GRIP Task 4

Exploratory Data Analysis of Global Terrorism Data

Samima Yasmin

Problem:

- As a security/ defense analyst my task is to find out **hot zone of terrorism** in the world.
- To check what all security issues and insights can be derived by EDA.

Data Preprocessing

We will first load the data file.

```
terror<-read.csv("C://Users//samim//OneDrive//Desktop//GRIP//globalterrorismdb_0718dist.csv")
```

Let's check the dimension of data.

```
dim(terror)
```

```
## [1] 181691    135
```

```
names(terror)
```

```
##   [1] "eventid"           "iyear"             "imonth"
##   [4] "iday"              "approxdate"        "extended"
##   [7] "resolution"        "country"           "country_txt"
##  [10] "region"            "region_txt"        "provstate"
##  [13] "city"              "latitude"          "longitude"
##  [16] "specificity"       "vicinity"          "location"
##  [19] "summary"           "crit1"             "crit2"
##  [22] "crit3"             "doubtterr"         "alternative"
##  [25] "alternative_txt"   "multiple"          "success"
##  [28] "suicide"           "attacktype1"       "attacktype1_txt"
##  [31] "attacktype2"       "attacktype2_txt"   "attacktype3"
##  [34] "attacktype3_txt"   "targtype1"         "targtype1_txt"
##  [37] "targsubtype1"      "targsubtype1_txt"  "corp1"
##  [40] "target1"           "natlty1"           "natlty1_txt"
##  [43] "targtype2"         "targtype2_txt"     "targsubtype2"
##  [46] "targsubtype2_txt"  "corp2"             "target2"
##  [49] "natlty2"           "natlty2_txt"       "targtype3"
##  [52] "targtype3_txt"     "targsubtype3"      "targsubtype3_txt"
##  [55] "corp3"             "target3"           "natlty3"
##  [58] "natlty3_txt"       "gname"             "gsubname"
##  [61] "gname2"            "gsubname2"         "gname3"
##  [64] "gsubname3"         "motive"            "guncertain1"
##  [67] "guncertain2"       "guncertain3"       "individual"
```

```
##  [70] "nperps"             "nperpcap"           "claimed"
##  [73] "claimmode"          "claimmode_txt"      "claim2"
##  [76] "claimmode2"         "claimmode2_txt"     "claim3"
##  [79] "claimmode3"         "claimmode3_txt"     "compclaim"
##  [82] "weaptype1"          "weaptype1_txt"      "weapsubtype1"
##  [85] "weapsubtype1_txt"   "weaptype2"          "weaptype2_txt"
##  [88] "weapsubtype2"       "weapsubtype2_txt"   "weaptype3"
##  [91] "weaptype3_txt"      "weapsubtype3"       "weapsubtype3_txt"
##  [94] "weaptype4"          "weaptype4_txt"      "weapsubtype4"
##  [97] "weapsubtype4_txt"   "weapdetail"         "nkill"
## [100] "nkillus"            "nkillter"           "nwound"
## [103] "nwoundus"           "nwoundte"           "property"
## [106] "propextent"         "propextent_txt"     "propvalue"
## [109] "propcomment"        "ishostkid"          "nhostkid"
## [112] "nhostkidus"         "nhours"             "ndays"
## [115] "divert"             "kidhijcountry"      "ransom"
## [118] "ransomamt"          "ransomamtus"        "ransompaid"
## [121] "ransompaidus"       "ransomnote"         "hostkidoutcome"
## [124] "hostkidoutcome_txt" "nreleased"          "addnotes"
## [127] "scite1"             "scite2"             "scite3"
## [130] "dbsource"           "INT_LOG"            "INT_IDEO"
## [133] "INT_MISC"           "INT_ANY"            "related"
```

We will work with a subset of data for our rest of analysis that is columns having data of our interest.

```
terror <- terror[, c("iyear", "imonth", "country_txt", "region_txt","city","latitude",
          "longitude", "attacktype1_txt","targtype1_txt","nkill", "nwound", "gname")]
```

Let's check if there are missing values.

```
missing_values_sum <- colSums(is.na(terror))
print(missing_values_sum)
```

```
##           iyear          imonth     country_txt      region_txt            city
##               0               0               0               0               0
##        latitude       longitude attacktype1_txt   targtype1_txt           nkill
##            4556            4557               0               0           10313
##          nwound           gname
##           16311               0
```

So, there are missing values in latitude ,longitude ,no of killed and wounded in this incidents.

