Analysis on Terrorism & Happiness

Guangyan Yu

Introduction

This report focused on global terrorism happened in 2017, and also explored the relationship between terrorism and happiness. Besides, With weird apperance in generosity data(countries with more attacks owned higher satisfactory scores in generosity), I tested data with Benford Law.

Data

Data Source

Terrorism data: The Global Terrorism Database

Happiness data: The happiness scores and rankings use data from the Gallup World Poll.

Data Structure

Terrorism data:

- eventid
- iyear
- imonth
- iday
- country_txt: country name
- region_txt: region name
- city
- longitude
- latitude
- attacktype1_txt: attack type
- targtype1_txt: target of the attack
- nkill: number of people been killed

- nwound: number of people been wounded
- summary: description of the attack

Happiness data:

- Country: country name
- Happiness Rank (the smaller the happier)
- Happiness Score: sum of the following 6 scores. (the higher the happier)
- GDP per Capital: score of economy (the higher the better)
- Family: score of family (the higher the better)
- Life Expectancy: score of health (the higher the better)
- Freedom: score of freedom (the higher the better)
- Generosity: score of generosity (the higher the better)
- Trust Government Corruption: score of trustness to government (the higher the better)

EDA

Overview(Map)

```r
library(tidyverse)
library(magrittr)
terrorism_2017<-read.csv("../Data/terrorism_2017.csv")
happiness_2017<-read.csv("../Data/2017.csv")
#data<-inner_join(terrorism,happiness_2017,by=c("country_txt"="Country"))
#data %<>% filter(iyear==2017)
```

```r
#library(shiny)
library(leaflet)
library(rgdal)
library(RColorBrewer)
library(plotly)

## Download data from Natural Earth
url <- "https://www.naturalearthdata.com/http//www.naturalearthdata.com/download/50m/cultural/ne_50m_admin_0_countries.zip"

tmp <- tempdir()
```

