2021-22 Semester B

GE2343/SDSC2004 Data Visualization Group Project

Topic: Happiness Score Analysis and Visualization

Lecturer*: Prof. Zhang Qingpeng*

Tutorial Class: T03 (CRN: 14379)

Word Count: 2974 words

Team Name: Data Killers

Written By Following Members:		
HA Quang Minh (57040174)	FOR Lek Shyuen (57038431)	DZHUNUSHOVA Naima (57109237)
Liu Ying Wo, Raymond (54818325)	Yu Wing Yin, Cherry (57147137)	Fung King Pui, Derek (56681127)

Presentation Link Data Visualization Project_Group Data Killers

Presentation Recorded Video

https://drive.google.com/file/d/1XrjEqGqC334ct4SIRqNI5Jdp0Q4wUa6D/view?usp=sharing

Table of Contents

1. Introduction

The Happiness Score becomes an important statistic for assessing citizens' levels of happiness in all the countries. Residents report feeling happier or more pleased when their score is high or climbing, and vice versa. This score is influenced by a range of characteristics, including economic, social, and health-related variables.

This report will study the World Happiness Report in conjunction with the United Nations high-level meeting on happiness, which listed 146 countries. The purpose of this paper is to study the underlying reasons for affecting the happiness score and find out if there is any correlation between the variables and the countries' happiness scores.

Research Objectives

- Discover the degree of correlation relationships between the happiness score and provided factors in the dataset Happiness World Report 2022, rank these categories, and find the most significant factors affecting the happiness score.
- Discover the level of relationships and correlation between the happiness score and extra factors added (found by external websites).
- Determine the top-ranked continents in terms of happiness rank, score, and its associated factors.
- Conduct multiple regression analysis between the happiness score and the associated variables and then determine the optimal model for the study.

2. Data & Methodology

Our team will use data of major countries contained in the happiness score datasets while some extra factors will be added for analyzing the level of correlation and relationship between happiness and different external datasets.

2.1 Data Understanding

- Happiness score: Polls assessing residents' assessments of their quality of life yield the happiness score. It is based on the results of a Cantril ladder survey as judged by inhabitants in each country and includes the seven elements stated below.

Provided Main Factors:

- *Dystopia*: A benchmark for the most unequal society or country with the lowest living standard on average, which was equal to 1.88 between 2016 and 2018 using each country's residual values.
- *Gross Domestic Product(GDP) per capita*: Interpreting a country's living standards and economic well-being.
- *Social support*: This index measures the network of resources available to respondents in times of need. Physical and mental health problems are associated with risk factors such as social isolation and loneliness.
- *Healthy life expectancy (HALE)*: Calculates the average number of years lived in a healthy state of mind using national mortality statistics. The HALE at birth was retrieved for analysis in this study.
- *Freedom to make life choices*: National average of binary responses to people's autonomy and ability to make and act on their own volitional decisions and behaviors.
- *Generosity*: Takes into account respondents' charitable giving and volunteering actions, which are then regressed on GDP per capita to determine the level of generosity.
- *Perceptions of corruption*: Corruption perceptions delve into the two facets of corruption: government and business. Respondents were asked to offer binary responses to questions about the prevalence of corruption in government and business.

Extra factors:

- *COVID-19 cases and deaths*: cumulative numbers of coronavirus confirmed cases and deaths.
- *Crime Index*: The degree of crime in a particular country or city, as determined by a crowd-sourced database called Numbeo. How likely it is that residents feel positively or negatively about the city's crime level.
- *Suicide rate (by country)*: Calculated the number of suicide cases per 100,000 populations.
- *Unemployment rate*: Assess the economic performance and well-being of a country.
- *Inflation rate*: A country's purchasing power or the purchasing value of a currency over
- *Density*: Population concentration by country. A country's population density indicates how concentrated its citizens are.

2.2 Data Preparation

Data Importing

- World Happiness Report 2022 from Kaggle.
- COVID-19 cases and deaths from Our World in Data
- Crime Index from NUMBEO
- Suicide rate from World Population Review
- Unemployment rate from TRADING ECONOMICS
- Inflation rate from TRADING ECONOMICS
- Population density from Our World in Data

Data Cleaning

Remove two attributes whisker-low and whisker-high of World Happiness Report 2022, which are superfluous to our study.

The COVID-19 and population density datasets collected from Our World in Data are clean, consistent, and don't have any missing values.

There are 20 missing values in the Crime index dataset, but it is unable to complete them because there is little information available.

Complete the missing values of the datasets of the unemployment rate, suicide rate, and inflation rate by finding their old data from the previous years, then using the Moving-average model to predict their values in 2021 and 2022.

Data Integration & Reduction

Due to missing many attribute values, 8 countries will be removed: Kosovo, Turkmenistan, Ivory Coast, Congo, Palestine Territories, Eswatini, and Northern Cyprus.

Data Transformation

```
'data.frame':   139 obs. of  18 variables:
 $ ï..Country                  : Factor w/ 139 levels "Afghanistan",..: 39 32 52 120 91 74 119 97 58 92 ...
 $ Continent                   : Factor w/ 6 levels "Africa","Asia",..: 3 3 3 3 3 3 3 3 2 5 ...
 $ RANK                        : int  1 2 3 4 5 6 7 8 9 10 ...
 $ Happiness.score             : num  7.82 7.64 7.56 7.51 7.42 ...
 $ Dystopia..1.83....residual  : num  2.52 2.23 2.32 2.15 2.14 ...
 $ GDP.per.capita              : num  1.89 1.95 1.94 2.03 1.95 ...
 $ Social.support              : num  1.26 1.24 1.32 1.23 1.21 ...
 $ Healthy.life.expectancy     : num  0.775 0.777 0.803 0.822 0.787 0.79 0.803 0.786 0.818 0.752 ...
 $ Freedom.to.make.life.choices: num  0.736 0.719 0.718 0.677 0.651 0.7 0.724 0.728 0.568 0.68 ...
 $ Generosity                  : num  0.109 0.188 0.27 0.147 0.271 0.12 0.218 0.217 0.155 0.245 ...
 $ Perceptions.of.corruption   : num  0.534 0.532 0.191 0.461 0.419 0.388 0.512 0.474 0.143 0.483 ...
 $ COVID.19.cases              : int  859477 3045935 179813 3453885 7869196 211280 2481736 1399714 3876115 627898 ...
 $ COVID.19.deaths             : int  3054 5585 97 13524 22015 1032 18189 2339 10485 266 ...
 $ Crime.index                 : num  27.2 26.6 23.5 21.7 27.9 ...
 $ Suicide.rate                : num  15.3 10.7 11.9 14.5 11.8 11.3 14.7 11.8 5.3 11 ...
 $ Unemployment.rate           : num  6.7 2.5 4.5 2.5 3.4 4.9 7.9 3.4 3.9 3.2 ...
 $ Inflation.rate              : num  4.5 4.8 6.7 2.4 6.2 6.6 4.3 3.7 3.5 5.9 ...
 $ Density                     : num  18.26 145.33 3.41 220.56 510.04 ...
```

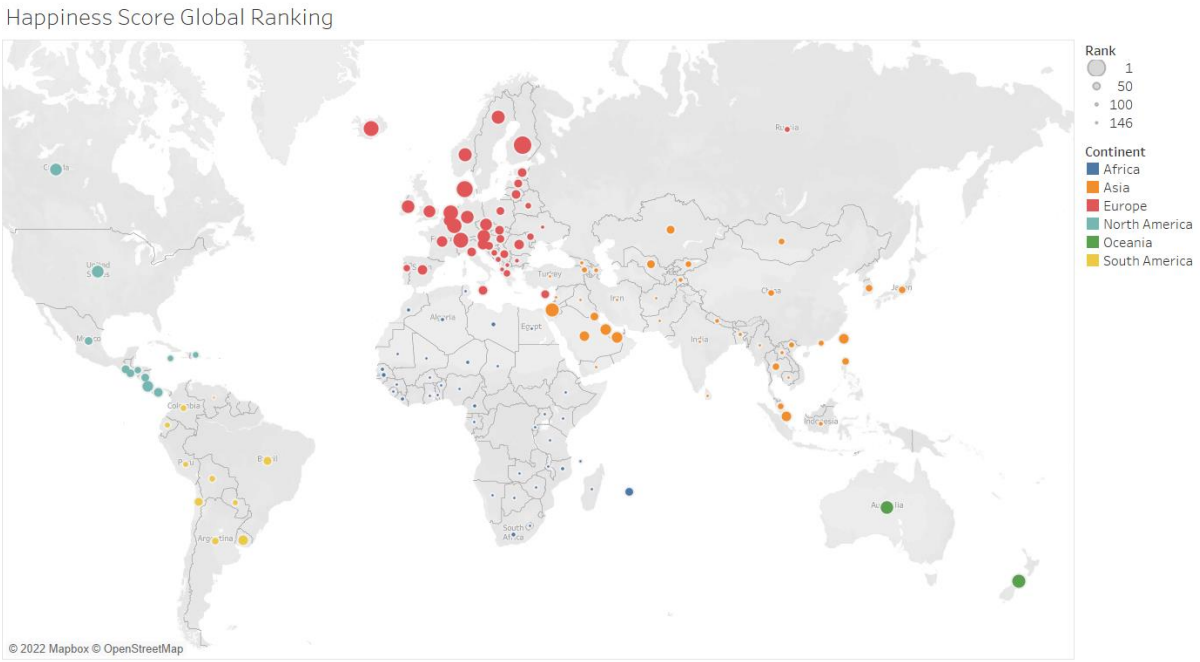
Sources: Happiness Dataset.xlsx

2.3 Methods

- Descriptive Analysis
- Regression Analysis (The R-squared values used in this report are mostly adjusted R-squared)
- Visualization Tools: R and Tableau

3. Analysis and Visualization

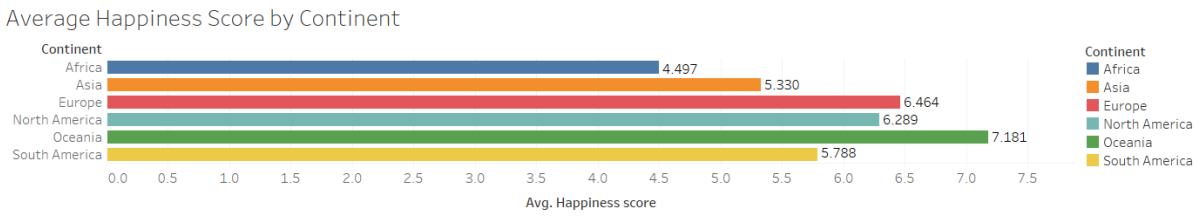
3.1. Global Happiness Ranking by Map Visualization



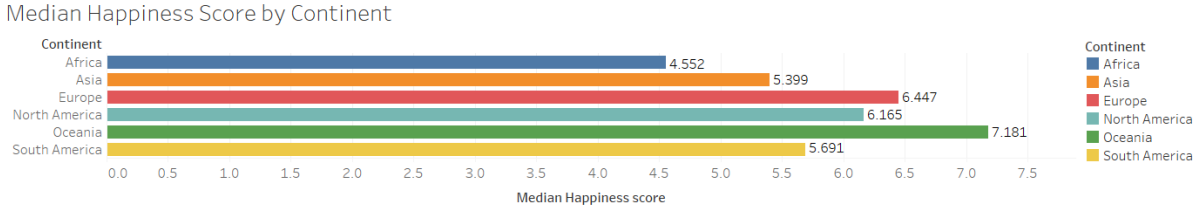
Map based on Longitude (generated) and Latitude (generated). Color shows details about Continent. Size shows sum of Rank. Details are shown for Country.