Now, we'll see the structure of data.

```
str(terror)
```

```
## 'data.frame':    181691 obs. of  12 variables:
##  $ iyear          : int  1970 1970 1970 1970 1970 1970 1970 1970 1970 1970 ...
##  $ imonth         : int  7 0 1 1 1 1 1 1 1 1 ...
##  $ country_txt    : chr  "Dominican Republic" "Mexico" "Philippines" "Greece" ...
##  $ region_txt     : chr  "Central America & Caribbean" "North America" "Southeast Asia" "Western Eur
##  $ city           : chr  "Santo Domingo" "Mexico city" "Unknown" "Athens" ...
##  $ latitude       : num  18.5 19.4 15.5 38 33.6 ...
##  $ longitude      : num  -70 -99.1 120.6 23.8 130.4 ...
##  $ attacktype1_txt: chr  "Assassination" "Hostage Taking (Kidnapping)" "Assassination" "Bombing/Expl
##  $ targtype1_txt  : chr  "Private Citizens & Property" "Government (Diplomatic)" "Journalists & Media
##  $ nkill          : int  1 0 1 NA NA 0 0 0 0 0 ...
##  $ nwound         : num  0 0 0 NA NA 0 0 0 0 0 ...
##  $ gname          : chr  "MANO-D" "23rd of September Communist League" "Unknown" "Unknown" ...
```

Exploratory Data Analysis:

Data Analysis

Now we will find the country, city, region, year, month, group, and attack type with the most attacks.

```
most_attacks_country <- names(sort(table(terror$country_txt), decreasing = TRUE)[1])
most_attacks_city <- names(sort(table(terror$city), decreasing = TRUE)[-1])[1]
most_attacks_region <- names(sort(table(terror$region_txt), decreasing = TRUE)[1])
most_attacks_year <- names(sort(table(terror$iyear), decreasing = TRUE)[1])
most_attacks_month <- names(sort(table(terror$imonth), decreasing = TRUE)[1])
most_attacks_group <- names(sort(table(terror$gname), decreasing = TRUE)[-1])[1]
most_attacks_targets <- names(sort(table(terror$targtype1_txt), decreasing = TRUE)[1])
most_common_attack_type <- names(sort(table(terror$attacktype1_txt), decreasing = TRUE)[1])

# Print the results
cat("Country:", most_attacks_country, "\nCity:", most_attacks_city, "\nRegion:",
    most_attacks_region, "\nYear:", most_attacks_year, "\nMonth:", most_attacks_month,
    "\nGroup:", most_attacks_group, "\nTargets:",most_attacks_targets,
    "\nAttack Type:" , most_common_attack_type, "\n")
```