```r
file <- basename(url)

download.file(url, file)

unzip(file, exdir = tmp)

## Read the data into R
countries <- readOGR(dsn=tmp,
                     layer = "ne_50m_admin_0_countries",
                     encoding = "UTF-8",
                     verbose=FALSE)
country_name<-intersect(countries$SUBUNIT,terrorism_2017$country_txt)
country_name<-intersect(country_name,happiness_2017$Country)

t_2017<-sp::merge(countries,happiness_2017%>%filter(Country%in%country
y_name),
                  by.y="Country",by.x="SUBUNIT",sort=FALSE,duplicateG
eoms =
                  TRUE,all.x=FALSE)

#map
pal <- colorNumeric("YlOrRd",domain = t_2017$Happiness.Score)
labels <- sprintf(
  "<strong>%s</strong><br/>%s Happiness",
  t_2017$SUBUNIT, round(t_2017$Happiness.Score,digits=2)
) %>% lapply(htmltools::HTML)

terrorism_2017_plot<-terrorism_2017 %>% filter(country_txt %in% count
ry_name)

leaflet() %>%
  addTiles() %>%

  addPolygons(data=t_2017,
  fillColor = ~pal(t_2017$Happiness.Score),
  weight = 2,
  opacity = 1,
  color = "white",
  dashArray = "3",
  fillOpacity = 0.5,
  label = labels
```
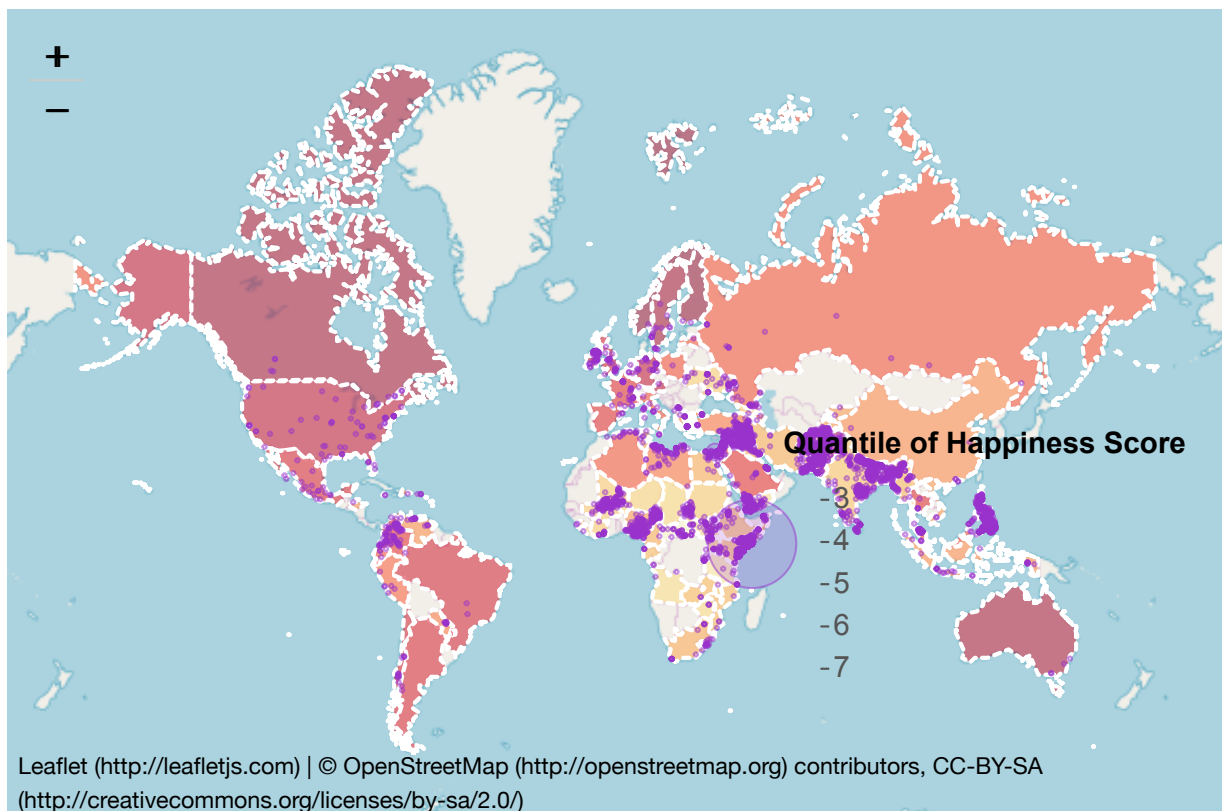
```
) %>%
    addLegend(pal = pal, values = t_2017$Happiness.Score, opacity = 0
.7, title = "Quantile of Happiness Score",bins = 5,
           position = "bottomright") %>%
  addCircles(data = terrorism_2017_plot,radius = ~(terrorism_2017_plo
t$nkill/5)^3, weight =1, fill=TRUE,color = "darkorchid",
             popup= paste("<strong>Date: </strong>", terrorism_2017_p
lot$iday,"/",terrorism_2017_plot$imonth,"/", terrorism_2017_plot$iyea
r,
                          "<br><br><strong>Place: </strong>", terrori
sm_2017_plot$city,"-",terrorism_2017_plot$country_txt,
                          "<br><strong>Killed: </strong>", terrorism_
2017_plot$nkill,
                          "<br><strong>Wounded: </strong>", terrorism
_2017_plot$nwound
             ))
```