The majority of the highest-ranking countries are located in Europe and Oceania, while the lowest-ranking countries are located in Africa. Additionally, the ranks of countries in the Middle East and North America are similar, and they are far greater than the remaining portion of Asia.

3.2 Descriptive Analysis Happiness Score by Continent



Average of Happiness score for each Continent. Color shows details about Continent. The marks are labeled by average of Happiness score.

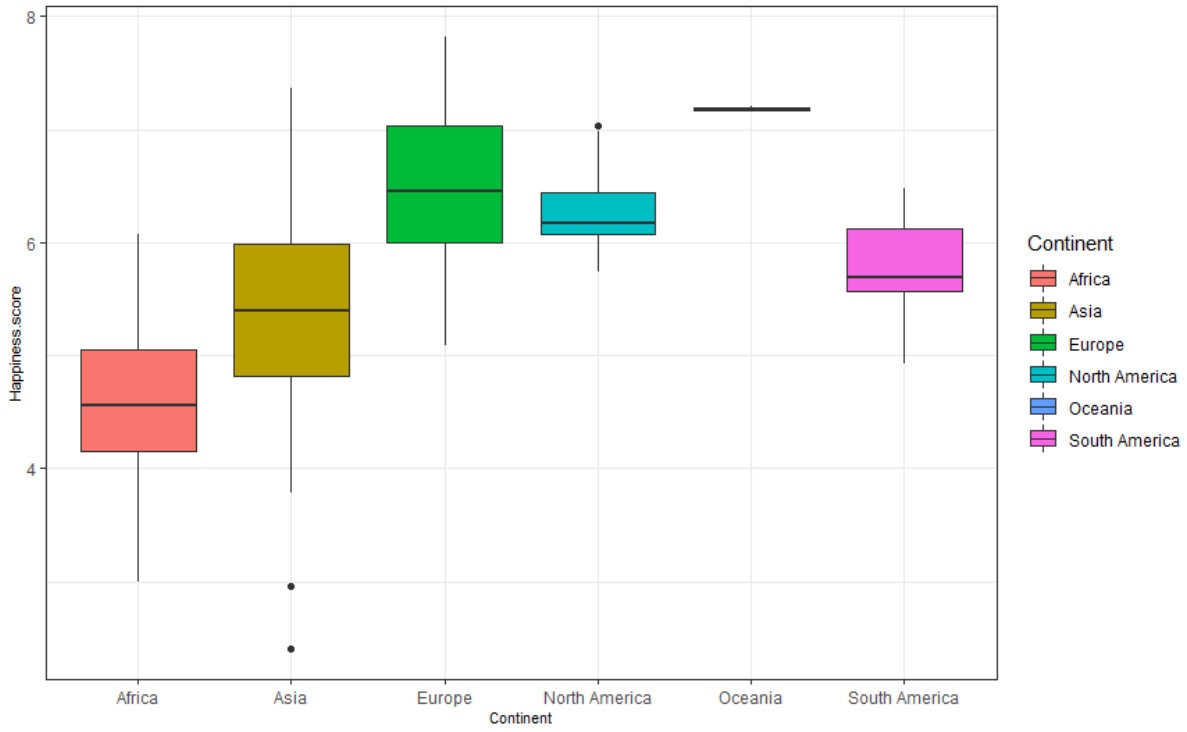


Median of Happiness score for each Continent. Color shows details about Continent. The marks are labeled by median of Happiness score.

As previously stated, Oceania has the highest median and average scores for happiness.

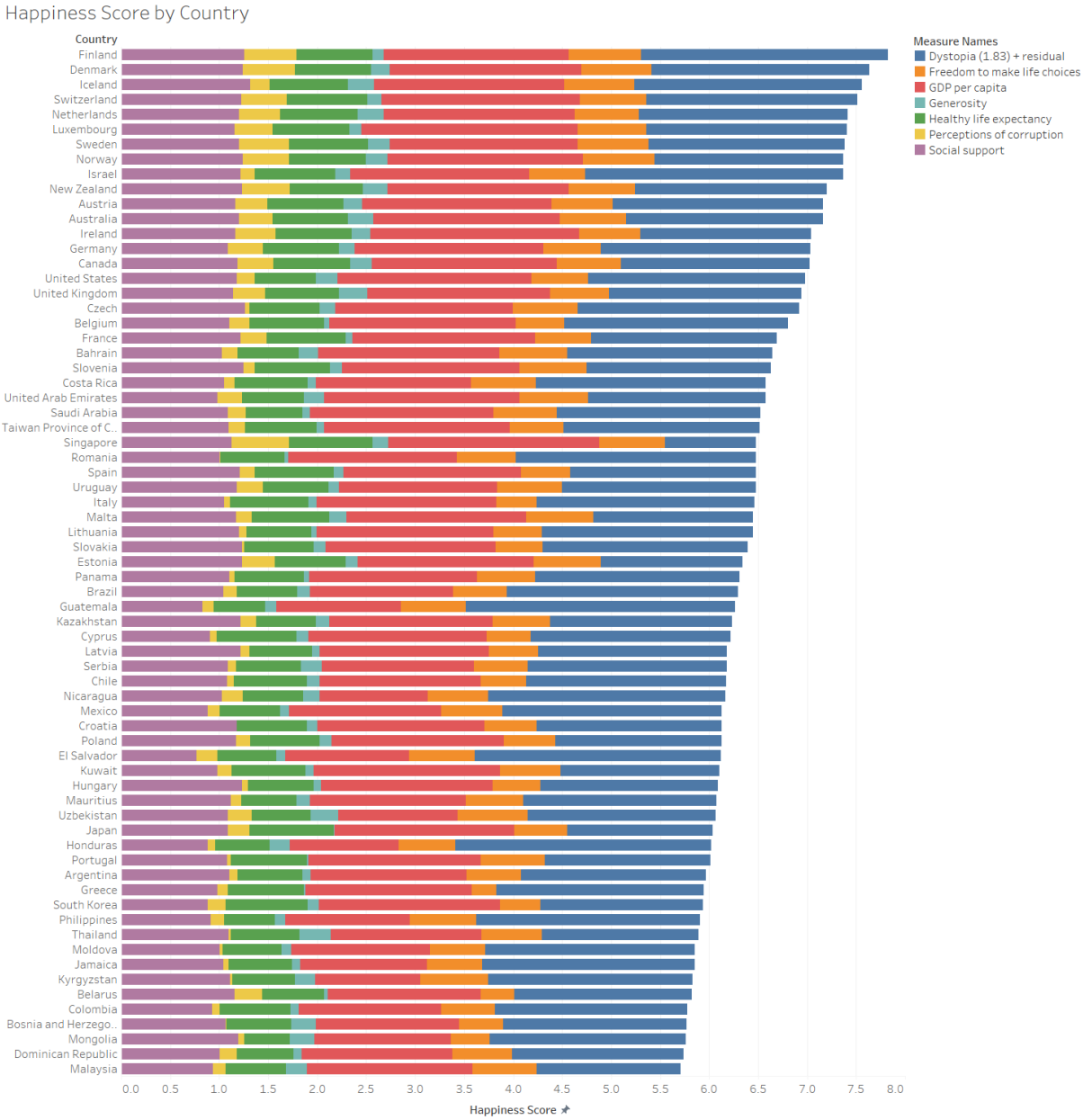
Europe is ranked second in terms of them, slightly higher than North and South America. Africa has the lowest median and average values in the world, following Asia.

Take a look to see how the happiness score is distributed across different countries and continents.



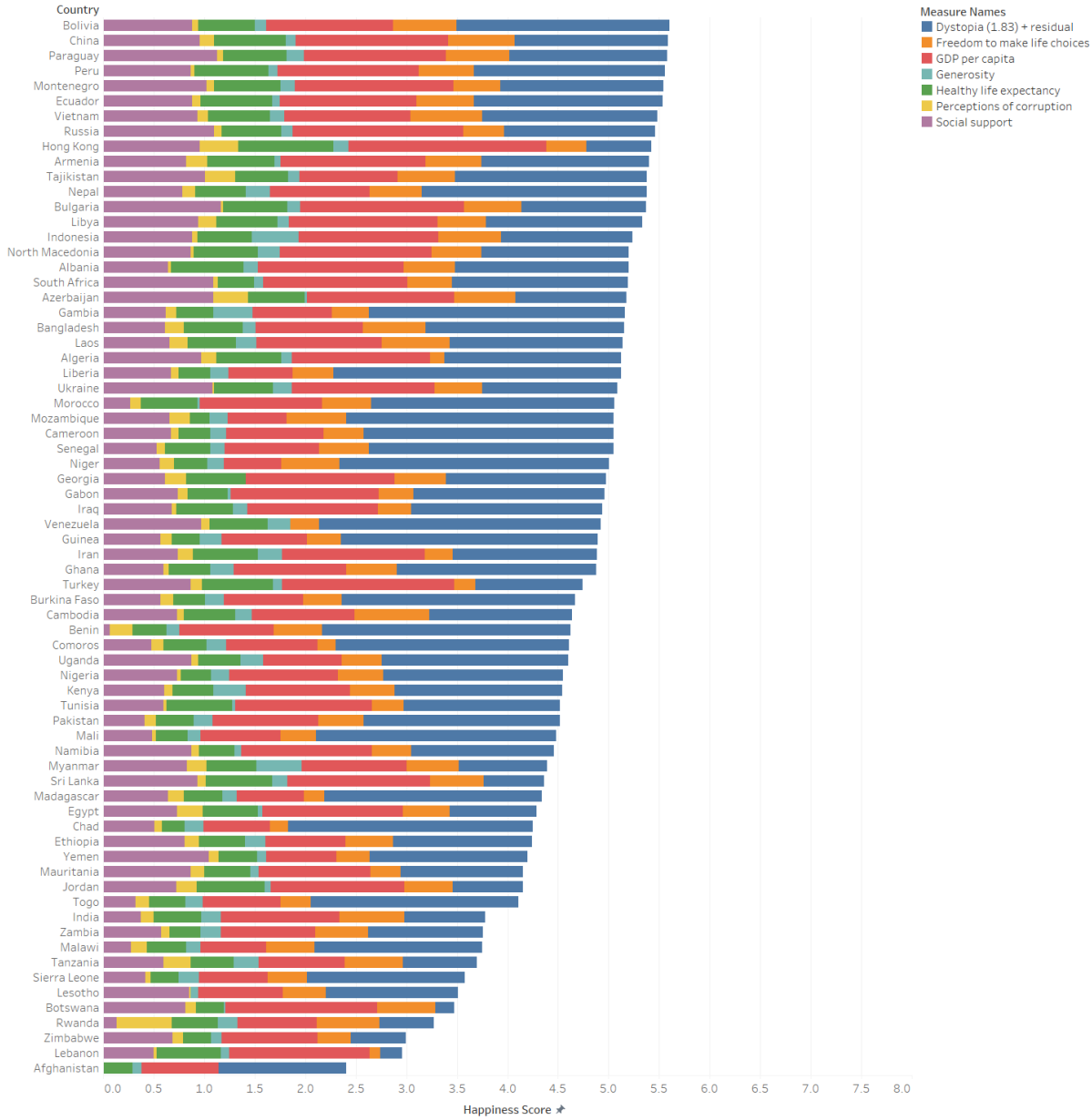
3.3 Happiness Score By Country Visualization

To illustrate the distribution of each factor of the Happiness Score by country in descending order of the total score, we create a stacked bar chart.



