

# Визуализация на Световното Щастие

В този проект ще анализирам световното щастие спрямо датата извадена от:

<https://www.kaggle.com/unsdsn/world-happiness/home>

Ще се фокусирам върху датата от 2015 година.

## Нужни библиотеки и зареждане на информацията

Като за начало трябва следните библиотеки да бъдат заредени както и прочитането на информацията от csv - то със следния код.

```
library(readr)
library(ggplot2)
library(corrplot)
library(plotly)
library(dplyr)
library(tidyr)

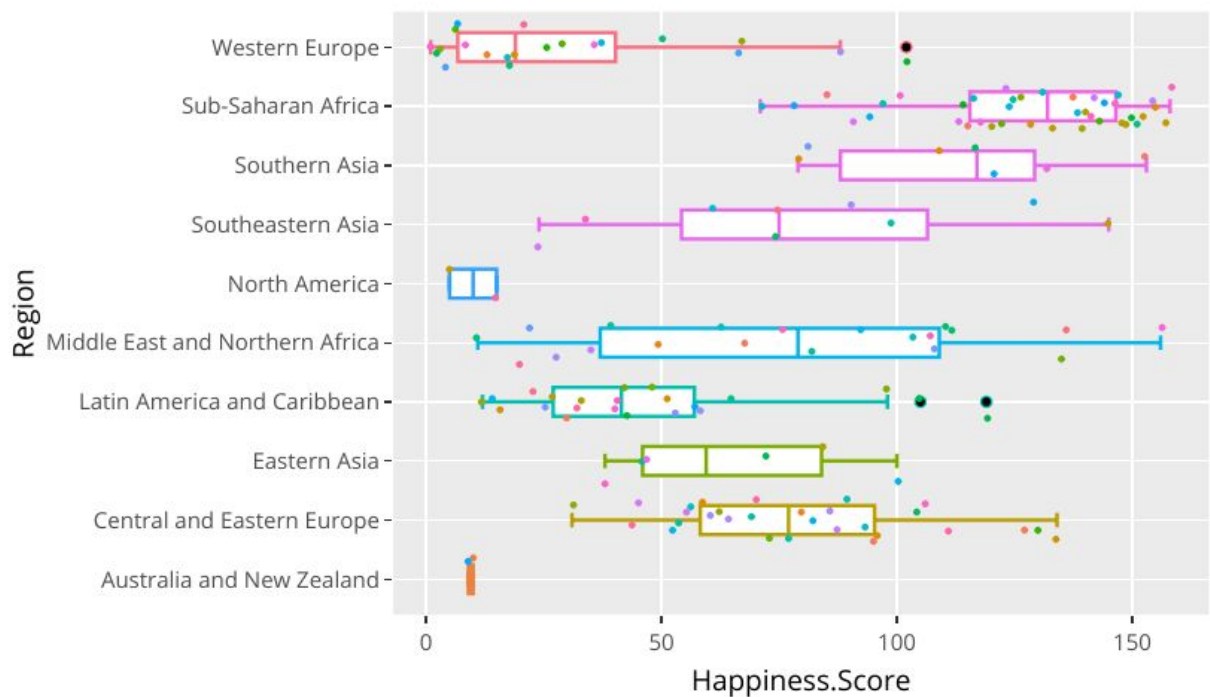
happy_2015 <- read.csv('/home/lt34507/Desktop/statistics_project/2015.csv')
names(happy_2015)
```

## Първоначален анализ на датата

След това ще анализирам средния 'Happiness Score' спрямо всички региони и да се фокусирам върху техните характеристики, като направя box plot и хистограма спрямо средния 'Happiness Score'.

