Studentized Residuals") ## The Studentized
```



lres(lm(detained_resp ~ poly(weapons_seized + organized.crime.wounded, degree = 3), data =

```
## residuals, like standardized residuals, are normalized to unit
## variance, but the Studentized version is fitted ignoring the current
## data point.
```

- **Residuals vs Fitted plot:** This plot shows whether the residuals have non-linear patterns. Although the cubic model had the highest prediction accuracy of all the regression models,the residuals do not show a cubic relationship on this plot, but rather two linear lines that have a clear turning point. Therefore, the cubic nature of the model may not actually be the most appropriate fit for the data. This may help explain why there was unbalanced prediction accuracy in that more “not detained” cases were predicted than “detained” cases.
- **Normal Q-Q plot:** This plot shows whether the residuals are normally distributed.
  Since the residuals on the graph do not follow the straight dashed line, the data are not normally distributed. This may represent bias or skewness in the data, which may be attributed to unbalanced cases of detained versus not detained. If this is the case, this may support the claim that there are more cases of people not being detained. To understand the distribution of the data, the histrogram of the `Distribution of Studentized Residuals` shows that the data are heavily skewed towards events when no one was detained.
- **Scale-Location plot:** This plot shows whether the residuals are spread equally along the ranges of predictors. Since the predictors are not equally spread throughout the graph, this suggests that the data suffer from heterscedasticity, meaning that variance among the residuals is not equal. This seriously undermines the validity of the model because the modeling errors may not be uncorrelated and uniform.

- **Residuals vs Leverage plot:** This plot illumiates outliers that may be influential to regresson results. The extreme outliers # 135 and 1429 are outside of the Cook's distance (represented by the dotted line). These cases are influential to the regression results, which would be altered if these cases are excluded from the model.
- **Conclusion:** Given the analysis of the diagnostic plots above, the cubic model has serious limitations that violate the basic regression assumptions. These limitations may invalidate the regression results. The severity of these limitations may become less influential as more data are collected. As more data are collected, theoretically, the distribution should normalize and become homoscedastic.

1d) did you find something interesting? what is that? does your finding suggest this question is worth pursuing further? why or why not?

- It is interesting that the cubic model predicts the best out of all the models but the predicitons are skewed towards predicting when no one will be detained. This skewness in the prediction may be attributed to the findings noted in #1c above related to the violations of regression assumptions, especially that the data are not normally distributed. Predicting whether or not people will be detained using all weapons seized and wounded organized crime members as predictors is not worth pursuing because the model poorly predicts cases when people will be detained. Moreover, since the model is not in compliance with the regression assumptions, the results may not be true findings.

1e) if you did not find something interesting, explain why, and whether there is some additional information that would help in answering your question

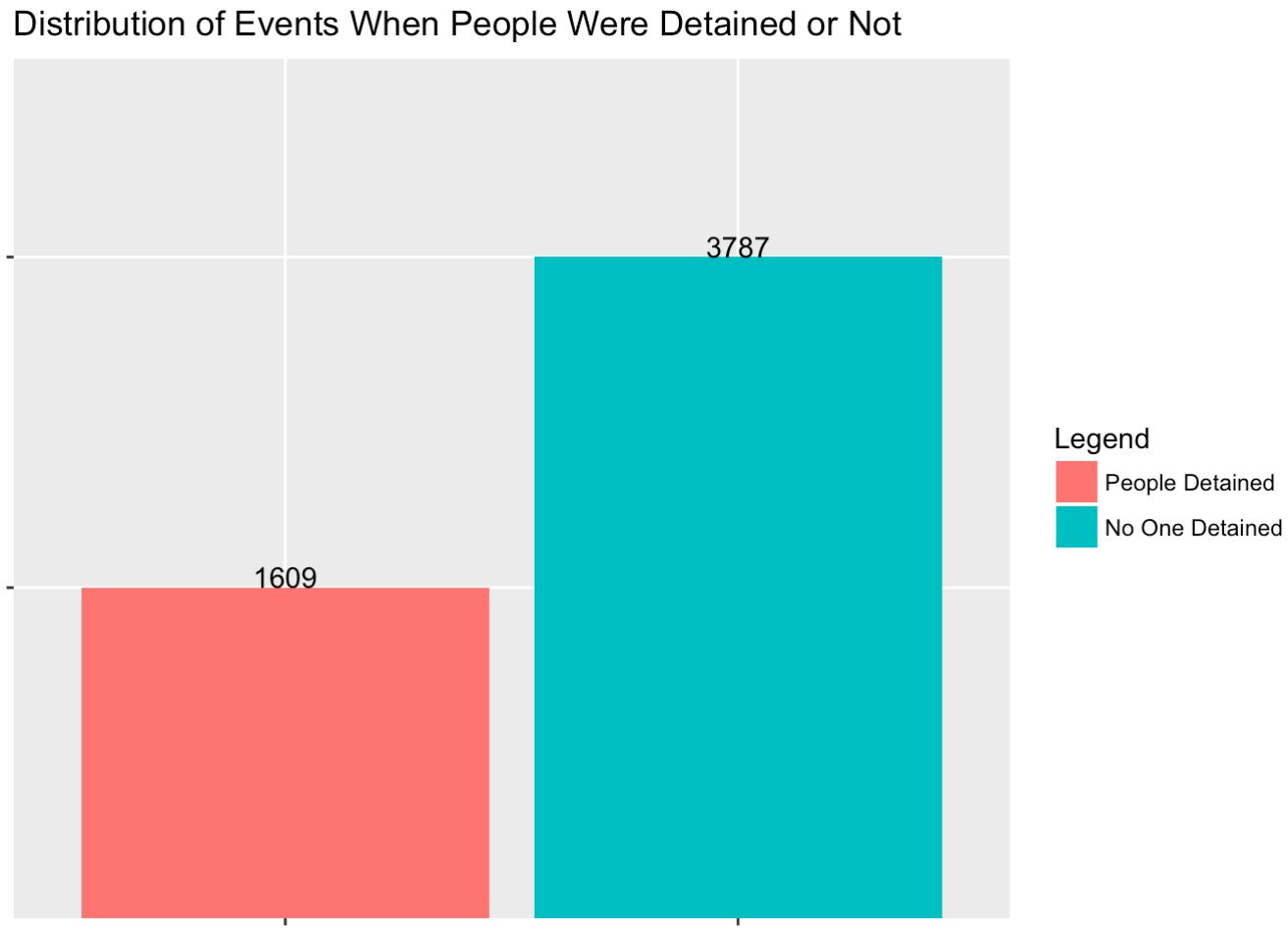
```
nrow(AllData[AllData$detained_resp == "0",]) # Number of events when no one was detained
```

```
## [1] 3787
```

```
nrow(AllData[AllData$detained_resp == "1",]) # Number of events when people were detaine
d
```

```
## [1] 1609
```

```
data_dist <- matrix(c("3787",
                      "1609",
                      "No One Detained",
                      "Detained"),
                    ncol = 2,
                    nrow = 2)

data_dist <- as.data.frame(data_dist)

library(plyr)
data_dist <- rename(data_dist, c("V1" = "totals", "V2" = "type"))

library(ggplot2)

ggplot(data = data_dist, aes(x = totals, y = type, fill = totals)) +
  geom_bar(stat = "identity") +
  ggtitle("Distribution of Events When People Were Detained or Not") +
  scale_fill_discrete(name = "Legend",
                      breaks = c("1609", "3787"),
                      labels = c("People Detained", "No One Detained")) +
  geom_text(aes(label = totals, vjust = 0)) +
  theme(
    axis.text.x = element_blank(),
    axis.text = element_blank(),
    axis.title.x = element_blank(),
    axis.title.y = element_blank()
  )
```



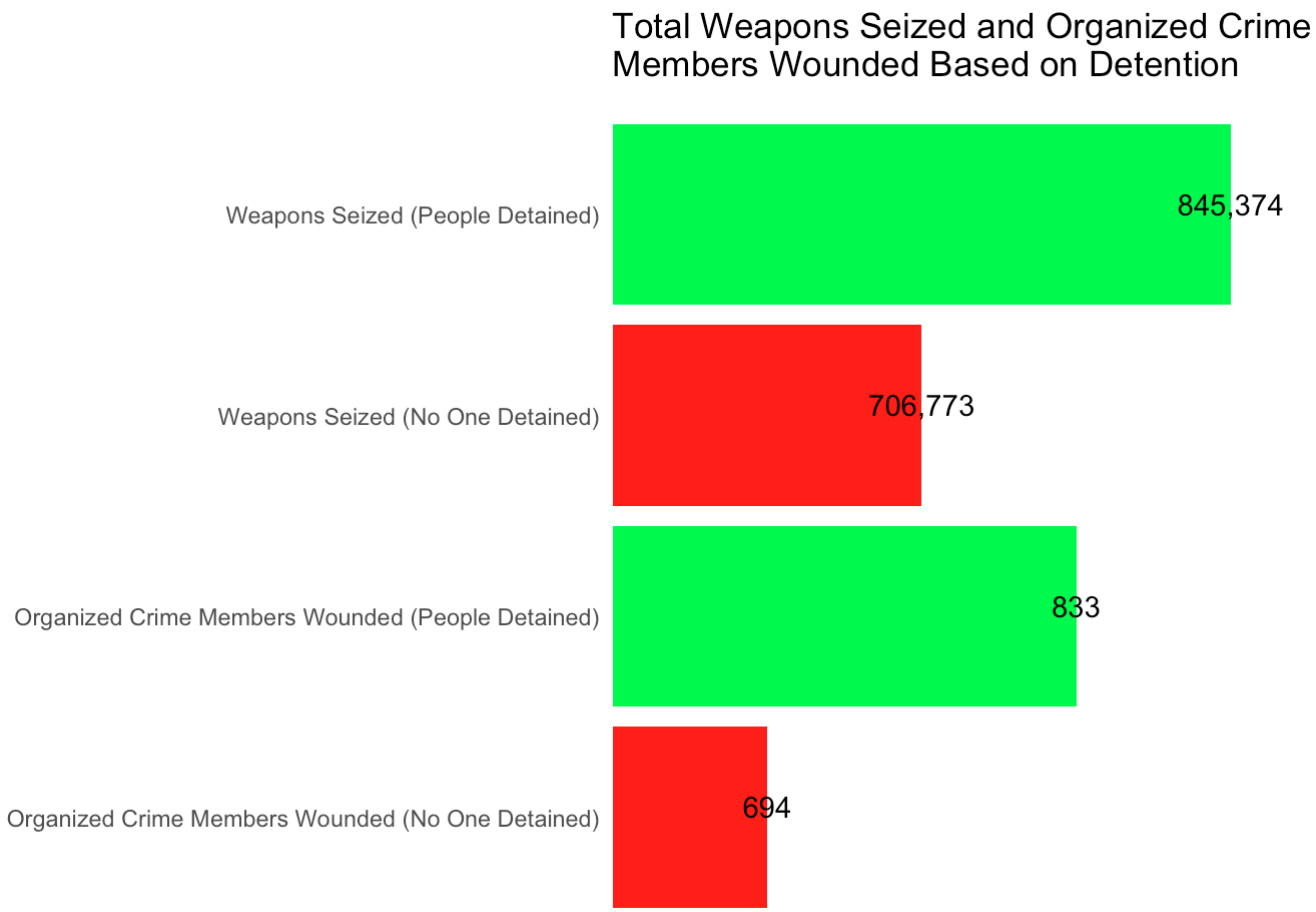
- The poor prediction accuracy of the cubic model in predicting when people will be detained may be attributed to the skewness of the data. Above, we see that there were 3,787 events when no one was detained and 1,609 when people were detained. Therefore, the model may not have enough data to better predict when people will be detained. The model may be more effective once more data are collected and the data become normally distributed.

1f) provide your code, and a single visualization per question that summarizes your finding

- To contextualize the usefulness of the prediction model, the following visualization shows the total number of weapons seized when people are detained versus not detained and the total number of organized crime members wounded when people are detained versus not detained. This graph should help the Mexican government decide whether it is worth pursuing a strategy centered around detaining more people if the goal is to maximize the number of weapons seized and organized crime members wounded.

