

The Happiest Place on Earth?

Author: Christopher Bradway

Bellevue University

DSC 680 Fall 2020

cbradway72@yahoo.com

<https://github.com/cbradway72>

11-02-2020

Table of Contents

Abstract:.....	3
Introduction:	3
Report:	4
Data	4
Methods.....	5
Analysis	6
Conclusion:.....	10
References:	Error! Bookmark not defined.

Abstract:

This project investigates the 153 countries and based on six factors; economic production, social support, life expectancy, freedom, absence of corruption, and generosity, have been giving a happiness rating. I set out to try to isolate the most important and the least important factors based on this rating system. I will also isolate the region (country, continent) that tends to be the happiest place on Earth. My analysis will include a lot of visuals to help compare and isolate variables. I will be using Python for my analysis, although I may create a few visuals in R. This is being written pre-analysis, so I do not have any results to report, but I do know that usually countries in Europe fair well in the happiness rating. Overall, this seems like a pretty simple project with little substance, but due to the prevalence of depression in the world, the findings could help us to create happier environments. Knowing what factors play the biggest roles in the happiness of a nation can allow us to address issues we have within our own country and allow for non-chemical depression to become more of a thing of the past.

Introduction:

What defines happiness? The pursuit of happiness has been a goal of every human that has walked this Earth, but we still debate over what can make us happy. Many people think that it may be wealth, while others feel that it could come from personal health. The people over at the "World Happiness Report" seem to think that we can measure a country's happiness by isolating six variables; GDP, Health, Support, Generosity, Corruption, and Freedom. Their approach was to look at life from a social environment perspective and allow the individuals to provide feedback on how they feel they are doing in the six variable areas. I will explain the six variables in more detail later in this paper.

The World Happiness Report set out with the goal of trying to judge happiness from a scientific perspective, and in doing so chose the aspects that they felt best correlated with ones happiness, which would allow us to then rank a country's happiness based on the scores of the individual's perception of their quality of life in these six areas. Having no better way to determine happiness I chose to use the same results, but to analyze them further to try to isolate the variable, or variables that seemed to weigh heaviest in this outcome. I would also like to look at the trend of happiness in countries over the last six years. Using data from multiple years will help to give a better perspective on which aspects affect our happiness the most, as well as allow us to monitor the trends of some of the happiest places on Earth.

Report:

Data

I first had to start with data collection, which I was able to retrieve off of Kaggle. You can find the same data by scrubbing the World Happiness Report website, but the data from Kaggle provided the same information already in a csv format. The original data was created by, John Helliwell, Richard Layard, Jeffrey D. Sachs, and Jan Emmanuel De Neve, Co-Editors; Lara Aknin, Haifang Huang and Shun Wang, Associate Editors; and Sharon Paculor, Production Editor. The data compares the results of a survey put out by the Gallup World Poll and using the Cantril Ladder. The results are averaged per country and then used to create a happiness score that was then used to give Happiness Rankings to 153 different countries.

The data over the years has varied. There have mostly been additions of variables, but in a few rare instances some variables have been randomly omitted. I chose to fine the common thread of variables through the six-year period that I analyzed and construct my data based on those six.

- 1) GDP: This is the GDP per capita of the country. This will give us a financial perspective, which was mentioned as some individuals concept of happiness.
- 2) Health: This is determined by the average life expectancy.
- 3) Support: This is the average person's perception of their own social support. This would include family, friends, and those in your regular surroundings.
- 4) Freedom: This is based on an individual's freedom to make life choices.
- 5) Generosity: This is based on whether or not a person has donated to a charity, either financially or physically (volunteering.)
- 6) Corruption: This variable is based on the persons perception of corruption within the government.