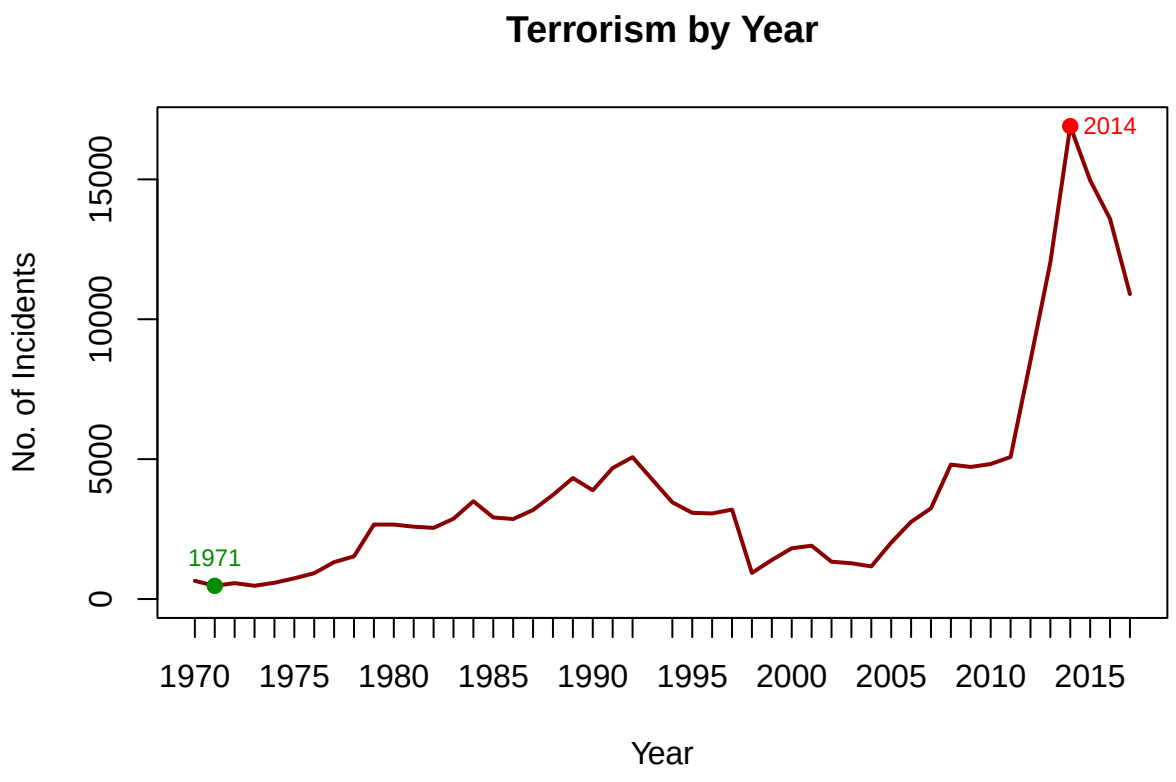
```
## Country: Iraq
## City: Baghdad
## Region: Middle East & North Africa
## Year: 2014
## Month: 5
## Group: Taliban
## Targets: Private Citizens & Property
## Attack Type: Bombing/Explosion
```

Data Visualization

Attacks by Year:

The below chart describes the number of attacks by year.

```
# Create a table of the number of attacks by year
attack_year <- table(terror$iyear)

# Plot the number of incidents by year
plot(attack_year, type='l',lwd=2, main='Terrorism by Year', xlab='Year',
     ylab='No. of Incidents', col='red4')

# Find the year with the highest and lowest number of attacks
max_year <- as.numeric(names(which.max(attack_year)))
min_year <- as.numeric(names(which.min(attack_year)))
max_attacks <- max(attack_year)
min_attacks <- min(attack_year)

# Add text labels for the year names at the arrow points
text(c(max_year+2, min_year), c(max_attacks , min_attacks +1000),
     labels = c(max_year, min_year), col = c('red', 'green4'), cex = 0.75)

points(max_year, max_attacks, col='red', pch=19)
points(min_year, min_attacks, col='green4', pch=19)
```



