MA304 Coursework 2: World Happiness Report

1705063, 1900716, 1906090, 1901181, 1901072, 1907683

December 6, 2019

Table of Contents

[Abstract 2](#_Toc26475183)

[Introduction 2](#_Toc26475184)

[Data 2](#_Toc26475185)

[Interesting Questions 3](#_Toc26475186)

[Methods 3](#_Toc26475187)

[Results 4](#_Toc26475188)

[Bar plots for the top-5 ranked countries in each region 4](#_Toc26475189)

[Visualising happiness scores using heat maps 8](#_Toc26475190)

[Exploring Happiness Score based on Regions 10](#_Toc26475191)

[Exploring factors with Strong Correlations with Happiness Score 10](#_Toc26475192)

[Regression Tree 12](#_Toc26475193)

[Correlation Plots 13](#_Toc26475194)

[Exploring the percentage change for the different factors of the World Happiness Report 18](#_Toc26475195)

[Discussion 19](#_Toc26475196)

[Appendix 19](#_Toc26475197)

[Code for ‘Bar plots for the top-5 ranked countries in each region’ 20](#_Toc26475198)

[Code for ‘Visualising happiness scores using heat maps’ 30](#_Toc26475199)

[Code for ‘Exploring Happiness Score based on Regions’ 33](#_Toc26475200)

[Code for ‘Exploring factors with Strong Correlations with Happiness Score’ 34](#_Toc26475201)

[Code for ‘Regression Tree’ 35](#_Toc26475202)

[Code for ‘Correlation Plots’ 36](#_Toc26475203)

[Code for ‘Exploring the percentage change for the different factors of the World Happiness Report’ 37](#_Toc26475204)

## Abstract

The World Happiness Report is a landmark survey of the state of global happiness that ranks countries all over the world by how happy their citizens perceive themselves to be. The World Happiness Report focuses on happiness and the community: how happiness has evolved over the past years, with a focus on the technologies, social norms, conflicts and government policies that have driven those changes and the estimated extent to which each of six factors - levels of GDP, life expectancy, generosity, social support, freedom, and corruption - contribute to making life evaluations higher in each country than they are in Dystopia, a hypothetical country that has values equal to the world’s lowest national averages for each of the six factors.

#### Word count: ~2180

## Introduction

This report explores happiness rankings and scores from different countries all over the world, in the span of 5 years (from 2015 to 2019). Would factors such as the economy and life expectancy affect the scores and rankings for a country? How did country ranks or scores change between the years 2015 to 2019? These would be relevant topics to discuss in the report.

## Data

The group is using the [World Happiness Report](https://www.kaggle.com/unsdsn/world-happiness) (available on Kaggle) as data for this coursework. The data is a landmark survey of the state of global happiness- we will be using years 2015 to 2019.

The .csv files contain information on a country’s happiness rank and happiness score, as well as factors that would contribute to the happiness score, such as economy, family, health (life expectancy), freedom, trust (government corruption), generosity, and dystopia residual.

The number of countries included in the datasets vary, but there are usually 155-158 countries listed. Below is a table focusing on the happiness score for each year:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Year | Mean | Median | Minimum | Maximum |
| **2015** | 5.376 | 5.232 | 2.839 | 7.587 |
| **2016** | 5.382 | 5.314 | 2.905 | 7.526 |
| **2017** | 5.354 | 5.279 | 2.693 | 7.537 |
| **2018** | 5.376 | 5.378 | 2.905 | 7.632 |
| **2019** | 5.407 | 5.380 | 2.853 | 7.769 |

There are only small variations to the mean and median of the years. Likewise, with the ranges between maximum and minimum, there are no big changes to it.

## Interesting Questions

These questions were explored/analysed in the report:

1. Can we design a regression tree diagram for the datasets with different happiness factors?
2. Can the happiness score be shown on a world map?
3. Can Top 5 ranking countries of each region be identified and visualised with their scores?
4. Can there be an interactive plot for countries and regions showing different statistics like mean, median, minimum, maximum?
5. Can we capture the percentage change of Happiness Score factors over 2015 to 2019?

## Methods

The rankings in figures and datasets of the World Happiness Report use data that come from the Gallup World Poll. The rankings are based on answers to the main life evaluation question asked in the poll.

We as a group came up with several brainstorming sessions and finalized on the following exploratory data analysis and visualizations methods:

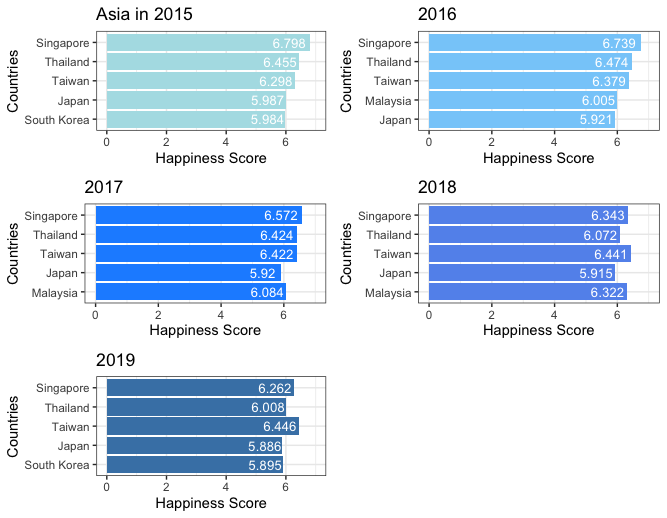
1. Bar plots of Top 5 ranking countries in 5 regions – Asia, Europe, Africa, South America , North America with Australia.
2. World Heat Map Plot depicting happiness score.
3. Interactive Scale Plot showing countries and continents with their respective minimum, maximum and mean values along with their individual scores.
4. Regression tree analysis for the World Happiness Score.
5. Correlation matrix plots for the factors.
6. Table for percentage change in the factors for the period 2015 to 2019.

The libraries used are:

* ggplot2
* ggmap
* maps
* mapdata
* dplyr
* stringr
* pdp
* plotly
* kableExtra
* knitr
* readr
* scales
* corrplot
* rpart
* rpart.plot
* gdata
* magrittr
* xlsx
* reshape
* corrgram
* plot3D
* ISLR
* data.table
* formattable
* tidyr

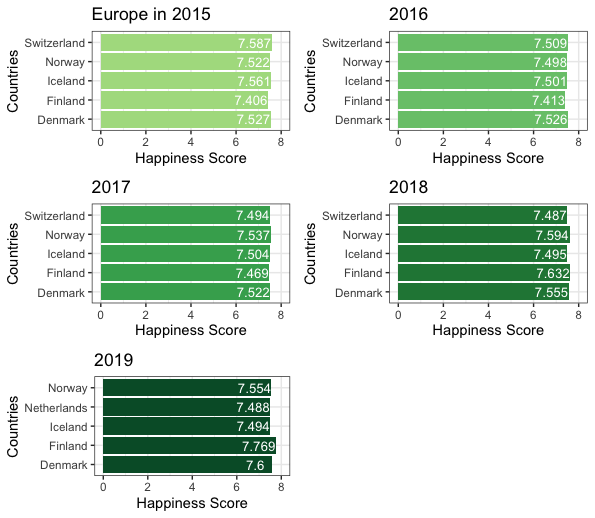
## Results

### Bar plots for the top-5 ranked countries in each region



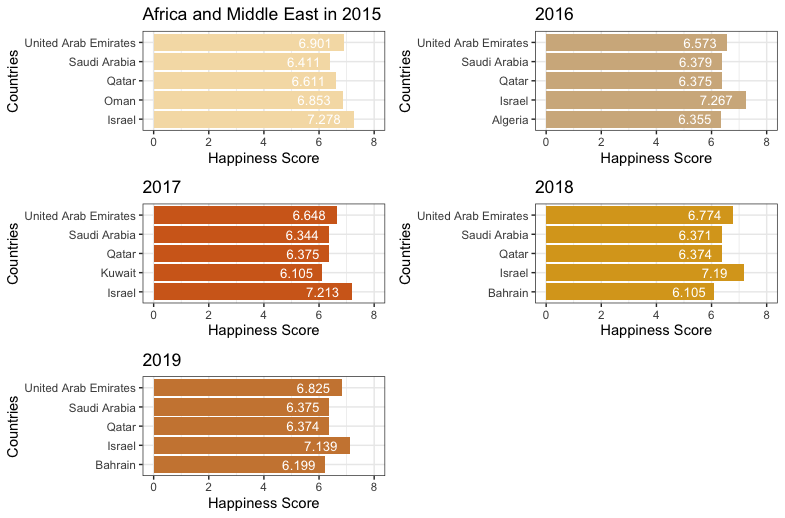
Asia from 2015 to 2019

It is clear that four of the five Asian countries are taking hold in the top five countries of the World Happiness Report, namely, Singapore, Thailand, Taiwan, Japan. While South Korea and Malaysia are trying to hold best five happiness score of Asia from 2015 to 2019. Although Singapore was ranked in the top for three years since 2015, the score of Singapore has decreased from 6.798 in 2015 to 6.262 in 2019, which declined by 8%. In contrast, Taiwan saw a significant increase in happiness score in this five years, and moreover, Taiwan was graded top in 2018 and 2019, at 6.441 and 6.446 respectively. The average scores of top five countries in Asia are about 6.