Looking at the first few rows of a sample of this data, it would look like this:

	Country	Regional	Score	GDP	Support	Health	Freedom	Generosity	Corruption	Rank	Year
0	Finland	Western Europe	7.8087	10.639267	0.954330	71.900825	0.949172	-0.059482	0.195445	1	2020
1	Denmark	Western Europe	7.6456	10.774001	0.955991	72.402504	0.951444	0.066202	0.168489	2	2020
2	Switzerland	Western Europe	7.5599	10.979933	0.942847	74.102448	0.921337	0.105911	0.303728	3	2020
3	Iceland	Western Europe	7.5045	10.772559	0.974670	73.000000	0.948892	0.246944	0.711710	4	2020
4	Norway	Western Europe	7.4880	11.087804	0.952487	73.200783	0.955750	0.134533	0.263218	5	2020

We can see the six variables that we took into consideration, as well as a couple other columns that were not mentioned. A brief explanation of the other columns:

- 1) Country: This is self-explanatory, it is the country that is being evaluated.
- 2) Region: This gives us the area of the world the country is located, by continent and at times geographical location within that continent.
- 3) Rank: this is where the country is rated based on happiness score, with 1 being the happiest, and 153 being the least happy.
- 4) Year: Simply the year the data was collected.

- 5) Score: This is the happiness score that is calculated by weighing in a countries average in each of the six variables.

There was data like this for each year from 2015-2020. Using this data I was able to combine them all into one data set, after taking a look at each year individually.

Methods

My first step in my analysis, was to look at the data and make sure that I had clean sets to work with. I had to make sure that the data was adjusted from year to year in order to make sure my variables were labeled in a like manner. Over the years the data addresses the same variables, but the titles will sometimes change based on the year and the individual processing the data. The values still mean the same thing.

After making sure that my data was clean and that the data from year to year would correlate with each other, I chose to then look at the data from each individual year before looking at it as a whole. I created histograms for each year to the variation of the variables from year to year. This also allows me to compare the composite dataset to see if there is a big variation from individual years to the composite. I then joined all of my data sets into one, but appended the column “Year”, which allowed me to keep track of the year the data was obtained.

I created a heat map of both the 2015 World Happiness Report and the dataset formed by the joining of the same report from 2015-2020. This shows the correlation of the various variables with one another. This helps to see which variables seem to be most closely related.

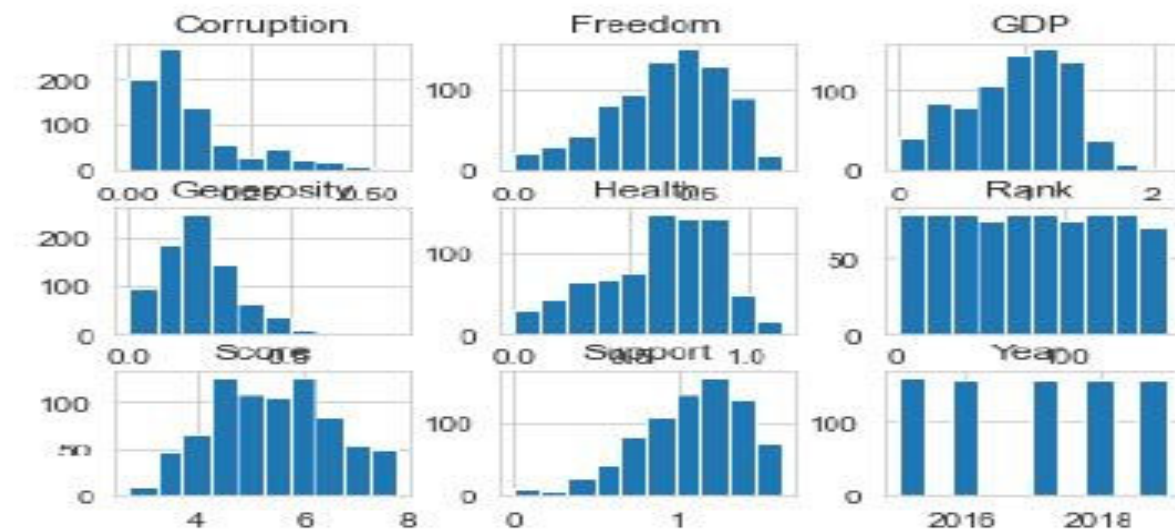
I then created a scatter plot for each of the six variables and their relationship to the Happiness Score. In this I found the linear regression which creates a linear equation that describes the relationship between two variables, and in this scenario, it is between the score and each variable individually.

I then created bar graphs to show the top rated countries for every variable. This was interesting because it gave a visual for the data that we could have seen in the table, but it is easier to see in a graph. This is also interesting because you start seeing where the happiest countries shine and the fact that some of the countries that are not in the top ten still score high in one or two categories. This was one of my least impressive visuals, but surprisingly was one of the most telling visuals for very unusual reasons, that I will get into further when we discuss results.

