MA304 Coursework 2

Registration numbers -

1701077
1900881
1907907
1908015
1900906

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: 2344

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.

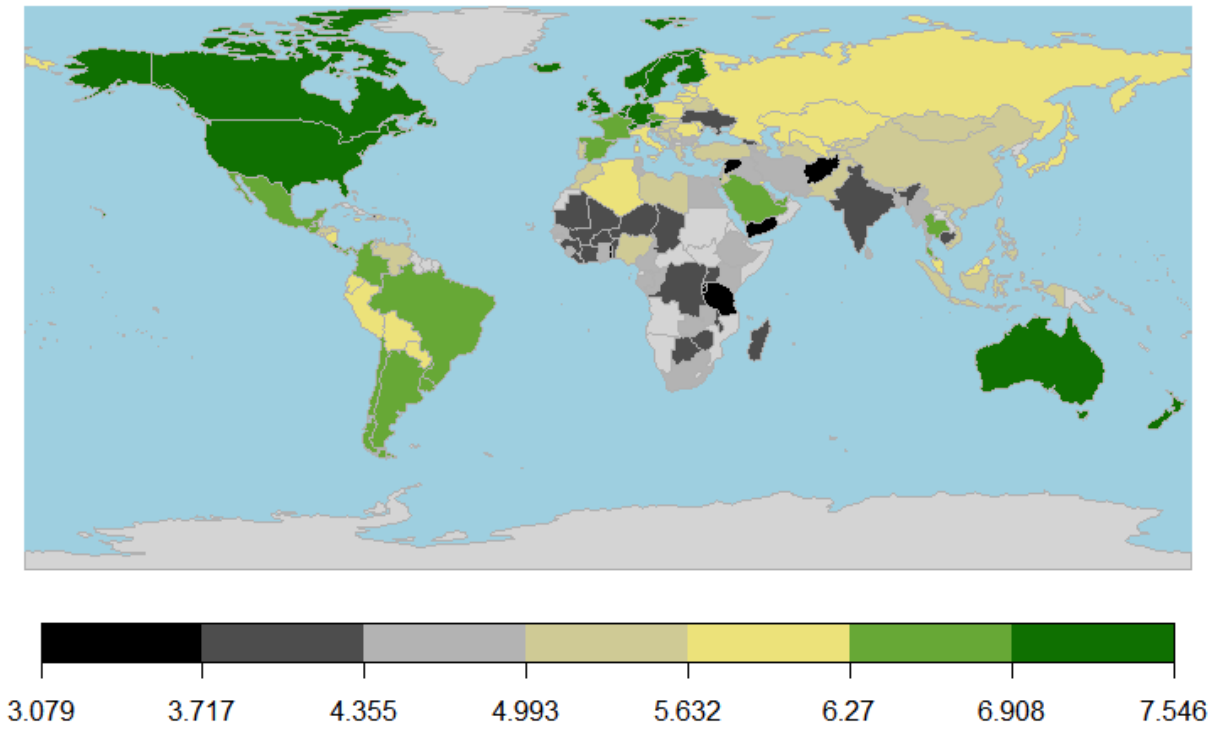

Figure 1: Happiness Scores by country (Average of 2015-2019)

