Rithvik Varma Datla
Zac Byrd
Lauren McLeod
Sotheara Narith

The World Happiness Report

The World Happiness Report is a landmark survey of the state of global happiness. The raw data provided was a list ranking 155 countries from happiest to least happy. Although they have released yearly reports since 2012, our group focused on the years 2015 to 2018. The report has gained global recognition as governments, organizations and civil society increasingly use happiness indicators to inform their policy-making decisions. This information is also used by leading experts across fields like economics, psychology, survey analysis, national statistics, health, public policy, etc. to assess the progress of nations based on the measurements of well-being.

We selected this data set for our project because it posed an interesting question: What makes a country happy? The Gallup World Happiness Survey asked participants to rank their current lives on a ladder scale from 0 to 10. The columns following the happiness score estimate the extent to which each of the six factors – economic production, social support, life expectancy, freedom, absence of corruption, and generosity – contribute to a country's happiness score used for the ranking. The data from the years 2015 to 2018 showed both change and stability in the results from the pooled survey. This information shows us how the correlation between happiness and each of the six factors has changed overtime.

The initial problem we faced when processing data was choosing which years to include in our analysis. We decided to move forward with the years 2015 through 2018 because the data provided for 2019 didn't match up with the previous years. Additionally, there are four people in our group so it made sense for each of us to take up the analysis for one year. Our second processing problem was understanding what each column represented and interpreting the scores given. Through research and reading more into the Gallup World survey, we found that the categories are factors that influence the overall happiness score and the data provided is the extent each column influenced the score. The highest ranked countries generally scored higher in each of the 6 categories.

When we first looked at the dataset, we wanted to know about the relationship between the happiness score and the different factors that it comprises. One of the best ways to do this is to find out the correlation between the factors. To do this we used the various correlation methods on the dataset and the one that we chose to answer our question with was the Pearson correlation model. This model does have its advantages and disadvantages. Some of the pros are the model is able to indicate the exact extent to which the variables are correlated. This model allows us to determine the direction of the correlation, i.e, positive or negative. The method allows us to estimate the value of a dependent variable with reference to a particular value of an independent variable. This method also allows for an easier calculation of the correlation coefficient due to its algebraic properties. Some of the cons are : Extreme items have a large influence on the final results. It can be easily misinterpreted in homogeneous data. This method is based on a large number of assumptions such as linear relationships etc. An alternative method that we considered using was the spearman's rank correlation. We considered spearman's model as we felt that we could have used ranked values for finding the relationships between the variables but we felt that maybe the ranked values would mean that variables would change too much with each other.

Happiness Rank	Happiness Score	Economy (GDP per Capita)	Health (Life Expectancy)	Dystopia Residual
158.000000	158.000000	158.000000	158.000000	158.000000
79.493671	5.375734	0.846137	0.630259	2.098977
45.754363	1.145010	0.403121	0.247078	0.553550
1.000000	2.839000	0.000000	0.000000	0.328580
40.250000	4.526000	0.545808	0.439185	1.759410
79.500000	5.232500	0.910245	0.696705	2.095415
118.750000	6.243750	1.158448	0.811013	2.462415

2015

	Happiness Rank	Happiness Score	Economy (GDP per Capita)	Health (Life Expectancy)	Dystopia Residual
count	157.000000	157.000000	157.000000	157.000000	157.000000
mean	78.980892	5.382185	0.953880	0.557619	2.325807
std	45.466030	1.141674	0.412595	0.229349	0.542220
min	1.000000	2.905000	0.000000	0.000000	0.817890
25%	40.000000	4.404000	0.670240	0.382910	2.031710
50%	79.000000	5.314000	1.027800	0.596590	2.290740
75%	118.000000	6.269000	1.279640	0.729930	2.664650
max	157.000000	7.526000	1.824270	0.952770	3.837720

