G5999_Data Challenge 2

Team 4

22 March 2017

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

library(readr)
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.3.2
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
library(ggplot2)
library(zoo)
library(dplyr)
library(plyr)
library(lubridate)
library(scales)
```

```
## Warning: package 'scales' was built under R version 3.3.2
```

```
inFileName1   <- "../../data/processed/AllViolenceData_170216.csv"     # cleaned data on violence
AllData <- read_csv(inFileName1)
```

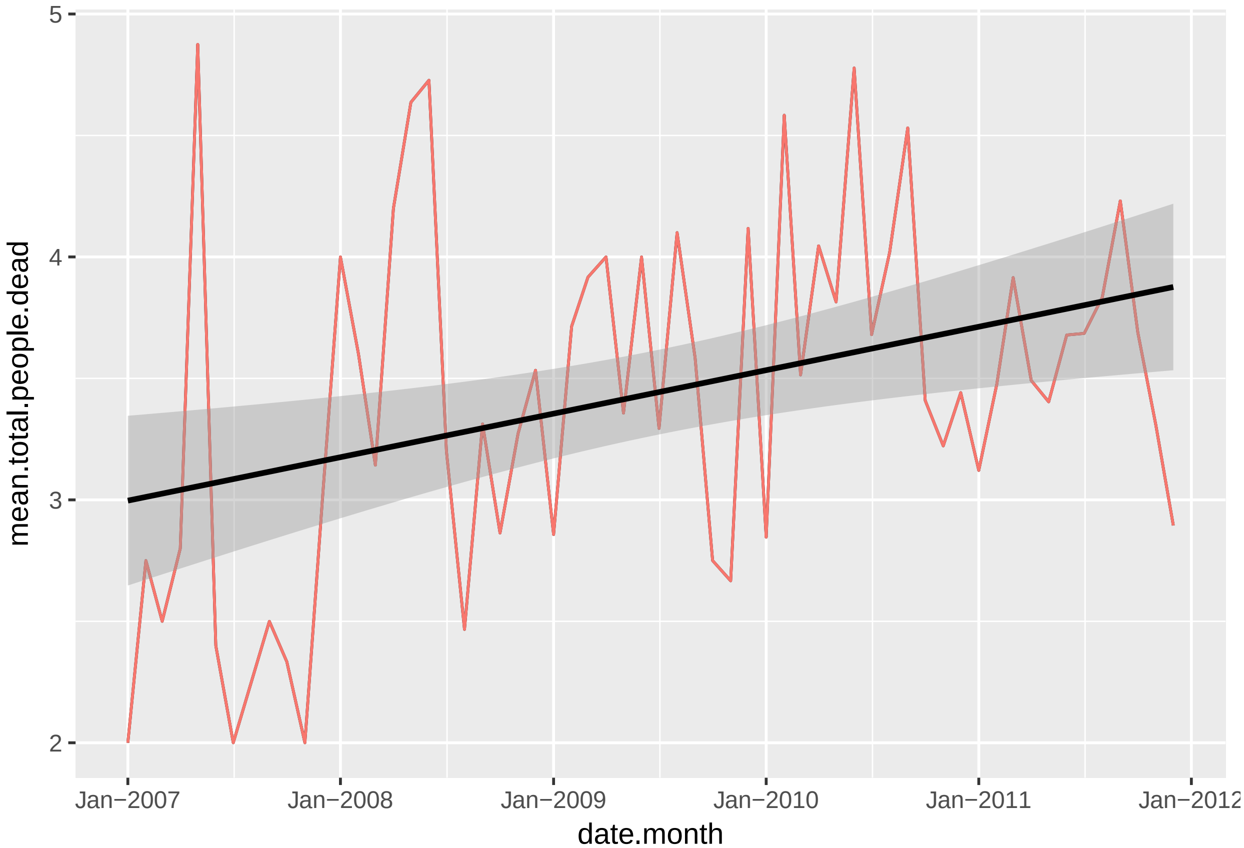
Q1

Time-series of Total Deaths

One question we wanted to investigate was whether any differences existed over time in the total number of deaths per confrontation.

To do so, we created a dataset which aggregated the mean number of total people dead by month. This only includes the events which have at least 1 fatality.

```
AllData$date.month <- round_date(AllData$date, "month")
monthly_agg <- ddply(AllData[AllData$total.people.dead > 1,], .(date.month), summarize, mean.total.peopl
ggplot(monthly_agg[2:nrow(monthly_agg),], aes(date.month, mean.total.people.dead)) +
  scale_x_date(labels = date_format("%b-%Y"), breaks = date_breaks("1 year")) +
  geom_line() +
  geom_line(aes(y = monthly_agg[2:nrow(monthly_agg),]$mean.total.people.dead, color = "red")) +
  geom_smooth(method = "lm", se = TRUE, color = "black", aes(group=1)) +
  theme(legend.position = "none")
```



From the above chart, we can see that there is a linear increasing trend in the number of deaths over time. However, there is quite a bit of fluctuation, even though we have smoothed out the trend by taking the average number of fatalities per month. The fluctuation month-to-month may be an indication of seasonality which would be interesting to explore in a regression analysis.

In addition, we built a linear model to predict the total number of people dead by month.

