Term Project: World Happiness Report

Deependra

5/23/2021

Introduction.

The World Happiness Report is being published every year since 2012. The report has continuously gained global recognitions as governments, organization, and civil society increasingly use the happiness indicators in policy-making decisions. The report reviews the state of happiness and explains the personal and national variations in happiness.

Research Question:

- What are the different factors that leads to the happiness of a country?
- How does economy of a country determine its happiness?
- Are the countries in certain region are correlated to average scores of the countries in the region?
- What are the similarities and differences between the countries with highest and lowest scores?
- Has the ranking changed from 2018 to 2021 reports? If yes, did any country experience a significant change?

Data source:

I will be analyzing the data from 2015 to 2018.

I am going to use the data from Kaggle:

- Aché, M. (2020, June 30). World Happiness Report up to 2020. Retrieved from https://www.kaggle.com/mathurinache/world-happiness-report

- Singh, A. (2021, March 22). World Happiness Report 2021. Retrieved from https://www.kaggle.com/ajaypalsinghlo/world-happiness-report-2021

The United Nations Sustainable Development Solutions Network (UNSDSN) creates the Global Happiness Report using the data from the Gallup World Poll.

SDSN (Sustainable development solutions network) is a UN group run by scholars from around the globe.

Gallup is an American analytics/data firm hired to collect information through happiness surveys around the world. The happiness score was received from Gallup's data, collected through randomized phone surveys, face-to-face surveys. While the possibility of bias is always possible,

the SDSN and all the funding companies would benefit more form the accurate data, so it seems that the data is highly likely to be accurate.

The people surveyed were asked to rate the variable between 0-10 where 0 represented the worst possible life and 10 the best possible. The six conditions the people were asked to rate were as follows:

- Economic condition.
- Physical and mental health.
- Freedom to make key life decisions.
- Social networks
- Public values or generosity
- Social trust.

Below is my work on the above mentioned data.

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

installing necessary packages

```
library(data.table)
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.0.5
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.
3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.0     v dplyr   1.0.5
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1
```

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

```
## Warning: package 'tidyr' was built under R version 4.0.5
```

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

```
## Warning: package 'forcats' was built under R version 4.0.5
```

```
## -- Conflicts ---------------------------------------- tidyverse_conflict
s() --
## x dplyr::between()   masks data.table::between()
## x dplyr::filter()    masks stats::filter()
## x dplyr::first()     masks data.table::first()
## x dplyr::lag()       masks stats::lag()
## x dplyr::last()      masks data.table::last()
## x purrr::transpose() masks data.table::transpose()

library(corrplot)

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

## corrplot 0.88 loaded

library(plotly)

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

##
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':
##
##     last_plot

## The following object is masked from 'package:stats':
##
##     filter

## The following object is masked from 'package:graphics':
##
##     layout

library(wildcard)

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

library(shiny)

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

library(PerformanceAnalytics)

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

## Loading required package: xts

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

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.0.5
```

```
##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':
##
##     first, last

## The following objects are masked from 'package:data.table':
##
##     first, last

##
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':
##
##     legend

library(DT)

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

##
## Attaching package: 'DT'

## The following objects are masked from 'package:shiny':
##
##     dataTableOutput, renderDataTable

library(dplyr)
```

importing and cleaning the data.

```
setwd("C:/Users/Deependra/Documents/Term Project")
data_2015 <- read.csv("2015.csv")
data_2016 <- read.csv ("2016.csv")
data_2017 <- read.csv("2017.csv")
data_2018 <- read.csv("2018.csv")

data_2015$year <- 2015
data_2016$year<- 2016
data_2017$year <- 2017
data_2018$year<- 2018
```