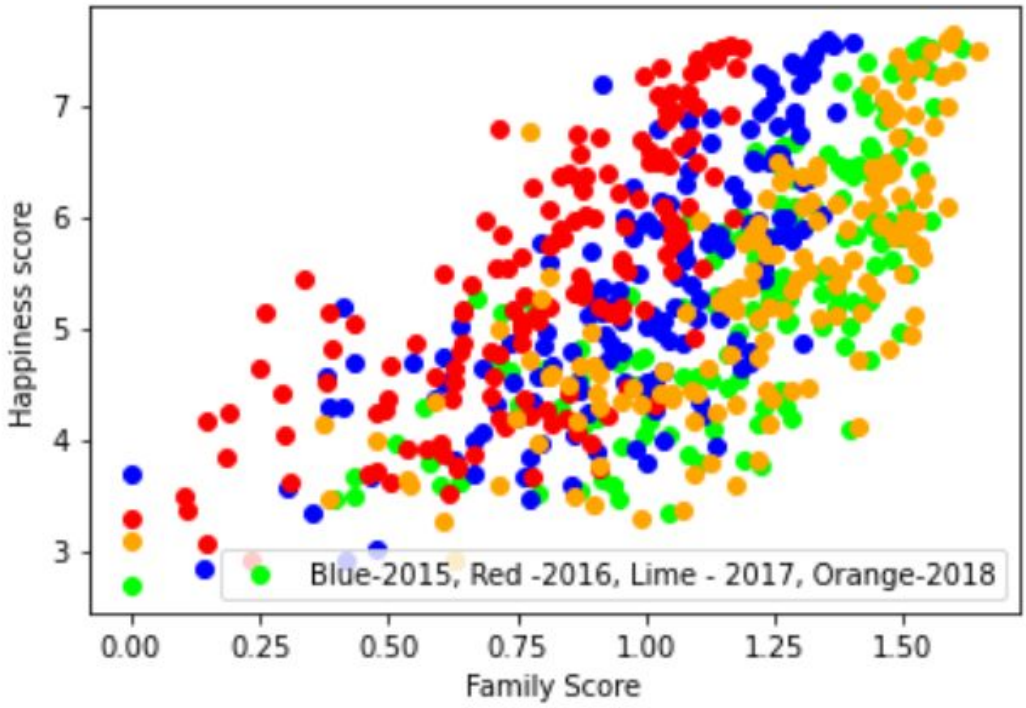
2016

	Happiness.Rank	Happiness.Score	Economy..GDP.per.Capita.	Health..Life.Expectancy.	Dystopia.Residual
count	155.000000	155.000000	155.000000	155.000000	155.000000
mean	78.000000	5.354019	0.984718	0.551341	1.850238
std	44.888751	1.131230	0.420793	0.237073	0.500028
min	1.000000	2.693000	0.000000	0.000000	0.377914
25%	39.500000	4.505500	0.663371	0.369866	1.591291
50%	78.000000	5.279000	1.064578	0.606042	1.832910
75%	116.500000	6.101500	1.318027	0.723008	2.144654
max	155.000000	7.537000	1.870766	0.949492	3.117485

2017

	Overall rank	Score	GDP per capita	Healthy life expectancy
count	156.000000	156.000000	156.000000	156.000000
mean	78.500000	5.375917	0.891449	0.597346
std	45.177428	1.119506	0.391921	0.247579
min	1.000000	2.905000	0.000000	0.000000
25%	39.750000	4.453750	0.616250	0.422250
50%	78.500000	5.378000	0.949500	0.644000
75%	117.250000	6.168500	1.197750	0.777250
max	156.000000	7.632000	2.096000	1.030000

