Terrorism Analysis

Anubhav Shankar

2022-11-13

Dataset Description: The Global Terrorism Database (GTD) is an open-source database including information on terrorist attacks around the world from 1970 through 2017. The GTD includes systematic data on domestic as well as international terrorist incidents that have occurred during this time period and now includes more than 180,000 attacks. The database is maintained by researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism (START), headquartered at the University of Maryland. The dataset can be found here -> https://www.kaggle.com/datasets/START-UMD/gtd?select=globalterrorismdb_0718dist.csv (https://www.kaggle.com/datasets/START-UMD/gtd?select=globalterrorismdb_0718dist.csv) . More information regarding the database overall can be found here -> https://www.start.umd.edu/gtd/ (https://www.start.umd.edu/gtd/)

```
tmp <- read.csv("globalterrorismdb_0718dist.csv")
#names(tmp)
glimpse(tmp)
```

```
## Rows: 181,691
## Columns: 135
## $ eventid            <dbl> 197000000001, 197000000002, 197001000001, 197001000…
## $ iyear              <int> 1970, 1970, 1970, 1970, 1970, 1970, 1970, 1970, 197…
## $ imonth             <int> 7, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …
## $ iday               <int> 2, 0, 0, 0, 0, 1, 2, 2, 2, 3, 1, 6, 8, 9, 9, 10, 11…
## $ approxdate         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ extended           <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ resolution         <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ country            <int> 58, 130, 160, 78, 101, 217, 218, 217, 217, 217, 217…
## $ country_txt        <chr> "Dominican Republic", "Mexico", "Philippines", "Gre…
## $ region             <int> 2, 1, 5, 8, 4, 1, 3, 1, 1, 1, 1, 1, 8, 1, 1, 9, 11,…
## $ region_txt         <chr> "Central America & Caribbean", "North America", "So…
## $ provstate          <chr> "", "Federal", "Tarlac", "Attica", "Fukouka", "Illi…
## $ city               <chr> "Santo Domingo", "Mexico city", "Unknown", "Athens"…
## $ latitude           <dbl> 18.45679, 19.37189, 15.47860, 37.99749, 33.58041, 3…
## $ longitude          <dbl> -69.95116, -99.08662, 120.59974, 23.76273, 130.3963…
## $ specificity        <int> 1, 1, 4, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 5, …
## $ vicinity           <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ location           <chr> "", "", "", "", "", "", "", "Edes Substation", "", …
## $ summary            <chr> "", "", "", "", "", "1/1/1970: Unknown African Amer…
## $ crit1              <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …
## $ crit2              <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, …
## $ crit3              <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1, 0, 1, …
## $ doubtterr          <int> 0, 0, 0, 0, -9, 0, 0, 1, 0, 0, 1, 1, -9, 0, 1, 0, 1…
## $ alternative        <int> NA, NA, NA, NA, NA, NA, NA, 2, NA, NA, 1, 2, NA, NA…
## $ alternative_txt    <chr> "", "", "", "", "", "", "", "Other Crime Type", "",…
## $ multiple           <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ success            <int> 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, …
## $ suicide            <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ attacktype1        <int> 1, 6, 1, 3, 7, 2, 1, 3, 7, 7, 3, 7, 4, 7, 7, 3, 9, …
## $ attacktype1_txt    <chr> "Assassination", "Hostage Taking (Kidnapping)", "As…
## $ attacktype2        <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ attacktype2_txt    <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ attacktype3        <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ attacktype3_txt    <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ targtype1          <int> 14, 7, 10, 7, 7, 3, 3, 21, 4, 2, 4, 4, 6, 2, 1, 2, …
## $ targtype1_txt      <chr> "Private Citizens & Property", "Government (Diploma…
## $ targsubtype1       <int> 68, 45, 54, 46, 46, 22, 25, 107, 28, 21, 27, 28, 42…
## $ targsubtype1_txt   <chr> "Named Civilian", "Diplomatic Personnel (outside of…
## $ corp1              <chr> "", "Belgian Ambassador Daughter", "Voice of Americ…
## $ target1            <chr> "Julio Guzman", "Nadine Chaval, daughter", "Employe…
## $ natlty1            <int> 58, 21, 217, 217, 217, 217, 218, 217, 217, 217, 217…
## $ natlty1_txt        <chr> "Dominican Republic", "Belgium", "United States", "…
## $ targtype2          <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ targtype2_txt      <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ targsubtype2       <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ targsubtype2_txt   <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ corp2              <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ target2            <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ natlty2            <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ natlty2_txt        <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
```

```
