Analysis of the 2017 World Happiness Report

Darrel Bott

Table of Contents

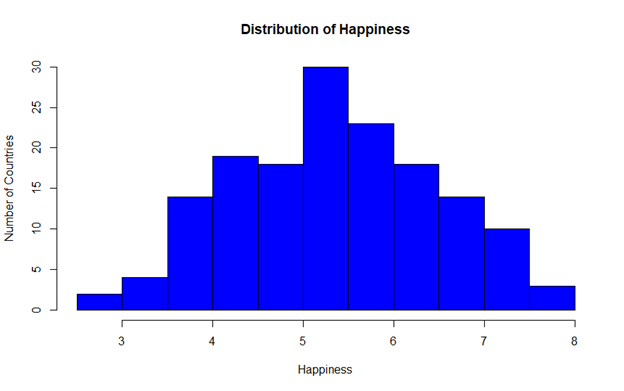
1. Introduction Page 2
2. Graphical Display and Summary Statistics Page 3
3. Statistical Analysis and Inference Page 9
4. Correlation and Regression Page 20
5. Conclusion Page 20

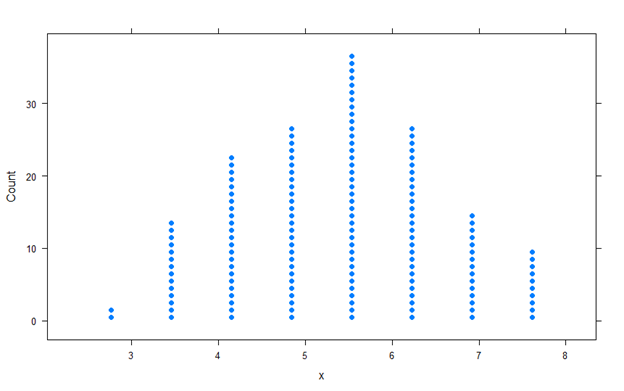
**Introduction**

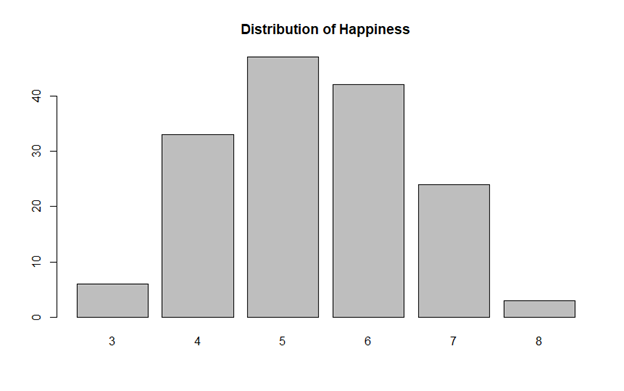
For something that has begun in 2012, world happiness has increasingly become an important topic in the United Nations, to the point where a world happiness day has been established, which is observed on March 20th. There are many factors that can contribute to this world happiness which includes the economy, family life, health/life expectancy, freedom, generosity, and trust in the government. The data adds the scores of each category to determine the overall happiness of the country. The countries with the lowest score in each category are considered to mimic a dystopian society, which is a society that contains the world’s unhappiest people. The countries with the highest score in each category are considered to mimic a utopian society, which is a ‘perfect’ society. Happiness is important to society because it measures the progress of the society.

Our data shows the happiness scores form 156 countries in ascending order. It includes the scores for the economy, family, health/life expectancy, freedom, generosity, and government trust. This data was found from the 2017 Gallup poll “World Happiness Report”. We are going to use the variables to find and describe patterns to the happiness and to see if the geographical location plays a key role to the countries happiness. All individuals were told to rate each factor from 0 to 10, zero being worst possible life and 10 being the best way to live. The sample data includes about 3000 people from each country, and the citizens are rating themselves. Even though everyone’s lives are different, the data shows the average of what the citizens have ranked the variables.