Dystopia (1.83) + residual, Freedom to make life choices, GDP per capita, Generosity, Healthy life expectancy, Perceptions of corruption and Social support for each Country. Color shows details about Dystopia (1.83) + residual, Freedom to make life choices, GDP per capita, Generosity, Healthy life expectancy, Perceptions of corruption and Social support. The view is filtered on Country, which has multiple members selected.

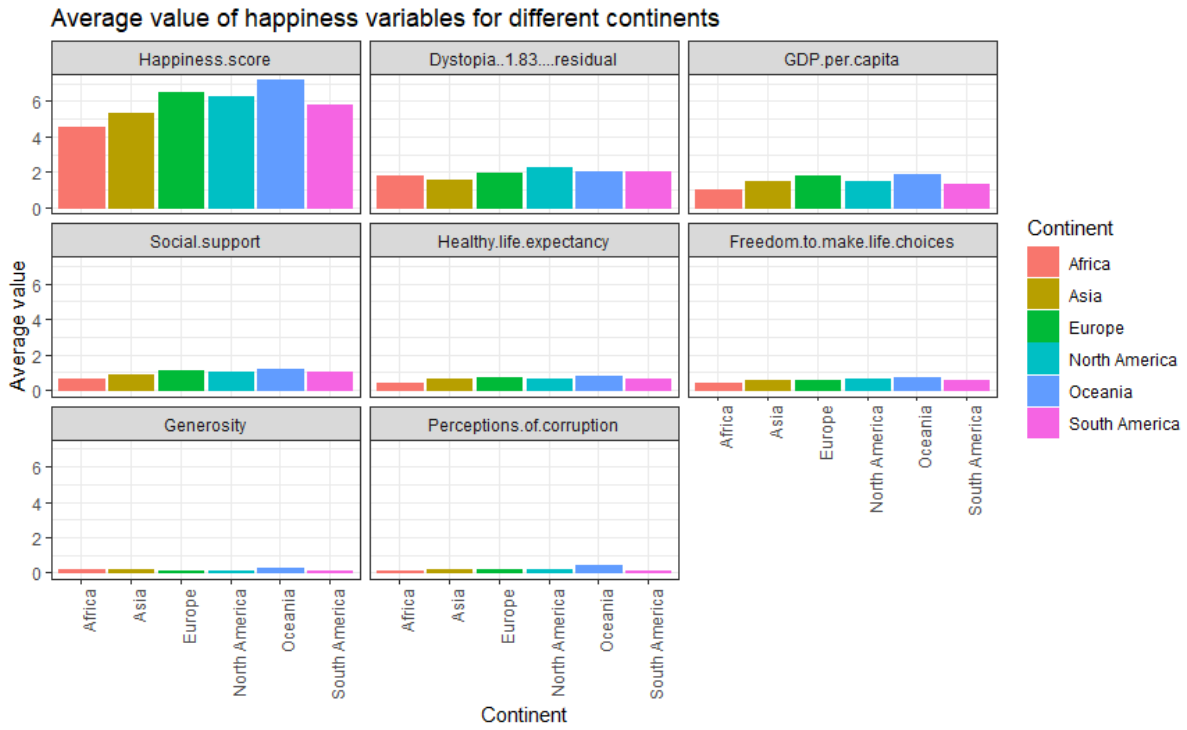
Happiness Score by Country



Dystopia (1.83) + residual, Freedom to make life choices, GDP per capita, Generosity, Healthy life expectancy, Perceptions of corruption and Social support for each Country. Color shows details about Dystopia (1.83) + residual, Freedom to make life choices, GDP per capita, Generosity, Healthy life expectancy, Perceptions of corruption and Social support. The view is filtered on Country, which has multiple members selected.

3.4 Happiness Ranks, Scores, and Their Seven Main Factors

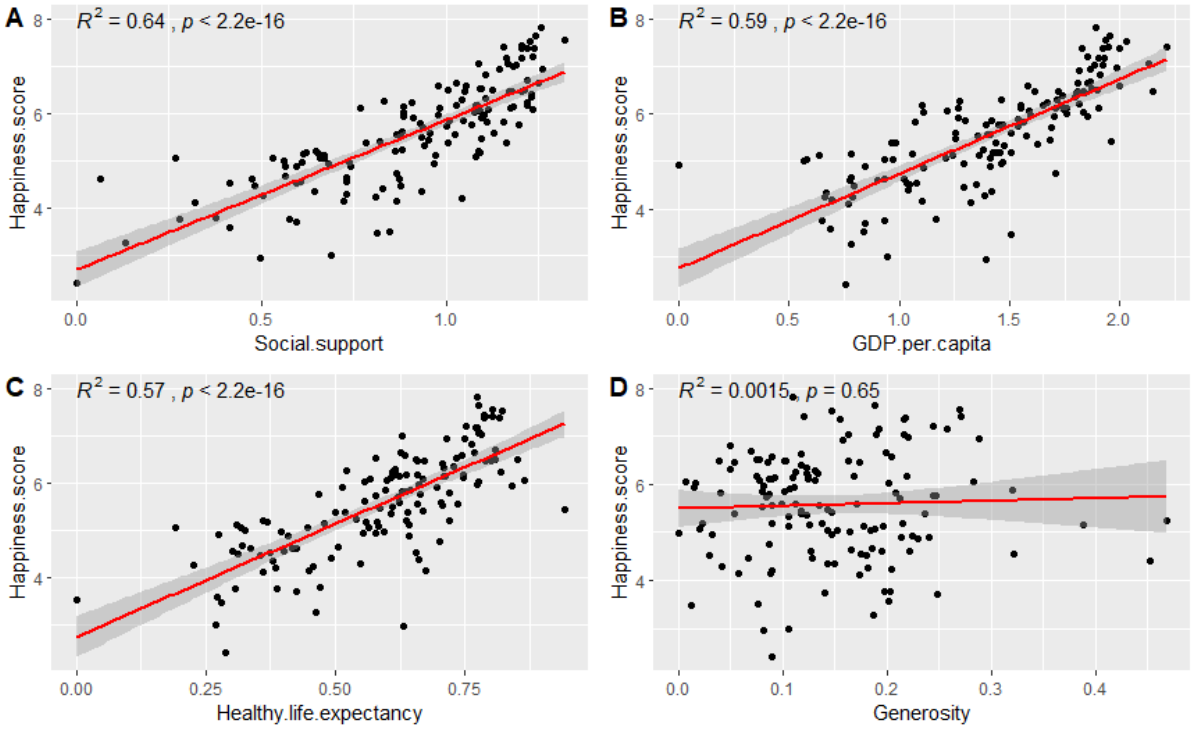
3.4.1 What is The Rank of The Average Values of Each Main Happiness Factor by Continent?



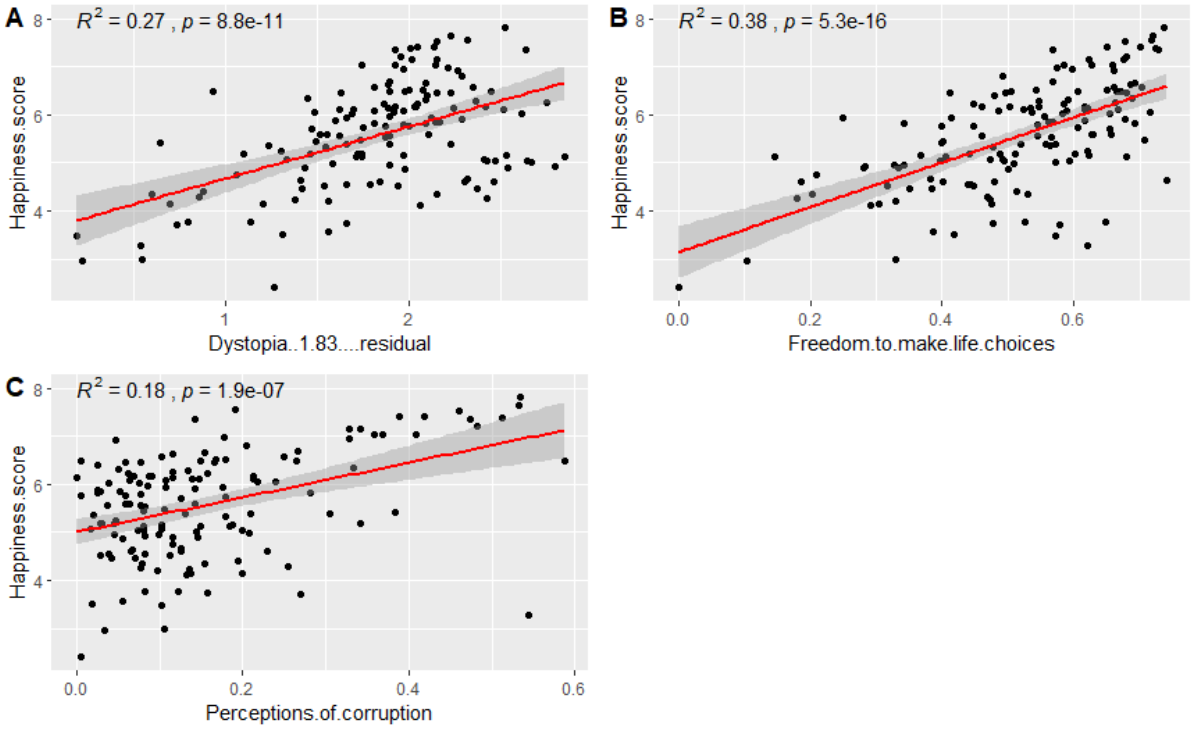
Oceania has the greatest average in all categories except dystopia + residual, followed by Europe, North America, and South America, which all have similar averages in terms of happiness and the other seven criteria. Finally, Asia and Africa rank the lowest in every category. The variations in average scores between those continents in the areas of generosity, freedom to make life choices, and perceptions of corruption, in particular, are insignificant.

3.4.2 The Relationship Between Happiness Score and Each of Its Main Factors

After we did seven simple regression analyses between happiness scores and all of its main factors, we discovered that social support has the most significant impact on our happiness and provides insight into why people are happy in particular nations. Additionally, GDP per capita and healthy life expectancy also have a substantial association with overall happiness in a country, followed by freedom to make life choices, dystopia + residual, and perceptions of corruption. Moreover, generosity doesn't have any relationship with happiness score.