## $ targtype3          <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ targtype3_txt      <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ targsubtype3       <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ targsubtype3_txt   <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ corp3              <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ target3            <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ natlty3            <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ natlty3_txt        <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ gname              <chr> "MANO-D", "23rd of September Communist League", "Un…
## $ gsubname           <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ gname2             <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ gsubname2          <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ gname3             <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ gsubname3          <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ motive             <chr> "", "", "", "", "", "To protest the Cairo Illinois …
## $ guncertain1        <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, …
## $ guncertain2        <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ guncertain3        <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ individual         <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ nperps             <int> NA, 7, NA, NA, NA, -99, 3, -99, 1, 1, NA, -99, 1, -…
## $ nperpcap           <dbl> NA, NA, NA, NA, NA, -99, NA, -99, 1, 1, NA, -99, NA…
## $ claimed            <int> NA, NA, NA, NA, NA, 0, NA, 0, 1, 0, NA, 0, NA, 0, 1…
## $ claimmode          <int> NA, NA, NA, NA, NA, NA, NA, NA, 1, NA, NA, NA, NA, …
## $ claimmode_txt      <chr> "", "", "", "", "", "", "", "", "Letter", "", "", "…
## $ claim2             <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ claimmode2         <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ claimmode2_txt     <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ claim3             <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ claimmode3         <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ claimmode3_txt     <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ compclaim          <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ weaptype1          <int> 13, 13, 13, 6, 8, 5, 5, 6, 8, 8, 6, 8, 5, 8, 8, 6, …
## $ weaptype1_txt      <chr> "Unknown", "Unknown", "Unknown", "Explosives", "Inc…
## $ weapsubtype1       <int> NA, NA, NA, 16, NA, 5, 2, 16, 19, 20, 16, 19, 4, 19…
## $ weapsubtype1_txt   <chr> "", "", "", "Unknown Explosive Type", "", "Unknown …
## $ weaptype2          <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 5, …
## $ weaptype2_txt      <chr> "", "", "", "", "", "", "", "", "", "", "", "", "Fi…
## $ weapsubtype2       <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 3, …
## $ weapsubtype2_txt   <chr> "", "", "", "", "", "", "", "", "", "", "", "", "Ha…
## $ weaptype3          <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ weaptype3_txt      <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ weapsubtype3       <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ weapsubtype3_txt   <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ weaptype4          <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ weaptype4_txt      <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ weapsubtype4       <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ weapsubtype4_txt   <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ weapdetail         <chr> "", "", "", "Explosive", "Incendiary", "Several gun…
## $ nkill              <int> 1, 0, 1, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, NA, …
## $ nkillus            <int> NA, NA, NA, NA, NA, 0, NA, 0, 0, 0, NA, 0, 0, 0, 0,…
## $ nkillter           <int> NA, NA, NA, NA, NA, 0, NA, 0, 0, 0, NA, 0, 0, 0, 0,…
## $ nwound             <dbl> 0, 0, 0, NA, NA, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, NA, …
```

```
## $ nwoundus          <int> NA, NA, NA, NA, NA, 0, NA, 0, 0, 0, NA, 0, 0, 0, 0,…
## $ nwoundte          <int> NA, NA, NA, NA, NA, 0, NA, 0, 0, 0, NA, 0, 0, 0, 0,…
## $ property          <int> 0, 0, 0, 1, 1, 1, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0, …
## $ propextent        <int> NA, NA, NA, NA, NA, 3, NA, 3, 3, 3, 3, 3, NA, 3, 2,…
## $ propextent_txt    <chr> "", "", "", "", "", "Minor (likely < $1 million)", …
## $ propvalue         <dbl> NA, NA, NA, NA, NA, NA, NA, 22500, 60000, NA, 0, 30…
## $ propcomment       <chr> "", "", "", "", "", "", "", "Three transformers wer…
## $ ishostkid         <int> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ nhostkid          <int> NA, 1, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ nhostkidus        <int> NA, 0, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ nhours            <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ ndays             <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ divert            <chr> "", "", "", "", "", "", "", "", "", "", "", "", "Be…
## $ kidhijcountry     <chr> "", "Mexico", "", "", "", "", "", "", "", "", "", "…
## $ ransom            <int> 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, …
## $ ransomamt         <dbl> NA, 8e+05, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, …
## $ ransomamtus       <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ ransompaid        <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ ransompaidus      <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ ransomnote        <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
## $ hostkidoutcome    <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, 2, …
## $ hostkidoutcome_txt <chr> "", "", "", "", "", "", "", "", "", "", "", "", "Ho…
## $ nreleased         <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,…
## $ addnotes          <chr> "", "", "", "", "", "The Cairo Chief of Police, Wil…
## $ scite1            <chr> "", "", "", "", "", "\"Police Chief Quits,\" Washin…
## $ scite2            <chr> "", "", "", "", "", "\"Cairo Police Chief Quits; De…
## $ scite3            <chr> "", "", "", "", "", "Christopher Hewitt, \"Politica…
## $ dbsource          <chr> "PGIS", "PGIS", "PGIS", "PGIS", "PGIS", "Hewitt Pro…
## $ INT_LOG           <int> 0, 0, -9, -9, -9, -9, 0, -9, 0, 0, 0, -9, -9, -9, 0…
## $ INT_IDEO          <int> 0, 1, -9, -9, -9, -9, 0, -9, 0, 0, 0, -9, -9, -9, 0…
## $ INT_MISC          <int> 0, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 1, …
## $ INT_ANY           <int> 0, 1, 1, 1, 1, -9, 0, -9, 0, 0, 0, -9, 1, -9, 0, 1,…
## $ related           <chr> "", "", "", "", "", "", "", "", "", "", "", "", "",…
```

