GLOBAL TERRORISM ATTACKS ANALYSIS

Introduction

Global terrorism attacks analysis is an analysis project in which GTD (Global terrorism dataset) also known as "most comprehensive unclassified data base on terrorist events in the world" and t includes over 190,000 terrorist attacks. This is performed using R studio and its various packages. The dataset used is terrorism.csv. These events undergo various processes to reach the final output which is terrorism attacks analysed. The output is visualised using bar plot, heatmap, themes, graphs, etc.

SCOPE

Global terrorism attacks analysis can be used for analysing various attacks by different terrorist organisations at various places of the world. The analysis has high accuracy which leads to tracing and prediction of such attacks well before in time.
This project can be used by intelligence agencies, government, authors, etc.

PLATFORM: R STUDIO

R was specifically designed for statistical analysis, which makes it highly suitable for data science applications. Although the learning curve for programming with R can be steep, especially for people without prior programming experience, the tools now available for carrying out text analysis in R make it easy to perform powerful, cutting-edge text analytics using only a few simple commands. One of the keys to R's explosive growth has been its densely populated collection of extension software libraries, known in R terminology as packages, supplied and maintained by R's extensive user community. Each package extends the functionality of the base R language and core packages, and in addition to functions and data must include documentation and examples, often in the form of vignettes demonstrating the use of the package. The best-known package repository, the Comprehensive R Archive Network (CRAN), currently has over 10,000 packages that are published.

Text analysis in particular has become well established in R. There is a vast collection of dedicated text processing and text analysis packages, from low-level string operations to advanced text modelling techniques such as fitting Latent Dirichlet Allocation models, R provides it all. One of the main advantages of performing text analysis in R is that it is often possible, and relatively easy, to switch between different packages or to combine them. Recent efforts among the R text analysis developers' community are designed to promote this interoperability to maximize flexibility and choice among users. As a result, learning the basics for text analysis in R provides access to a wide range of advanced text analysis features.

PROJECT SPECIFICATION

- R Studio version 1.2.5033

HARDWARE SPECIFICATIONS

- Microsoft® Windows® 7/8/10 (32- or 64-bit)
- 3 GB RAM minimum, 8 GB RAM recommended;
- 2 GB of available disk space minimum
- core processor of i3 minimum or above.

DATASET

- Terrorism.csv

CHAPTER 1

Reading events and Cleaning events

READING EVENTS FROM TERRORISM.CSV

Before going to terrorism attacks analysis, the first step is to read the data for performing analysis on. The data is saved in dataset named as terrorism.csv. This dataset contains 1,90,000 attacks that have took place all across the world. The events saved in dataset are unstructured. To perform analysis, reading of data set is done using command “read.csv”.

Steps to read data:

1) Get the current working directory using **getwd()** function. It returns the current directory.
2) Set the working directory to the directory in which the dataset is saved. This is implemented using **setwd()** function.
3) To successfully load this file into R, you can use **read.csv** function. The read.csv() as well as the read.csv2() function are almost identical to the read.table() function, with the sole difference that they have the header and fill arguments set as TRUE by default.

```
> getwd()
[1] "C:/Users/varun/Documents"
> setwd('/Users/varun/Desktop')
>
>
> # Read the file
>
> # /  [The Gloabl Terrorism database (GDS)](http://start.umd.edu/gtd/) is an open-source database including information on terrorist attacks around the
 world from 1970 through 2015. The dataset I am using contains more than 150,000 terrorism attacks worldwide from 1970 to 2015 except 1993.
>
> terrorism <- read.csv('terrorism.csv', stringsAsFactors = F)
>
> dim(terrorism)
[1] 191464    135
>
> str(terrorism)
'data.frame':   191464 obs. of  135 variables:
 $ eventid           : num  1.97e+11 1.97e+11 1.97e+11 1.97e+11 1.97e+11 ...
 $ iyear             : int  1970 1970 1970 1970 1970 1970 1970 1970 1970 1970 ...
 $ imonth            : int  7 0 1 1 1 1 1 1 1 1 ...
 $ iday              : int  2 0 0 0 0 1 2 2 2 3 ...
 $ approxdate        : chr  "" "" "" "" ...
 $ extended          : int  0 0 0 0 0 0 0 0 0 0 ...
```

Figure 1. Read File

There are 137 columns in the dataset, to make it neat, I will have to do some sub setting, to only keep the columns I need.

```
 $ weapdetail        : chr  "" "" "" "Explosive" ...
 $ nkill             : int  1 0 1 NA NA 0 0 0 0 0 ...
  [list output truncated]
>
> # There are 137 variables in the dataset, to make it neat, I will have to do some subsetting, only keep the columns I need.
>
> terrorism <- terrorism[,c("iyear","imonth", "iday", "country_txt", "region_txt", "provstate", "city", "latitude", "longitude", "attacktype1_txt", "tar
gtype1_txt", "corp1", "target1", "natlty1_txt", "gname", "weaptype1_txt", "weapsubtype1_txt")]
>
>
> |
```