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 -

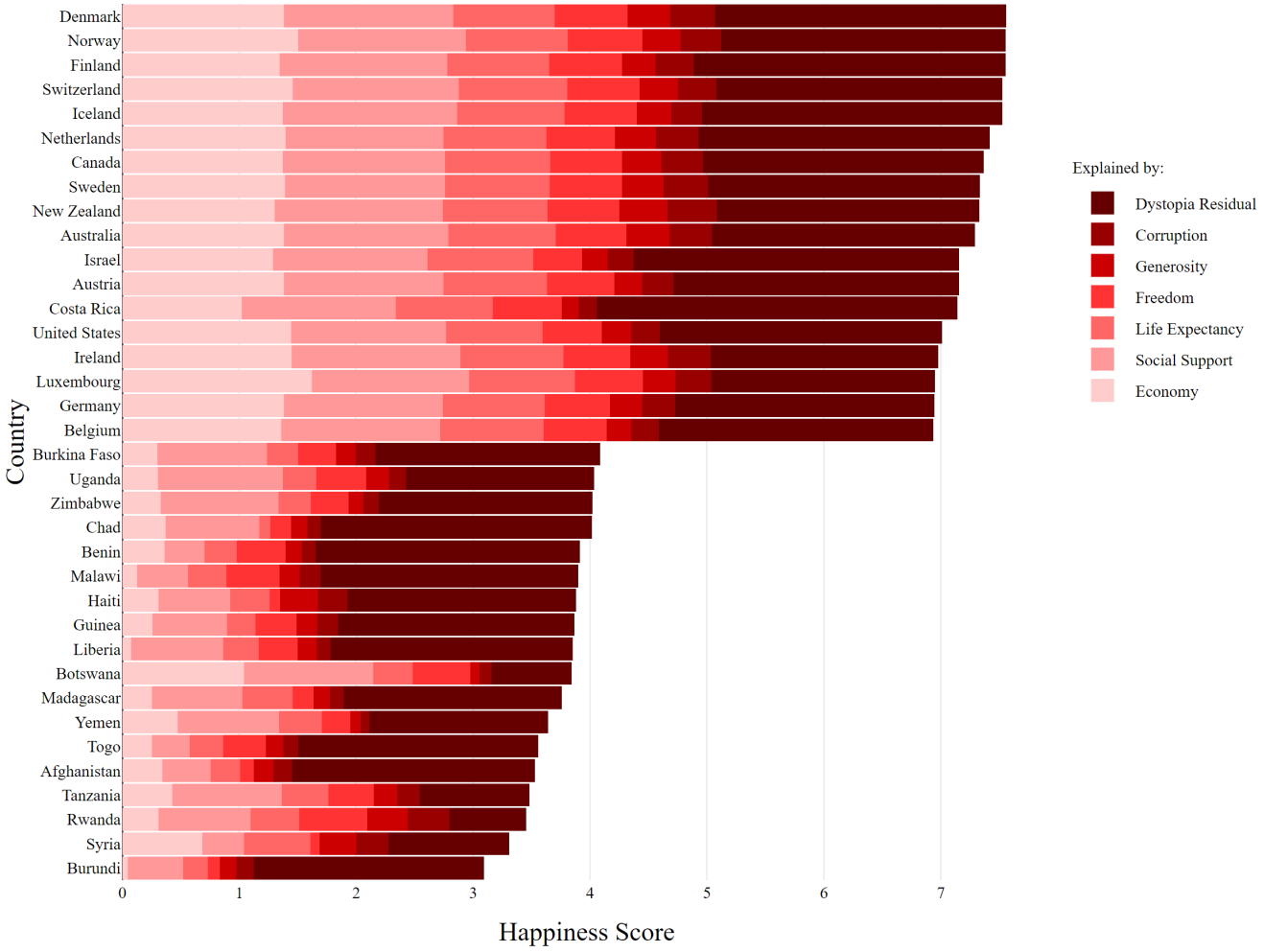

Figure 2: Ranking of Happiness from 2015-2019 (Top and bottom 18)

Each colour in the bar shows the proportion of the happiness score which is explained by the corresponding variable. It is evident that, in most cases, Dystopia Residual is the largest contributing factor. This applies both to the top and bottom ranking countries. We can see that the second and third largest contributing factors are Economy and Social Support, after which the remaining factors contribute significantly less. The graph shows that there is a strong trend between all countries' score composition, with the exception of a few countries (such as Botswana and Rwanda), where Dystopia Residual contributes a far smaller percentage. The happiest country (on average) across the years 2015-19 was Denmark, with 4 of the top 5 countries being Scandinavian. The least happy country was Burundi, with 14 of the bottom 18 countries being African. The remaining members of the bottom 18 are from Asia (Syria, Afghanistan, Yemen) and the Caribbean (Haiti).