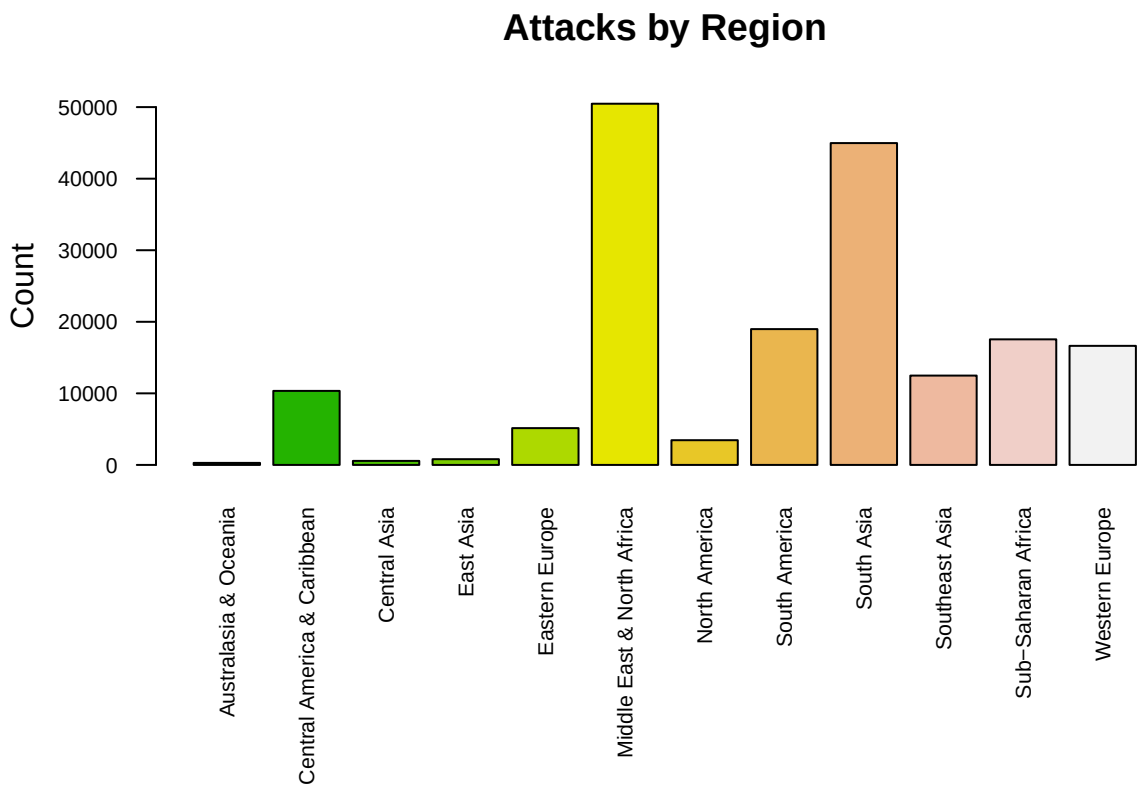
Interpretation

From the above chart, we observe that no. of attacks increased over the years since 1971 till the early 90s. Then, there was a decline in the no. of attacks till the late 90s. Moving forward, we again see increase in the no. of attacks since the late 90s till 2014(which also records the highest number of attacks) and there after we note a decline in the no. of attacks.

Attacks by Region:

We next look at the frequency of attacks based on regions.

```
attack_reg<-table(terror$region_txt)
par(mar=c(9,5,4,1))
barplot(attack_reg,las=2,cex.axis = 0.7,cex.names=0.7, col=terrain.colors(nrow(attack_reg)),
        ylab='Count',main="Attacks by Region")
```



Interpretation

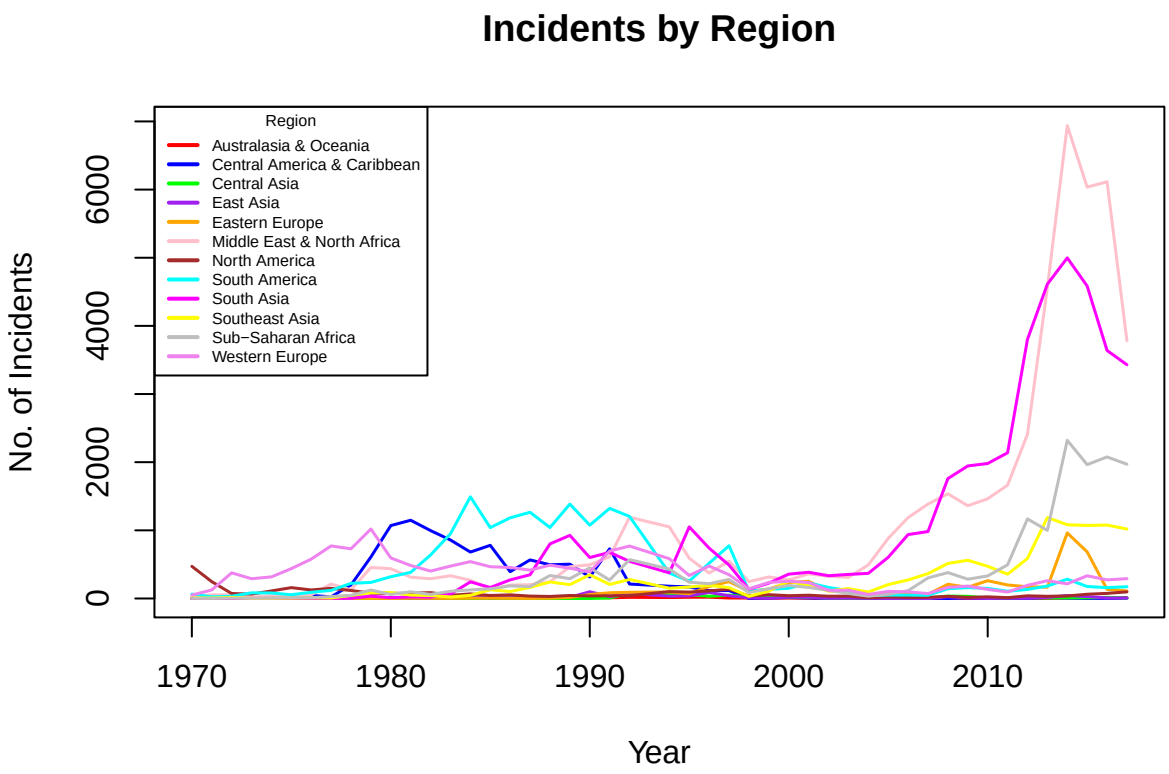
Most affected region by terrorism in the world is **Middle East & North Africa** followed closely by **South Asia**.

