Hertie Data Management Final Project

Jeffrey Pu

December 12, 2017

Dataset Selection

For my final project, I decided to analyze the Armed Conflict Location & Event Data dataset for Asia in the year 2016. A very quick look at the summary of the data revealed something informative for my analysis moving forward.

```
summary(acled)
```

```
##       GWNO       EVENT_ID_CNTY      EVENT_ID_NO_CNTY
##  Min.   :750.0   Length:14196       Min.   :    1
##  1st Qu.:750.0   Class :character   1st Qu.: 3550
##  Median :750.0   Mode  :character   Median : 7098
##  Mean   :757.2                      Mean   : 7098
##  3rd Qu.:750.0                      3rd Qu.:10647
##  Max.   :816.0                      Max.   :14196
##
##    EVENT_DATE                         YEAR      TIME_PRECISION
##  Min.   :2016-01-01 00:00:00   Min.   :2016   Min.   :1.000
##  1st Qu.:2016-05-03 00:00:00   1st Qu.:2016   1st Qu.:1.000
##  Median :2016-08-01 00:00:00   Median :2016   Median :1.000
##  Mean   :2016-07-20 20:12:16   Mean   :2016   Mean   :1.058
##  3rd Qu.:2016-10-14 00:00:00   3rd Qu.:2016   3rd Qu.:1.000
##  Max.   :2016-12-31 00:00:00   Max.   :2016   Max.   :3.000
##
##   EVENT_TYPE           ACTOR1          ALLY_ACTOR_1           INTER1
##  Length:14196       Length:14196       Length:14196       Min.   :1.000
##  Class :character   Class :character   Class :character   1st Qu.:5.000
##  Mode  :character   Mode  :character   Mode  :character   Median :6.000
##                                                           Mean   :5.406
##                                                           3rd Qu.:6.000
##                                                           Max.   :8.000
##
##     ACTOR2          ALLY_ACTOR_2          INTER2        INTERACTION
##  Length:14196       Length:14196       Min.   :0.0000   Min.   :10.0
##  Class :character   Class :character   1st Qu.:0.0000   1st Qu.:37.0
##  Mode  :character   Mode  :character   Median :0.0000   Median :60.0
##                                        Mean   :0.8962   Mean   :48.3
##                                        3rd Qu.:1.0000   3rd Qu.:60.0
##                                        Max.   :8.0000   Max.   :80.0
##
##    COUNTRY             ADMIN1             ADMIN2
##  Length:14196       Length:14196       Length:14196
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
```

```
##
##     ADMIN3            LOCATION            LATITUDE        LONGITUDE
##  Length:14196       Length:14196       Min.   : 5.925   Min.   : 62.03
##  Class :character   Class :character   1st Qu.:22.552   1st Qu.: 75.10
##  Mode  :character   Mode  :character   Median :26.433   Median : 77.54
##                                        Mean   :24.728   Mean   : 80.62
##                                        3rd Qu.:30.270   3rd Qu.: 85.51
##                                        Max.   :35.921   Max.   :112.25
##                                        NA's   :24
##  GEO_PRECISION      SOURCE             NOTES              FATALITIES
##  Min.   :1.000   Length:14196       Length:14196       Min.   : 0.0000
##  1st Qu.:1.000   Class :character   Class :character   1st Qu.: 0.0000
##  Median :1.000   Mode  :character   Mode  :character   Median : 0.0000
##  Mean   :1.206                                         Mean   : 0.2072
##  3rd Qu.:1.000                                         3rd Qu.: 0.0000
##  Max.   :3.000                                         Max.   :93.0000
##
```