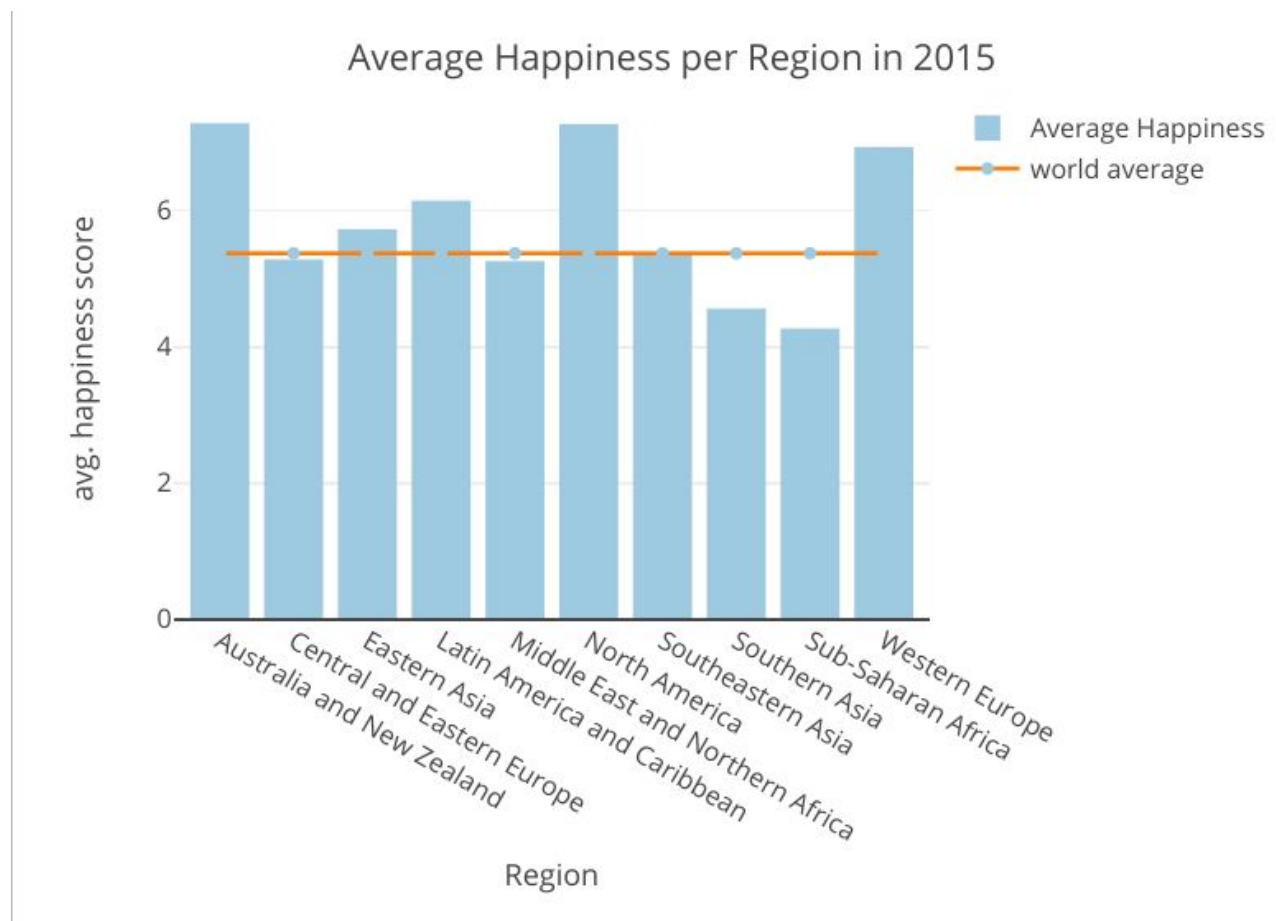
```
box <- ggplot(happy_2015, aes(x = Region, y = Happiness.Score, color = Region))+
  geom_boxplot() +
  geom_jitter(aes(color=Country), size = 0.5) +
  ggtitle("Happiness Score for Regions and Countries") +
  coord_flip() +
  theme(legend.position="none")
ggplotly(box)
```

## Happiness Score for Regions and Countries



```
avg_happiness_region <- happy_2015 %>%
  group_by(Region) %>%
  summarise(avg_happiness = mean(Happiness.Score, round(1)))

p_avg_happiness_region <- plot_ly(avg_happiness_region, x = ~Region,
  y = ~avg_happiness,
  type = 'bar',
  name = 'Average Happiness',
  marker = list(color = 'rgb(158,202,225)')) %>%
  add_lines(y = ~mean(happy_2015$Happiness.Score), name = 'world average') %>%
  layout(title="Average Happiness per Region in 2015", yaxis = list(title = "avg.
happiness score"))
ggplotly(p_avg_happiness_region)
```