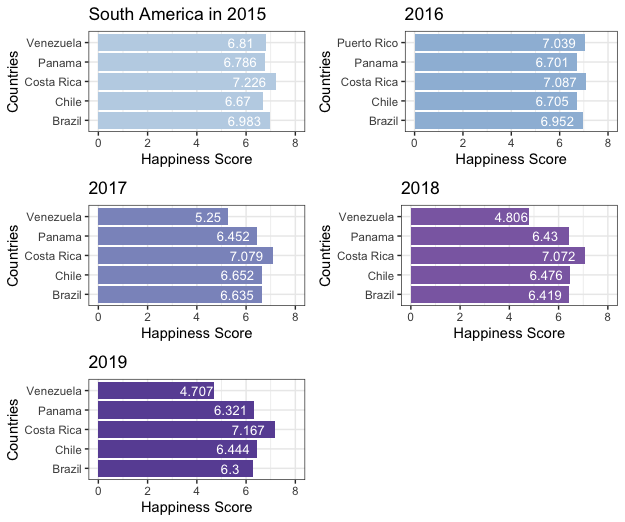
Europe from 2015 to 2019

Overall, the scores of European countries are at least 7.4. The happiness score of Finland increased significantly by approximately 5%. Finland also had the highest score at 7.632 in 2018 and 7.769 in 2019. However, Iceland saw a considerable decline from roughly 7.561 in 2015 to 7.494 in 2019. It is the first time that Netherlands was graded the fifth best (with a score of 7.488) among Europe in 2019 and at the same year, surprisingly, Switzerland was out of the top five europe countries, which was the champion of happiness score in 2015. Norway and Denmark remained stable.



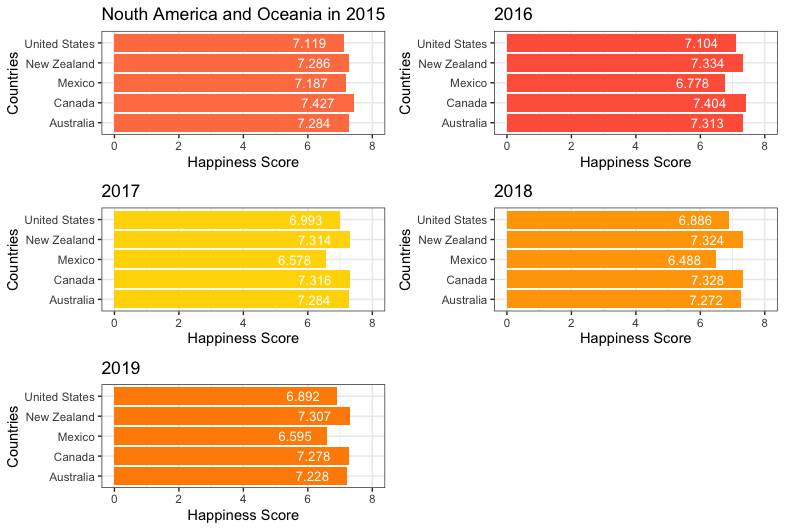
Africa and Middle East from 2015 to 2019

It is apparent that Israel has the highest happiness score in the Middle-East and Africa region in all the five years. Although the score of Israel has dropped by 0.5% slightly by years, it still keeps more than 7 each year. The countries in the Middle East keep occupied at the top five countries; however, Algeria which is located in North Africa was the only African country that was graded fifth best score at roughly 6.355 in 2016. United Arab Emirates remains the second best happiness score in Africa and Middle East countries in the five years.



South America from 2015 to 2019

Costa Rica has stayed on the best in happiness score in South America for five years, but the score has decreased by approximately 2% from 2015 to 2018 and increased by 1.3% in 2019. Other than Costa Rica, the scores of other countries in South America are less than 7. Chile and Brazil have similar happiness scores in the five years; however, Chile was lower about 4% than Brazil in the first two years and it became higher about 1% since 2017.



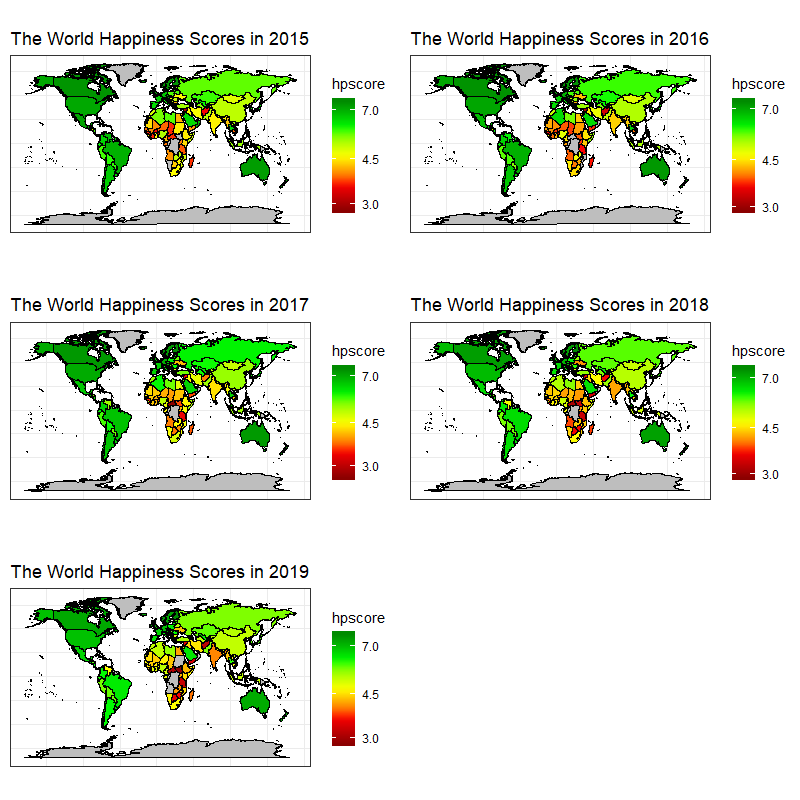
North America and Oceania from 2015 to 2019

Canada has had the highest happiness score in the region of North America and Oceania from 2015 (7.427) to 2018 (7.328); however, it became the second top in 2019, at about 7.278. For New Zealand, it has been graded second best for four years since 2015, while it got the highest score at 7.307 in 2019. There are two countries that remain the fourth and fifth best score, which is United States and Mexico, respectively.

### Visualising happiness scores using heat maps

The maps below illustrate the happiness scores of the world from 2015 to 2019. A hpscore (happiness score) of 7.0 means the country has a relatively high happiness score (and therefore will be coloured green in the hpscore colour spectrum), while a score of 3.0 means the country has a low happiness score (and thus coloured red according to the hpscore colour spectrum).

There are some countries greyed out/missing due to a couple reasons: the countries were not present in the data used, or the countries were not recognised in the mapdata package. Greenland is a result of the former reason, where it is not present in the data used, and therefore does not have a happiness score or ranking. Antartica, of course, would also not have a happiness score assigned to it as the population is made up of penguins, polar bears and other wild life.



World Maps illustrating Happiness Score from 2015 to 2019

The maps above show some change in happiness scores between the years 2015 and 2019. While countries like Canada and Australia are relatively green throughout (meaning their scores are in the 7.0 value range - a high happiness score), countries in the region of Africa are either coloured red or orange. This means their happiness scores are below 5.0.

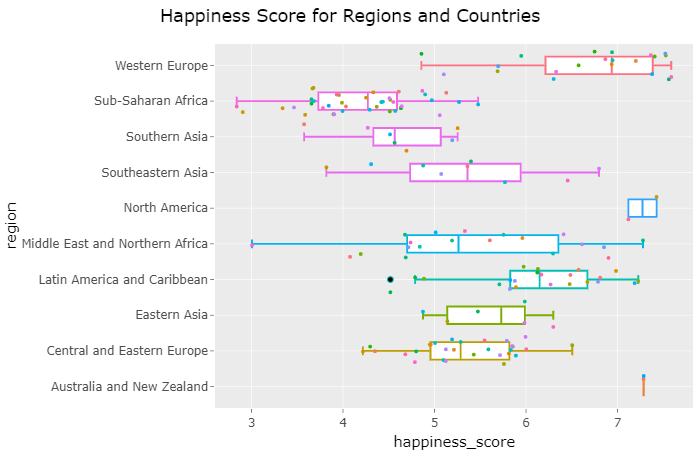
Comparing the happiness scores for the year 2015 with the happiness scores for 2019, we can see that India has a noticeable difference in colour (and therefore happiness score) in the two years. In 2015, India is coloured yellow. However, in 2019 it is coloured orange. We can assume that the happiness score for India in 2015 has decreased in a four-year period, resulting in the orange colour of India that is displayed in 2019.

Niger is another country that has a noticeable difference in colour in the years 2015 and 2019. However, unlike India, the country went from an orange red colour in 2015 to a brighter yellow in 2019, therefore, from the illustrations above, the happiness score for Niger increased substantially from 2015 to 2019.

### Exploring Happiness Score based on Regions

The scale plots (found in the Appendix) can be interacted with in HTML once eval is set to TRUE.

Below, the average happiness scores among regions are analysed.



Above is a screenshot of the interactive plot. The top 3 happiest regions based on average happiness scores are: Australia and New Zealand, North America, and Western Europe.