We see that this is a huge dataset with a lot of information and from the preliminary view, it looks like a lot of the features/fields/variables are categorical. So, in the next step we'll clean the data a bit.

```
terror$casualties[is.na(terror$casualties)] <- median(terror$casualties,na.rm = T) #Impute the m
edian value into the required numeric field to replace the NA's

sum(is.na(terror$casualties)) #Ensure that the output is 0.It indicates there are no NA's or any
other undesired values
```

```
## [1] 0
```

```
terror$severity <- " " #We need to categorize the attacks basis the casualties racked up
terror$severity <- ifelse(terror$casualties >= 10, "Major", terror$severity) #For Major terroris
t attacks
terror$severity <- ifelse(terror$casualties < 3, "Minor", terror$severity) #For Minor terrorist
 attacks
terror$severity <- ifelse(terror$casualties >= 3 & terror$casualties < 10, "Small", terror$sever
ity) #For small-scale terrorist attacks
table(terror$severity) #Check the pivot output to ensure that there are no blank rows
```
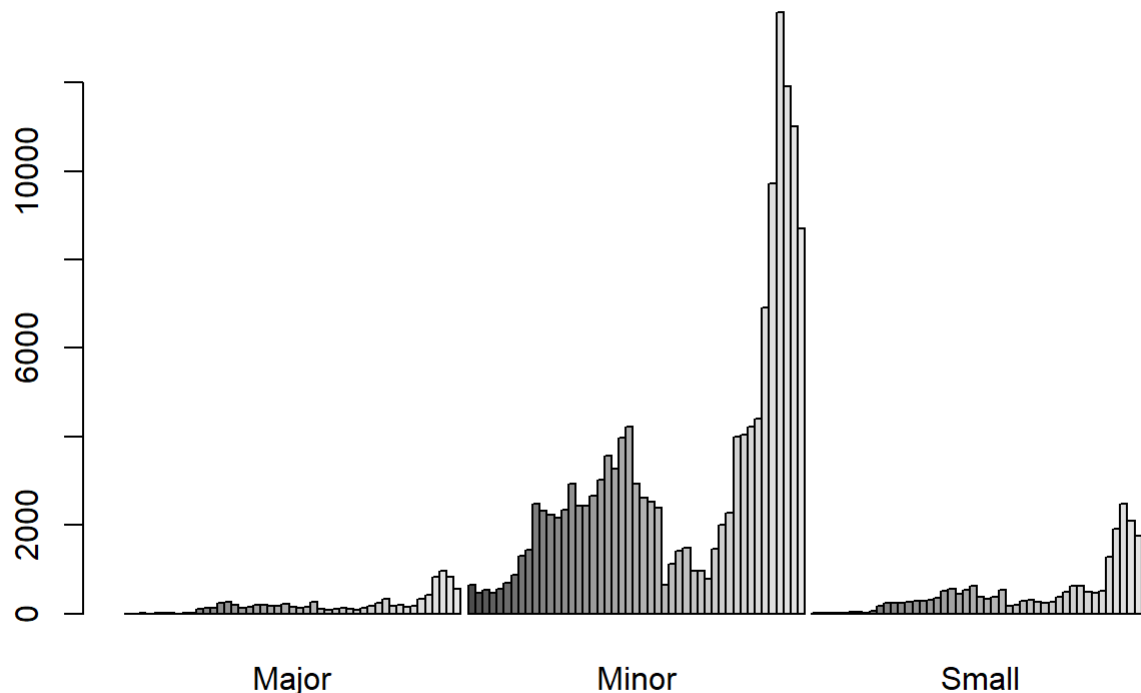
```
##
##  Major  Minor  Small
##   9181 149185  23325
```

```
terror_tb <- table(terror$year,terror$severity) #Pivot the data basis the year and severity
barplot(terror_tb,beside = T, legend.text = F)
```


