Forces and Factors Driving Global Terrorism - Insights for a Safer World

Final Report

Applied Statistical Analysis with R
Master of IT in Business
Singapore Management University

G3 - Group 7:
BANDA NIKITHA
CHEE KAH WEN GERALD
GABRIELLA PAULINE DJOJOSAPUTRO
GORDY ADIPRASETYO
JESSICA TAN SOO CHENG
KENDRA LUISA BAYLON GADONG

Contents

1 Introduction 3
2 Methodology 3
2.1 Review of correlated factors that influence terrorist attacks 4
2.2 Theoretical Framework 4
2.3 Data Sources 5
2.3.1 Exploratory Data Analysis 5
2.3.2 Inferential Statistical Analysis 5
2.4 Model Specification 6
2.5 Model Assumptions 6
2.6 Difference with Existing Studies 6
3 Analysis and Discussions 7
3.1 Descriptive Statistics 7
3.1.1 Interactive World Map 7
3.1.2 Interactive Bar Charts 8
3.1.3 Interactive Word Cloud 12
3.2 Inferential Statistics 15
4 Conclusions and Recommendations 18
Appendices 19
Appendix 1 - Linear Regression from Years 2012 – 2018 19
Appendix 2- Linear Regression for Year 2017 20

1 Introduction

Terrorism is a perpetration of criminal act (such as murder, kidnapping, arson, bombing) with the intention to spread fear among the population or to coerce the authorities to take certain actions (or to refrain from it)[1]. This fear has changed our everyday lives, such as how we do airport checks[2] and the lack of garbage bins in Tokyo[3] and London[4]. Understanding terrorism and the factors correlated to it helps us see the attacks more objectively than how the media portrays them in the world today.

This study aims to reveal objective insights on terrorism and its correlated factors. Analysis in this study consists of two parts: descriptive and inferential statistics. Descriptive statistics explains current trends of terrorist attacks through the years, as well as the motives, targets, and terrorist groups responsible. From the initial findings, this study would then identify potential relationships between global terrorism and possible economic, social, or political driving factors. Next, these preliminary findings would then be tested using inferential statistics to determine the validity and significance of such relationships, and the strength of correlation between different variables to provide insights on factors affecting global terrorism.

2 Methodology

The database consisting of all global terrorism attacks from 1970 - 2018 was extracted from Global Terrorism Database (GTD)[5]. The database consists of 135 variables ranging from the date and location of the incident, weapons used, nature of target, number of casualties, and the group or individual responsible. To be included in the GTD, a terrorist attack is a *threatened or actual use of illegal force and violence by a non-state actor to attain a political, economic, religious, or social goal through fear, coercion, or intimidation*. This study has only limited the terrorist attacks for the past 20 years (1998-2018) to be able to extract insights on the most recent attacks, and cumulate the key findings in the form of an interactive RShiny application to highlight the significant relationships to the user. Given the wealth of data available on this topic, the app would feature interactive functionalities such as:

- Interactive World Map: to display the magnitude of the number of attacks for a country with a selection of a range of years
- Interactive Bar Charts: to describe the profile of each country by the targets of the attacks, the types of weapons used, and terrorist groups involved
- Interactive Word Cloud: to identify the motives behind the attacks

From the initial analysis through the descriptive statistics, this study then utilizes the Global Terrorism Index (GTI) which is calculated from the information in the GTD. GTI combines a number of factors to provide an ordinal ranking of countries to build a picture of the impact of terrorism. In a given year, each country has a GTI score which is calculated from the following four factors:

1. Total number of terrorist incidents

[1] United Nations Office on Drugs and Crime (UNODC). (2018). E4J University Module Series: Counter-Terrorism - Module 4: Criminal Justice Responses to Terrorism. Retrieved from https://www.unodc.org/e4j/en/terrorism/module-4/key-issues/defining-terrorism.html.
[2] Leiter, M., Carlin, J., Fong, I., Marcus, D., Vladeck, S. (2012). Ten Years after 9/11: The Changing Terrorist Threat. American University National Security Law Brief, 2(1), 113–147.
[3] Jacobs, P. (2016, April 4). Here's why there are so few trash cans in London. Business Insider. Retrieved from https://www.businessinsider.com/why-few-trash-cans-in-london-2016-4.
[4] Richarz, A. (2019, May 23). Carefully, Japan Reconsiders the Trash Can. Bloomberg CityLab. Retrieved from https://www.bloomberg.com/news/articles/2019-05-23/where-are-all-the-trash-cans-in-japanese-cities.
[5] Global Terrorism Database. Retrieved October 1, 2020 from https://www.start.umd.edu/gtd/access/

2. Total number of fatalities caused by terrorism
3. Total number of injuries caused by terrorism
4. The approximate level of total property damage from terrorist incidents

The Global Terrorism Index (GTI) is used as the aggregate variable to describe the extent of terrorism in a country and acts as the dependent variable for Multiple Linear Regression (MLR). From the review of related literature, economic, social, and political factors are key determinants of terrorism.

2.1 Review of correlated factors that influence terrorist attacks

This study explores possible correlated factors that can influence terrorist attacks. Some researchers believe that economic variables such as a country's GDP and their involvement in international trade have a negative correlation with the terrorist attacks. A country's economic development and economic growth, like inflation, enable salary rises and improve education, transportation services, health care, and the like[6], so it would less likely lead to a potential terrorist attack. Additionally, the gradual increase of the gap between the rich and the poor impacted by population can lead to a discontent among citizens and potentially lead to a growing number of terrorist attacks. Similarly, social factors such as growing unemployment rates, life satisfaction can deepen social conflicts[7], and impact terrorism. Internet also has enabled activities involving terrorism[8], such as propaganda, planning, financing, and execution. Lastly, motives of terrorist attacks reveal that the government or the failure of effective leadership is a key determinant of an attack. A country's level of corruption, democracy, and government efficiency are also explored for political factors influencing terrorism.