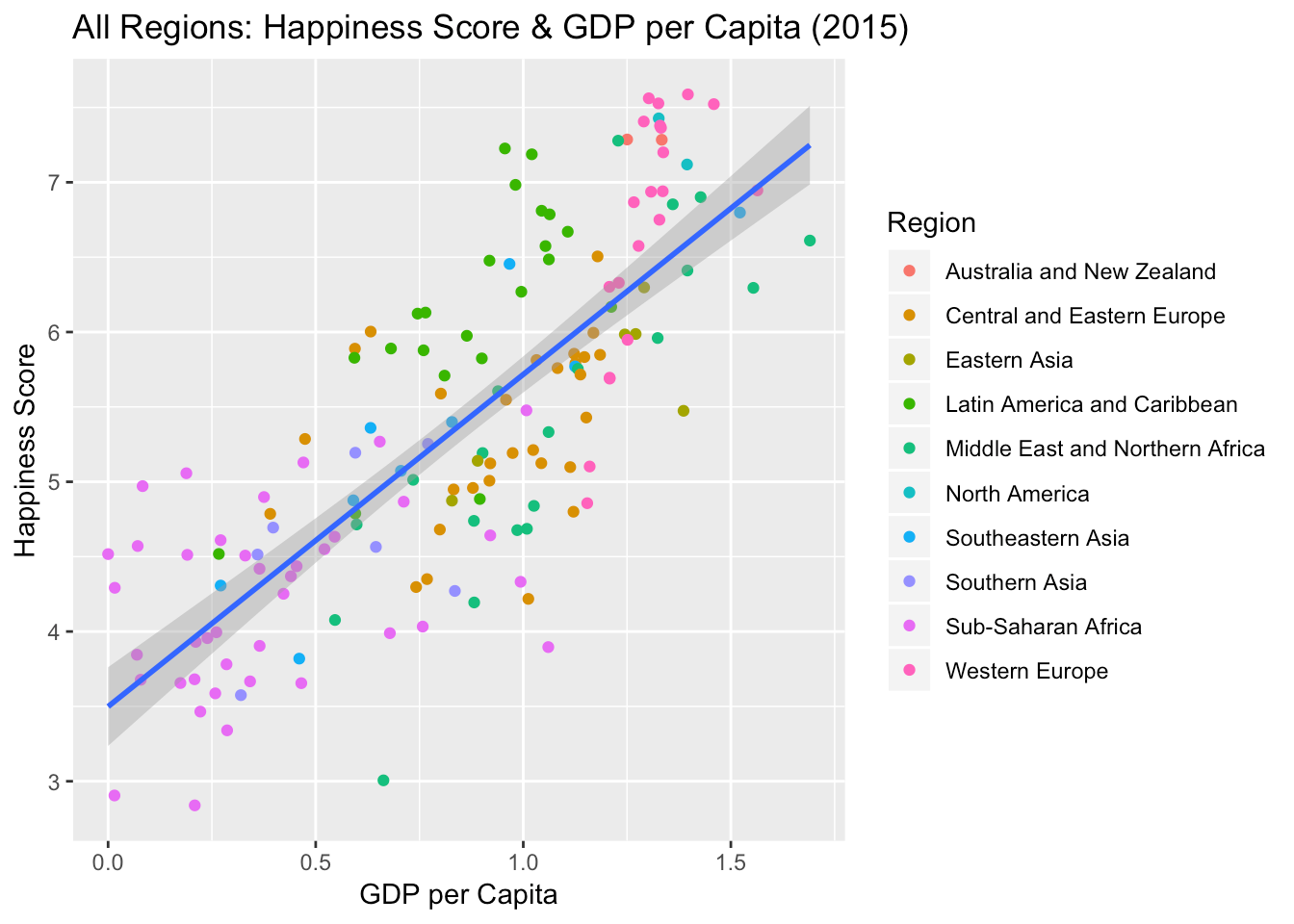
It is important to note that both of the first two regions include only two countries while Western Europe has 21 countries. Additionally, all of these regions include countries with developed economies.

The unhappiest region is Sub-Saharan Africa, which includes 40 different countries. The second unhappiest region is Southern Asia.

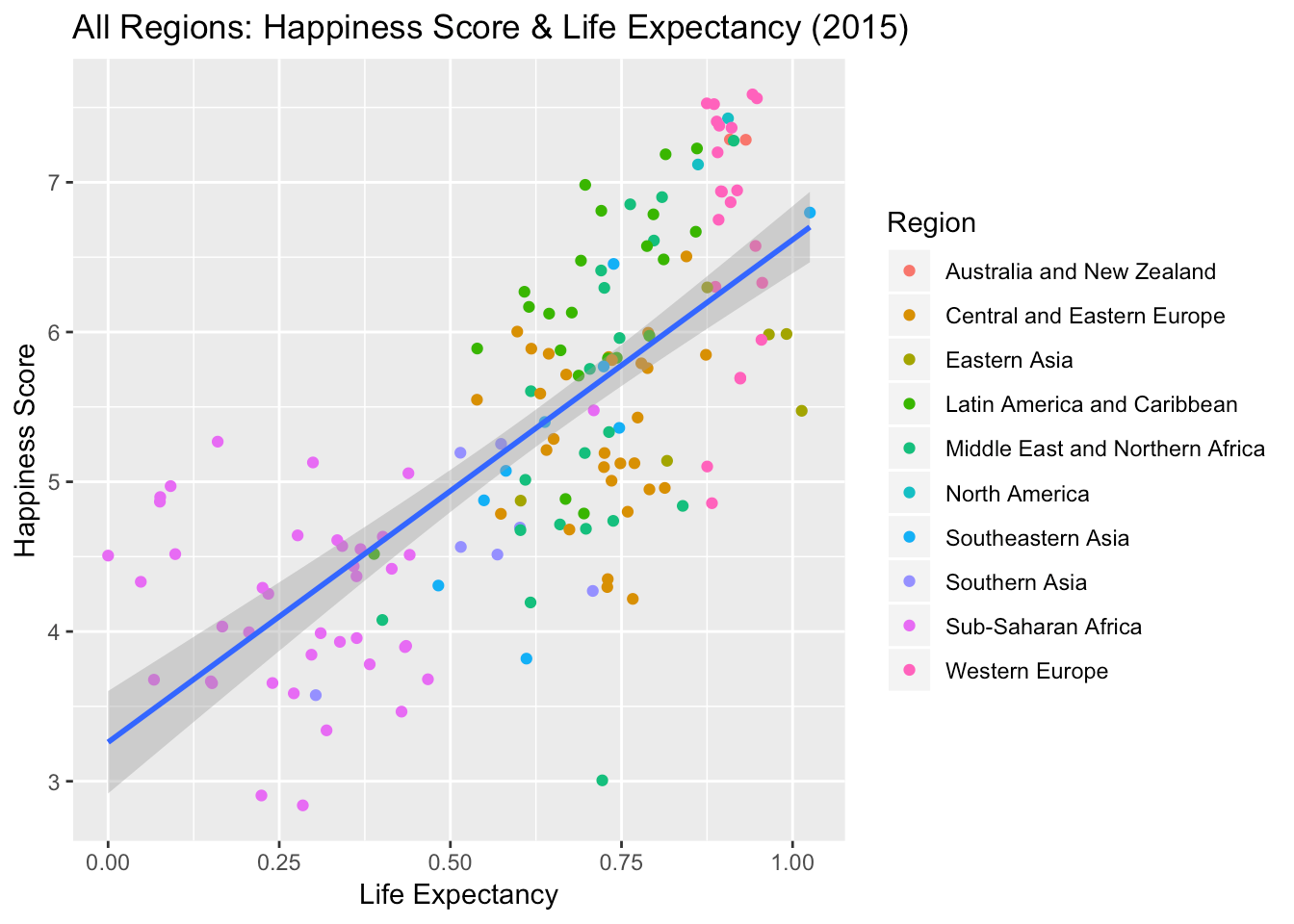
### Exploring factors with Strong Correlations with Happiness Score

The characteristics of the happiest and unhappiest regions are focused on in this section.

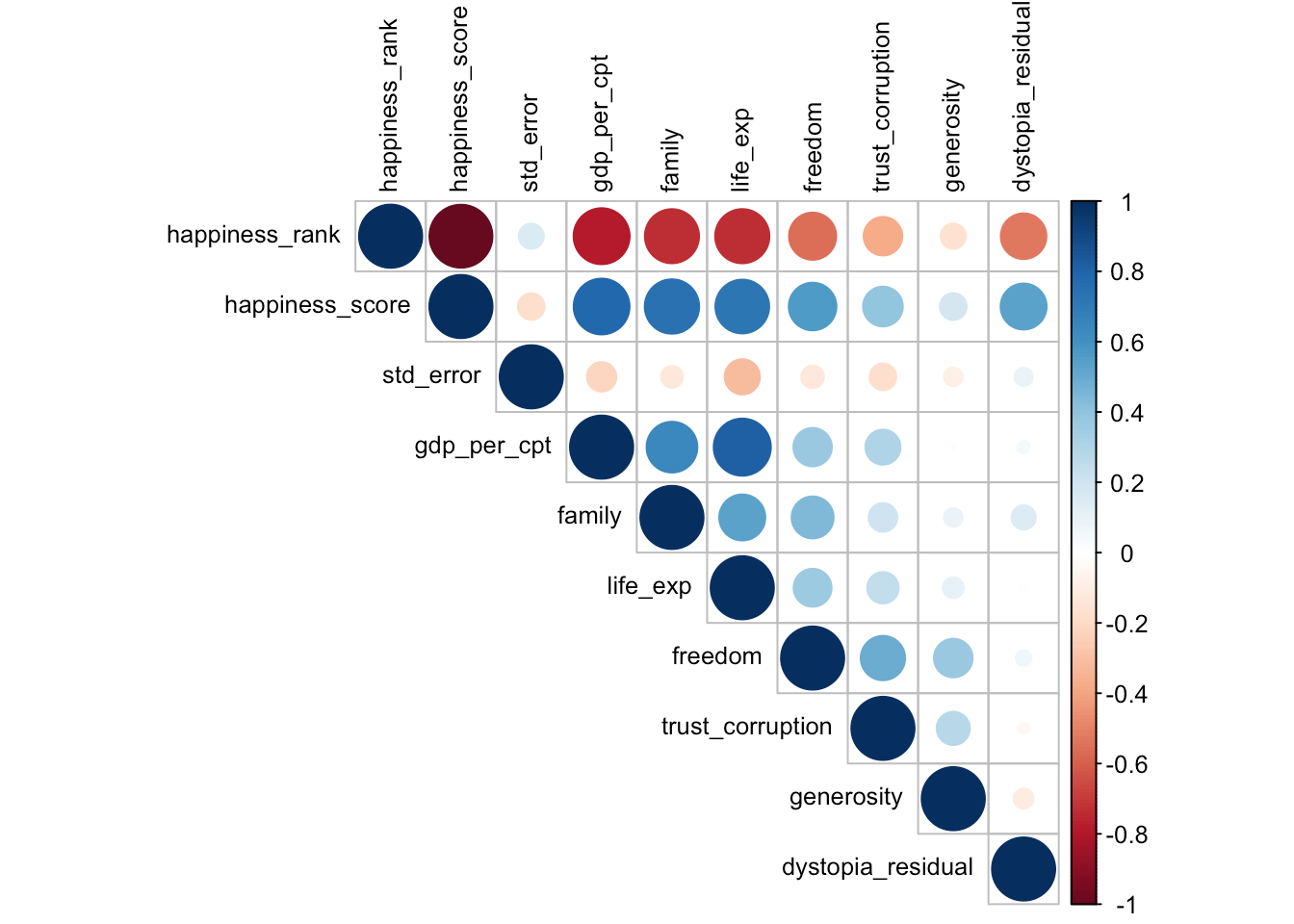
Strong correlations will give an idea on which factors and attributes contributed more towards the happiness score. It is always important to note that correlation and causation are different things!