Leaflet (http://leafletjs.com) | © OpenStreetMap (http://openstreetmap.org) contributors, CC-BY-SA
(http://creativecommons.org/licenses/by-sa/2.0/)

In the map, bigger circles mean greater number of people died in that attack.

The happiness scores of North Africa, Middle East, and South Asia region are low.

Word cloud for terrorism summary

```r
library(tidytext)
library(wordcloud)
library(RColorBrewer)
library(wordcloud2)

terrorism_2017$summary<-as.character(terrorism_2017$summary)
text<-terrorism_2017 %>%
  select(summary)
colnames(text)<-"text"
text$text<-str_sub(text$text, start = 13)

tidy_text<-text %>%
  unnest_tokens(word,text) %>%
  anti_join(stop_words)

#pal = brewer.pal(9,"Dark2")
tidy_text %<>%
  count(word) %>%
  arrange(desc(n))
wordcloud2(tidy_text[1:100,])
```

From the wordcloud, we can see that "Iraq" is a country with frequent attacks.

And police is always the target of attacks.

The islamic has a great proportion in the participants of attacks.
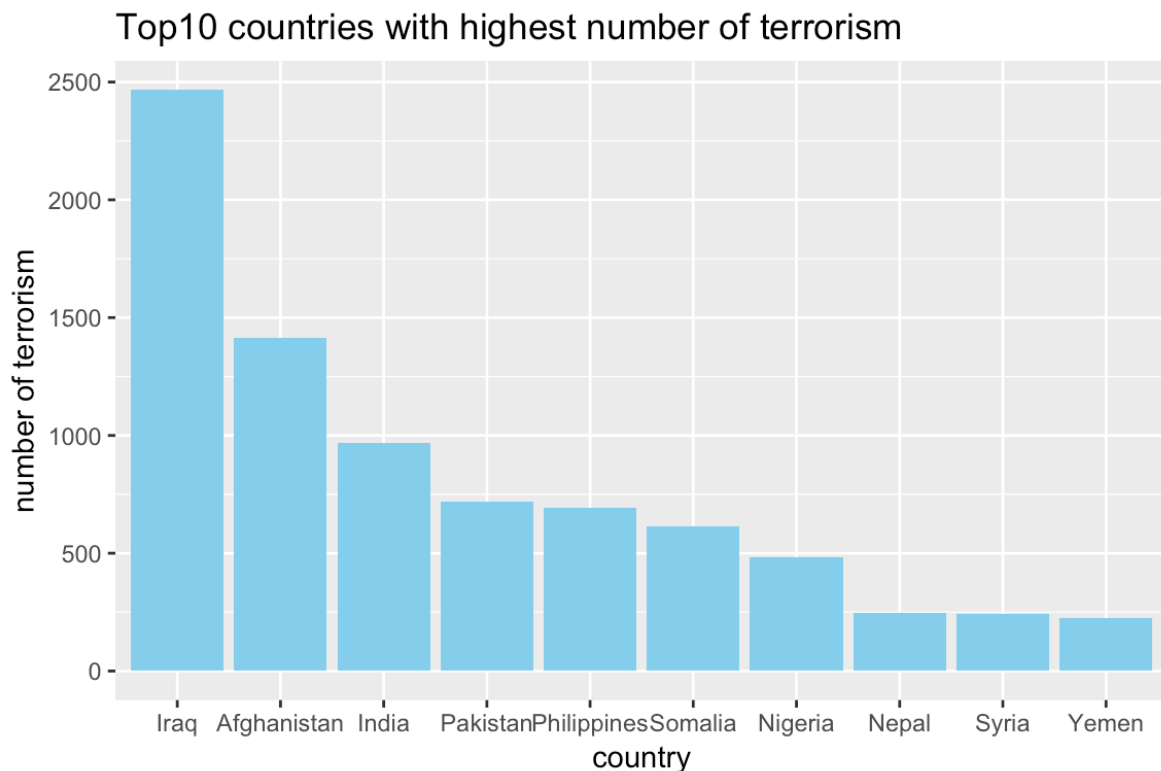
Number of attacks by country

```
num_attack<-terrorism_2017 %>%
  group_by(country_txt) %>%
  summarise(count = n())
plotdata<-num_attack %>%
  ungroup() %>%
  top_n(10,count)

ggplot(plotdata) + geom_bar(aes(x = reorder(country_txt, -count),y =
count),stat="Identity",fill = "skyblue") +
  labs(title = "Top10 countries with highest number of terrorism",
       x = "country",
       y = "number of terrorism")
```


Top10 countries with highest number of terrorism

From this plot, we can see that in 2017, Iraq is the country that have the most number of terrorism.

Target of terrorism

```
num_target<-terrorism_2017 %>%
  group_by(targtype1_txt) %>%
  summarise(count = n())
plotdata<-num_target %>%
  ungroup() %>%
  top_n(10,count)
ggplot(plotdata) + geom_bar(aes(x = reorder(targtype1_txt, count),y =
count),stat="Identity",fill = "pink") +
  labs(title = "Top10 target with highest number of terrorism",
       x = "country",
       y = "number of terrorism") +
  coord_flip()
```