```
# Create a smaller dataset to build the visualization:

# Subset of `weapons_seized` and `organized.crime.wounded` when peoeple were detained:
detained_subset <- subset(AllData,
                          detained_resp == 1,
                          select = c(weapons_seized, organized.crime.wounded))

# Subset of `weapons_seized` and `organized.crime.wounded` when no one was detained:
not_detained_subset <- subset(AllData,
                              detained_resp == 0,
                              select = c(weapons_seized, organized.crime.wounded))

det_weap_total <- sum(detained_subset$weapons_seized) # total weapons seized
# when people were detained

n_det_weap_total <- sum(not_detained_subset$weapons_seized) # total weapons seized
# when people were NOT detained

det_ocw_total <- sum(detained_subset$organized.crime.wounded) # total number of
# organized crime members wounded when people were detained

n_det_ocw_total <- sum(not_detained_subset$organized.crime.wounded) # total number of
# organized crime members wounded when people were NOT detained

total_figures <- data.frame(det_weap_total,
                            n_det_weap_total,
                            det_ocw_total,
                            n_det_ocw_total) # convert to data.frame

library(plyr)
# rename column headings:
total_figures <- rename(total_figures,
                        c("det_weap_total" = "Weapons Seized (People Detained)",
                          "n_det_weap_total" = "Weapons Seized (No One Detained)",
                          "det_ocw_total" =
                            "Organized Crime Members Wounded (People Detained)",
                          "n_det_ocw_total" =
                            "Organized Crime Members Wounded (No One Detained)"))

# Have numbers appear with commas:
total_figures$`Weapons Seized (People Detained)` <- prettyNum(
  total_figures$`Weapons Seized (People Detained)`,
  big.mark = ",",
  scientific = FALSE)

total_figures$`Weapons Seized (No One Detained)` <- prettyNum(
  total_figures$`Weapons Seized (No One Detained)`,
  big.mark = ",",
  scientific = FALSE)

total_figures$`Organized Crime Members Wounded (People Detained)` <- prettyNum(
  total_figures$`Organized Crime Members Wounded (People Detained)`,
  big.mark = ",",
```

```
  scientific = FALSE)

total_figures$`Organized Crime Members Wounded (No One Detained)` <- prettyNum(
  total_figures$`Organized Crime Members Wounded (No One Detained)`,
  big.mark = ",",
  scientific = FALSE)

total_figures_new <- t(total_figures) # transpose total_figures for graphing purposes
total_figures_new <- as.data.frame(total_figures_new) # convert to data.frame
total_figures_new$subject <- row.names(total_figures_new)
total_figures_new <- rename(total_figures_new,
                            c("V1" = "Totals"))

ggplot(data = total_figures_new, aes(y = Totals, x = subject, fill = subject)) +
  geom_bar(stat = "identity") +
  scale_fill_manual(values = c("red", "green", "red", "green")) +
  coord_flip() +
  ggtitle("Total Weapons Seized and Organized Crime\nMembers Wounded Based on
Detention") + # title line break
  geom_text(aes(label = Totals), vjust = 0, colour = "black") +
  theme(axis.line = element_blank(),
        axis.text.x = element_blank(),
        axis.ticks = element_blank(),
        axis.title.x = element_blank(),
        axis.title.y = element_blank(),
        legend.position = "none",
        panel.background = element_blank(),
        panel.border = element_blank(),
        panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(),
        plot.background = element_blank())
```



- The graph of all the data shows that when people are detained, more weapons are seized and organized crime members wounded. If the Mexican government's goal is to maximize the number of weapons seized and organized crime members wounded, they should pursue a strategy that focuses on detaining people in each event. Although the cubic model does not predict when people will be detained based on thee varaibles as well as it predicts when people will not be detained, this issue may be solved once more data are collected. The current dataset is heavily skewed towards cases when no one was detained, as seen above in question ***1e***. Once more data are colelcted, the distribution should normalize and the cubic model may have a more balanced success rate.

1g) phrase your finding for each question in two ways:

1g-1) one sentence that summarizes your insight

- A cubic model used the number of organized crime members wounded and weapons seized per event to successfully predict when people will not be detained but not when people will be detained with the same level of accuracy.

1g-2) one paragraph that reflects all nuance in your insight

- A cubic model used the number of organized crime members wounded and weapons seized per event to successfully predict when people will not be detained but not when people will be detained with the same level of accuracy. The distribution of the data is not normal and skewed towards more cases when people were not detained. Additionally, the data suffers from heteroscedasticity and extreme outliers that influence the regression results. Overall, the cubic model violates many regression assumptions and the ouput is not useful in predicting when people will be detained.

1h) make sure to also include your code

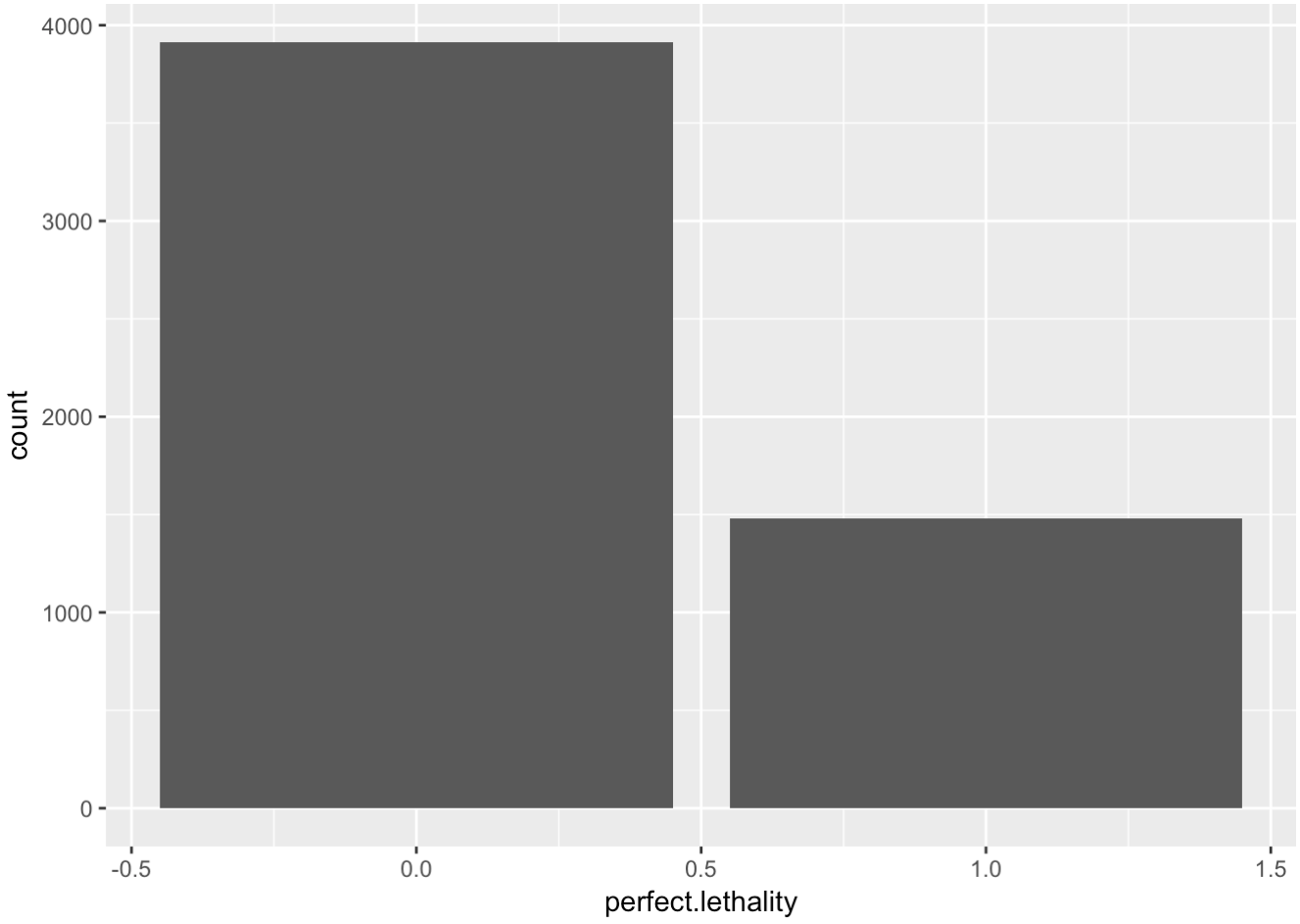
- The code is included under each quetsion above.

Model #2 for question #1:

1.) Ask two (2) questions that might help you understand better the dynamics of violence contained on our datas et. Apply one algorithm per question and share your insights from each analysis. [50 pts] Remember: a non-finding is also a finding! It tells you whether a question is worth pursuing further or not.

- Does the specific government group involved with an event (AFI, Army, Federal Police, Ministerial Police, Municipal Police, Navy, Other, and State Police) have a positive or negative relationship with Perfect Lethality? We can see from the graph below that there are about 1500 cases with perfect lethality therefore I believe it is acceptable to examine this question.

```
library(ggplot2)
pl <- ggplot(AllData, aes(perfect.lethality))
pl + geom_bar()
```



1a) perform the necessary transformations in your data - if any are needed, and explain why you did that

- We used the same code used above to create a variable for total weaponry seized (`weapons_seized`) by using the sum of the variables `clips.seized` , `cartridge.sezied` , `small.arms.seized` , `long.guns.seized` . This variable will be used in order to control for all weaponary seized in our analysis.

```
# used Stephanie's code to create one variable to sum weaponry seized
AllData$weapons_seized <- transform(AllData$clips.seized +
                                      AllData$cartridge.sezied +
                                      AllData$small.arms.seized +
                                      AllData$long.guns.seized)
AllData$weapons_seized <- as.numeric(unlist((AllData$weapons_seized)))
```

1b) show the output from your analysis in a consumable form

- In order to answer our hypothesis we ran a multiple logistic regression on perfect lethality by the involvment of the AFI, Army, Federal Police, Ministerial Police, Municipal Police, Navy, Other, and State Police. A number of control variables were also included which are municipality code, number of detained in the events, the date of each event, the total people dead in the events, the total people wounded in the events, weaponry seized in the events, and the source of the data for the events (Confrontations or Aggresions database). We are using logistic regression becuase the variables being used are binary (0, 1).