**Quantitative Variable: Happiness Score**



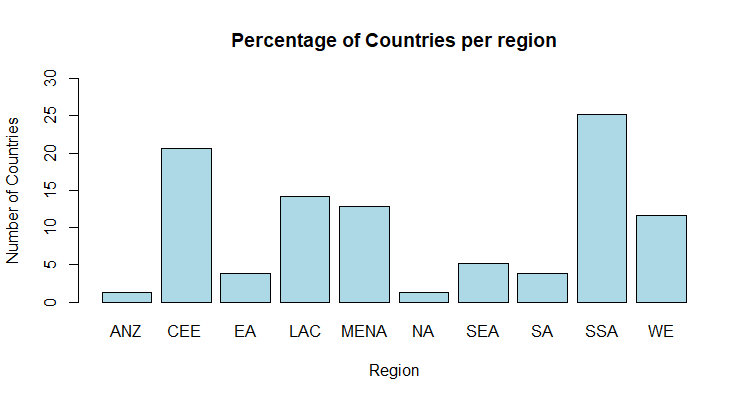


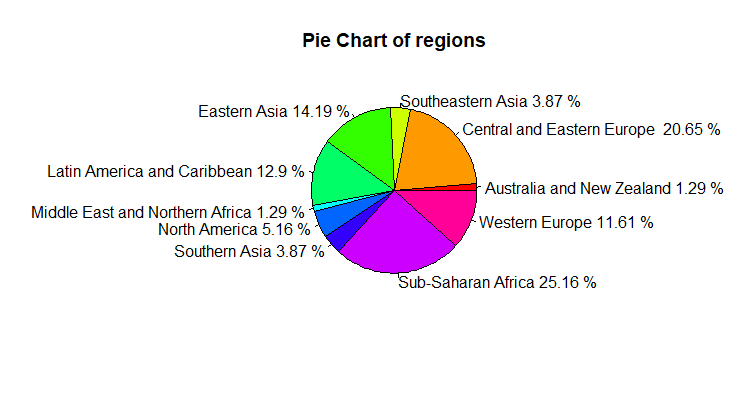


The distribution of the happiness score is symmetric, where the mean happiness score is close to 5.5.