Top10 target with highest number of terrorism
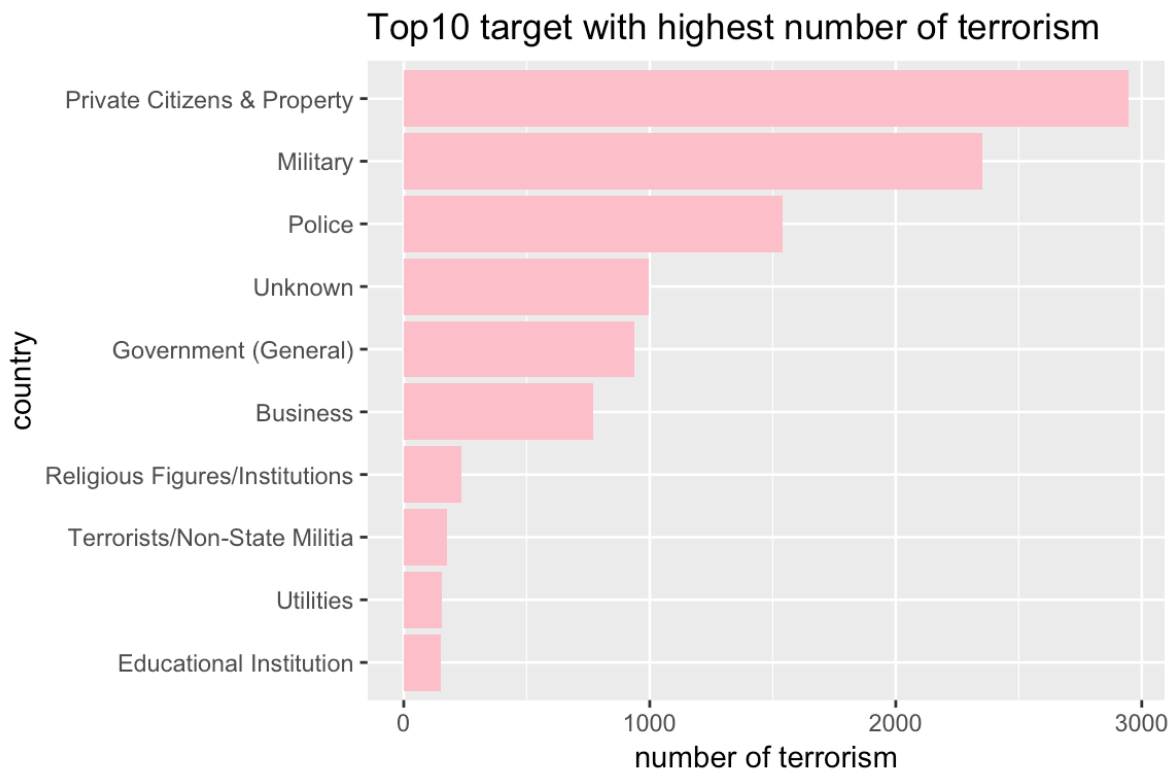
```
#theme(axis.text.x = element_text(angle = 30, vjust = 0.5))
```

From the plot we can see that "Private Citizens & Property","Military" and "Police" are the top3 frequent targets of terrorism.

Attack type & region

```
ggplot(terrorism_2017) + geom_bar(aes(x = region_txt, fill = attackty
pe1_txt)) +
  guides(fill=guide_legend(
    keywidth=0.1,
    keyheight=0.3,
    default.unit="inch"))+
  labs(title = "Attack type and region",
       x = "region",
       y = "count") +
  scale_fill_discrete(name="attack type") +
  coord_flip()
```
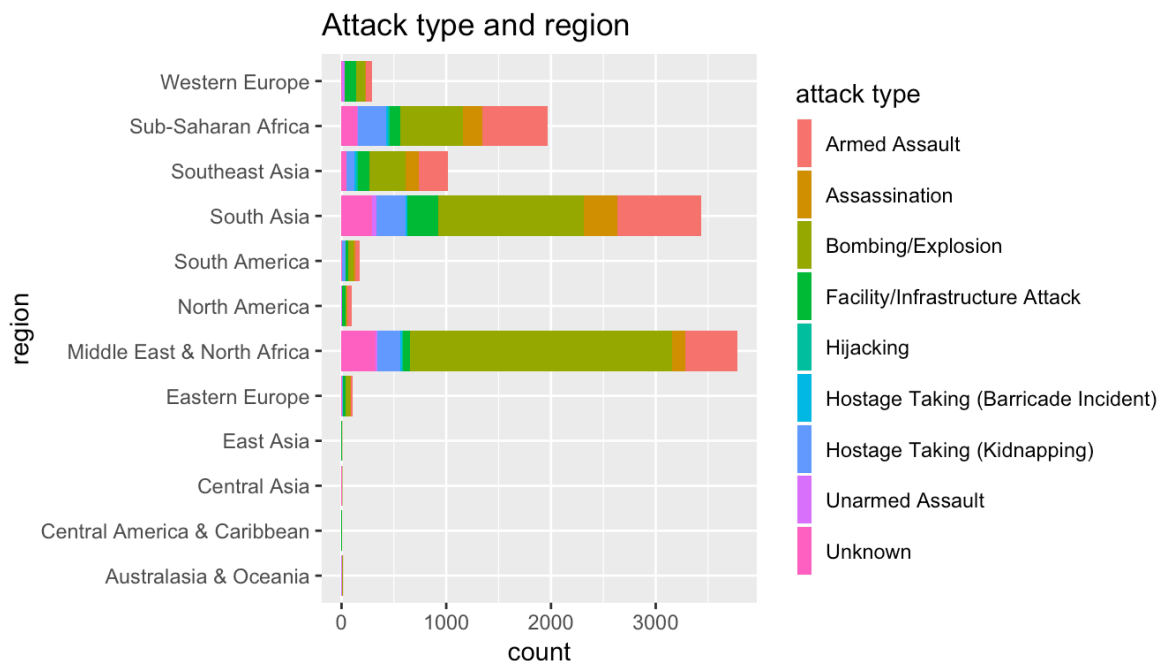

From this figure we can see that Middle Easr & North Africa amd South Asia have more than 3000 terrorism in 2017. And most of them are Bombing/Explosion and Armed Assault.