From there I created a map that is similar in coloring to a heatmap. The darker the country the higher the happiness score. The other fun thing about this is that when you hover over a country the score for that country will show up. It is an interesting visual and allows us to look for regions of happiness as well as check on individual scores.

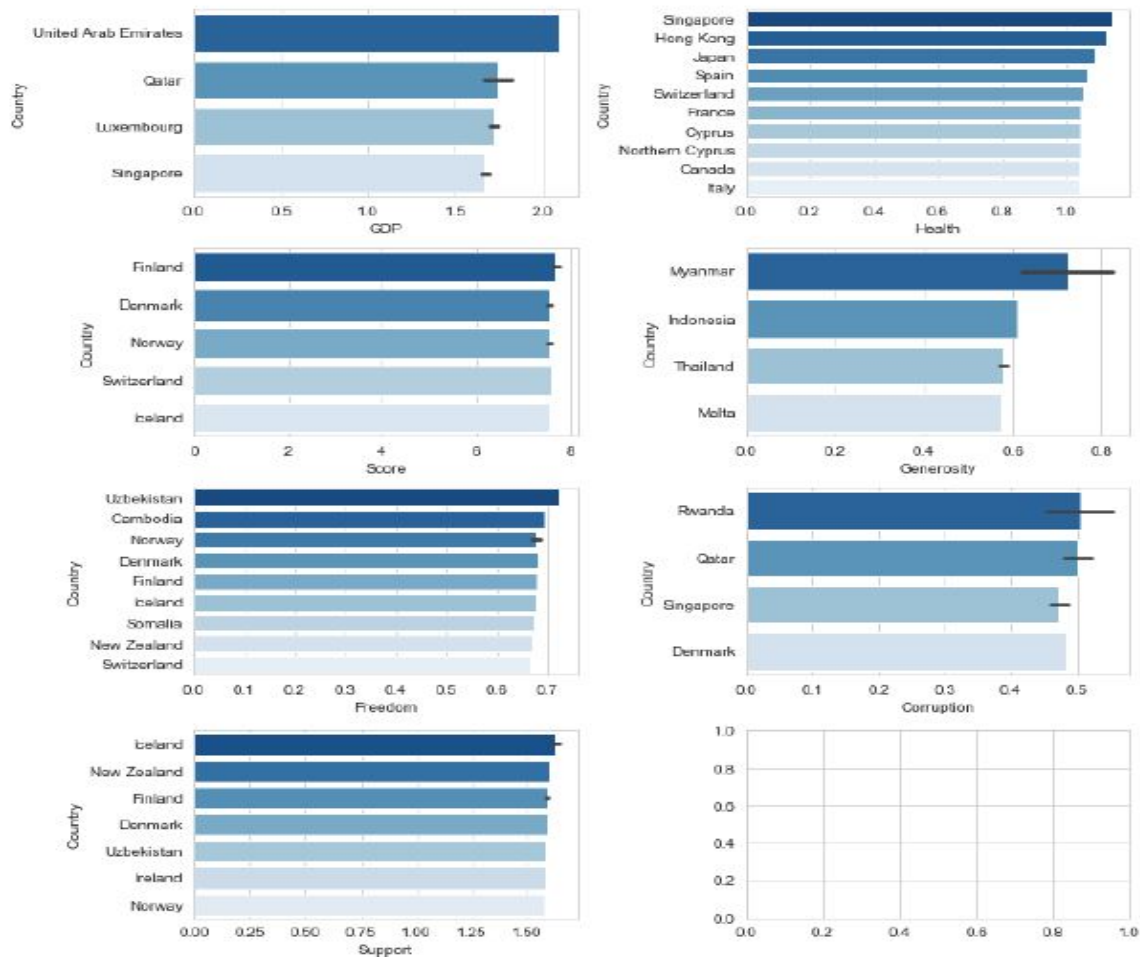
Analysis

As previously stated, we were looking at six variables that reflect an individual's perception of their social environment. We use these to create a "Happiness Score". What I wanted to look at is the weight of each variable on the country's overall placement by happiness score. I first chose to look at the distribution of countries in each category. I created a histogram for every year for 2015-2020. I then compared the individual years to the data frame created by combine the data from each year into one. The combined histogram did not vary greatly from the year to year, so I chose to look at that one as an overview of distribution. The following histogram is from the combined data frame.