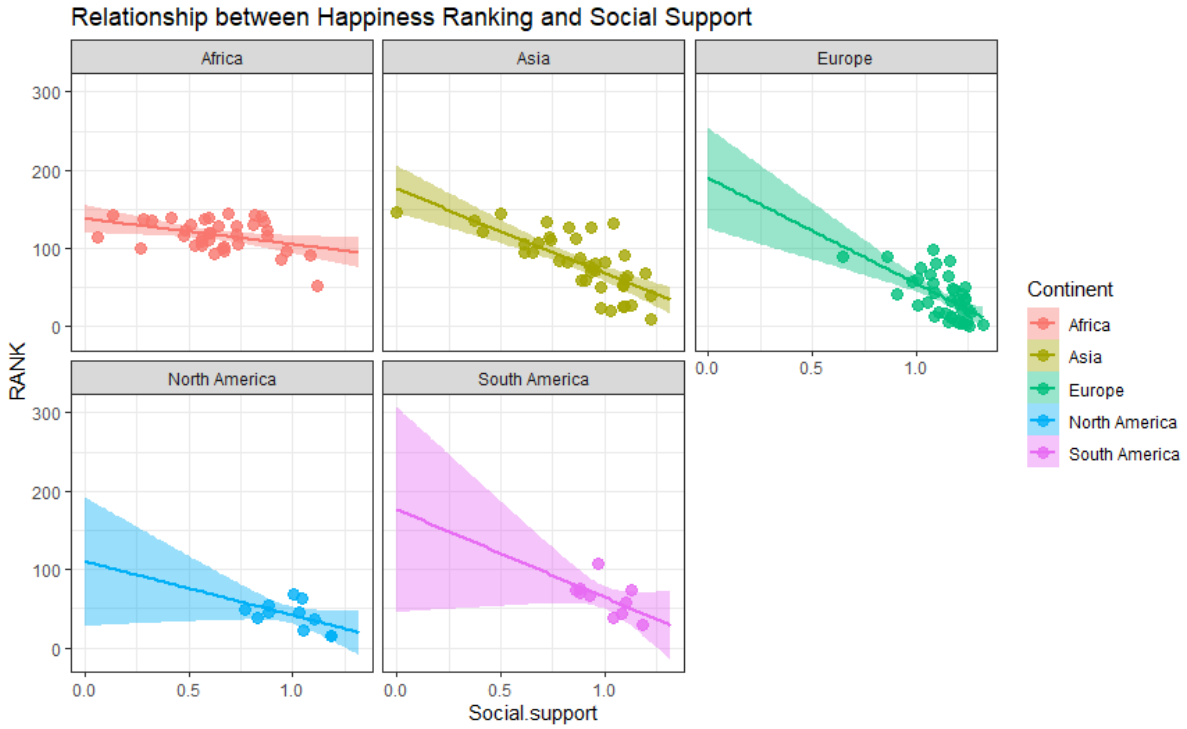


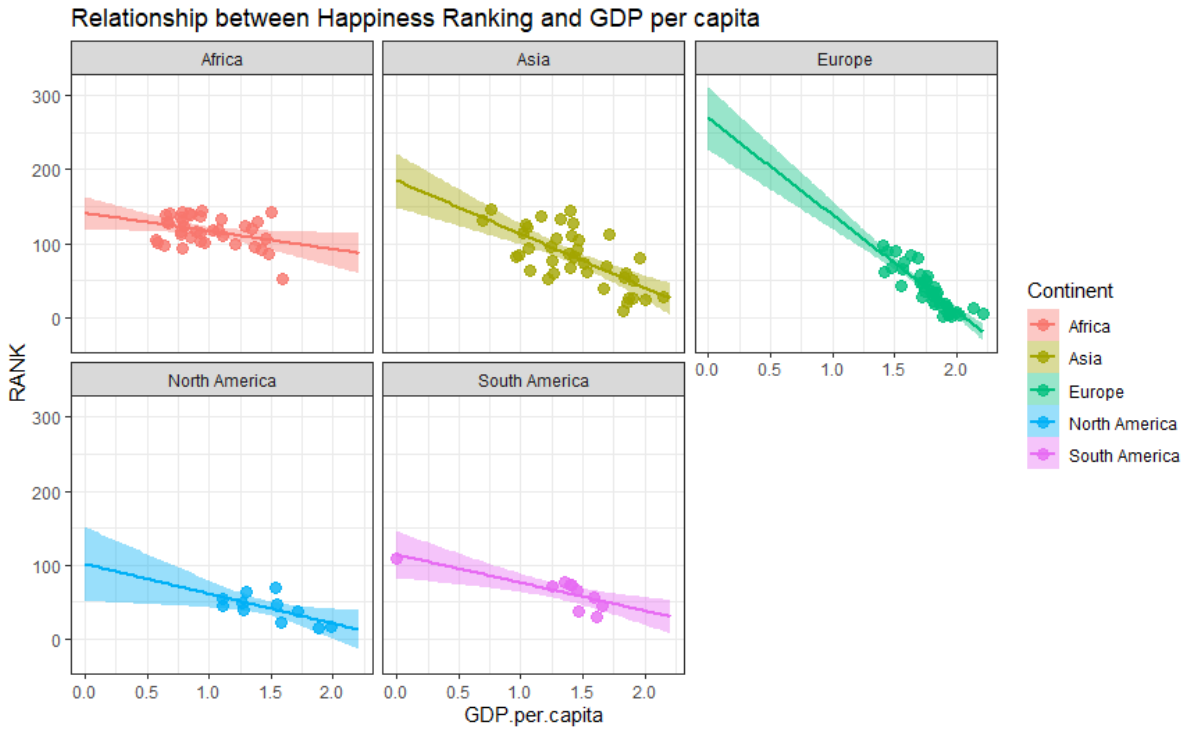
With the three strongest factors of happiness, each has a p-value **> 0.0001** and the R-squared value **> 0.5**, which means we can reject the hypothesis that these factors have no association with the happiness score and they explain quite much of the variance in happiness overall. Each of these elements has a strong and significant positive relationship with happiness score. Otherwise, because generosity has a p-value **> 0.5**, we cannot reject the hypothesis, indicating that there is no correlation between happiness and generosity.

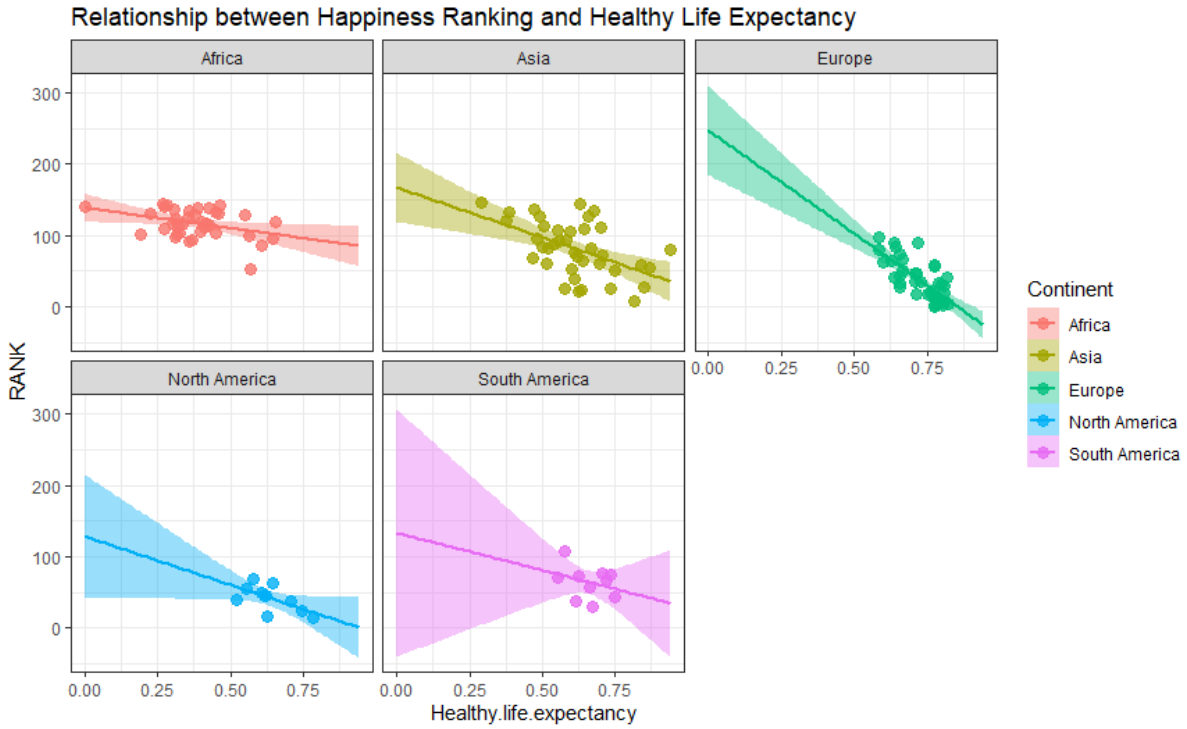


With the three remaining factors, dystopia + residual, freedom to make life choices, and perceptions of corruption, their p-values are < **0.0001** and R-squared values are < **0.4**, so they show the weak but significant positive relationships between happiness score and each of them. The higher one of these attributes is associated with a higher happiness score of a country.

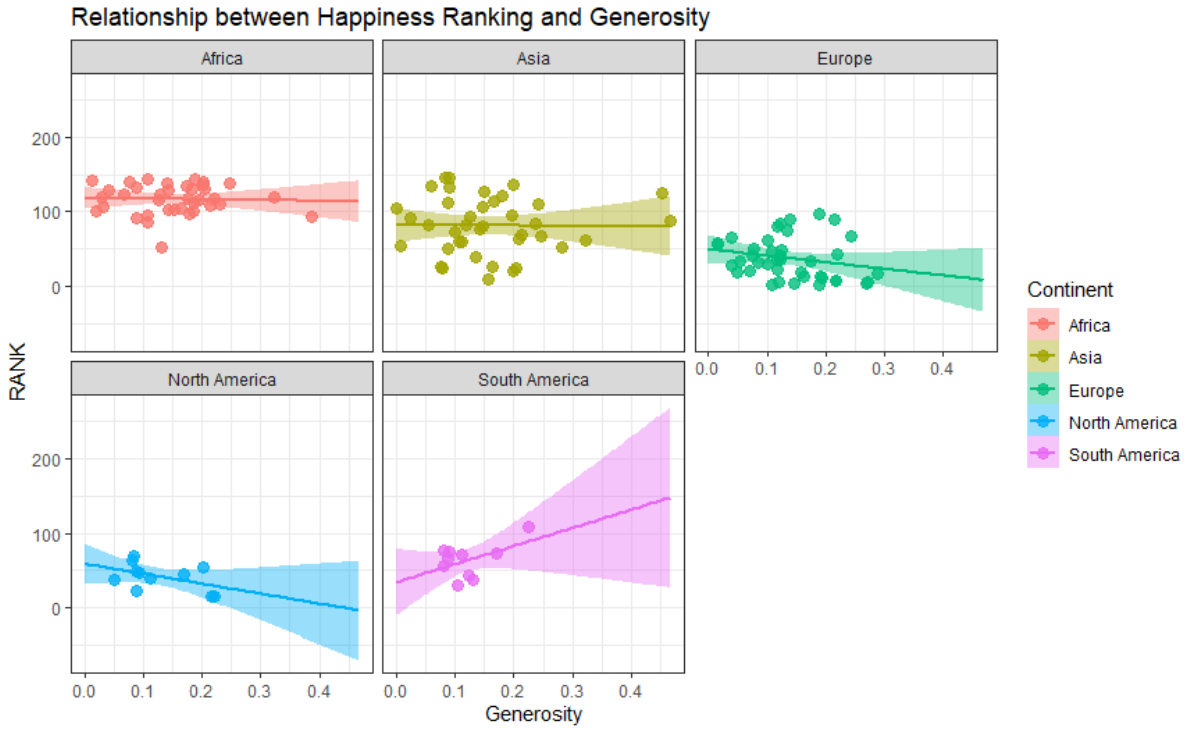
3.4.3 Relationships between Happiness Rank and Some of Its Main Factors by Continent