```
summary(lm(total.people.dead ~ factor(date.month), data = AllData[AllData$total.people.dead > 0,]))
```

```
##
## Call:
## lm(formula = total.people.dead ~ factor(date.month), data = AllData[AllData$total.people.dead >
##     0, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.952 -1.320 -0.650  0.434 49.385
##
## Coefficients:
##                              Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     7.000      2.544   2.751  0.00598 **
## factor(date.month)2007-01-01   -5.500      3.116  -1.765  0.07770 .
## factor(date.month)2007-02-01   -5.364      2.658  -2.018  0.04367 *
## factor(date.month)2007-03-01   -5.077      2.641  -1.923  0.05463 .
## factor(date.month)2007-04-01   -4.714      2.720  -1.733  0.08320 .
## factor(date.month)2007-05-01   -3.048      2.604  -1.170  0.24203
## factor(date.month)2007-06-01   -5.417      2.648  -2.045  0.04093 *
## factor(date.month)2007-07-01   -5.000      3.598  -1.389  0.16480
```

```
## factor(date.month)2007-08-01   -6.000      3.116  -1.925  0.05430 .
## factor(date.month)2007-09-01   -5.250      2.845  -1.845  0.06508 .
## factor(date.month)2007-10-01   -5.500      2.699  -2.038  0.04166 *
## factor(date.month)2007-11-01   -5.556      2.682  -2.071  0.03843 *
## factor(date.month)2007-12-01   -5.111      2.682  -1.906  0.05681 .
## factor(date.month)2008-01-01   -4.333      2.682  -1.616  0.10629
## factor(date.month)2008-02-01   -3.400      2.669  -1.274  0.20276
## factor(date.month)2008-03-01   -4.846      2.641  -1.835  0.06658 .
## factor(date.month)2008-04-01   -4.222      2.614  -1.615  0.10641
## factor(date.month)2008-05-01   -4.182      2.602  -1.607  0.10810
## factor(date.month)2008-06-01   -3.437      2.623  -1.311  0.19010
## factor(date.month)2008-07-01   -4.935      2.585  -1.909  0.05636 .
## factor(date.month)2008-08-01   -5.312      2.584  -2.056  0.03989 *
## factor(date.month)2008-09-01   -4.943      2.581  -1.915  0.05555 .
## factor(date.month)2008-10-01   -5.024      2.575  -1.951  0.05113 .
## factor(date.month)2008-11-01   -4.913      2.599  -1.890  0.05884 .
## factor(date.month)2008-12-01   -4.812      2.584  -1.862  0.06265 .
## factor(date.month)2009-01-01   -5.278      2.614  -2.019  0.04360 *
## factor(date.month)2009-02-01   -4.690      2.588  -1.812  0.07009 .
## factor(date.month)2009-03-01   -4.704      2.591  -1.815  0.06960 .
## factor(date.month)2009-04-01   -4.429      2.604  -1.700  0.08917 .
## factor(date.month)2009-05-01   -4.680      2.595  -1.804  0.07142 .
## factor(date.month)2009-06-01   -4.350      2.607  -1.668  0.09536 .
## factor(date.month)2009-07-01   -4.946      2.579  -1.918  0.05522 .
## factor(date.month)2009-08-01   -4.524      2.604  -1.737  0.08250 .
## factor(date.month)2009-09-01   -4.852      2.591  -1.872  0.06126 .
## factor(date.month)2009-10-01   -5.323      2.585  -2.059  0.03961 *
## factor(date.month)2009-11-01   -4.966      2.588  -1.919  0.05513 .
## factor(date.month)2009-12-01   -4.233      2.587  -1.637  0.10182
## factor(date.month)2010-01-01   -5.314      2.581  -2.059  0.03956 *
## factor(date.month)2010-02-01   -3.676      2.579  -1.425  0.15415
## factor(date.month)2010-03-01   -4.842      2.561  -1.891  0.05879 .
## factor(date.month)2010-04-01   -4.424      2.559  -1.728  0.08405 .
## factor(date.month)2010-05-01   -4.566      2.568  -1.778  0.07556 .
## factor(date.month)2010-06-01   -4.187      2.561  -1.635  0.10227
## factor(date.month)2010-07-01   -4.405      2.560  -1.721  0.08539 .
## factor(date.month)2010-08-01   -4.431      2.557  -1.733  0.08320 .
## factor(date.month)2010-09-01   -3.907      2.568  -1.522  0.12823
## factor(date.month)2010-10-01   -4.605      2.561  -1.798  0.07228 .
## factor(date.month)2010-11-01   -4.913      2.558  -1.920  0.05491 .
## factor(date.month)2010-12-01   -4.683      2.565  -1.826  0.06799 .
## factor(date.month)2011-01-01   -4.986      2.563  -1.945  0.05185 .
## factor(date.month)2011-02-01   -4.406      2.557  -1.723  0.08500 .
## factor(date.month)2011-03-01   -4.558      2.558  -1.782  0.07488 .
## factor(date.month)2011-04-01   -4.744      2.555  -1.856  0.06352 .
## factor(date.month)2011-05-01   -4.731      2.556  -1.851  0.06429 .
## factor(date.month)2011-06-01   -4.611      2.556  -1.804  0.07137 .
## factor(date.month)2011-07-01   -4.645      2.556  -1.817  0.06933 .
## factor(date.month)2011-08-01   -4.350      2.557  -1.701  0.08903 .
## factor(date.month)2011-09-01   -4.385      2.557  -1.715  0.08647 .
## factor(date.month)2011-10-01   -4.186      2.559  -1.636  0.10203
## factor(date.month)2011-11-01   -5.100      2.557  -1.994  0.04621 *
## factor(date.month)2011-12-01   -5.163      2.574  -2.006  0.04498 *
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.544 on 2610 degrees of freedom
## Multiple R-squared:  0.02465,    Adjusted R-squared:  0.002225
## F-statistic: 1.099 on 60 and 2610 DF,  p-value: 0.2815
```

In the output we can see that there are significant coefficients on some of the date.month. However, the model overall is not statistically significant at the 10% level since the F-statistic has a p-value of 0.2815. This indicates that while time isn't the only factor that can predict the total number of deaths for fatal events, it does appear that there are differences over time (possibly seasonal).

From the dataset we constructed (restricting to only those events with at least one casualty), we can't infer how the average deaths per month resulting from military conflicts has trended over time. For example, there could be an increasing in conflicts which result in injuries but no fatalities which would decrease the average deaths per event. So while our finding here is interesting, future exploration could look to take into account the probability of an event having at least one fatality to see if there is a temporal trend there as well.

Perfect Lethality at State level

The second question we are concerned with - does the perfect lethality event ratio vary by state? If so, is this due to a higher frequency of confrontations or some other state-specific factor?

To do so, we wrangled a dataset into a data frame which shows the number of perfect lethality and non-perfect lethality events by state code.

(1) We use the variable "perfect.lethality" provided in the data set. The "perfect.lethality" variable is a dummy variable of 0 and 1; 1 indicates a perfect lethality event and 0 indicates a non-perfect lethality event at the time point of the event in the dataset. A possible limitation is that the variable might miss out some perfect lethality events e.g. when some individuals who are severely wounded during the confrontation are later classified as dead after wounded.