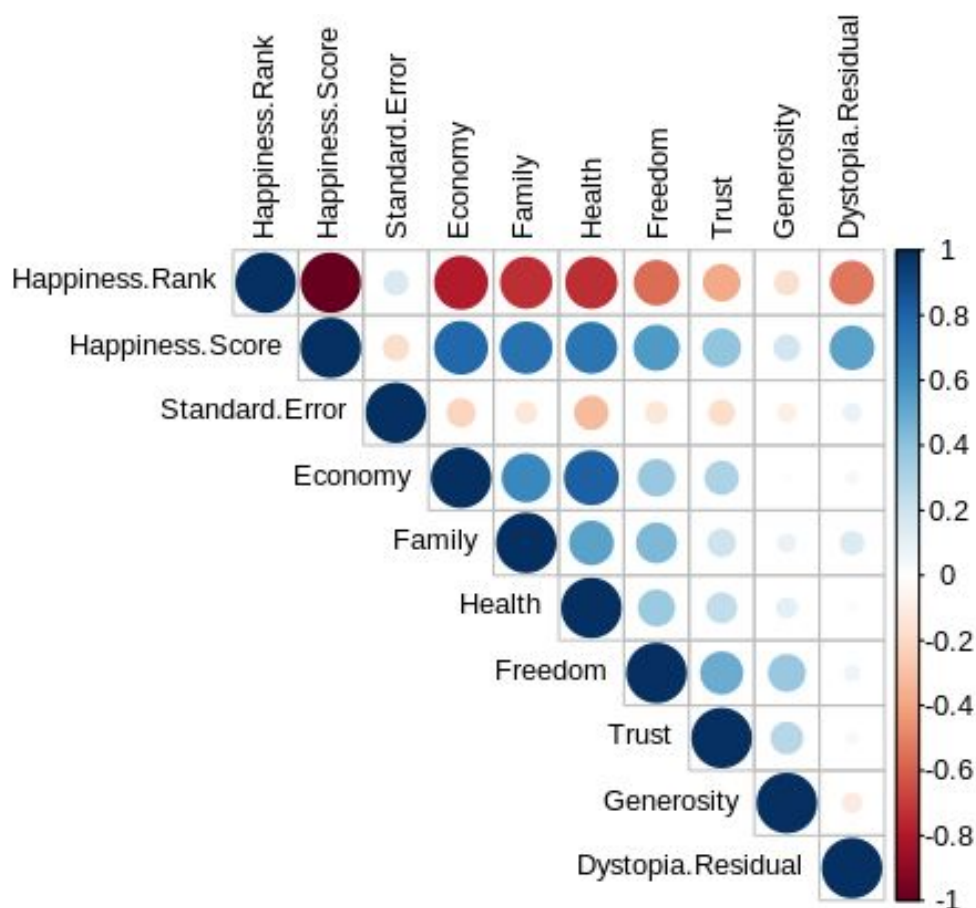
Както се вижда от горните визуализации можем да заключим, че топ 3 най-щастливи региона са Австралия и Нова Зеландия, Северна Америка и Западна Европа.

Най-нещастните региони са Африка отвъд Сахара и Южна Азия.

Можем да забележим, че и трите най-щастливи региона имат силно развити икономики, което е предпоставка затова, че колкото по-развита е една страна, толкова за по-щастливи хората в нея. Но тъй като не сме сигурни дали това е така ще направя една корелограма, за да анализирам връзките между различните атрибути, като, разбира се, оставя само числовите параметри.

