Exploratory Data Analysis-Global Terrorism

Introduction:

The Global Terrorism Database (GDS) is an open-source database that contains information on terrorist acts worldwide from 1970 to 2015. Except for 1993, the dataset I'm using contains almost 150,000 terrorist acts from 1970 to 2015.

From 1970 to 2017, this dataset offers a wealth of information regarding terrorism. There are 181691 entries with 135 columns covering date, time, location, number of hostages, dead, wounded, if a ransom was paid, the outcome, if a suicide attack occurred, claims, and weapons used.

Missing Data:

There is a significant amount of data that is missing. There are 4557 missing longitude data and 4556 missing latitude values, for example. A type of claim was recorded only for 19083 incidents, and only 514 ransom demands are reported.

Analysis

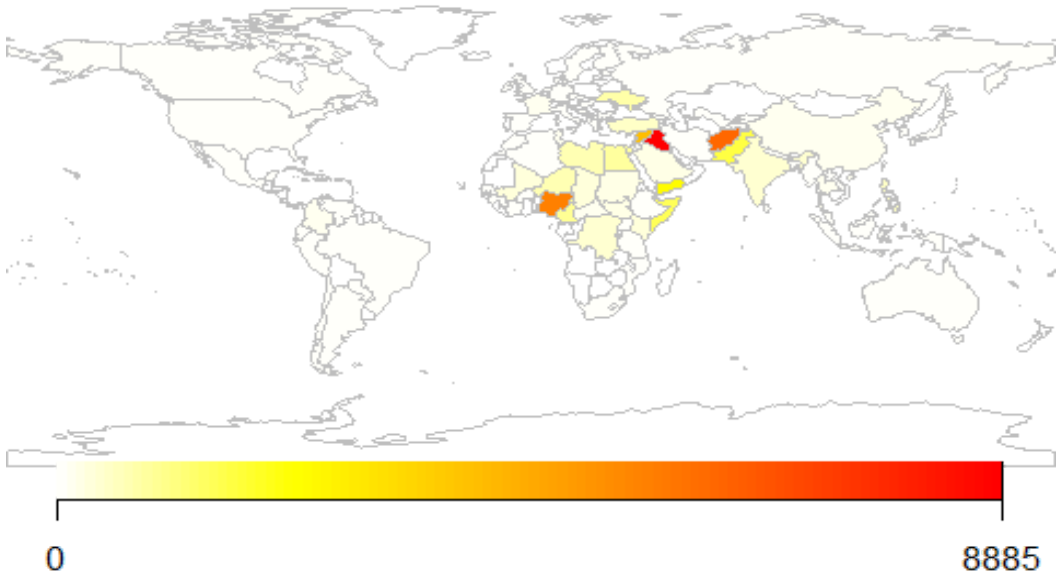
I have analyzed Terrorist Attack Trends, Terrorist Attack Trends by Region and Year, Terrorist Attack Trends by Region, Tactics used by Terrorists, Weapons used by Terrorists, Targeted Victims, Casualties, Countries with Most Terrorist Attacks, Dangerous Countries in 2015, Terrorist Attacks, and Attack Type in one of the most dangerous cities “Baghdad”, the Total number of people killed and wounded in India by Terrorists, Casualties in India and Pakistan, Attacks and Casualties in Selected Countries like India, Pakistan, United States, and Japan.

Code:

```
#Load Packages
library(readr)

## Warning: package 'readr' was built under R version 4.0.5

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.0.5

library(dplyr)

## Warning: package 'dplyr' was built under R version 4.0.5

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(rworldmap)

## Warning: package 'rworldmap' was built under R version 4.0.5

## Loading required package: sp

## Warning: package 'sp' was built under R version 4.0.5

## ### Welcome to rworldmap ###

## For a short introduction type :   vignette('rworldmap')
```

```
library(rworldxtra)

## Warning: package 'rworldxtra' was built under R version 4.0.5

library(maps)

## Warning: package 'maps' was built under R version 4.0.5

library(ggmap)

## Warning: package 'ggmap' was built under R version 4.0.5

## Google's Terms of Service: https://cloud.google.com/maps-platform/terms/.

## Please cite ggmap if you use it! See citation("ggmap") for details.

library(highcharter)

## Warning: package 'highcharter' was built under R version 4.0.5

## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo

#Load Dataset
terrorism<-read.csv("C:\\Users\\SivaLalitha Chikkala\\Downloads\\Global
Terrorism.csv")

#Data Cleaning
terrorism=rename(terrorism,id=eventid,year=iyear,nation=country_txt,Region=re
gion_txt,attack=attacktype1_txt,
                 target=targtype1_txt,weapon=weaptype1_txt,Killed=nkill,
wounded=nwound)

terrorism$Killed=as.integer(terrorism$Killed)
terrorism$wounded=as.integer(terrorism$wounded)

terrorism$Killed[which(is.na(terrorism$Killed))]=0
terrorism$wounded[which(is.na(terrorism$wounded))]=0


#Renaming
US<-filter(terrorism,nation =="United States")
US <- rename(terrorism, long=longitude, lat=latitude)
India<-filter(terrorism,nation=="India")
wEurope<-filter(terrorism,Region=="Western Europe")
Pakistan<-filter(terrorism,nation=="Pakistan")
SEAsia<-rbind(India,Pakistan)

countries<-filter(terrorism,nation %in% c("United
States","India","Pakistan","Japan"))
countries_m<-rbind(countries,wEurope)
```

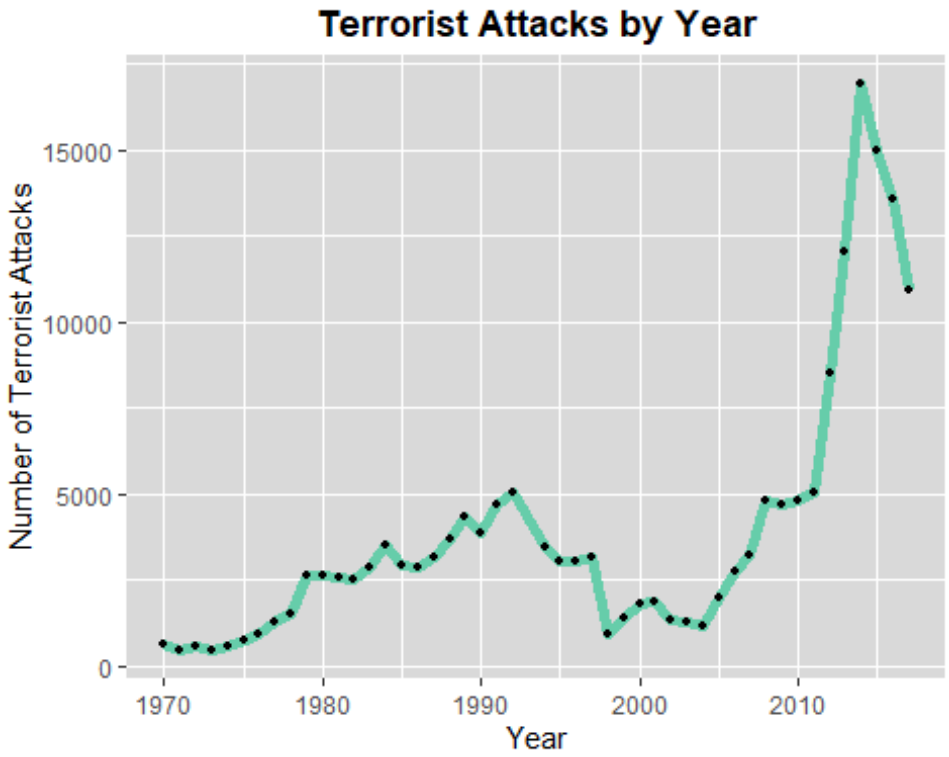
```
#Heatmap of terrorist attack deaths-2015
gtd <- read.csv("C:\\Users\\SivaLalitha Chikkala\\Downloads\\Global
Terrorism.csv")
gtd2015 <- gtd[gtd$iyear==2015, ]
gtd2015 <- aggregate(nkill~country_txt,gtd2015,sum)
gtdMap <- joinCountryData2Map( gtd2015,
                               nameJoinColumn="country_txt",
                               joinCode="NAME" )

## 97 codes from your data successfully matched countries in the map
## 1 codes from your data failed to match with a country code in the map
## 146 codes from the map weren't represented in your data

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



Terrorist Attacks By Year:

```
#Terrorist Attack Trends
by_year<-terrorism %>% group_by(year) %>% summarize(n=n())
ggplot(data=by_year,aes(x = year, y = n)) +
  geom_line(size = 2, alpha = 1, color = "aquamarine3") +
  geom_point(size = 1)+labs(title="Terrorist Attacks by
Year",x="Year",y="Number of Terrorist Attacks")+
```

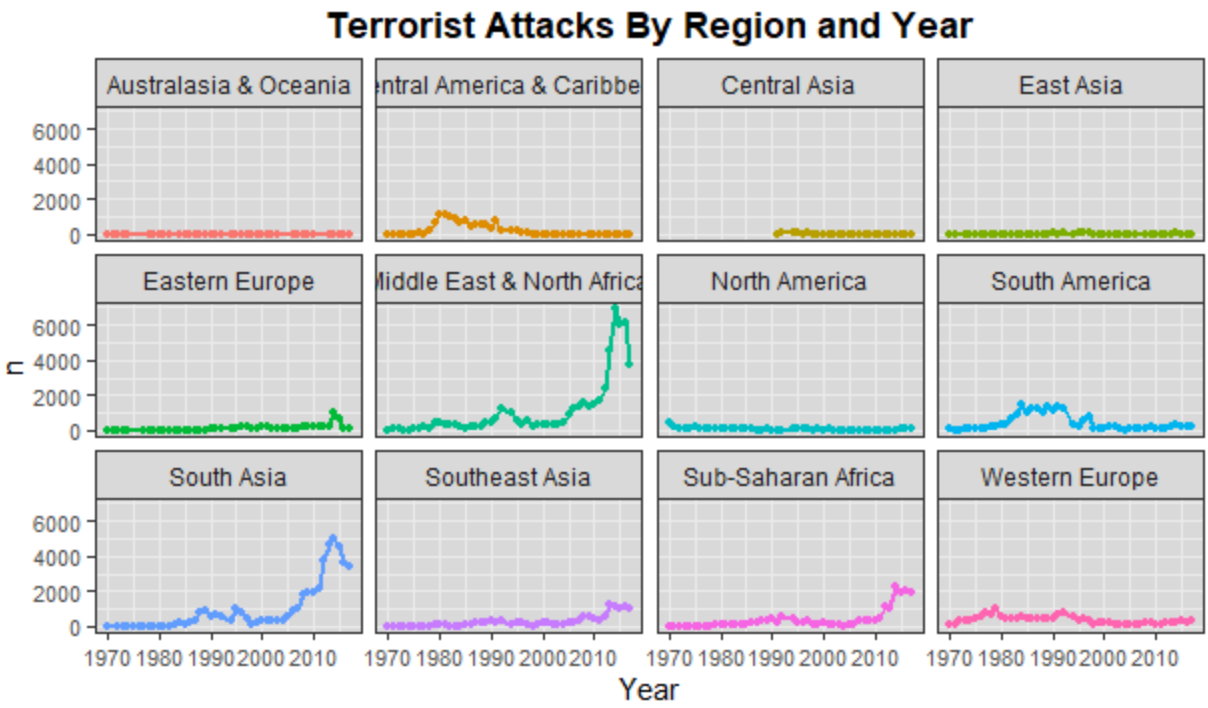
```
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"))
```



➢ Globally, terrorist attacks have increased dramatically since 2010

Terrorist Attacks By Region and Year:

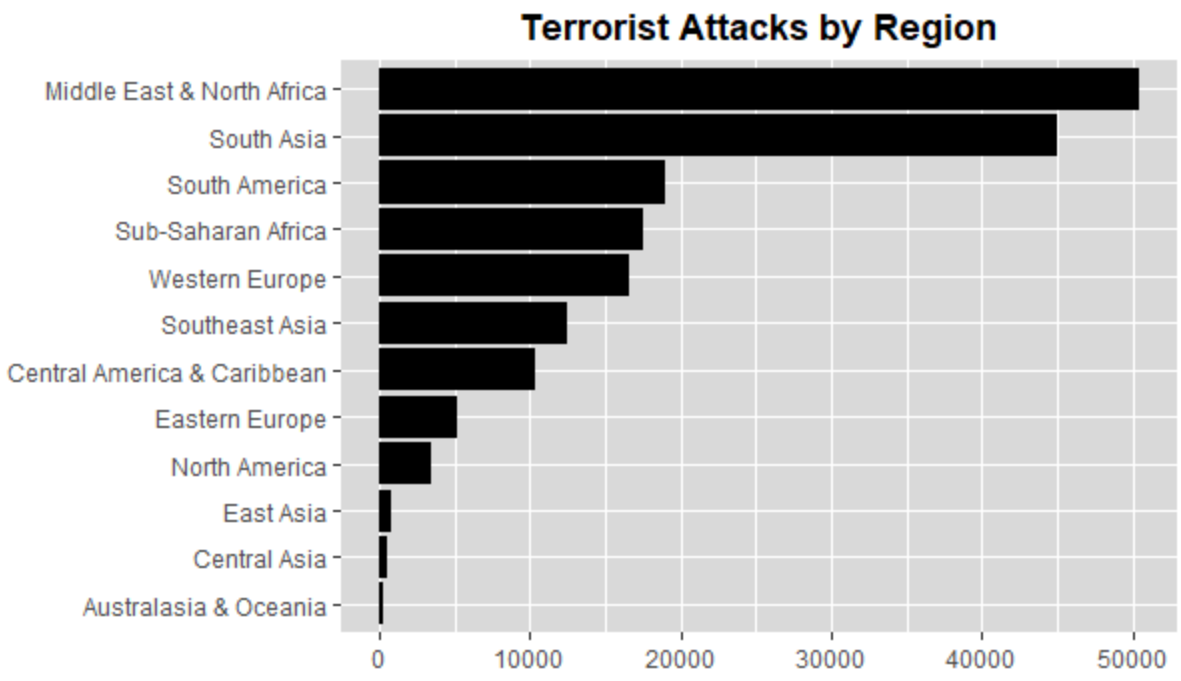
```
#Terrorist Attack Trends By Region n Year
by_region<-terrorism %>% group_by(Region,year) %>% summarize(n=n())