What this tells us is where most countries are landing in regard to the score for that particular variable. This can start to give us an idea of what variables really come into play. Couple things I would like to point out are that when it comes to a low score for the corruption variable that is better. The less corruption the better the country is perceived to be. The only other variable that is grotesquely skewed right is generosity. I feel this is starting to shine light on the fact that the variables that weigh heaviest to a person's opinion are the ones that more directly effect that person's life. I feel this becomes more apparent with my next visual.

I also created bar graphs that would show the countries that had the highest scores in each variable, again keeping in mind that a high score in corruption would be a negative.



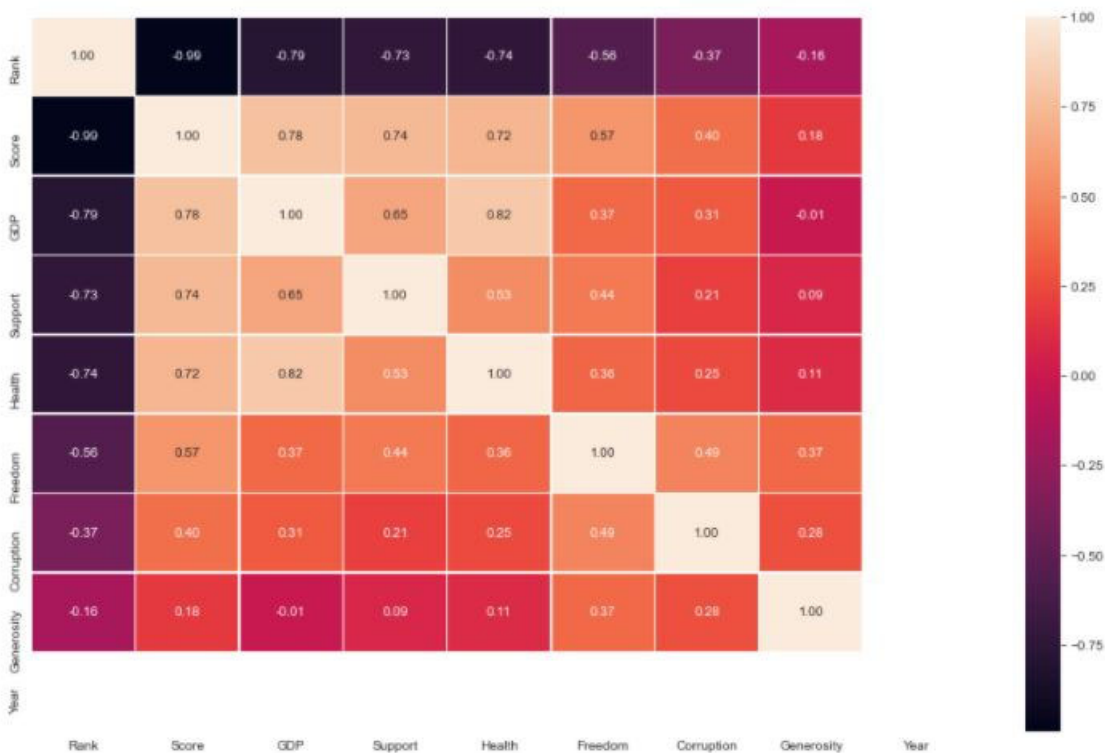
This is where some of our more interesting discoveries in our analysis take place. In five out of six of our variables, a high score would be understood as a positive towards our happiness score. If you do not do the analysis, then you assume the countries with the best happiness scores should be leading in all of the variables except corruption. If we look at our bar graphs, that is not necessarily the case. Our happiest places on Earth show up sporadically and at times in the order we would expect. This shows us that some countries that did not score high will still have strong scores in one or two areas. For instance, United Arab Emirates scores well above everyone else in GDP, but they do not finish in the top five happiest places at the end of the tallying. We see a similar situation when we look at health. In fact if we look at the graph showing the top five countries (happiest countries), which would be the score graph, we notice they do not even show up in the top scores for GDP and health. They do however show up in freedom and support.