```
# sort total counts of perfect lethality and non-perfect lethality events by state
lethality_index <- AllData %>%
  group_by(state_code, state, perfect.lethality) %>%
  dplyr::summarise(n = n()) %>%
  arrange(desc(perfect.lethality))
```

(2) Once we get the subset, we divide this dataset into two groups of data fields. The first group covers states with perfect lethality events and the second group covers states with non-perfect lethality events.

```
# top 5 states with highest perfect lethality events
p_lethality_index <- lethality_index %>%
  filter(perfect.lethality == 1)  %>%
  arrange(desc(n))

head(p_lethality_index, n=5)    # View top 5 states with highest perfect lethality events
```

```
## Source: local data frame [5 x 4]
## Groups: state_code, state [5]
##
##   state_code                       state perfect.lethality     n
##        <dbl>                       <chr>             <dbl> <int>
## 1         28                  Tamaulipas                 1   386
## 2         19            Nuevo Le<U+00F3>n                 1   202
## 3         12                    Guerrero                 1   122
## 4         16 Michoac<U+00E1>n de Ocampo                 1    98
## 5          8                   Chihuahua                 1    74
```

- Top 5 states with highest perfect lethality event total sum of counts
  State code: 28, 19, 12, 16, 8
  State name: Tamaulipas, Nuevo Leon, Guerrero, Michoacan de Ocampo, Chihuahua

```
# top 5 states with highest average non-perfect lethatlity ratio
np_lethality_index <- lethality_index %>%
  filter(perfect.lethality == 0)  %>%
  arrange(desc(n))

