Analysis of the Factors Affecting World Happiness

By: Devanshi Pratiher

Background

Happiness is one of the most driving and important reasons to be alive. Happiness is something everyone desires and strives for. However it is important to note that many factors affect happiness.

In this report, I will use the World Happiness Reports from different years to ask important questions and analyze results that would hopefully answer these questions.

The concept of the World Happiness Report was founded in the United Nations. The UN General Assembly adopted a resolution where country members were assembled and invited them to measure the happiness of their people. This data would then be used to create new policies that would help increase people's happiness in their states.

The first Happiness Report was created in 2012 and summarized how states were happy, causes why people were happy and sad, and different policies that could affect these factors. The Happiness Reports have been annual

Description of Data

There are 5 main datasets that I have used for my research and analysis. They are all Happiness World Reports. I used the Happined World Reports of 2015, 2016, 2017, 2018, and 2019.

The variables in the dataset are provided in a manner that is easy to analyze. The first column is the Overall Rank of the Country or region in Happiness. The second column is the Country or region. The third column is the Happiness Score and the rest of the variables are different factors that affect the Happiness Score and Rank of each country/ region. Here is a screenshot of the 2019 dataset:

2019

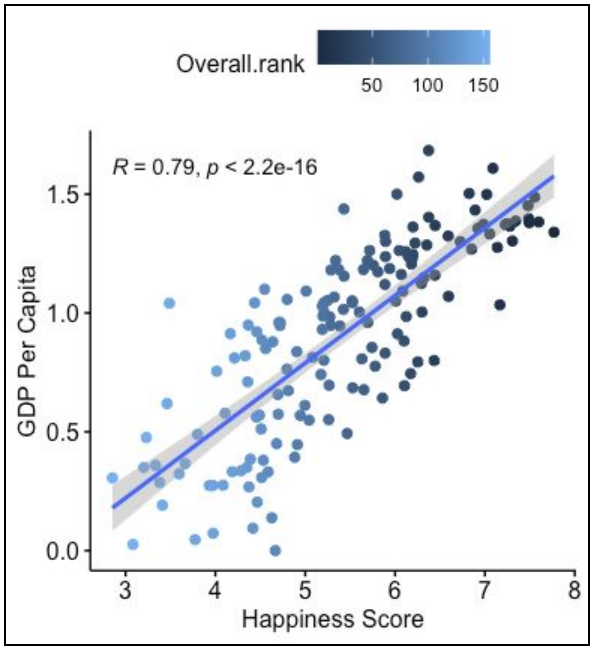
Overall rank	Country or region	Score	GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	Generosity	Perceptions of corruption
1	Finland	7.769	1.340	1.587	0.986	0.596	0.153	0.393
2	Denmark	7.600	1.383	1.573	0.996	0.592	0.252	0.410
3	Norway	7.554	1.488	1.582	1.028	0.603	0.271	0.341
4	Iceland	7.494	1.380	1.624	1.026	0.591	0.354	0.118
5	Netherlands	7.488	1.396	1.522	0.999	0.557	0.322	0.298
6	Switzerland	7.480	1.452	1.526	1.052	0.572	0.263	0.343
7	Sweden	7.343	1.387	1.487	1.009	0.574	0.267	0.373
8	New Zealand	7.307	1.303	1.557	1.026	0.585	0.330	0.380
9	Canada	7.278	1.365	1.505	1.039	0.584	0.285	0.308
10	Austria	7.246	1.376	1.475	1.016	0.532	0.244	0.226
11	Australia	7.228	1.372	1.548	1.036	0.557	0.332	0.290
12	Costa Rica	7.167	1.034	1.441	0.963	0.558	0.144	0.093
13	Israel	7.139	1.276	1.455	1.029	0.371	0.261	0.082
14	Luxembourg	7.090	1.609	1.479	1.012	0.526	0.194	0.316
15	United Kingdom	7.054	1.333	1.538	0.996	0.450	0.348	0.278
16	Ireland	7.021	1.499	1.553	0.999	0.516	0.298	0.310
17	Germany	6.985	1.373	1.454	0.987	0.495	0.261	0.265
18	Belgium	6.923	1.356	1.504	0.986	0.473	0.160	0.210
19	United States	6.892	1.433	1.457	0.874	0.454	0.280	0.128
20	Czech Republic	6.852	1.269	1.487	0.920	0.457	0.046	0.036
21	United Arab Emirates	6.825	1.503	1.310	0.825	0.598	0.262	0.182
22	Malta	6.726	1.300	1.520	0.999	0.564	0.375	0.151
23	Mexico	6.595	1.070	1.323	0.861	0.433	0.074	0.073
24	France	6.592	1.324	1.472	1.045	0.436	0.111	0.183
25	Taiwan	6.446	1.368	1.430	0.914	0.351	0.242	0.097
26	Chile	6.444	1.159	1.369	0.920	0.357	0.187	0.056
27	Guatemala	6.436	0.800	1.269	0.746	0.535	0.175	0.078
28	Saudi Arabia	6.375	1.403	1.357	0.795	0.439	0.080	0.132
29	Qatar	6.374	1.684	1.313	0.871	0.555	0.220	0.167
30	Spain	6.354	1.286	1.484	1.062	0.362	0.153	0.079
31	Panama	6.321	1.149	1.442	0.910	0.516	0.109	0.054
32	Brazil	6.300	1.004	1.439	0.802	0.390	0.099	0.086
33	Uruguay	6.293	1.124	1.465	0.891	0.523	0.127	0.150