From the above plot, countries that have a higher *GDP per Capita* seem to be happier. We can assume that a higher *GDP per Capita* indicates a better standard of living for a country.



From the plot above, we can assume that the average life expectancy contributes to a happier country.



The plot above allows us to analyze the relationships between the different factors.

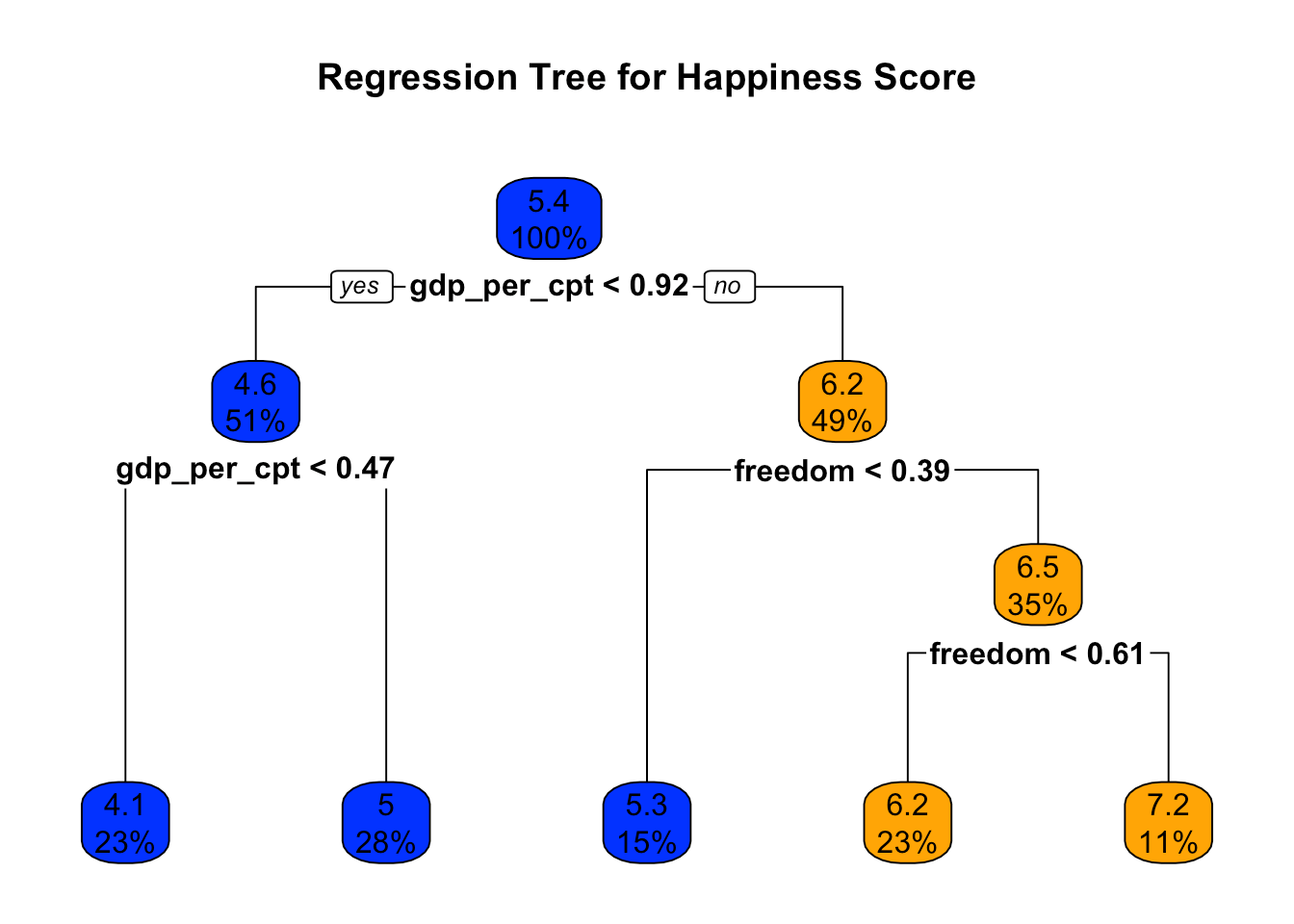
### Regression Tree

From the regression tree below, the most optimal complexity parameter for regression analysis were:

1. GDP per Capita
2. Freedom

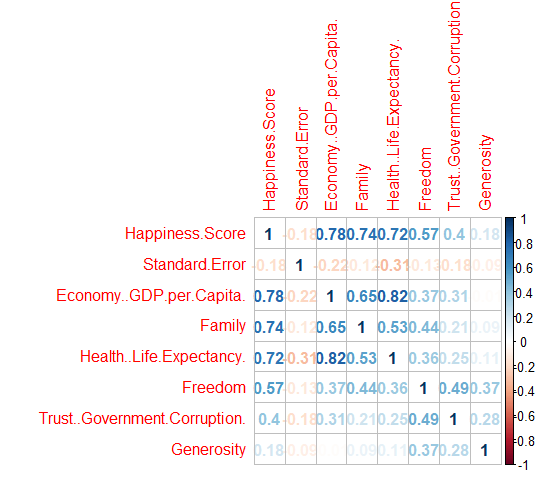
Out of 157 countries that were included in the World Happiness Report in 2015, the average happiness score turned out to be 5.4 out of 10.

80 countries had a GDP per capita of < 0.92, while 77 countries had a GDP per capita of > 0.92. Their happiness score was higher (6.2) as compared to their counterparts having a score of 4.6



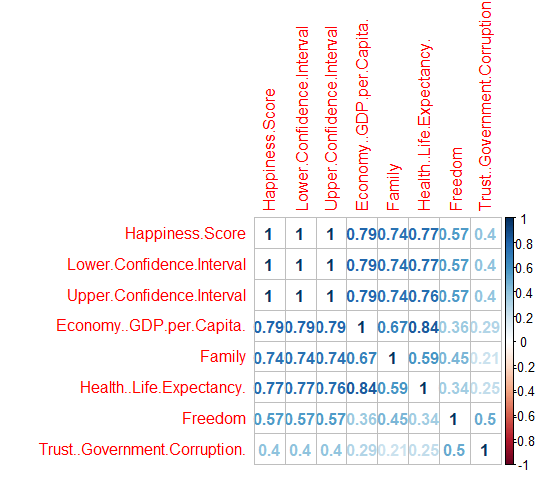
### Correlation Plots

A correlation matrix for the World Happiness data would help us understand the relationship between the variables present in the dataset.



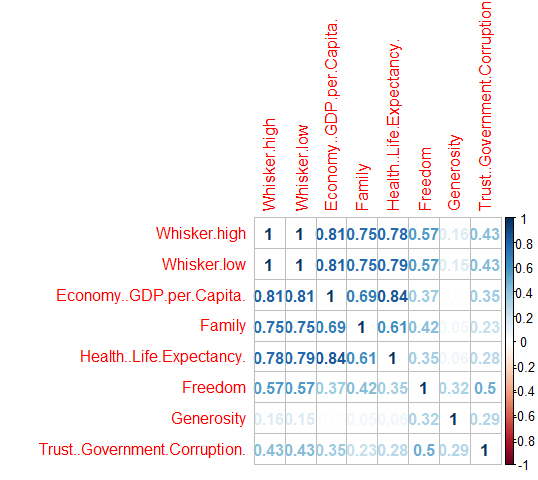
Correlation Plot for 2015

The correlation between each attribute of the dataset can be viewed in the image above. A strong correlation of 0.82 exists between *health (life expectancy)* and *economy (GDP per capita)*. The data suggests that in the year 2015, a correlation of 0.78 and 0.74 exists between the happiness score to *economy* and *family* respectively.



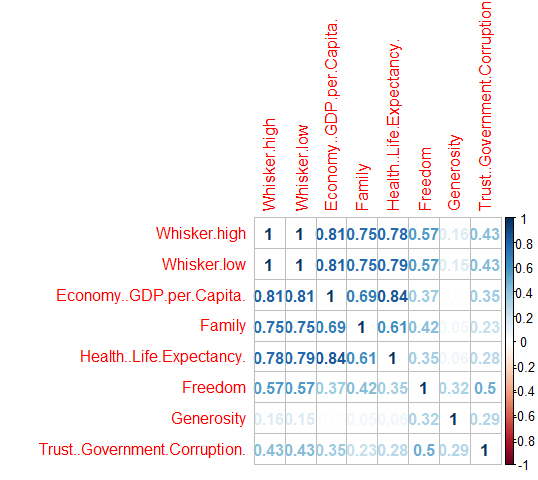
Correlation Plot for 2016

The data for the year 2016 shows that there is a strong correlation of 0.84 between *health* and *economy*. Similarly, there exists a correlation between the *trust (government corruption)* and *family* which is numerically equivalent to 0.21.