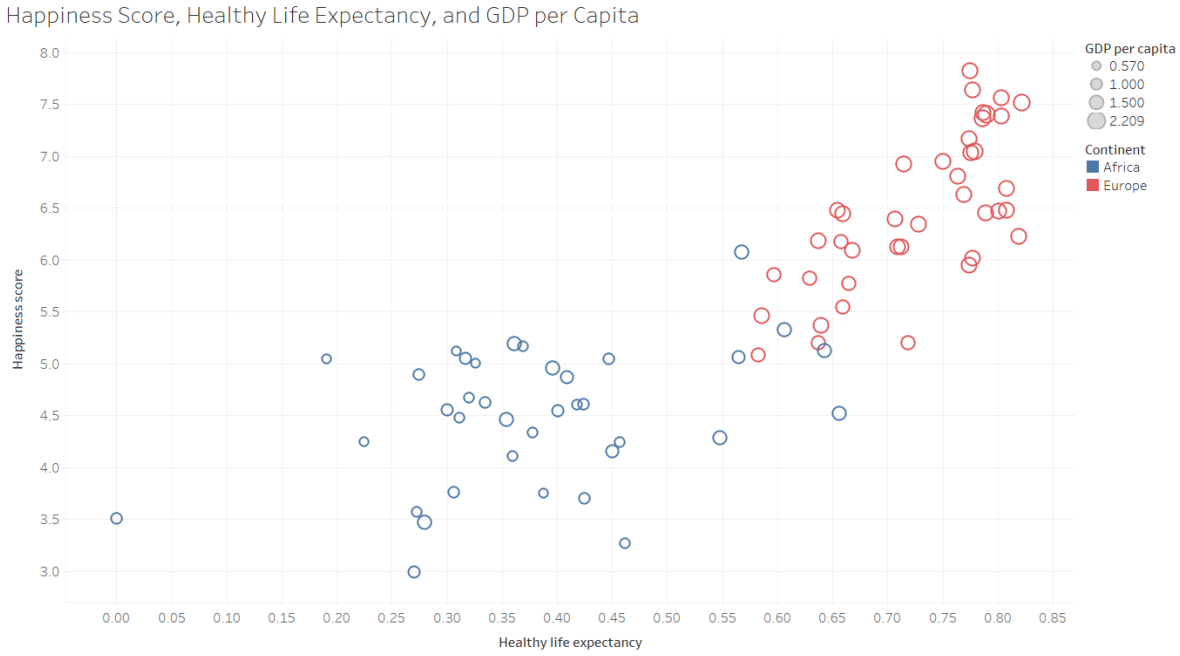
They all exhibit similar tendencies. Europe demonstrates the greatest and most negative association - the higher a country's score on each of these three factors, the higher its happiness score, followed by Asia and America. Africa shows the weakest relationship.



Africa and Asia exhibit a similar pattern and they show that there's no relationship between generosity in their countries and happiness scores. However, in Europe and North America, there is still a weak but significant negative linear relationship between generosity and happiness - the generosity score of European and North American countries is related to a higher rank in happiness. South America demonstrates the inverse trend.

3.4.4 Further Analysis of Healthy Life Expectancy and GDP per Capita

Because there are so many parallels between these two disciplines, we attempt to draw more highlights between Europe and Africa's strongest and weakest relationships.

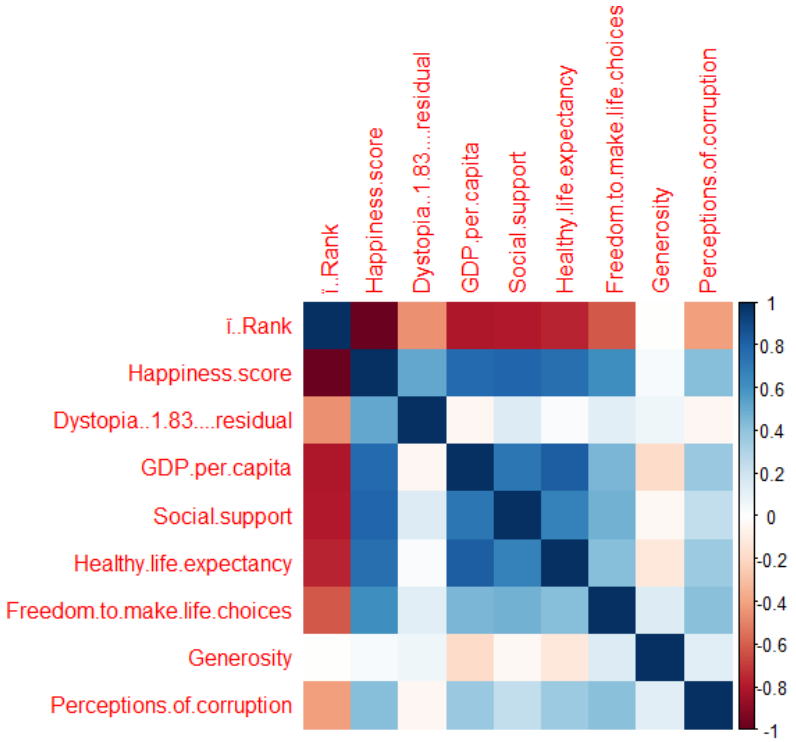


Healthy life expectancy vs. Happiness score. Color shows details about Continent. Size shows sum of GDP per capita. The view is filtered on Continent, which keeps Africa and Europe.

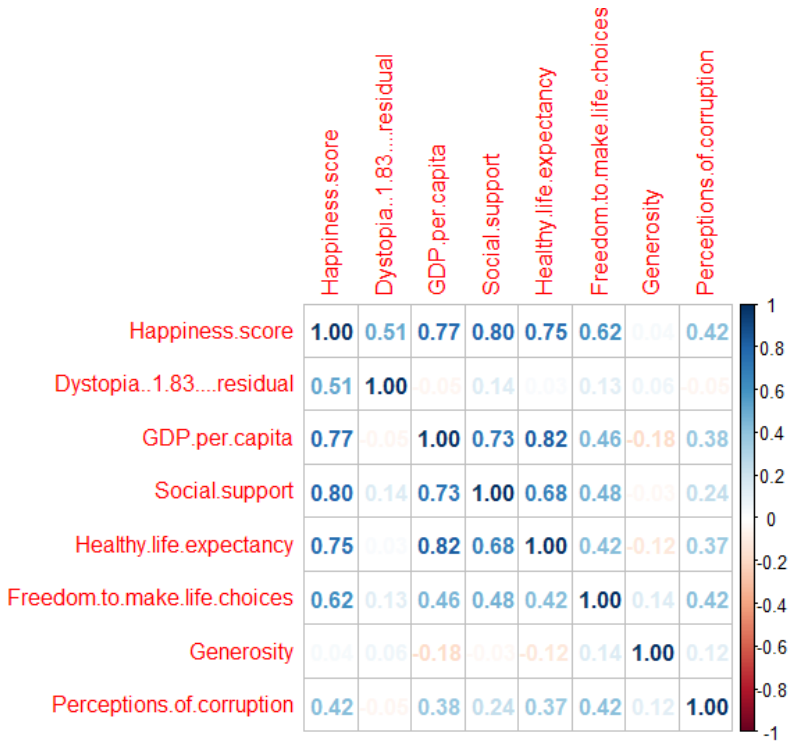
It is clear that happier countries tend to be those with healthier life expectancies, and a higher GDP per capita and the differences between Europe and Africa are significant.

3.4.5 Examine the correlation between Happiness Rank, Score, and any two of their primary variables in aggregate

Let's see the correlation between the main factors, happiness scores, and happiness ranks in our dataset.



Clearly, there is an inverse relationship between "Happiness Rank" and the other elements. Therefore, let us delete the happiness rank and examine the correlation once more.



According to the above correlation plot, the most significant factors contributing to happiness are social support, followed by GDP per capita, and healthy life expectancy scores. Corruption perceptions and generosity scores have the smallest effect on happiness, which is the same as our simple regression analysis.

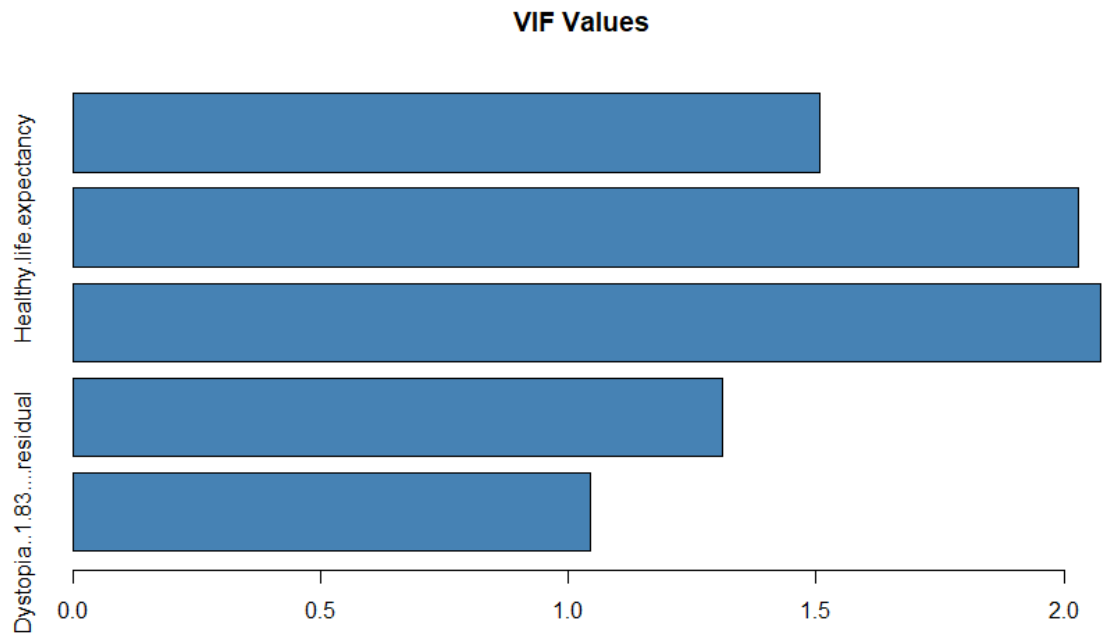
There's one correlation > 0.8 between any two factors GDP per capita and healthy life expectancy, so multicollinearity can occur when conducting multiple regression analysis. Moreover, the correlation between GDP per capita and social support is 0.73, which is also very high. Besides them, the correlations between any two of the main factors of happiness score are insignificant.

3.4.6 The Multiple Regression Analysis Between All of The Main Factors with Happiness Score

GDP per capita was removed due to the correlation between GDP per capita and healthy life expectancy is 0.82, which can cause multicollinearity in our regression analysis. Substitution of either GDP per capita or healthy life expectancy by another uncorrelated statistic is thus advised to optimize the happiness score calculation from a data analytics perspective.