Terrorism & Happiness Score

```
library(gridExtra)
attack_happy<-left_join(happiness_2017,terrorism_2017,by = c("Country
" = "country_txt"))
attack_happy %<>%
```

```
  group_by(Country,Happiness.Score) %>%
  summarise(count = n())
attack_happy$count[attack_happy$Country %in% setdiff(happiness_2017$C
ountry,terrorism_2017$country_txt)]<-0

attack_num<-rep("0",dim(attack_happy)[1])
logical<-(attack_happy$count>=1) & (attack_happy$count<=5)
attack_num[logical]<-"(0,5]"
logical<-(attack_happy$count>5) & (attack_happy$count<=15)
attack_num[logical]<-"(5,15]"
logical<-(attack_happy$count>15) & (attack_happy$count<=55)
attack_num[logical]<-"(15,55]"
logical<-(attack_happy$count>55)
attack_num[logical]<-"(55,2466]"
#attack_num<-as.data.frame(attack_num)
attack_happy["attack_num"]<-attack_num
ave<-attack_happy%>%
  group_by(attack_num) %>%
  summarise(ave_happy = mean(Happiness.Score))

## set the levels in order we want
attack_happy <- within(attack_happy,
                  attack_num <- factor(attack_num,
                                        levels=c("0","(0,5]","(5,15]"
,"(15,55]","(55,2466]")))
p1<-ggplot(attack_happy) + geom_boxplot(aes(x = attack_num, y = Happi
ness.Score,fill = attack_num)) + geom_point(data = ave, aes(x = attac
k_num,y= ave_happy)) +
  labs(title = "Terrorism & happiness score",
       x = "Range of attacks",
       y = "Happiness Score") +
  theme(axis.text.x = element_text(angle = 30, vjust = 0.5))


p2<-ggplot(attack_happy) + geom_bar(aes(x = attack_num,fill = attack_
num)) +
   labs(title = "Distribution of attacks count",
         x = "Range of attacks",
         y = "Count") +
  theme(axis.text.x = element_text(angle = 30, vjust = 0.5))
grid.arrange(p1,p2,nrow = 1)
```

Terrorism & happiness score   Distribution of attacks count
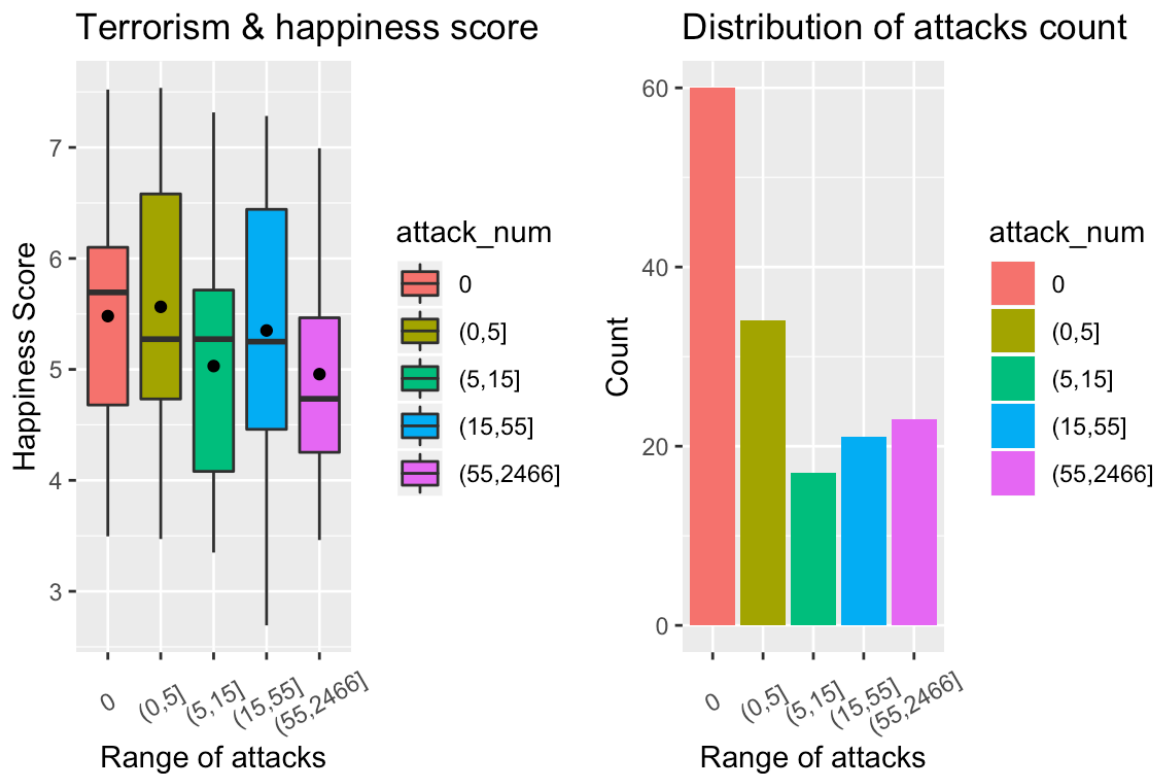
Considering the right skewed of distribution of attacks count, we divide the number of attacks in 4 groups: (0,5],(5,15],(15,55],(55,2466].