```
logit.perf_leth3 <- glm(perfect.lethality ~ afi + army + federal.police + ministerial.po
lice + municipal.police + navy + other + state.police + mun_code + detained + date + tot
al.people.dead + total.people.wounded + weapons_seized + source, family = binomial (link
 = logit), data = AllData)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

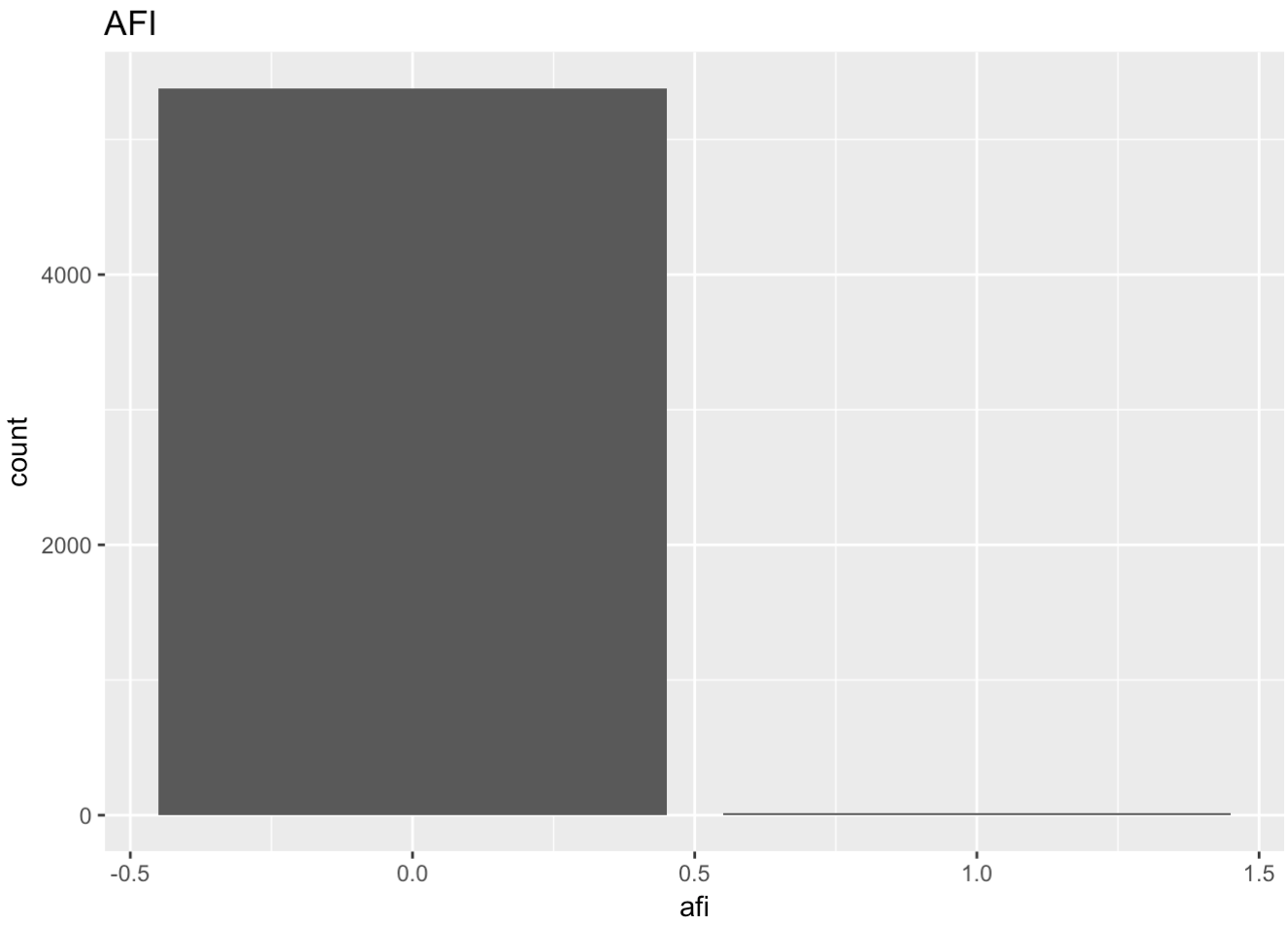
```
summary(logit.perf_leth3)
```

```
##
## Call:
## glm(formula = perfect.lethality ~ afi + army + federal.police +
##     ministerial.police + municipal.police + navy + other + state.police +
##     mun_code + detained + date + total.people.dead + total.people.wounded +
##     weapons_seized + source, family = binomial(link = logit),
##     data = AllData)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -8.4904  -0.6525  -0.2044   0.3100   4.3585
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -9.104e+00  1.545e+00  -5.891 3.83e-09 ***
## afi                   6.573e-02  7.191e-01   0.091  0.92716
## army                  6.110e-01  9.737e-02   6.275 3.49e-10 ***
## federal.police        2.116e-01  1.378e-01   1.535  0.12472
## ministerial.police   -5.084e-02  1.782e-01  -0.285  0.77537
## municipal.police     -6.147e-02  1.356e-01  -0.453  0.65024
## navy                  5.555e-01  2.216e-01   2.507  0.01218 *
## other                -1.486e-02  2.585e-01  -0.057  0.95416
## state.police         -2.491e-02  1.698e-01  -0.147  0.88338
## mun_code              2.358e-03  9.525e-04   2.476  0.01329 *
## detained             -3.775e-02  1.691e-02  -2.232  0.02564 *
## date                  3.269e-04  1.029e-04   3.176  0.00149 **
## total.people.dead     8.258e-01  3.202e-02  25.793  < 2e-16 ***
## total.people.wounded -6.113e-01  3.782e-02 -16.162  < 2e-16 ***
## weapons_seized        1.018e-05  1.700e-05   0.599  0.54918
## sourceconfrontations  2.784e+00  1.852e-01  15.035  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6341.9  on 5395  degrees of freedom
## Residual deviance: 4052.7  on 5380  degrees of freedom
## AIC: 4084.7
##
## Number of Fisher Scoring iterations: 7
```

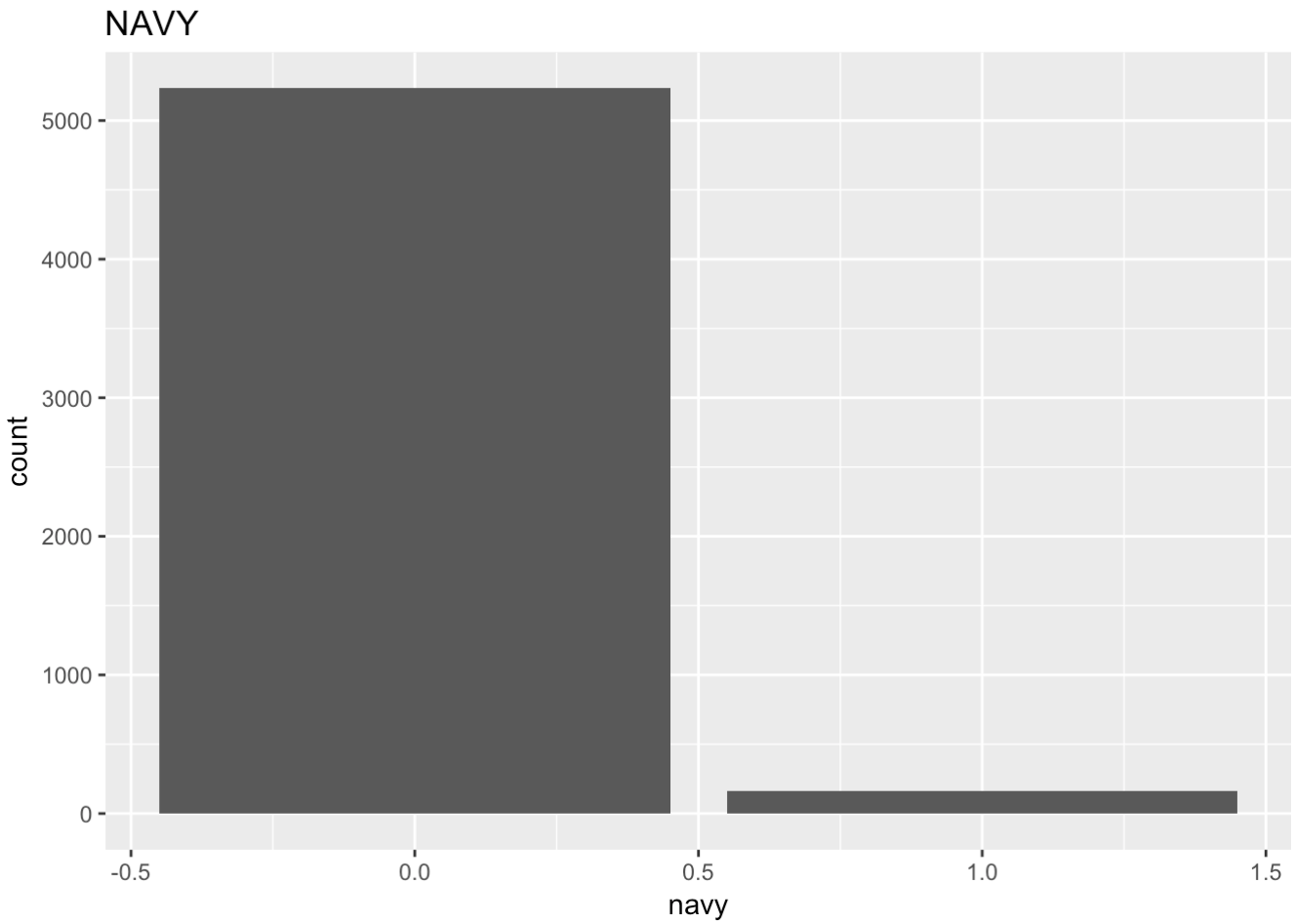
1c be explicit about the limitations of your anaylisis, due to estimation or to the data itself

- There are a number of limitations with the anaylisis. One limitation is that not all government groups were involved in a large sum of events. For example, the Navy was only involved in 89 cases and the AFI was only involved in 13 cases. The graphs below display the cases in which the navy and AFI were involved. Having such a small sample can be problematic.

```
afi <- ggplot(AllData, aes(afi))
afi + geom_bar() +
      ggtitle("AFI")
```




```
n <- ggplot(AllData, aes(navy))
n + geom_bar() +
    ggtitle("NAVY")
```