This brings us to our next concern - **how have the different regions performed on an average with respect to the six key factors contributing to happiness?**

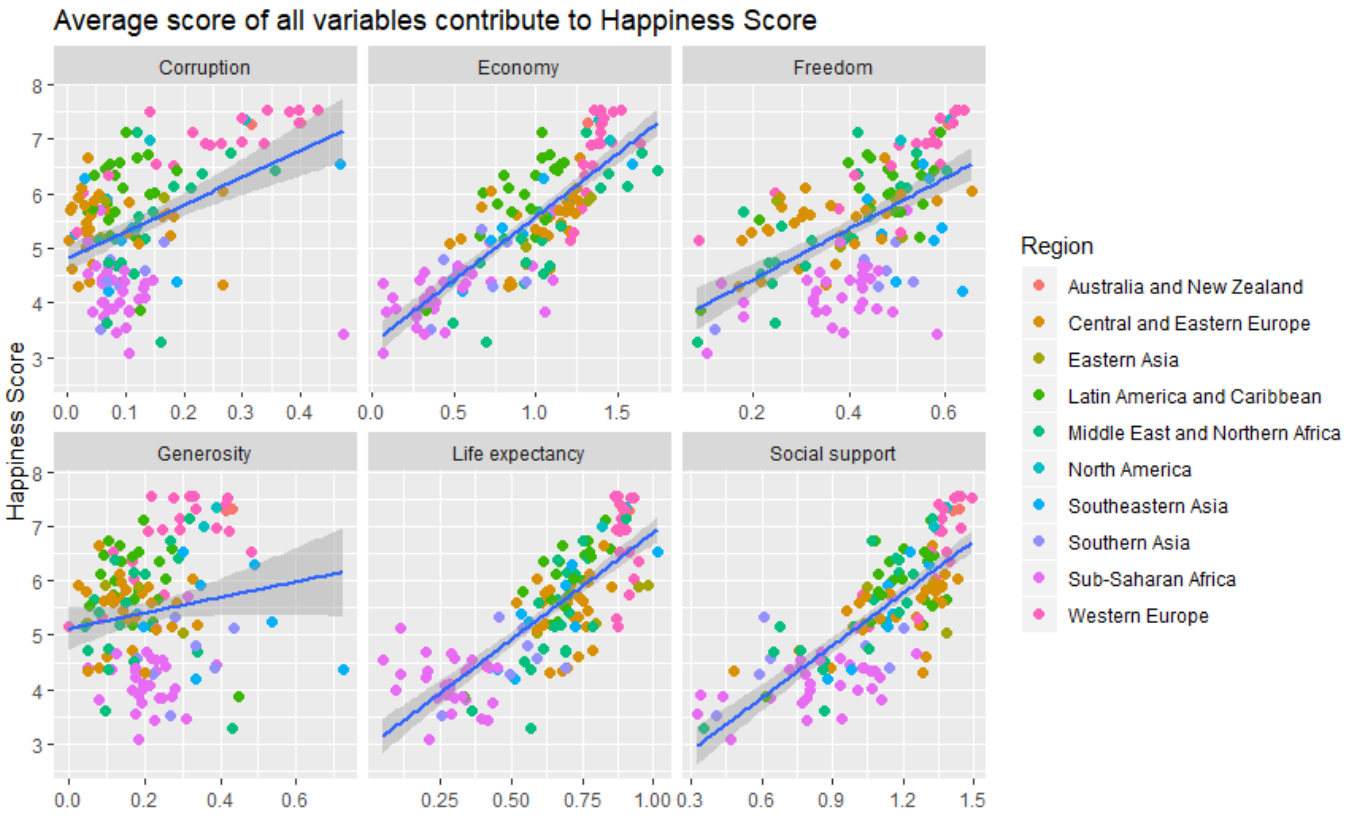

Figure 3: Distribution of Countries across six features of Happiness

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.

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.