Figure 2. Clean Dataset

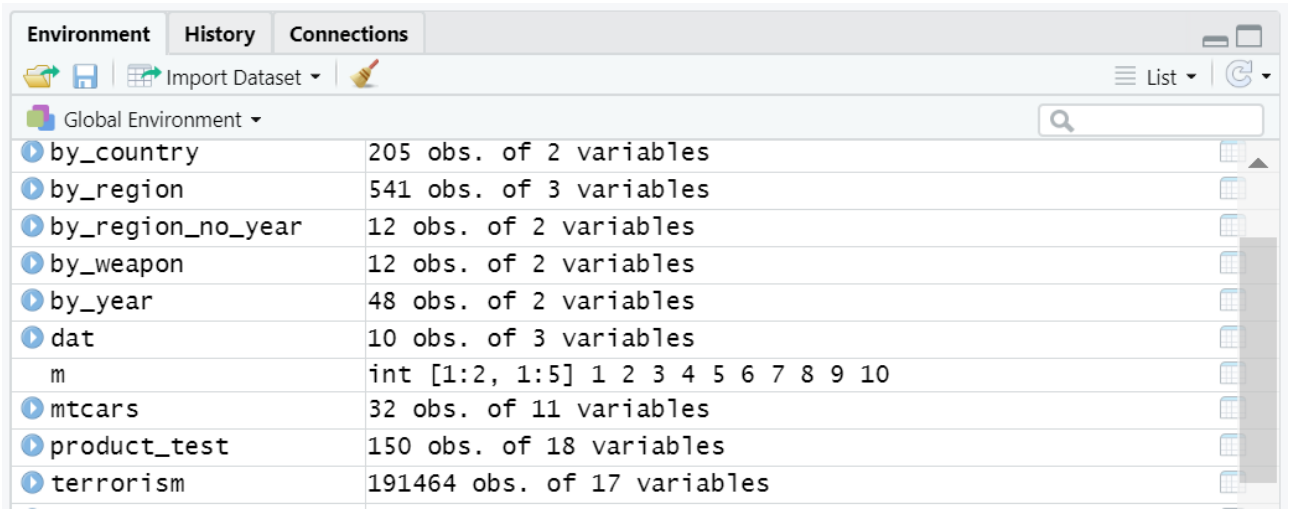

Figure 3. Dataset imported

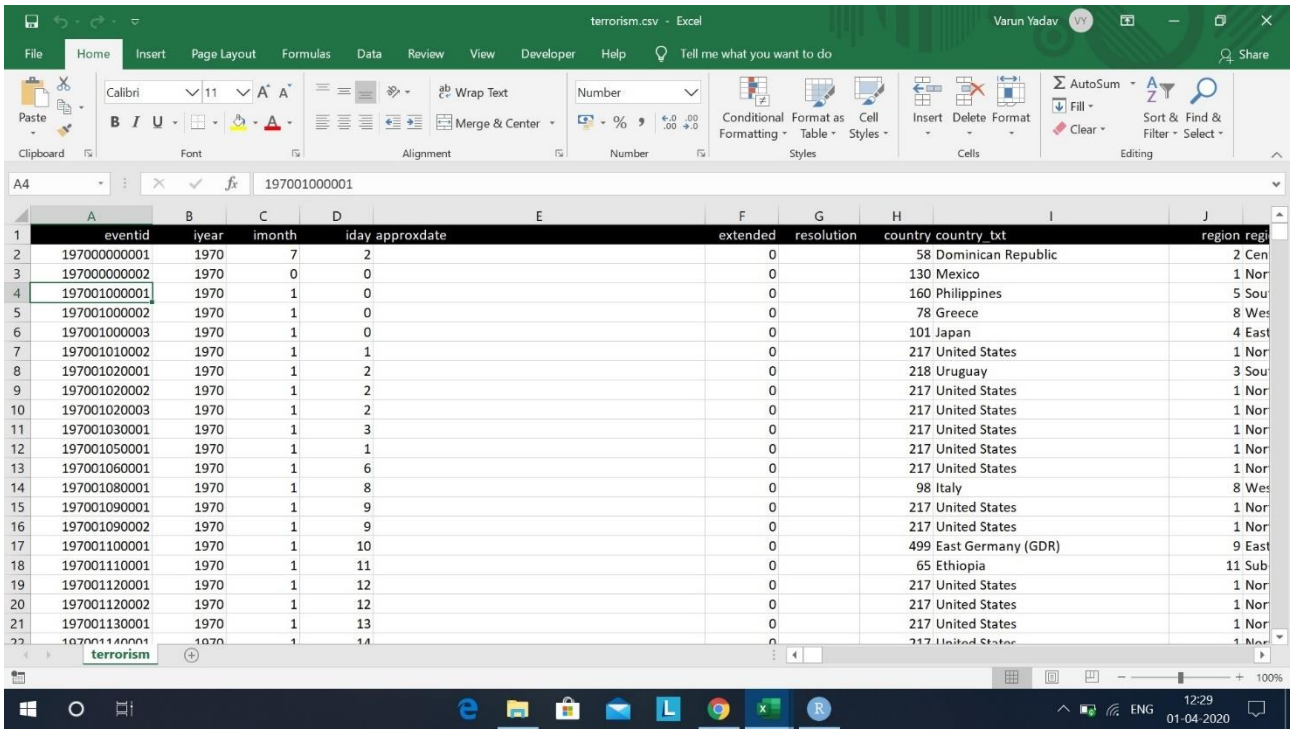

Figure 4. terrorism.csv

There are a set of libraries and packages needed for this analysis of the dataset which would help us to produce various visualizations of the events and help us better understand the various scenarios and crisis the world is facing because of these terrorist attacks.