- It should also be noted that the 'source' of the data was highly significant in the model and if that variable is not included Ministerial Police, Municipal Police, and State Police involvment all become significant. I beleive this is problematic becuase the two datasets might not be collected in the same fashion. This model with and without the variable 'source' can be viewed below. It is also a limitation that logistic models are more difficult to explain/interpret in comparison to simple linear regression models.

```
logit.perf_leth2 <- glm(perfect.lethality ~ afi + army + federal.police +
                          ministerial.police + municipal.police + navy +
                          other + state.police + mun_code + detained +
                          date + total.people.dead + total.people.wounded +
                          weapons_seized, family = binomial (link = logit),
                        data = AllData)
```

```
## Warning: glm.fit: fitted probabilities numerically 0 or 1 occurred
```

```
summary(logit.perf_leth2)
```

```
##
## Call:
## glm(formula = perfect.lethality ~ afi + army + federal.police +
##     ministerial.police + municipal.police + navy + other + state.police +
##     mun_code + detained + date + total.people.dead + total.people.wounded +
##     weapons_seized, family = binomial(link = logit), data = AllData)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -8.4904  -0.6247  -0.3738   0.3400   4.1966
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -4.546e+00  1.491e+00  -3.049 0.002296 **
## afi                  -5.667e-03  7.505e-01  -0.008 0.993975
## army                  7.997e-01  1.013e-01   7.894 2.92e-15 ***
## federal.police        6.097e-02  1.369e-01   0.445 0.656092
## ministerial.police   -6.103e-01  1.722e-01  -3.543 0.000396 ***
## municipal.police     -9.595e-01  1.303e-01  -7.363 1.79e-13 ***
## navy                  6.522e-01  2.152e-01   3.030 0.002445 **
## other                -2.982e-01  2.358e-01  -1.265 0.206013
## state.police         -4.357e-01  1.638e-01  -2.661 0.007796 **
## mun_code              2.111e-03  8.955e-04   2.357 0.018427 *
## detained             -8.242e-04  1.644e-02  -0.050 0.960015
## date                  1.855e-04  1.005e-04   1.845 0.065080 .
## total.people.dead     8.491e-01  3.084e-02  27.530  < 2e-16 ***
## total.people.wounded -5.772e-01  3.695e-02 -15.622  < 2e-16 ***
## weapons_seized        1.015e-05  1.682e-05   0.604 0.545928
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6341.9  on 5395  degrees of freedom
## Residual deviance: 4418.9  on 5381  degrees of freedom
## AIC: 4448.9
##
## Number of Fisher Scoring iterations: 6
```

```
summary(logit.perf_leth3)
```

```
##
## Call:
## glm(formula = perfect.lethality ~ afi + army + federal.police +
##     ministerial.police + municipal.police + navy + other + state.police +
##     mun_code + detained + date + total.people.dead + total.people.wounded +
##     weapons_seized + source, family = binomial(link = logit),
##     data = AllData)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -8.4904  -0.6525  -0.2044   0.3100   4.3585
##
## Coefficients:
##                          Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -9.104e+00  1.545e+00  -5.891 3.83e-09 ***
## afi                   6.573e-02  7.191e-01   0.091  0.92716
## army                  6.110e-01  9.737e-02   6.275 3.49e-10 ***
## federal.police        2.116e-01  1.378e-01   1.535  0.12472
## ministerial.police   -5.084e-02  1.782e-01  -0.285  0.77537
## municipal.police     -6.147e-02  1.356e-01  -0.453  0.65024
## navy                  5.555e-01  2.216e-01   2.507  0.01218 *
## other                -1.486e-02  2.585e-01  -0.057  0.95416
## state.police         -2.491e-02  1.698e-01  -0.147  0.88338
## mun_code              2.358e-03  9.525e-04   2.476  0.01329 *
## detained             -3.775e-02  1.691e-02  -2.232  0.02564 *
## date                  3.269e-04  1.029e-04   3.176  0.00149 **
## total.people.dead     8.258e-01  3.202e-02  25.793  < 2e-16 ***
## total.people.wounded -6.113e-01  3.782e-02 -16.162  < 2e-16 ***
## weapons_seized        1.018e-05  1.700e-05   0.599  0.54918
## sourceconfrontations  2.784e+00  1.852e-01  15.035  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6341.9  on 5395  degrees of freedom
## Residual deviance: 4052.7  on 5380  degrees of freedom
## AIC: 4084.7
##
## Number of Fisher Scoring iterations: 7
```

```
# explain limitation with datasets and also explain how logit are hard to explain/interp
ret
```

1d) did you find something interesting? what is that? does your finding suggest this question is worth pursuing further? why or why not?

- We found that the navy and the army seem to a statistically significant and positive relationship to perfect lethality, controlling for all other variables included in the model. This means that if the Army or Navy is involved in an event the event has higher odds of being an event of perfect lethality. I feel this finding does suggest this question is worth pursuing further because it is important to know the effectivness of each individual group to know what groups to send into specific situations.

1e) if you did not find something interesting, explain why, and whether there is some additional information that would help in answering your question

- Not applicable.

1f) provide your code, and a single visualization per question that summarizes your finding

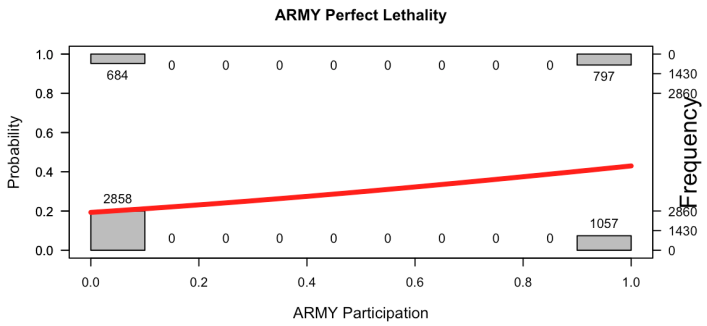
```
library(popbio)
```

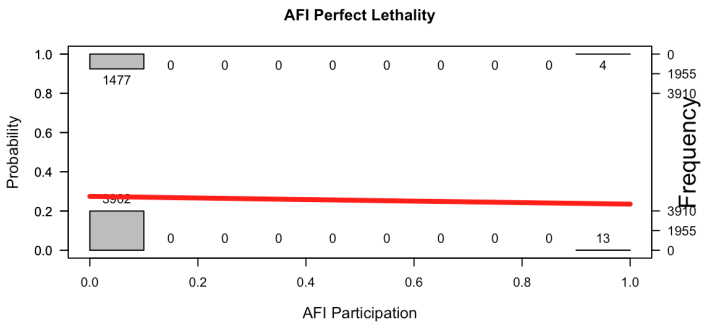
```
##
## Attaching package: 'popbio'
```

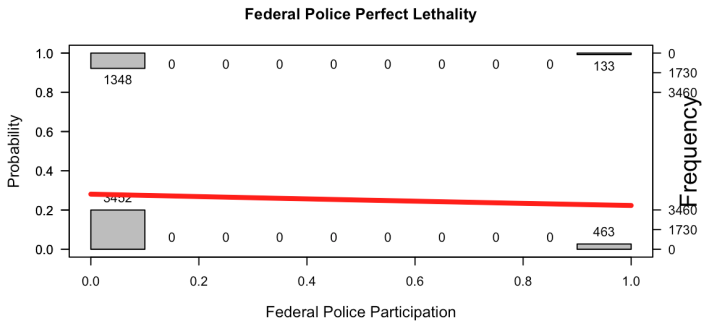
```
## The following object is masked from 'package:caret':
##
##     sensitivity
```

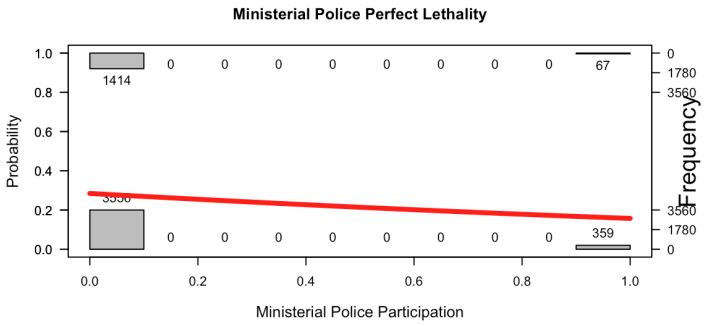
```
par(mfrow = c(4, 2))

logi.hist.plot(AllData$army,
               AllData$perfect.lethality,
               boxp = FALSE,
               rug = FALSE,
               logi.mod = 1,
               type = "hist",
               col = "gray",
               counts = TRUE,
               mainlabel = "ARMY Perfect Lethality",
               xlabel = "ARMY Participation")

logi.hist.plot(AllData$afi,
               AllData$perfect.lethality,
               boxp = FALSE,
               rug = FALSE,
               logi.mod = 1,
               type = "hist",
               col = "gray",
               counts = TRUE,
               mainlabel = "AFI Perfect Lethality",
               xlabel = "AFI Participation")

logi.hist.plot(AllData$federal.police,
               AllData$perfect.lethality,
               boxp = FALSE,
               rug = FALSE,
               logi.mod = 1,
               type = "hist",
               col="gray",
               counts = TRUE,
               mainlabel = "Federal Police Perfect Lethality",
               xlabel = "Federal Police Participation")

logi.hist.plot(AllData$ministerial.police,
               AllData$perfect.lethality,
               boxp = FALSE,
               rug = FALSE,
               logi.mod = 1,
               type = "hist",
               col = "gray",
               counts = TRUE,
               mainlabel = "Ministerial Police Perfect Lethality",
               xlabel = "Ministerial Police Participation")

logi.hist.plot(AllData$municipal.police,
               AllData$perfect.lethality,
               boxp = FALSE,
               rug = FALSE,
               logi.mod = 1,
               type = "hist",
               col = "gray",
```

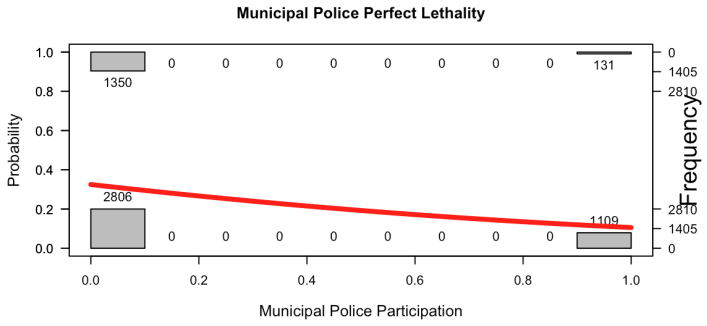
```
               counts = TRUE,
               mainlabel = "Municipal Police Perfect Lethality",
               xlabel = "Municipal Police Participation")

logi.hist.plot(AllData$navy,
               AllData$perfect.lethality,
               boxp = FALSE,
               rug = FALSE,
               logi.mod = 1,
               type = "hist",
               col = "gray",
               counts = TRUE,
               mainlabel = "NAVY Perfect Lethality",
               xlabel = "NAVY Participation")

logi.hist.plot(AllData$other,
               AllData$perfect.lethality,
               boxp = FALSE,
               rug = FALSE,
               logi.mod = 1,
               type = "hist",
               col = "gray",
               counts = TRUE,
               mainlabel ="OtherPerfect Lethality" ,
               xlabel = "Other Participation")

logi.hist.plot(AllData$state.police,
               AllData$perfect.lethality,
               boxp = FALSE,
               rug = FALSE,
               logi.mod = 1,
               type = "hist",
               col = "gray",
               counts = TRUE,
               mainlabel = "State Police Perfect Lethality",
               xlabel = "State Police Participation")
```

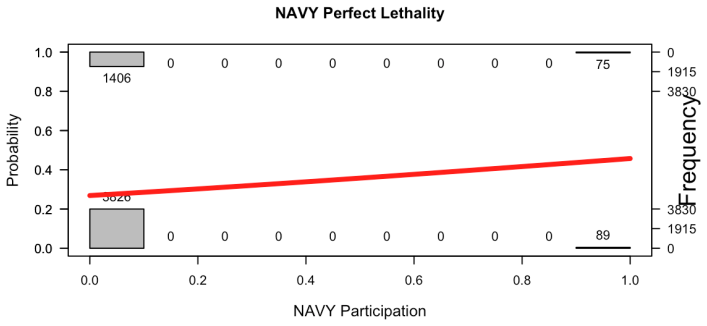


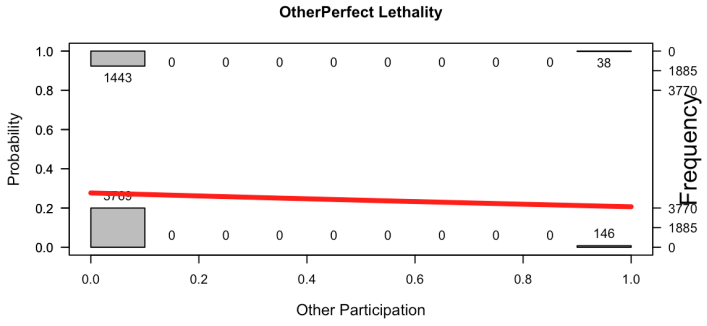


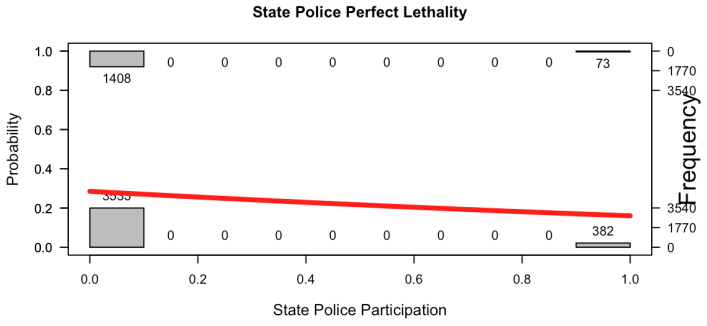












1g) phrase your finding for each question in two ways:

1g-1) one sentence that summarizes your insight

- The invovlment of the Navy or Army in an event on average increase the chances of said event being one of perfect lethality.