##changing data type to bind the rows.

```
data_2018$Perceptions.of.corruption <- as.double(data_2018$Perceptions.of.cor
ruption)

## Warning: NAs introduced by coercion
```

selecting only the columns available in all the data set.

```
library(dplyr)
merged_data <- bind_rows(data_2015, data_2016, data_2017, data_2018)
names(merged_data)

##  [1] "Country"                       "Region"
##  [3] "Happiness.Rank"                "Happiness.Score"
##  [5] "Standard.Error"                "Economy..GDP.per.Capita."
##  [7] "Family"                        "Health..Life.Expectancy."
##  [9] "Freedom"                       "Trust..Government.Corruption."
## [11] "Generosity"                    "Dystopia.Residual"
## [13] "year"                          "Lower.Confidence.Interval"
## [15] "Upper.Confidence.Interval"     "Whisker.high"
## [17] "Whisker.low"                   "Overall.rank"
## [19] "Country.or.region"             "Score"
## [21] "GDP.per.capita"                "Social.support"
## [23] "Healthy.life.expectancy"       "Freedom.to.make.life.choices"
## [25] "Perceptions.of.corruption"

merged_data <- merged_data %>% select(Country:year, -Region, -Standard.Error)
```

##renaming the columns

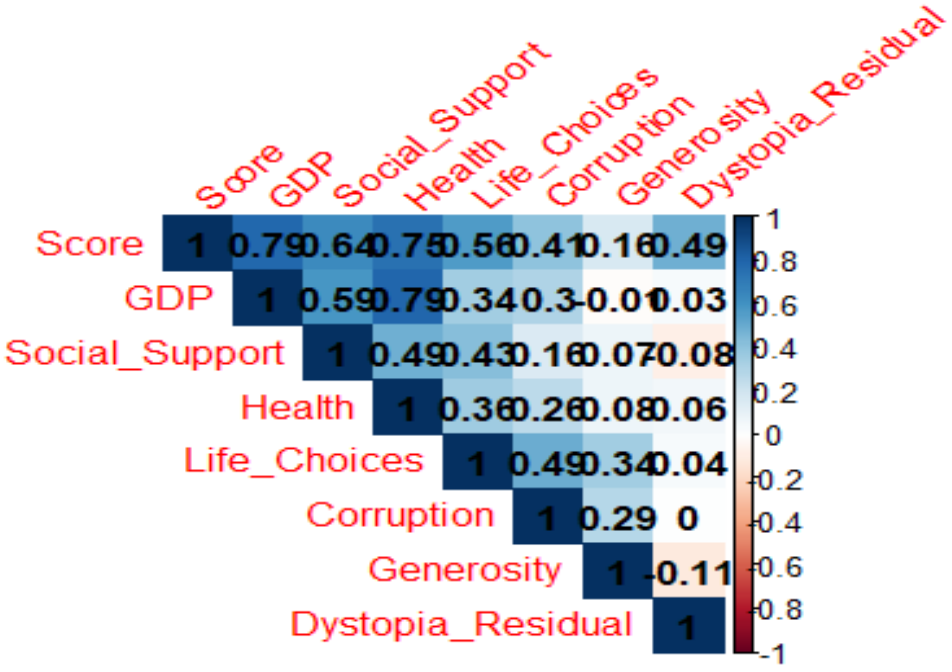
```
names(merged_data) <- c("Country", "Happiness_Rank", "Score", "GDP", "Social_
Support", "Health", "Life_Choices","Corruption", "Generosity","Dystopia_Resid
ual" , "Year")
names(merged_data)

##  [1] "Country"           "Happiness_Rank"    "Score"
##  [4] "GDP"               "Social_Support"    "Health"
##  [7] "Life_Choices"      "Corruption"        "Generosity"
## [10] "Dystopia_Residual" "Year"

head(merged_data)

##       Country Happiness_Rank Score     GDP Social_Support  Health Life_Cho
ices
## 1 Switzerland             1 7.587 1.39651        1.34951 0.94143      0.6
6557
## 2     Iceland             2 7.561 1.30232        1.40223 0.94784      0.6
2877
## 3     Denmark             3 7.527 1.32548        1.36058 0.87464      0.6
4938
## 4      Norway             4 7.522 1.45900        1.33095 0.88521      0.6
6973
## 5      Canada             5 7.427 1.32629        1.32261 0.90563      0.6
```

```
3297
## 6     Finland              6 7.406 1.29025        1.31826 0.88911      0.6
4169