CHAPTER 2

Line Plot

A line chart or line plot or line graph or curve chart is a type of chart which displays information as a series of data points called 'markers' connected by straight line segments. It is a basic type of chart common in many fields. Used across many fields, this type of graph can be quite helpful in depicting the changes in values over time. We are going to use ggplot for depicting the line plot.

Syntax of line plot:

The basic syntax to create a line-chart in R is −

```
ggplot(aes(x = column_name, y = n), data = by_year) +
geom_line(size = 2.5, alpha = 0.7, color = "mediumseagreen") +
geom_point(size = 0.5) + xlab("Year") + ylab("Number of terrorist Attacks") +
ggtitle("Terrorist Attacks Worldwide by Year 1970-2015") + theme_fivethirtyeight()
```

Following is the description of the parameters used −

- Aes is the keyword used before the declaration of the x axis and the y axis.
- Column_name is the column of whose parameter is to be shown
- By_year is the variable which summarises the plot
- Geom_line is the parameter holding line attributes
- Geom_point is used to create scatterplots.
- Ggtitle is used to give title to the theme.

Line plot is used in global terrorism attacks analysis to get the frequency of attacks with respect to the countries that these took place and the methods used to perform these. This is required so that patterns can be identified. This will give the country-wise attacks that has took place. A summarization on region will further help the summits in the world to act against these forces.

```r
# Globally, terrorist attacks have increased dramatically since 2010,

by_region <- terrorism %>% group_by(region_txt, iyear) %>%
  summarise(n=n())
ggplot(by_region, aes(x = iyear, y = n, colour = region_txt)) +
  geom_line() +
  geom_point() +
  facet_wrap(~region_txt) + xlab('Year') +
  ggtitle('Terrorist Attacks by Region and Year 1970-2015') +
  theme(legend.position="none")
```