Checking Fatalities

According to the summary, the vast number of conflict events in Asia in the year 2016 resulted in 0 fatalities, however, the range is quite large, maxing out at 93. To explore this further, I decided to focus specifically on cases that resulted in at least 1 fatality.

```
acled_fatal <- acled %>%
  filter(FATALITIES > 0) %>%
  arrange(desc(FATALITIES))

str(acled_fatal)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    1206 obs. of  25 variables:
##  $ GWNO            : num  770 770 770 770 770 775 750 770 771 775 ...
##  $ EVENT_ID_CNTY   : chr  "711PAK" "306PAK" "1015PAK" "915PAK" ...
##  $ EVENT_ID_NO_CNTY: num  13078 12673 13382 13282 12383 ...
##  $ EVENT_DATE      : POSIXct, format: "2016-08-08" "2016-03-27" ...
##  $ YEAR            : num  2016 2016 2016 2016 2016 ...
##  $ TIME_PRECISION  : num  1 1 1 1 1 1 1 1 1 1 ...
##  $ EVENT_TYPE      : chr  "Violence against civilians" "Violence against civilians" "Violence agains
##  $ ACTOR1          : chr  "Jamaat-ul-Ahrar" "Jamaat-ul-Ahrar" "IS: Islamic State" "LeJ: Lashkar-e-Ja
##  $ ALLY_ACTOR_1    : chr  NA NA NA "IS: Islamic State" ...
##  $ INTER1          : num  3 3 2 3 1 4 3 2 2 1 ...
##  $ ACTOR2          : chr  "Civilians (Pakistan)" "Civilians (Pakistan)" "Civilians (Pakistan)" "Poli
##  $ ALLY_ACTOR_2    : chr  NA "Christian Community" NA "Trainee Policemen" ...
##  $ INTER2          : num  7 7 7 1 2 1 1 7 1 2 ...
##  $ INTERACTION     : num  37 37 27 13 12 14 13 27 12 12 ...
##  $ COUNTRY         : chr  "Pakistan" "Pakistan" "Pakistan" "Pakistan" ...
##  $ ADMIN1          : chr  "Balochistan" "Punjab" "Balochistan" "Balochistan" ...
##  $ ADMIN2          : chr  "Quetta" "Lahore" "Khuzdar" "Quetta" ...
##  $ ADMIN3          : chr  "Quetta" "Lahore" "Khuzdar" "Quetta" ...
##  $ LOCATION        : chr  "Quetta" "Lahore" "Khuzdar" "Quetta" ...
##  $ LATITUDE        : num  30.2 31.5 27.7 30.2 32.9 ...
##  $ LONGITUDE       : num  67 74.3 66.6 67 69.7 ...
##  $ GEO_PRECISION   : num  1 1 2 1 1 1 1 1 1 3 ...
##  $ SOURCE          : chr  "Daily Regional Times" "Daily Times" "The News" "Daily Regional Times" ...
```

```
##  $ NOTES           : chr  "At least 93 people were killed and 117 others injured when a bomb explode
##  $ FATALITIES      : num  93 74 62 60 38 30 29 29 28 28 ...
```

Data Manipulation

Now that I have chosen the subsection of the dataset that I want to work with, I work on getting rid of unnecessary columns, including those that are used for administrative data entry.

```
acled_fatal2 <- acled_fatal %>%
  select(EVENT_DATE, EVENT_TYPE, ACTOR1, ALLY_ACTOR_1, INTER1, ACTOR2, ALLY_ACTOR_2, INTER2, INTERACTIO
```

I am also interested in looking at how the different seasons might affect the presence of fatal, armed conflict, so I transform the date column to reflect the season in which each particular incident took place.

```
d <- as.Date(cut(as.Date(acled_fatal2$EVENT_DATE, "%m/%d/%Y"), "month")) + 32
```

```
## Warning in as.POSIXlt.POSIXct(x, tz = tz): unknown timezone '%m/%d/%Y'
```

```
acled_fatal2$Season <- factor(quarters(d), levels = c("Q1", "Q2", "Q3", "Q4"),
   labels = c("winter", "spring", "summer", "fall"))

acled_fatal3 <- acled_fatal2 %>%
  select(Season, EVENT_TYPE, ACTOR1, ALLY_ACTOR_1, INTER1, ACTOR2, ALLY_ACTOR_2, INTER2, INTERACTION, C
```

Exploratory Data Analysis

Now that the data is prepared, I can do some exploratory data analysis in order to get an intuitive sense of what the big picture looks like with regards to fatal armed conflicts in Asia in 2016. I start by looking at how each country fares in terms of fatal conflicts.

```
country_count <- acled_fatal3 %>%
  group_by(COUNTRY) %>%
  summarize(count = n())
country_count
```