##   Corruption Generosity Dystopia_Residual Year
## 1    0.41978    0.29678           2.51738 2015
## 2    0.14145    0.43630           2.70201 2015
## 3    0.48357    0.34139           2.49204 2015
## 4    0.36503    0.34699           2.46531 2015
## 5    0.32957    0.45811           2.45176 2015
## 6    0.41372    0.23351           2.61955 2015

str(merged_data)

## 'data.frame':    626 obs. of  11 variables:
##  $ Country          : chr  "Switzerland" "Iceland" "Denmark" "Norway" ...
##  $ Happiness_Rank   : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Score            : num  7.59 7.56 7.53 7.52 7.43 ...
##  $ GDP              : num  1.4 1.3 1.33 1.46 1.33 ...
##  $ Social_Support   : num  1.35 1.4 1.36 1.33 1.32 ...
##  $ Health           : num  0.941 0.948 0.875 0.885 0.906 ...
##  $ Life_Choices     : num  0.666 0.629 0.649 0.67 0.633 ...
##  $ Corruption       : num  0.42 0.141 0.484 0.365 0.33 ...
##  $ Generosity       : num  0.297 0.436 0.341 0.347 0.458 ...
##  $ Dystopia_Residual: num  2.52 2.7 2.49 2.47 2.45 ...
##  $ Year             : num  2015 2015 2015 2015 2015 ...

summary(merged_data)

##    Country          Happiness_Rank       Score             GDP
##  Length:626         Min.   :  1.00   Min.   :2.693   Min.   :0.0000
##  Class :character   1st Qu.: 40.00   1st Qu.:4.509   1st Qu.:0.6053
##  Mode  :character   Median : 79.00   Median :5.282   Median :0.9954
##                     Mean   : 78.83   Mean   :5.371   Mean   :0.9278
##                     3rd Qu.:118.00   3rd Qu.:6.234   3rd Qu.:1.2524
##                     Max.   :158.00   Max.   :7.587   Max.   :1.8708
##                     NA's   :156      NA's   :156     NA's   :156
##  Social_Support       Health        Life_Choices      Corruption
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.7930   1st Qu.:0.4023   1st Qu.:0.2976   1st Qu.:0.05978
##  Median :1.0257   Median :0.6300   Median :0.4183   Median :0.09950
##  Mean   :0.9903   Mean   :0.5800   Mean   :0.4028   Mean   :0.13479
##  3rd Qu.:1.2288   3rd Qu.:0.7683   3rd Qu.:0.5169   3rd Qu.:0.17316
##  Max.   :1.6106   Max.   :1.0252   Max.   :0.6697   Max.   :0.55191
##  NA's   :156      NA's   :156      NA's   :156      NA's   :156
##    Generosity     Dystopia_Residual      Year
##  Min.   :0.0000   Min.   :0.3286    Min.   :2015
##  1st Qu.:0.1373   1st Qu.:1.7380    1st Qu.:2015
##  Median :0.2086   Median :2.0946    Median :2016
##  Mean   :0.2270   Mean   :2.0927    Mean   :2016
##  3rd Qu.:0.2909   3rd Qu.:2.4556    3rd Qu.:2017
```

```
##  Max.   :0.8381   Max.   :3.8377    Max.   :2018
##                    NA's   :156
```