## Корелограма за различните атрибути

```
num_hap <- happy_2015[, c("Happiness.Rank", "Happiness.Score", "Standard.Error",
"Economy", "Family", "Health", "Freedom", "Trust", "Generosity",
"Dystopia.Residual")]
m <- cor(num_hap)
corrplot(m, method="circle", type='upper', tl.cex=0.8, tl.col = 'black')
```



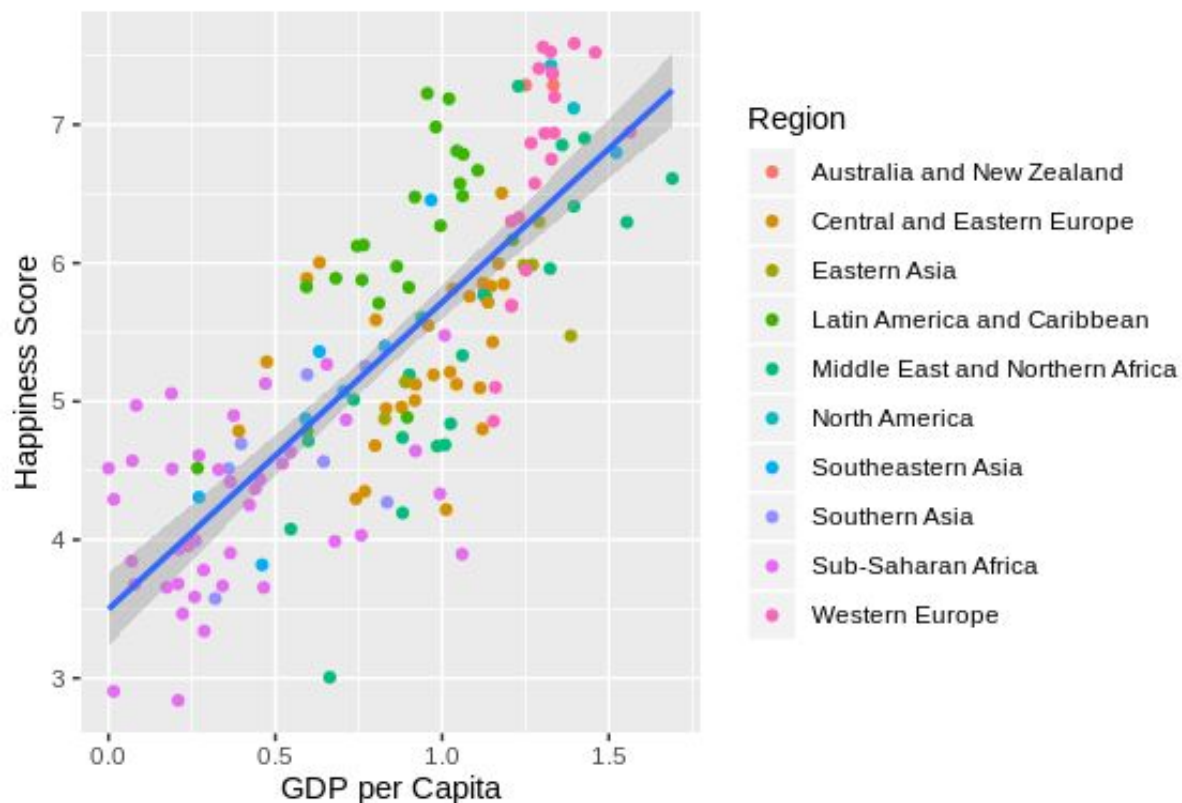
Сега можем да се фокусираме върху атрибутите, които имат силна връзка с 'Happiness Score'. Те са именно 'Economy', 'Family' и 'Health'. Те представляват брутния вътрешен продукт на даден регион, силни родствени връзки и семейство и средната продължителност на живот. Както виждаме веднага се появява атрибута с Икономиката, което е още една предпоставка за по-горното ни предположение за страните с развита икономика.