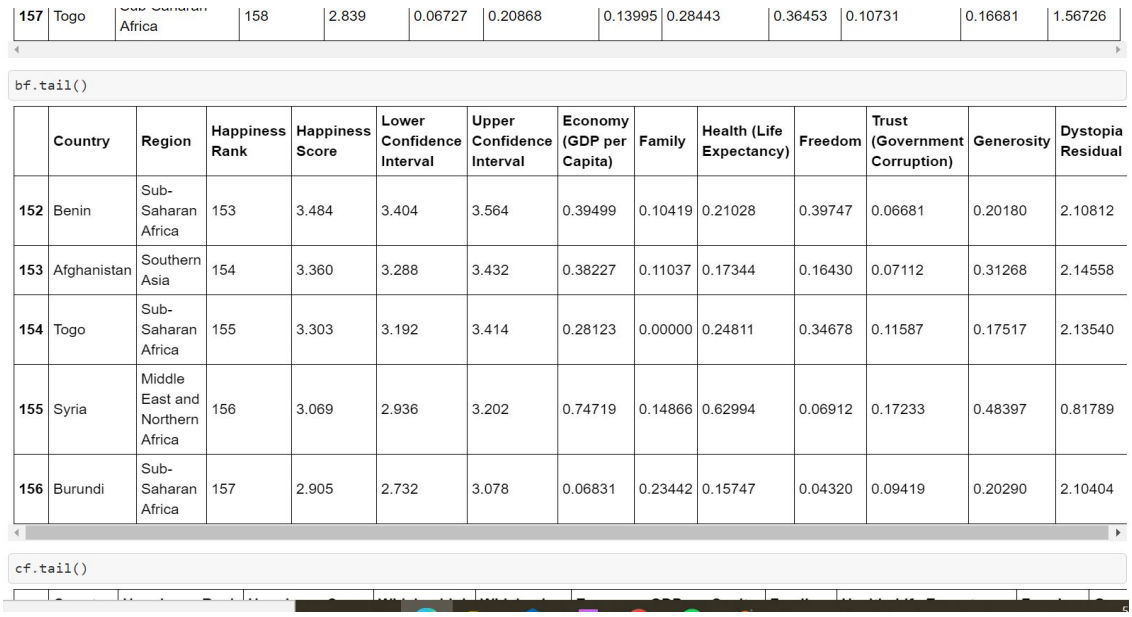
2018



There is a moderately strong correlation between the happiness score and the family scores. A trend that was noticed with countries towards the top had relatively high family scores. This coincided with the happiness scores of the top countries also being high. This is expected, as having more family support means that the person in general feels more at ease and happier in general. One country that is often near the top of the happiness rankings is Norway. The Scandanavian country was top of the charts in 2017 and a close second in 2018. The country is financially stable and offers its citizens a plethora of public services, for example, free healthcare and public education. The country is known to be very peaceful, with low crime rates in all major areas. Life expectancy throughout the country is high and the citizens are able to enjoy a lot of freedom.

On the other hand, a country that consistently finds itself towards, or at the bottom of the happiness ranking is Burundi, a country out of East Africa. In 2016 and 2018, Burundi ranked last among all the other countries polled for this dataset, and recorded the lowest feeling of freedom among its citizens. Burundi is a landlocked country in the Great Rift Valley of

Eastern Africa that has a major issue with their constant increase in population. The eastern African country is not on any travel recommendation list, as petty and violent crimes are common throughout the country. The sense of freedom among the citizens in Burundi is significantly lower than every other country reported on the Happiness Rank. The figure shown below is from the 2016 Happiness Report.