From the distribution plot we can see that every group have enough sample size.

The dot of each group represents the averange happiness score of that group.

From the boxplot we can see that with growing of number of attacks, if there is no big gap in the number of attacks, the happiness score show no decreasing trend. But if the attacks count increases significantly, the happiness score would correspondingly decrease.

Terrorism & Other Social Scores

```
radar_data<-left_join(happiness_2017,terrorism_2017,by = c("Country"
= "country_txt"))
radar_data %<>%

group_by(Country,Happiness.Score,Economy..GDP.per.Capita.,Family,Heal
th..Life.Expectancy.,Freedom,Generosity,Trust..Government.Corruption.
) %>%
```

```r
  summarise(count = n())
radar_data$count[radar_data$Country %in% setdiff(happiness_2017$Count
ry,terrorism_2017$country_txt)]<-0

attack_num<-rep("0",dim(radar_data)[1])

logical<-(radar_data$count>=1) & (radar_data$count<=5)
attack_num[logical]<-"(0,5]"
logical<-(radar_data$count>5) & (radar_data$count<=15)
attack_num[logical]<-"(5,15]"
logical<-(radar_data$count>15) & (radar_data$count<=55)
attack_num[logical]<-"(15,55]"
logical<-(radar_data$count>55)
attack_num[logical]<-"(55,2466]"
#attack_num<-as.data.frame(attack_num)
radar_data["attack_num"]<-attack_num
radar_data %<>%