I feel our analysis of this visual can lead us to believe three things. One, the two variables that people mention as being keys to happiness, health and wealth, actually do not weigh as heavy on our minds when interpreting our own happiness. Two, the two variables that more directly effect our day to day lives, support and freedom, seem to impact our happiness more. Remember support represents your friends and family that you feel you can turn to when

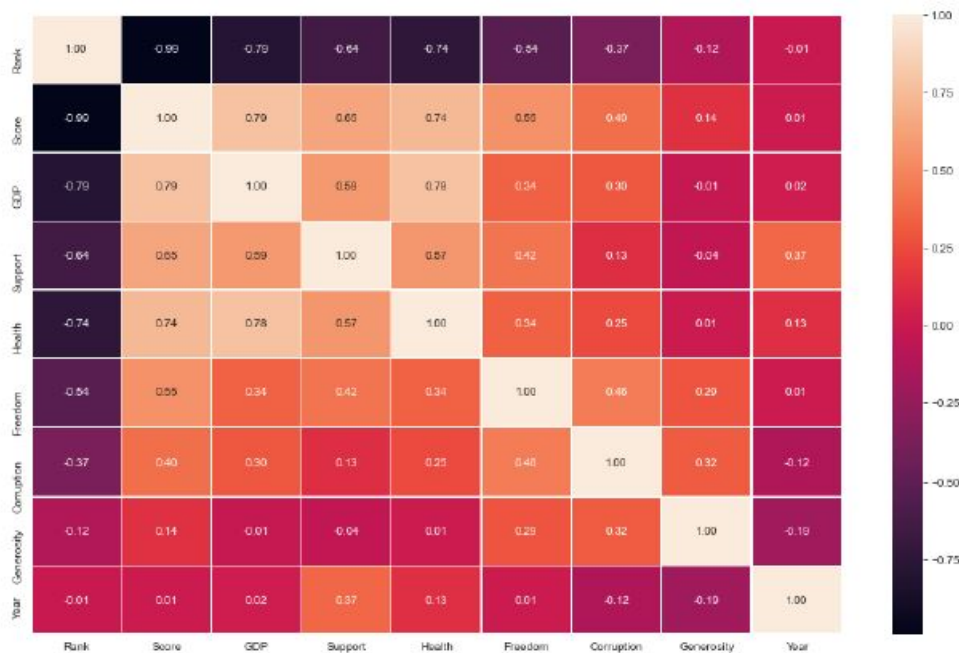
in need, and freedom is your right to choose. Three, the countries that are considered the happiest, that are scoring highest, have more of a balance of all six variables. Again I stated above that the countries that consistently finish in the top five happiest places do not lead any of the categories, and rarely show up with the top scores in any of them.

Our heat maps that were created were done so to show the correlation between the variables. I chose to create two heat maps, again this is to make sure that correlation between variables is not changing drastically from year to year. If this is the case, then the happiness score would be more of a result of what is trending at the time. I chose to do 2015 and our joined data base and make sure they are fairly in line with one another.

2015 Heatmap



2015-2020 Heatmap

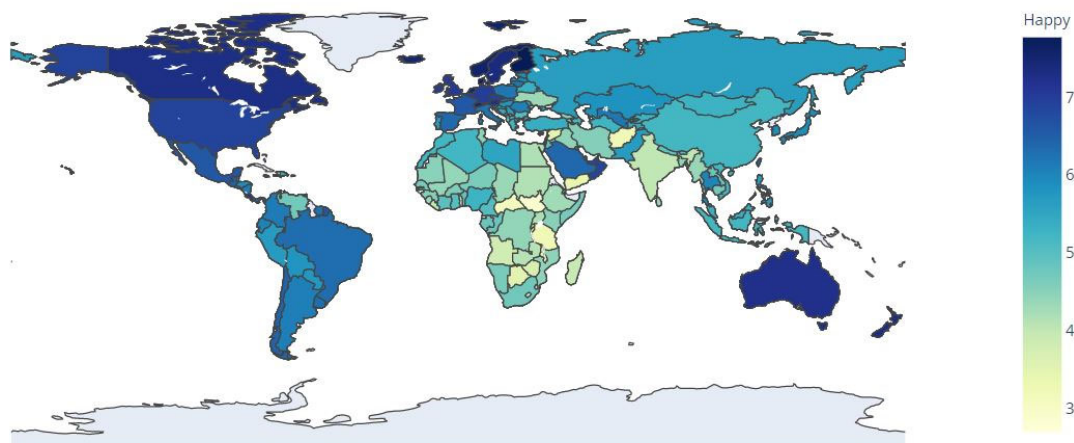


We do see that even though some of the correlations vary slightly from overall to the individual year, they are still close enough that we can analyze the data as a whole and still be accurate with our analysis.

