MA304 Coursework 3

Pirawas Maleemongkol :: Registration number: 1908015

Abstract

The World Happiness Report is a comprehensive output of a survey that deals with global happiness. It ranks a total of 156 countries of the world based on how happy their citizens say they are across 6 variables or features that contribute to the overall happiness - **Economy, Social support, Life expectancy, Freedom, Corruption, and Generosity**. The happiness scores of all countries are also compared with that of **Dystopia** (a hypothetical country with the least happiness score). This report is an analysis of the world happiness from 2015-2019 for 141 countries in which several comparisons of the happiness scores have been done taking into account the countries, regions, and the variables contributing to happiness over the span of 5 years.

Word Count:

Contents

Introduction

The World Happiness Report is produced by the United Nations Sustainable Development Solutions Network that compiles the data collected by Gallup's World Poll, which surveys people across 140+ countries to understand their lives. Answers to the questions in the survey are converted into scores for the six main variables - Economy, Social support, Life expectancy, Freedom, Corruption, and Generosity. The scores from these variables are then aggregated to form the Happiness Scores, where the country with the highest happiness score is said to be the happiest. The countries are then ranked based on these Happiness Scores - Rank 1 being the happiest country. All countries are also compared to Dystopia, an imaginary country having the world's least happy people, which also affects the Happiness Score of each country.

The Happiness Score lies between 0 and 10, with all 7 features (including Dystopia Residuals) adding up to this score. In this report, we have considered multiple analyses such as the average happiness score for each country from 2015-2019. We have observed the countries having the highest and the lowest happiness scores, corresponding scores for the 7 features, and also how their score has changed across five years. Comparisons have also been made between 10 regions in which all countries have been divided. Findings have been noted and explained in detail after each visualisation.

The World Happiness Data

The Happiness datasets from 2015-2019 is a collection of the answers to Gallup's Word Poll. Each of the datasets has a column for Country, Happiness Score, Happiness Rank, GDP per Capita (Economy), Family (Social Support), Health (Life Expectancy), Freedom, Generosity, and Trust in Government (Corruption). The data from 2015 and 2016 additionally classifies all countries into their respective Regions. Datasets of 2015, 2016 and 2017 also consider Dystopia Residuals, a column not found in the Happiness datasets of 2018 and 2019.

Country

Each year from 2015-2019 sees participation from at most 158 countries, out of which only 141 unique countries found in all the five datasets have been taken into consideration in this report.

Region

All countries considered in this report have been classified into a total of 10 Regions -
1. Western Europe
2. North America
3. Australia and New Zealand
4. Middle East and Northern Africa
5. Latin America and Caribbean
6. Southeastern Asia
7. Central and Eastern Europe
8. Eastern Asia
9. Sub-Saharan Africa
10. Southern Asia

Happiness Score

A score between 0 and 10 allotted to each country that contributed to Gallup's World Poll. This score is the sum of the individual scores of the 7 features listed in the dataset that contribute to happiness.

Happiness Rank

All the countries are given a rank based on their happiness scores - The country with the highest happiness score is assigned Rank 1, and so on.

GDP per capita (Economy)

This factor indicates how much a country's economy contributes to its overall happiness.

Social Support (Family)

The contribution of support from family and peers in the overall happiness.

Health (Life Expectancy)

How the life expectancy of each of the countries affects the total happiness.

Freedom

This variable measures the amount of freedom given to the citizens of each country to make their life choices.

Trust in Government (Corruption)

This feature tells us the amount of trust that a country's citizens have in their government.

Generosity

A factor used to understand how generous the people in a country perceive themselves to be.

Dystopia Residuals

As mentioned before, Dystopia is the imaginary country having the least happy people and hence, the least possible happiness score i.e. **1.88**. Dystopia serves as a benchmark against all the countries and no country is said to perform worse than Dystopia. Hence, it can also be implied that Dystopia performs the worst in the 6 main factors of happiness as well.
Residuals are the 'unknown components' in each country that may affect happiness. The average value of all the residuals is approximately 0.