  group_by(attack_num) %>%
  summarise_all(mean)%>%

select(attack_num,Economy..GDP.per.Capita.,Family,Health..Life.Expect
ancy.,Freedom,Generosity,Trust..Government.Corruption.)
colnames(radar_data)<-c("attack_num","Economy","Family","Health","Fre
edom","Generosity","Trust to Government")

plotdata<-gather(radar_data,key=dimensions,value = score,-attack_num)
plotdata<-plotdata %>% arrange(dimensions)
## set the levels in order we want
plotdata <- within(plotdata,
                 attack_num <- factor(attack_num,
                                    levels=c("0","(0,5]","(5,15]"
,"(15,55]","(55,2466]")))


ggplot(plotdata)+
  geom_polygon(aes(x = dimensions, y = score, color = attack_num, gro
up = attack_num),alpha=0.5,fill = NA) +
  coord_polar(theta = "x", start=0, direction = 1) +
  theme_minimal()+
  scale_x_discrete(labels = function(x) str_wrap(x, width = 10))+
  theme(axis.title.y = element_text(margin = unit(c(0, 1, 0, 0), "mm"
), angle = 0),
```
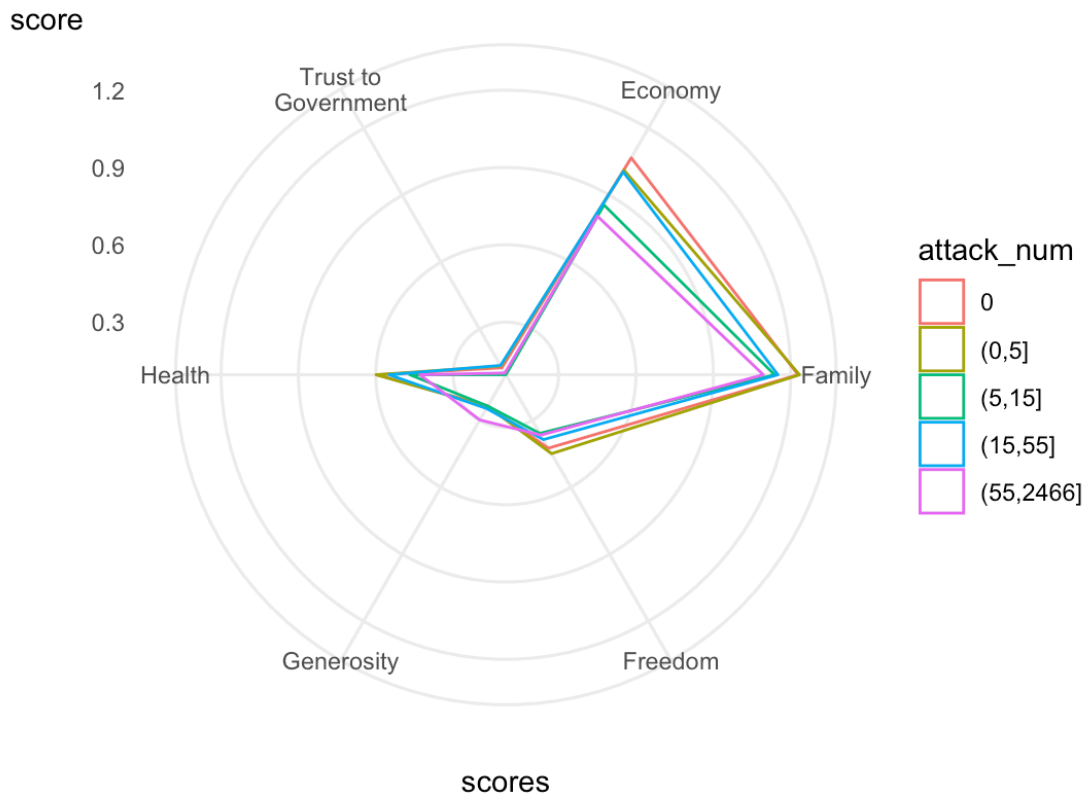
```
        plot.margin = margin(0, 0, 0, 1, "cm"))+
    labs(x="scores")
```
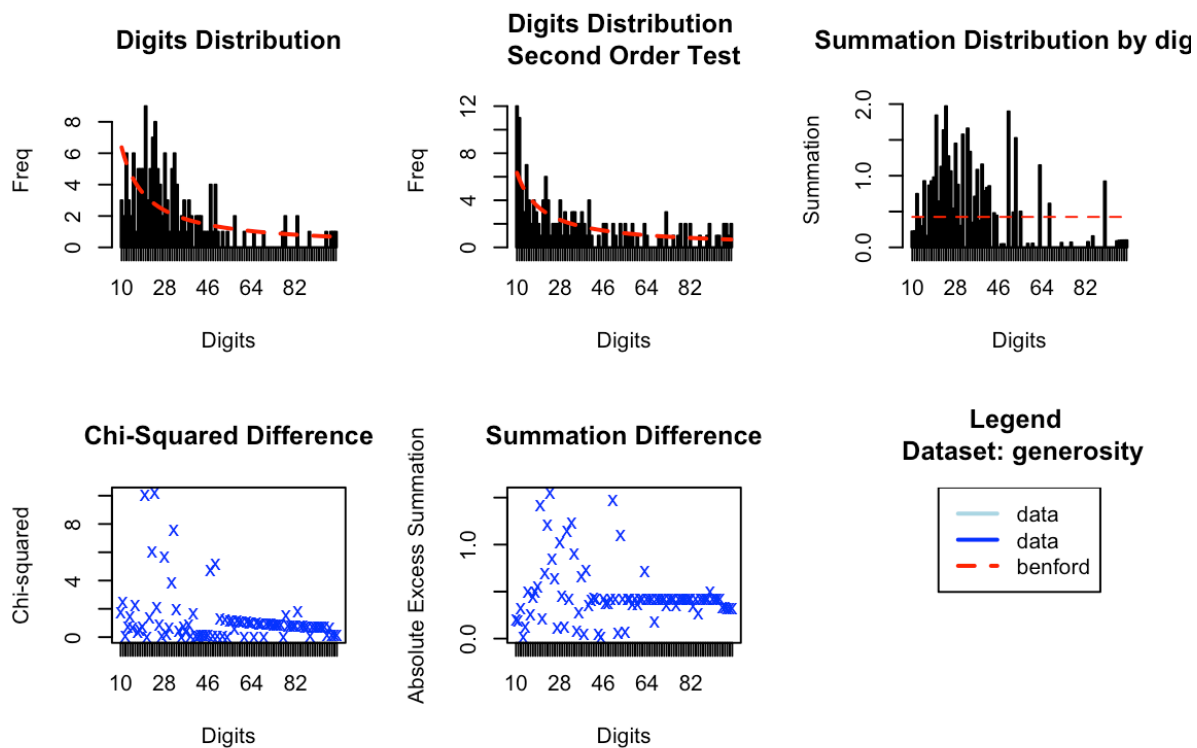
score

scores

In the radar plot, 6 dimensions are 6 aspects related to the happiness of people.The more the scores, the higher extend to which people satisfied with that aspect.