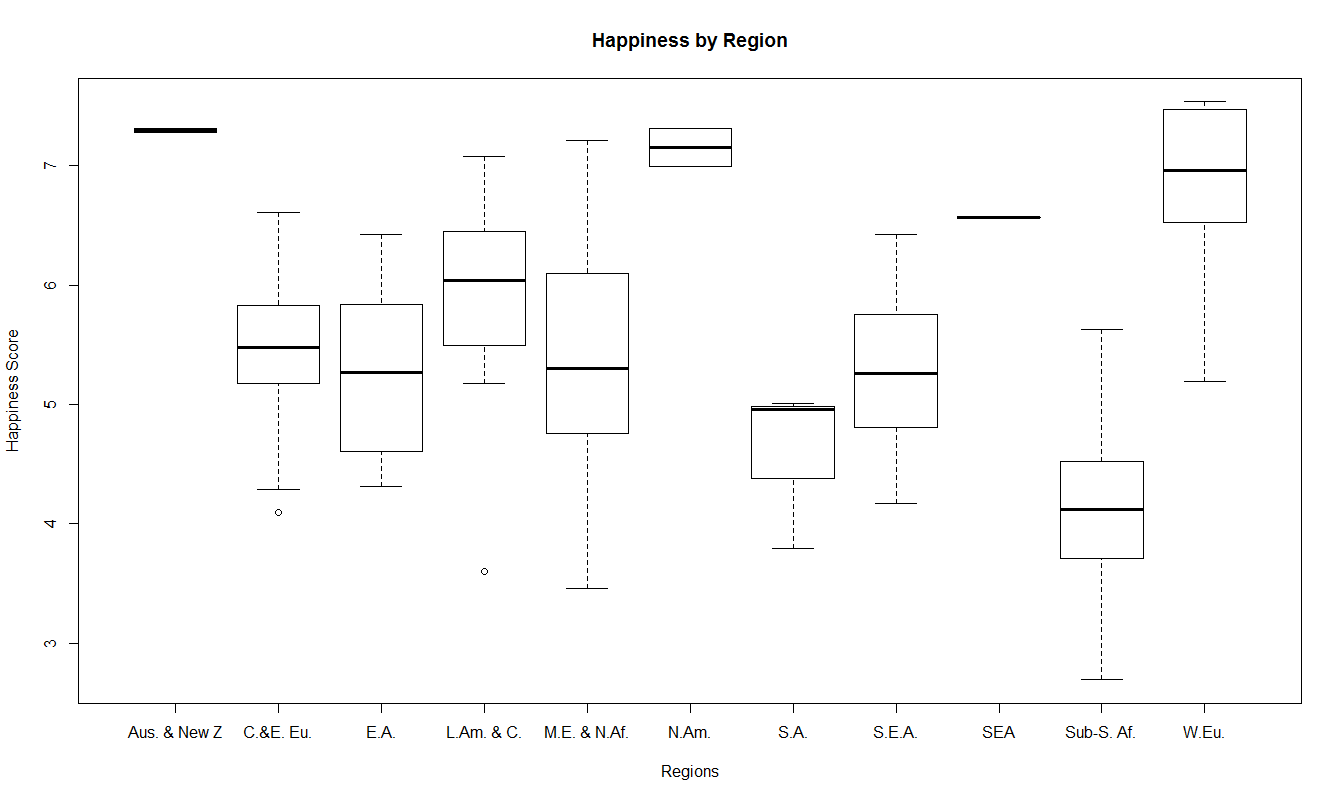
**Categorical Variable: Region**

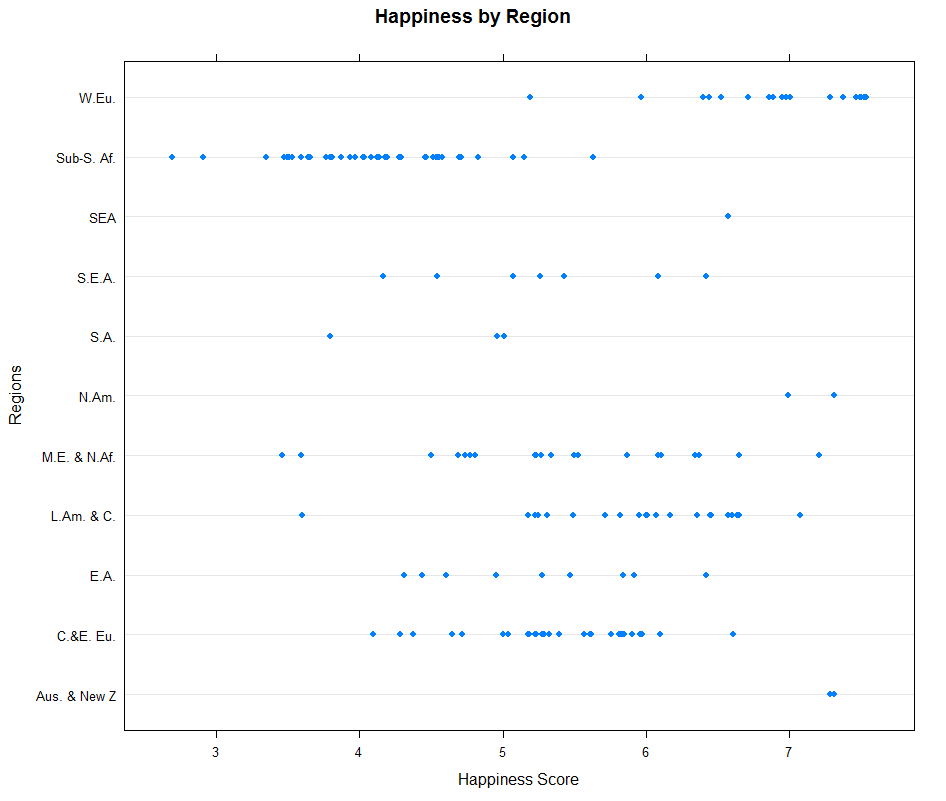
|  |  |  |
| --- | --- | --- |
| Country | Count | Percent |
| Australia and New Zealand | 2 | 1.29 |
| Central and Eastern Europe | 32 | 20.64 |
| Eastern Asia | 6 | 3.87 |
| Latin America and Caribbean | 22 | 14.19 |
| Middle East and Northern Africa | 20 | 12.9 |
| North America | 2 | 1.29 |