157	Togo	Sub-Saharan Africa	158	2.839	0.06727	0.20868	0.13995	0.28443	0.36453	0.10731	0.16681	1.56726	

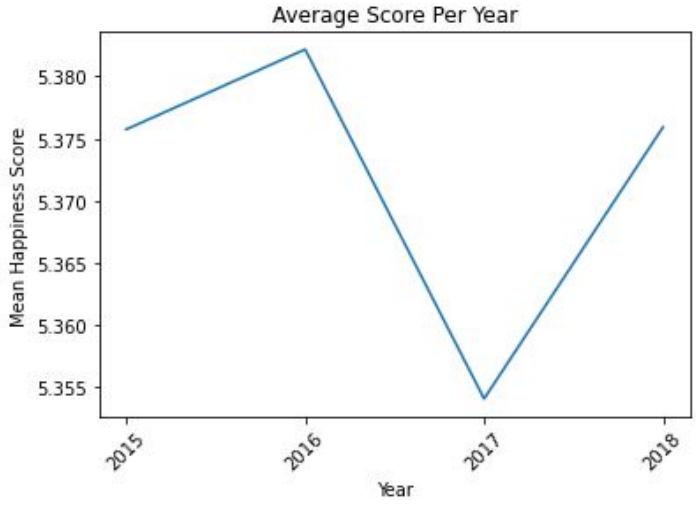
```
bf.tail()
```

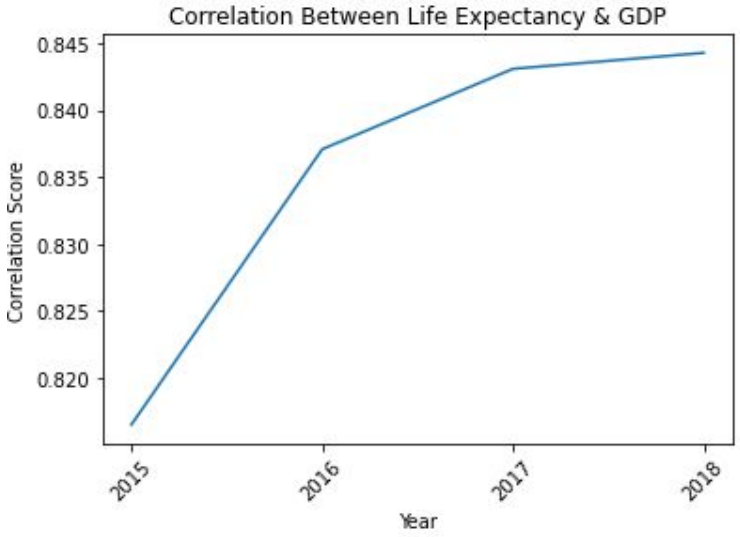
	Country	Region	Happiness Rank	Happiness Score	Lower Confidence Interval	Upper Confidence Interval	Economy (GDP per Capita)	Family	Health (Life Expectancy)	Freedom	Trust (Government Corruption)	Generosity	Dystopia Residual
152	Benin	Sub-Saharan Africa	153	3.484	3.404	3.564	0.39499	0.10419	0.21028	0.39747	0.06681	0.20180	2.10812
153	Afghanistan	Southern Asia	154	3.360	3.288	3.432	0.38227	0.11037	0.17344	0.16430	0.07112	0.31268	2.14558
154	Togo	Sub-Saharan Africa	155	3.303	3.192	3.414	0.28123	0.00000	0.24811	0.34678	0.11587	0.17517	2.13540
155	Syria	Middle East and Northern Africa	156	3.069	2.936	3.202	0.74719	0.14866	0.62994	0.06912	0.17233	0.48397	0.81789
156	Burundi	Sub-Saharan Africa	157	2.905	2.732	3.078	0.06831	0.23442	0.15747	0.04320	0.09419	0.20290	2.10404

```
cf.tail()
```

Some additional analyses we would have liked to factor into our final report is how the homeless rate of a country affects the happiness rank. Specifically, if areas with a higher homeless population rank lower on the happiness scale. For this we would need additional data that would provide what city the respondents live in, or the IP address of where the answers were recorded. For example, a happiness ranking of all the states in the United States, with emphasis on the major cities and capitals to determine whether homelessness has a major impact on a countries overall happiness score. Another piece we would have liked to do additional research on is how the happiness score for each country differs based on the age groups that respond. This would help in comparing whether different age groups are more or less satisfied with their country and living conditions. For this we would need the same data given before, but with additional text that includes the participants date of birth. From there

we would be able to categorize the respondents data by placing them in their proper age group.





The Happiness Rank was a very interesting and thought-provoking dataset to analyze for our final project. To answer our question of what makes the citizens of a country happy we decided to use correlation models to determine common variables among the dataset. By using the Pearson correlation model, we were able to determine exactly how closely related the happiness score was to other social and economical factors. Our group found that two of the most correlated variables from our dataset were ‘Health (Life Expectancy)’ and ‘Economy (GDP per Capita)’, which we were not surprised to find. These two variables were consistency

correlated through all four of the years we analyzed, and their correlation even increased in this time. Through our research, factors that result in a high ranking 'Happiness Score' are economy (GDP per capita) of the country, as well as health (life expectancy), and family life of the citizens.

Additional visuals and Jupyter notebook cells can be found at:

https://github.com/byrdzac30/Final-Project.git