## `summarise()` has grouped output by 'Region'. You can override using the
`.groups` argument.

ggplot(data=by_region,aes(x=year,y=n,color=Region))+
  geom_line(size=1,alpha=1)+
  geom_point(size=1,alpha=1)+
  facet_wrap(~Region)+
  labs(title="Terrorist Attacks By Region and Year",x="Year")+
  theme_bw()+
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"),legend.position =
"none",axis.text=element_text(size=8))
```



- ➢ Central America was very unstable starting from the late 1970s, it got better with time, and it has been stabilized since around 1995.
- ➢ Western Europe had a rough past, experienced many attacks until the early 2000s.
- ➢ South America had a similar pattern, was very dangerous from early 1980 until just before 2000.
- ➢ Middle East, North Africa, and South Asia had a relatively quiet past until around 1980, the terrorist attacks in those regions had a steady increase from the 1980s to around 2005 and had surged dramatically since then.

Terrorist Attacks By Region:

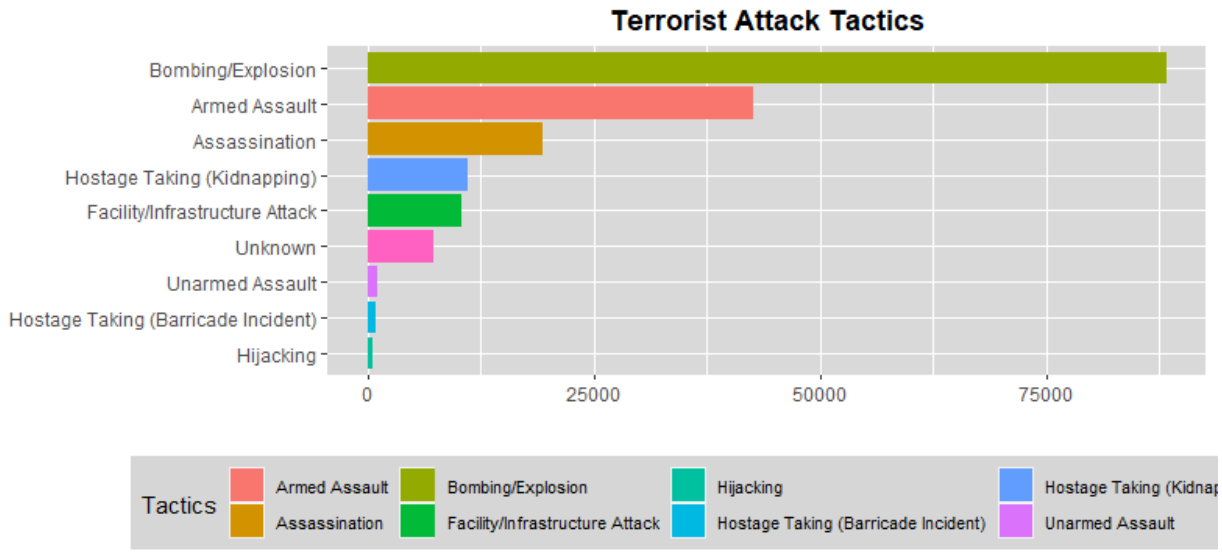
```
#Terrorist Attack Trends By Region
by_region_no_year<-terrorism %>% group_by(Region) %>% summarize(n=n())
ggplot(data=by_region_no_year,aes(x=reorder(Region,n),y=n))+geom_bar(stat =
"identity",fill="black")+
  labs(title="Terrorist Attacks by Region",x="",y="")+
  coord_flip()+
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"))
```



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

Tactics Used:

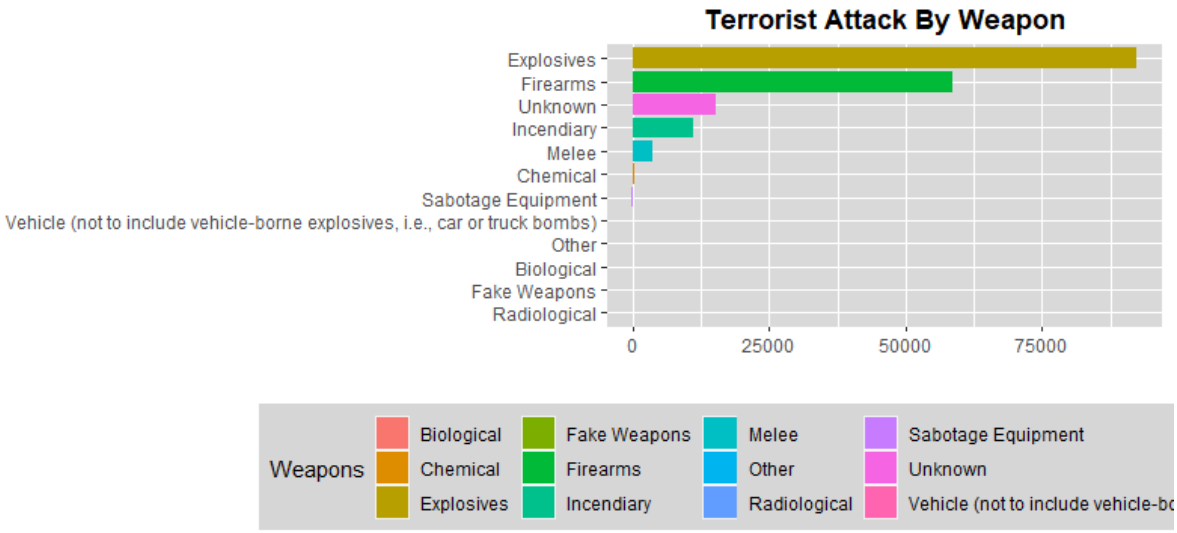
```
#Tactics
by_attack<-terrorism %>% group_by(attack) %>% summarize(n=n())
ggplot(data=by_attack,aes(x=reorder(attack,n),y=n,fill=attack))+
  geom_bar(stat="identity")+
  labs(title="Terrorist Attack Tactics",x="",y="",fill="Tactics")+
  coord_flip()+
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"),legend.background =
element_rect(fill="gray84"),legend.position = "bottom",legend.text =
element_text(size = 8))
```



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

Weapons Used:

```
#Weapons
by_weapon<-terrorism %>% group_by(weapon) %>% summarize(n=n())
ggplot(data=by_weapon,aes(x=reorder(weapon,n),y=n,fill=weapon))+
  geom_bar(stat="identity")+
  labs(title="Terrorist Attack By Weapon",x="",y="",fill="Weapons")+
  coord_flip()+
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"),legend.background =
element_rect(fil="gray84"),legend.position = "bottom")
```



➢ The most Commonly Weapon used are Explosives and followed by Firearms.

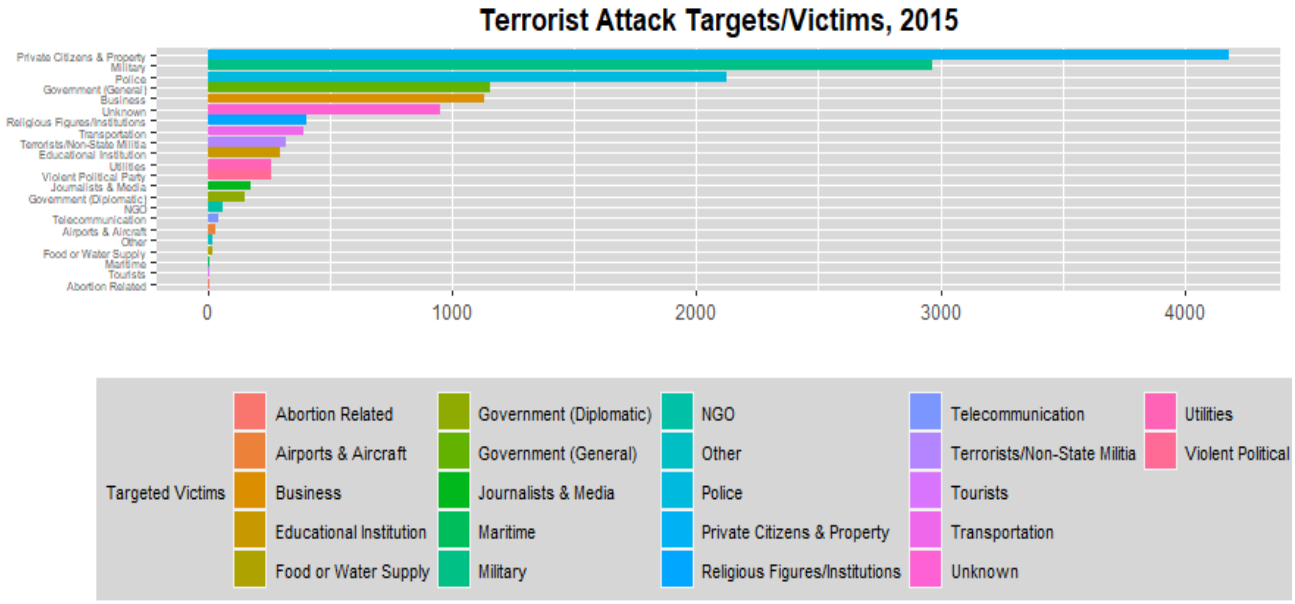
Targeted Victims:

```
#Target Type
attack2015<-terrorism[terrorism$year==2015,]
by_target<-attack2015 %>% group_by(target) %>% summarize(n=n())
by_target<-arrange(by_target,desc(n))
by_target
```

```
## # A tibble: 22 x 2
##    target                           n
##    <chr>                        <int>
##  1 Private Citizens & Property     4184
##  2 Military                        2968
##  3 Police                          2124
##  4 Government (General)            1158
##  5 Business                        1134
##  6 Unknown                          953
##  7 Religious Figures/Institutions   404
##  8 Transportation                   390
##  9 Terrorists/Non-State Militia     320
## 10 Educational Institution          293
## # ... with 12 more rows
```

```
ggplot(data=by_target,aes(x=reorder(target,n),y=n,fill=target))+
  geom_bar(stat="identity")+
  labs(title="Terrorist Attack Targets/Victims,
2015",x="",y="",fill="Targeted Victims")+
  coord_flip()+
```

```
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"),legend.background =
element_rect(fill="gray84"),legend.position = "bottom",axis.text.y =
element_text(size=5),legend.text = element_text(size = 8),legend.title =
element_text(size = 8))
```

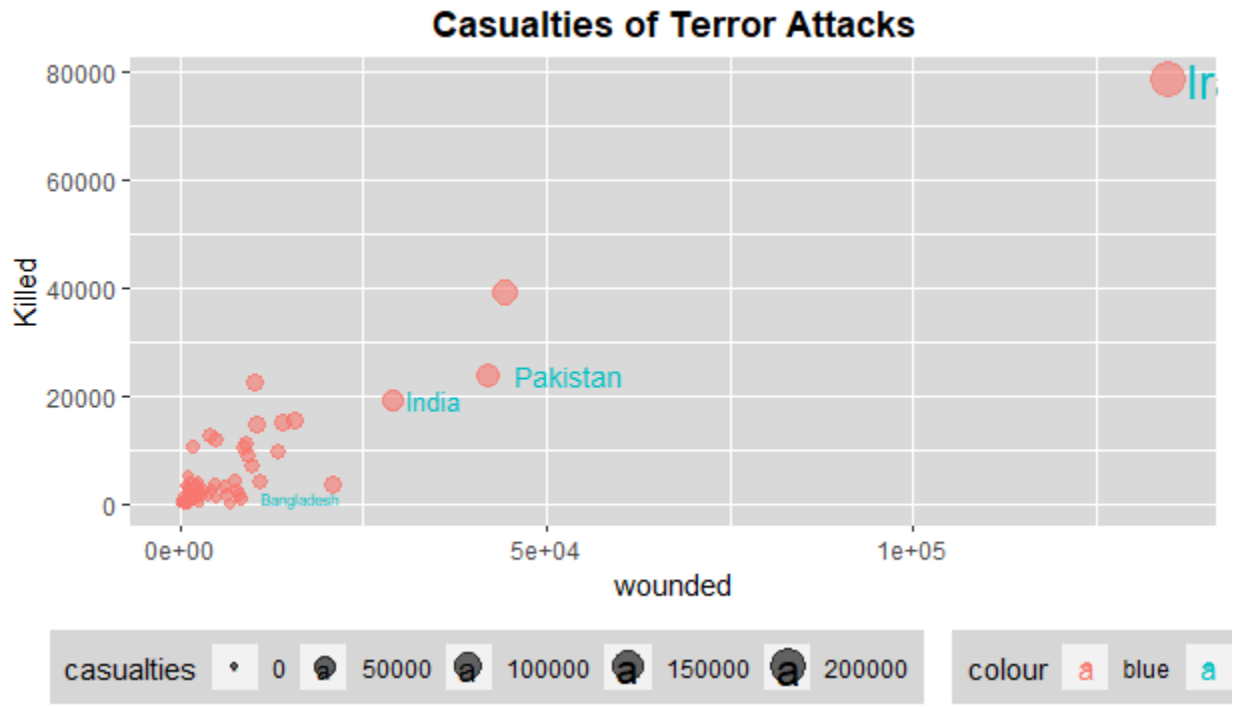


➢ Most Targeted Victims are Private Citizens and Properties and followed by the Military.

Casualties:

```
#Casualties
terror_f=terrorism%>%
  group_by(nation,Region)%>%
  summarise(Killed=sum(Killed),wounded=sum(wounded))%>%
  mutate(casualties=Killed+wounded)%>%

mutate(nation=ifelse(nation=="India","India",ifelse(nation=="Pakistan","Pakis
tan",ifelse(nation=="Iraq","Iraq",ifelse(nation=="Bangladesh","Bangladesh",""
)))))

## `summarise()` has grouped output by 'nation'. You can override using the
`.groups` argument.

posnjd <- position_jitterdodge(jitter.width = 1, dodge.width = 1)

ggplot(terror_f,aes(x=wounded,y=Killed,colour="blue", size=casualties))+
```

```
  geom_point(position = posnjd,alpha=0.6, show.legend = TRUE)+
  labs(title="Casualties of Terror
Attacks",x="Killed",y="Wounded",fill="Casualties")+
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"),legend.background =
element_rect(fill="gray84"),legend.position = "bottom")+
  geom_text(aes(label=nation,hjust=-.25, colour="red"))+
  scale_x_continuous("wounded") +
  scale_y_continuous("Killed")
```



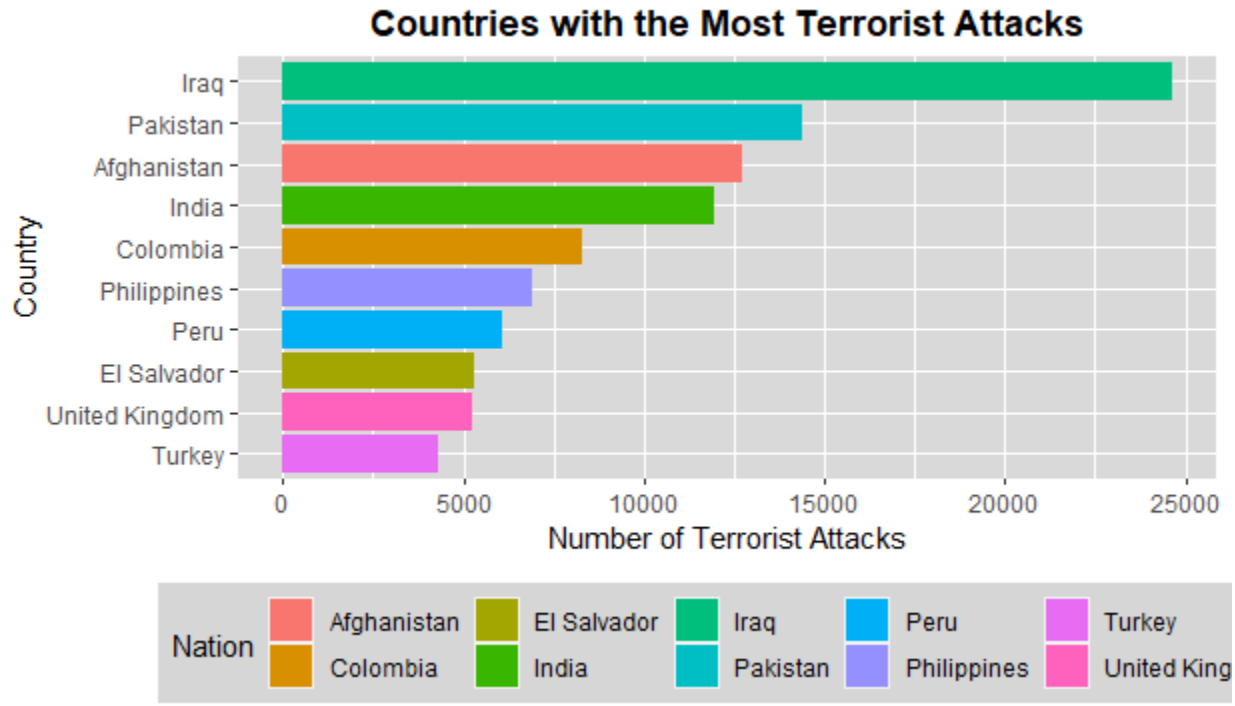
➢ The most number of casualties are in Iraq,Pakistan and followed by Afghanistan and India.

Countries with the most attacks:

```
#Countries with Most Terrorist Attacks
by_country<-terrorism %>% group_by(nation) %>% summarize(n=n())
by_country<-arrange(by_country,desc(n))
top10<-head(by_country,10)
top10

## # A tibble: 10 x 2
##    nation             n
##    <chr>          <int>
##  1 Iraq           24636
##  2 Pakistan       14368
```

```
##  3 Afghanistan    12731
##  4 India          11960
##  5 Colombia        8306
##  6 Philippines     6908
##  7 Peru            6096
##  8 El Salvador     5320
##  9 United Kingdom  5235
## 10 Turkey          4292
```

```
ggplot(data=top10,aes(x=reorder(nation,n),y=n,fill=nation))+
  geom_bar(stat="identity")+
  labs(title="Countries with the Most Terrorist
Attacks",x="Country",y="Number of Terrorist Attacks",fill="Nation")+
  coord_flip()+
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"),legend.position =
"bottom",legend.background = element_rect(fill="gray84"))
```



Dangerous Countries in 2015:

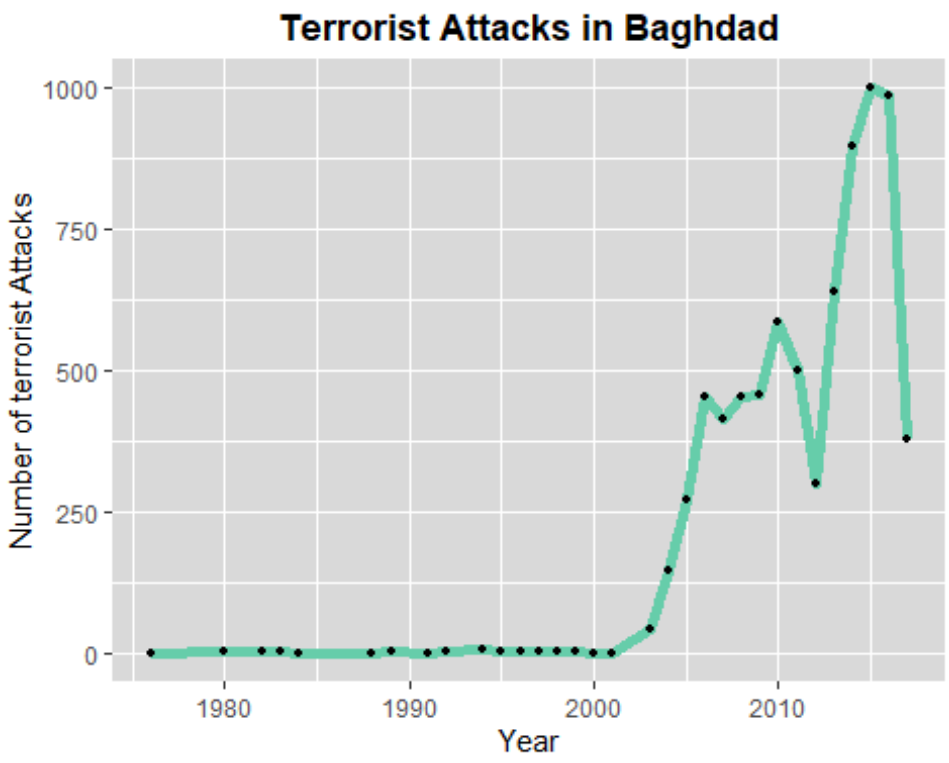
```
#Which countries/cities were the most dangerous in 2015?
attack2015_bycity<-attack2015 %>% group_by(nation,city) %>% summarise(n=n())
```