Country	Happiness Score	Economy	Social Support	Life Expectancy	Generosity	Freedom	Corruption	Dystopia Residual
Benin	46.2	37.1	23.5	24.41	-4.16	-27.97	2.37	86.74
Togo	43.89	31.78	308.72	44.15	6.11	-19.62	-20.79	45.03
Ivory Coast	35.27	22.28	4.78	52.78	-23.63	-24.89	-49.78	93.26
Burundi	29.95	200.65	7.49	69.67	-10.78	85.65	78.89	26.89
Burkina Faso	27.88	28.23	23.96	40.09	-18.61	-35.43	-11.94	55.3
Guinea	24.02	118.18	78.38	56.19	-27.77	-11.99	-29.15	16.73
Gabon	23.18	-0.31	30.68	31.65	-36.97	-7.56	-50.41	59.67
Cambodia	23.07	24.68	78.84	4.23	-42.52	-8.07	-14.45	49.09
Honduras	22.39	7.84	29.63	19.12	6.83	26.28	14.29	25.97
Congo (Brazzaville)	20.63	-0.83	20.53	63.6	-15.24	-10.29	-20.42	34.53
India	-12.05	17.06	100.4	14.11	-24.46	25.17	0.09	-50.6
Liberia	-13.04	2.53	16.76	29.53	-4.36	29.68	-47.05	-31.55
Tanzania	-14.55	66.9	-11.74	30.58	-19.71	26.83	155.79	-61.54
Yemen	-17.1	-47.48	70.8	15.57	18.28	-59.8	-1.96	-40.77
Botswana	-19.48	4.78	3.65	1026.47	-76.1	-8.07	-19.83	-87.41
Zambia	-19.93	22.88	15.49	42.36	26.08	-11.73	-30.22	-51.41
Haiti	-20.39	21.1	-7.4	15.58	-9.28	-89.36	-35.95	-29.43
Zimbabwe	-20.54	35.06	7.87	29.35	-20.47	39.59	10.16	-52.95
Malawi	-20.55	1090.77	36.14	119.4	-34.19	2.89	27.56	-49.64
Venezuela	-30.88	-8.07	13.62	11.72	9.57	-64.11	-57.54	-60.83

Figure 4: Most Significant Percentage Change in Happiness Score from 2015-2019

Out of the 141 countries, 66 countries experienced decrease in the happiness score from 2015 to 2019 while 75 countries showed improvement. The top 10 countries which have the maximum percentage change in their happiness through the five years have been considered.

There is a common trend across all 20 countries in a few features. The factors responsible for happiness like Economy, Family and Life Expectancy are all on the rise in almost all the countries. This could point towards the fact that people are starting to focus on the basic necessities of life. However, Yemen shows a drastic decrease in the Economy score. This change can be attributed to the prevailing Humanitarian crisis in the country. As for Generosity, Freedom, and Corruption, the trend is alarming. More than 65% of the countries can be seen having major decrease in scores of those variables. It could be concluded that the people are becoming more self-centric, they feel that their freedom is being taken away, and that they are losing trust in their government.

Another interesting observation is that Dystopia residual is directly proportional to the Happiness Score of the countries. It can be implied that the countries with increase in Dystopia Residual score have an increase in the Happiness Score and a similar case with the decrease in Dystopia Residuals.

This problem can be better visualised with the following plots -

On the x-axis we have the contributing factors and on y-axis we have the happiness score. The size of the bubble signifies the percentage change in happiness score and its colour shows either an improvement or a deterioration of happiness score.

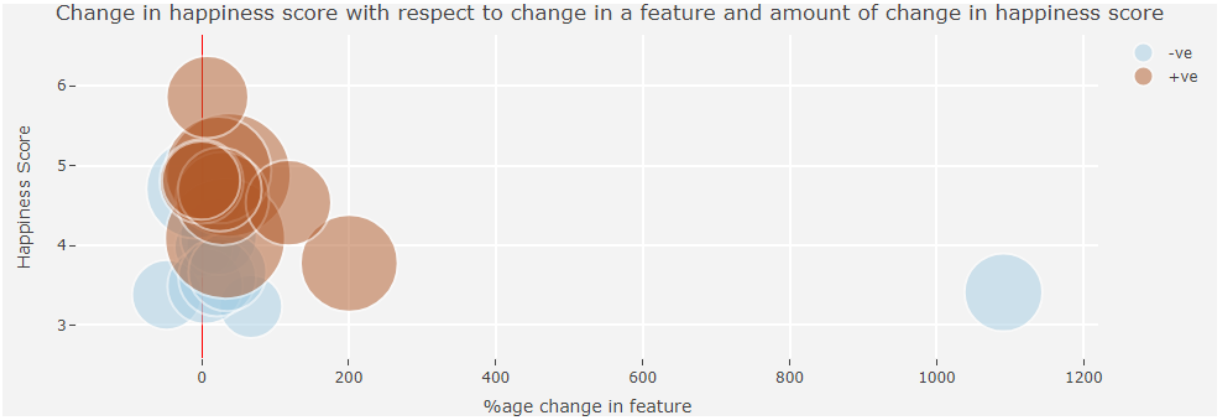

Figure 5: Percentage change in Economy and corresponding Happiness Score

The bubble seen on the right of the graph is that of Malawi. It is seen having more than 1000% increase in its score for Economy. Despite such a tremendous increase, the happiness score of this country has decreased over the years. The brown bubble at the top of the graph shows Honduras, which has seen an increased happiness over the years, even though there is negligible increment in its corresponding score for Economy. Therefore, increase in Economy score does not contribute enough to increase the happiness of this country.