2.2 Theoretical Framework

Figure 1: Variables Included in the Multiple Linear Regression Model

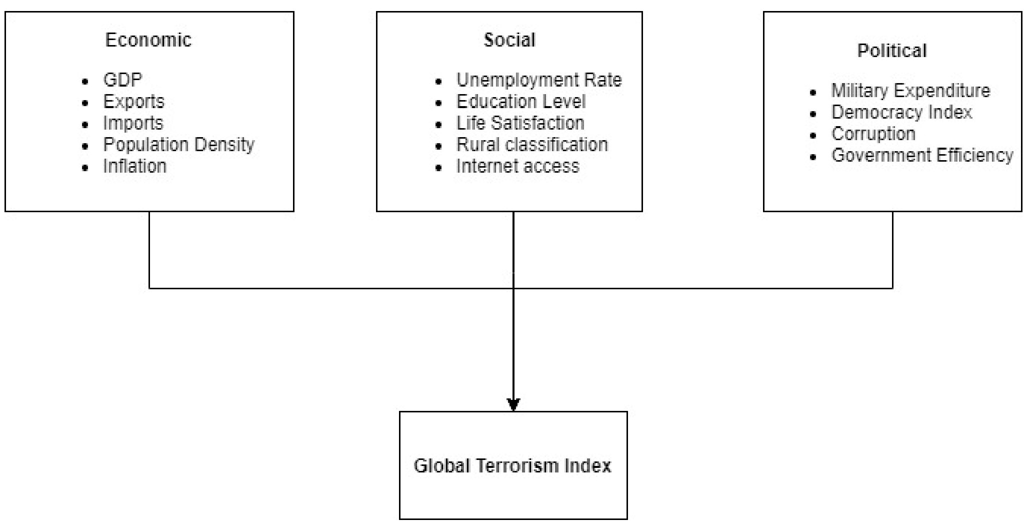


Figure 1 summarizes the economic, social, and political variables included in the initial model of the regression to describe the global terrorism index.

---

[6] She, S., Wang, Q., & Weimann-Saks, D. (2019). Correlation factors influencing terrorist attacks: Political, social or economic? A study of terrorist events in 49 "Belt and Road" countries. Quality & Quantity, 54(1), 125-146. doi:10.1007/s11135-019-00946-x

[7] Ibid.

[8] UNODC (2012). The use of the Internet for terrorist purposes. https://www.unodc.org/documents/terrorism/Publications/Use_of_Internet_for_Terrorist_Purposes/ebook_use_of_the_internet_for_terrorist_purposes.pdf

2.3 Data Sources

2.3.1 Exploratory Data Analysis

Thirty-four (34) reduced variables from the Global Terrorism Database (https://www.start.umd.edu/gtd/access/) are included for the analysis.

2.3.2 Inferential Statistical Analysis

The table below lists all economic, social, and political data used to build predictive regression model in this study, their definitions, and sources.

Table 1: Data Description and Data Sources

Data Label	Data Description	Data Source
GTI	Global Terrorism Index	http://visionofhumanity.org/indexes/terrorism-index/
GDP	Gross Domestic Products of a country	https://data.worldbank.org/indicator/NY.GDP.MKTP.CD
Exports	Exports of a country expressed as percentage of GDP	https://data.worldbank.org/indicator/NE.EXP.GNFS.ZS
Imports	Imports of a country expressed as percentage of GDP	https://data.worldbank.org/indicator/NE.IMP.GNFS.ZS
PopDen	Population Density	https://data.worldbank.org/indicator/EN.POP.DNST
MiltExp	Military expenditure expressed as percentage GDP	https://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS?view=chart
LifeSatis	Life satisfaction in Cantril Ladder (World Happiness Report 2019)	https://worldhappiness.report/archive/
UnEmp	Unemployment Rate	https://data.worldbank.org/indicator/SL.UEM.TOTL.ZS
Dem	Democracy Index	https://www.eiu.com/public/topical_report.aspx?campaignid=democracyindex2019
HCI	Human Capital Index measures productivity relative to a benchmark of complete education and full health, and ranges from 0 to 1.	https://datacatalog.worldbank.org/dataset/human-capital-index
Internet	Fixed broadband subscriptions (per 100 people)	https://data.worldbank.org/indicator/IT.NET.BBND.P2
Inflation	Consumer Price Index	https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG
Rural	Rural population (% of total population)	https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS
Corruption	Control of Corruption; Estimate of governance (ranges from approximately -2.5 (weak) to 2.5 (strong) governance performance) Reflects perceptions of the extent to which public power is exercised for private gain, including both petty and grand forms of corruption, as well as "capture" of the state by elites and private interests.	http://info.worldbank.org/governance/wgi/
GovtEff	Government Efficiency; Reflects perceptions of the quality of public services, the quality of the civil service and the degree of its independence from political pressures, the quality of policy formulation and implementation, and the credibility of the government's commitment to such policies.	http://info.worldbank.org/governance/wgi/

2.4 Model Specification

The proposed model is a multiple linear regression with Global Terrorism Index as the dependent variable and 14 quantitative variables as independent variables. Considering that this study wants to provide insights on the factors affecting terrorism for a country. The study tests three different models: (1) years 2012 – 2018 data, (2) 2017 data, and (3) 2018 data with lag by one period for the independent variables. The model specification ca be summarized in Formula 1.

$$GTI = \beta_O + \beta_1 GDP + \beta_2 Exports + \beta_3 Imports + \beta_4 PopDen + \beta_5 Inflation + \beta_6 UnEmp + \beta_7 HCI + \beta_8 LifeSatis + \beta_9 Rural + \beta_{10} Internet + \beta_{11} MiltExp + \beta_{12} Dem + \beta_{13} Corruption + \beta_{14} GovtEff + \varepsilon$$

2.5 Model Assumptions

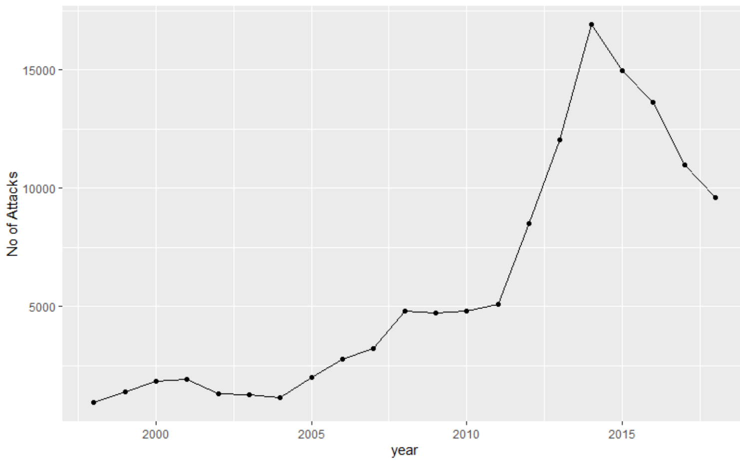
- Countries and years with incomplete information on some variables were omitted from the observations.
- All tests were tested with a significance level of 0.05.
- Relationship between terrorism index and the predictor variables are linear and subject to random error.
- Error terms are assumed to be normal, homoscedastic, and uncorrelated unless there is no sufficient evidence to conclude otherwise.
- Assume no linear dependence in the predictor variables unless there is no sufficient evidence to conclude otherwise.
- There is no autocorrelation between observations unless there is no sufficient evidence to conclude otherwise.