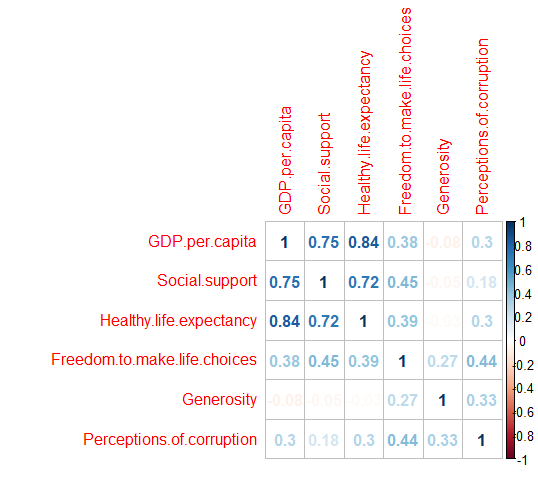
Correlation Plot for 2017

There exists no correlation between *economy* and *generosity*. However, *economy* and *health* are strongly correlated.



Correlation Plot for 2018

There exists a 0.84 correlation between *economy* and *health*. However, the correlation between *economy* and *freedom*, *generosity* and *trust* is almost nill.



Correlation Plot for 2019

There exists a 0.84 correlation between *economy* and *health*. Alternatively, *generosity* and *economy* are negatively correlated.

Therefore, from the data, between the years 2015 and 2019, there exists a strong correlation between *Economy (GDP per capita)* and *Health (life expectancy)* which is at around 0.84.

### Exploring the percentage change for the different factors of the World Happiness Report

The top 5 ranked countries in 2019:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Country | economychange | socialchange | lifechange | genechange | freechange | corruchange | improvement |
| Finland | +3.86 | +20.39 | +10.49 | -34.48 | -7.12 | -5.01 | +4.9 |
| Denmark | +4.34 | +15.61 | +13.88 | -26.18 | -8.84 | -15.21 | +0.97 |
| Norway | +1.99 | +18.86 | +16.13 | -21.9 | -9.96 | -6.58 | +0.43 |
| Iceland | +5.96 | +15.82 | +8.25 | -18.86 | -6.01 | -16.58 | -0.89 |
| Netherlands | +5.01 | +18.89 | +11.89 | -32.38 | -9.54 | -6.33 | +1.49 |

The bottom 5 ranked countries in 2019:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Country | economychange | socialchange | lifechange | genechange | freechange | corruchange | improvement |
| Yemen | -47.48 | +70.8 | +15.57 | +18.28 | -59.8 | -1.96 | -17.1 |
| Rwanda | +61.65 | -8.1 | +43.24 | -4.1 | -6.25 | -25.53 | -3.78 |
| Tanzania | +66.9 | -11.74 | =30.58 | -19.71 | +26.83 | +155.79 | -14.55 |
| Afghanistan | +9.44 | +70.71 | +19 | -56.72 | -100 | -74.28 | -10.41 |
| Central African Republic | -66.88 | NaN | +56.74 | -1.41 | -53.97 | -57.78 | -16.18 |

From 2015 to 2019, most of the countries have an incremental change in the *economychange*, *socialchange*, and *lifechange* factors. Whereas there is a decremental change for *genechange*, *freechange*, and *corruchange* factors from 2015 to 2019.

## Discussion

There were limitations to the data used as it did not present all the countries in the world. If there was an opportunity to expand on the map work done for this coursework, possibly finding the average happiness score for each region/continent and illustrating that in a map would be informational. However, even without mapping out the different regions, the map plots above provide a clear illustration of the happiness scores in a general area of the map.

Compared to the five regions, Europe has the highest average happiness score, with North America being the second highest.

In the [World Happiness Report website](https://worldhappiness.report/faq/), it explains that the six factors used to explain life evaluations reflect what has been found in research. As explained in the Results section above, the factors illustrate some correlation but obviously does not give a full explanation on cause.

## Appendix

Loading the .csv files that will be used for this coursework:

# The .csv files should be found in the data folder  
  
data2015 = read.csv("data/2015.csv", header = TRUE, sep = ",")  
data2016 = read.csv("data/2016.csv", header = TRUE, sep = ",")  
data2017 = read.csv("data/2017.csv", header = TRUE, sep = ",")  
data2018 = read.csv("data/2018.csv", header = TRUE, sep = ",")  
data2019 = read.csv("data/2019.csv", header = TRUE, sep = ",")

### Code for ‘Bar plots for the top-5 ranked countries in each region’

#Asia group 2015  
asia2015\_index = grep("Asia", data2015$Region) #Select the group of Asia in data2015  
asia2015\_index = asia2015\_index[1:5] #Take rank 5 of Asia from dataset  
asia2015 = data2015[asia2015\_index, ] #take asia\_index for row and all the columns  
asia2015 = select(asia2015, Country, Happiness.Score) #Choose the columns of country and happiness score from asia2015  
  
#Asia group 2016  
asia2016\_index = grep("Asia", data2016$Region)  
asia2016\_index = asia2016\_index[1:5]  
asia2016 = data2016[asia2016\_index,]  
asia2016 = select(asia2016, Country, Happiness.Score)  
  
#Asia group 2017  
asia2017\_a = c('Afghanistan','Cambodia','Sri Lanka','Myanmar','Nepal',  
 'India','Bangladesh','Mongolia','Laos','Philippines',  
 'China','Pakistan','Bhutan','Vietnam','Indonesia',  
 'Hong Kong','Malaysia','South Korea','Japan','Taiwan',  
 'Thailand','Singapore')  
asia2017 = filter(data2017, grepl(paste(asia2017\_a, collapse="|"),data2017$Country)) #Select the countries from asia2017\_a in data2017  
asia2017 = asia2017[1:5,]  
asia2017 = select(asia2017, Country, Happiness.Score)  
asia2017$Country = factor(asia2017$Country,   
 level = c('Malaysia','Japan','Taiwan','Thailand','Singapore'))  
  
#Asia group 2018  
asia2018 = filter(data2018, grepl(paste(asia2017\_a, collapse="|"),data2018$Country))  
asia2018 = asia2018[1:5,]  
asia2018 = select(asia2018, Country, Score)  
  
#Asia group 2019  
asia2019 = filter(data2019, grepl(paste(asia2017\_a, collapse="|"),data2019$Country.or.region))  
asia2019 = asia2019[1:5,]  
asia2019 = select(asia2019, Country.or.region, Score)  
  
#Asia plot  
as2015gg = ggplot() + geom\_bar(data = asia2015, aes(Country, Happiness.Score),  
 fill = "#B0E0E6", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("Asia in 2015") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,7))+  
 geom\_text(aes(asia2015$Country, asia2015$Happiness.Score,   
 label = asia2015$Happiness.Score),   
 position = position\_stack(vjust = 0.9), color="white", size=3.5)  
  
as2016gg = ggplot() + geom\_bar(data = asia2016, aes(Country, Happiness.Score),   
 fill = "#87CEFA", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2016") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,7))+  
 geom\_text(aes(asia2016$Country, asia2016$Happiness.Score,   
 label = asia2016$Happiness.Score),   
 position = position\_stack(vjust = 0.9), color="white", size=3.5)  
  
  
as2017gg = ggplot() + geom\_bar(data = asia2017, aes(Country, Happiness.Score),  
 fill = "#1E90FF", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2017") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,7))+  
 geom\_text(aes(asia2017$Country, asia2017$Happiness.Score,   
 label = asia2017$Happiness.Score),   
 position = position\_stack(vjust = 0.9), color="white", size=3.5)  
  
as2018gg = ggplot() + geom\_bar(data = asia2018, aes(Country, Score),  
 fill = "#6495ED", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2018") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,7))+  
 geom\_text(aes(asia2018$Country, asia2018$Score,   
 label = asia2018$Score),   
 position = position\_stack(vjust = 0.9), color="white", size=3.5)  
  
as2019gg = ggplot() + geom\_bar(data = asia2019, aes(Country.or.region, Score),  
 fill = "#4682B4", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2019") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,7))+  
 geom\_text(aes(asia2019$Country, asia2019$Score,   
 label = asia2019$Score),   
 position = position\_stack(vjust = 0.9), color="white", size=3.5)  
  
  
grid.arrange(as2015gg,as2016gg,as2017gg,as2018gg,as2019gg) #combine with all plots from 2015 to 2019

#Europe group 2015  
europe\_index = grep("Europe", data2015$Region)  
europe2015 = data2015[europe\_index,]  
europe2015 = europe2015[1:5,]  
europe2015 = select(europe2015, Country, Happiness.Score)  
  
europe\_level=c('Bulgaria','Georgia','Armenia','Ukraine','Tajikistan',  
 'Hungary','Greece','Bosnia and Herzegovina',  
 'Albania','Macedonia','Latvia','Portugal',  
 'Serbia','Romania','Montenegro','Azerbaijan',  
 'Kyrgyzstan','Estonia','Turkmenistan','Kosovo',  
 'Cyprus','North Cyprus','Russia','Croatia',  
 'Poland','Belarus','Lithuania','Slovenia',  
 'Kazakhstan','Moldova','Italy','Slovakia',  
 'Uzbekistan','Malta','Spain','Czech Republic',  
 'France','Germany','United Kingdom','Belgium',  
 'Ireland','Luxembourg','Austria','Sweden',  
 'Netherlands','Finland','Norway','Denmark',  
 'Iceland','Switzerland')  
  