##Correlation matrix

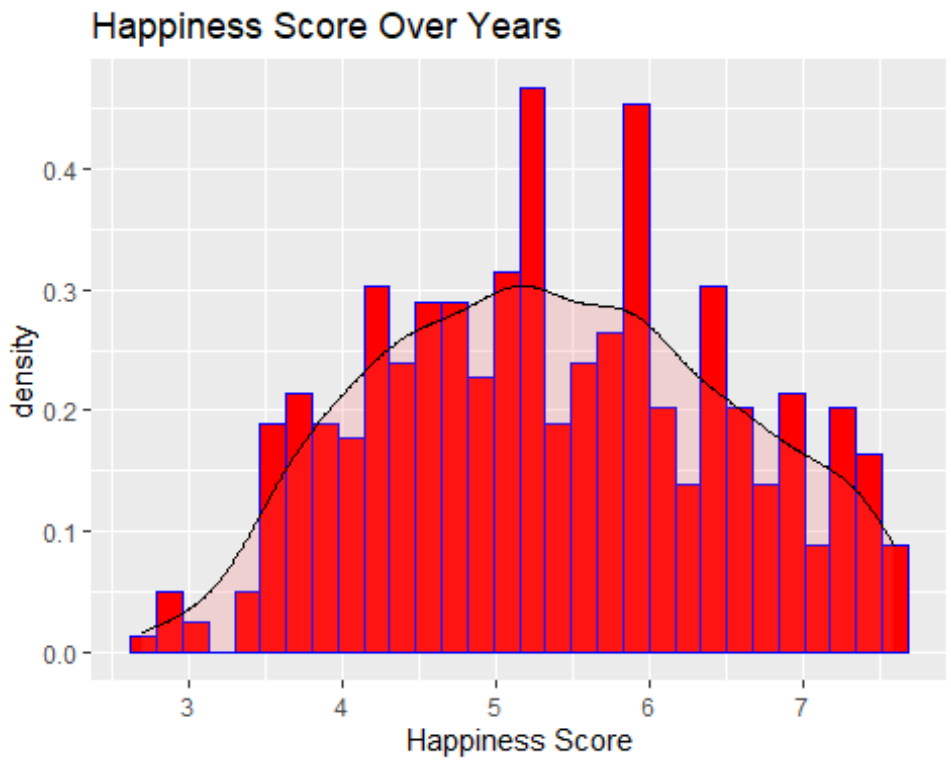
```
corrplot(cor(merged_data
             %>%
               select(Score : Dystopia_Residual),
             use = "complete.obs"),
         method="color",
         sig.level = 0.01, insig = "blank",
         addCoef.col = "black",
         tl.srt=45,
         type="upper"
         )
```



##World Happiness Score over the years(histogram)

```
ggplot(merged_data, aes(x=Score)) +
  geom_histogram(aes(y =..density..),colour='blue', fill='red')+
  geom_density(alpha=0.2, fill='#FF6666')+ggtitle('Happiness Score Over Years
')+
  xlab("Happiness Score")

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 156 rows containing non-finite values (stat_bin).
```

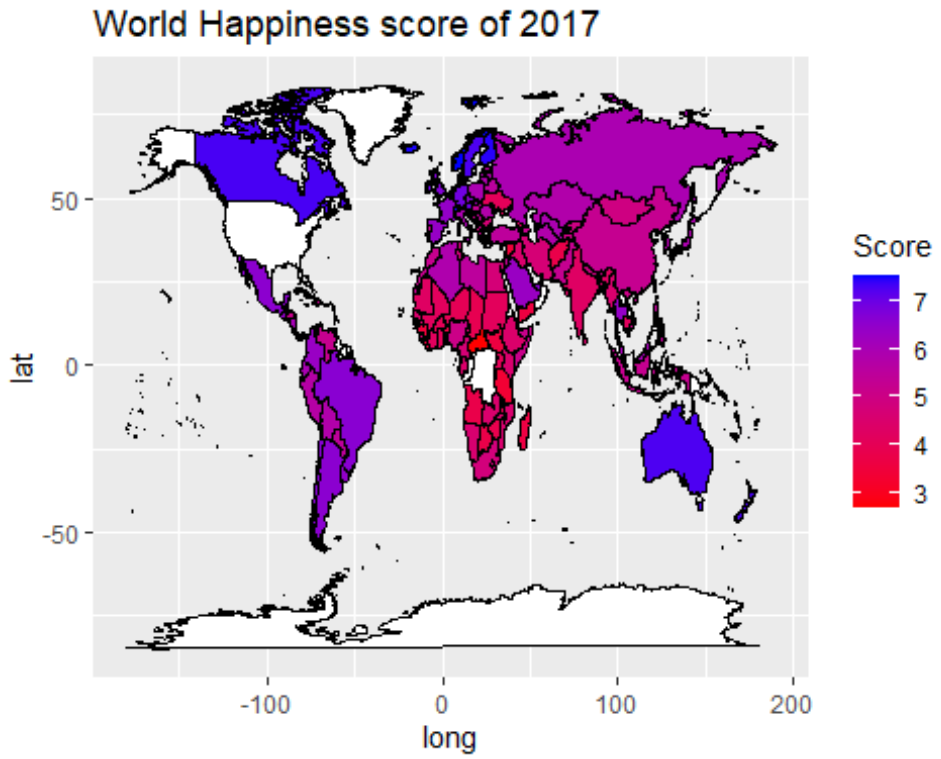
```
## Warning: Removed 156 rows containing non-finite values (stat_density).
```