2.6 Difference with Existing Studies

There are similar studies conducted in online forums such as Kaggle using similar dataset obtained from GTD. This study is different as, in addition to GTD dataset, we also explore GTI as well as other datasets containing the different economic, political, and social variables from different sources and link the findings of the two. In comparison, other studies that have been done will usually focus on descriptive statistics or choose a different set of independent variables for regression.

3 Analysis and Discussions

Figure 2 & 3: Number of Attacks trend line, Word Cloud of summary of all countries



The following macro statistics are explored using R as a visualization tool before the study zoomed in into key factors to analyze in inferential statistics.

From 1998 to 2018, the line graph shows a big jump in number of attacks year-on-year, except a notable decrease by 2014. Statistically, this major upward trend from 2011–2014 is of interest to determine possible driving factors. The word cloud above also reveals the common motives of the terrorist attack which are political and religious reasons.

3.1 Descriptive Statistics

3.1.1 Interactive World Map

An interactive map was created in shiny using "maps' package which displays the color of the count of attacks that took place in the range of years that the user selects. There exists a legend that show the color scheme of the representation of the number of attacks. One can also hover over any area on the map plot to identify which country it is and the exact number of attacks that took place.

The UI of the application is such that the user selects the range of years. By default, 2014 – 2018 will be selected. It is also possible to select just one year with the sliderInput.

Rshiny script:

```
ui = navbarPage("Global Terrorism", theme=shinytheme("sandstone") ,
                tabPanel(
                  title = "World Map",icon = icon("fas fa-globe") ,
                  # Sidebar panel for inputs
                  sidebarLayout(
                    sidebarPanel(
                      #input: year
                      sliderInput(inputId="iyear","Year",min=1998,max=2018,value=c(2014,2018),sep="",animate=FALSE)
                    ),
                    # Main panel for displaying outputs
                    mainPanel(
                      # Hide errors
                      tags$style(type = "text/css",
                                 ".shiny-output-error { visibility: hidden; }",
                                 ".shiny-output-error:before { visibility: hidden; }"),

                      # Output: interactive world map
                      girafeOutput("distPlot")
                    )

                  )
                ),
```

The server part of the applications takes in the input parameters and calls a function called “worldMaps”. It is in the “worldMaps” function where the input of the user is matched with the previously obtained world map data frame to display only those count of attacks that the user selects.

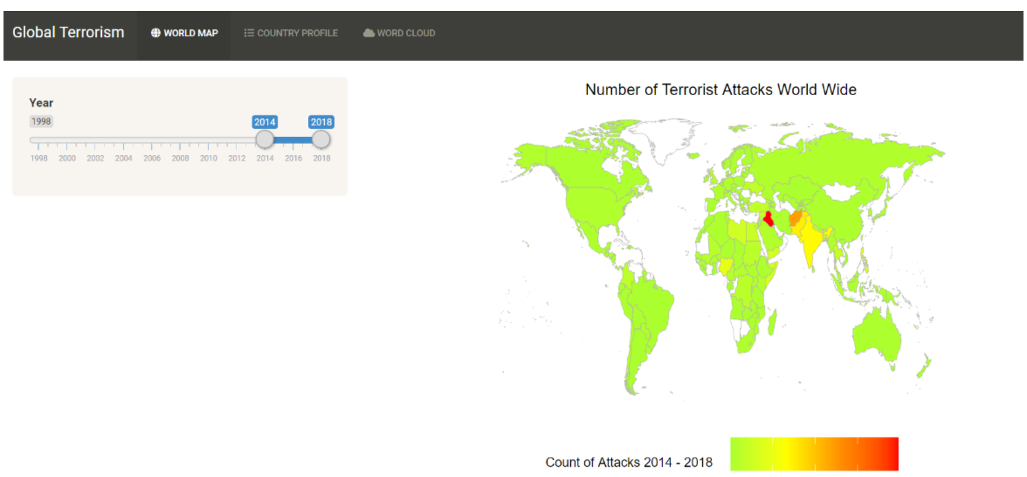
<u>*Rshiny script:*</u>

```
#World map server
# Create the interactive world map
output$distPlot <- renderGirafe({
  ggiraph(code = print(worldMaps(sub_df, world_data, input$iyear[1],input$iyear[2])))
})
```

```
# Select only the data that the user has selected to view
plotdf <- sub_df[sub_df$iyear >= yearStart & sub_df$iyear <= yearEnd ,]
plotdf <- plotdf[!is.na(plotdf$country_txt), ]
df=sqldf(paste("select country_txt,sum(count) as sum_val from plotdf group by country_txt"))

# Add the data the user wants to see to the geographical world data
world_data['Count'] <- df$sum_val[match(world_data$region, df$country_txt)]

if(yearStart!=yearEnd)
{
  yearSpan=paste(yearStart," - ",yearEnd)
}
else{
yearSpan=yearStart
}
# Specify the plot for the world map
library(RColorBrewer)
library(ggiraph)
g <- ggplot() +  ggtitle("Number of Terrorist Attacks World Wide") +
  geom_polygon_interactive(data = subset(world_data, lat >= -60 & lat <= 90), color = 'gray70', size = 0.2,
                           aes(x = long, y = lat, fill = Count, group = group,
                               tooltip = sprintf("%s<br/>%s", region, Count))) +
  scale_fill_gradientn(colours = brewer.pal(5, "Paired"), na.value = 'white',name=paste("Count of Attacks ", yearSpan)) +
  labs(fill = yearStart, color = yearStart, x = NULL, y = NULL) +
  my_theme()

return(g)
}
```

The resultant would then look like this:

<u>Figure 4</u>: RShiny application outcome for “World map”



From the map plot in Figure 4, it is visible that for the years 2014 – 2018, Iraq has had the highest count of terrorist attacks of 11902 followed by Afghanistan with 7268 attacks. (country name and number of attacks can be viewed by hovering mouse over the plot).