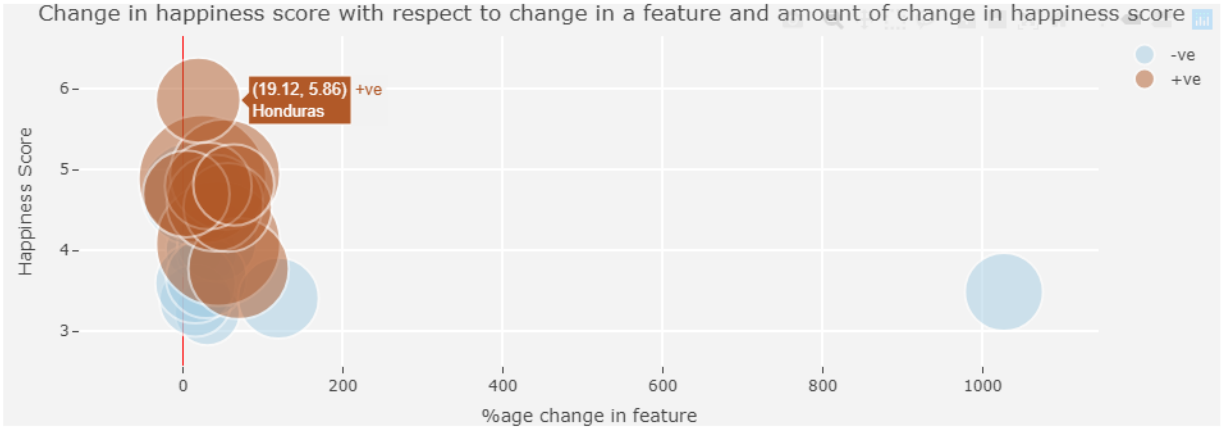

Figure 6: Percentage change in Life Expectancy and corresponding Happiness Score

Similarly, we can consider Botswana. Even with a huge increment in the life expectancy (more than 1000%) of this country, its happiness score decreased by 19.48% over five years.

And finally after considering all these aspects, we wished to know where the overall world happiness was moving towards -

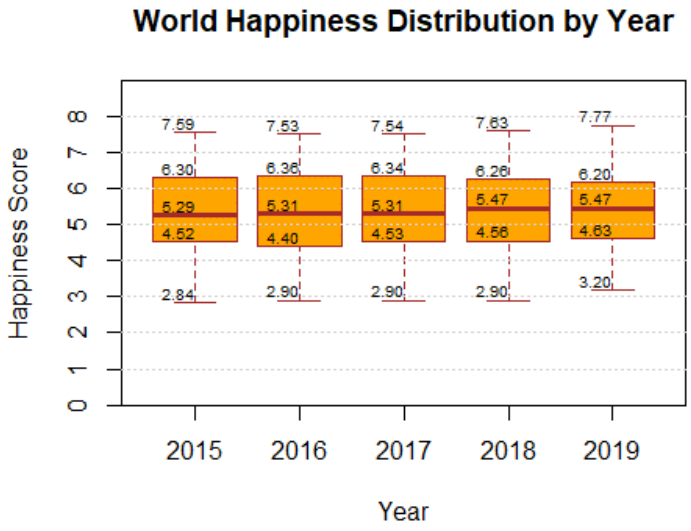

Figure 7: World Happiness Distribution by Year

It can be seen that the minimum, maximum and median value has increased from 2015 to 2019, as well as the interquartile range has become smaller. This shows that overall world happiness has increased for both the happiest and least happy countries. The minimum, maximum and median values only either increased or remain constant between the 2015-19 period- these values do not decrease between any of the years.