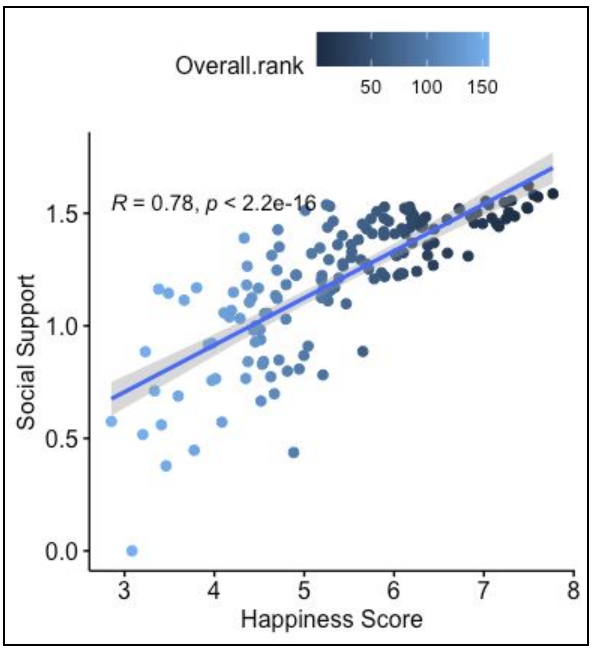
Questions and Analysis

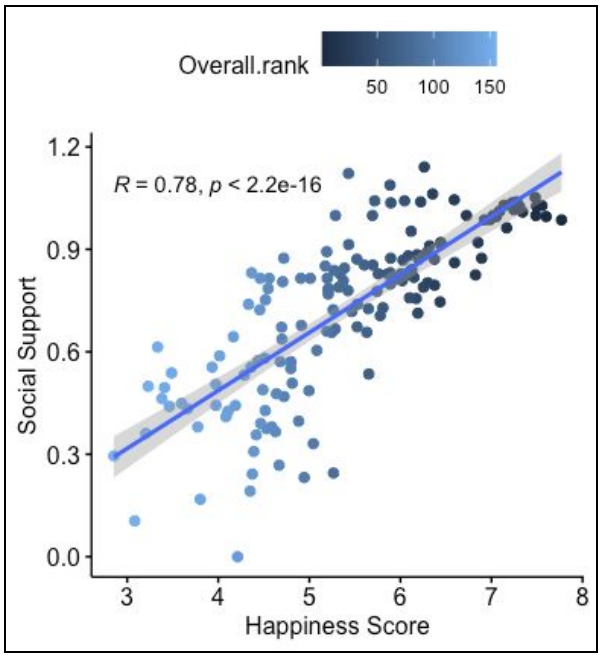
Q1: Which factor has the least correlation to the Happiness Score of a country?