##Happiness Score

world map

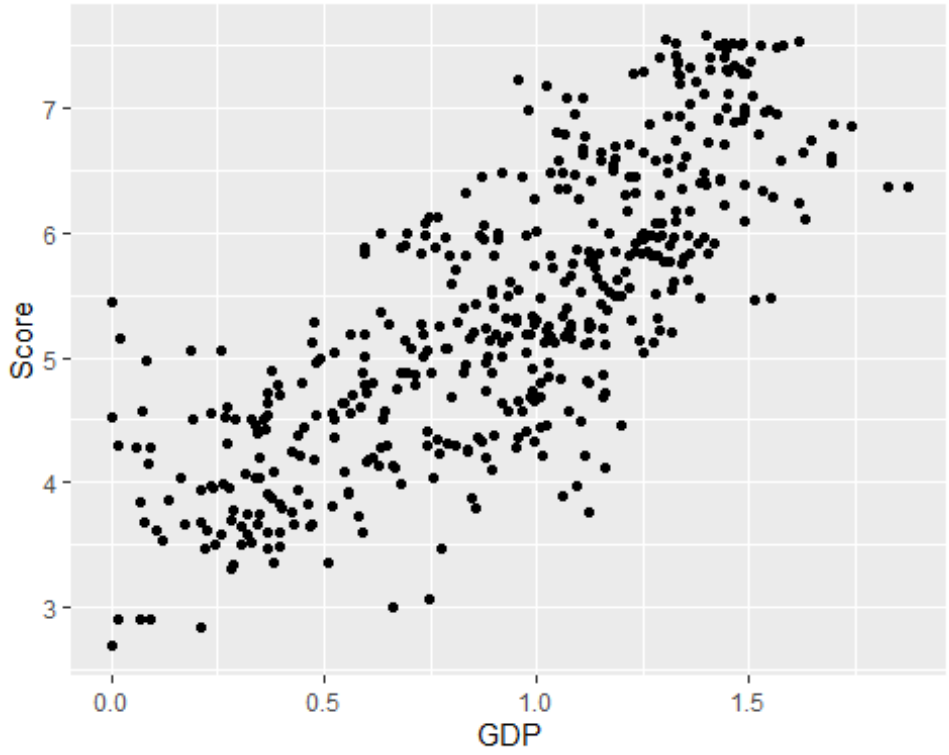
```
World <- map_data('world')
World <- fortify(World)
Happiness_Score <- merged_data %>% select(Country, Score,Year ) %>% filter (Y
ear == 2017)
Happiness_Score <- wildcard(df = Happiness_Score, wildcard = 'United states',
values = 'USA',
                            expand = TRUE, rules = NULL)
Happiness_Score <- wildcard(df = Happiness_Score, wildcard = 'United Kingdom'
, values = 'UK',
                            expand = TRUE, rules = NULL)
Happiness_Score <- wildcard(df = Happiness_Score, wildcard = 'Democratic Repu
blic of the Congo', values = 'Congo (Kinshasa)',
                            expand = TRUE, rules = NULL)
ggplot() +
  geom_map(data= World, map = World, aes(x=long, y=lat, group=group, map_id=r
egion),
           fill = 'white',col = 'black' )+
  geom_map(data = Happiness_Score, map=World,
           aes(fill= Score, map_id = Country),
           col = 'black')+
  scale_fill_continuous(low = 'red', high = 'blue', guide='colorbar')+
  labs(title = 'World Happiness score of 2017')

## Warning: Ignoring unknown aesthetics: x, y
```



##looking at the scatter plot of the data with GDP and Score.