3.1.2 Interactive Bar Charts

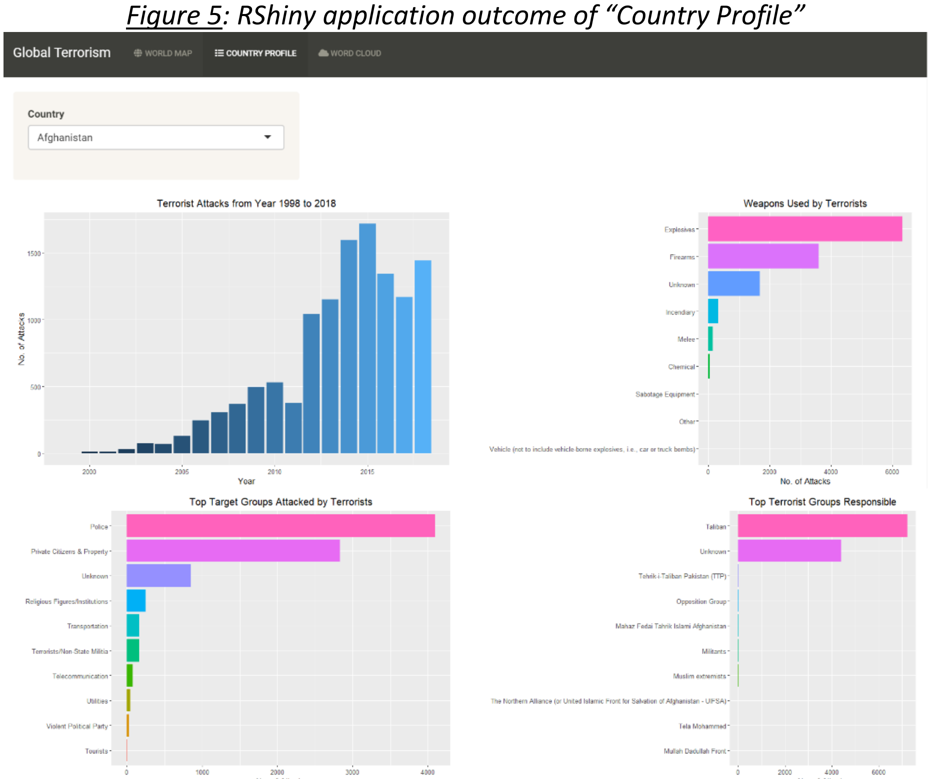
The “**Country Profile**” tab of the application displays interactive bar charts profiling country-level terrorist statistics. By selecting the country name in the selection tab, the user will be able to access and view the individual country level profiles of (i) Terrorist Attacks from Year 1998 to 2018, (ii) Top Terror Targets, (iii) Weapons Used, and (iv) Top Terror Groups in the country.

Rshiny script:

```
tabPanel(
  title = "Country Profile",icon = icon("fas fa-list") ,
  sidebarLayout(
    sidebarPanel(
    selectInput("v_country", "Country", choices = globalterror %>%
                  select(country_txt) %>%
                  distinct() %>%
                  arrange(country_txt) %>%
                  drop_na()),
    width = 4

    ),
  mainPanel(
    fluidRow(box(plotOutput("Year_of_Attacks")), box(plotOutput("Weapons_Used"))),
    fluidRow(box(plotOutput("Terror_Targets")),box(plotOutput("Terrorist_Groups"))),
    width = 12
    )
  )
),
```

Running the Rshiny script will load the following country-level terrorist statistical profiles. In the example below, Afghanistan is selected, and the interactive dashboard generates the bar charts showing the terrorist statistics for Afghanistan:

Figure 5: RShiny application outcome of "Country Profile"



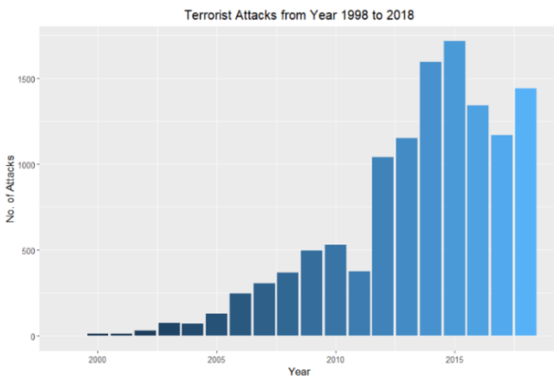
(i) Terrorist Attacks from Year 1998 to 2018

The 1st interactive bar chart on "Terror Attacks from Year 1998 to 2018" plots the number of terror attacks for each selected country for each year from 1998 to 2018:

Rshiny script:

```
#Country Profile server
output$Year_of_Attacks <- renderPlot({
  globalterror %>%
    filter(country_txt == input$v_country) %>%
    select(iyear) %>%
    drop_na() %>%
    count(iyear) %>%
    ggplot(aes(x = iyear, y = n, fill = iyear)) +
    geom_col() +
    ylab("No. of Attacks") + xlab("Year")+
    labs(title = "Terrorist Attacks from Year 1998 to 2018") +
    theme(plot.title = element_text(hjust = 0.5))+
    theme(legend.position = "none")
})
```

Figure 6: Bar Chart distribution of number of terrorist attacks in Afghanistan



For the selected country of Afghanistan, Figure 6 shows the attacks had built up from a low base in year 1998 to more than 1000 attacks per year from year 2012 onwards, which continued unabated to year 2018.

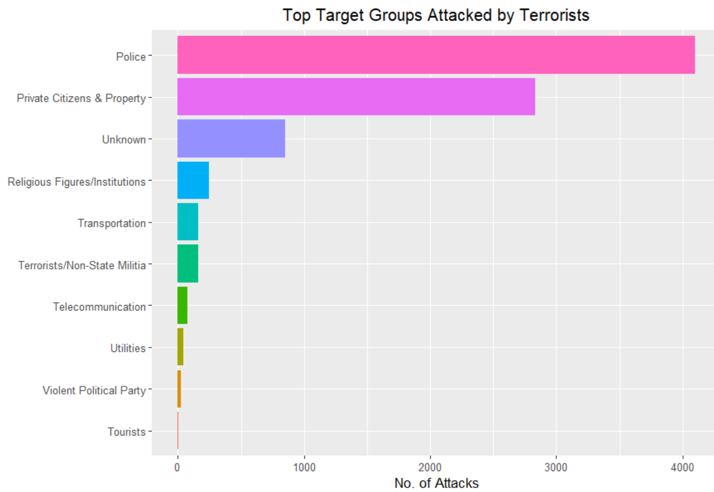
(ii) Top Terror Targets Groups

The 2nd chart shows the "Top Target Groups Attacked by Terrorists" for each selected country over the past 20 years from 1998 to 2018 by number of attacks:

Rshiny script:

```
output$Terror_Targets <- renderPlot({
  globalterror %>%
    filter(country_txt == input$v_country) %>%
    select(targtype1_txt) %>%
    drop_na() %>%
    count(targtype1_txt) %>%
    slice_max(order_by = targtype1_txt,n=10)%>%
    mutate(targtype1_txt = fct_reorder(targtype1_txt, n)) %>%
    ggplot(aes(x = n, y = targtype1_txt, fill = targtype1_txt)) +
    geom_col() +
    ylab("") + xlab("No. of Attacks")+
    labs(title = "Top Target Groups Attacked by Terrorists") +
    theme(plot.title = element_text(hjust = 0.5))+
    theme(legend.position = "none")
})
```

Figure 7: Bar Chart distribution of target groups in Afghanistan



For the selected country of Afghanistan, Figure 7 shows most of the terror attacks were targeted on the Police, which bore the brunt of more than 4000 terrorist attacks over the past 20 years to 2018; followed by Private Citizens & Property, which suffered more than 2700 attacks during this period.