| Southeastern Asia | 8 | 5.16 |
| Southern Asia | 6 | 3.87 |
| Sub-Saharan Africa | 39 | 25.16 |
| Western Europe | 18 | 11.61 |



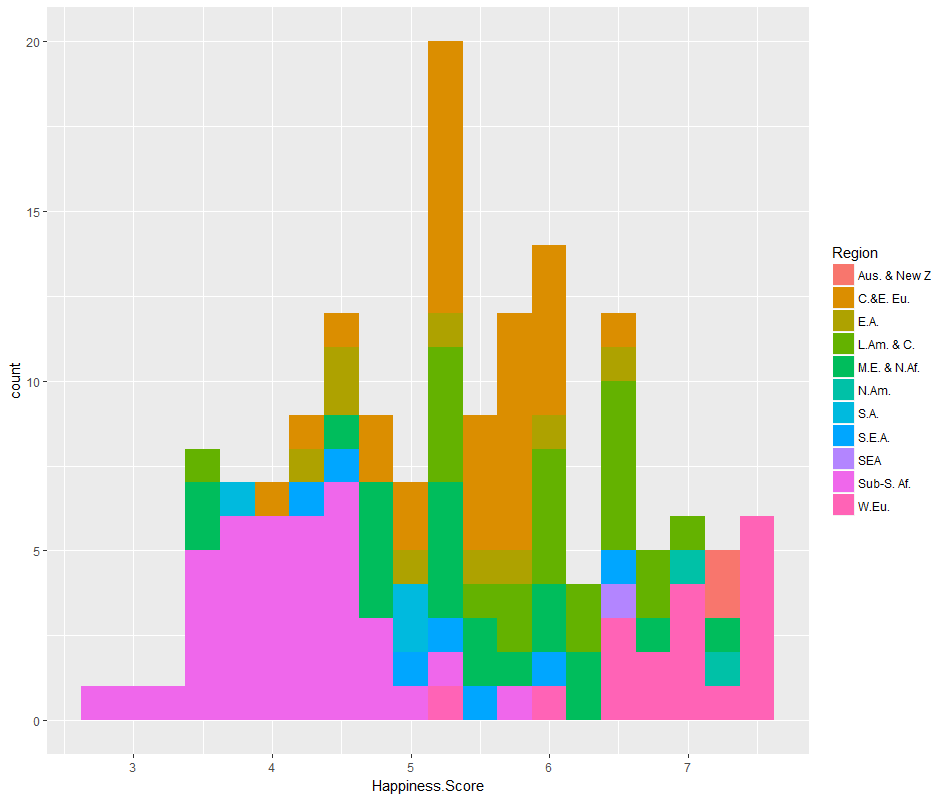


**Categorical and Quantative variable: Region and Happiness Score**





There appears to be a correlation which shows the relationship between regions. Most countries in a given region have similar happiness scores, and generally range between three units.