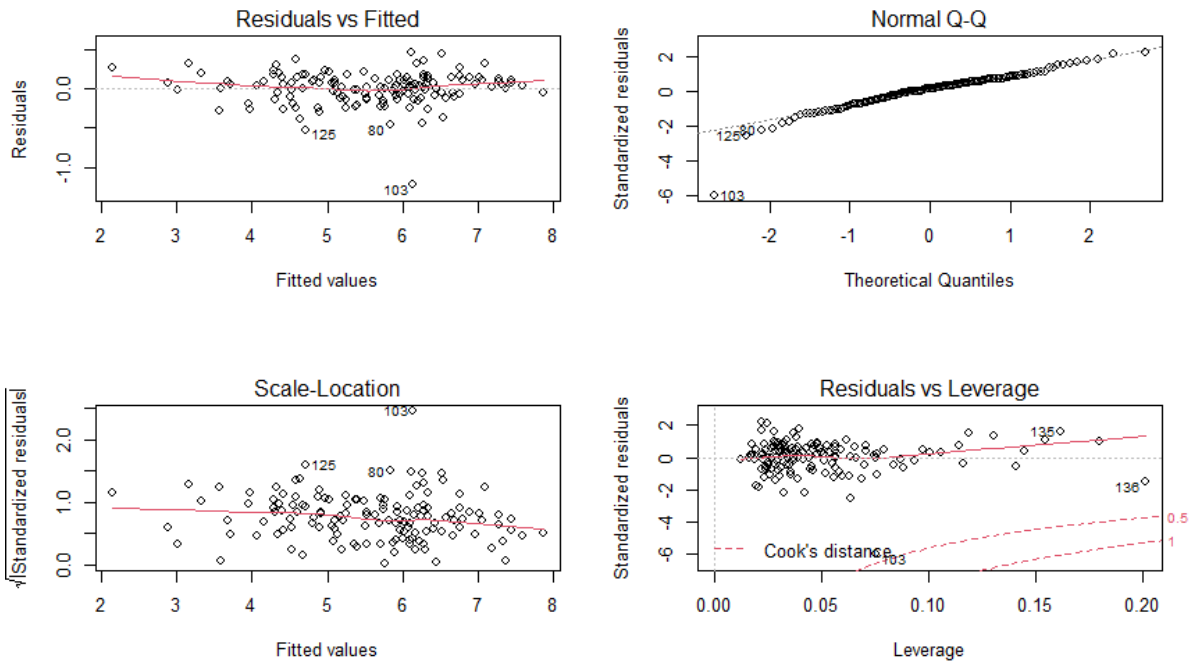
```
Call:
lm(formula = Happiness.score ~ Dystopia..1.83....residual + Perceptions.of.corruption +
    Social.support + Healthy.life.expectancy + Generosity + Freedom.to.make.life.choices,
    data = Happiness)

Residuals:
     Min       1Q   Median       3Q      Max
-1.22186 -0.10619  0.02383  0.12627  0.46323

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   0.28721    0.10026   2.865  0.00486 **
Dystopia..1.83....residual    0.91197    0.03506  26.014  < 2e-16 ***
Perceptions.of.corruption     1.27533    0.16061   7.940 7.73e-13 ***
Social.support                1.49909    0.09317  16.090  < 2e-16 ***
Healthy.life.expectancy       2.30337    0.15052  15.302  < 2e-16 ***
Generosity                    0.37384    0.22961   1.628  0.10588
Freedom.to.make.life.choices  1.21592    0.15310   7.942 7.65e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2118 on 132 degrees of freedom
Multiple R-squared:  0.9647,    Adjusted R-squared:  0.9631
F-statistic:   601 on 6 and 132 DF,  p-value: < 2.2e-16
```

The p-value of attribute “generosity” is higher than 0.05, so it should be deleted. Then, the VIF values of the remaining attributes can be checked to complete the process of detecting multicollinearity below.



All of them are smaller than 5, so they are appropriate for the multiple regression analysis. This is the final result.

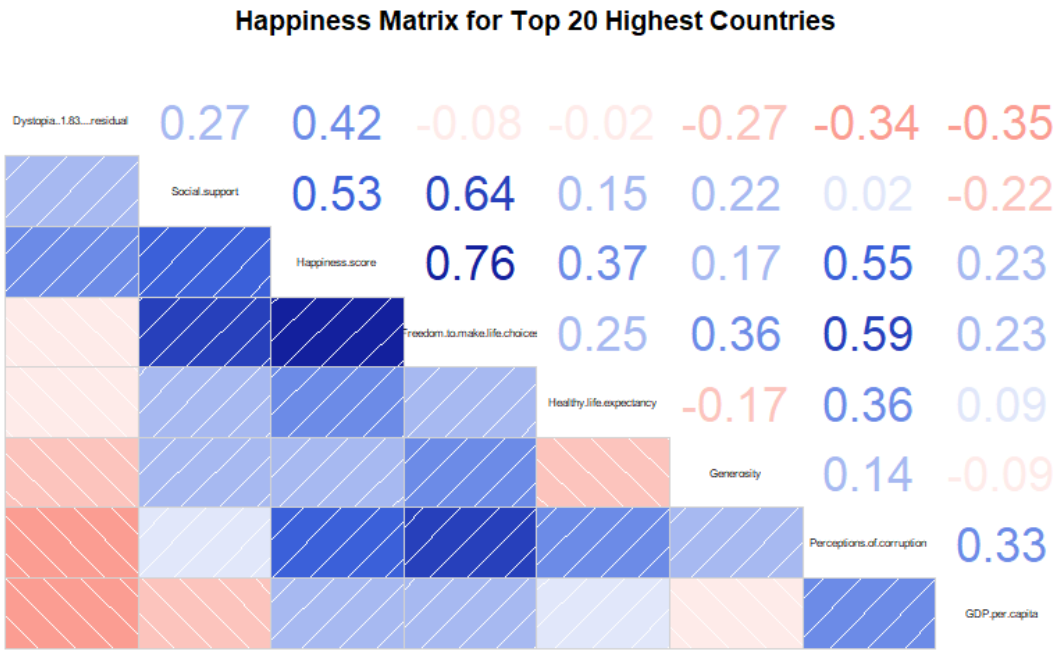
```
Call:
lm(formula = Happiness.score ~ Dystopia..1.83....residual + Perceptions.of.corruption +
    Social.support + Healthy.life.expectancy + Freedom.to.make.life.choices,
    data = Happiness)

Residuals:
     Min       1Q   Median       3Q      Max
-1.18495 -0.11659  0.02792  0.13878  0.47667

Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                   0.33779    0.09591   3.522 0.000588 ***
Dystopia..1.83....residual    0.91489    0.03523  25.971  < 2e-16 ***
Perceptions.of.corruption     1.30993    0.16019   8.177 2.03e-13 ***
Social.support                1.50155    0.09373  16.019  < 2e-16 ***
Healthy.life.expectancy       2.25662    0.14867  15.178  < 2e-16 ***
Freedom.to.make.life.choices  1.25222    0.15240   8.217 1.63e-13 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2131 on 133 degrees of freedom
Multiple R-squared:  0.964, Adjusted R-squared:  0.9626
F-statistic: 711.8 on 5 and 133 DF,  p-value: < 2.2e-16
```



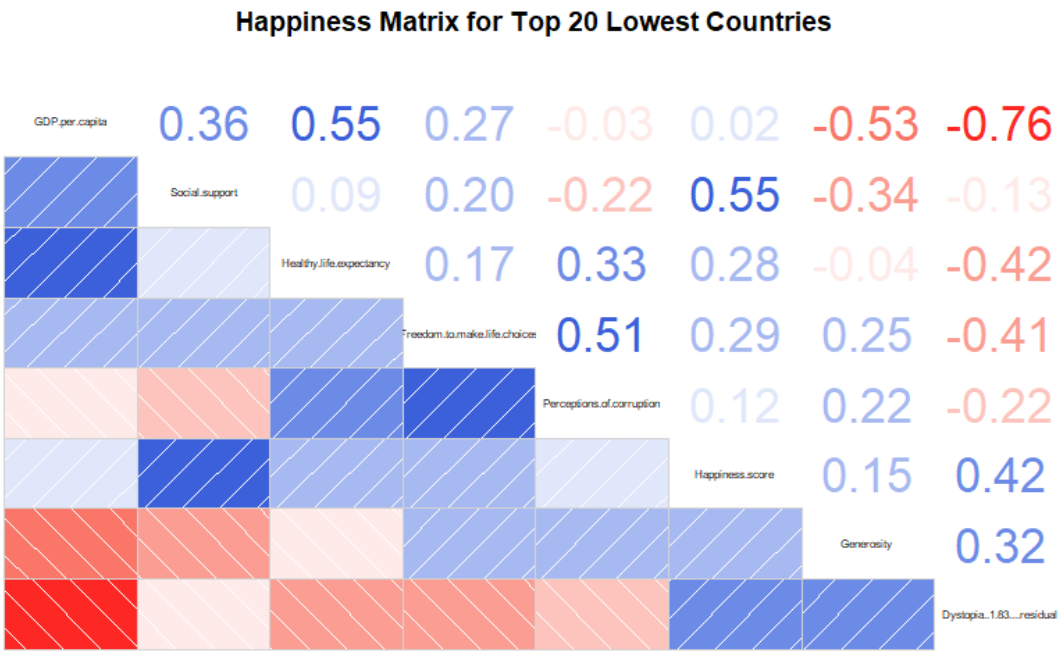
The p-value is **< 0.0001** and the R-squared value is **0.9626**, indicating a strong and significant positive linear relationship between the happiness score and the 5 remaining independent values. This is also the most optimal model to predict the happiness score of a country. However, statisticians may be concerned about whether the exceptionally high R-squared value indicated an overfitting problem. As a result, our group proposes that this possible issue

be re-examined using lasso regression or stepwise regression (backward elimination or forward selection), which both address the issue of hidden variable selection.

3.5 Comparing the Top 20 highest and Top 20 Lowest Happiness Scores



From highest to lowest, the link between happiness scores and their primary element includes freedom to make life choices, perceptions of corruption, social support, dystopia + residual, healthy life expectancy, GDP per capita, and generosity. It's a little bit different from the trend in 3.4.2, in which social support, GDP per capita, and healthy life expectancy correlation are smaller than the others.

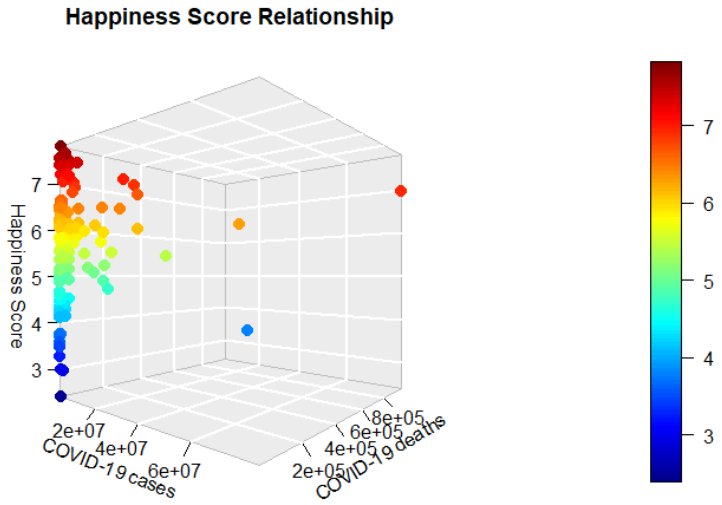


The order of the correlation between happiness score and its main factor of the top 20 highest countries from the highest to the lowest is social support, dystopia + residual, freedom to make life choices, healthy life expectancy, generosity, perception of corruption, GDP per capita. Overall, the trend is also different from section 3.4.2 and most of the correlations in the top 20 lowest countries are smaller than 0.5, which means the relationships are weak.