Attacks by Region over the years

We now look at the attacks over the years on various regions.

```
# To create a table of the number of attacks by region and year
reg_year <- table(terror$region_txt, terror$iyear)
```

```r
ylim <- range(reg_year)
xlim <- c(1970, 2017)
custom_colors <- c("red", "blue", "green", "purple", "orange", "pink", "brown", "cyan",
                   "magenta", "yellow", "gray", "violet", "turquoise")

# Plot the first region
plot(x = colnames(reg_year), y = reg_year[1, ], type = 'l' , ylim = ylim, xlim = xlim,
     col = custom_colors[1], xlab = 'Year',
     ylab = 'No. of Incidents', main = 'Incidents by Region',lwd= 1.5)

# Loop through the other regions
for (i in 2:nrow(reg_year))
    {
  lines(x = colnames(reg_year), y = reg_year[i, ], type = 'l',
        col = custom_colors[i],lwd=1.5)
    }

# To Add a legend
legend("topleft", legend = rownames(reg_year), lwd = 2,
       col = custom_colors, cex = 0.5, title = 'Region')
```



Observation

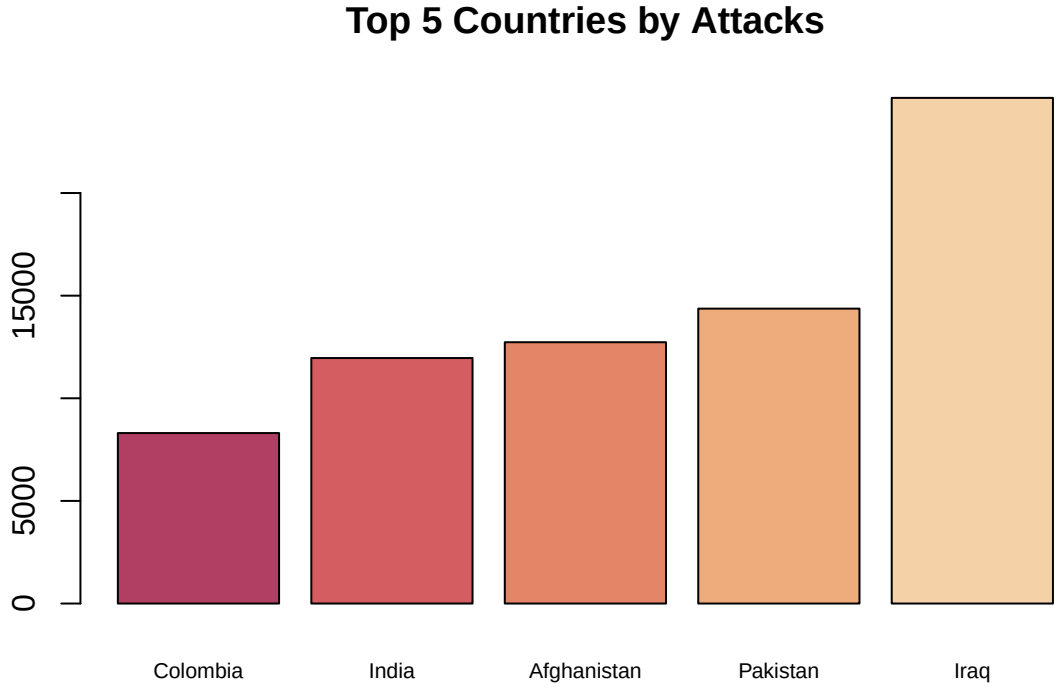
- While initially, **South America** saw more unrests in the period between 1980 to 2000.
- South Asia and Middle East & North Africa saw a huge rise in terror activities from around 2003-2004 which has only increased in upcoming years.

- We also notice that in the year 2014, there has been a global unrest.
- After that there is significant decrease overall except **South-East Asia**.