1g-2) one paragraph that reflects all nuance in your insight

- It is very interesting that the conttrol variables municipality code, number of detained in the events, the date of each event, the total people dead in the events, the total people wounded in the events, and the source of the data for the events (Confrontations or Aggresions database) all were statistically significant. This means that the municipality in which the event took place is related to whether or not the event was one of perfect lethality. This is a question that could be further examined. It is also worth noting that the navy was only included in 89 events yet 75 of the events were ones of perfect lethality. When seeing those numbers it

does seem clear that the navy is important when it come to perfect lethality, yet they are not involved in many cases whatsover. The Army on the other hand was involved in 1057 events and 797 of said events had perfect lethality. The army and the Navy seem to be well trained at killing in comparison to the AFI, Federal Police, Ministerial Police, Municipal Police, State Police, and other.

1h) make sure to also include your code

- The code is included under each quetsion above.

Model #1 for question #2:

2.) Formulate two (2) conditional hypotheses that you seek to investigate with the data. One of your hypotheses should condition on two variables (as the example on the slides), and the other should condition on three variables. [50 pts]

- ZACH

```
fit <- lm(total.people.dead ~ cartridge.sezied + army + federal.police + ministerial.pol
ice + municipal.police + navy + state.police + long.guns.seized * source + small.arms.se
ized + clips.seized + vehicles.seized, data = AllData)
```

2a) formulate each one of your hypotheses explicitly in substantive terms (as opposed to statistical terms) using 2-3 lines at most

- Our first hypothesis is such that there is a relationship between long guns seized and number of people dead such that as more long guns are seized (a proxy for severity of the incident as long guns are most lethal of the arms listed), there will be more people dead; further, becuase of the clear differences in datasets (seen in exploratory analysis), we expect that this relationship may differ as a function of the source of the dataset.

2b) show exactly how each one of your hypotheses translates into the marginal effect that you will seek to estimate from the data

- Given that the Aggressions dataset appears to have less lethality, we may expect that the relationship between long guns seized and people dead would be lower in this dataset as opposed to the confrontations dataset – therefore the interaction term would be positive and the marginal effect of ‘source’ being the confrontations (as opposed to aggression) would be higher than if we were to estimate the relationship without this conditional variable. Rather, we expect the positive relationship between long guns seized and people dead to be lower for the aggressions data, therefore making the marginal effect an effect lower than the effect for 1) the confrontation dataset and 2) the overall dataset aggregating between confrontations and aggresions.

2c) show the output from your analysis in a consumable form

```
summary(fit)
```

```
##
## Call:
## lm(formula = total.people.dead ~ cartridge.sezied + army + federal.police +
##     ministerial.police + municipal.police + navy + state.police +
##     long.guns.seized * source + small.arms.seized + clips.seized +
##     vehicles.seized, data = AllData)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -11.869  -1.012  -0.532   0.443  51.024
##
## Coefficients:
##                                       Estimate Std. Error t value
## (Intercept)                           9.756e-01  8.191e-02  11.910
## cartridge.sezied                     -1.820e-04  2.287e-05  -7.956
## army                                 -4.816e-01  7.595e-02  -6.341
## federal.police                       -4.942e-01  9.774e-02  -5.057
## ministerial.police                   -3.159e-01  1.132e-01  -2.790
## municipal.police                     -4.184e-01  8.471e-02  -4.939
## navy                                  7.700e-02  1.692e-01   0.455
## state.police                         -3.732e-01  1.075e-01  -3.470
## long.guns.seized                      5.065e-02  5.145e-02   0.984
## sourceconfrontations                  5.175e-01  7.283e-02   7.106
## small.arms.seized                    -1.043e-02  2.195e-02  -0.475
## clips.seized                          1.253e-04  3.733e-04   0.336
## vehicles.seized                       9.760e-03  4.132e-03   2.362
## long.guns.seized:sourceconfrontations 1.011e-01  5.073e-02   1.992
##                                      Pr(>|t|)
## (Intercept)                           < 2e-16 ***
## cartridge.sezied                     2.15e-15 ***
## army                                 2.47e-10 ***
## federal.police                       4.41e-07 ***
## ministerial.police                   0.005284 **
## municipal.police                     8.10e-07 ***
## navy                                 0.649023
## state.police                         0.000525 ***
## long.guns.seized                     0.324967
## sourceconfrontations                 1.35e-12 ***
## small.arms.seized                    0.634695
## clips.seized                         0.737161
## vehicles.seized                      0.018218 *
## long.guns.seized:sourceconfrontations 0.046389 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.049 on 5382 degrees of freedom
## Multiple R-squared:  0.09518,    Adjusted R-squared:  0.093
## F-statistic: 43.55 on 13 and 5382 DF,  p-value: < 2.2e-16
```

- This model includes a series of controls in order to adequately control for other possible mechanisms related to the outcome (see assumptions section for further discussion of this). For the purpose of the output, we will focus on the two constituent main effects, long guns seized and source, as well as their interaction.

- First, the long guns seized term indicates that for every one more long guns seized, .984 more people are dead net of all other factors and when source is the aggressions data, but this is not significant at the .05 level. This is the opposite of what we originally hypothesized.
- The source term indicates that on average in the confrontations dataset, net of all other factors and when long guns seized is 0, there are 7.106 more people dead than in the aggressions dataset and it is significant at the .001 level. As aforementioned, conditional hypotheses give new meanings to the constituent variables. Rather, the terms for 'source' and 'long guns seized' indicate net of all other factors and when the other variable is at zero.

- The interaction term is testing for a differences in slopes; specifically whether or not the relationship between long guns seized and total people dead is significantly different for the aggression versus the confrontations dataset. This term indicates that when the dataset is confrontations, every one unit increase in long guns seized (or for every additional long gun seized), there are 1.992 more dead than the original long gun seized term (which is .984). Substantively, this indicates that the slope between long guns seized and total people dead is 1.992 more for the confrontations dataset than the aggressions dataset. The interaction term is significant at the .05 level.

2d) show all your computations to estimate the corresponding marginal effect and its standard error

- The marginal effect, as indicated in class, can be found by compounding the original slope term (in our case, the long gun seized term) and the interaction term. Therefore, this would be .984 + 1.992 = 2.976. This indicates that when the dataset is confrontations (as opposed to aggressions), for every long gun seized, 2.976 more people are dead.
- The standard error of the marginal effect is calculated using the following formula: var(B1) + Z^2 * var(B3) + 2ZCov(B1, B3), where in our case: B1: long guns seized slope Z: source (0 = aggressions, 1 = confrontations) B3: long guns seized * source interaction term

We can first extract the variance - covariance matrix necessary to calculate this using the following code:

```
vcov(fit)
```

```
##                                       (Intercept) cartridge.sezied
## (Intercept)                          6.710039e-03    -5.325002e-08
## cartridge.sezied                    -5.325002e-08     5.231649e-10
## army                                -2.616837e-03     1.006302e-07
## federal.police                      -3.165742e-03     7.542938e-08
## ministerial.police                  -4.015943e-03     5.243026e-08
## municipal.police                    -4.836634e-03     4.438084e-08
## navy                                -2.626316e-03     1.099716e-07
## state.police                        -3.518808e-03     7.091795e-08
## long.guns.seized                    -4.353787e-04    -1.709094e-07
## sourceconfrontations                -4.117572e-03     8.663958e-08
## small.arms.seized                    6.966012e-05     9.577649e-08
## clips.seized                        -5.386948e-07    -5.937481e-10
## vehicles.seized                     -3.833115e-06     8.692281e-10
## long.guns.seized:sourceconfrontations  4.700265e-04     1.546523e-08
##                                              army federal.police
## (Intercept)                         -2.616837e-03  -3.165742e-03
## cartridge.sezied                     1.006302e-07   7.542938e-08
## army                                 5.768119e-03   2.595223e-03
## federal.police                       2.595223e-03   9.553023e-03
## ministerial.police                   2.578860e-03   2.397200e-03
## municipal.police                     2.634917e-03   2.620137e-03
## navy                                 2.944267e-03   2.548609e-03
## state.police                         2.164396e-03   1.879591e-03
## long.guns.seized                    -2.270099e-04  -1.590297e-05
## sourceconfrontations                -7.448981e-04   4.220301e-04
## small.arms.seized                   -5.441164e-05  -1.216452e-04
## clips.seized                        -6.038693e-07   8.178105e-07
## vehicles.seized                     -4.293661e-06   9.786336e-07
## long.guns.seized:sourceconfrontations  1.147411e-04  -3.192296e-05
##                                     ministerial.police municipal.police
## (Intercept)                              -4.015943e-03    -4.836634e-03
## cartridge.sezied                          5.243026e-08     4.438084e-08
## army                                      2.578860e-03     2.634917e-03
## federal.police                            2.397200e-03     2.620137e-03
## ministerial.police                        1.281803e-02     3.158790e-03
## municipal.police                          3.158790e-03     7.175609e-03
## navy                                      2.291663e-03     2.567167e-03
## state.police                              2.342796e-03     2.580173e-03
## long.guns.seized                         -1.121451e-05     1.421911e-04
## sourceconfrontations                      1.284931e-03     2.106150e-03
## small.arms.seized                        -7.287763e-05    -9.064498e-05
## clips.seized                              5.756020e-07     5.840350e-07
## vehicles.seized                           3.227502e-06     4.621845e-06
## long.guns.seized:sourceconfrontations    -1.359821e-05    -1.624817e-04
##                                                 navy  state.police
## (Intercept)                         -2.626316e-03 -3.518808e-03
## cartridge.sezied                     1.099716e-07  7.091795e-08
## army                                 2.944267e-03  2.164396e-03
## federal.police                       2.548609e-03  1.879591e-03
## ministerial.police                   2.291663e-03  2.342796e-03
## municipal.police                     2.567167e-03  2.580173e-03
## navy                                 2.862345e-02  2.148852e-03
```

```
## state.police                              2.148852e-03  1.156587e-02
## long.guns.seized                         -9.445780e-05  1.266830e-04
## sourceconfrontations                     -3.821794e-04  1.157797e-03
## small.arms.seized                        -5.151855e-05 -9.945373e-05
## clips.seized                              1.017059e-07  8.569213e-07
## vehicles.seized                          -3.720852e-05  5.016431e-06
## long.guns.seized:sourceconfrontations -1.737634e-05 -1.658131e-04
##                                       long.guns.seized
## (Intercept)                              -4.353787e-04
## cartridge.sezied                         -1.709094e-07
## army                                     -2.270099e-04
## federal.police                           -1.590297e-05
## ministerial.police                       -1.121451e-05
## municipal.police                          1.421911e-04
## navy                                     -9.445780e-05
## state.police                              1.266830e-04
## long.guns.seized                          2.647219e-03
## sourceconfrontations                      5.261867e-04
## small.arms.seized                        -1.999081e-04
## clips.seized                             -3.223272e-07
## vehicles.seized                          -6.269431e-06
## long.guns.seized:sourceconfrontations    -2.557878e-03
##                                       sourceconfrontations
## (Intercept)                                  -4.117572e-03
## cartridge.sezied                              8.663958e-08
## army                                         -7.448981e-04
## federal.police                                4.220301e-04
## ministerial.police                            1.284931e-03
## municipal.police                              2.106150e-03
## navy                                         -3.821794e-04
## state.police                                  1.157797e-03
## long.guns.seized                              5.261867e-04
## sourceconfrontations                          5.304735e-03
## small.arms.seized                            -1.132056e-04
## clips.seized                                  3.497723e-07
## vehicles.seized                              -9.890482e-06
## long.guns.seized:sourceconfrontations        -5.867215e-04
##                                       small.arms.seized  clips.seized
## (Intercept)                              6.966012e-05 -5.386948e-07
## cartridge.sezied                         9.577649e-08 -5.937481e-10
## army                                    -5.441164e-05 -6.038693e-07
## federal.police                          -1.216452e-04  8.178105e-07
## ministerial.police                      -7.287763e-05  5.756020e-07
## municipal.police                        -9.064498e-05  5.840350e-07
## navy                                    -5.151855e-05  1.017059e-07
## state.police                            -9.945373e-05  8.569213e-07
## long.guns.seized                        -1.999081e-04 -3.223272e-07
## sourceconfrontations                    -1.132056e-04  3.497723e-07
## small.arms.seized                        4.816048e-04 -1.034672e-07
## clips.seized                            -1.034672e-07  1.393606e-07
## vehicles.seized                         -2.212652e-06 -7.840377e-09
## long.guns.seized:sourceconfrontations    9.138058e-05 -4.391337e-07
##                                       vehicles.seized
## (Intercept)                             -3.833115e-06
```

```
## cartridge.sezied                          8.692281e-10
## army                                     -4.293661e-06
## federal.police                            9.786336e-07
## ministerial.police                        3.227502e-06
## municipal.police                          4.621845e-06
## navy                                     -3.720852e-05
## state.police                              5.016431e-06
## long.guns.seized                         -6.269431e-06
## sourceconfrontations                     -9.890482e-06
## small.arms.seized                        -2.212652e-06
## clips.seized                             -7.840377e-09
## vehicles.seized                           1.707715e-05
## long.guns.seized:sourceconfrontations     3.818035e-06
##                                    long.guns.seized:sourceconfrontations
## (Intercept)                                                    4.700265e-04
## cartridge.sezied                                               1.546523e-08
## army                                                           1.147411e-04
## federal.police                                                -3.192296e-05
## ministerial.police                                            -1.359821e-05
## municipal.police                                              -1.624817e-04
## navy                                                          -1.737634e-05
## state.police                                                  -1.658131e-04
## long.guns.seized                                              -2.557878e-03
## sourceconfrontations                                          -5.867215e-04
## small.arms.seized                                              9.138058e-05
## clips.seized                                                  -4.391337e-07
## vehicles.seized                                                3.818035e-06
## long.guns.seized:sourceconfrontations                          2.573099e-03
```