Figure 5. Line Plot code

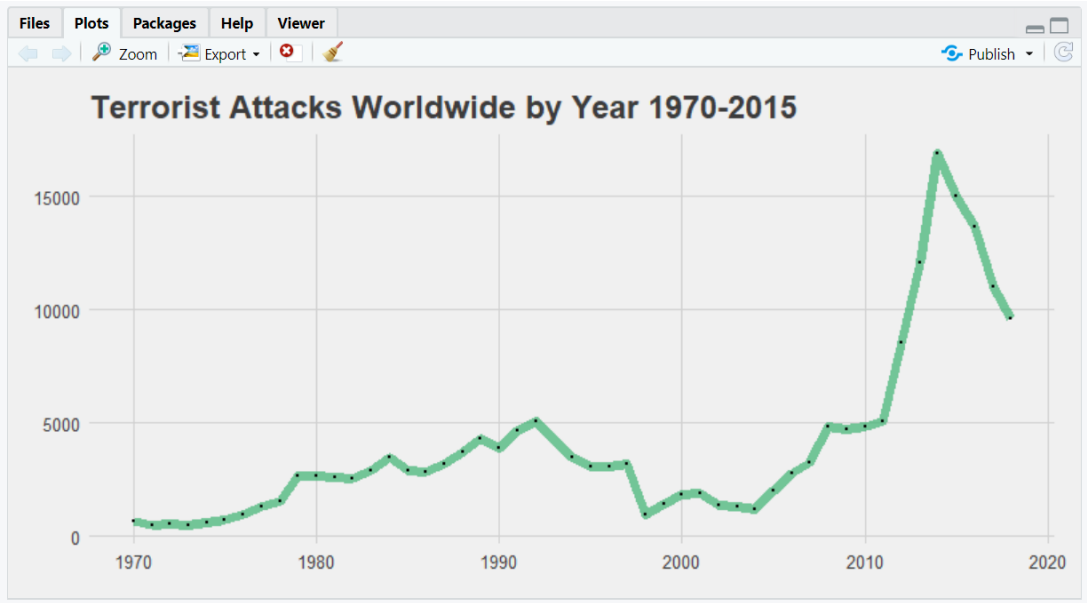

Figure 6. Line plot of Worldwide attacks

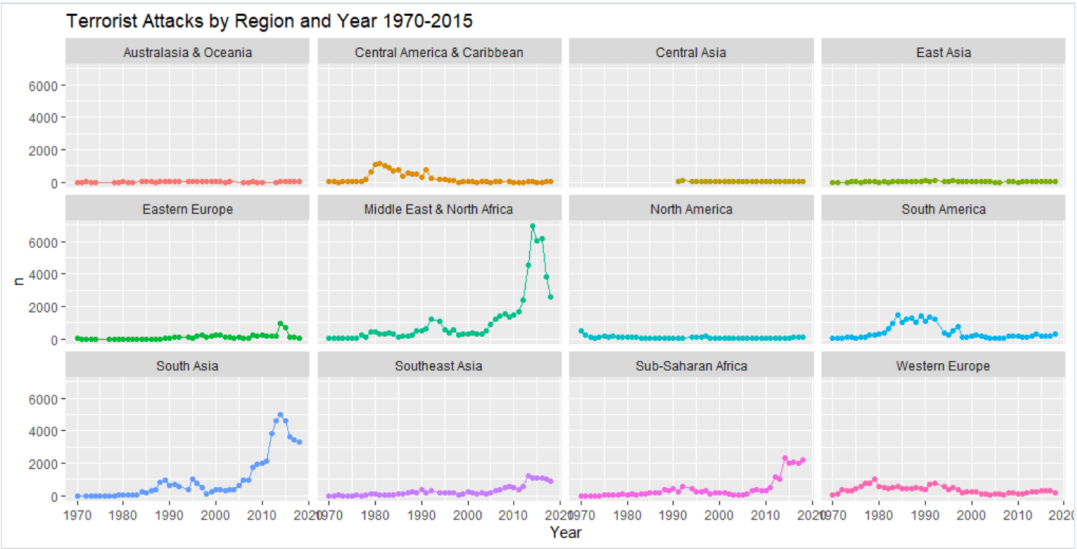

Figure 7. Line plot of attacks by region and year

CHAPTER 3

Bar Chart

A bar chart represents data in rectangular bars with length of the bar proportional to the value of the variable. R uses the function **barplot()** to create bar charts. R can draw both vertical and Horizontal bars in the bar chart. In bar chart each of the bars can be given different colors.

```
by_region_no_year <- terrorism %>% group_by(region_txt) %>%
  summarise(n=n())
ggplot(aes(x=reorder(region_txt, n), y=n), data=by_region_no_year) +
  geom_bar(stat = 'identity') +
  ggtitle('Terrorist Attacks by Region 1970-2015') + coord_flip() + theme_fivethirtyeight()
```

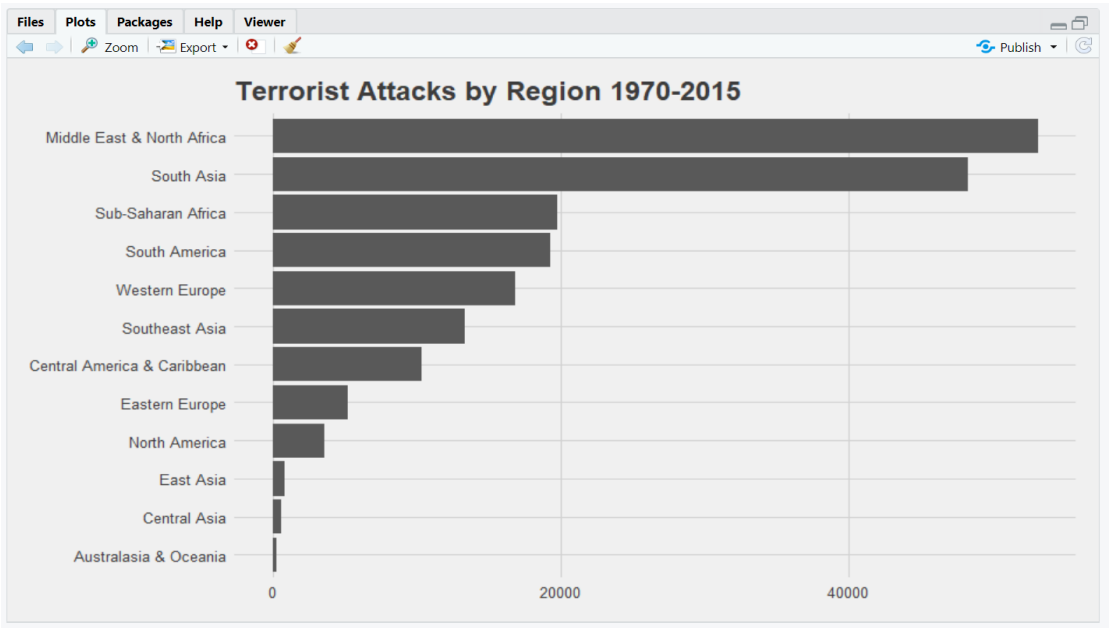

Figure 8. Bar plot of attacks by region and year

A small fraction of the terrorist attacks happened in the Western countries. Most attacks were heavily concentrated geographically in Middle East, North Africa and South Asia.

Let's look at the countries.

```
by_country <- terrorism %>% group_by(country_txt) %>%
  summarise(n=n())
by_country <- arrange(by_country, desc(n))
top10 <- head(by_country, 10)
top10
ggplot(aes(x=reorder(country_txt, n), y=n), data=top10) +
  geom_bar(stat = 'identity') + xlab('Country') + ylab('Number of Terrorist Attacks') + ggtitle('Countries with the most terrorist attacks, 1970-2015') +
  coord_flip() + theme_fivethirtyeight()
```