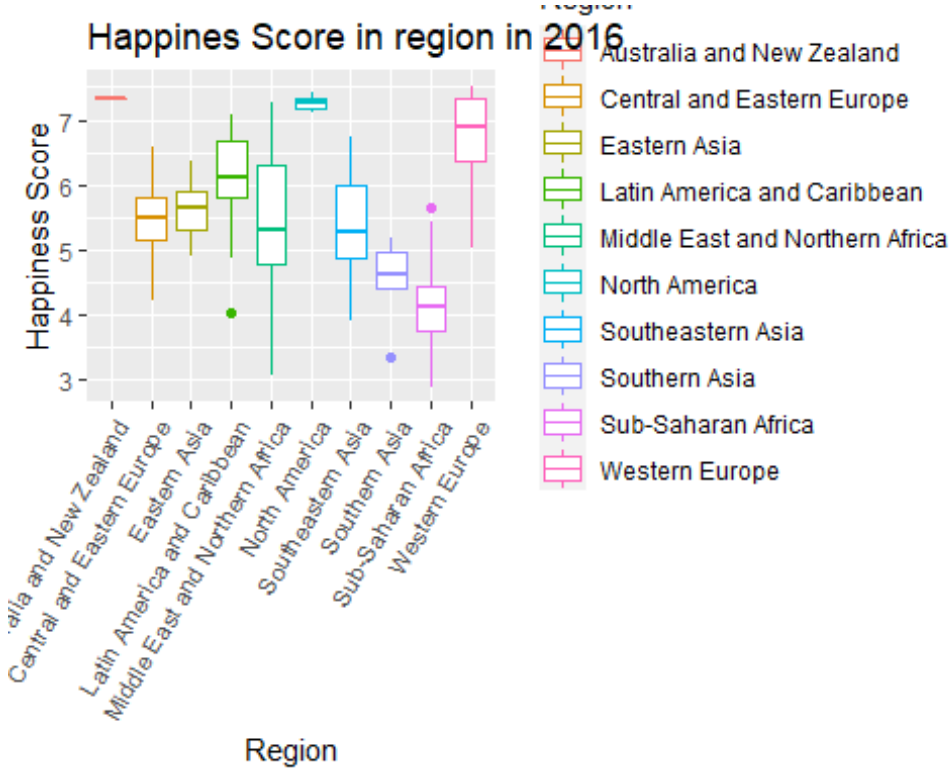
```
ggplot(merged_data, aes(x = GDP, y = Score)) + geom_point()
```

```
## Warning: Removed 156 rows containing missing values (geom_point).
```