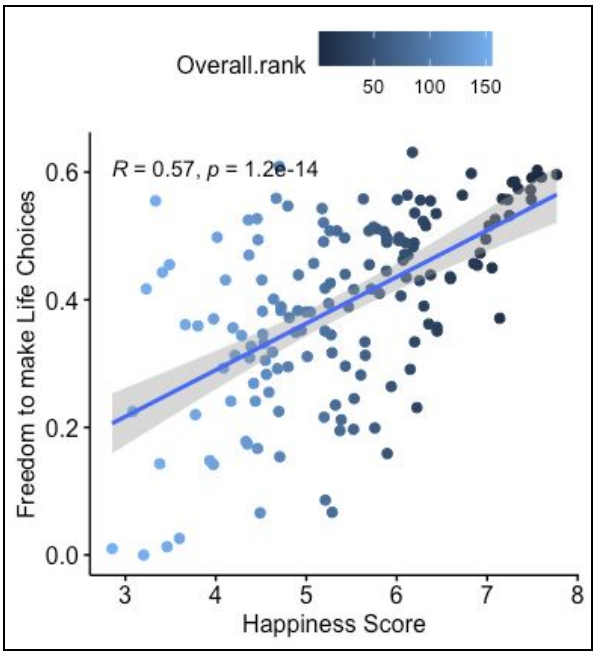
The Happiness Score of a country is determined by many different variables as seen by the Happiness Report. However, I wanted to see out of all the factors, which ones were the least related to affecting the happiness score, i.e had close to no correlation with the happiness score. This question is interesting to the data because it shows us which variable is least correlated with the happiness score.

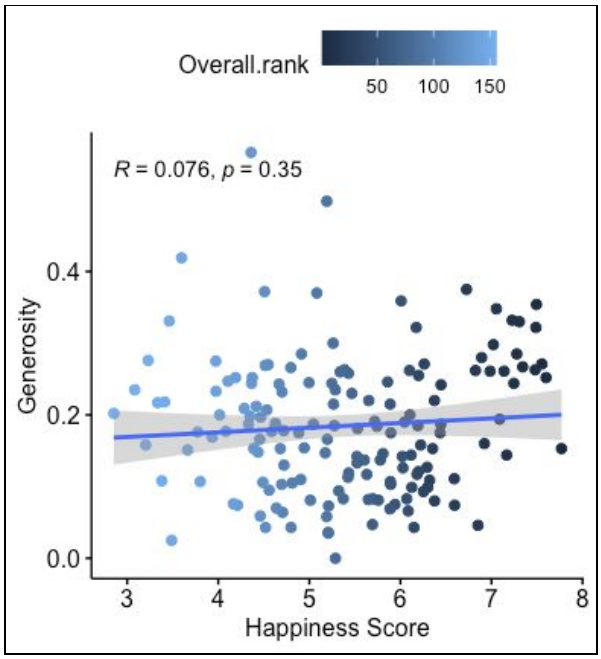
I performed correlation coefficient tests to test each factor with the happiness score and plot them. This method is appropriate as we can analyze the relation between two variables and find out how correlated they are based on coefficients. If the values are less than 0.70, this shows that the variables do not have a significant correlation. This would help me understand the relations between the factor and the Score well. I also used ggplot to help plot the correlations of each factor to the happiness score. This helped me look at the correlations and analyze how each factor ties into the happiness score.