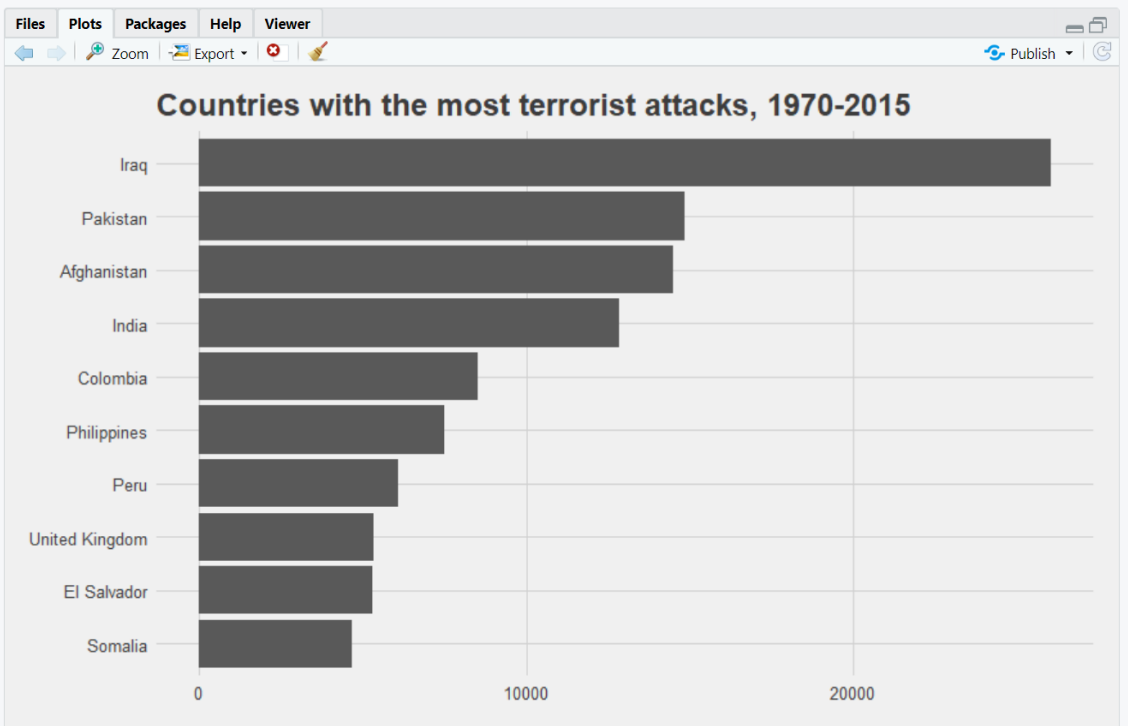

Figure 9. Bar plot of countries affected

Iraq, Afghanistan and pakistan, India have suffered the most from terrorism. Surprisingly, United Kingdom tops the list in Europe, with almost 5000 attackes from 1970 to 2015.

Tactics and Weapons

```
by_attacktype <- terrorism %>% group_by(attacktype1_txt) %>%
  summarise(n=n())
ggplot(aes(x=reorder(attacktype1_txt, n), y=n), data=by_attacktype) +
  geom_bar(stat = 'identity') + xlab('Attack Type') + ylab('Number of Attacks') + ggtitle('Terrorist Attack Tactics Wordwide, 1970-2015') + coord_flip() +
  theme_fivethirtyeight()
```

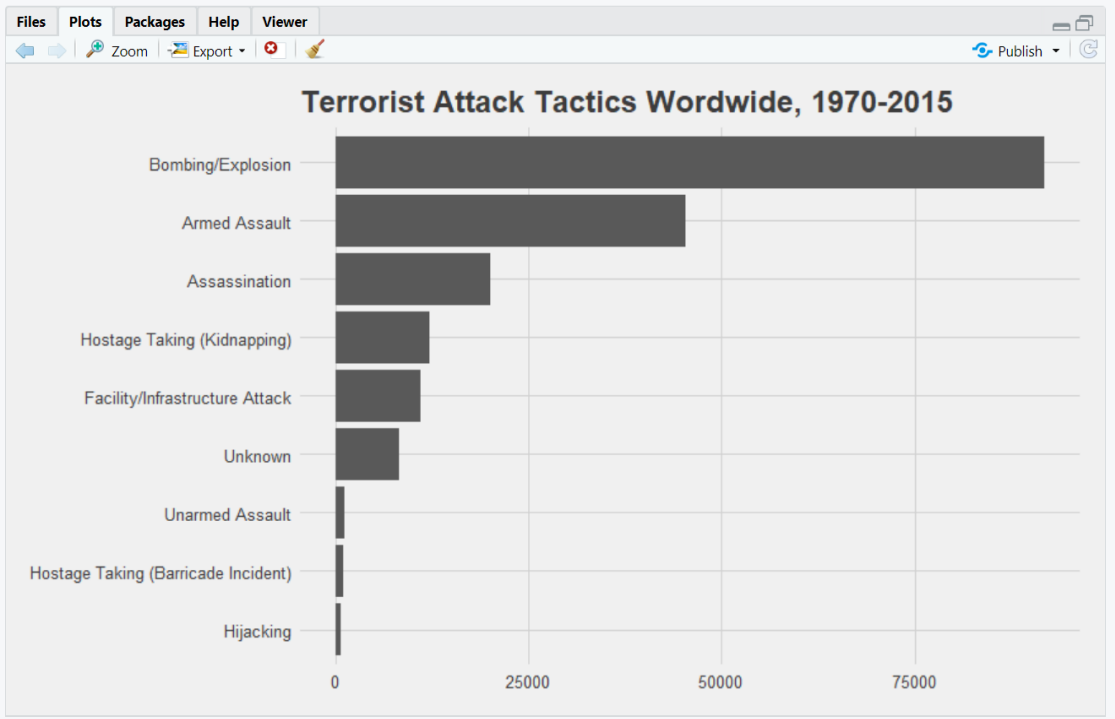

Figure 10. Bar plot of attack tactics

```
by_weapon <- terrorism %>% group_by(weaptype1_txt) %>%
  summarise(n=n())
ggplot(aes(x=reorder(weaptype1_txt, n), y=n), data=by_weapon) +
  geom_bar(stat = 'identity') + xlab('Weapon') + ylab('Number of Attacks') + ggtitle('Terrorist Attack By Weapon Wordwide, 1970-2015') + coord_flip() +
  theme_fivethirtyeight()
```