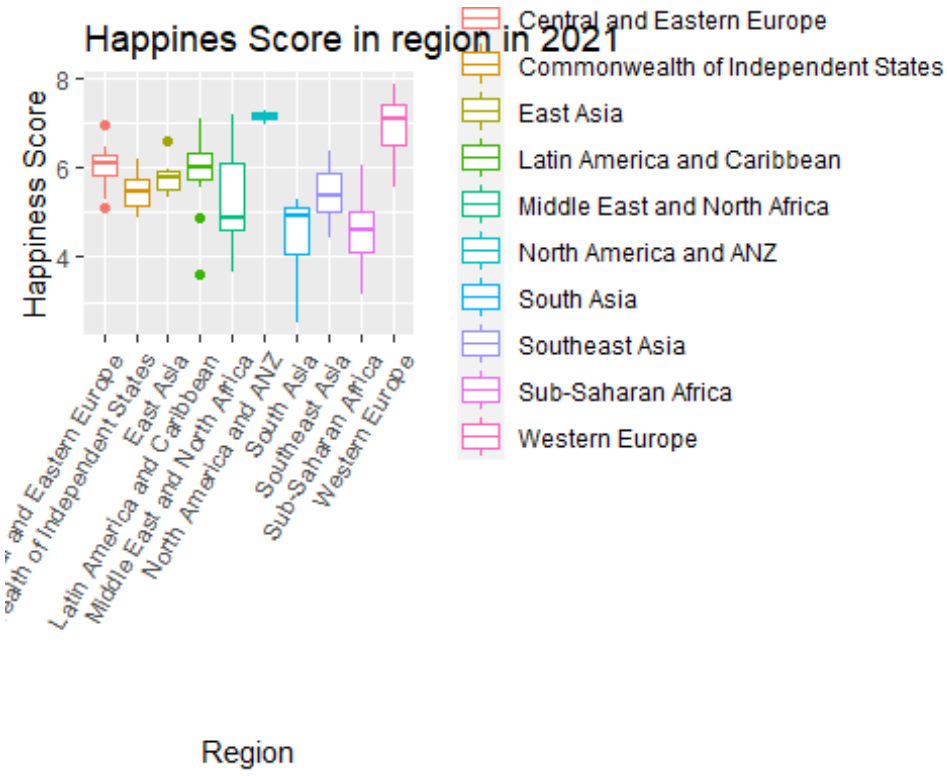
##looking at the Happiness score by region in 2016

```
names(data_2016)[4] <- 'Score'
ggplot(data_2016, aes(x= Region, y=Score, color=Region))+
  geom_boxplot()+
  theme(axis.text.x = element_text(angle=60, hjust=1))+
  labs(title = 'Happines Score in region in 2016',
       x= 'Region',
       y= 'Happiness Score')
```



##checking how is the Happiness score by region in 2021

```
year_2021 <-  read.csv("world-happiness-report-2021.csv")
names(year_2021)[2] <- 'Region'
names(year_2021)[3] <- 'Score'
ggplot(year_2021, aes(x= Region, y=Score, color=Region))+
  geom_boxplot()+
  theme(axis.text.x = element_text(angle=60, hjust=1))+
  labs(title = 'Happines Score in region in 2021',
       x= 'Region',
       y= 'Happiness Score')
```



After analyzing the given data, I have been able to interpret:

1. Correlation plot shows that the economy(money) has the largest of impact in determining the overall happiness followed by health and social support.
2. Scatter plot also shows that strong correlation between economy and the overall happiness score.
3. Box plot of 2016 shows that Western Europe is the happiest region and Sub-Saharan Africa as the least happy region.
4. Box plot of 2021 shows that Sub-Saharan Africa have shown some improvement in the last five years but they still the least happy countries.

Conclusion:

After analyzing the World Happiness Report, I was able to figure out the impact of different factors in determining the overall happiness of a nation. Economic (GDP) tends to be the strongest of all factors in determining happiness and followed by health (life expectancy).

Looking at the data, I could suggest that countries in the similar region often have similar living conditions which affects the overall score accordingly. For example, Scandinavian Countries were the happiest and the countries in the Africa and the Middle East the least happy.

I analyzed the recent World Happiness Report of 2021 to compare the results of 2016 and 2021, what and how much of the difference been made but to little surprise, the countries lying at the bottom have done much to address the issues of their citizen and Western Europe continue to dominate the world as being the happiest countries, Scandinavian Countries in particular.

References:

References:

- Aché, M. (2020, June 30). World Happiness Report up to 2020. Retrieved from https://www.kaggle.com/mathurinache/world-happiness-report
- Singh, A. (2021, March 22). World Happiness Report 2021. Retrieved from https://www.kaggle.com/ajaypalsinghlo/world-happiness-report-2021
- Home. (n.d.). Retrieved from https://worldhappiness.report/
- Koki25ando. (2018, June 09). Data Analysis of World Happiness Report. Retrieved from https://www.kaggle.com/koki25ando/data-analysis-of-world-happiness-report