head(np_lethality_index, n = 5)          # View top 5 states with highest non-perfect lethality events
```

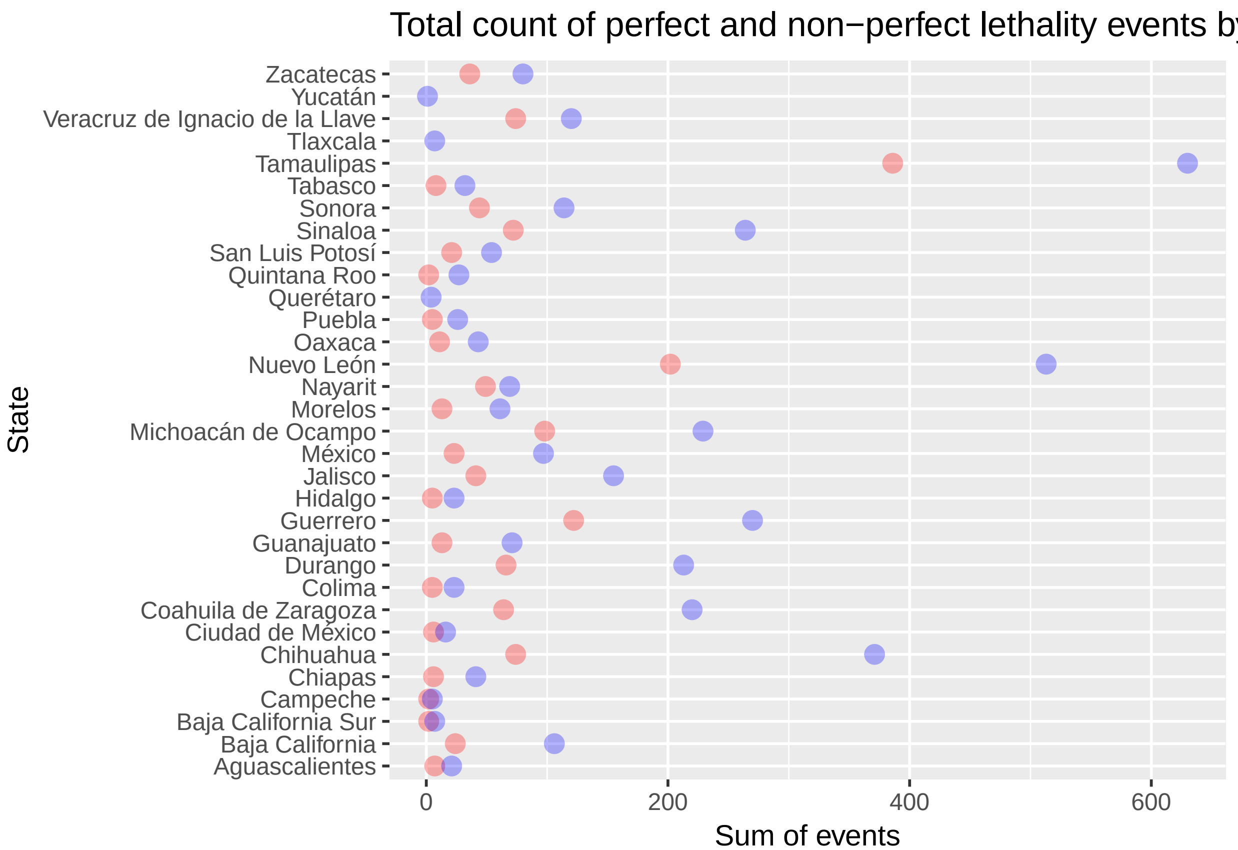
```
## Source: local data frame [5 x 4]
## Groups: state_code, state [5]
##
##   state_code             state perfect.lethality     n
##        <dbl>             <chr>             <dbl> <int>
## 1         28        Tamaulipas                 0   630
## 2         19 Nuevo Le<U+00F3>n                 0   513
## 3          8         Chihuahua                 0   371
## 4         12          Guerrero                 0   270
## 5         25           Sinaloa                 0   264
```

- Top 5 states with highest non-perfect lethality event total sum of counts:
  State code: 28, 19, 8, 12, 25
  State name: Tamaulipas, Nuevo Leon, Chihuahua, Guerrero, Sinaloa

We then draw a plot by count of lethality event per state code [Feature 1].

```
ggplot(p_lethality_index) +
  geom_point(aes(x = n, y = state, size = pop), alpha = 0.3, size = 3, color = "red") +
  geom_point(data = np_lethality_index, aes(x = n, y = state, size = pop), alpha = 0.3, size = 3, color
  labs(x = "Sum of events", y = "State") +
  ggtitle("Total count of perfect and non-perfect lethality events by state")
```

```
## Warning: Removed 2 rows containing missing values (geom_point).
```




```
# Red: Perfect lethality events
# Blue: Non-perfect lethality events
```

From the above graph, we see that the number of perfect lethality events and non-perfect lethality events varied by state. Also, total sum of perfect lethality events varied by state as well. States with higher perfect lethality events are also likely to have more non-perfect lethality events, possibly suggesting that it is the frequency of confrontations, rather than some other state-specific reason, which results in higher perfect lethality events in each state.

```
logit0 <- glm(perfect.lethality ~ as.factor(state), family = binomial(link = "logit"), data = AllData)
summary(logit0)
```

```
##
## Call:
## glm(formula = perfect.lethality ~ as.factor(state), family = binomial(link = "logit"),
##     data = AllData)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0359  -0.8441  -0.6945   1.3913   2.3126
##
## Coefficients:
##                                                    Estimate Std. Error
## (Intercept)                                        -1.09861    0.43644
## as.factor(state)Baja California                    -0.38677    0.49150
## as.factor(state)Baja California Sur                -0.15415    0.91287
## as.factor(state)Campeche                            0.18232    0.94365
## as.factor(state)Chiapas                            -0.82320    0.61768
```