#Europe grope 2016  
europe2016\_index = grep("Europe", data2016$Region)  
europe2016 = data2016[europe2016\_index,]  
europe2016 = europe2016[1:5,]  
europe2016 = select(europe2016, Country, Happiness.Score)  
  
#Europe group 2017  
europe2017 = filter(data2017, grepl(paste(europe\_level, collapse="|"),Country))  
europe2017 = europe2017[1:5,]  
europe2017 = select(europe2017, Country, Happiness.Score)  
  
#Europe group 2018  
europe2018 = filter(data2018, grepl(paste(europe\_level, collapse="|"),Country))  
europe2018 = europe2018[1:5,]  
europe2018 = select(europe2018, Country, Score)  
  
#Europe group 2019  
europe2019 = filter(data2019, grepl(paste(europe\_level, collapse="|"),data2019$Country))  
europe2019 = europe2019[1:5,]  
europe2019 = select(europe2019, Country.or.region, Score)  
  
  
eu2015gg = ggplot() + geom\_bar(data = europe2015, aes(Country, Happiness.Score),  
 fill = "#addd8e", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("Europe in 2015") + labs(x="Countries", y = "HappinessScore")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(europe2015$Country, europe2015$Happiness.Score,   
 label = europe2015$Happiness.Score),   
 position = position\_stack(vjust = 0.9), color="white", size=3.5)  
  
eu2016gg = ggplot() + geom\_bar(data = europe2016, aes(Country, Happiness.Score),   
 fill = "#78c679", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2016") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(europe2016$Country, europe2016$Happiness.Score,   
 label = europe2016$Happiness.Score),   
 position = position\_stack(vjust = 0.9), color="white", size=3.5)  
  
eu2017gg = ggplot() + geom\_bar(data = europe2017, aes(Country, Happiness.Score),  
 fill = "#41ab5d", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2017") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(europe2017$Country, europe2017$Happiness.Score,   
 label = europe2017$Happiness.Score),   
 position = position\_stack(vjust = 0.9), color="white", size=3.5)  
  
eu2018gg = ggplot() + geom\_bar(data = europe2018, aes(Country, Score),  
 fill = "#238443", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2018") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(europe2018$Country, europe2018$Score,   
 label = europe2018$Score),   
 position = position\_stack(vjust = 0.9), color="white", size=3.5)  
  
eu2019gg = ggplot() + geom\_bar(data = europe2019, aes(Country.or.region, Score),  
 fill = "#005a32", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2019") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(europe2019$Country, europe2019$Score,   
 label = europe2019$Score),   
 position = position\_stack(vjust = 0.9), color="white", size=3.5)  
  
grid.arrange(eu2015gg,eu2016gg,eu2017gg,eu2018gg,eu2019gg)

#Africa group 2015  
africa\_index = grep("Africa", data2015$Region)  
africa2015 = data2015[africa\_index,]  
africa\_country = africa2015$Country  
africa2015 = africa2015[1:5,]  
africa2015 = select(africa2015, Country, Happiness.Score)  
  
#Afica group 2016  
africa\_index = grep("Africa", data2016$Region)  
africa2016 = data2016[africa\_index,]  
africa2016 = africa2016[1:5,]  
africa2016 = select(africa2016, Country, Happiness.Score)  
  
#Africa group 2017  
africa2017 = filter(data2017, grepl(paste(africa\_country, collapse="|"),Country))  
africa2017 = africa2017[1:5,]  
africa2017 = select(africa2017, Country, Happiness.Score)  
  
#Africa group 2018  
africa2018 = filter(data2018, grepl(paste(africa\_country, collapse="|"),Country))  
africa2018 = africa2018[1:5,]  
africa2018 = select(africa2018, Country, Score)  
  
#Africa group 2019  
africa2019 = filter(data2019, grepl(paste(africa\_country, collapse="|"),Country.or.region))  
africa2019 = africa2019[1:5,]  
africa2019 = select(africa2019, Country.or.region, Score)  
  
  
af2015gg = ggplot() + geom\_bar(data = africa2015, aes(Country, Happiness.Score),  
 fill = "#F5DEB3", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("Africa and Middle East in 2015") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(africa2015$Country, africa2015$Happiness.Score,   
 label = africa2015$Happiness.Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
af2016gg = ggplot() + geom\_bar(data = africa2016, aes(Country, Happiness.Score),   
 fill = "#D2B48C", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2016") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(africa2016$Country, africa2016$Happiness.Score,   
 label = africa2016$Happiness.Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
af2017gg = ggplot() + geom\_bar(data = africa2017, aes(Country, Happiness.Score),  
 fill = "#D2691E", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2017") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(africa2017$Country, africa2017$Happiness.Score,   
 label = africa2017$Happiness.Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
af2018gg = ggplot() + geom\_bar(data = africa2018, aes(Country, Score),  
 fill = "#DAA520", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2018") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(africa2018$Country, africa2018$Score,   
 label = africa2018$Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
af2019gg = ggplot() + geom\_bar(data = africa2019, aes(Country.or.region, Score),  
 fill = "#CD853F", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2019") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(africa2019$Country, africa2019$Score,   
 label = africa2019$Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
grid.arrange(af2015gg,af2016gg,af2017gg,af2018gg,af2019gg)

#South America 2015  
sudamerica\_index = grep("Latin America", data2015$Region)  
sudamerica2015 = data2015[sudamerica\_index,]  
sudamerica\_country = sudamerica2015$Country  
sudamerica2015 = sudamerica2015[1:5,]  
sudamerica2015 = select(sudamerica2015, Country, Happiness.Score)  
  
#South America 2016  
sudamerica\_index = grep("Latin America", data2016$Region)  
sudamerica2016 = data2016[sudamerica\_index,]  
sudamerica2016 = sudamerica2016[1:5,]  
sudamerica2016 = select(sudamerica2016, Country, Happiness.Score)  
  
#South America 2017  
sudamerica2017 = filter(data2017, grepl(paste(sudamerica\_country, collapse="|"),Country))  
sudamerica2017 = sudamerica2017[1:5,]  
sudamerica2017 = select(sudamerica2017, Country, Happiness.Score)  
  
#South America 2018  
sudamerica2018 = filter(data2018, grepl(paste(sudamerica\_country, collapse="|"),Country))  
sudamerica2018 = sudamerica2018[1:5,]  
sudamerica2018 = select(sudamerica2018, Country, Score)  
  
#South America 2019  
sudamerica2019 = filter(data2019, grepl(paste(sudamerica\_country, collapse="|"),Country.or.region))  
sudamerica2019 = sudamerica2019[1:5,]  
sudamerica2019 = select(sudamerica2019, Country.or.region, Score)  
  
sa2015gg = ggplot() + geom\_bar(data = sudamerica2015, aes(Country, Happiness.Score),  
 fill = "#bfd3e6", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("South America in 2015") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(sudamerica2015$Country, sudamerica2015$Happiness.Score,   
 label = sudamerica2015$Happiness.Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
sa2016gg = ggplot() + geom\_bar(data = sudamerica2016, aes(Country, Happiness.Score),   
 fill = "#9ebcda", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2016") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(sudamerica2016$Country, sudamerica2016$Happiness.Score,   
 label = sudamerica2016$Happiness.Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
sa2017gg = ggplot() + geom\_bar(data = sudamerica2017, aes(Country, Happiness.Score),  
 fill = "#8c96c6", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2017") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(sudamerica2017$Country, sudamerica2017$Happiness.Score,   
 label = sudamerica2017$Happiness.Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
sa2018gg = ggplot() + geom\_bar(data = sudamerica2018, aes(Country, Score),  
 fill = "#8c6bb1", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2018") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(sudamerica2018$Country, sudamerica2018$Score,   
 label = sudamerica2018$Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
sa2019gg = ggplot() + geom\_bar(data = sudamerica2019, aes(Country.or.region, Score),  
 fill = "#6a51a3", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2019") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(sudamerica2019$Country, sudamerica2019$Score,   
 label = sudamerica2019$Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
grid.arrange(sa2015gg,sa2016gg,sa2017gg,sa2018gg,sa2019gg)

#North America, Australia and New Zealand 2015  
noramerica\_country = c("Australia","New Zealand","United States","Canada","Mexico")  
noramerica2015 = filter(data2015, grepl(paste(noramerica\_country, collapse="|"),Country))  
noramerica2015 = select(noramerica2015, Country, Happiness.Score)  
  
#North America, Australia and New Zealand 2016  
noramerica2016 = filter(data2016, grepl(paste(noramerica\_country, collapse="|"),Country))  
noramerica2016 = select(noramerica2016, Country, Happiness.Score)  
  
#North America, Australia and New Zealand 2017  
noramerica2017 = filter(data2017, grepl(paste(noramerica\_country, collapse="|"),Country))  
noramerica2017 = select(noramerica2017, Country, Happiness.Score)  
  
#North America, Australia and New Zealand 2018  
noramerica2018 = filter(data2018, grepl(paste(noramerica\_country, collapse="|"),Country))  
noramerica2018 = select(noramerica2018, Country, Score)  
  