From this, we can calculate:

```
var <- .002647219 + 1 * 1 *.002573099 + 2 * 1 * .002557878
var
```

```
## [1] 0.01033607
```

```
sqrt(var)
```

```
## [1] 0.1016665
```

Therefore the standard error of the marginal effect is .10167 (which matches the interaction term of the regression).

2e) be explicit in your assumptions

- There are quite a number of assumptions within this model. First off, a linear regression model was used – which indicates that we are assuming the dependent variable to be continuous and that there is a linear relationship between the predictors and the dependent variable. The dependent variable is not properly continuous as it can only have integer values and it cannot have any negative values; however for the purpose of drawing overall insight we may still be able to use this method to gain some understanding of

the relationship between the variables, but a poisson or zero-inflated binomial regression would theoretically be better suited for the nature of the data.

- A linear model, by definition, assumes that the predictors are not significantly correlated to one another (rather, that they are orthogonal). In this case, it is likely that these variables are at least somewhat correlated as the presence of one type of armed force (i.e. army) is likely correlated with another, as is the type of weaponry seized is likely correlated as well (long guns, for example, require cartridges in order to operate, therefore these metrics are likely correlated). However, this assumption may be somewhat malleable depending on how correlated the variables are, and insight can still be discerned overall from the model – although one must be skeptical of the insight gleaned.
- Because the dependent variable is a composite, we are collapsing different forms of death. As this model does not condition on the depenvent variable, we are making the implicit assumption that the predictors predict different types of death (i.e. civilian vs. non-civilian) to be equivalent, as the two are collapsed into total and the model is predicting this aggregated value. This perhaps is not necessarily the case and one could, with more complicated machinery, condition on the dependent variable as well to account for the potentially differential relationship.
- Furthermore, it is worth noting that because this model does not condition on any other variables or does not include an interaction between 'source' and any other variables, it is implicitly assuming that the relationship between any other predictor and total dead is constant as a function of source, which is likely not the case.

2f) be explicit in the limitations of your inferences

- The model has a series of limitations insofar as the dependent variable is not necessarily distributed ideally for the model used. Because this is not a count variable, one must take the interpretation of the model with skepticism (for example, some combination of predictors could result in non-integer or negative number of deaths, both of which are non-possible values).
- Additionally, one must recognize that the inter-correlation between the predictors means that these estimates may be even more unstable than one might think and therefore we are limited in the extent to which we are confident in any of the model's parameters due to this inter-correlation. *It is worth noting that there is a censorship on the dependent variable, as only the cases in which they are defined as a confrontation are documented, and therefore this censorship means that the inferences that one can make from the data are limited by the collection and scope of the data itself, indicating that perhaps we should be skeptical of the extent to which the estimates and insights from this dataset may extrapolate out of the scope of time, location, and data gathering method that this data possesses.

2g) phrase your finding for each question in two ways:

2g-1) one sentence that summarizes your insight

- The more long guns seized, the more people dead in the incidence and this relationship is even stronger for the data gathered in the confrontations dataset than the aggressions dataset.

2g-2) one paragraph that reflects all nuance in your insight

- There appears to be some evidence of a relationship that indicates the more long guns seized, the more dangerous (lethal) the situation is overall. Furthermore, the relationship between long gun seizure and death is three times stronger for the cases in the confrontations dataset than the aggressions dataset, controlling for all of the other types of ammunition seized and types of armed forces involved. The relationship between long guns and death is not statistically significant in the aggressions dataset, but it is in the confrontations set and the two relationships are significantly different – indicating that these two datasets are clearly examining quite different cases and capturing quite different relationships, while claiming to be coming from the same data-generating process. This is case for further investigation of the data gathering

processes of these two datasets and further examination of why lethality differs so greatly, as well as seizure lethality relationships, between two ostensibly identical data-generating processes.

2h) make sure to also include your code

- The code is included under each quetsion above.

Model #2 for question #2:

2.) Formulate two (2) conditional hypotheses that you seek to investigate with the data. One of your hypotheses should condition on two variables (as the example on the slides), and the other should condition on three variables. [50 pts]

- ZACH:

```
fit <- lm(total.people.dead ~ cartridge.sezied + federal.police + ministerial.police + m
unicipal.police + navy + state.police + long.guns.seized * source * army + small.arms.se
ized + clips.seized + vehicles.seized, data = AllData)
```

2a) formulate each one of your hypotheses explicitly in substantive terms (as opposed to statistical terms) using 2-3 lines at most

- As before, there is reason to believe that the relationship between seizure of long guns and death varies as a function of armed forces present; we hypothesize that the difference in this relationship previously found with different data sources is further delineated by whether or not the army is present – with differences in the relationship being observable only in the confrontation set.

2b) show exactly how each one of your hypotheses translates into the marginal effect that you will seek to estimate from the data

- We will estimate whether or not the previous two-way interaction can be extended into a three-way interaction such that this differential relationship between data sets of the relationship between long-gun seizure and total deaths is stronger when the army is present only in the confrontation set and not necessarily the aggression set. Rather, the marginal effect will be if conditioning on presence of army adds additional information about the relationship between long-gun seizure and death that may differ between the two datasets.

2c) show the output from your analysis in a consumable form

```
summary(fit)
```

```
##
## Call:
## lm(formula = total.people.dead ~ cartridge.sezied + federal.police +
##     ministerial.police + municipal.police + navy + state.police +
##     long.guns.seized * source * army + small.arms.seized + clips.seized +
##     vehicles.seized, data = AllData)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.365 -1.044 -0.532  0.434 51.021
##
## Coefficients:
##                                              Estimate Std. Error t value
## (Intercept)                                 0.9788875  0.0841806  11.628
## cartridge.sezied                           -0.0001563  0.0000232  -6.738
## federal.police                             -0.5278304  0.0976215  -5.407
## ministerial.police                         -0.3174379  0.1130876  -2.807
## municipal.police                           -0.4130403  0.0851754  -4.849
## navy                                       -0.0705032  0.1703932  -0.414
## state.police                               -0.3853555  0.1074034  -3.588
## long.guns.seized                            0.0452041  0.0760280   0.595
## sourceconfrontations                        0.4496398  0.0776240   5.793
## army                                       -0.4576034  0.2285776  -2.002
## small.arms.seized                          -0.0085839  0.0219676  -0.391
## clips.seized                                0.0001443  0.0003722   0.388
## vehicles.seized                             0.0097782  0.0041197   2.374
## long.guns.seized:sourceconfrontations       0.1683411  0.0763176   2.206
## long.guns.seized:army                      -0.0055546  0.1036553  -0.054
## sourceconfrontations:army                   0.1233911  0.2333687   0.529
## long.guns.seized:sourceconfrontations:army -0.0845933  0.1045816  -0.809
##                                            Pr(>|t|)
## (Intercept)                                 < 2e-16 ***
## cartridge.sezied                           1.77e-11 ***
## federal.police                             6.69e-08 ***
## ministerial.police                         0.005018 **
## municipal.police                           1.27e-06 ***
## navy                                       0.679061
## state.police                               0.000336 ***
## long.guns.seized                           0.552154
## sourceconfrontations                       7.33e-09 ***
## army                                       0.045339 *
## small.arms.seized                          0.695997
## clips.seized                               0.698253
## vehicles.seized                            0.017654 *
## long.guns.seized:sourceconfrontations      0.027440 *
## long.guns.seized:army                      0.957266
## sourceconfrontations:army                  0.597009
## long.guns.seized:sourceconfrontations:army 0.418624
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.042 on 5379 degrees of freedom
```

```
## Multiple R-squared:  0.1014, Adjusted R-squared:  0.09876
## F-statistic: 37.95 on 16 and 5379 DF,  p-value: < 2.2e-16
```

- This model includes a series of controls in order to adequately control for other possible mechanisms related to the outcome (see assumptions section for further discussion of this). For the purpose of the output, we will focus on the three constituent main effects, long guns seized, source and army, as well as their two-way interactions, and then the three way interaction term (the main conditional hypothesis).
- The first main effect is long guns seized; the term in the model indicates that when army and source are 0 (indicating an instance from the aggression set with no army presence), for every long gun seized, .045 more people are dead net of all other factors, but this is not significant at the .05 level.
- The second main effect is source; the term in the model indicates that when long guns seized and army are 0 (indicating an instance with no long guns seized and no army presence), the confrontations set on average has .449 more deaths than the aggression set per instance net of all other factors and this is significant at the .001 level.
- The third main effect is army; the term in the model indicates that when long guns seized and source are 0 (indicating an instance from the aggression set with no long guns seized), having army presence indicates on average .458 less deaths net of all other factors and this is significant at the .05 level.
- The first two-way interaction is between long guns seized and source such that the slope between long guns seized and deaths is .168 higher for the confrontations data set than the aggressions dataset net of all other factors and when army is not present; it is significant at the .01 level.
- The second two-way interaction is between long guns seized and army such that the slope between long guns seized and deaths is .0056 lower when the army is present as opposed to when the army is present net of all other factors and when source is 0 (meaning in the aggressions dataset). This term, however, is not statistically significant at the .05 level.
- The third two-way interaction is between source and army such that when looking at the confrontations dataset and the army is present, the average number of deaths is .1233 higher than when neither or only one of them is present net of all other factors and when long gun seizure is 0. This term, however, is not statistically significant at the .05 level.
- The three-way interaction term (or three way conditional hypothesis) is testing whether the slope of long guns predicting death varies as a function of both source and army. The term indicates that when source and army are 1 (indicating the army is present and the source is the confrontations set), the effect of long guns seizes on total deaths is -.085 less when neither or only one is present net of all other factors. This term, however, is not statistically significant at the .05 level. See next section for further discussion for the marginal effect estimated by this term.