Methods

Since all the five datasets analysed were not consistent with the variables and number of entries, our first approach was to clean the data. Out of about 156 countries found in the five datasets, only 141 unique countries were seen in all. The datasets were cleaned such that each of them contained the same 141 countries. Variables such as Region and Dystopia Residuals which were important for our analysis but were missing from certain datasets, were added to each of the datasets through simple if-conditions and for-loops. The columns containing Standard Error and Confidence Intervals were not taken into consideration for the purpose of this report.

A variety of libraries have been used to develop our visuals. The **ggplot2** library was used along with plotly for most of the visualisations in this report such as bar charts, bubble plots, boxplots and scatter plots. To produce the world map, we used the **rworldmap** library, and also **dplyr** for using the filter and join functions. Our scatterplot could be best represented in the form of a grid, for which the **grid** and **gridExtra** libraries were used. The scatterplot also shows the regression line to easily observe some outliers. Finally, the library **data.table** has been used for one of our table outputs.

Results

The overall happiness score can be visualised in a variety of ways. We begin by looking at a map of the world, highlighting the countries for which happiness scores have been calculated in our analysis.

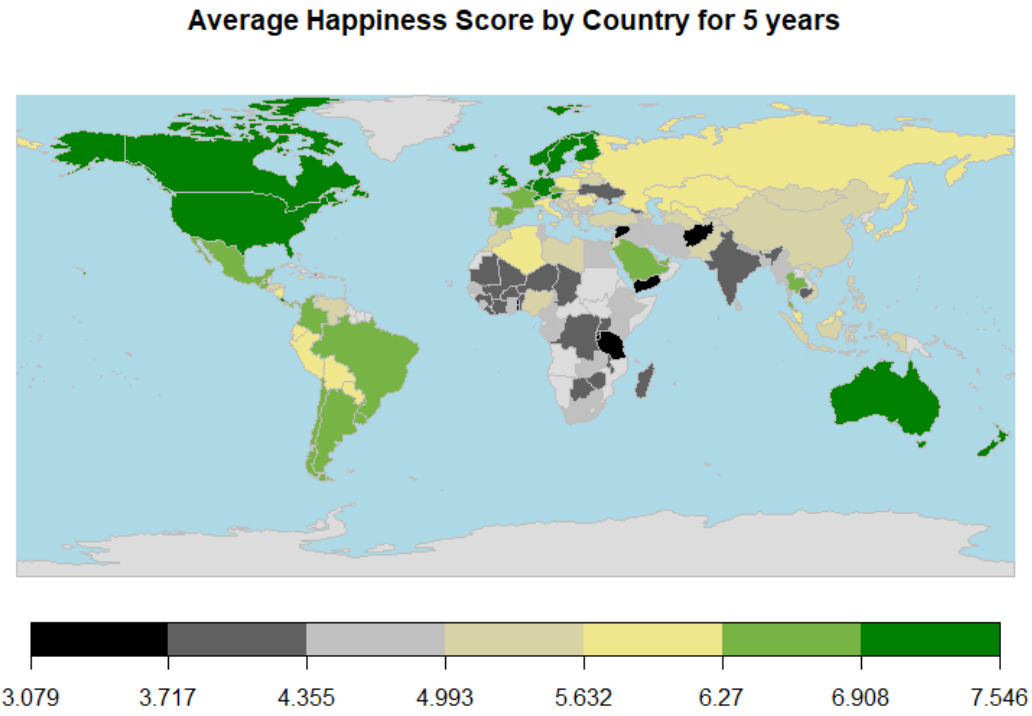

Figure 1: Happiness Scores by country (Average of 2015-2019)

This map displays the average happiness scores gathered for 141 countries from the year 2015 to 2019. The dark green portions are countries having the the highest happiness scores. The countries coloured in black are the ones that have the least happiness scores averaged over the five years. The countries in yellow show a moderate level of happiness and those in light grey are the ones for which we have no data.

It can also be observed that the Scandinavian countries, North America, United Kingdom, Iceland, Australia and New Zealand are generally the happiest countries. On the other hand, countries like Afghanistan, Tanzania, Yemen and some Sub-Saharan African countries have performed the worst in happiness in all five years.

This map gives us some idea about the overall happiness of the different regions of the world. Later in this

report, there will be a region-wise analysis that will give us some more understanding of how all the regions fair in the case of happiness.