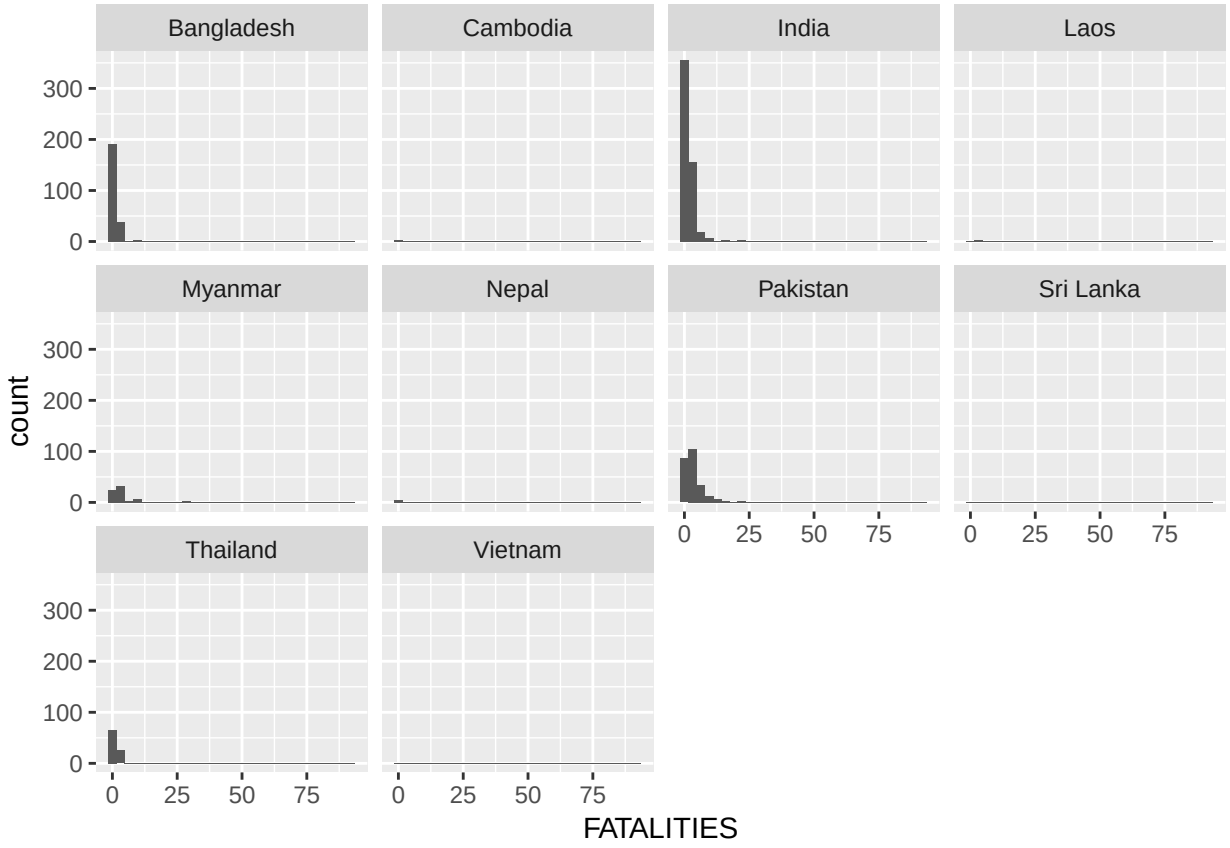
```
## # A tibble: 10 x 2
##       COUNTRY count
##         <chr> <int>
##  1 Bangladesh   236
##  2   Cambodia     4
##  3      India   542
##  4       Laos     3
##  5    Myanmar    67
##  6      Nepal     6
##  7   Pakistan   254
##  8  Sri Lanka     2
##  9   Thailand    91
## 10    Vietnam     1
```

```
country_max <- acled_fatal3 %>%
  group_by(COUNTRY) %>%
  summarize(max = max(FATALITIES))
country_max
```

```
## # A tibble: 10 x 2
##       COUNTRY   max
```

```
##          <chr> <dbl>
##  1 Bangladesh    28
##  2   Cambodia     2
##  3      India    29
##  4       Laos     2
##  5    Myanmar    30
##  6      Nepal     3
##  7   Pakistan    93
##  8  Sri Lanka     2
##  9   Thailand     5
## 10    Vietnam     3
```