2d) show all your computations to estimate the corresponding marginal effect and its standard error

- The marginal effect for a three variable conditioning hypothesis will be formulated using tables from Aiken & West (1991; see attached picture for formulas) which give the marginal effects and standard errors for a three-way interaction.

Table 1: Marginal Effects and Variances for Various Multiplicative Interaction Models (Double and Triple Interaction Terms)

Case	Equation	Marginal Effect	Variance
1a	$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ$	$\frac{\partial Y}{\partial X} = \beta_1 + \beta_3 Z$	$\hat{\sigma}^2_{\frac{\partial Y}{\partial X}} = \text{var}(\hat{\beta}_1) + Z^2\text{var}(\hat{\beta}_3) + 2Z\text{cov}(\hat{\beta}_1\hat{\beta}_3)$
1b	$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ$	$\frac{\partial Y}{\partial Z} = \beta_2 + \beta_3 X$	$\hat{\sigma}^2_{\frac{\partial Y}{\partial X}} = \text{var}(\hat{\beta}_2) + X^2\text{var}(\hat{\beta}_3) + 2X\text{cov}(\hat{\beta}_2\hat{\beta}_3)$
2	$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_6 ZW$	$\frac{\partial Y}{\partial X} = \beta_1 + \beta_4 Z$	$\hat{\sigma}^2_{\frac{\partial Y}{\partial X}} = \text{var}(\hat{\beta}_1) + Z^2\text{var}(\hat{\beta}_4) + 2Z\text{cov}(\hat{\beta}_1\hat{\beta}_4)$
3	$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_5 XW + \beta_6 ZW$	$\frac{\partial Y}{\partial X} = \beta_1 + \beta_4 Z + \beta_5 W$	$\hat{\sigma}^2_{\frac{\partial Y}{\partial X}} = \text{var}(\hat{\beta}_1) + Z^2\text{var}(\hat{\beta}_4) + W^2\text{var}(\hat{\beta}_5) + 2Z\text{cov}(\hat{\beta}_1\hat{\beta}_4) + 2W\text{cov}(\hat{\beta}_1\hat{\beta}_5) + 2ZW\text{cov}(\hat{\beta}_4\hat{\beta}_5)$
4	$\hat{Y} = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 W + \beta_4 XZ + \beta_5 XW + \beta_6 ZW + \beta_7 XZW$	$\frac{\partial Y}{\partial X} = \beta_1 + \beta_4 Z + \beta_5 W + \beta_7 ZW$	$\hat{\sigma}^2_{\frac{\partial Y}{\partial X}} = \text{var}(\hat{\beta}_1) + Z^2\text{var}(\hat{\beta}_4) + W^2\text{var}(\hat{\beta}_5) + Z^2W^2\text{var}(\hat{\beta}_7) + 2Z\text{cov}(\hat{\beta}_1\hat{\beta}_4) + 2W\text{cov}(\hat{\beta}_1\hat{\beta}_5) + 2ZW\text{cov}(\hat{\beta}_1\hat{\beta}_7) + 2ZW\text{cov}(\hat{\beta}_4\hat{\beta}_5) + 2WZ^2\text{cov}(\hat{\beta}_4\hat{\beta}_7) + 2ZW^2\text{cov}(\hat{\beta}_5\hat{\beta}_7)$

Formulas for Marginal Effects

The formula for marginal effect is the long gun term added with the army – long gun interaction term, the source – long gun interaction term, and the three-way interaction term like such:

.0452 + .1683 - .0056 - .0846 = .1233

This indicates that when army is present and it is the confrontations dataset, each long gun seized results in .1233 more deaths. Let's now calculate the standard error of this marginal effect.

- The standard error of the marginal effect is calculated using the following formula: var(B1) + Z^2 var(B4) + W^2 var(B5) + Z^2 W^2 var(B7) + 2ZCov(B1, B4) + 2Wcov(B1,B5) + 2ZWcov(B1,B7) + 2ZWcov(B4,B5) + 2WZ^2cov(B4, B7) + 2ZW^2cov(B5, B7), where in our case: B1: long guns seized slope Z: source (0 = aggressions, 1 = confrontations) W: army (0 = no army, 1 = army) B4: long guns seized * source interaction term B5: long guns seized * army interaction term B7: long guns seized, army, source interaction term