Following is a feature-wise breakdown of the top and the bottom 18 countries with respect to the overall happiness score -

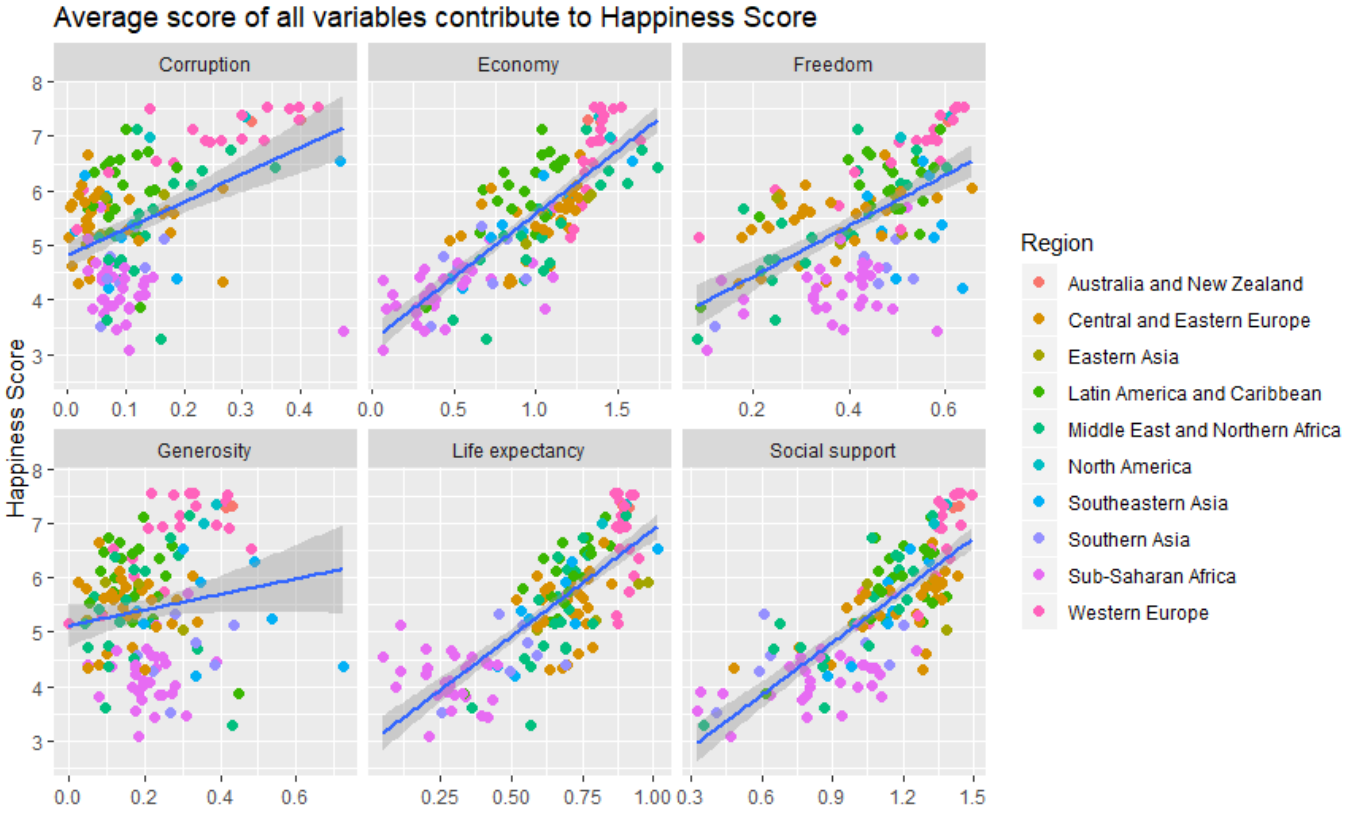

Figure 2: Distribution of Countries across six features of Happiness

In the above compilation of graphs, the y-axis shows Happiness Score and the x-axis takes six values, one for each contributing factor. Every country in the scatterplot has been classified into the 10 regions through a colour-code.

We again observe that Economy, Social Support, and Life Expectancy are the most significant contributors to the overall happiness score. For instance, larger the score for social support, more will be the happiness score.

We observe a trend in all these graphs showing Western European countries having the highest happiness scores no matter what variable we choose. On the other hand, countries belonging to the Sub-Saharan African region are constantly seen having the lowest scores in happiness.