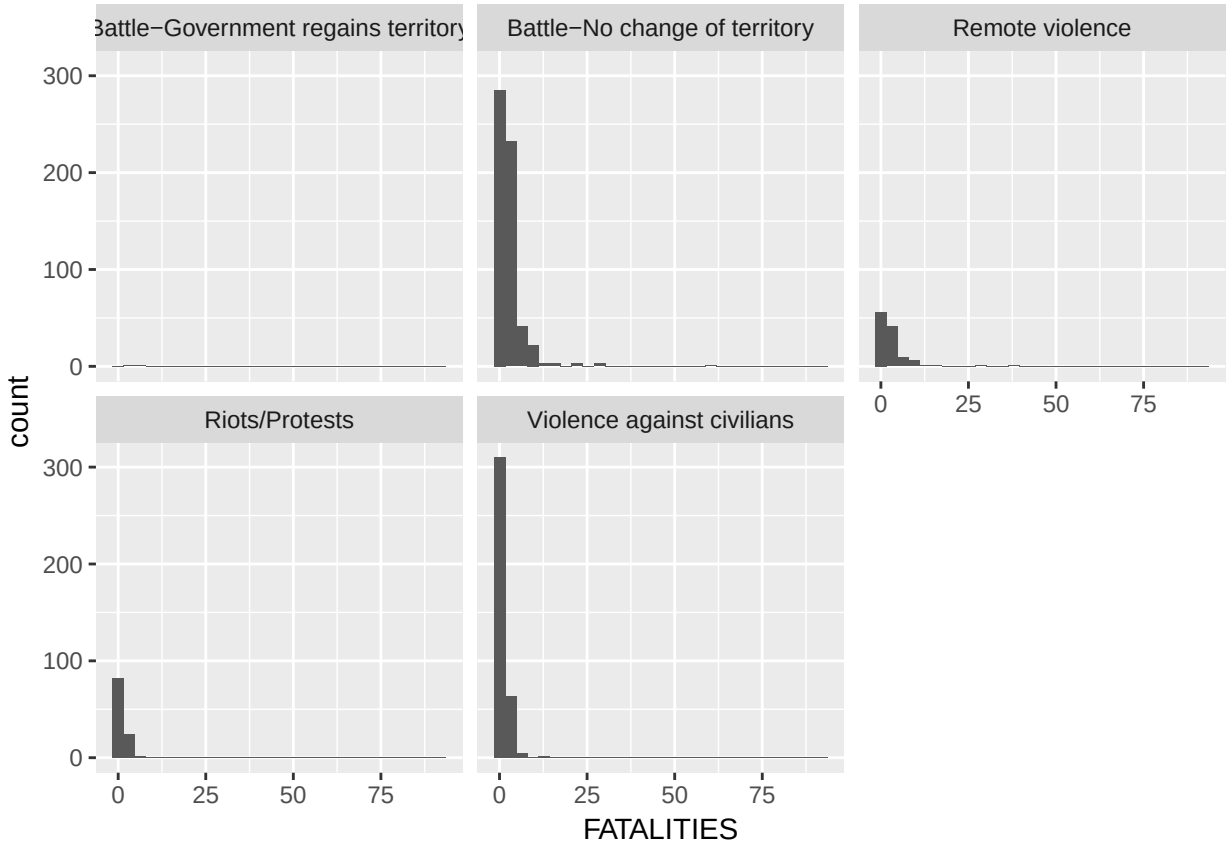
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



According to this graph, we see that India has by far the largest number of incidents that have resulted in at least 1 death, but Pakistan has a lot more deadlier cases.

We can also see if there are any patterns in terms of what type of incidents generate more violence.