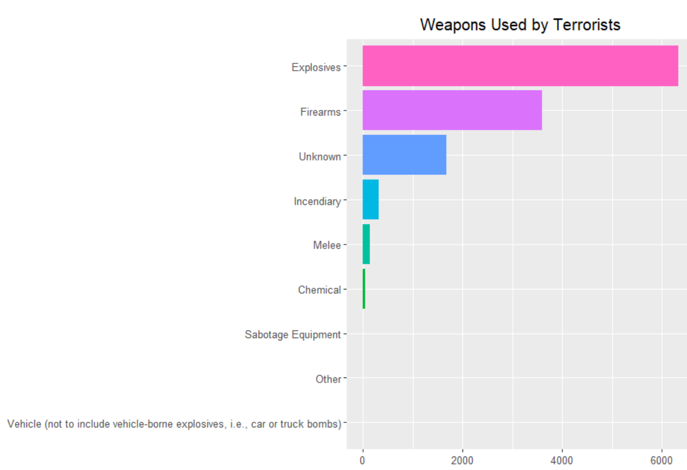
(iii) Weapons Used

The next chart highlights the "Weapons Used by Terrorists" in attacks over the 20 years to 2018 for each selected country:

Rshiny script:

```
output$Weapons_Used <- renderPlot({
  globalterror %>%
    filter(country_txt == input$v_country) %>%
    select(weaptype1_txt) %>%
    drop_na() %>%
    count(weaptype1_txt) %>%
    mutate(weaptype1_txt = fct_reorder(weaptype1_txt, n)) %>%
    ggplot(aes(x = n, y = weaptype1_txt, fill = weaptype1_txt)) +
    geom_col() +
    ylab("") + xlab("No. of Attacks")+
    labs(title = "Weapons Used by Terrorists") +
    theme(plot.title = element_text(hjust = 0.5))+
    theme(legend.position = "none")
})
```

Figure 8: Bar Chart distribution of weapon used by terrorist in Afghanistan



For the selected country of Afghanistan, Figure 8 shows explosives as the key attack weapon, with more than 6000 instances over the 20 years to 2018; followed by firearms, which profiled more than 3500 instances during the same period.

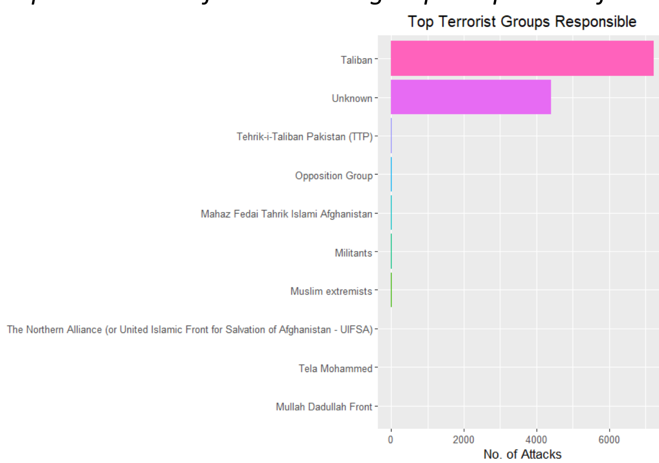
(iv) Top Terror Groups

The final interactive bar chart shows "Top Terrorist Groups" responsible for the attacks in each selected country for 20 years from 1998 to 2018:

Rshiny script:

```
output$Terrorist_Groups <- renderPlot({
  globalterror %>%
    filter(country_txt == input$v_country) %>%
    select(gname) %>%
    drop_na() %>%
    count(gname) %>%
    slice_max(order_by = gname,n=10)%>%
    mutate(gname = fct_reorder(gname, n)) %>%
    ggplot(aes(x = n, y = gname, fill = gname)) +
    geom_col() +
    ylab("") + xlab("No. of Attacks")+
    labs(title = "Top Terrorist Groups Responsible") +
    theme(plot.title = element_text(hjust = 0.5))+
    theme(legend.position = "none")
})
```

Figure 9: Bar Chart representation of the terrorist groups responsible for attacks in Afghanistan



For the selected country of Afghanistan, Figure 9 pointed to Taliban as the main perpetuator of terrorist attacks in the country with more than 7000 attacks over the past 20 years to 2018; followed by unknown/unprofiled groups with more than 4000 attacks made in the country over the same period.

3.1.3 Interactive Word Cloud

The word cloud displays the words that are mentioned most often in the motives of the terrorist attacks, with the size of the words determined by the frequency. The package "wordcloud2" is used for this graph.

Rshiny script:

```
tabPanel(
  title = "Word Cloud",icon = icon("fas fa-cloud") ,
  sidebarLayout(
    sidebarPanel(
      radioButtons(
        inputId = "source",
        label = "Word source",
        choices = c(
          "Motives" = "motives"
        )
      ),
      hr(),
      selectInput(inputId="country","Country",country,selected="Afghanistan"),
      hr(),
      sliderInput(inputId="year","Year",min=1998,max=2018,value=2018,sep="",animate=FALSE),
      hr(),
      checkboxInput("remove_words", "Remove specific words?", FALSE),
      conditionalPanel(
        condition = "input.remove_words == 1",
        textAreaInput("words_to_remove1", "Words to remove (one per line)", rows = 1)
      ),

    numericInput("num", "Maximum number of words",
                 value = 100, min = 5
    ),
    hr(),
    colourInput("col", "Background color", value = "white")
  ),
  mainPanel(
    textOutput("alt"),
    wordcloud2Output("cloud"),
    br(),
    br()
  )
```

There are several ways the users can interactively customize this word cloud. The users have the option to choose the country, or the year to extract the data from. Only one year can be examined at a time in the word cloud because using data from multiple years might cause the Shiny app to hang due to the large number of observations. Additionally, there are dynamic textboxes where the users can input words that they want to be removed from the word cloud. A new textbox will appear every time the users enter a value in the previous textbox. The users can also choose the maximum number of words to be displayed to prevent unnecessary clutter, as well as the background color.