```
## `summarise()` has grouped output by 'nation'. You can override using the
`.groups` argument.
```

```
attack2015_bycity<-arrange(attack2015_bycity,desc(n))
top10_city_2015 <- head(attack2015_bycity, 20)
top10_city_2015
```

```
## # A tibble: 20 x 3
## # Groups:   nation [10]
##    nation      city              n
##    <chr>       <chr>         <int>
##  1 Iraq        Baghdad        1000
##  2 Bangladesh  Dhaka           211
##  3 Afghanistan Unknown         160
##  4 Libya       Benghazi        151
##  5 Pakistan    Karachi         146
##  6 Egypt       Arish           107
##  7 Somalia     Mogadishu        95
##  8 Yemen       Aden             95
##  9 Yemen       Unknown          95
## 10 Yemen       Sanaa            94
## 11 Egypt       Sheikh Zuweid    88
## 12 Iraq        Ramadi           87
## 13 Yemen       Taizz            87
## 14 Syria       Damascus         86
## 15 Pakistan    Quetta           84
## 16 Syria       Aleppo           83
## 17 Afghanistan Kabul            80
## 18 Egypt       Unknown          77
## 19 Burundi     Bujumbura        74
## 20 Libya       Sirte            69
```

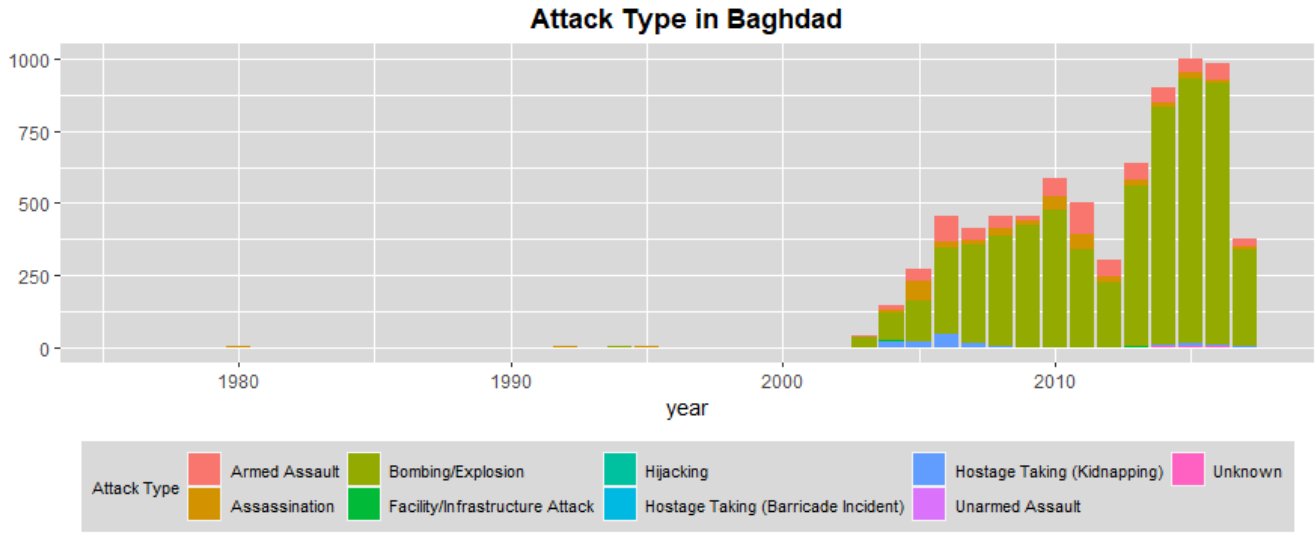
```
#Terrorist Attacks in Baghdad
baghdad <- terrorism[terrorism$city=='Baghdad', ]
baghdad_year <- baghdad %>% group_by(year) %>% summarise(n=n())
ggplot(data=baghdad_year,aes(x=year,y = n))+
  geom_line(size = 2, alpha = 1, color = "aquamarine3") +
  geom_point(size = 1)+
  labs(title="Terrorist Attacks in Baghdad",x="Year",y="Number of terrorist
Attacks")+
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"))
```



➢ Baghdad is one of the most dangerous cities in the world and terrorist attacks have drastically increased from the year 2000.

Attack Type in Baghdad:

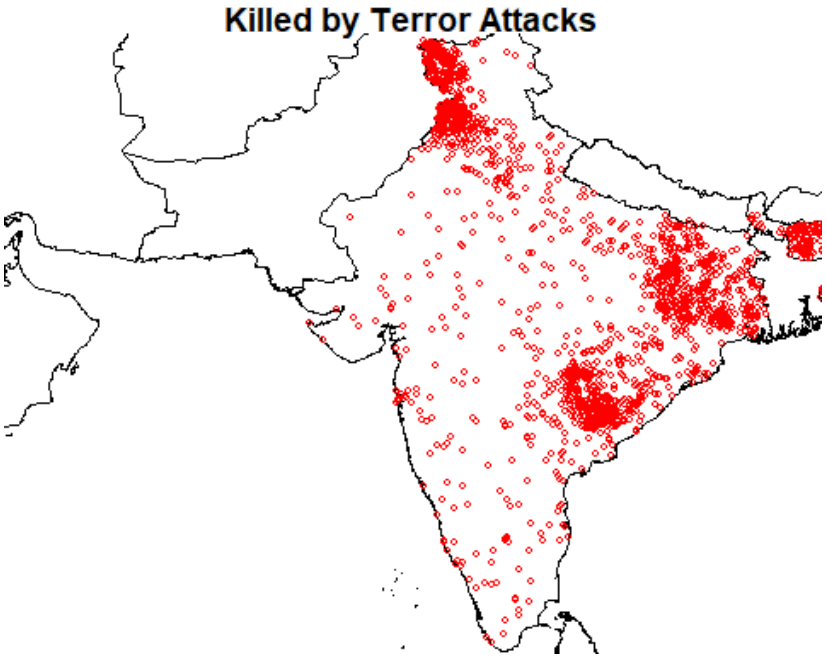
```
#Attack type in Baghdad
baghdad_type <- baghdad %>% group_by(attack, year) %>% summarise(n=n())