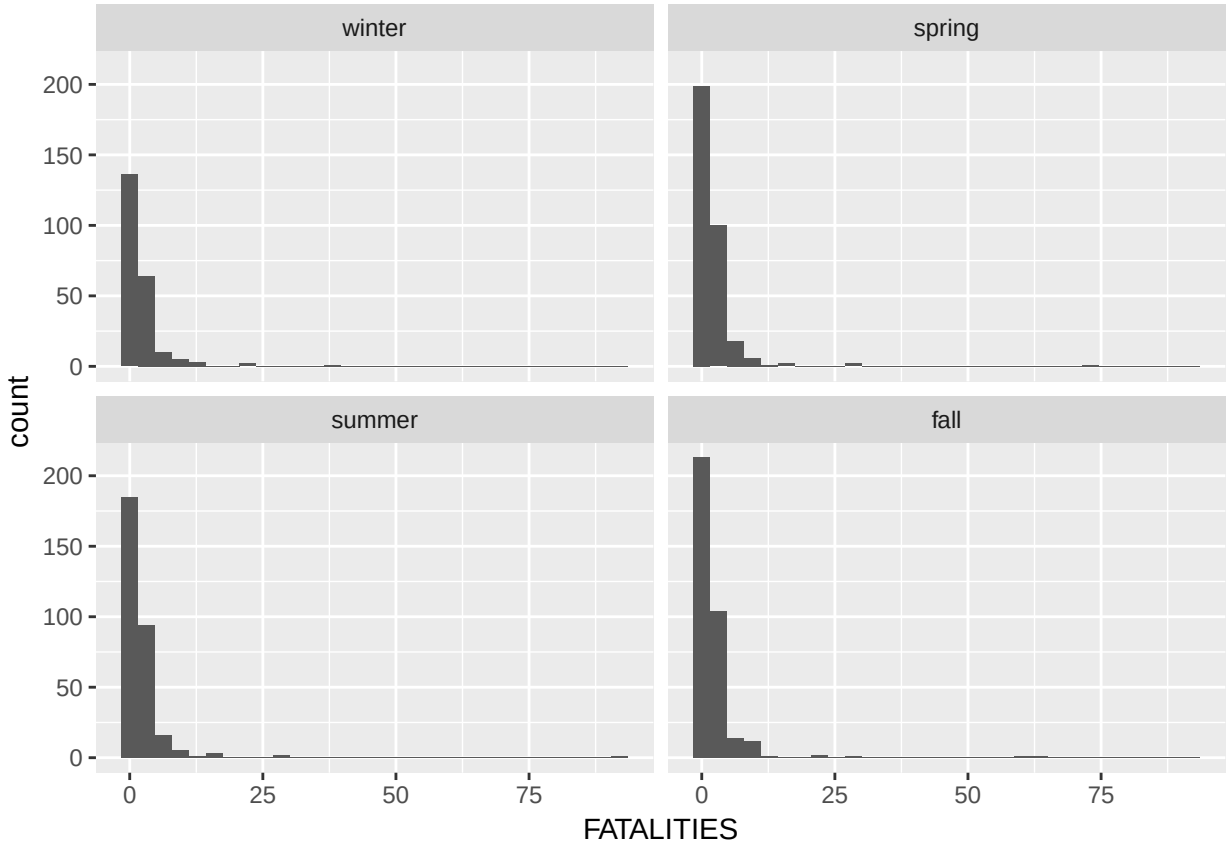
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



According to this chart, we see that battles between armed groups represent the greatest number of incidences of fatal conflicts, however, the most fatal incidences unfortunately involve violence against civilians.

Finally, let's see if there is a correlation between fatal incidents and seasons:

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



From here, it seems that incidences of fatal, armed conflict are pretty well-distributed throughout the year.

Data Analysis

Now that we have a sense of what the data looks like, let's run a regression analysis to see if we can make any statistical inference with regards to the relationship between the factors we've looked at (country, incident type, and season) and the presence of fatal, armed conflict.