3.6 Relationship between Happiness Scores and Other Related Attributes

3.6.1 COVID-19 Cases and Deaths

A multiple regression analysis is conducted with happiness score as a dependent variable and COVID-19 cases/deaths are independent variables.

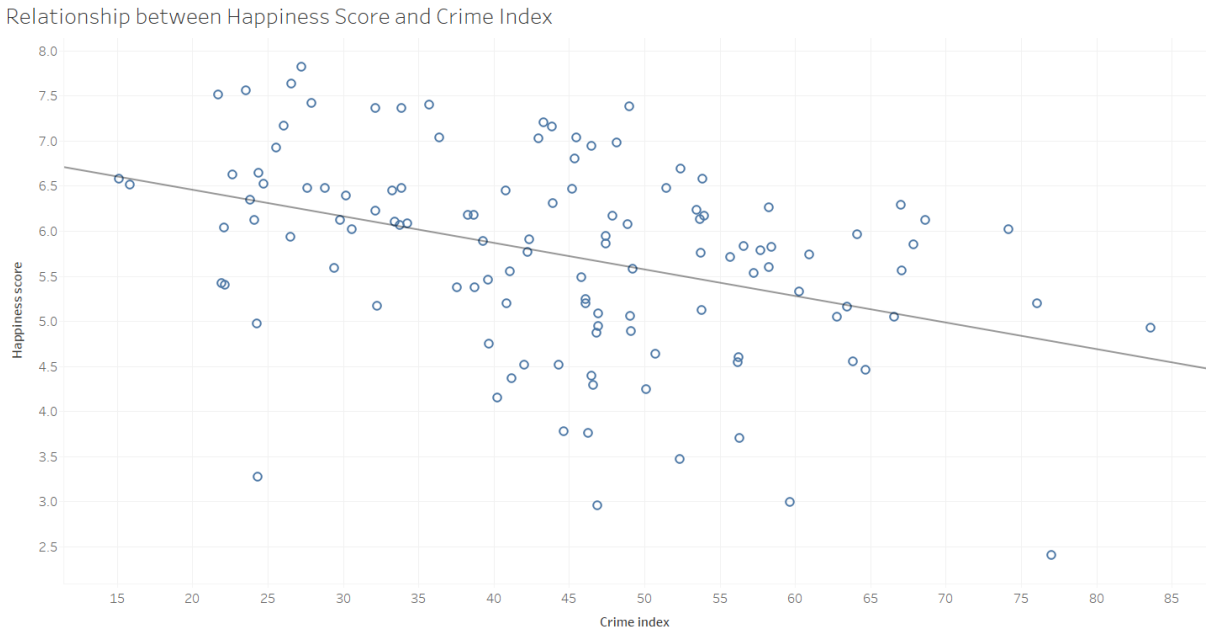



```
Call:
lm(formula = Happiness.score ~ COVID.19.cases + COVID.19.deaths,
    data = Happiness)

Residuals:
     Min       1Q   Median       3Q      Max
-3.08203 -0.61656  0.09188  0.70756  2.28140

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.498e+00  9.864e-02  55.742   <2e-16 ***
COVID.19.cases   5.855e-08  2.555e-08   2.291   0.0235 *
COVID.19.deaths -2.960e-06  1.904e-06  -1.555   0.1223
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.083 on 136 degrees of freedom
Multiple R-squared:  0.04877,   Adjusted R-squared:  0.03478
F-statistic: 3.486 on 2 and 136 DF,  p-value: 0.03339
```

The p-value is **< 0.05** and the R-squared value is **0.04877**. This shows a weak but positive linear relationship between happiness score and COVID-19 cases/deaths - the more COVID cases and deaths of a country is associated with a higher happiness score.

This relationship is not explained fully in our life, but it may be because the highest-ranking countries in Europe, Oceania, and America have a disproportionate number of COVID-19 cases and deaths in comparison to the rest of the globe, resulting in a ‘confusing’ analysis.

3.6.2 Crime Index



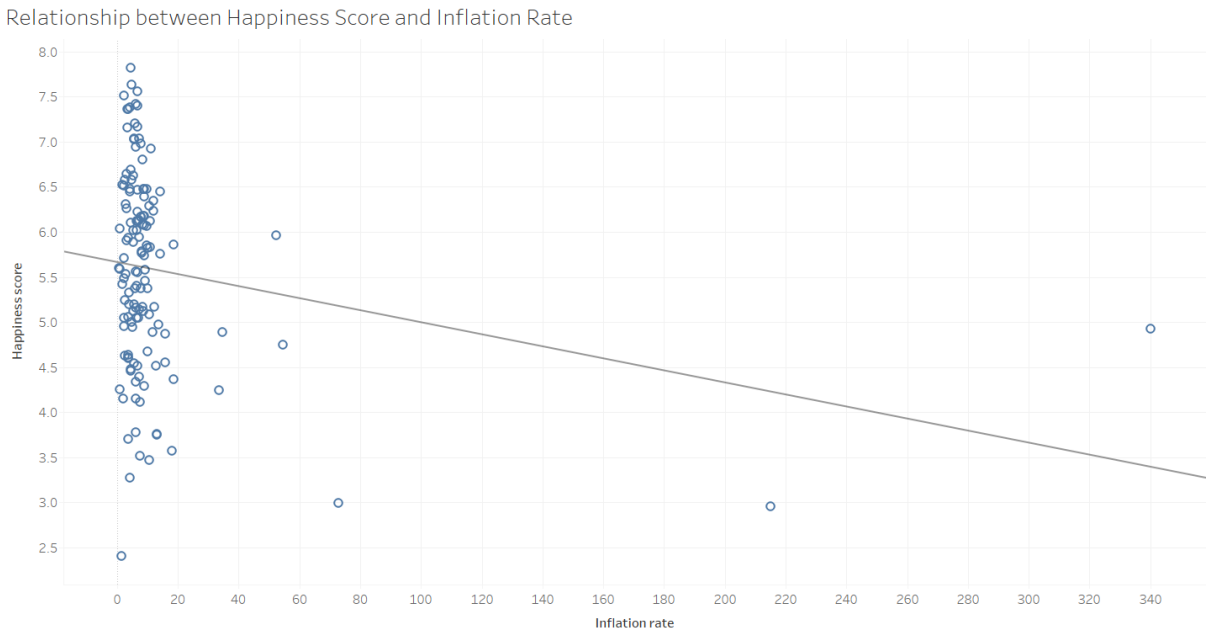
The p-value of the model is **< 0.0001**. This means we can *reject* the hypothesis that the crime index of a country has no association with its score of happiness. The R-squared is **0.152983** so the crime index explains very little of the variance in happiness scores overall.

There is a weak but significant negative linear relationship - the lower crime index of a country is associated with a higher score of happiness.

3.6.3 Unemployment Rate and Inflation Rate

Unemployment rate vs. Happiness score.

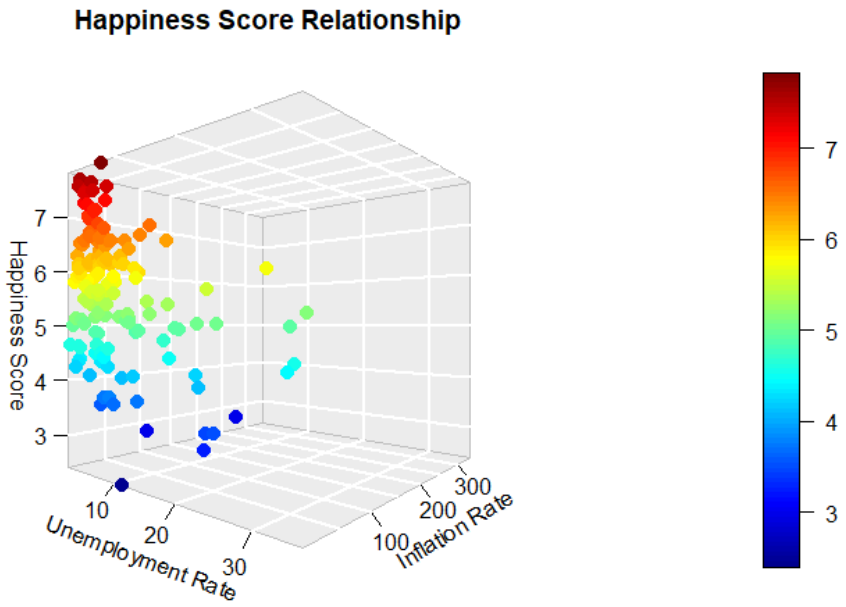
```
Call:
lm(formula = Happiness.score ~ Unemployment.rate, data = Happiness)

Residuals:
    Min      1Q  Median      3Q     Max
-3.0334 -0.7116  0.1220  0.7388  2.1616

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        5.99004    0.14065  42.589  < 2e-16 ***
Unemployment.rate -0.04934    0.01279  -3.859 0.000175 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.05 on 137 degrees of freedom
Multiple R-squared:  0.09805,   Adjusted R-squared:  0.09147
F-statistic: 14.89 on 1 and 137 DF,  p-value: 0.0001746
```



Inflation rate vs. Happiness score. The data is filtered on Country, which keeps 139 of 139 members.

```
Call:
lm(formula = Happiness.score ~ Inflation.rate, data = Happiness)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2362 -0.6757  0.0750  0.7845  2.2001

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    5.650449   0.097559  57.918   <2e-16 ***
Inflation.rate -0.006563   0.002688  -2.442   0.0159 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.083 on 137 degrees of freedom
Multiple R-squared:  0.04172,   Adjusted R-squared:  0.03472
F-statistic: 5.964 on 1 and 137 DF,  p-value: 0.01588
```