```
## as.factor(state)Chihuahua                       -0.51352    0.45463
## as.factor(state)Ciudad de M<U+00E9>xico          0.11778    0.64780
## as.factor(state)Coahuila de Zaragoza            -0.13613    0.45896
## as.factor(state)Colima                          -0.42744    0.65875
## as.factor(state)Durango                         -0.07303    0.45861
## as.factor(state)Guanajuato                      -0.59912    0.53055
## as.factor(state)Guerrero                         0.30421    0.44986
## as.factor(state)Hidalgo                         -0.42744    0.65875
## as.factor(state)Jalisco                         -0.23124    0.47044
## as.factor(state)M<U+00E9>xico                   -0.34060    0.49423
## as.factor(state)Michoac<U+00E1>n de Ocampo       0.24986    0.45282
## as.factor(state)Morelos                         -0.44731    0.53272
## as.factor(state)Nayarit                          0.75633    0.47474
## as.factor(state)Nuevo Le<U+00F3>n                0.16660    0.44427
## as.factor(state)Oaxaca                          -0.26469    0.55194
## as.factor(state)Puebla                          -0.55005    0.65493
## as.factor(state)Quer<U+00E9>taro               -12.46745  267.70594
## as.factor(state)Quintana Roo                    -1.50408    0.85294
## as.factor(state)San Luis Potos<U+00ED>           0.15415    0.50657
## as.factor(state)Sinaloa                         -0.20067    0.45624
## as.factor(state)Sonora                           0.14660    0.47114
## as.factor(state)Tabasco                         -0.28768    0.58883
## as.factor(state)Tamaulipas                       0.60873    0.44120
## as.factor(state)Tlaxcala                       -12.46745  202.36687
## as.factor(state)Veracruz de Ignacio de la Llave  0.61519    0.46079
## as.factor(state)Yucat<U+00E1>n                 -12.46745  535.41135
## as.factor(state)Zacatecas                        0.30010    0.48037
##                                                z value Pr(>|z|)
## (Intercept)                                     -2.517   0.0118 *
## as.factor(state)Baja California                 -0.787   0.4313
## as.factor(state)Baja California Sur             -0.169   0.8659
## as.factor(state)Campeche                         0.193   0.8468
## as.factor(state)Chiapas                         -1.333   0.1826
## as.factor(state)Chihuahua                       -1.130   0.2587
## as.factor(state)Ciudad de M<U+00E9>xico          0.182   0.8557
## as.factor(state)Coahuila de Zaragoza            -0.297   0.7668
## as.factor(state)Colima                          -0.649   0.5164
## as.factor(state)Durango                         -0.159   0.8735
## as.factor(state)Guanajuato                      -1.129   0.2588
## as.factor(state)Guerrero                         0.676   0.4989
## as.factor(state)Hidalgo                         -0.649   0.5164
## as.factor(state)Jalisco                         -0.492   0.6230
## as.factor(state)M<U+00E9>xico                   -0.689   0.4907
## as.factor(state)Michoac<U+00E1>n de Ocampo       0.552   0.5811
## as.factor(state)Morelos                         -0.840   0.4011
## as.factor(state)Nayarit                          1.593   0.1111
## as.factor(state)Nuevo Le<U+00F3>n                0.375   0.7077
## as.factor(state)Oaxaca                          -0.480   0.6315
## as.factor(state)Puebla                          -0.840   0.4010
## as.factor(state)Quer<U+00E9>taro                -0.047   0.9629
## as.factor(state)Quintana Roo                    -1.763   0.0778 .
## as.factor(state)San Luis Potos<U+00ED>           0.304   0.7609
## as.factor(state)Sinaloa                         -0.440   0.6601
## as.factor(state)Sonora                           0.311   0.7557
```

```
## as.factor(state)Tabasco                            -0.489   0.6252
## as.factor(state)Tamaulipas                          1.380   0.1677
## as.factor(state)Tlaxcala                           -0.062   0.9509
## as.factor(state)Veracruz de Ignacio de la Llave    1.335   0.1818
## as.factor(state)Yucat<U+00E1>n                     -0.023   0.9814
## as.factor(state)Zacatecas                           0.625   0.5321
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6340.6  on 5393  degrees of freedom
## Residual deviance: 6165.8  on 5362  degrees of freedom
##   (2 observations deleted due to missingness)
## AIC: 6229.8
##
## Number of Fisher Scoring iterations: 12
```

Running a simple logistic regression may also give us an insight on the correlation between states and the probability of a confrontation having perfect lethality. We see that there is no statistically significant relationship (10%) between states and the probability of having a perfect lethality event, except for the baseline state (Aguascalientes) and Quintana Roo.

In the meantime, it would be interesting to investigate in those top five states with the highest perfect lethality events at which time point in time series did the perfect lethality events increase. We can also try and focus on what attributes within the state which may have triggered higher likelihood of perfect lethality event per state.

Q2

Cartridges and Civilian Deaths

Hypothesis:

1. The more cartridges seized, the fewer the amount of civilians deaths and wounded.
2. When the army is present in the confrontation, the fewer the amount of civilians deaths and wounded.
3. In addition, when the army is present in the confrontation, given that it is a military force with stronger power and might, they are able to be more successful than other agencies in seizing cartridges, and therefore resulting in fewer organised crime-induced civilian deaths and wounded.

Models:

$$(civilian.dead + civilian.wounded) = \alpha + \beta_1 cartidges.seized + \beta_2 army + \beta_3 army * cartidges.seized + controls$$

In our model, β_1 indicates the effect of cartridges seized on the number of civilian wounded and dead when there is no army presence. β_2 indicates the effect of army involvement on the number of civilian wounded and dead, when there is no cartridges seized. β_3 indicates the marginal effect of cartridges seized on the number of civilian wounded and dead when the army is present in the confrontation. This relates to the hypothesis we would like to test.

We run a regression without interactions first to see the direction of the relationship between cartridges seized with total civilian deaths and wounded, and that between army presence and total civilian deaths and wounded. We control for the presence of other agencies in the confrontation.