## `summarise()` has grouped output by 'attack'. You can override using the
`.groups` argument.

ggplot(data=baghdad_type,aes(x=year,y=n,fill=attack))+
  geom_bar(stat="identity")+
  labs(title="Attack Type in Baghdad",x="year",y="",fill="Attack Type")+
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"),legend.background =
element_rect(fill="gray85"),legend.position =
"bottom",legend.text=element_text(size=8),legend.title = element_text(size =
8))
```



➢ Most of the Attack Type in Baghdad is Bombing and Explosion.

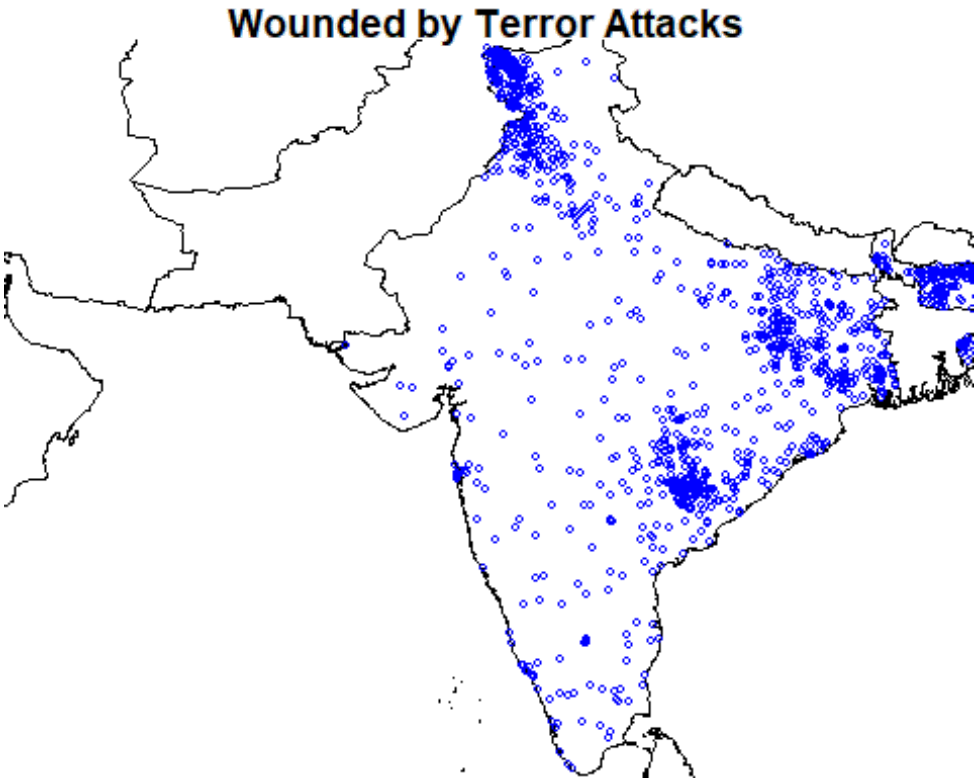
Killed in India:

```
#Killed In India
IndiaK=filter(India,Killed!=0)
india_map<-getMap(resolution="high")
plot(india_map, xlim = c(72,75), ylim = c(8, 35), asp = 1,main = "Killed by
Terror Attacks")
points(IndiaK$longitude,IndiaK$latitude,col="red",cex=.6)
```



Wounded in India:

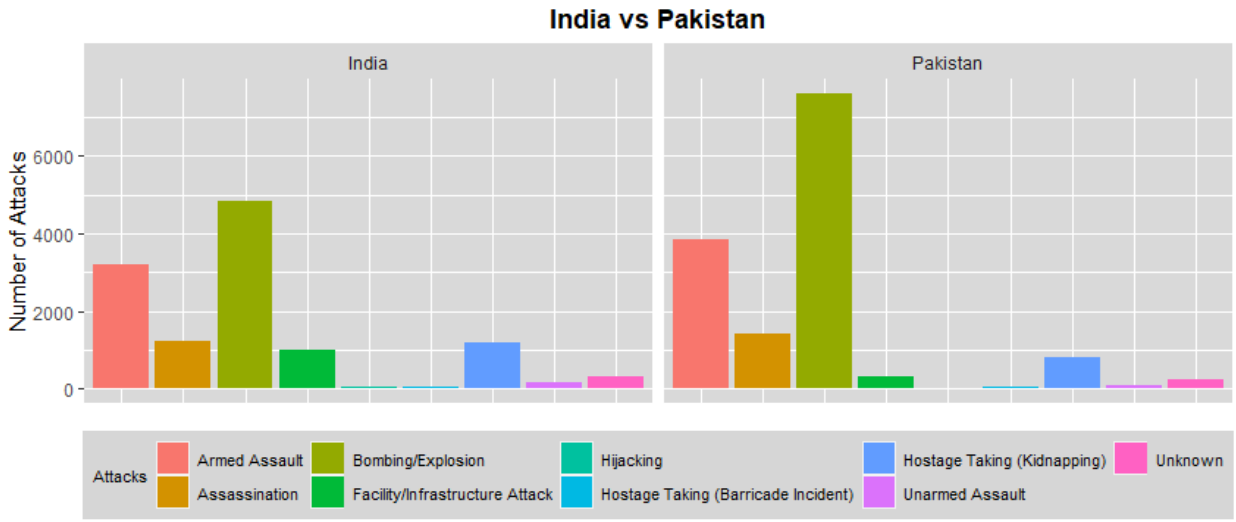
```
#Wounded in India
IndiaW<-filter(India,wounded!=0)
india_map<-getMap(resolution="high")
plot(india_map, xlim = c(73,75), ylim = c(8, 35), asp = 1,main = "Wounded by
Terror Attacks")
points(IndiaW$longitude,IndiaW$latitude,col="blue",cex=.6)
```



Attacks in India & Pakistan:

```
#India vs Pakistan
ggplot(SEAsia,aes(x=attack,fill=attack))+
  geom_bar(position = "dodge")+
  labs(title="India vs Pakistan",y="Number of Attacks",x="Type of
Attacks",fill="Attacks")+
  facet_grid(~nation)+
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"),legend.position =
"bottom",legend.background = element_rect(fill="gray84"),legend.text
=element_text(size = 8), legend.title = element_text(size = 8))+
  theme(axis.title.x=element_blank(),
```

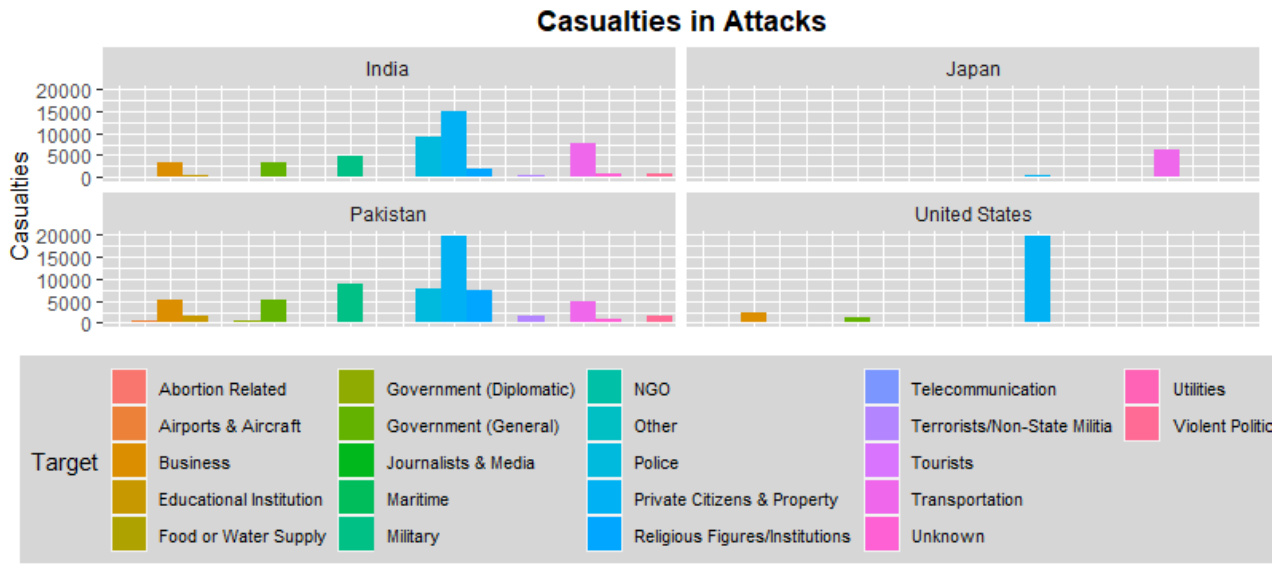
```
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())
```



➢ Pakistan has more explosions when compared to India whereas India has more Infrastructure Attacks and Barricade incidents when compared to Pakistan.

Attacks in Selected Countries:

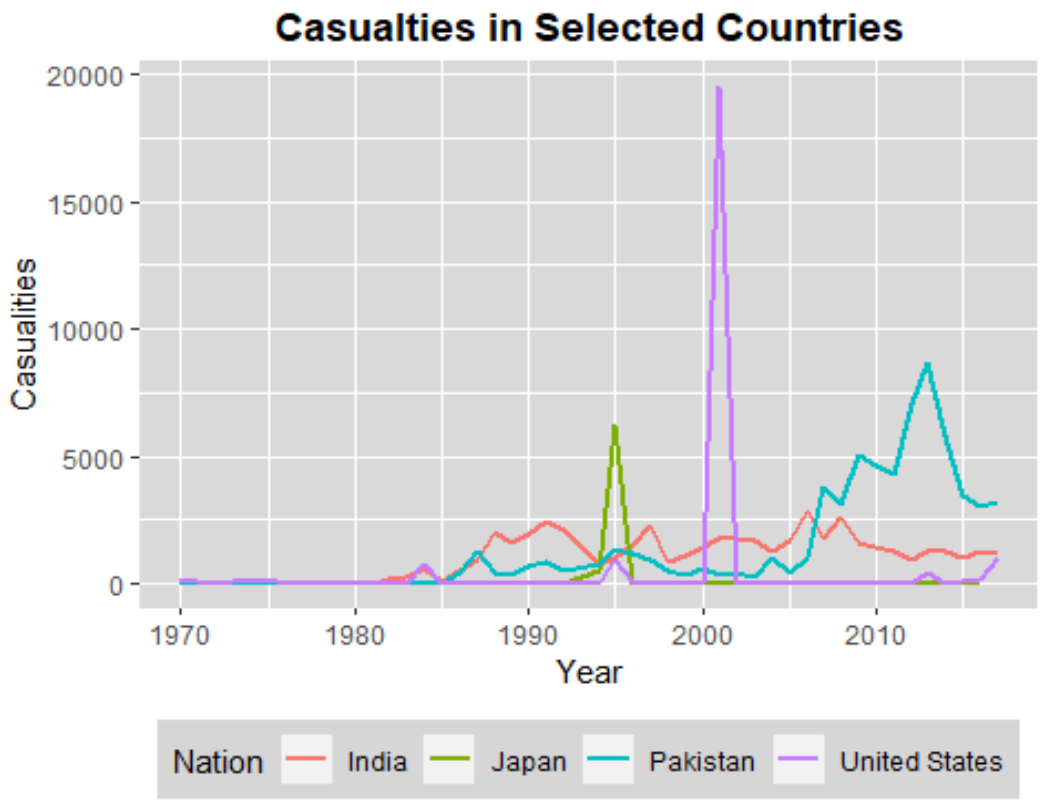
```
#Attacks in Selected Countries
countries_f<-countries%>%
  group_by(nation,target)%>%
  summarise(Killed=sum(Killed),wounded=sum(wounded))%>%
  mutate(casualties = Killed+wounded)

## `summarise()` has grouped output by 'nation'. You can override using the
`.groups` argument.

ggplot(countries_f,aes(x=target,y=casualties,fill=target))+
  geom_bar(width = 1,stat="identity")+
  facet_wrap(~nation)+
  labs(title="Casualties in Attacks",y="Casualties",fill="Target")+
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"),legend.background =
element_rect(fill="gray84"),legend.position = "bottom",legend.text =
element_text(size=8))+
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())
```



➢ Pakistan has the highest number of casualties and Japan has the least number of casualties.

Casualties in Selected Countries:

```
#Casualties in Selected Countries
countries_s<-countries%>%
  group_by(nation,year)%>%
  summarise(Killed=sum(Killed),wounded=sum(wounded))%>%
  mutate(casualties=Killed + wounded)

## `summarise()` has grouped output by 'nation'. You can override using the
`.groups` argument.

ggplot(countries_s,aes(x=year,y=casualties,group=nation))+
  geom_line(aes(color=nation),size=1)+
  labs(title="Casualties in Selected
Countries",x="Year",y="Casualities",color="Nation")+
  theme(panel.background = element_rect(fill="gray85"),plot.title =
element_text(hjust=0.5,face="bold",color="black"),legend.background =
element_rect(fill="gray84"),legend.position = "bottom")
```



- Japan has more casualties in the year 1995.
- The United States has more casualties in the year 2001.
- Pakistan has more casualties after 2005.

Data Source:

- Dataset Link

References:

- Explore Global Terrorist Attacks
- Global Terrorist Analysis
- Global Terrorist Attacks
- Terror Attacks