#North America, Australia and New Zealand 2019  
noramerica2019 = filter(data2019, grepl(paste(noramerica\_country, collapse="|"),Country.or.region))  
noramerica2019 = select(noramerica2019, Country.or.region, Score)  
  
  
na2015gg = ggplot() + geom\_bar(data = noramerica2015, aes(Country, Happiness.Score),  
 fill = "#FF7F50", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("Nouth America and Oceania in 2015") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(noramerica2015$Country, noramerica2015$Happiness.Score,   
 label = noramerica2015$Happiness.Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
na2016gg = ggplot() + geom\_bar(data = noramerica2016, aes(Country, Happiness.Score),   
 fill = "#FF6347", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2016") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(noramerica2016$Country, noramerica2016$Happiness.Score,   
 label = noramerica2016$Happiness.Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
na2017gg = ggplot() + geom\_bar(data = noramerica2017, aes(Country, Happiness.Score),  
 fill = "#FFD700", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2017") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(noramerica2017$Country, noramerica2017$Happiness.Score,   
 label = noramerica2017$Happiness.Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
na2018gg = ggplot() + geom\_bar(data = noramerica2018, aes(Country, Score),  
 fill = "#FFA500", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2018") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(noramerica2018$Country, noramerica2018$Score,   
 label = noramerica2018$Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
na2019gg = ggplot() + geom\_bar(data = noramerica2019, aes(Country.or.region, Score),  
 fill = "#FF8C00", stat = "identity")+theme\_bw()+coord\_flip()+  
 ggtitle("2019") + labs(x="Countries", y = "Happiness Score")+  
 scale\_y\_continuous(limits =c(0,8))+  
 geom\_text(aes(noramerica2019$Country, noramerica2019$Score,   
 label = noramerica2019$Score),   
 position = position\_stack(vjust = 0.85), color="white", size=3.5)  
  
grid.arrange(na2015gg,na2016gg,na2017gg,na2018gg,na2019gg)

### Code for ‘Visualising happiness scores using heat maps’

data2015 = read.csv("data/2015.csv", header = TRUE, sep = ",")  
data2016 = read.csv("data/2016.csv", header = TRUE, sep = ",")  
data2017 = read.csv("data/2017.csv", header = TRUE, sep = ",")  
data2018 = read.csv("data/2018.csv", header = TRUE, sep = ",")  
data2019 = read.csv("data/2019.csv", header = TRUE, sep = ",")  
  
## this will be used throughout creating the map visualisations  
  
# using mapdata package to get data of the world map - this includes information on latitude, longitude, etc.  
world\_map <- map\_data("world")  
  
# creating the base of the map - this acts as the foundation for all map visualisations  
map\_base <- ggplot( data = world\_map, mapping = aes( x = long, y = lat, group = group ) ) + coord\_fixed(1.3) + geom\_polygon(color = "black", fill = "gray") + theme\_nothing()  
  
# we don't want to display certain things like latitude and longitude values in the axes  
ignore\_axes <- theme( axis.text = element\_blank(), axis.line = element\_blank(), axis.ticks = element\_blank(), axis.title = element\_blank() )

## preparing the data for the year 2015   
  
mod\_dat2015 <- cbind(data2015)  
  
# we need to rename some countries in the data so it would match with the names used in the mapdata package  
# as the countries column is of class factor, we need to create a few new levels to add in order to rename the countries  
levels(mod\_dat2015$Country) <- c( levels(mod\_dat2015$Country), c("UK","USA","Palestine","Somalia") )  
mod\_dat2015[ which(mod\_dat2015[,1] == "United States") ,1] <- "USA"  
mod\_dat2015[ which(mod\_dat2015[,1] == "United Kingdom") ,1] <- "UK"  
mod\_dat2015[ which(mod\_dat2015[,1] == "Somaliland region") ,1] <- "Somalia"  
mod\_dat2015[ which(mod\_dat2015[,1] == "Palestinian Territories") ,1] <- "Palestine"  
  
# we want to subset mod\_dat2015 and only get the Country and Happiness Score columns  
country\_hpscore15 <- subset( mod\_dat2015, select = c("Country","Happiness Score") )  
# rename the columns to fit in with map\_data  
names(country\_hpscore15) <- c("region","hpscore")  
  
# add hpscore to the world\_map variable  
world\_hpscore15 <- inner\_join(world\_map, country\_hpscore15, by = "region")  
  
# create the map for 2015  
map\_hp2015 <- map\_base + geom\_polygon( data = world\_hpscore15, aes(fill = hpscore), color = "white" ) + geom\_polygon(color = "black", fill = NA) + theme\_bw() + scale\_fill\_gradientn(colors = c("red4","red","yellow","green","green4"), breaks = c(3,4.5,7), trans = "log10") + ggtitle("The World Happiness Scores in 2015") + ignore\_axes

## preparing the data for the year 2016   
  
mod\_dat2016 <- cbind(data2016)  
  
levels(mod\_dat2016$Country) <- c( levels(mod\_dat2016$Country), c("UK","USA","Palestine","Somalia") )  
mod\_dat2016[ which(mod\_dat2016[,1] == "United States") ,1] <- "USA"  
mod\_dat2016[ which(mod\_dat2016[,1] == "United Kingdom") ,1] <- "UK"  
mod\_dat2016[ which(mod\_dat2016[,1] == "Somaliland region") ,1] <- "Somalia"  
mod\_dat2016[ which(mod\_dat2016[,1] == "Palestinian Territories") ,1] <- "Palestine"  
  
country\_hpscore16 <- subset( mod\_dat2016, select = c("Country","Happiness Score") )  
names(country\_hpscore16) <- c("region","hpscore")  
  
world\_hpscore16 <- inner\_join(world\_map, country\_hpscore16, by = "region")  
  
map\_hp2016 <- map\_base + geom\_polygon( data = world\_hpscore16, aes(fill = hpscore), color = "white" ) + geom\_polygon(color = "black", fill = NA) + theme\_bw() + scale\_fill\_gradientn(colors = c("red4","red","yellow","green","green4"), breaks = c(3,4.5,7), trans = "log10") + ggtitle("The World Happiness Scores in 2015") + ignore\_axes

## preparing the data for the year 2017   
  
mod\_dat2017 <- cbind(data2017)  
  
levels(mod\_dat2017$Country) <- c( levels(mod\_dat2017$Country), c("UK","USA","Palestine","Somalia") )  
mod\_dat2017[ which(mod\_dat2017[,1] == "United States") ,1] <- "USA"  
mod\_dat2017[ which(mod\_dat2017[,1] == "United Kingdom") ,1] <- "UK"  
mod\_dat2017[ which(mod\_dat2017[,1] == "Somaliland region") ,1] <- "Somalia"  
mod\_dat2017[ which(mod\_dat2017[,1] == "Palestinian Territories") ,1] <- "Palestine"  
  
country\_hpscore17 <- subset( mod\_dat2017, select = c("Country","Happiness Score") )  
names(country\_hpscore17) <- c("region","hpscore")  
  
world\_hpscore17 <- inner\_join(world\_map, country\_hpscore17, by = "region")  
  
map\_hp2017 <- map\_base + geom\_polygon( data = world\_hpscore17, aes(fill = hpscore), color = "white" ) + geom\_polygon(color = "black", fill = NA) + theme\_bw() + scale\_fill\_gradientn(colors = c("red4","red","yellow","green","green4"), breaks = c(3,4.5,7), trans = "log10") + ggtitle("The World Happiness Scores in 2015") + ignore\_axes

## preparing the data for the year 2018   
  
mod\_dat2018 <- cbind(data2018)  
  
levels(mod\_dat2018$Country) <- c( levels(mod\_dat2018$Country), c("UK","USA","Palestine","Somalia") )  
mod\_dat2018[ which(mod\_dat2018[,1] == "United States") ,1] <- "USA"  
mod\_dat2018[ which(mod\_dat2018[,1] == "United Kingdom") ,1] <- "UK"  
mod\_dat2018[ which(mod\_dat2018[,1] == "Somaliland region") ,1] <- "Somalia"  
mod\_dat2018[ which(mod\_dat2018[,1] == "Palestinian Territories") ,1] <- "Palestine"  
  
country\_hpscore18 <- subset( mod\_dat2018, select = c("Country","Happiness Score") )  
names(country\_hpscore18) <- c("region","hpscore")  
  
world\_hpscore18 <- inner\_join(world\_map, country\_hpscore18, by = "region")  
  
map\_hp2018 <- map\_base + geom\_polygon( data = world\_hpscore18, aes(fill = hpscore), color = "white" ) + geom\_polygon(color = "black", fill = NA) + theme\_bw() + scale\_fill\_gradientn(colors = c("red4","red","yellow","green","green4"), breaks = c(3,4.5,7), trans = "log10") + ggtitle("The World Happiness Scores in 2015") + ignore\_axes

## preparing the data for the year 2019   
  
mod\_dat2019 <- cbind(data2019)  
  
levels(mod\_dat2019$Country) <- c( levels(mod\_dat2019$Country), c("UK","USA","Palestine","Somalia") )  
mod\_dat2019[ which(mod\_dat2019[,1] == "United States") ,1] <- "USA"  
mod\_dat2019[ which(mod\_dat2019[,1] == "United Kingdom") ,1] <- "UK"  
mod\_dat2019[ which(mod\_dat2019[,1] == "Somaliland region") ,1] <- "Somalia"  
mod\_dat2019[ which(mod\_dat2019[,1] == "Palestinian Territories") ,1] <- "Palestine"  
  
country\_hpscore19 <- subset( mod\_dat2019, select = c("Country","Happiness Score") )  
names(country\_hpscore19) <- c("region","hpscore")  
  
world\_hpscore19 <- inner\_join(world\_map, country\_hpscore19, by = "region")  
  
map\_hp2019 <- map\_base + geom\_polygon( data = world\_hpscore19, aes(fill = hpscore), color = "white" ) + geom\_polygon(color = "black", fill = NA) + theme\_bw() + scale\_fill\_gradientn(colors = c("red4","red","yellow","green","green4"), breaks = c(3,4.5,7), trans = "log10") + ggtitle("The World Happiness Scores in 2015") + ignore\_axes

### Code for ‘Exploring Happiness Score based on Regions’

# AIM - Exploring Happiness Score for 2015 based on Regions. SCALE PLOT - You can see the countries individually, as you hover on the points.  
  