```r
dataset1 <- AllData %>%
  filter(afi == 0 & federal.police == 0 & ministerial.police == 0 & municipal.police == 0 & navy == 0 &
dataset2 <- AllData[-dataset1$event.id, ]

lm1 <- lm(I(civilian.dead + civilian.wounded) ~ cartridge.sezied + army  + afi + navy + federal.police +
summary(lm1)
```

```
##
## Call:
## lm(formula = I(civilian.dead + civilian.wounded) ~ cartridge.sezied +
##     army + afi + navy + federal.police + state.police + municipal.police +
##     ministerial.police + other, data = dataset2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.352 -0.239 -0.213 -0.213 59.209
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)          1.208e-01  9.045e-02   1.335   0.1819
## cartridge.sezied    -8.667e-06  3.392e-05  -0.255   0.7984
## army                 6.794e-02  1.080e-01   0.629   0.5295
## afi                  2.070e-01  3.578e-01   0.579   0.5630
## navy                 2.785e-02  1.435e-01   0.194   0.8461
## federal.police       1.080e-01  9.913e-02   1.089   0.2762
## state.police         1.744e-01  1.000e-01   1.743   0.0814 .
## municipal.police     9.272e-02  9.390e-02   0.987   0.3235
## ministerial.police   1.185e-01  1.070e-01   1.108   0.2679
## other                6.705e-01  1.308e-01   5.124 3.18e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.463 on 2958 degrees of freedom
## Multiple R-squared:  0.01078,    Adjusted R-squared:  0.007775
## F-statistic: 3.583 on 9 and 2958 DF,  p-value: 0.0001905
```

We first filtered out confrontations with 0 agencies involved at all - since in these scenarios, we are not supposed to have any cartridges seized. We then filter out events which only had army participation and no participation from other agencies, in order to see the marginal effect of army presence (on top of other agencies) on cartridges seized on civilians death and wounded. This allows us to see the extra influence the army has when it is present at confrontations with other agencies.

As the first model shows, controlling for other factors, on average, cartridge.sezied is negatively related to the total number of civilians dead and wounded. Surprisingly, with other factors held constant, on average, army presence has a positive relationship with total civilian deaths and wounded, which does not support our hypothesis. However, the coefficients on both cartridge.sezied and army are not statistically significant at the 10% level, indicating that the relationship might be spurious.

We add an interaction term to the original model and explore if there was support for our third hypothesis.

```r
lm2 <- lm(I(civilian.dead + civilian.wounded) ~ cartridge.sezied + army + army * cartridge.sezied + afi
summary(lm2)
```

```
##
## Call:
## lm(formula = I(civilian.dead + civilian.wounded) ~ cartridge.sezied +
```

```
##     army + army * cartridge.sezied + afi + navy + federal.police +
##     state.police + municipal.police + ministerial.police + other,
##     data = dataset2)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -1.346 -0.239 -0.214 -0.214 59.208
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)             1.214e-01  9.051e-02   1.341   0.1800
## cartridge.sezied       -1.241e-05  3.841e-05  -0.323   0.7467
## army                    6.155e-02  1.124e-01   0.548   0.5839
## afi                     2.080e-01  3.579e-01   0.581   0.5611
## navy                    2.919e-02  1.437e-01   0.203   0.8390
## federal.police          1.080e-01  9.915e-02   1.090   0.2760
## state.police            1.737e-01  1.001e-01   1.736   0.0826 .
## municipal.police        9.225e-02  9.394e-02   0.982   0.3262
## ministerial.police      1.181e-01  1.070e-01   1.103   0.2700
## other                   6.704e-01  1.309e-01   5.123  3.2e-07 ***
## cartridge.sezied:army   1.697e-05  8.172e-05   0.208   0.8355
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.463 on 2957 degrees of freedom
## Multiple R-squared:  0.0108, Adjusted R-squared:  0.007454
## F-statistic: 3.228 on 10 and 2957 DF,  p-value: 0.0003781
```

From the results, we see that controlling for other factors, on average, when there is no army presence (vs. army presence), each additional cartridge seized is associated with 1.241e-05 fewer civilian deaths and wounded. However, the coefficient is not statistically significant at the 10% level.

Controlling for other factors, on average, when there are no cartridges seized, army presence (vs. no army presence) is associated with 6.155e-02 more civilian deaths and wounded. However, the coefficient is not statistically significant at the 10% level. Again, this is similar to the finding from the previous model.

Controlling for other factors, on average, when the army is present (vs. no army presence), each additional cartridge seized results in 1.697e-05 more civilian deaths and wounded i.e. ultimately, each additional cartridge seized in a confrontation with army presence results in 4.56e-06 more (-1.241e-05 + 1.697e-05) civilian deaths or wounded. However, the coefficient is not statistically significant at the 10% level.