## Щастие спрямо развита Икономика

Сега ще представя корелацията между колко са щастливи хората в даден регион и колко е развита дадената икономика като използвам 'GDP per capita' вместо 'Economy', защото е така е по-ясно.

```
ggplot(happy_2015, aes(x=happy_2015$Economy, y=happy_2015$Happiness.Score))+
  geom_point(aes(color = happy_2015$Region)) +
  geom_smooth(method="lm") +
  xlab("GDP per Capita") +
  ylab("Happiness Score") +
  labs(colour="Region") +
  ggtitle("All Regions: Happiness Score & GDP per Capita (2015)")
```

## All Regions: Happiness Score & GDP per Capita (2015)

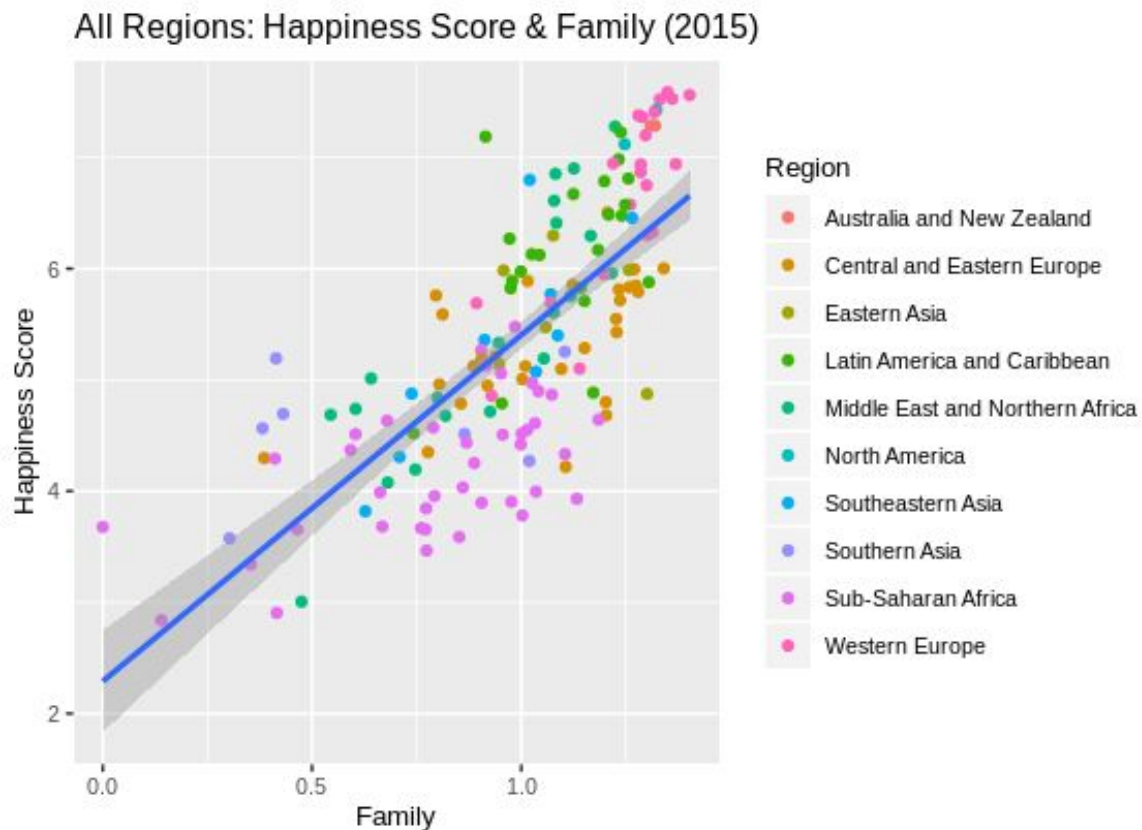


И отново очакванията ни са оправдани. Очевидно се вижда, че колкото по-развита е дадена страна в Икономичен план, то толкова са по-щастливи хората. Това е така, защото колкото по-богата е дадена държава, толкова е по-висок стандартът на живот в нея.