We now have the yearly breakdown of the terrorist strikes basis the severity. The barplot provides us a bird's eye view of the data and we see that there are a lot of **Minor** attacks which happen quite frequently. However, let's do a deeper dive into the data.
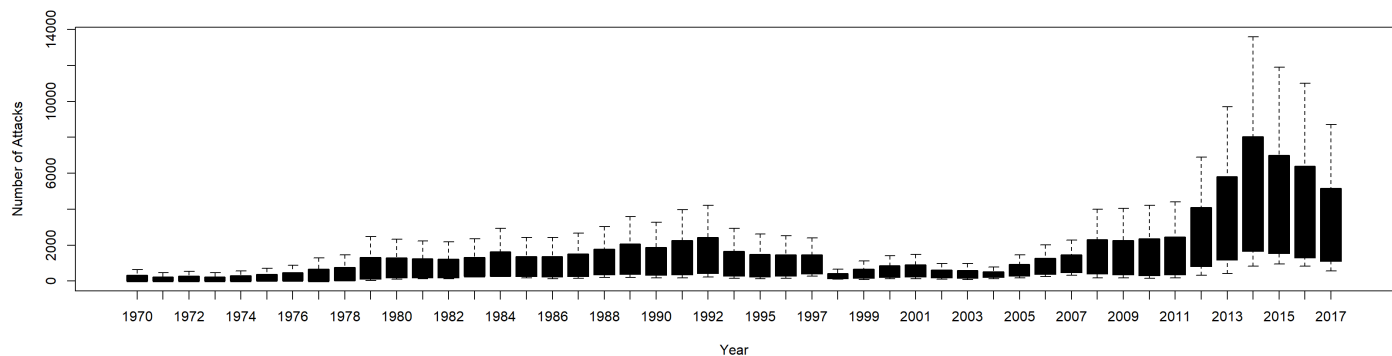
```
sum(is.na(terror_df$year)) #Ensure that the output is 0. It indicates there are no NA's or any o
ther undesired values
```

```
## [1] 0
```

```
sum(is.na(terror_df$nAttacks)) #Ensure that the output is 0.It indicates there are no NA's or an
y other undesired values
```

```
## [1] 0
```

```
plot(terror_df$year,terror_df$nAttacks,col=terror_df$severity,pch=20,xlab="Year",ylab="Number of
Attacks") #Creates a boxplot
```



```
ggplot(terror_df,aes(year,nAttacks,group=severity,fill=factor(severity))) + geom_point(aes(color
=severity)) + scale_color_manual(values = c("red","blue","purple")) + xlab("Year") + ylab("Numbe
r of Attacks") #Scatter plot
```
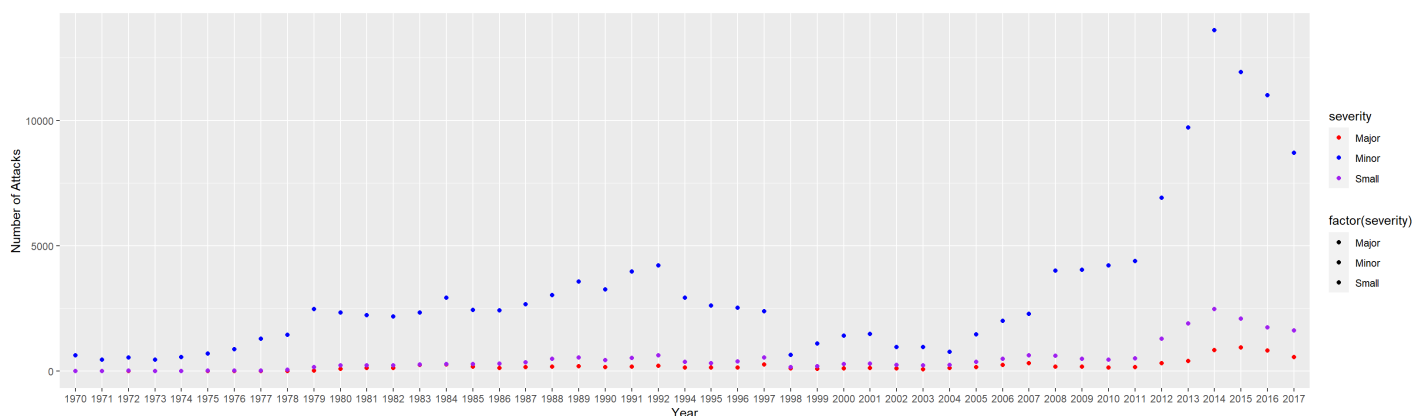