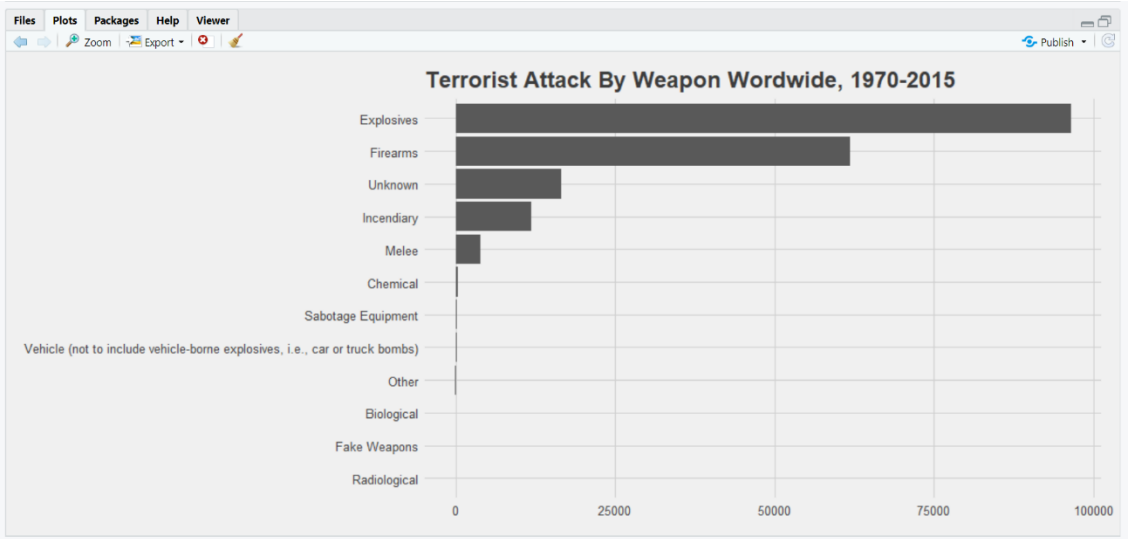

Figure 11. Bar plot of weapon types used in attacks

The most commonly used attack tactic from 1970 to 2015 involved bomb and explosives, followed by armed assault.

Who are they targeting?

```
attack2015 <- terrorism[terrorism$iyear==2015, ]
by_target <- attack2015 %>% group_by(targtype1_txt) %>%
  summarise(n=n())
by_target <- arrange(by_target, desc(n))
by_target
ggplot(aes(x=reorder(targtype1_txt, n), y=n), data=by_target) +
  geom_bar(stat = 'identity') + ggtitle('Terrorist Attack Targets/Victims, 2015') +
  coord_flip() + theme_fivethirtyeight()
```

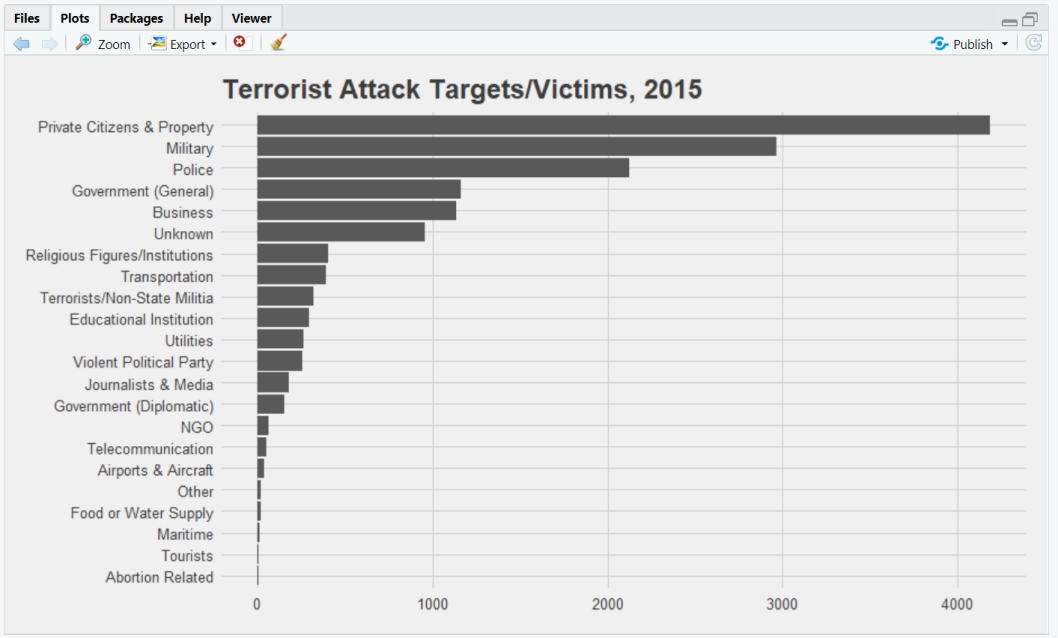

Figure 12. Bar plot of victims of such attacks

Attack type in Baghdad:

```
baghdad_type <- baghdad %>% group_by(attacktype1_txt, iyear) %>%
  summarise(n=n())
ggplot(aes(x=iyear, y=n, fill=attacktype1_txt), data=baghdad_type) +
  geom_bar(stat = 'identity') +
  ggtitle('Attack Type in Baghdad') + theme_fivethirtyeight()
```