We can first extract the variance - covariance matrix necessary to calculate this using the following code:

```
vcov(fit)
```

```
##                                             (Intercept) cartridge.sezied
## (Intercept)                                7.086381e-03    -4.607461e-08
## cartridge.sezied                           -4.607461e-08     5.382140e-10
## federal.police                             -3.199882e-03     5.043803e-08
## ministerial.police                         -4.132252e-03     4.775181e-08
## municipal.police                           -5.030244e-03     4.371274e-08
## navy                                       -2.581405e-03     1.071673e-08
## state.police                               -3.625475e-03     6.083896e-08
## long.guns.seized                           -4.646958e-04    -1.259836e-07
## sourceconfrontations                       -4.611159e-03     3.319278e-08
## army                                       -7.007749e-03     7.447815e-08
## small.arms.seized                           8.051928e-05     9.492434e-08
## clips.seized                               -6.630114e-07    -5.746510e-10
## vehicles.seized                            -4.398949e-06     9.104778e-10
## long.guns.seized:sourceconfrontations       4.994002e-04     1.389319e-08
## long.guns.seized:army                       4.999100e-04    -8.436395e-08
## sourceconfrontations:army                   4.767567e-03     1.342643e-07
## long.guns.seized:sourceconfrontations:army -5.138984e-04     2.241846e-08
##                                            federal.police
## (Intercept)                                 -3.199882e-03
## cartridge.sezied                             5.043803e-08
## federal.police                               9.529949e-03
## ministerial.police                           2.402870e-03
## municipal.police                             2.628207e-03
## navy                                         2.657240e-03
## state.police                                 1.894117e-03
## long.guns.seized                            -6.638868e-05
## sourceconfrontations                         5.441746e-04
## army                                         3.089621e-03
## small.arms.seized                           -1.218748e-04
## clips.seized                                 8.068014e-07
## vehicles.seized                              9.780934e-07
## long.guns.seized:sourceconfrontations       -3.607914e-05
## long.guns.seized:army                        5.023772e-05
## sourceconfrontations:army                   -6.906234e-04
## long.guns.seized:sourceconfrontations:army   3.156954e-05
##                                            ministerial.police
## (Intercept)                                     -4.132252e-03
## cartridge.sezied                                 4.775181e-08
## federal.police                                   2.402870e-03
## ministerial.police                               1.278880e-02
## municipal.police                                 3.218326e-03
## navy                                             2.264058e-03
## state.police                                     2.371412e-03
## long.guns.seized                                -1.147214e-04
## sourceconfrontations                             1.455092e-03
## army                                             4.038382e-03
## small.arms.seized                               -7.236410e-05
## clips.seized                                     6.061028e-07
## vehicles.seized                                  3.308402e-06
## long.guns.seized:sourceconfrontations            8.924627e-05
## long.guns.seized:army                            3.647956e-05
## sourceconfrontations:army                       -1.608607e-03
```

```
## long.guns.seized:sourceconfrontations:army      -3.060868e-05
##                                           municipal.police          navy
## (Intercept)                                  -5.030244e-03 -2.581405e-03
## cartridge.sezied                              4.371274e-08  1.071673e-08
## federal.police                                2.628207e-03  2.657240e-03
## ministerial.police                            3.218326e-03  2.264058e-03
## municipal.police                              7.254851e-03  2.503176e-03
## navy                                          2.503176e-03  2.903385e-02
## state.police                                  2.630483e-03  2.171029e-03
## long.guns.seized                              5.461226e-05  6.613615e-05
## sourceconfrontations                          2.360498e-03 -1.645661e-04
## army                                          4.975400e-03  2.416815e-03
## small.arms.seized                            -9.228823e-05 -6.218776e-05
## clips.seized                                  6.445485e-07  2.431985e-08
## vehicles.seized                               4.824573e-06 -3.697910e-05
## long.guns.seized:sourceconfrontations        -6.325545e-05 -4.216119e-04
## long.guns.seized:army                        -8.474448e-05 -1.832510e-04
## sourceconfrontations:army                    -2.530927e-03 -2.706980e-05
## long.guns.seized:sourceconfrontations:army    7.531661e-05  5.380455e-04
##                                            state.police long.guns.seized
## (Intercept)                                -3.625475e-03    -4.646958e-04
## cartridge.sezied                            6.083896e-08    -1.259836e-07
## federal.police                              1.894117e-03    -6.638868e-05
## ministerial.police                          2.371412e-03    -1.147214e-04
## municipal.police                            2.630483e-03     5.461226e-05
## navy                                        2.171029e-03     6.613615e-05
## state.police                                1.153550e-02     1.455794e-04
## long.guns.seized                            1.455794e-04     5.780256e-03
## sourceconfrontations                        1.329446e-03     5.254270e-04
## army                                        3.491582e-03     5.261841e-04
## small.arms.seized                          -1.031083e-04    -3.068185e-04
## clips.seized                                8.854754e-07    -6.985281e-08
## vehicles.seized                             5.166709e-06    -3.490848e-06
## long.guns.seized:sourceconfrontations      -2.030219e-04    -5.699216e-03
## long.guns.seized:army                      -1.659672e-04    -5.687154e-03
## sourceconfrontations:army                  -1.494342e-03    -5.462181e-04
## long.guns.seized:sourceconfrontations:army  1.971722e-04     5.697308e-03
##                                            sourceconfrontations
## (Intercept)                                      -4.611159e-03
## cartridge.sezied                                  3.319278e-08
## federal.police                                    5.441746e-04
## ministerial.police                                1.455092e-03
## municipal.police                                  2.360498e-03
## navy                                             -1.645661e-04
## state.police                                      1.329446e-03
## long.guns.seized                                  5.254270e-04
## sourceconfrontations                              6.025481e-03
## army                                              4.621713e-03
## small.arms.seized                                -1.279366e-04
## clips.seized                                      4.682802e-07
## vehicles.seized                                  -9.173251e-06
## long.guns.seized:sourceconfrontations            -6.915321e-04
## long.guns.seized:army                            -5.212641e-04
## sourceconfrontations:army                        -6.059141e-03
```

```
## long.guns.seized:sourceconfrontations:army         6.924325e-04
##                                                          army small.arms.seized
## (Intercept)                                     -7.007749e-03      8.051928e-05
## cartridge.sezied                                 7.447815e-08      9.492434e-08
## federal.police                                   3.089621e-03     -1.218748e-04
## ministerial.police                               4.038382e-03     -7.236410e-05
## municipal.police                                 4.975400e-03     -9.228823e-05
## navy                                             2.416815e-03     -6.218776e-05
## state.police                                     3.491582e-03     -1.031083e-04
## long.guns.seized                                 5.261841e-04     -3.068185e-04
## sourceconfrontations                             4.621713e-03     -1.279366e-04
## army                                             5.224771e-02     -1.845904e-04
## small.arms.seized                               -1.845904e-04      4.825767e-04
## clips.seized                                     8.077294e-07     -1.135990e-07
## vehicles.seized                                  2.339366e-06     -2.305227e-06
## long.guns.seized:sourceconfrontations           -5.439411e-04      2.018622e-04
## long.guns.seized:army                           -6.032327e-03      2.052982e-04
## sourceconfrontations:army                       -5.004404e-02      1.384122e-04
## long.guns.seized:sourceconfrontations:army       6.038379e-03     -2.097880e-04
##                                                  clips.seized vehicles.seized
## (Intercept)                                     -6.630114e-07   -4.398949e-06
## cartridge.sezied                                -5.746510e-10    9.104778e-10
## federal.police                                   8.068014e-07    9.780934e-07
## ministerial.police                               6.061028e-07    3.308402e-06
## municipal.police                                 6.445485e-07    4.824573e-06
## navy                                             2.431985e-08   -3.697910e-05
## state.police                                     8.854754e-07    5.166709e-06
## long.guns.seized                                -6.985281e-08   -3.490848e-06
## sourceconfrontations                             4.682802e-07   -9.173251e-06
## army                                             8.077294e-07    2.339366e-06
## small.arms.seized                               -1.135990e-07   -2.305227e-06
## clips.seized                                     1.385440e-07   -7.390038e-09
## vehicles.seized                                 -7.390038e-09    1.697185e-05
## long.guns.seized:sourceconfrontations           -6.536839e-07    1.092426e-06
## long.guns.seized:army                           -5.945794e-07   -5.552067e-06
## sourceconfrontations:army                       -1.420544e-06   -6.809334e-06
## long.guns.seized:sourceconfrontations:army       5.503749e-07    5.514895e-06
##                                                 long.guns.seized:sourceconfrontations
## (Intercept)                                                              4.994002e-04
## cartridge.sezied                                                         1.389319e-08
## federal.police                                                          -3.607914e-05
## ministerial.police                                                       8.924627e-05
## municipal.police                                                        -6.325545e-05
## navy                                                                    -4.216119e-04
## state.police                                                            -2.030219e-04
## long.guns.seized                                                        -5.699216e-03
## sourceconfrontations                                                    -6.915321e-04
## army                                                                    -5.439411e-04
## small.arms.seized                                                        2.018622e-04
## clips.seized                                                            -6.536839e-07
## vehicles.seized                                                          1.092426e-06
## long.guns.seized:sourceconfrontations                                    5.824379e-03
## long.guns.seized:army                                                    5.670264e-03
## sourceconfrontations:army                                                6.919337e-04
```

```
## long.guns.seized:sourceconfrontations:army                         -5.829394e-03
##                                                 long.guns.seized:army
## (Intercept)                                              4.999100e-04
## cartridge.sezied                                        -8.436395e-08
## federal.police                                           5.023772e-05
## ministerial.police                                       3.647956e-05
## municipal.police                                        -8.474448e-05
## navy                                                    -1.832510e-04
## state.police                                            -1.659672e-04
## long.guns.seized                                        -5.687154e-03
## sourceconfrontations                                    -5.212641e-04
## army                                                    -6.032327e-03
## small.arms.seized                                        2.052982e-04
## clips.seized                                            -5.945794e-07
## vehicles.seized                                         -5.552067e-06
## long.guns.seized:sourceconfrontations                    5.670264e-03
## long.guns.seized:army                                    1.074442e-02
## sourceconfrontations:army                                5.970715e-03
## long.guns.seized:sourceconfrontations:army              -1.073217e-02
##                                                 sourceconfrontations:army
## (Intercept)                                                  4.767567e-03
## cartridge.sezied                                             1.342643e-07
## federal.police                                              -6.906234e-04
## ministerial.police                                          -1.608607e-03
## municipal.police                                            -2.530927e-03
## navy                                                        -2.706980e-05
## state.police                                                -1.494342e-03
## long.guns.seized                                            -5.462181e-04
## sourceconfrontations                                        -6.059141e-03
## army                                                        -5.004404e-02
## small.arms.seized                                            1.384122e-04
## clips.seized                                                -1.420544e-06
## vehicles.seized                                             -6.809334e-06
## long.guns.seized:sourceconfrontations                        6.919337e-04
## long.guns.seized:army                                        5.970715e-03
## sourceconfrontations:army                                    5.446095e-02
## long.guns.seized:sourceconfrontations:army                  -6.337378e-03
##                                                 long.guns.seized:sourceconfrontations:army
## (Intercept)                                                                  -5.138984e-04
## cartridge.sezied                                                              2.241846e-08
## federal.police                                                                3.156954e-05
## ministerial.police                                                           -3.060868e-05
## municipal.police                                                              7.531661e-05
## navy                                                                          5.380455e-04
## state.police                                                                  1.971722e-04
## long.guns.seized                                                              5.697308e-03
## sourceconfrontations                                                          6.924325e-04
## army                                                                          6.038379e-03
## small.arms.seized                                                            -2.097880e-04
## clips.seized                                                                  5.503749e-07
## vehicles.seized                                                               5.514895e-06
## long.guns.seized:sourceconfrontations                                        -5.829394e-03
## long.guns.seized:army                                                        -1.073217e-02
```

```
## sourceconfrontations:army                                  -6.337378e-03
## long.guns.seized:sourceconfrontations:army                  1.093731e-02
```

From this, we can calculate (because Z and W are 1, they are dropped from this equation for simplicity's sake):

```
var <- 5.780256e-03 + 5.824379e-03 + 1.074442e-02 + 1.093731e-02 + 2 * -5.699216e-03 + 2
 * -5.687154e-03 + 2 * 5.697308e-03 + 2 * 5.670264e-03 + 2 * -5.829394e-03 + 2 * -1.0732
17e-02
var
```

```
## [1] 0.000125641
```

```
sqrt(var)
```

```
## [1] 0.01120897
```

Becuase we are just evaluating the standard error of the marginal effect, it does not necessarily match the standard error of the three-way interaciton term itself (as this is also including the two-way interaction between source and army, which does not impact the marginal effect of long guns on total death in this instance), the standard error of the marginal effect is .0112.

2e) be explicit in your assumptions

- Because of similarities between this model and the last, some of the assumptions stated are repeated here.
- There are quite a number of assumptions within this model. First off, a linear regression model was used – which indicates that we are assuming the dependent variable to be continuous and that there is a linear relationship between the predictors and the dependent variable. The dependent variable is not properly continuous as it can only have integer values and it cannot have any negative values; however for the purpose of drawing overall insight we may still be able to use this method to gain some understanding of the relationship between the variables, but a poisson or zero-inflated binomial regression would theoretically be better suited for the nature of the data.
- A linear model, by definition, assumes that the predictors are not significantly correlated to one another (rather, that they are orthogonal). In this case, it is likely that these variables are at least somewhat correlated as the presence of one type of armed force (i.e. army) is likely correlated with another, as is the type of weaponry seized is likely correlated as well (long guns, for example, require cartridges in order to operate, therefore these metrics are likely correlated). However, this assumption may be somewhat malleable depending on how correlated the variables are, and insight can still be discerned overall from the model – although one must be skeptical of the insight gleaned.
- Because the dependent variable is a composite, we are collapsing different forms of death. As this model does not condition on the depenvent variable, we are making the implicit assumption that the predictors predict different types of death (i.e. civilian vs. non-civilian) to be equivalent, as the two are collapsed into total and the model is predicting this aggregated value. This perhaps is not necessarily the case and one could, with more complicated machinery, condition on the dependent variable as well to account for the potentially differential relationship.
- Furthermore, it is worth noting that because this model does not condition on any other variables or does not include an interaction between 'source' and any other variables, it is implicitly assuming that the relationship between any other predictor and total dead is constant as a function of source, which is likely not the case.

- Additionally, this model is making an assumption or rather including a term such that the two-way relationship between long gun seizure and total deaths varies as a function of whether or not the army is present (which is what the three-way interaction term includes), but this assumption that these relationships vary as a function of whether or not army is present may not be true (and given the context of our model and the p-value, does not appear to hold true). Therefore, in allowing for this three-way interaction we are making an implicit assumption of allowing the relationships to vary as a function of the conditioning on the third variable.

2f) be explicit in the limitations of your inferences

```
* Because of similarities between this model and the last, some of the limitations state
d are repeated here.
```

- The model has a series of limitations insofar as the dependent variable is not necessarily distributed ideally for the model used. Because this is not a count variable, one must take the interpretation of the model with skepticism (for example, some combination of predictors could result in non-integer or negative number of deaths, both of which are non-possible values).
- Additionally, one must recognize that the inter-correlation between the predictors means that these estimates may be even more unstable than one might think and therefore we are limited in the extent to which we are confident in any of the model's parameters due to this inter-correlation. *It is worth noting that there is a censorship on the dependent variable, as only the cases in which they are defined as a confrontation are documented, and therefore this censorship means that the inferences that one can make from the data are limited by the collection and scope of the data itself, indicating that perhaps we should be skeptical of the extent to which the estimates and insights from this dataset may extrapolate out of the scope of time, location, and data gathering method that this data possesses.
- There is a limitation in the three-way interaction term because of the relative number of scenarios in each cell. For example, there is not an equivalent number of instances whether army = 1 and source = 0 as compared with when army = 0 and source = 1. The interaction term is allowing for there to be differential long gun seizure – death count slopes amongst the four categories and a marginal effect when both are present. However, the different sample sizes in the various cells may limit the extent to which particular slopes may be estimated and whether or not a subsequent interaction term can be estimated; making the overall inferences from the interaction term limited in nature. A more deliberately balanced dataset may ameliorate this problem, although this then ignores the naturalistic relationship between these variables occurring together.

2g) phrase your finding for each question in two ways:

2g-1) one sentence that summarizes your insight

- The more long guns seized, the more deaths overall in the confrontations data but not the aggressions data and this effect does not change whether the army is there or not regardless of dataset.

2g-2) one paragraph that reflects all nuance in your insight

- The more long guns seized, the more deaths overall in the confrontations data but not the aggressions data. However, there is no significant relationship between long guns and deaths in the aggressions dataset when army is present or army is not present. Furthermore, the presence of the army in the confrontations set does not significantly impact the relationship between long gun seizure and deaths above and beyond the previously observed relationship between the two in the confrontations dataset irrespective of the presence of the army.

2h) make sure to also include your code

- The code is included under each quetsion above.