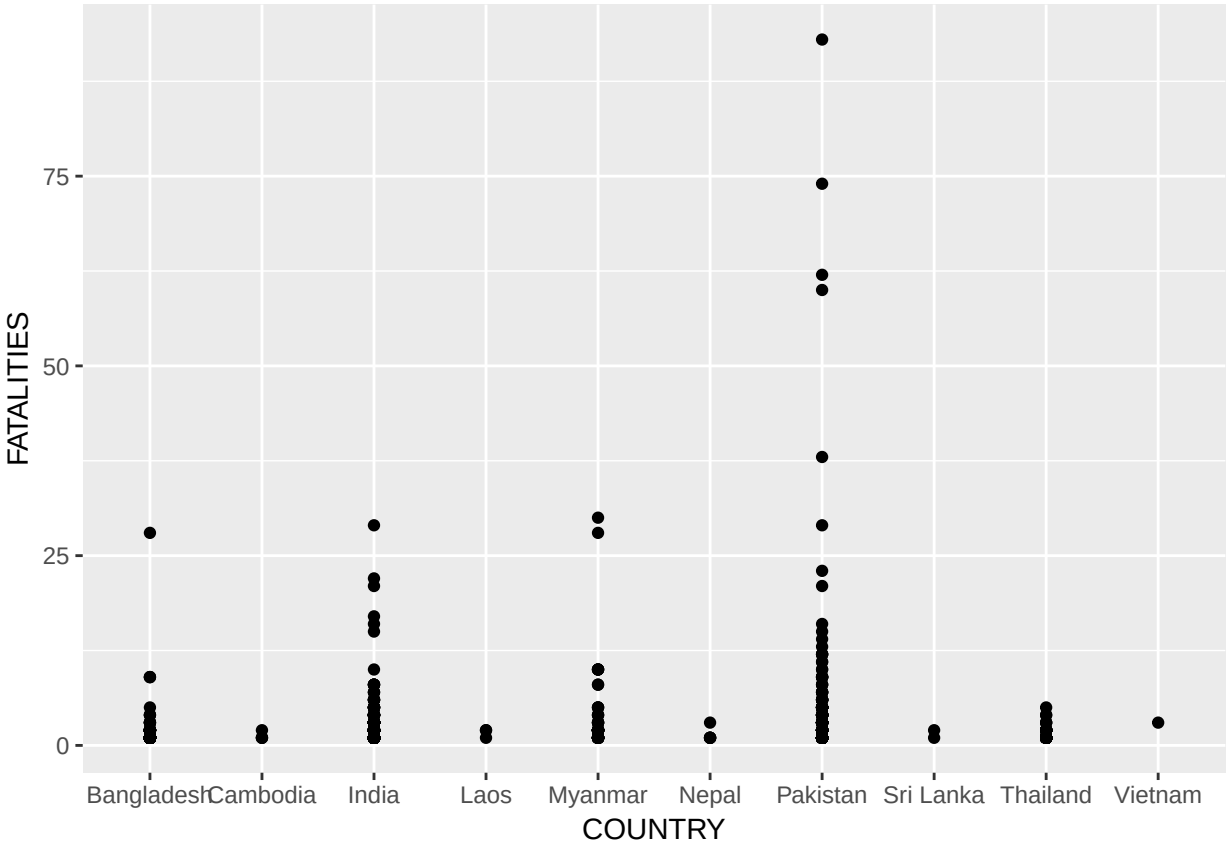
```
acled_analysis <- lm(FATALITIES ~ COUNTRY + EVENT_TYPE + Season, data = acled_fatal3)
summary(acled_analysis)
```

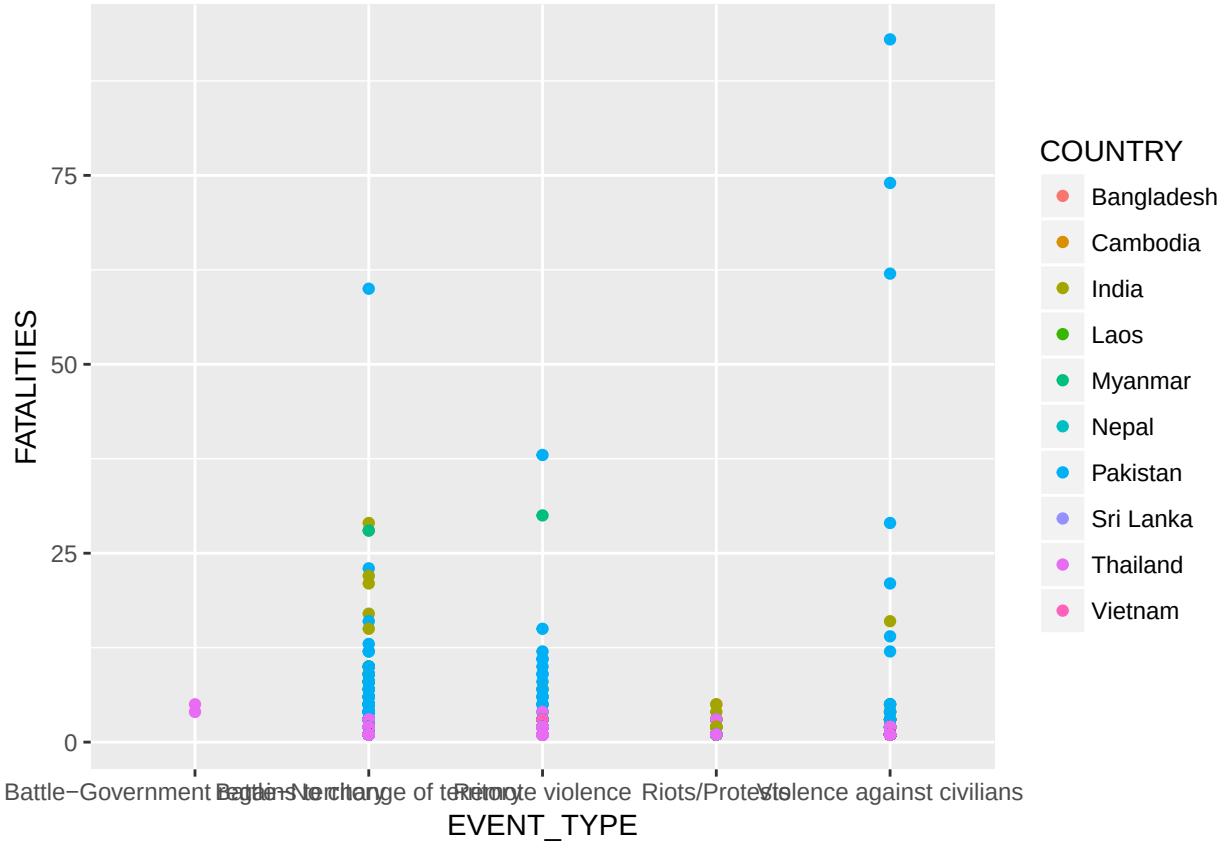
```
##
## Call:
## lm(formula = FATALITIES ~ COUNTRY + EVENT_TYPE + Season, data = acled_fatal3)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.913 -1.046 -0.572  0.034 88.175
##
## Coefficients:
##                                     Estimate Std. Error t value
## (Intercept)                           4.2540     3.6129   1.177
## COUNTRYCambodia                      -0.2616     2.5258  -0.104
## COUNTRYIndia                          0.4795     0.3958   1.211
## COUNTRYLaos                           0.3462     2.9189   0.119
## COUNTRYMyanmar                        1.9238     0.6996   2.750
```