## Щастие спрямо Семейство

Следващата корелация е между щастиято и семейството.

```
ggplot(happy_2015, aes(x=happy_2015$Family, y=happy_2015$Happiness.Score))+  
geom_point(aes(color=happy_2015$Region)) +  
geom_smooth(method="lm") +  
xlab("Family") +  
ylab("Happiness Score") +  
labs(colour="Region")+  
ggtitle("All Regions: Happiness Score & Family (2015)") +  
theme(text = element_text(size=10))
```



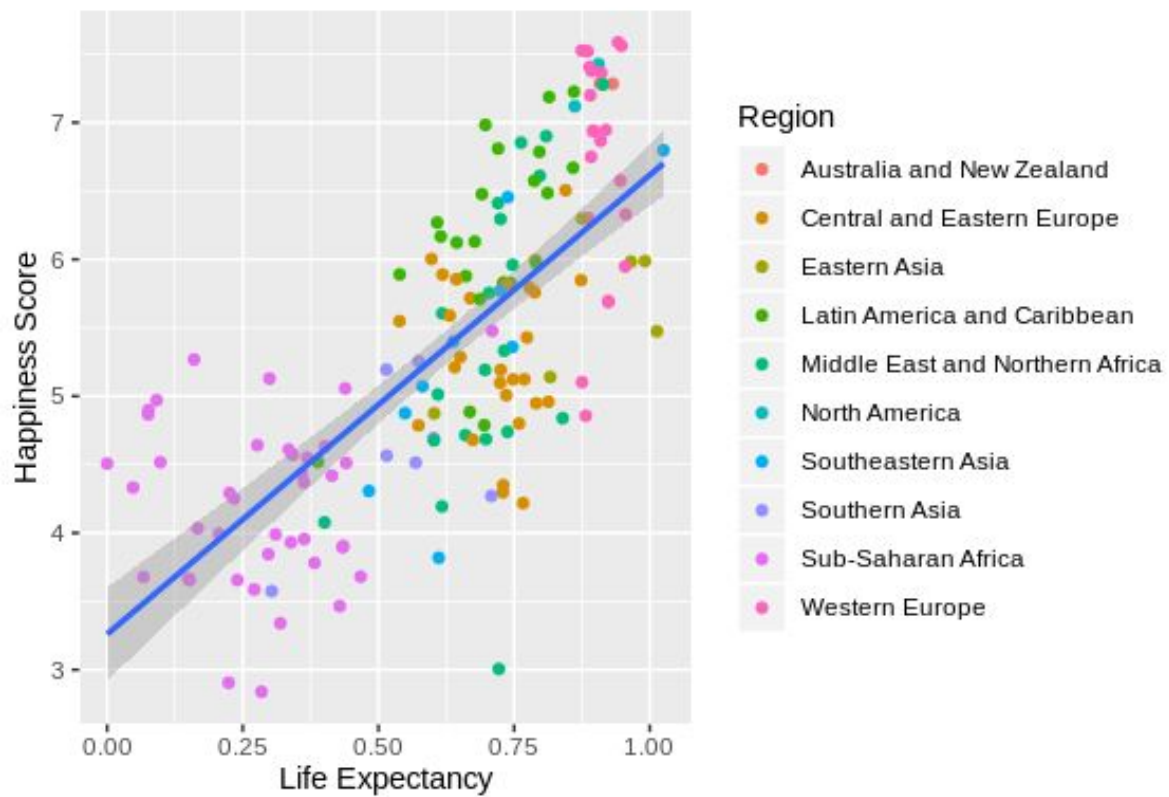
От графиката се вижда, че колкото е по-развито дадено семейство, толкова по-щастливи са хората в дадения регион.

## Щастие спрямо очакване за живот

Остана и последната корелация, която трябва да направим а именно щастието спрямо очакването за живот.

```
ggplot(happy_2015, aes(x=happy_2015$Health, y=happy_2015$Happiness.Score))+
  geom_point(aes(color = happy_2015$Region)) +
  geom_smooth(method="lm") +
  xlab("Life Expectancy") +
  ylab("Happiness Score") +
  labs(colour="Region")+
  ggtitle("All Regions: Happiness Score & Life Expectancy (2015)")
```

## All Regions: Happiness Score & Life Expectancy (2015)



И отново регионите с по-висока продължителност на живот излизат отгоре като по-щастливи от останалите, което е логично, защото по този начин хората се чувстват много по-сигурни и по-добре, което води до това те, да са по-щастливи.