Happiness 95% confidence interval:

(5.174521, 5.533517)

With a 95% level of confidence we can say that the average happiness score for the world lies between 5.18 and 5.53.

The average that we calculated is 5.35.

To divide into two means, we decided to compare the world happiness by using two hemispheres, Northern and Southern. The Northern hemisphere includes Europe, Northern Africa and the Middle East, North America and Eastern Asia. The Southern hemisphere includes Sub-Saharan Africa, Southern Asia, South Eastern Asia, Latin America and the Caribbean, and Australia and New Zealand. When analyzing the happiness score between the two hemispheres, we use the following hypothesis:

Ho: Average happiness score of the Northern Hemisphere = Average happiness score of the Southern Hemisphere

HA: Average happiness score of the Northern Hemisphere > Average happiness score of the Southern Hemisphere

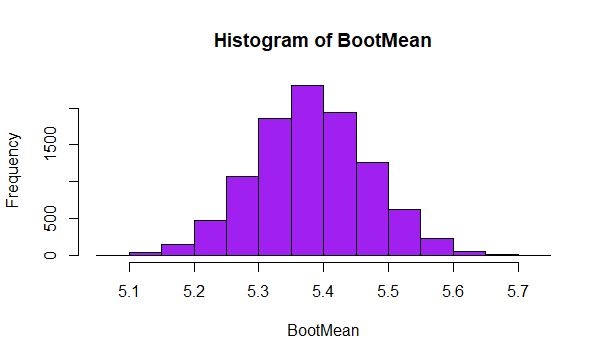
The average happiness score of the Northern Hemisphere is 5.78, while the average happiness of the Southern Hemisphere is 4.89. Using a two-sampled t-test we calculate a p-value of 0.81 \* 10^-7. Based on this result, we derive the following conclusion:

Using a 95% confidence level, there is evidence to conclude that the Northern Hemisphere is generally happier than the Southern Hemisphere. Therefore, we reject the null hypothesis that the both hemispheres are equally happy.

**Bootstrap with one mean**

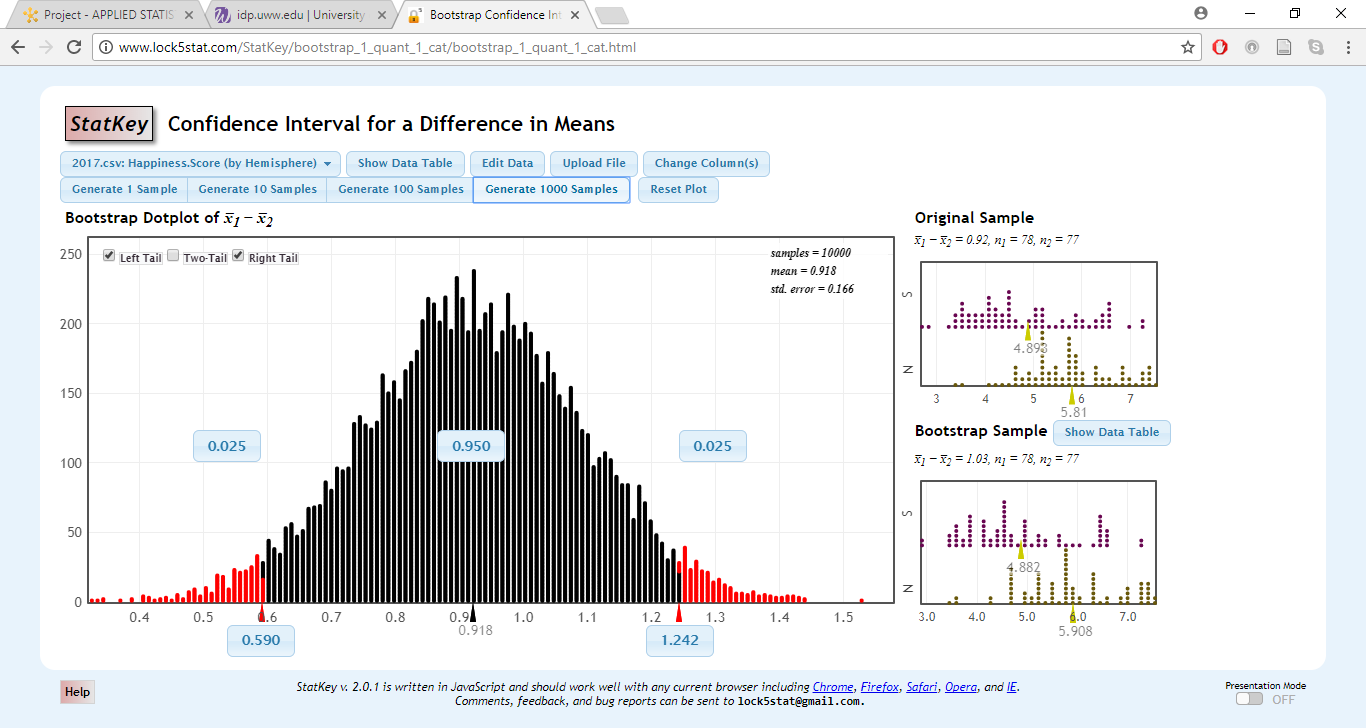
Happiness score 95% confidence interval:

(5.177 – 5.536)



Using a 95% level of significance, the average happiness score of the world lies between 5.177 and 5.536

**Bootstrap with 2 means**

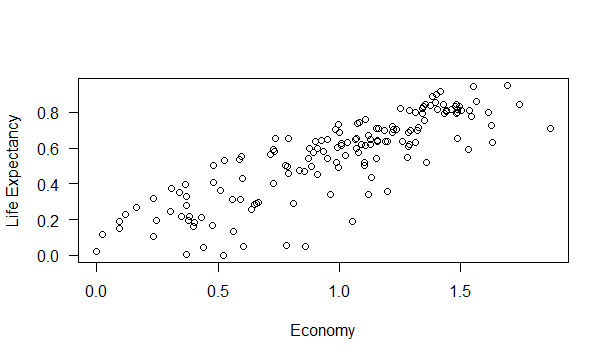


Confidence interval .590 – 1.242

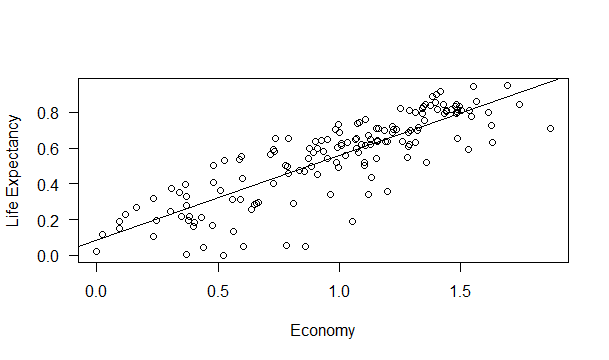
Mean = 0.918

The lower limit of the confidence interval with bootstrapping increased slightly by about .003 from the confidence interval found without bootstrapping. The upper limit of the confidence interval with bootstrapping also increased by .003 from the confidence interval found without bootstrapping.

The variables being tested are life expectancy and economy. This is important for answering the question, “Do people life longer if they have more money?” The population that this data is representing would be the citizens from 155 countries. The following scatter plots will show the trend, and any outliers that may be present.



The scatterplot shows a strong correlation between the economy and life expectancy of citizens of a specific country. There are a few outliers where the economy is around 0.7-0.8 and the life expectancy is close to 0. There appears to be a positive correlation and without calculating the correlation the coefficient looks to be around 0.75-0.80.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| n | median | 1st Quartile | Mean | 3rd Quartile | Maximum |
| 155 | 1.0646 | 0.6634 | 0.9847 | 1.3180 | 1.8708 |

The above table shows the summary statistics for the economy

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| n | median | 1st Quartile | Mean | 3rd Quartile | Maximum |
| 155 | 0.6060 | 0.3699 | 0.5513 | 0.7230 | 0.9495 |

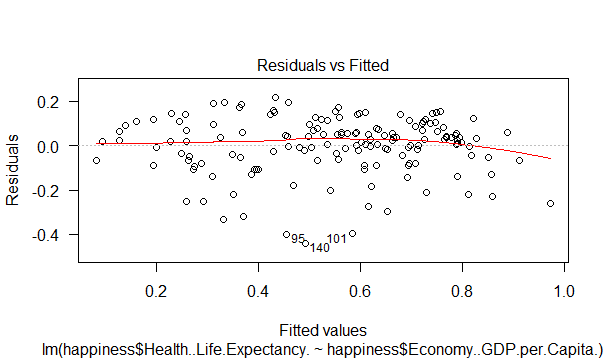
The above table shows the summary statistics for the average life expectancy

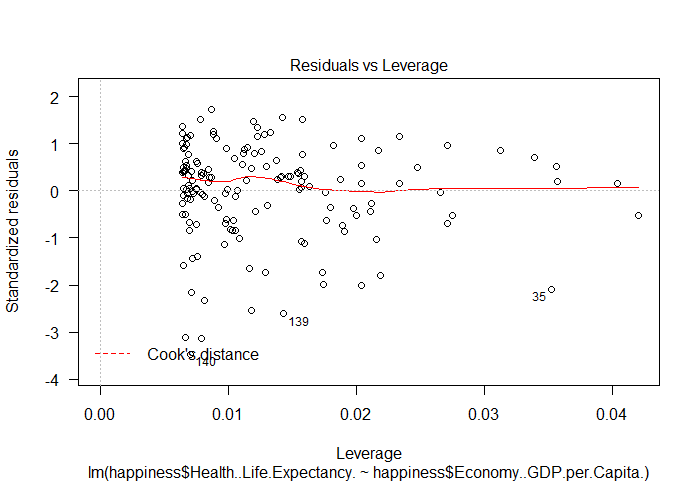
The null hypothesis being tested states that there is no correlation between the economy and average life expectancy.