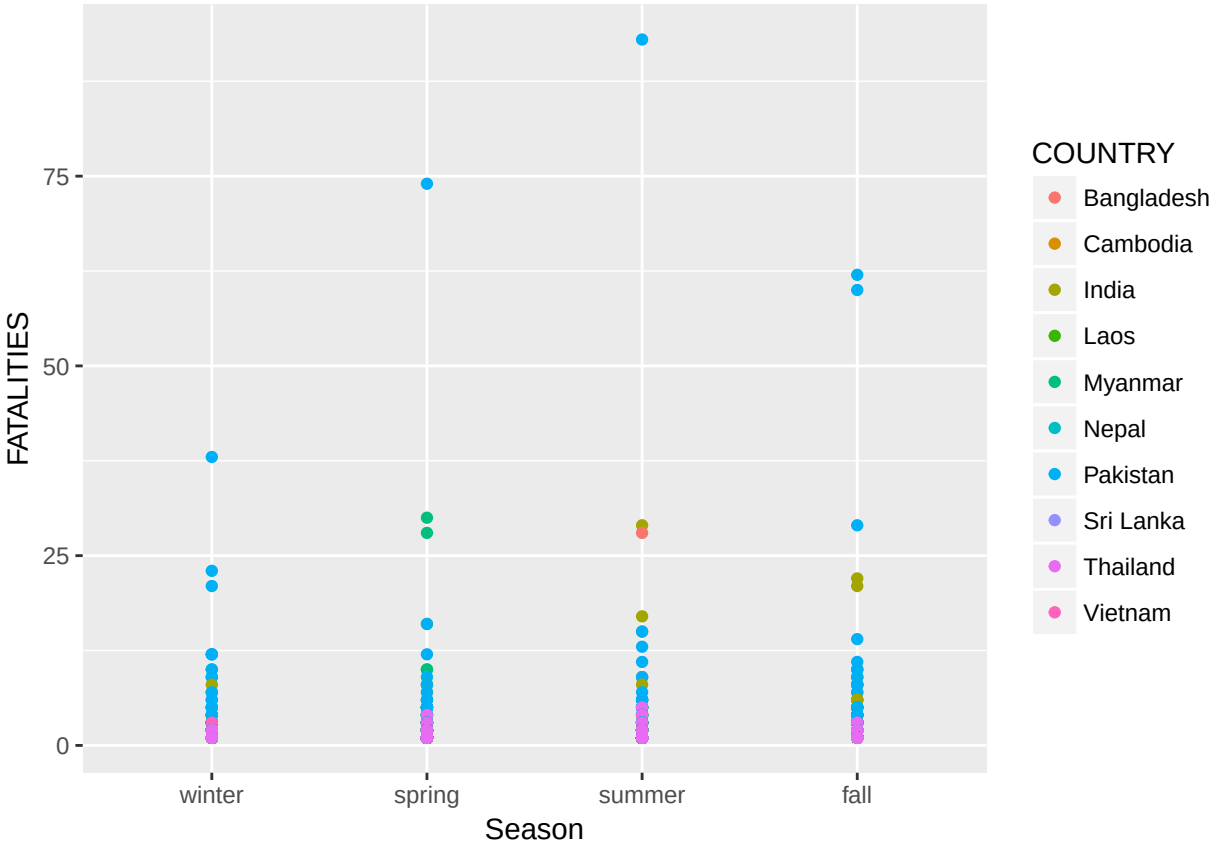
```
## COUNTRYNepal                                0.2363     2.0925   0.113
## COUNTRYPakistan                             3.2531     0.4782   6.803
## COUNTRYSri Lanka                            0.3958     3.5519   0.111
## COUNTRYThailand                            -0.1074     0.6380  -0.168
## COUNTRYVietnam                              1.8482     5.0323   0.367
## EVENT_TYPEBattle-No change of territory    -2.9473     3.5801  -0.823
## EVENT_TYPERemote violence                  -3.1021     3.5938  -0.863
## EVENT_TYPERiots/Protests                   -3.4436     3.6106  -0.954
## EVENT_TYPEViolence against civilians       -3.0358     3.5768  -0.849
## Seasonspring                                0.2598     0.4450   0.584
## Seasonsummer                                0.3534     0.4491   0.787
## Seasonfall                                  0.1798     0.4330   0.415
##                                           Pr(>|t|)
## (Intercept)                                0.23926
## COUNTRYCambodia                            0.91752
## COUNTRYIndia                               0.22596
## COUNTRYLaos                                0.90561
## COUNTRYMyanmar                             0.00605 **
## COUNTRYNepal                               0.91009
## COUNTRYPakistan                           1.61e-11 ***
## COUNTRYSri Lanka                           0.91129
## COUNTRYThailand                            0.86634
## COUNTRYVietnam                             0.71349
## EVENT_TYPEBattle-No change of territory    0.41054
## EVENT_TYPERemote violence                  0.38821
## EVENT_TYPERiots/Protests                   0.34042
## EVENT_TYPEViolence against civilians       0.39619
## Seasonspring                               0.55942
## Seasonsummer                               0.43141
## Seasonfall                                 0.67806
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.983 on 1189 degrees of freedom
## Multiple R-squared:  0.06162,    Adjusted R-squared:  0.04899
## F-statistic:  4.88 on 16 and 1189 DF,  p-value: 8.459e-10
```

According to these summary statistics, it seems like the only statistically significant relationships involve fatalities that occur in Pakistan or Myanmar. These results are likely driven by the ongoing militant and terrorist violence that takes place in Pakistan, and the Rohingya conflict in Myanmar.

However, if we take a look at the R-squared values, we can see that that the model has very little explanatory value. Suffice it to say that there are a lot of factors when it comes to predicting violent conflicts that are not represented in this model, and that understanding the driving forces behind what leads to violent, armed conflicts requires a much more sophisticated model than I am currently able to produce.


FATALITIES
75
50
25
0
Bangladesh
Cambodia
India
Laos
Myanmar
Nepal
Pakistan
Sri Lanka
Thailand
Vietnam
COUNTRY


FATALITIES
75
50
25
0
Battle−Government regains territory
Battle−No change of territory
Remote violence
Riots/Protests
Violence against civilians
EVENT_TYPE
COUNTRY
Bangladesh
Cambodia
India
Laos
Myanmar
Nepal
Pakistan
Sri Lanka
Thailand
Vietnam


FATALITIES
75
50
25
0
winter
spring
summer
fall
Season
COUNTRY
Bangladesh
Cambodia
India
Laos
Myanmar
Nepal
Pakistan
Sri Lanka
Thailand
Vietnam