Rshiny script:

```
##Word cloud server
#Get data depending on input (country, year, motive/summary)
data_source <- reactive({
  if(input$country=="All"){
    subdata <- subset(globalterror,globalterror$iyear==input$year)
  }else{
    subdata <- subset(globalterror,globalterror$iyear==input$year&globalterror$country_txt==input$country)
  }

  #if (input$source == "summary") {
    #data <- paste(subdata$summary[subdata$summary!=""],collapse=" ")
 # }
  #else if (input$source == "motives") {
    data <- paste(subdata$motive[subdata$motive!=""],collapse=" ")
  #}

  return(data)
})
```

The data is extracted based on the values specified by the users. Paste function is used to concatenate all the motives from different records into a single string. Originally, the word cloud is designed to allow the users to choose between displaying the motives or summary. However, displaying summary often cause the application to crash due to the large number of data, so it is commented out.

Rshiny script:

```
create_wordcloud <- function(data, num_words = 100, background = "white") {

    # If text is provided, convert it to a dataframe of word frequencies
    if (is.character(data)) {
        corpus <- Corpus(VectorSource(data))
        corpus <- tm_map(corpus, tolower)
        corpus <- tm_map(corpus, removePunctuation)
        corpus <- tm_map(corpus, removeNumbers)
        corpus <- tm_map(corpus, removeWords, stopwords(tolower("English")))
        corpus <- tm_map(corpus, removeWords, c(input$words_to_remove1))
        corpus <- tm_map(corpus, removeWords, c(input$words_to_remove2))
        corpus <- tm_map(corpus, removeWords, c(input$words_to_remove3))
        corpus <- tm_map(corpus, removeWords, c(input$words_to_remove4))
        corpus <- tm_map(corpus, removeWords, c(input$words_to_remove5))
        corpus <- tm_map(corpus, removeWords, c(input$words_to_remove6))
        corpus <- tm_map(corpus, removeWords, c(input$words_to_remove7))
        corpus <- tm_map(corpus, removeWords, c(input$words_to_remove8))
        corpus <- tm_map(corpus, removeWords, c(input$words_to_remove9))
        corpus <- tm_map(corpus, removeWords, c(input$words_to_remove10))
        tdm <- as.matrix(TermDocumentMatrix(corpus))
        data <- sort(rowSums(tdm), decreasing = TRUE)
        data <- data.frame(word = names(data), freq = as.numeric(data))
    }
```

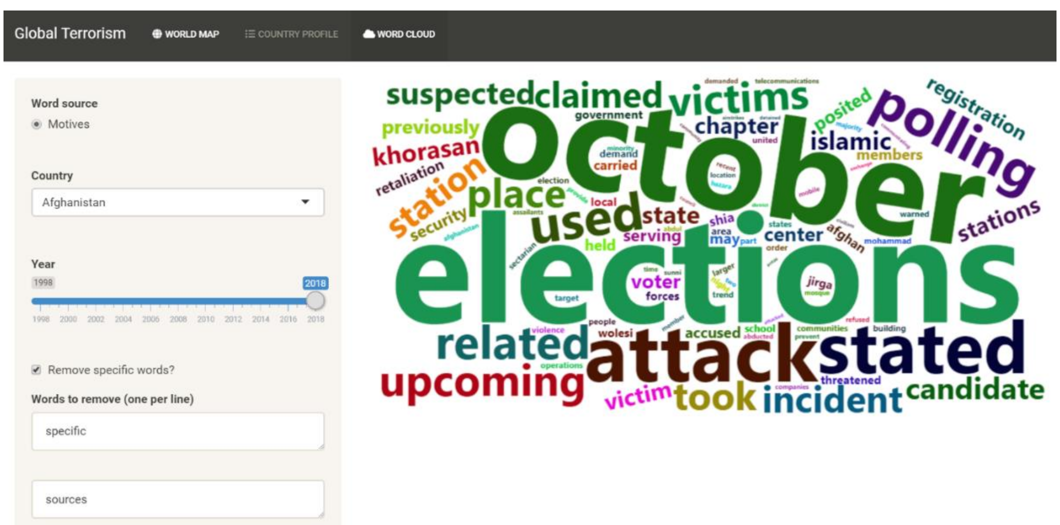
Corpus function from tm package allows the string to be prepared for the word cloud by standardizing the format and removing unwanted characters and common words that might not be insightful. English stop words and specific words from the input text box are removed. The data is then converted into a data frame that contains the words and the frequencies in descending order.

Rshiny script:

```
wordcloud2(data, backgroundColor = background)
```

The word cloud is then generated.

Figure 10: RShiny application outcome of "Word Cloud"



This example is taken from motives of attack in Afghanistan in 2018, removing the word "specific", "sources", "motives", and "noted". The motives of majority of the attacks seem to be related to elections, with words such as "elections", "polling", "candidate", and "registration" dominating the space. "October" is also prominent in the display, and by cross-checking with online resources, we found that the parliamentary election was held in October 2018. The election was finally held after three years delays due to security issues, and still receive numerous security threats. At least 28 people was killed in incidents related to the voting[9].

3.2 Inferential Statistics

As explained in section 2.4, 3 multiple regression models for different periods were developed. To construct the models, the following approach was taken:

1. Model was first run using all available independent variables
2. Correlated and statistically insignificant variables were removed
3. Model was re-run using significant variables resultant from (2)

The results are summarized below:

Table 2: Results obtained for different models in Multilinear Regression

Model	Equation	P-value	R-squared
Model (1) Years 2012-2018	$GTI = -21.26 + 0.82 logGDP - 0.018 Exports + 3.7e^{-4} PopDen + 0.022 LifeSatis + 2.29e^{-2} Unemp + 0.33 MiltExp + 4.13e^{-2} RuralPop - 0.90 GovtEff$	<2.2 e-16	40%
Model (2) Year 2017	$GTI = -16.37 + 0.81 logGDP - 0.85 logExports + 0.41 MiltExp + 0.02 RuralPop - 1.08 GovtEff$	5.707e-15	43%
Model (3) Year 2018 lag	$GTI = -16.95 + 0.84 logGDP - 0.91 logExports + 0.39 MiltExp + 0.02 RuralPop - 1.04 GovtEff$	<2.2 e-16	46%

The final model selected is Model (3) as it has the highest adjusted R-Squared value. The results for the two other models are found in Appendix 1 & 2. The third model uses 2018 GTI score with a lag of one period for the independent variables (2017). This is to resolve the issue of endogeneity of variables where the response variable, GTI, could affect one of the independent variables[10] or vice versa. Detailed steps on how this model was developed is explained below.