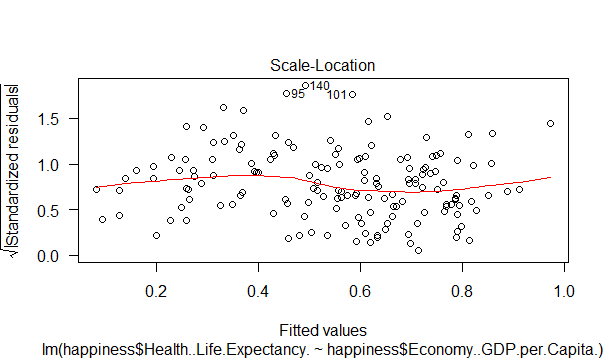
The alternative hypothesis states that there is a correlation between the economy and average life expectancy

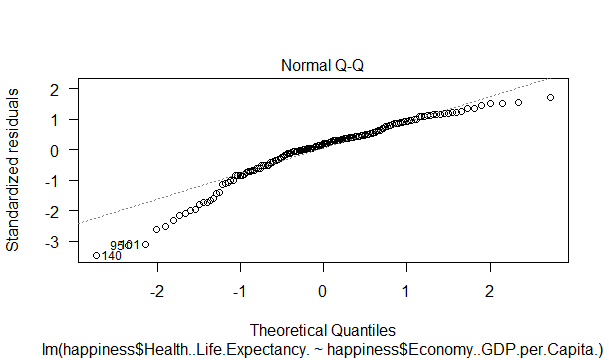
After calculating the correlation between the economy and the average life expectancy, we have found it to be 0.843, which confirms that there is a relationship between the economy and a persons’ life expectancy. Which in turn allows us to form the following conclusion:

With a 95% significance level and a p-value of 0, we found that there is a correlation, as the economy grows stronger, the average life expectancy increases. Therefore, we reject the null hypothesis.



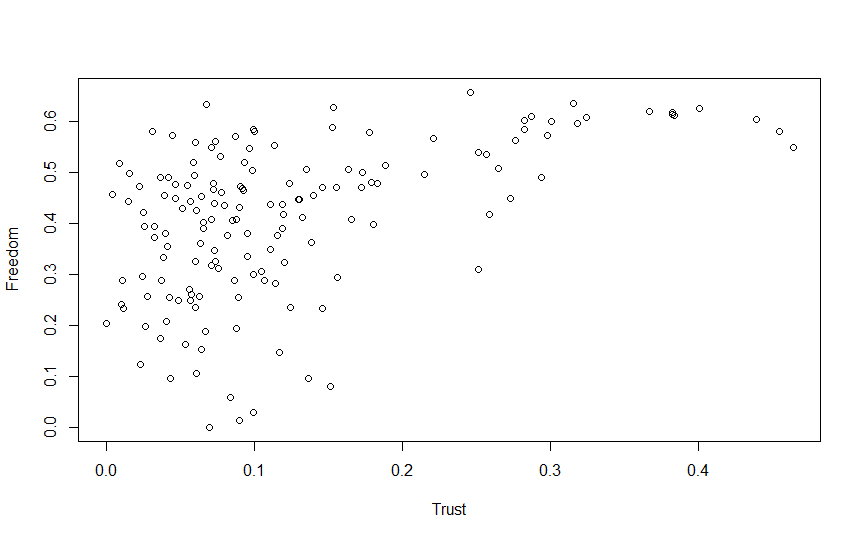


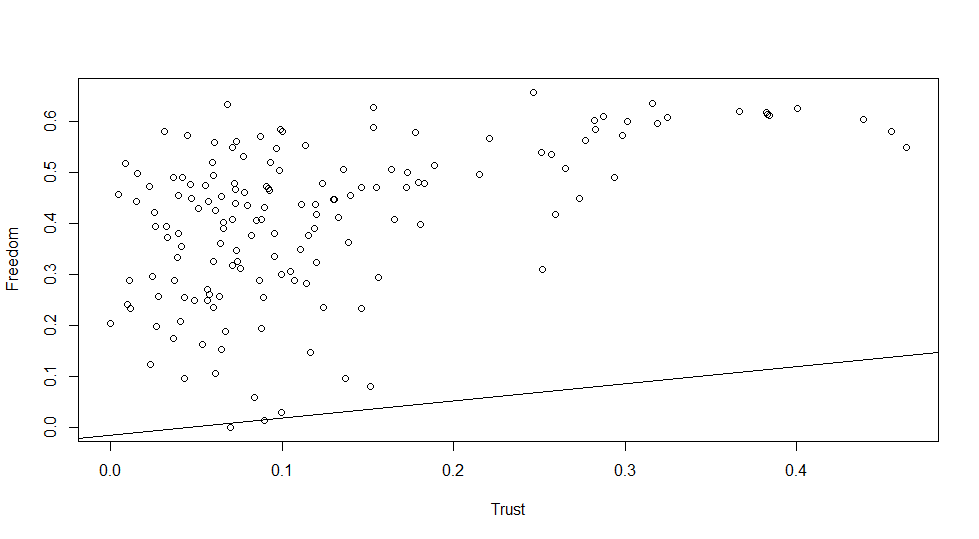




|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F Value | p |
| Freedom | 1 | 6.1520 | 6.1520 | 376.01 | 2.2\*10^-16 |
| Residuals | 153 | 2.5033 | 0.0164 |  |  |