library(kableExtra)  
library(knitr)  
library(readr)  
library(plotly)  
X2015 <- read\_csv("data/2015.csv")  
happy\_2015 <- as.data.frame(X2015)  
  
#Changing column names  
names(happy\_2015) <- c("country", "region", "happiness\_rank", "happiness\_score", "std\_error", "gdp\_per\_cpt", "family", "life\_exp", "freedom", "trust\_corruption", "generosity", "dystopia\_residual")  
  
  
library(scales)  
box1 <- 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 for 2015") +   
 coord\_flip() +   
 theme(legend.position="none")  
ggplotly(box1)  
  
  
# AIM - Exploring Happiness Score for 2016 based on Regions. SCALE PLOT - You can see the countries individually, as you hover on the points.  
  
library(kableExtra)  
library(knitr)  
library(readr)  
X2016 <- read\_csv("data/2016.csv")  
happy\_2016 <- as.data.frame(X2016)  
  
#Changing column names  
names(happy\_2016) <- c("country", "region", "happiness\_rank", "happiness\_score", "std\_error", "gdp\_per\_cpt", "family", "life\_exp", "freedom", "trust\_corruption", "generosity", "dystopia\_residual")  
  
  
library(scales)  
box2 <- ggplot(happy\_2016, 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 for 2016") +   
 coord\_flip() +   
 theme(legend.position="none")  
ggplotly(box2)

### Code for ‘Exploring factors with Strong Correlations with Happiness Score’

X2015 <- read.csv("data/2015.csv")  
happy\_2015 <- as.data.frame(X2015)  
  
#Changing column names  
names(happy\_2015) <- c("country", "region", "happiness\_rank", "happiness\_score", "std\_error", "gdp\_per\_cpt", "family", "life\_exp", "freedom", "trust\_corruption", "generosity", "dystopia\_residual")  
  
cor(happy\_2015$happiness\_score, happy\_2015$gdp\_per\_cpt)  
  
ggplot(happy\_2015, aes(x=happy\_2015$gdp\_per\_cpt, y=happy\_2015$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)")  
  
X2016 <- read.csv("data/2016.csv")  
happy\_2016 <- as.data.frame(X2016)  
  
#Changing column names  
names(happy\_2016) <- c("country", "region", "happiness\_rank", "happiness\_score", "std\_error", "gdp\_per\_cpt", "family", "life\_exp", "freedom", "trust\_corruption", "generosity", "dystopia\_residual")  
  
cor(happy\_2015$happiness\_score, happy\_2015$life\_exp)  
  
ggplot(happy\_2015, aes(x=happy\_2015$life\_exp, y=happy\_2015$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)")

X2015 <- read.csv("data/2015.csv")  
happy\_2015 <- as.data.frame(X2015)  
  
#Changing column names  
names(happy\_2015) <- c("country", "region", "happiness\_rank", "happiness\_score", "std\_error", "gdp\_per\_cpt", "family", "life\_exp", "freedom", "trust\_corruption", "generosity", "dystopia\_residual")  
  
library(corrplot)  
  
num\_hap <- happy\_2015[, c("happiness\_rank", "happiness\_score", "std\_error", "gdp\_per\_cpt", "family", "life\_exp", "freedom", "trust\_corruption", "generosity", "dystopia\_residual")]  
  
m <- cor(num\_hap) #creating correlation matrix  
corrplot(m, method="circle", type='upper', tl.cex=0.8, tl.col = 'black')

### Code for ‘Regression Tree’

library(rpart)  
library(rpart.plot)  
  
X2015 <- read.csv("data/2015.csv")  
happy\_2015 <- as.data.frame(X2015)  
  
#Changing column names  
names(happy\_2015) <- c("country", "region", "happiness\_rank", "happiness\_score", "std\_error", "gdp\_per\_cpt", "family", "life\_exp", "freedom", "trust\_corruption", "generosity", "dystopia\_residual")  
  
#Removing happiness Rank  
happy\_2015\_tree <- happy\_2015[,c("country", "region", "happiness\_rank", "happiness\_score", "std\_error", "gdp\_per\_cpt", "family", "life\_exp", "freedom", "trust\_corruption", "generosity", "dystopia\_residual")]  
  
fit <- rpart(formula = happiness\_score~gdp\_per\_cpt+family+life\_exp-std\_error-trust\_corruption+  
 freedom+generosity-dystopia\_residual,   
 data = happy\_2015\_tree,  
 method="anova")   
  
#finding the optimal cp  
plotcp(fit)  
fit <- rpart(formula = happiness\_score~gdp\_per\_cpt+family+life\_exp-std\_error-trust\_corruption+  
 freedom+generosity-dystopia\_residual,   
 data = happy\_2015\_tree,  
 method="anova", #regression tree  
 control=rpart.control(cp=0.043)) #optimal cp value   
  
rpart.plot(fit, box.palette = c("blue","orange"), main="Regression Tree for Happiness Score")

### Code for ‘Correlation Plots’

library(gdata)  
library(dplyr)  
library(magrittr)  
library(ggplot)  
library(corrplot)  
library(xlsx)  
library(reshape)  
library(corrgram)  
library(plot3D)  
library(rpart.plot)  
library(ISLR)  
  
data1 <- read.csv("data/2015.csv", header = TRUE, sep=",")  
data2 <- read.csv("data/2016.csv", header = TRUE, sep=",")  
data3 <- read.csv("data/2017.csv", header = TRUE, sep=",")  
data4 <- read.csv("data/2018.csv", header = TRUE, sep=",")  
data5 <- read.csv("data/2019.csv", header = TRUE, sep=",")  
  
# Finding the correlation between numerical columns for 2015  
Num.cols <- sapply(data1, is.numeric)  
Cor.data <- cor(data1[, Num.cols])  
  
corrplot(Cor.data, method = 'color')  
  
# Finding the correlation between numerical columns for 2016  
Num.cols <- sapply(data2, is.numeric)  
Cor.data <- cor(data2[, Num.cols])  
  
corrplot(Cor.data, method = 'color')  
  
# Finding the correlation between numerical columns for 2017  
Num.cols <- sapply(data3, is.numeric)  
Cor.data <- cor(data3[, Num.cols])  
  
corrplot(Cor.data, method = 'color')  
  
# Finding the correlation between numerical columns for 2018  
Num.cols <- sapply(data3, is.numeric)  
Cor.data <- cor(data3[, Num.cols])  
  
corrplot(Cor.data, method = 'color')  
  
# Finding the correlation between numerical columns for 2019  
Num.cols <- sapply(data3, is.numeric)  
Cor.data <- cor(data3[, Num.cols])  
  
corrplot(Cor.data, method = 'color')  
  
# Create a correlation plot for 2015  
newdatacor = cor(data1[c(4:11)])  
corrplot(newdatacor, method = "number")  
  
# Create a correlation plot for 2016  
newdatacor = cor(data2[c(4:11)])  
corrplot(newdatacor, method = "number")  
  
# Create a correlation plot for 2017  
newdatacor = cor(data3[c(4:11)])  
corrplot(newdatacor, method = "number")  
  
# Create a correlation plot for 2018  
newdatacor = cor(data4[c(4:10)])  
corrplot(newdatacor, method = "number")  
  
# Create a correlation plot for 2019  
newdatacor = cor(data5[c(4:10)])  
corrplot(newdatacor, method = "number")

### Code for ‘Exploring the percentage change for the different factors of the World Happiness Report’

library(data.table)  
library(dplyr)  
library(formattable)  
library(tidyr)  
  
datat2015 <- read.csv("data/2015.csv", header = TRUE)  
datat2019 <- read.csv("data/2019.csv", header = TRUE)  
  
#merge(datat2015, datat2019, by = "Country")  
totalData <- merge(datat2015, datat2019, by = "Country")  
  
customGreen0="#DeF7E9"  
customGreen="#71CA97"  
customRed="#ff7f7f"  
  
improvement=round((totalData$Score-totalData$Happiness.Score)/totalData$Happiness.Score\*100,2)  
  
economychange=round((totalData$GDP.per.capita-totalData$Economy..GDP.per.Capita.)/totalData$Economy..GDP.per.Capita.\*100,2)  
socialchange=round((totalData$Social.support-totalData$Family)/totalData$Family\*100,2)  
lifechange=round((totalData$Healthy.life.expectancy-totalData$Health..Life.Expectancy.)/totalData$Health..Life.Expectancy.\*100,2)  
genechange=round((totalData$Generosity.y-totalData$Generosity.x)/totalData$Generosity.x\*100,2)  
freechange=round((totalData$Freedom.to.make.life.choices-totalData$Freedom)/totalData$Freedom\*100,2)  
corruchange=round((totalData$Perceptions.of.corruption-totalData$Trust..Government.Corruption.)/totalData$Trust..Government.Corruption.\*100,2)  
  
totalData<-cbind(totalData,economychange,socialchange,lifechange,genechange,freechange,corruchange)  
  
final2 = totalData[c(1,21,22,23,24,25,26)]  
final2  
formattable(final2)  
improvement\_formatter<-formatter("span",style=z~style(font.weight="bold",color=ifelse(z>0,customGreen,ifelse(z<0,customRed,"black"))),z~icontext(ifelse(z>0,"arrow-up","arrow-down"),z))  
formattable(final2,list('Country'=formatter("span",style=~style(color="grey",font.weight="bold")),'improvement'=improvement\_formatter,  
 'economychange'=improvement\_formatter,'socialchange'=improvement\_formatter,'lifechange'=improvement\_formatter,'genechange'=improvement\_formatter,'freechange'=improvement\_formatter,'corruchange'=improvement\_formatter))  
  
final2<-cbind(final2,improvement)