```r
beta <- coef(lm2)
varcov <- as.matrix(vcov(lm2))
var <- diag(varcov)

cartridge.sezied <- mean(dataset2$cartridge.sezied)
mfx.1 <- as.numeric(beta["army"]) + as.numeric(beta["cartridge.sezied:army"]) * cartridge.sezied
mfx.1.se <- sqrt(var["army"] + cartridge.sezied ^ 2 * var["cartridge.sezied:army"] + 2 * cartridge.sezie
c(mfx.1, mfx.1.se)
```

```
##                  army
## 0.06355067 0.11010360
```

Assuming the mean number of cartridges were seized in each confrontation, we calculated the total effect of army presence in the confrontation. Army presence (vs. no army presence) is positively associated with 0.064 total civilian deaths and wounded, with a standard error of 0.110. This refutes our second hypothesis that army presence leads to a lower number of civilian deaths and wounded.

Conclusions & Limitations:

We do not find support for our hypotheses. However, the coefficients in the regression are statistically insignificant (likely due to the limited number of cases), and as such might potentially be spurious.

Our assumptions are that the cartridges, if not seized, would have been used by criminals to kill or wound civilians. In addition, we assume that events with and without army presence are similar in magnitude (i.e. similar number of people involved, including civilians and organised criminals etc.), such that the marginal effect of army presence can be causally determined. However, we feel that this is unlikely to be the case in real life, given that army presence likely signifies that the confrontation is bigger in magnitude. This would have explained the opposite association we found between army presence and total civilian deaths and wounded.

Teamwork Effect between Police and Military in Achieving Perfect Lethality

Hypothesis:

1. When the federal police is present in the confrontation, there is a higher probability of perfect lethality.
2. When the army is present in the confrontation, there is a higher probability of perfect lethality.
3. When there are more long guns seized, there is a lower probability of perfect lethality - since there is a higher level of combat involved. In this case, we assume that long guns seized is a proxy for the level of combat intensity in the confrontation.
4. When both the army and federal police are present at the confrontation, we believe that there is an additional 'teamwork effect' i.e. even higher probability of perfect lethality. This is the marginal effect we are interested in studying.
5. When both the army and federal police are present at the confrontation, we believe that additional long guns seized i.e. higher combat intensity, will result in lower probability of perfect lethality, since the 'teamwork effect' will diminish under chaotic conditions.

Models:

$$(perfect.lethality) = \alpha + \beta_1 army + \beta_2 federal.police + \beta_3 long.guns.seized + \beta_4 army * federal.police +$$

$$\beta_5 army*long.guns.seized+\beta_6 federal.police*long.guns.seized+\beta_7 army*federal.police*long.guns.seized+controls$$

In this case, the coefficient on the interaction term between army and federal police, β_4, indicates the 'teamwork effect'. The coefficient on the interaction term with the three variables, β_7, represents the diminished 'teamwork effect' from higher combat intensity.

First, we run a logit model with no interaction included to see the effect direction of the predictors.

```r
logit1 <- glm(perfect.lethality ~ army + federal.police + long.guns.seized + afi + navy + state.police +
summary(logit1)
```

```
##
## Call:
## glm(formula = perfect.lethality ~ army + federal.police + long.guns.seized +
##     afi + navy + state.police + municipal.police + ministerial.police +
##     other, family = binomial(link = "logit"), data = AllData)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.7650  -0.8171  -0.5733   1.1175   2.6204
##
## Coefficients:
##                      Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)        -0.925565   0.072264 -12.808  < 2e-16 ***
## army                0.429654   0.087079   4.934 8.05e-07 ***
## federal.police     -0.407911   0.119084  -3.425 0.000614 ***
## long.guns.seized    0.079816   0.009261   8.618  < 2e-16 ***
## afi                -0.378899   0.605219  -0.626 0.531280
## navy                0.487924   0.173654   2.810 0.004958 **
## state.police       -0.766579   0.142798  -5.368 7.95e-08 ***
## municipal.police   -1.233460   0.114612 -10.762  < 2e-16 ***
## ministerial.police -0.796841   0.149089  -5.345 9.06e-08 ***
## other              -0.474838   0.194235  -2.445 0.014499 *
## ---
## 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: 5722.5  on 5386  degrees of freedom
## AIC: 5742.5
##
## Number of Fisher Scoring iterations: 4
```

As the results show, controlling for participation of other agencies, on average, army presence and additional long guns seized increases the logit of perfect lethality. On the contrary, federal police presence decreases the logit of perfect lethality. The coefficients of these three variables are statistically significant at the 1% level.

Given this, the directions are not entirely the same as we posited. Federal police presence reduces the probability of perfect lethality, while more long guns seized increases the probability of perfect lethality. Only our second hypothesis regarding army presence is supported.

To further learn about the teamwork effects, we run a second logistic regression with interactions.