From the above plots we see that the **Minor** attacks seem to ebb & flow over the years with the peak in 2016. However, the **Small** and **Major** attacks seem to have increased during 2014-2017.

Let's do a simple regression analysis to analyse the relationship between number of attacks and year.

```
set.seed(123) #For reproducability
terror_model <- lm(nAttacks~year,data = terror_df) #Regression model
summary(terror_model) #Check output
```

file:///C:/Users/anubh/Desktop/Course Requisites/Fall 2022/MTH 522 - Advanced Mathematical Stats/Homeworks/Project 1/Project1.html

6/8

```
##
## Call:
## lm(formula = nAttacks ~ year, data = terror_df)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
## -4797.3  -847.7  -333.3   647.0  7954.7
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    217.00    1304.96   0.166  0.86829
## year1971       -60.00    1845.49  -0.033  0.97413
## year1972       -27.67    1845.49  -0.015  0.98807
## year1973       -59.33    1845.49  -0.032  0.97442
## year1974       -23.33    1845.49  -0.013  0.98994
## year1975        29.67    1845.49   0.016  0.98721
## year1976        90.67    1845.49   0.049  0.96092
## year1977       222.67    1845.49   0.121  0.90422
## year1978       291.67    1845.49   0.158  0.87476
## year1979       670.33    1845.49   0.363  0.71725
## year1980       670.33    1845.49   0.363  0.71725
## year1981       645.00    1845.49   0.350  0.72750
## year1982       631.00    1845.49   0.342  0.73318
## year1983       739.67    1845.49   0.401  0.68948
## year1984       948.00    1845.49   0.514  0.60868
## year1985       754.67    1845.49   0.409  0.68353
## year1986       736.33    1845.49   0.399  0.69081
## year1987       844.00    1845.49   0.457  0.64849
## year1988      1023.33    1845.49   0.555  0.58055
## year1989      1224.33    1845.49   0.663  0.50869
## year1990      1078.67    1845.49   0.584  0.56029
## year1991      1344.00    1845.49   0.728  0.46826
## year1992      1473.33    1845.49   0.798  0.42668
## year1994       935.00    1845.49   0.507  0.61359
## year1995       810.00    1845.49   0.439  0.66174
## year1996       802.33    1845.49   0.435  0.66474
## year1997       848.67    1845.49   0.460  0.64668
## year1998        94.33    1845.49   0.051  0.95934
## year1999       248.00    1845.49   0.134  0.89339
## year2000       387.67    1845.49   0.210  0.83407
## year2001       418.33    1845.49   0.227  0.82117
## year2002       227.33    1845.49   0.123  0.90223
## year2003       209.00    1845.49   0.113  0.91007
## year2004       171.67    1845.49   0.093  0.92609
## year2005       455.33    1845.49   0.247  0.80566
## year2006       702.33    1845.49   0.381  0.70438
## year2007       863.67    1845.49   0.468  0.64088
## year2008      1384.67    1845.49   0.750  0.45495
## year2009      1356.67    1845.49   0.735  0.46409
## year2010      1391.67    1845.49   0.754  0.45268
## year2011      1475.00    1845.49   0.799  0.42616
## year2012      2623.67    1845.49   1.422  0.15843
```

```
## year2013        3795.00      1845.49    2.056   0.04252 *
## year2014        5417.33      1845.49    2.935   0.00419 **
## year2015        4771.33      1845.49    2.585   0.01126 *
## year2016        4312.00      1845.49    2.337   0.02159 *
## year2017        3416.33      1845.49    1.851   0.06728 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2260 on 94 degrees of freedom
## Multiple R-squared:  0.3198, Adjusted R-squared:  -0.01312
## F-statistic: 0.9606 on 46 and 94 DF,  p-value: 0.5509
```