[9] Popalzai, E., Latifi, A. M., & Smith-Spark, L. (2018, October 21). 4 million vote in Afghanistan despite violence and technical glitches. *CNN.* Retrieved November 10, 2020 from https://edition.cnn.com/2018/10/20/asia/afghanistan-elections-intl/index.html

[10] She, S., Wang, Q., & Weimann-Saks, D. (2019). Correlation factors influencing terrorist attacks: Political, social or economic? A study of terrorist events in 49 “Belt and Road” countries. Quality & Quantity, 54(1), 125-146. doi:10.1007/s11135-019-00946-x

Figure 11 & 12: Significant variables of Model (3) Year 2018 lag

```
gti2018lag <- na.omit(gti2018laga)
#2018 lag 1 data
fit2018 <- lm(GTI ~ LogGDP + Exports + Imports + PopDen + LifeSatis + UnEmp + Dem + HCI +
                 MiltExp + Corruption + Inflation + RuralPop + Internet + GovtEff, data=gti2018lag)
summary(fit2018)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -2.278e+01  4.478e+00  -5.087 1.41e-06 ***
LogGDP       9.852e-01  1.486e-01   6.628 1.12e-09 ***
Exports     -1.121e-02  1.788e-02  -0.627  0.53196    
Imports     -3.855e-03  1.968e-02  -0.196  0.84507    
PopDen       2.341e-04  2.976e-04   0.787  0.43314    
LifeSatis   -8.249e-02  2.971e-01  -0.278  0.78174    
UnEmp        2.973e-02  3.691e-02   0.805  0.42224    
Dem         -8.676e-02  1.517e-01  -0.572  0.56838    
HCI          2.315e-01  1.605e+00   0.144  0.88552    
MiltExp      3.506e-01  1.447e-01   2.424  0.01691 *  
Corruption   9.563e-01  5.206e-01   1.837  0.06878 .  
Inflation    2.314e-02  4.489e-02   0.516  0.60718    
RuralPop     2.962e-02  1.362e-02   2.174  0.03174 *  
Internet     6.555e-03  2.824e-02   0.232  0.81684    
GovtEff     -2.037e+00  7.192e-01  -2.833  0.00545 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 13: Result of Multi Linear Regression

```
Residual standard error: 2.003 on 116 degrees of freedom
Multiple R-squared:  0.4609,	Adjusted R-squared:  0.3958 
F-statistic: 7.082 on 14 and 116 DF,  p-value: 2.234e-10
```

The full model is significant with a p-value less than 0.05. The individual p-values of GDP, MiltExp, RuralPop and GovtEff are significant with p-values less than 0.05. Variables with collinearity might be present in the current model so a pairwise partial correlation coefficient matrix among the independent variables is outputted in Fig. 14.

Figure 14: Pairwise Partial Correlation Matrix

```
> X2018<-gti2018lag[,4:17]
> res <- cor(X2018,use="complete.obs")
> round(res, 2)
```

	LogGDP	Exports	Imports	PopDen	LifeSatis	UnEmp	Dem	HCI	MiltExp	Corruption	Inflation	RuralPop	Internet	GovtEff
LogGDP	1.00	0.03	-0.30	0.09	0.53	-0.14	0.39	0.53	0.14	0.47	-0.15	-0.51	0.57	0.59
Exports	0.03	1.00	0.84	0.47	0.35	-0.01	0.20	0.40	-0.05	0.40	-0.22	-0.31	0.44	0.45
Imports	-0.30	0.84	1.00	0.41	0.10	0.06	0.06	0.20	-0.09	0.19	-0.06	-0.05	0.19	0.18
PopDen	0.09	0.47	0.41	1.00	0.07	-0.11	0.02	0.18	0.12	0.19	-0.06	-0.16	0.07	0.21
LifeSatis	0.53	0.35	0.10	0.07	1.00	-0.19	0.58	0.59	-0.05	0.67	-0.41	-0.68	0.74	0.73
UnEmp	-0.14	-0.01	0.06	-0.11	-0.19	1.00	0.06	-0.02	0.09	-0.03	0.00	-0.15	0.00	-0.05
Dem	0.39	0.20	0.06	0.02	0.58	0.06	1.00	0.64	-0.20	0.76	-0.21	-0.45	0.64	0.75
HCI	0.53	0.40	0.20	0.18	0.59	-0.02	0.64	1.00	0.09	0.69	-0.21	-0.50	0.73	0.79
MiltExp	0.14	-0.05	-0.09	0.12	-0.05	0.09	-0.20	0.09	1.00	0.01	-0.09	-0.23	-0.02	0.03
Corruption	0.47	0.40	0.19	0.19	0.67	-0.03	0.76	0.69	0.01	1.00	-0.27	-0.56	0.81	0.93
Inflation	-0.15	-0.22	-0.06	-0.06	-0.41	0.00	-0.21	-0.21	-0.09	-0.27	1.00	0.30	-0.33	-0.32
RuralPop	-0.51	-0.31	-0.05	-0.16	-0.68	-0.15	-0.45	-0.50	-0.23	-0.56	0.30	1.00	-0.64	-0.59
Internet	0.57	0.44	0.19	0.07	0.74	0.00	0.64	0.73	-0.02	0.81	-0.33	-0.64	1.00	0.86
GovtEff	0.59	0.45	0.18	0.21	0.73	-0.05	0.75	0.79	0.03	0.93	-0.32	-0.59	0.86	1.00

The pairwise partial correlation matrix in Fig. 14 reveals that exports and imports are highly correlated with ρ=0.84. Corruption is also highly correlated with internet and government efficiency with ρ=0.81 and ρ=0.93, respectively, and Internet is high correlated with government efficiency at ρ=0.86. We then again manually take out imports, corruption, and internet from the model and transform GDP and Exports.

Figure 15 & 16: Model output upon removal of non-significant and correlated variables

```
fit2018a <- lm(GTI ~ LogGDP + log(Exports) + PopDen + LifeSatis + UnEmp + Dem +
                 HCI + MiltExp + Inflation + RuralPop + GovtEff, data=gti2018lag)
summary(fit2018a)
```

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -1.666e+01  4.755e+00  -3.503 0.000649 ***
LogGDP        8.602e-01  1.317e-01   6.529  1.7e-09 ***
log(Exports) -9.451e-01  3.786e-01  -2.496 0.013922 *  
PopDen        1.359e-04  2.533e-04   0.537 0.592496    
LifeSatis    -8.820e-02  2.894e-01  -0.305 0.761067    
UnEmp         2.793e-02  3.592e-02   0.778 0.438336    
Dem          -4.173e-02  1.397e-01  -0.299 0.765706    
HCI          -1.686e-01  1.493e+00  -0.113 0.910285    
MiltExp       3.548e-01  1.374e-01   2.581 0.011053 *  
Inflation     2.895e-02  4.323e-02   0.670 0.504402    
RuralPop      2.149e-02  1.292e-02   1.663 0.098918 .  
GovtEff      -8.612e-01  4.223e-01  -2.039 0.043658 *  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.89 on 119 degrees of freedom
Multiple R-squared:  0.4941,	Adjusted R-squared:  0.4473 
F-statistic: 10.57 on 11 and 119 DF,  p-value: 2.522e-13
```

The model has improved its R-squared after eliminating correlated variables and transformation, currently at 45%. A stepwise selection of variables is performed to retain only the significant variables.