Attacks by Country:

The below barplot shows the most attacked countries.

```
country<-table(terror$country_txt)
max_country<-tail(sort(country),5)
barplot(max_country,col=hcl.colors(5,'RedOr'),main='Top 5 Countries by Attacks',cex.names = 0.7)
```



Observation:

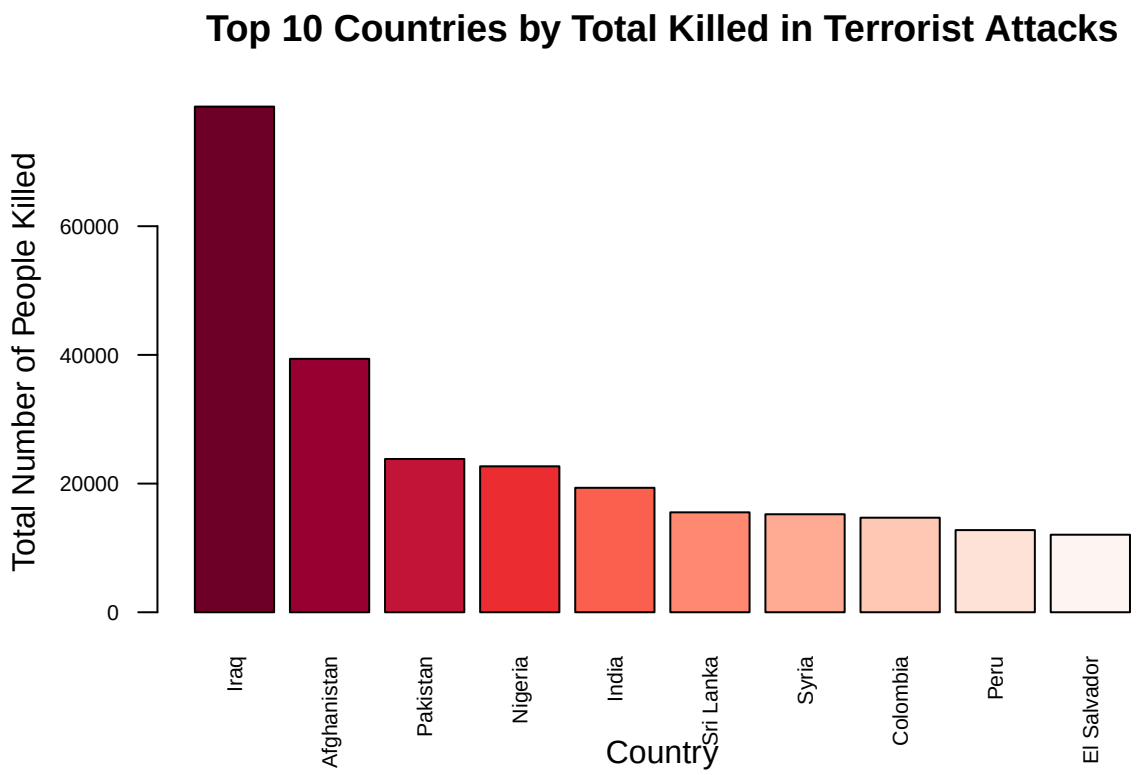
- We find that Iraq is the most affected country in the world.
- It has a huge gap with Pakistan which is placed second in the list.

Loss of life in Terrorist Attacks:

Now we will see Top 10 Countries by Total Killed humans in Terrorist Attacks.

```
country_kills <- aggregate(nkill ~ country_txt, data = terror, FUN = sum, na.rm = TRUE)
colnames(country_kills) <- c("Country", "Total_Killed")
# To Select the top 10 countries with the highest total number of people killed
top_10_countries <- head(country_kills[order(-country_kills$Total_Killed), ], 10)
barplot(top_10_countries$Total_Killed, names.arg = top_10_countries$Country,
        col = hcl.colors(10, 'Reds'), las = 2,
        main = 'Top 10 Countries by Total Killed in Terrorist Attacks',
```

```
        xlab = 'Country', ylab = 'Total Number of People Killed',
        cex.names  = 0.7,cex.axis = 0.7)
```