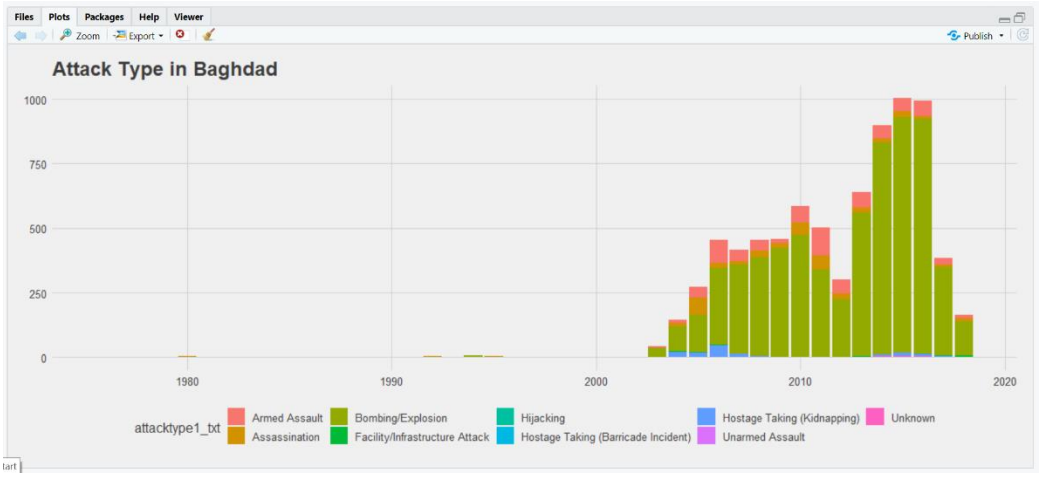

Figure 13. Bar plot of most affected country

CHAPTER 4

Heatmap

Heatmap in R: Static and Interactive Visualization. A **heatmap** (or heat map) is another way to visualize hierarchical clustering. It's also called a false coloured image, where data values are transformed to colour scale. Heat maps allow us to simultaneously visualize clusters of samples and features. A heatmap is a graphical representation of data where the individual values contained in a matrix are represented as colours. This page displays many examples built with R, both static and interactive. The heatmap() function is natively provided in R. It produces high quality matrix and offers statistical tools to normalize input data, run clustering algorithm and visualize the result with dendrograms.

```
gtd <- read.csv("terrorism.csv")
gtd2015 <- gtd[gtd$iyear==2015, ]
gtd2015 <- aggregate(nkill~country_txt,gtd2015,sum)
library(rworldmap)
gtdMap <- joinCountryData2Map( gtd2015,
                    nameJoinColumn="country_txt",
                    joinCode="NAME" )

mapDevice('x11')
mapCountryData( gtdMap,
          nameColumnToPlot='nkill',
          catMethod='fixedWidth',
          numCats=100 )
```

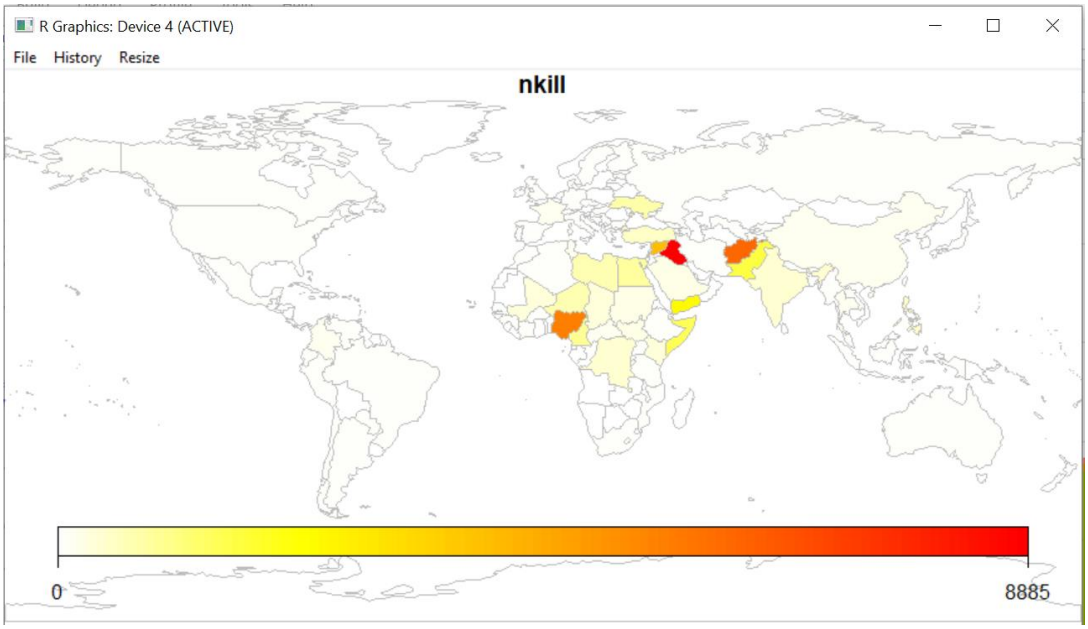

Figure 14. Heat map of attacks for the year 2015

CHAPTER 5

Global terrorism

A terrorist group is fundamentally a social system. The group is formed as a unified whole by fanatics determined to inflict civilian and economic damage on specific targets in pursuit of their extremist goals. Currently, as terrorist groups are becoming more international, allied, and networked, the social systems are transforming into complex sociotechnical systems. For a continuous active terrorist group, the behaviour is influenced by both the interactional environment and the organizational mechanism and thus is a continuously evolving phenomenon. The structure and efficiency of the group will fluctuate and finally reach a relatively stable state because of its covert nature and self-defence mechanism.