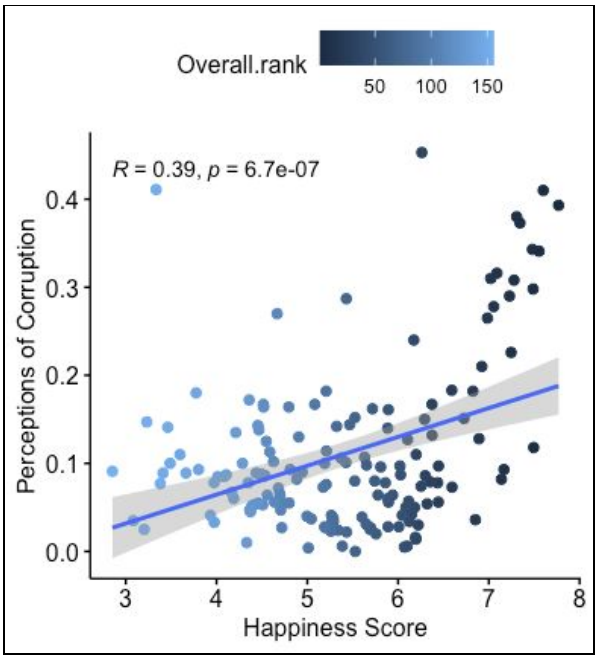








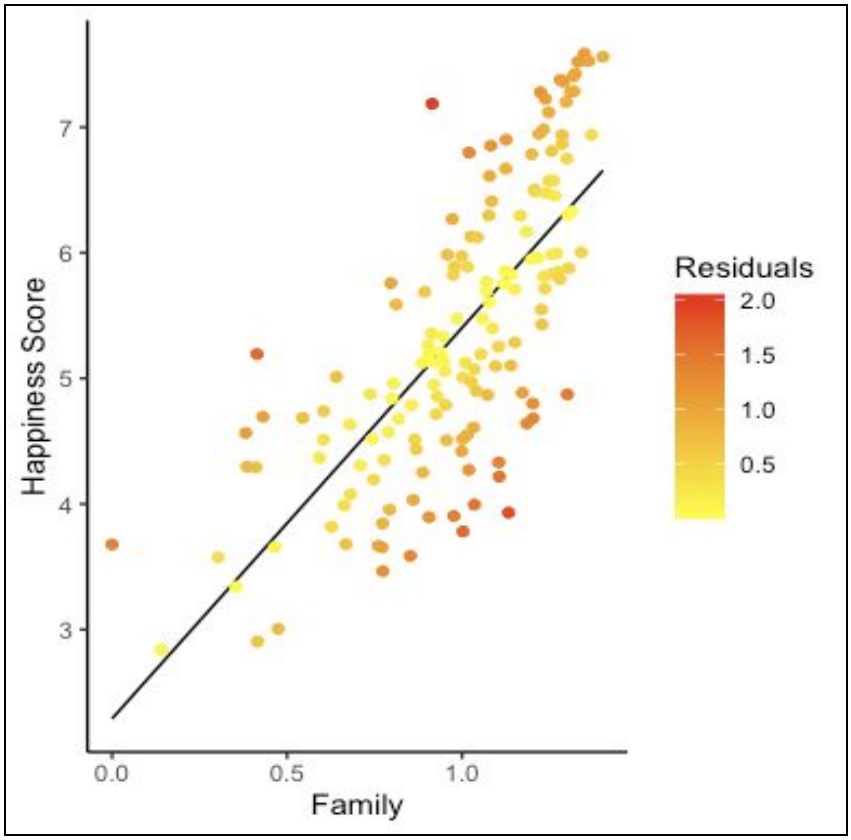




From the above plots and the correlation tests, we see that the factor that has little to no correlation with the happiness scores of the countries is Generosity. All the factors have p-value < 0.05, which shows significance and a positive correlation. However, the p-value of Generosity is 0.35 which is greater than alpha 0.05 and the correlation coefficient is 0.07582369. These two reasons help us conclude that Generosity shows little to no correlation with the Happiness Scores.

Q2: Can we predict if having a family increases Happiness?

In the 2015 World Happiness Report, there was a factor that wasn't present in the latest report- Family. This factor accounted for the World Happiness Score and Rank and I wanted to visualize if having a family would increase one's chance of happiness. I used a linear regression model to approach this question and visualize whether this was true. I wanted to use this method as I could see how the increasing values of Family would affect the Happiness Score and this model was a good fit for that.

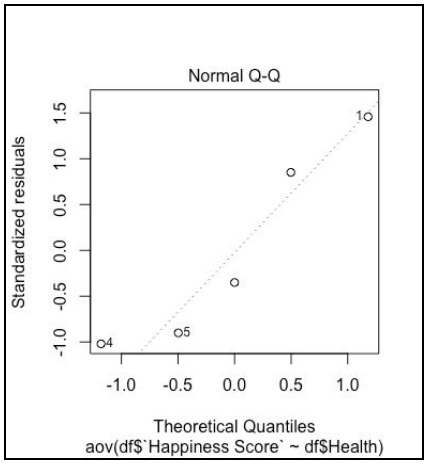


The scatterplot shows that Family does have an affect on the Happiness Score. Families have a greater chance of being happy than non-family members.