Figure 17 & 18: Result of Stepwise Regression using Stepwise Selection

```
step2018a <- stepAIC(fit2018a, direction ="both")
summary(step2018a)
```

```
Call:
lm(formula = GTI ~ LogGDP + log(Exports) + MiltExp + RuralPop + 
    GovtEff, data = gti2018lag)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.4782 -1.3476 -0.1729  1.1161  4.6251 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  -16.94948    3.94258  -4.299 3.42e-05 ***
LogGDP         0.84496    0.12265   6.889 2.44e-10 ***
log(Exports)  -0.91254    0.34902  -2.615 0.010031 *  
MiltExp        0.38867    0.11901   3.266 0.001409 ** 
RuralPop       0.02258    0.01027   2.199 0.029696 *  
GovtEff       -1.03900    0.26745  -3.885 0.000165 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.86 on 125 degrees of freedom
Multiple R-squared:  0.4852,	Adjusted R-squared:  0.4646 
F-statistic: 23.56 on 5 and 125 DF,  p-value: < 2.2e-16
```

The stepwise model, having significant variables GDP, Exports, MiltExp, RuralPop, GovtEff, is significant with p-value less than 0.05. The 2018lag model now has an adjusted R-squared equal to 46%. The final model for 2018lag is described by:

$$GTI = -16.95 + 0.84logGDP - 0.91logExports + 0.39MiltExp + 0.02RuralPop - 1.04GovtEff$$

The model is interpreted, *having all other variables constant*:

- GDP is positively correlated with the GTI score, e.g. a 10-unit increase of a country's GDP can increase the GTI score by 0.84
- Exports is negatively correlated with GTI score, e.g. a 10-percent increase in Exports decreases the GTI score by 0.91
- Military Expenditure is positively correlated with GTI score, e.g. a 1-percent increase in Military Expenditure increases the GTI score by 0.39
- Rural Population is positively correlated with GTI score, e.g. a 1-percent increase in Rural Population increases the GTI score by 0.02, and
- Government Efficiency is negatively correlated with GTI score, e.g. a 1-unit increase in Government Efficiency decreases the GTI score by 1.04

4 Conclusions and Recommendations

The best multiple linear regression model chosen in this study can explain around 46% of terrorism activities and damages in a country. There are other variables to be considered that might further improve this model further but is currently not covered in this study due to lack of data. For instance, variables related to religion can be explored further as per insights discovered in exploratory data analysis stage of this study. Certain geographical characteristics of a country have also been studied to favor terrorist activities[11] and in future suitable measurements of geographical characteristics might be included.

Another key challenge faced during this study was availability of data for countries and years. Observations were frequently removed due to missing data and no suitable replacement could be found. Having more complete set of indices data for all countries and all years might further improve the model developed. Having seen the robustness of using RShiny for exploratory data analysis, more features can also be added in the future should new variables be added for meaningful analysis.

[11] Abadie, Alberto, Poverty, Political Freedom, and the Roots of Terrorism (October 2004). Available at SSRN: https://ssrn.com/abstract=617542 or http://dx.doi.org/10.2139/ssrn.617542

Appendices

Appendix 1 - Linear Regression from Years 2012 – 2018

<u>Figure 19</u>: Final Result of Multi Linear Regression of all years

```
Call:
lm(formula = GTI ~ LogGDP + Exports + PopDen + LifeSatis + UnEmp +
    MiltExp + RuralPop + GovtEff, data = allyears)

Residuals:
    Min      1Q  Median      3Q     Max
-4.8891 -1.3972 -0.1839  1.2764  5.7800

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.126e+01  1.630e+00 -13.038  < 2e-16 ***
LogGDP       8.234e-01  5.588e-02  14.736  < 2e-16 ***
Exports     -1.821e-02  4.277e-03  -4.258 2.36e-05 ***
PopDen       3.765e-04  1.263e-04   2.981  0.00298 **
LifeSatis    2.251e-01  8.588e-02   2.621  0.00898 **
UnEmp        2.293e-02  1.403e-02   1.634  0.10273
MiltExp      3.305e-01  5.671e-02   5.829 8.74e-09 ***
RuralPop     4.134e-02  5.277e-03   7.834 1.89e-14 ***
GovtEff     -9.022e-01  1.298e-01  -6.952 8.66e-12 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.903 on 661 degrees of freedom
Multiple R-squared:  0.4116,    Adjusted R-squared:  0.4045
F-statistic: 57.79 on 8 and 661 DF,  p-value: < 2.2e-16
```

The stepwise model, having significant variables GDP, Exports, PopDen, LifeSatis, MiltExp, RuralPop, GovtEff, is significant with p-value less than 0.05. The 2012-2018 model has an adjusted R-squared equal to 40%. The final model for 2012-2018 is described by:

$$GTI = -21.26 + 0.82logGDP - 0.018Exports + 3.7e^{-4}PopDen + 0.022LifeSatis + 2.29e^{-2}Unemp + 0.33MiltExp + 4.13e^{-2}RuralPop - 0.90GovtEff$$

Appendix 2- Linear Regression for Year 2017

Figure 20 & 21: Result of Stepwise Regression using Stepwise Selection strategy

```
step2017a <- stepAIC(fit2017a, direction ="both")
summary(step2017a)
```

```
Call:
lm(formula = GTI ~ LogGDP + log(Exports) + MiltExp + RuralPop +
    GovtEff, data = gti2017)

Residuals:
    Min      1Q  Median      3Q     Max
-4.2945 -1.3900 -0.1735  1.3320  4.6937

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -16.37050    4.12381  -3.970 0.000121 ***
LogGDP         0.81850    0.12829   6.380 3.13e-09 ***
log(Exports)  -0.84501    0.36506  -2.315 0.022260 *
MiltExp        0.41483    0.12448   3.332 0.001133 **
RuralPop       0.01907    0.01074   1.776 0.078247 .
GovtEff       -1.08000    0.27974  -3.861 0.000180 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.946 on 125 degrees of freedom
Multiple R-squared:  0.4519,    Adjusted R-squared:   0.43
F-statistic: 20.61 on 5 and 125 DF,  p-value: 5.707e-15
```

The stepwise model, having significant variables GDP, Exports, MiltExp, RuralPop, GovtEff, is significant with p-value less than 0.05. The 2017 model has an adjusted R-squared equal to 43%. The final model for 2017 is described by:

$$GTI = -16.37 + 0.81 logGDP - 0.85 logExports + 0.41 MiltExp + 0.02 RuralPop - 1.08 GovtEff$$