However, the skewed observations in Corruption and Generosity are also notable. In Corruption, most countries except the ones in the Western Europe have a lesser score in Corruption. This theoretically means that these countries have lesser trust in their government. A noteworthy outlier however is Rwanda, which seems to enjoy the least corruption overall. Despite the low score in corruption, many countries from other regions such as Latin American, Caribbean, and Central and Eastern Europe have a high happiness score. A similar trend can be observed in the graph for Generosity where most regions have countries with low Generosity score, but high happiness score. One outlier in this graph is Myanmar, a Southern Asian country

which has the highest Generosity score but still couldn't achieve a high overall happiness score.

The citizens of Sub-Saharan African countries, however, do enjoy a little more than the moderate level of Freedom and Social Support but again fail to achieve a high happiness score.

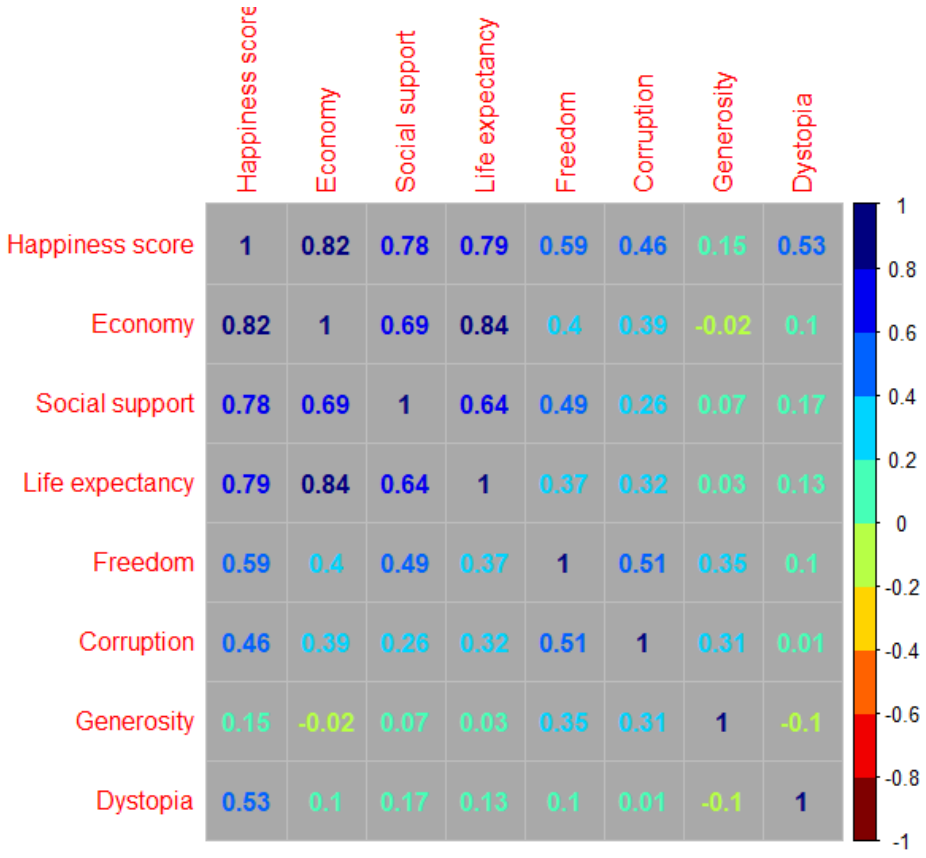

Figure 3: Correlation of Happiness score across all features

With all these peculiar observations in the analysis, we wanted to know the countries showing the most significant change in their Happiness Score from 2015 to 2019.

We can observe that the Economy variable play the most important role on happiness score following by life expectancy and social support of 0.82, 0.79, and 0.78 respectively.

The two table which Figure 4 and 5 show top 10 countries which have the most happiness score.

We can see that in 2015, Finland is placed in sixth with 7.406 of happiness score and finally in 2019, Finland finally can be the first rank of happiness score with happiness score of 7.769. And as we mentioned above that which variables play the most important role. Finland have increased in score of economy, social support and life expectancy significantly.By the way, another factors did not change that much.