What we notice in the heatmap is that this sheds more light on the information we derived from the bar graphs. According to the heatmap we notice that the variables with the closest correlation with the score are GDP and health. We also notice that GDP and health are also closely correlated with one another, although they seemed to have less and less of a correlation. So how do we explain this discrepancy from our analysis of the bar graphs, after all we stated that GDP and health have less impact, but here it is showing that health and GDP are closer correlated with the score than the other variables. If we go back and look at the histograms as well as the bar graphs, we will notice that in the variables of support and freedom, the scores are not as skewed. The difference in scores are less and there are more scores in the higher range, therefore they will impact the score variation less than the GDP and health which seem to have greater differences from country to country as well as less countries score high in those variables. Based on this second analysis and the addition of the heatmaps, I still feel very confident in stating that the variables impacting an individual's perceived happiness are still closer related to their day to day life, such as support and freedom.

The last visual I will present gives us a look at the world so that we can isolate the regions that tend to have the happiest countries.

Happiest Place on Earth



This map does show which countries are considered the happiest based on their score, but I feel its real benefit is in showing us the regions that have multiple countries scoring high. The Northern European countries tend to score high regularly, as well as the North American countries. We also notice that the countries that we often hear about having civil unrest, or financial and economic short comings tend to have lower happiness scores. We can see regions that seem to have this issue such as the African nations and areas towards the southern part of Asia.

Conclusion:

We analyzed the data provided in the world happiness report, not so much to see which countries were happiest because this was provided, but more so to isolate the variables that we feel might impact an individual's happiness.

We set out to analyze the six variables that reflect how a person feels from a social environment perspective. Looking at the six variables, the two that stood out as being reasons people feel happy or depressed, were GDP and health. A lot of people are under the impression that if they were rich, they would find happiness, or that if you have your health you should be happy.

In our analysis we discovered that on the surface and looking at the correlation between the happiness score and GDP and health, we seemed to be focusing on the right two variables. The GDP and Health had the closest correlation with a country's happiness score. When we analyzed all the data and visuals that we created we found that there were some hidden influences. The correlation between the GDP and Health may have had the biggest impact on the overall score of the country, but that was due to the larger variation from score to score between each country. One argument I think we can make, and this is based on the fact that the countries with the highest happiness scores would show up in these categories, is that Support

and Freedom actual are a better indicator of happiness. The other thing that we notice in the results about the countries that tend to always be in the top five happiest places, is that they are more balanced in all six areas. As we stated before the United Arab Emirates scored well above all others in GDP, but because they were not balanced in all other variables, even scoring in the bottom in a couple, they do not show up in the top five.

There are some strong takeaways from our analysis. We can be encouraged that the information we analyzed gives us hope in trying to deal with the onslaught of depression in the world. We notice that the two variables that seemed to loom large, Support and Freedom, are two variables that we can aid with and try to influence. If we found GDP to be the biggest factor, we would be hard fought to make people richer. We can provide Support. The other thing is to find balance in many aspects of your life to achieve happiness.

I do not believe that this will end depression worldwide, but it gives us some insight into the age-old question of what is the key to happiness. With our best scientific approach we have attempted to tackle that question and shed light on what we find important to enjoying our existence.

References

<https://www.kaggle.com/mathurinache/world-happiness-report>

<https://towardsdatascience.com/happiness-and-life-satisfaction-ecdc7d0ab9a5>

<https://towardsdatascience.com/top-9-data-science-projects-for-a-beginner-in-2020-26eb7d42b116>

<https://www.editage.com/insights/4-step-approach-to-writing-the-introduction-section-of-a-research-paper>

<https://medium.com/@yarnigam/how-to-write-a-data-science-report-181bd49d8f4d>

<https://www.aje.com/en/arc/make-great-first-impression-6-tips-writing-strong-abstract/>