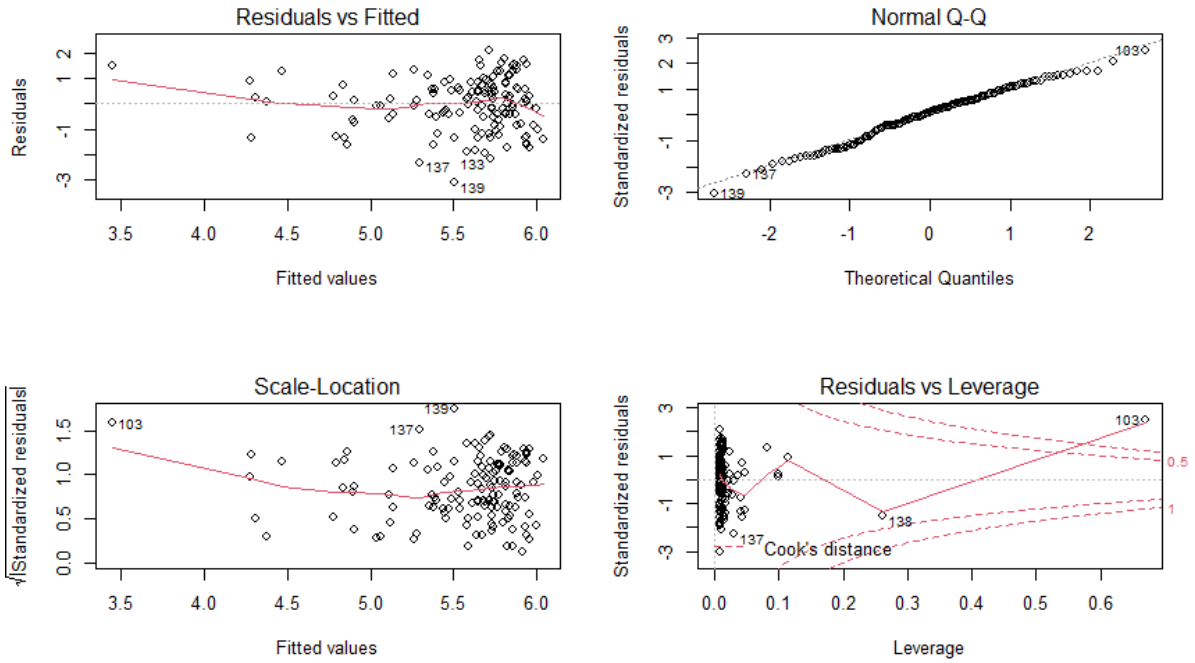
From the plots, there are weak but negative linear relationships between happiness score and unemployment rate/inflation rate. Multiple regression analysis is conducted with the happiness score as a dependent variable, and the others are independent variables. The following 3D scatter plot visualizes the relationship.



It is obvious that the lower the unemployment rate and inflation rate, the higher the happiness score.

Result of Regression Analysis:

```
Call:
lm(formula = Happiness.score ~ Unemployment.rate + Inflation.rate,
    data = Happiness)

Residuals:
    Min      1Q  Median      3Q     Max
-3.1040 -0.6353  0.1006  0.7294  2.1075

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       6.079865   0.141689  42.910  < 2e-16 ***
Unemployment.rate -0.050110   0.012514  -4.004 0.000102 ***
Inflation.rate    -0.006799   0.002552  -2.664 0.008651 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.028 on 136 degrees of freedom
Multiple R-squared:  0.1428,    Adjusted R-squared:  0.1302
F-statistic: 11.33 on 2 and 136 DF,  p-value: 2.818e-05
```



The p-value of the model is **< 0.0001** and the R-squared value is **0.1302**, which means there is a weak but significant negative linear relationship between them - the lower the unemployment rate and inflation rate, the higher the happiness score.

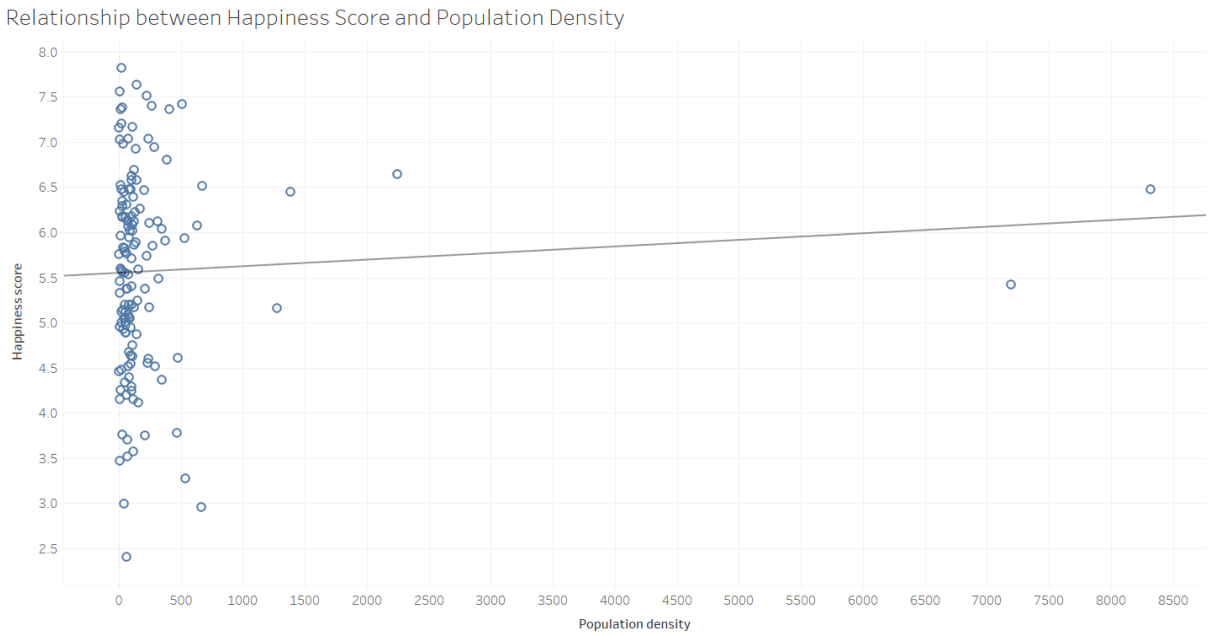
3.6.4 Suicide Rate

Suicide rate vs. Happiness score.

```
Call:
lm(formula = Happiness.score ~ Suicide.rate, data = Happiness)

Residuals:
    Min      1Q  Median      3Q     Max
-3.2452 -0.6411  0.0421  0.7619  2.1421

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.38993    0.14768   36.50   <2e-16 ***
Suicide.rate  0.01888    0.01203    1.57    0.119
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.096 on 137 degrees of freedom
Multiple R-squared:  0.01767,   Adjusted R-squared:  0.0105
F-statistic: 2.465 on 1 and 137 DF,  p-value: 0.1187
```

There’s no relationship between happiness score and suicide rate because the p-value is > **0.05**.

3.6.5 Population Density



Population density vs. Happiness score.

```
Call:
lm(formula = Happiness.score ~ Density, data = Happiness)

Residuals:
     Min       1Q   Median       3Q      Max
-3.15069 -0.66464  0.04897  0.74768  2.26943

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.550e+00  9.742e-02  56.972   <2e-16 ***
Density     7.293e-05  9.907e-05   0.736    0.463
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.104 on 137 degrees of freedom
Multiple R-squared:  0.00394,   Adjusted R-squared:  -0.00333
F-statistic: 0.5419 on 1 and 137 DF,  p-value: 0.4629
```

There’s no relationship between happiness score and population density because the p-value is > **0.05**.

4. Discussions and Conclusions

Though examining the relationships between happiness scores and supplied variables, we determined that social support has the greatest effect on the score. Increased social support, such as public housing and allowances, provides citizens in need with additional resources to improve their housing and consumption quality of life. They report a higher level of satisfaction with social support as a result of their improved quality of life, resulting in a higher score. GDP per capita and healthy life expectancy come in second and third place, respectively. A higher indicator shows that the country has advanced economically, with increased productivity and job opportunities for inhabitants, resulting in a greater quality of life and happiness score. Residents maintain their health for a longer period of time due to a higher healthy life expectancy. They may worry less about their health and feel happy as a result of having fewer diseases.

There are correlations between happiness score and the majority of additional factors, with the exception of suicide rate and population density, which are lower than our group expectation; since a high suicide rate implies residents are unhappy and dissatisfied with their lives, and people frequently exclaim "The more the merrier."

There are positive correlations between the happiness score and COVID-19 cases and deaths, but this is out of real-life situations. Since the epidemic has a harmful effect on the mental health of citizens, they avoid going out, resulting in a fall in countries' GDP and an increase in unemployment, lowering their standard of living and contentment. As a result, if COVID-19 becomes more contagious, the happiness score should be reduced.

There are negative linear relationships between the index and the crime rate, unemployment rate, and inflation rate. As crime increases, people's sense of security erodes, leading to a loss of faith in their own country. A higher unemployment rate means that more people are unemployed for a variety of reasons, including the recession or inability to work. Unemployed individuals will worry about their financial stability and will feel restless as a result of their loss of money, lowering their degree of enjoyment. Increased inflation reduces citizens' purchasing power. Citizens will be concerned about their financial well-being and consumption. They tend to save more money by reducing spending, which lowers their quality of life and consequently their degree of pleasure.

From Section 3.2 and 3.4.1, Oceania has the highest average and median happiness scores, followed by Europe, America, and Asia. However, because Oceania only has two countries, there is little data and it is questionable whether it deserves to be ranked. Africa has the lowest values in all of the 7 continents. Additionally, Oceania has the highest average in all categories except healthy life expectancy, followed by Europe, North America, and South America, all of which have comparable averages in terms of happiness and the other seven criteria. Finally, Asia and Africa are ranked worst in all factors.

Comparing the happiness rankings of the top 20 highest- and lowest-ranked countries, we discovered that the descending order association between happiness scores and its primary seven components is significantly different in these two groups. Both the orders are diametrically opposed to the order of the top-most significant correlations discovered in the first objective.

To summarize, countries can improve their rankings by focusing on socio-economic factors. It will improve inhabitants' quality of life through improved social support and economic development. Apart from socio-economic considerations, countries should prioritize public health, specifically reducing the spread of contagious diseases, in order to alleviate residents' dread. Citizens who have a higher quality of life will be happier, resulting in a higher happiness score.

5. Appendices

The R code and comments for this report are published:
https://gist.github.com/wibubunbo/c47bb19bddb8dde10e736e2eca010227

6. References

About crime indices at this website. (n.d.). Retrieved April 11, 2022, from https://www.numbeo.com/crime/indices_explained.jsp

Changing world happiness. (n.d.). Retrieved April 11, 2022, from https://worldhappiness.report/ed/2019/changing-world-happiness/

Happiness, benevolence, and trust during COVID-19 and beyond. (n.d.). Retrieved April 11, 2022, from https://worldhappiness.report/ed/2022/happiness-benevolence-and-trust-during-covid-19-and-beyond/

Indicator metadata registry details. (n.d.). Retrieved April 11, 2022, from https://www.who.int/data/gho/indicator-metadata-registry/imr-details/66

Inflation rate—Countries—List. (n.d.). Retrieved April 11, 2022, from https://tradingeconomics.com/country-list/inflation-rate

OECD. (2020). *Social support*. Organisation for Economic Co-operation and Development. https://www.oecd-ilibrary.org/social-issues-migration-health/lack-of-social-support/indicator/english_0cfbe26f-en

Population density (People per sq. Km of land area) | Data. (n.d.). Retrieved April 11, 2022, from https://data.worldbank.org/indicator/EN.POP.DNST

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D., & Roser, M. (2020a). Coronavirus pandemic(COVID-19). *Our World in Data*. https://ourworldindata.org/covid-cases

Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D., & Roser, M. (2020b). Coronavirus pandemic(COVID-19). *Our World in Data*. https://ourworldindata.org/covid-deaths

Suicide rate by country 2022. (n.d.). Retrieved April 11, 2022, from https://worldpopulationreview.com/country-rankings/suicide-rate-by-country

Unemployment rate—Countries—List. (n.d.). Retrieved April 11, 2022, from https://tradingeconomics.com/country-list/unemployment-rate