Therefore, I would like to suggest to the government of those countries which have less happiness score to focus on developing of economy, life expectancy, and social support more for improving of happiness of the population in their own countries.

Country	Rank	Score	Economy	Social.support	life.expectancy	Freedom	Corruption	Generosity	Dystopia
Switzerland	1	7.587	1.39651	1.34951	0.94143	0.66557	0.41978	0.29678	2.51738
Iceland	2	7.561	1.30232	1.40223	0.94784	0.62877	0.14145	0.43630	2.70201
Denmark	3	7.527	1.32548	1.36058	0.87464	0.64938	0.48357	0.34139	2.49204
Norway	4	7.522	1.45900	1.33095	0.88521	0.66973	0.36503	0.34699	2.46531
Canada	5	7.427	1.32629	1.32261	0.90563	0.63297	0.32957	0.45811	2.45176
Finland	6	7.406	1.29025	1.31826	0.88911	0.64169	0.41372	0.23351	2.61955
Netherlands	7	7.378	1.32944	1.28017	0.89284	0.61576	0.31814	0.47610	2.46570
Sweden	8	7.364	1.33171	1.28907	0.91087	0.65980	0.43844	0.36262	2.37119
New Zealand	9	7.286	1.25018	1.31967	0.90837	0.63938	0.42922	0.47501	2.26425
Australia	10	7.284	1.33358	1.30923	0.93156	0.65124	0.35637	0.43562	2.26646

Figure 4: Top 10 countries of happiness score in 2015

Country	Rank	Score	Economy	Social.support	life.expectancy	Freedom	Generosity	Corruption	Dystopia
Finland	1	7.769	1.340	1.587	0.986	0.596	0.153	0.393	2.714
Denmark	2	7.600	1.383	1.573	0.996	0.592	0.252	0.410	2.394
Norway	3	7.554	1.488	1.582	1.028	0.603	0.271	0.341	2.241
Iceland	4	7.494	1.380	1.624	1.026	0.591	0.354	0.118	2.401
Netherlands	5	7.488	1.396	1.522	0.999	0.557	0.322	0.298	2.394
Switzerland	6	7.480	1.452	1.526	1.052	0.572	0.263	0.343	2.272
Sweden	7	7.343	1.387	1.487	1.009	0.574	0.267	0.373	2.246
New Zealand	8	7.307	1.303	1.557	1.026	0.585	0.330	0.380	2.126
Canada	9	7.278	1.365	1.505	1.039	0.584	0.285	0.308	2.192
Austria	10	7.246	1.376	1.475	1.016	0.532	0.244	0.226	2.377

Figure 5: Top 10 countries of happiness score in 2019

Discussion

The analysis can be concluded saying that certain countries or regions perform very well in some of the factors. However, that does not always majorly affect their respective happiness score. We constantly see the introduction of Dystopia Residuals as a seventh feature in the aggregation of the happiness score, one which highly contributes to the score. There still is not much information regarding how the residuals have been taken into consideration. This limits our analysis to only the remaining six factors.
Some research literature also states other important factors that could contribute to happiness, but not much data is available for all the years and countries in order for them to be counted among the important

variables affecting the overall happiness.

Appendix

Shiny App

This responsive shiny app has two sections - the first section allows us to see the happiness of all 141 countries in our dataset plotted over a world map. The color code indicates the country's happiness. Filters can be applied to see only the happiest countries, moderately happy countries, or countries with low happiness score. Happiness can also be seen for each year over this world map or the average happiness score over the course of 5 years as well. In the second section we can see what is the actual contribution of a single feature as a percentage to the happiness score of a particular country for all 5 years. You can select a feature and a country to see how much of an impact this feature is making to the contribution of happiness of a particular country. Link to the live app is given below -
https://worldhapinessreport.shinyapps.io/shiny_code/