Below is the scatterplot made for the variables trust and freedom as marked on the axes. This answers the question that if a country trusts their government, does it make them more free? The trend that this plot is showing is exponential. This means that there is as a country trust for a government grows, their freedom score also grows, but only up to a specific point. Once a country’s trust score reaches around 0.2, the trust score keeps growing while the freedom score evens out be between 0.5 and 0.6. The interesting part about trust was that the scores for trust were completely unexpected. For example, Rwanda has one of the lowest happiness scores, but they have the highest trust score. Also, the United States is not in the top 10 for trust, but they have one of the highest happiest scores.





Below are the five statistic summaries of both variables. The first is trust:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N | Mean | Minimum | 1st Quartile | Median | 3rd Quartile | Maximum |
| 155 | 0.123 | 0.0 | 0.057 | 0.0899 | 0.1533 | 0.464 |

The second is freedom:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N | Mean | Minimum | 1st Quartile | Median | 3rd Quartile | Maximum |
| 155 | 0.4088 | 0.0 | 0.3037 | 0.4375 | 0.5166 | 0.6583 |

The regression model to predict trust is given by:

T = -0.01518 \* 0.3383F

T denotes Trust while F denotes Freedom

While freedom increases by one unit, trust then increases by 0.3383.

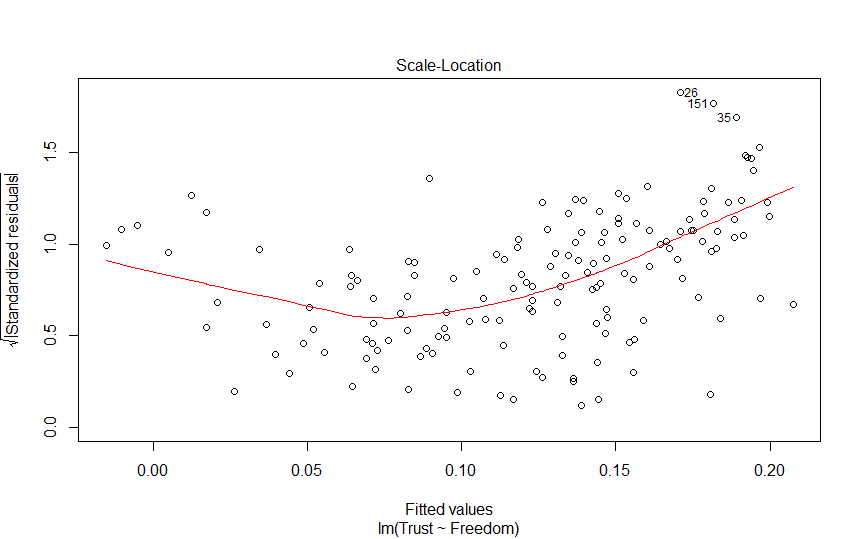
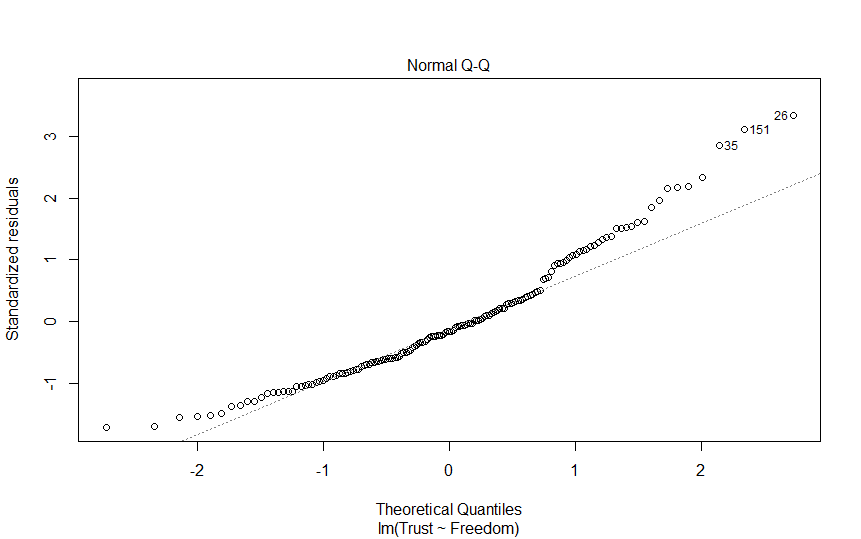
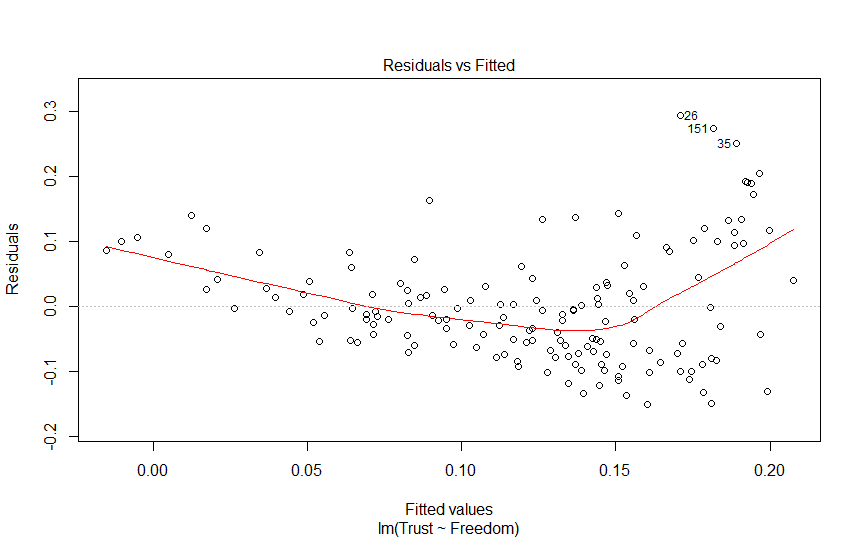
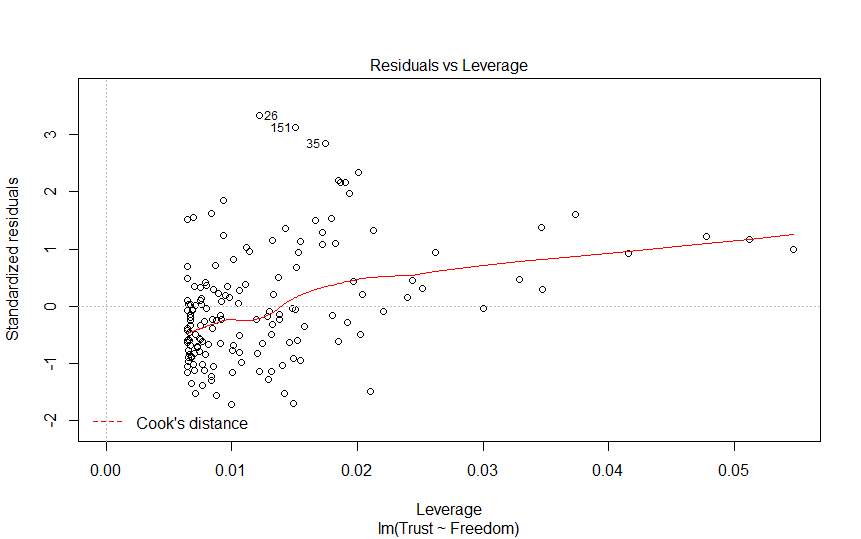
We used the following hypothesis to test the correlation between trust and freedom:

H0: cor = 0

HA: cor ≠ 0

cor is the correlation between trust and freedom. The population that this hypothesis is covering would be 155 countries put together to represent the population of the world. After calculation the correlation, we computed 0.499 and we got a p-value of 3.81 \* 10^-11, which helps us to derive the following conclusion:

Using a 95% level of confidence, we have evidence to support that there is a relationship between trust and freedom. Therefore, we reject the null hypothesis that a relationship does not exist.



Above are the residual plots. For ANOVA we get the following output table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Df | Sum Sq | Mean Sq | F Value | p |
| Freedom | 1 | 0.3965 | 0.3965 | 50.78 | 3.81\*10^-11 |
| Residuals | 153 | 1.19498 | 0.00781 |  |  |

**Conclusion**

In conclusion, throughout this analysis we looked at the average happiness score between, and individual country, region and hemisphere. Our first null hypothesis stated that the average happiness score of the Northern and Southern Hemispheres were equal. After our calculations we found that the Northern Hemisphere was generally happier than the Southern Hemisphere. We dug further into the data to look for correlations between the variables that effect the happiness score, and the strongest correlation that we found was between life expectancy and economy. This may be due to the fact that countries that have stronger economies will have more money to spend on medicine, food and technology to help keep people healthier. The r-value we found for this correlation was 0.843, which shows to be a strong correlation. This proved our null hypothesis wrong, which was that there was no correlation between life expectancy and economy. This analysis helped us to understand how countries are ranked and what is involved when deciding how happy a country is. The most interesting fact was that not one country had a perfect score in any of the categories. Can full happiness actually be achieved?

Data obtained from: <https://www.kaggle.com/unsdsn/world-happiness>