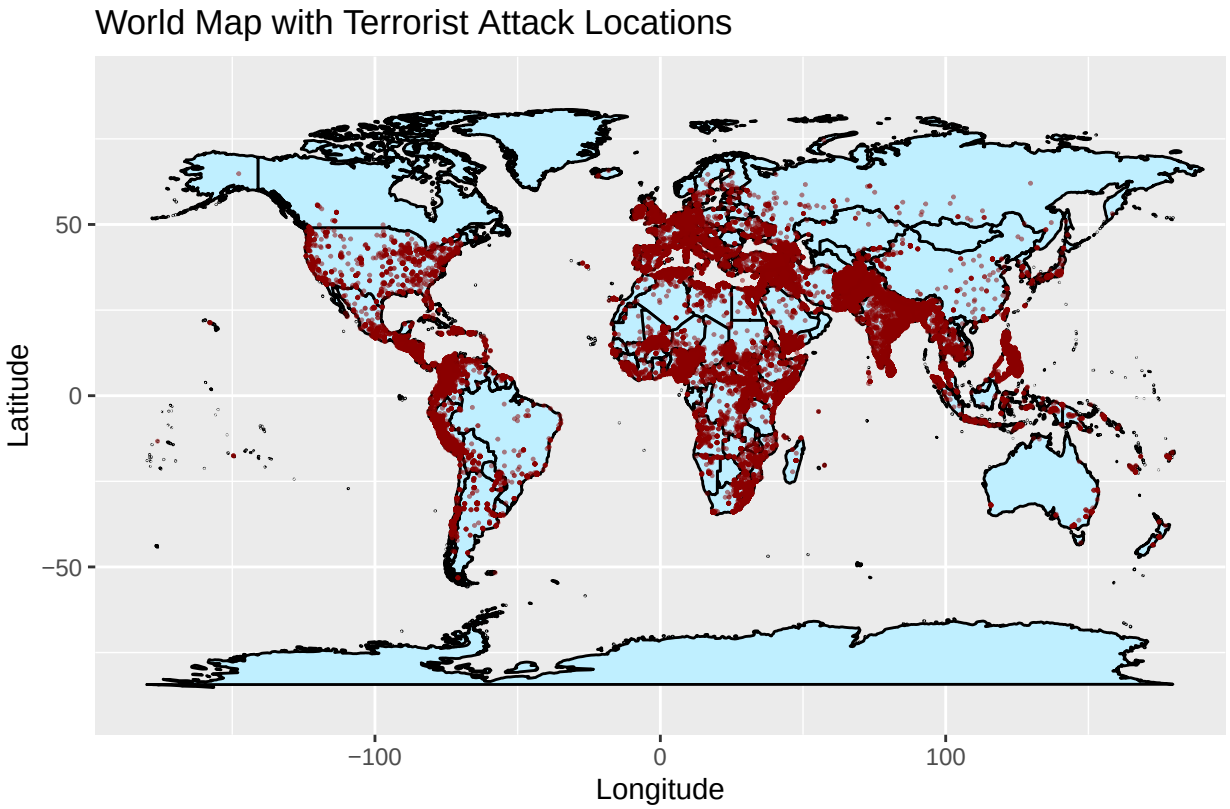


Spread of Terrorist Attacks Globally

Now we will visualize Terrorist Attack Locations in **World Map**.

```
# Load required libraries
library(ggplot2)
library(maps)

# Get world map data
world_map <- map_data("world")

# To create the base world map plot
base_map <- ggplot() +
  geom_polygon(data = world_map, aes(x = long, y = lat, group = group),
               fill = "lightblue1", color = "black") +

  # To add terrorist attack locations as points
  geom_point(data = terror, aes(x = longitude, y = latitude),
             color = "red4", size = 0.2, alpha = 0.5) +

  # To Set the aspect ratio and limits for the map
  coord_fixed(ratio = 1.2, xlim = c(-180, 180), ylim = c(-90, 90)) +
```

```r
  # To add labels and title
  labs(title = "World Map with Terrorist Attack Locations",
       x = "Longitude", y = "Latitude")

# Display the map
print(base_map)
```



Conclusion from the EDA:

- The **Middle East region** is the **hot zone of terrorism** in the world with **Iraq** being the **most affected** one in this region (and across the globe).
- Since 2014, the overall terror activity across the globe has **decreased significantly** till the time point we have data that is 2017.
- Most of the terrorism activities were done by **Taliban**.
- **Bombing/ Explosion** is attack type mostly used.
- **Private Citizens & Property** are most vulnerable targets of terrorist attacks.