Scatterplot: Figure 2

combine the rows all 5 years df1617 = rbind(region15,region16,region17,region18,region19) **calculate the mean in each column by country** avghap2 = aggregate(df1617[,c(3,4,5,6,7,8,9,10)],list(df1617$Country$), mean)$*$
$*renamethecolumns**colnames(avghap2) = c("Country", "Score", "Economy", "Social.support", "life.expectancy", "Free$
$*mergetheaverage5yrsdataframewithregion**regionavg = merge(as.data.frame(avghap2), as.data.frame(dfregion), by =$
$Country', all = FALSE)**rearrangethecoloumnorder**regionavg2 = regionavg[, c(1, 10, 2, 3, 4, 5, 6, 7, 8, 9)]*$
$*scatterploteveryvariablewithHappinessScoreinaverage5years * *multig < -data.frame(Happys =$
$regionavg2$Score, independent.variable = c(regionavg2$Economy, regionavg2$Social.support,regionavg2$life.expectancy, regio$
Region = rep(regionavg2$Region,times=6), Variable.name = rep(c("Economy","Social support","Life expectancy","Freedom","Corruption","Generosity"),each=nrow(regionavg2))) ggplot(multig,aes(independent.variable,Happys)
+ geom_point(aes(colour=Region),size=2) + facet_wrap(~Variable.name,scales="free_x") + xlab("")+geom_smooth(meth
="lm")+ylab("Happiness Score")+labs(title ="Average score of all variables contribute to Happiness Score")

World Map Figure 1

create data frame of all rows with Country and Score columns ddf <- subset(df15[c("Country","Score")]) ddf16 <- subset(df16[c("Country","Score")]) ddf17 <- subset(df17[c("Country","Score")]) ddf18 <- subset(df18[c("Country","Score")]) ddf19 <- subset(df19[c("Country","Score")]) **combine all row for 5 years** ddf1617 = rbind(ddf,ddf16,ddf17,ddf18,ddf19) **calculate mean all column by country** avghap = aggregate(ddf1617[,2],list(ddf1617$Country),mean) **change the column names** colnames(avghap) = c("Country","Score")

set columns to years years = c(2015,2016,2017,2018,2019,"avg") **mapping the name of countries to the world map** spdf15 <- joinCountryData2Map(ddf, joinCode="NAME", nameJoinColumn="Country", verbose = FALSE) spdf16 <- joinCountryData2Map(ddf16, joinCode="NAME", nameJoinColumn="Country", verbose = FALSE) spdf17 <- joinCountryData2Map(ddf17, joinCode="NAME", nameJoinColumn="Country", verbose = FALSE) spdf18 <- joinCountryData2Map(ddf18, joinCode="NAME", nameJoinColumn="Country", verbose = FALSE) spdf19 <- joinCountryData2Map(ddf19, joinCode="NAME", nameJoinColumn="Country", verbose = FALSE) spdfavg <- joinCountryData2Map(avghap, joinCode="NAME", nameJoinColumn="Country", verbose = FALSE)

demonstrating the happiness score by countries in average five years which green is the most happiness score and grey is the least. avgmap = mapCountryData(spdfavg, nameColumnToPlot="Score", addLegend = FALSE,catMethod="fixedWidth",mapTitle = "Average Happiness Score by Country for 5 years", colourPalette = c('#000000','#C0C0C0','#F0E68C','#008000'),missingCountryCol = '#DCDCDC',oceanCol="lightblue", mapRegion='world') do.call(addMapLegend,c(avgmap,labelFontSize=1 ,legendLabels = "all",digits = 4))

Table of 10 top happiness score of year 2015 and 2019

reorder column to be the same as year 2015 df19 = df19[,c(2,1,3,4,5,6,7,8,9,10)] **make top 10 table 2015 and 2019** df15 %>% head(10) %>% formattable(list(Rank = color_bar("yellow"), Score = color_bar("lightgreen"), Economy = color_bar("deepskyblue"), Social.support = color_bar("deepskyblue"), life.expectancy = color_bar("deepskyblue"), Freedom = color_bar("deepskyblue"), Corruption = color_bar("deepskyblue"), Generosity = color_bar("deepskyblue"), Dystopia = color_bar("deepskyblue")), align = "l") df19 %>% head(10) %>% formattable(list(Rank = color_bar("yellow"), Score = color_bar("lightgreen"), Economy = color_bar("deepskyblue"), Social.support = color_bar("deepskyblue"), life.expectancy = color_bar("deepskyblue"), Freedom = color_bar("deepskyblue"), Corruption = color_bar("deepskyblue"), Generosity = color_bar("deepskyblue"), Dystopia = color_bar("deepskyblue")),align = "l")