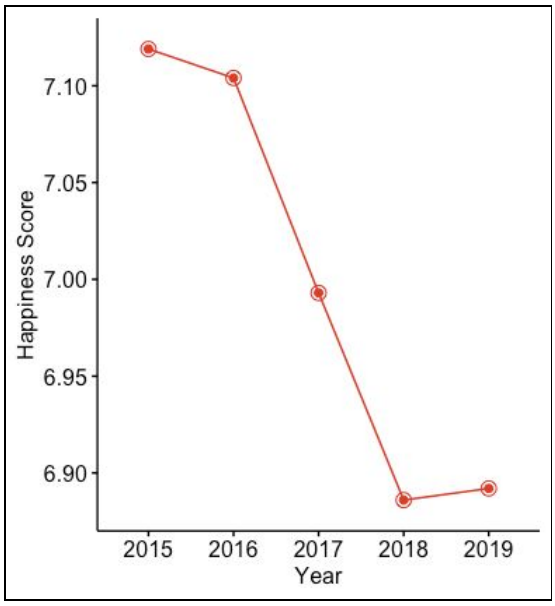
Q3: Is the change of the Happiness Scores in the United States significant with the Health and Life Expectancy Scores of the US throughout the years?

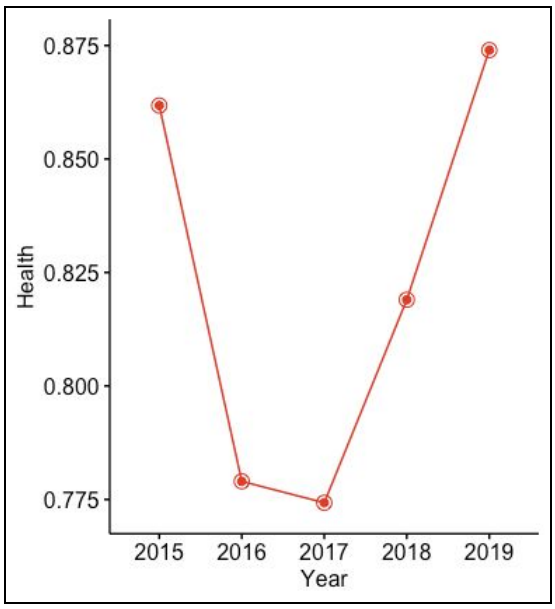
For this test I wanted to take the Happiness Score and the Health/ Life Expectancy of the United States throughout the Years- (2015-2019). Using the scores of the health each year and comparing it with the happiness scores gave me an understanding of whether there was any significance between these scores.

I used an ANOVA test to check whether there was a significant difference between these two variables. I wanted to use this test because it would be the best fit to compare the variance of different variable values of the US. I checked the normality and checked if the residuals are normally distributed.



I also made a plot to show the Health Scores of the United States throughout the years and another plot that shows Happiness Scores of the U.S. throughout the years.





ANOVA Test:

```
> res.aov <- aov(df$`Happiness Score` ~ df$Health, data = df)
> summary(res.aov)
             Df  Sum Sq Mean Sq F value Pr(>F)
df$Health     1 0.00260 0.00260   0.166  0.711
Residuals     3 0.04708 0.01569
```

Based on the ANOVA Test, the p-value (0.711) is not less than the significance level 0.05, we can conclude that there are no significant differences between the two groups (Health and Happiness Scores).

There is no significant difference and based on the line plots there does not seem to be any trends to indicate Health is solely creating a significant different in the factor for Happiness Score in the U.S.

Conclusion:

Overall, I learned a lot about the different major factors that affect the Happiness Scores of different Countries and Regions. This was important as it helped me understand how different countries have different strengths and weaknesses. This report is important to decide policies in the world and it makes me happy that the UN has decided to collect the data to create these data sets annually.

I learned how there were many different important factors that are correlated to create happiness scores. Generosity is one of the variables that do not have a significant correlation to the Happiness Score. I also know that Family is a strong predicting factor of happiness. Finally, I learned that the United States' Happiness Scores and the United States' Health scores have no particular significance. This makes sense as there are many factors that go into calculating the happiness score.

Sources:

Data Source: https://www.kaggle.com/unsdsn/world-happiness?select=2019.csv

Background: https://worldhappiness.report/ed/2020/#read