We find that countries with less terrorism have higher scores in "Freedom","Health Life Expectacy","Economy GDP","Family", and "Trust to Government, indicating that countries with less terrorism have bettehr situation in these 5 aspects.

Only the aspect "Generosity" have the highest score in countries with attacks more than 55, which is weird.

Benford Law

```
library(benford.analysis)
generosity<-happiness_2017$Generosity
bfd.generosity <- benford(generosity)
plot(bfd.generosity)
```

Digits Distribution

Digits Distribution Second Order Test

Summation Distribution by dig

Chi-Squared Difference

Summation Difference

Legend Dataset: generosity

From the plots of benford law, we can see some difference between our data and benford law from the first digits distribution plot. However, the Second order of digits distribution is almost consist with Benford Law. Then we explore the distribution of generosity.

```
ggplot(happiness_2017) + geom_density(aes(x = Generosity),color = "or
ange") +
  labs(title = "Distribution of Generosity")
```

Distribution of Generosity
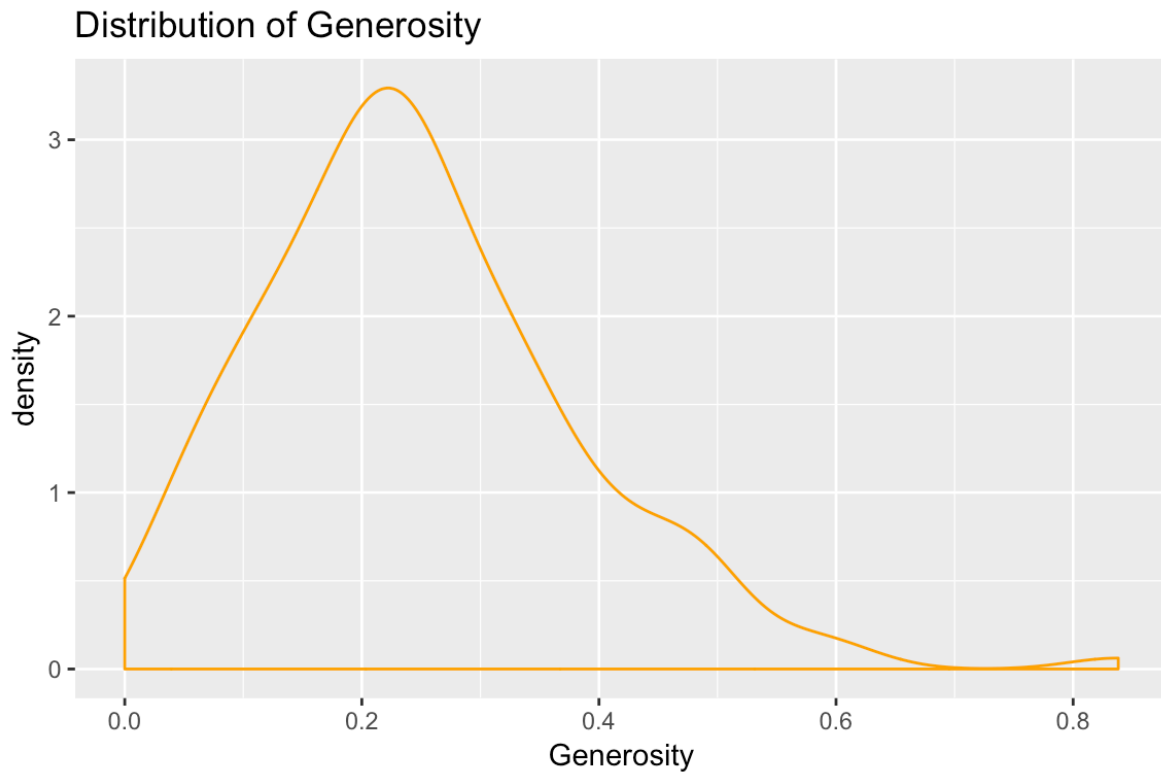
The distribution of generosity score is almost symmetric, consist with our common sense. We believe that the data is reliable.

Conclusion

1. Iraq is a country with most frequent attacks.

2. Police is always the target of attacks.

3. The islamic has a great proportion in the participants of attacks.

4. Middle East, North Africa, and South Asia have concentrated attacks, and at the same time, happiness scores in these regions are low.

5. Happiness score is negative correlated with number of terrorism when comparing countries with enough difference in number of terrorism.

6. Countries with less terrorism are more satisfied in Freedom, Health Life Expectacy, Economy GDP, Family, and Trust to Government.

7. Data of generosity in Happiness is reliable.