The inter-quartile range has become smaller each year since 2016, suggesting that happiness scores are becoming more similar, with more scores residing in a smaller range. It can be inferred from this analysis that the equality of world happiness is increasing.

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.

Individual Contribution

1701077 - R code and analysis of Figure 2 and Figure 7
1900881 - R code and analysis of Figure 5 and Figure 6, Shiny App (Appendix)
1907907 - R code and analysis of Figure 4
1908015 - R code and analysis of Figure 1 and Figure 3
1900906 - Research on related literature, conceptualisation of the report, and compilation of Rmd document

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/

Stacked bar graph: Figure 2

y <- Country p <- plot_ly(Top_Bottom_18, x = ~Economy, y = ~y, type = 'bar', orientation = 'h', name = 'Economy', marker = list(color = 'FFCCCC', line = list(color = 'FFCCCC', width = 2))) %>%

add_trace(x = ~Social.support, name = 'Social Support', marker = list(color = 'FF9999', line = list(color = 'FF9999', width = 2))) %>%

add_trace(x = ~life.expectancy, name = 'Life Expectancy', marker = list(color = 'FF6666', line = list(color = 'FF6666', width = 2))) %>%

add_trace(x = ~Freedom, name = 'Freedom', marker = list(color = 'FF3333', line = list(color = 'FF3333', width = 2))) %>%

add_trace(x = ~Generosity, name = 'Generosity', marker = list(color = 'CC0000', line = list(color = 'CC0000', width = 2))) %>%

add_trace(x = ~Corruption, name = 'Corruption', marker = list(color = '990000', line = list(color = '990000', width = 2))) %>%

add_trace(x = ~Dystopia, name = 'Dystopia Residual', marker = list(color = '660000', line = list(color = '660000', width = 2))) %>%

add_annotations(text="Explained by:", xref="paper", yref="paper", x=1.02, xanchor="left", y=0.8, yanchor="bottom",
legendtitle=TRUE, showarrow=FALSE) %>%

#Create correct formatting (N.B. x/y axes are flipped) layout(barmode = 'stack', #title = 'Average Composition of Happiness Scores (2015-19) for top 18 and bottom 18 Countries', font = textfont, xaxis = list(title = "Happiness Score", titlefont = list(size = 16)), yaxis = list(title ="Country", titlefont = list(size = 16)), legend = list(y=0.8, yanchor="top"), autosize = T)

print(p)

Scatterplot: Figure 3

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)$*
$*rename the columns**colnames(avghap2) = c("Country", "Score", "Economy", "Social.support", "life.expectancy", "Free$
$*merge the average 5yrs data frame with region**regionavg = merge(as.data.frame(avghap2), as.data.frame(dfregion), by =$
$Country', all = FALSE)**rearrange the coloumn order**regionavg2 = regionavg[, c(1, 10, 2, 3, 4, 5, 6, 7, 8, 9)]*$
$*scatterplot every variable with Happiness Score in average 5 years * *multig < -data.frame(Happys =$
$regionavg2$Score, independent.variable = c(regionavg2$Economy, regionavg2$Social.support,regionavg2$life.expectancy, regiona$
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(method ="lm")+ylab("Happiness Score")+labs(title ="Average score of all variables contribute to Happiness Score")

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")

scatterplot set the name of button for shiny 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)

Boxplot: Figure 7

df6 <- rbind(df1, df2, df3, df4, df5)

library(reshape2)
data_melted <- melt(df3, id.var="condition")

library(ggplot2)

box <- boxplot(df6$Score~condition,data=df6, yaxt="n", main="World Happiness Distribution by Year", xlab="Year", ylab="Happiness Score", col = "orange",
border = "brown", ylim=c(0,9) ,xaxs="i", yaxs="i")

Create custom y-axis
ticks<-seq(from = 0, to = 8, by = 1)
axis(2,at=ticks,labels=ticks)

Add dotted horizontal lines
grid(nx = 0, ny = 9, col = "lightgray", lty = "dotted", lwd = par("lwd"), equilogs = TRUE)

Add numbered labels for box statistics
text(x = col(box$stats)$-0.15, y = box$stats + 0.27, labels = format(round(box$stats, 2), nsmall = 2), cex=0.60, col = 'black')