From the above model it becomes clear that there is no particular relationship between a calendar year and the frequency of terrorist attacks. This is **indicated particularly well by the respective p-values of the variable and the low R-squared value**. However, we do see that **2014 was particularly significant which is consistent with the peak of attack frequencies across the three varying severities**.

We'll now superpose the regression lines over our earlier scatterplot.

```
## `geom_smooth()` using formula = 'y ~ x'
```
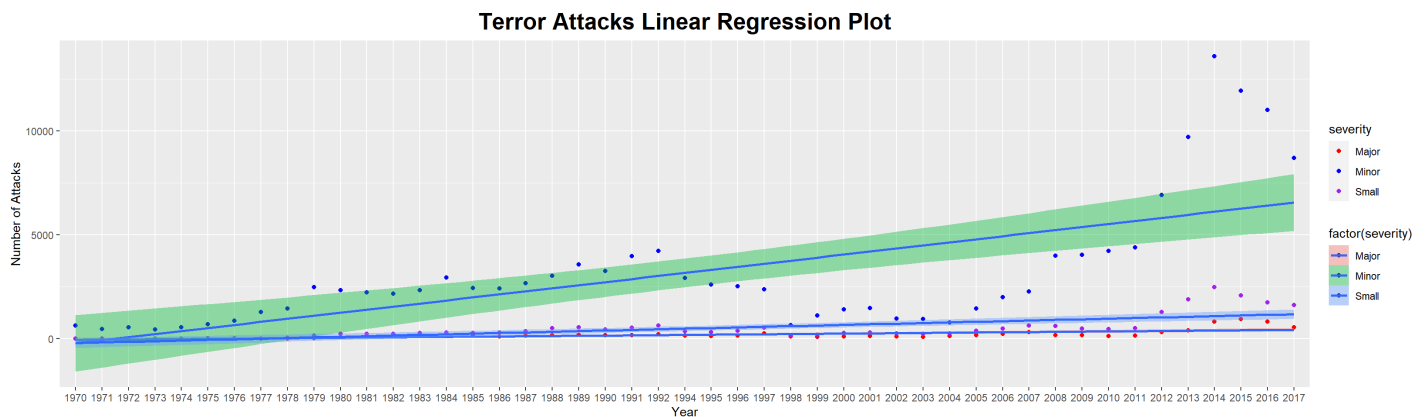

We see that the error margin for **Minor** terrorist attacks is very large. However, not so much for the **Major** and **Small** severities which is quite understandable as the trend for both of them are **fairly constant with low fluctuation**.

References: 1.) https://dplyr.tidyverse.org/ (https://dplyr.tidyverse.org/) 2.) http://courses.atlas.illinois.edu/spring2016/STAT/STAT200/RProgramming/RegressionFactors.html (http://courses.atlas.illinois.edu/spring2016/STAT/STAT200/RProgramming/RegressionFactors.html) 3.) https://statisticsglobe.com/r-error-continuous-value-supplied-to-discrete-scale (https://statisticsglobe.com/r-error-continuous-value-supplied-to-discrete-scale) 4.) https://stackoverflow.com/questions/29587881/increase-plot-size-width-in-ggplot2 (https://stackoverflow.com/questions/29587881/increase-plot-size-width-in-ggplot2)