Predicting terrorist attacks by group networks is an important but difficult issue in intelligence and security informatics. Effective prediction of the behaviour not only facilitates the understanding of the dynamics of organizational behaviours but also supports homeland security's missions in prevention, preparedness, and response to terrorist acts. There are certain dynamic characteristics of terrorist groups, such as periodic features and correlations between the behaviour and the network.

Terrorism attack analysis is a process of extracting the visualizations of the affected places in the world due to various terrorist attacks by different terrorist organizations. This will help classify the data as well as assist us better understand these events with the help of patterns created by the visualizations. This can be useful in understanding the goals of such organizations and will also help us prepare to take better actions against them well before they even plan on doing such attack on any country, organization, places, etc. This can also be helpful in taking precautionary measures to reduce such activities in the future and also discourage and stop people from becoming part of such groups or organizations.

Following are the packages used for terrorism attacks analysis:

- **Ggplot2 library**
  A system for 'declaratively' creating graphics, based on "The Grammar of Graphics". You provide the data, tell 'ggplot2' how to map variables to aesthetics, what graphical primitives to use, and it takes care of the details.

```
> # Adding necessary libraries
>
> install.packages("ggplot2")
Error in install.packages : Updating loaded packages
> install.packages("ggplot2")
WARNING: Rtools is required to build R packages but is not currently installed. Please download and install the appropriate version of
 Rtools before proceeding:

https://cran.rstudio.com/bin/windows/Rtools/
Installing package into 'C:/Users/varun/Documents/R/win-library/3.6'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/3.6/ggplot2_3.3.0.zip'
Content type 'application/zip' length 4021158 bytes (3.8 MB)
downloaded 3.8 MB

package 'ggplot2' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\varun\AppData\Local\Temp\RtmpqwZOb9\downloaded_packages
>
> library(ggplot2)
>
```

Figure 15. Library ggplot2

- **Dplyr library**
  dplyr is designed to abstract over how the data is stored. That means as well as working with local data frames, you can also work with remote database tables, using exactly the same **R** code.

- **Ggthemes library**
  Some extra themes, geoms, and scales for 'ggplot2'. Provides 'ggplot2' themes and scales that replicate the look of plots by Edward Tufte, Stephen Few, 'Fivethirtyeight', 'The Economist', 'Stata', 'Excel', and 'The Wall Street Journal', among others. Provides 'geoms' for Tufte's box plot and range frame.

- **Rworldmap library**
  Enables mapping of country level and gridded user datasets by facilitating joining to modern world maps and offering visualisation options. Country borders are derived from Natural Earth data v 1.4.0.

- **Reshape2 library**
  Reshape2 is a reboot of the reshape package. It's been over five years since the first release of reshape, and in that time, I've learned a tremendous amount about R programming, and how to work with data in R. Reshape2 uses that knowledge to make a new package for reshaping data that is much more focused and much faster.

- **Xlsx library**
  The xlsx package gives programmatic control of Excel files using R. A high level API allows the user to read a sheet of an xlsx document into a data.frame and write a data.frame to a file. Lower level functionality permits the direct manipulation of sheets, rows and cells. For example, the user has control to set colours, fonts, data formats, add borders, hide/unhide sheets, add/remove rows, add/remove sheets, etc.

```
>
>
> # The rest of the packages required are:
>
> #install.packages("dplyr")
> # library(dplyr)
>
> # install.packages("ggthemes")
> # library(ggthemes)
>
> # install.packages("rworldmap")
> # library(rworldmap)
>
> # install.packages("reshape2")
> # library(reshape2)
>
> # install.packages("xlsx")
> # library(xlsx)
>
>
```

Figure 16. Libraries required

CONCLUSION

Global terrorism attacks analysis plays an important role in today's world as such types of events are on the rise. People of different caste, creed, religion are getting attracted too such organizations who are willing to brain wash their targets. This global terrorism attacks analysis covers terrorism attacks from all over the world. This analysis will help the government and intelligence agencies identify such events well before they take place. R studio helps us to make this process simpler with-it flexibility and supporting packages. Visualising the output is also possible in R-Studio.

BIBLIOGRAPHY

1) https://gtd.terrorismdata.com/files/gtd-1970-2018/
2) https://cran.r-project.org
3) https://cran.r-project.org/web/packages/dplyr/vignettes/dplyr.html
4) https://www.rdocumentation.org/packages/reshape2/versions/1.4.3
5) https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf
6) https://susanli2016.github.io/Terrorist-Attack/
7) https://www.r-graph-gallery.com/heatmap
8) https://cran.r-project.org/web/packages/xlsx/xlsx.pdf