```
logit2 <- glm(perfect.lethality ~ army + federal.police + long.guns.seized + army * long.guns.seized * f
summary(logit2)
```

```
##
## Call:
## glm(formula = perfect.lethality ~ army + federal.police + long.guns.seized +
##     army * long.guns.seized * federal.police + afi + navy + state.police +
##     municipal.police + ministerial.police + other, family = binomial(link = "logit"),
##     data = AllData)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2851  -0.8330  -0.5735   1.1409   2.7304
##
## Coefficients:
##                                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)                           -0.88003    0.07518 -11.706  < 2e-16
## army                                   0.42851    0.09442   4.539 5.66e-06
## federal.police                        -0.69958    0.14498  -4.825 1.40e-06
## long.guns.seized                       0.08417    0.01804   4.665 3.09e-06
## afi                                   -0.59081    0.63083  -0.937  0.34899
## navy                                   0.43454    0.17694   2.456  0.01406
## state.police                          -0.83536    0.14512  -5.756 8.59e-09
## municipal.police                      -1.28820    0.11658 -11.050  < 2e-16
## ministerial.police                    -0.84175    0.15000  -5.612 2.00e-08
```

```
## other                              -0.52371    0.19497  -2.686  0.00723
## army:long.guns.seized              -0.02266    0.02117  -1.071  0.28437
## army:federal.police                 0.65542    0.48948   1.339  0.18057
## federal.police:long.guns.seized     0.11422    0.04137   2.761  0.00576
## army:federal.police:long.guns.seized -0.02581    0.07593  -0.340  0.73395
##
## (Intercept)                        ***
## army                               ***
## federal.police                     ***
## long.guns.seized                   ***
## afi
## navy                               *
## state.police                       ***
## municipal.police                   ***
## ministerial.police                 ***
## other                              **
## army:long.guns.seized
## army:federal.police
## federal.police:long.guns.seized    **
## army:federal.police:long.guns.seized
## ---
## 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: 5702.6  on 5382  degrees of freedom
## AIC: 5730.6
##
## Number of Fisher Scoring iterations: 5
```

Controlling for all other factors, on average, when there is no federal police presence and no long guns seized, army presence increases the logit of perfect lethality by 0.429. The coefficient is statistically significant at the 0.1% level.

Controlling for all other factors, on average, when there is no army presence and no long guns seized, federal police presence decreases the logit of perfect lethality by 0.700. The coefficient is statistically significant at the 0.1% level.

Controlling for all other factors, on average, when there is no federal police and no army presence, an additional long gun seized increases the logit of perfect lethality by 0.084. The coefficient is statistically significant at the 0.1% level. This possibly suggests that long guns seized is a measure of weakened enemy power, rather than combat intensity as we have assumed.

Controlling for all other factors, on average, when there is no federal police presence but army presence, an additional long gun seized decreases the logit of perfect lethality by 0.023. However, the coefficient is not statistically significant at the 10% level.

Controlling for all other factors, on average, when there is no army presence but federal police presence, an additional long gun seized increases the logit of perfect lethality by 0.114. The coefficient is statistically significant at the 1% level.

Controlling for all other factors, on average, when there is no long guns seized, army and federal police presence increases the logit of perfect lethality by 0.655. However, the coefficient is not statistically significant at the 10% level.

Controlling for all other factors, on average, when there is army and federal police presence, an additional long gun seized decreases the logit of perfect lethality by 0.026. The coefficient is not statistically significant at the 10% level.

```
beta_logit <- coef(logit2)
varcov_logit <- as.matrix(vcov(logit2))
se_logit <- sqrt(diag(vcov(logit2)))

long.guns.seized <- mean(AllData$long.guns.seized)
mfx.2 <- as.numeric(beta_logit["army"]) + as.numeric(beta_logit["federal.police"]) + as.numeric(beta_log

mfx.2.se <- mfx.2.se <- sqrt(
  se_logit["army"] ^ 2 + se_logit["federal.police"] ^ 2 + se_logit["army:federal.police"] ^ 2 + se_logi
    2 * varcov_logit["army", "federal.police"] + 2 * varcov_logit["army", "army:federal.police"] + 2 *
    2 * varcov_logit["federal.police", "army:federal.police"] + 2 * varcov_logit["federal.police", "arm
    2 * varcov_logit["army:federal.police", "army:federal.police:long.guns.seized"] * long.guns.seized
)

c(mfx.2, mfx.2.se)
```

```
##                army
## 0.3412817 0.4157074
```

Assuming the mean number of long guns were seized in each confrontation, we calculated the total effect of army and federal police presence in the confrontation. At the combat intensity of mean long guns seized, army presence and federal police presence increases the logit of perfect lethality by 0.341, with a standard error of 0.416. While the standard error is indeed large, this provides some support for our hypothesis that there is indeed the presence of some synergy when both agencies are present at a confrontation.

Conclusion & Limitations:

We assume that long guns seized is a proxy for combat intensity i.e. the more long guns there were seized meant that there were more long guns present in combat, and hence higher intensity. However, this may not be a good proxy because it can also meant that the fighting power of the criminals is reduced when more guns are seized, i.e. less combat intensity. Nonetheless, given the lack of any other variables that could serve as a proxy for combat intensity, we think this is the closest measure.

In addition, we assume that events are similar in magnitude with or without army and/or federal police presence. However, we feel that this is unlikely to be the case in real life. As mentioned before, the more agencies involved in a confrontation, the more likely the confrontation is greater in magnitude.

Even though some of the interaction terms are statistically insignificant (possibly due to the lack of confrontations with both army and federal police involvement), there is some support for our hypotheses. The direction of the 'teamwork effect' is positive in the regression, and that of the marginal effect of combat intensity on the 'teamwork effect' is negative.