Happiness: An analyzation

Lucas Bell, Sam Marchand, and Fadi Rached

**Abstract**

We are researching the happiness ranking of countries from The World Happiness Report for 2019. These rankings are based on GDP per capita, social support, healthy life expectancy, freedom to make choices, generosity, and perceptions of corruption. This data helped determine what influences a country's happiness. We assessed data from each country in 2019, determining which factors are the most influential towards happiness. This data provided valuable insight into the major sources of happiness.

**Introduction**

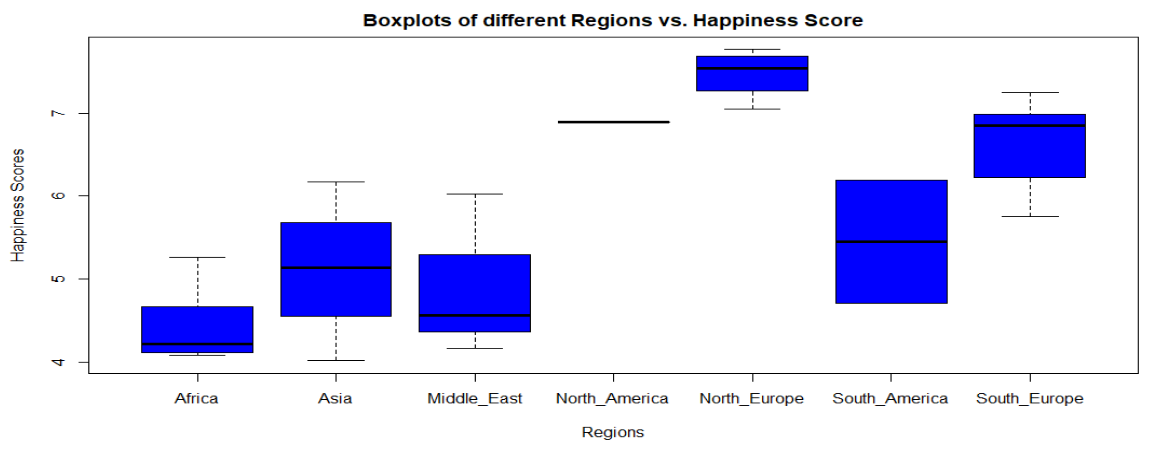
We asked ourselves multiple questions revolving around the World Happiness dataset. The most prominent being which factors are most and least influential in determining a country’s happiness. The dataset consists of 156 countries all across the world. The happiness score for each country is between 1 and 8. Each factor had a score between 0 and approximately 1.8. The factors added together provides the happiness score of the country. We started by conducting exploratory data analysis to get a sense of the most important factors and how they relate to one another.

**Exploratory data analysis**

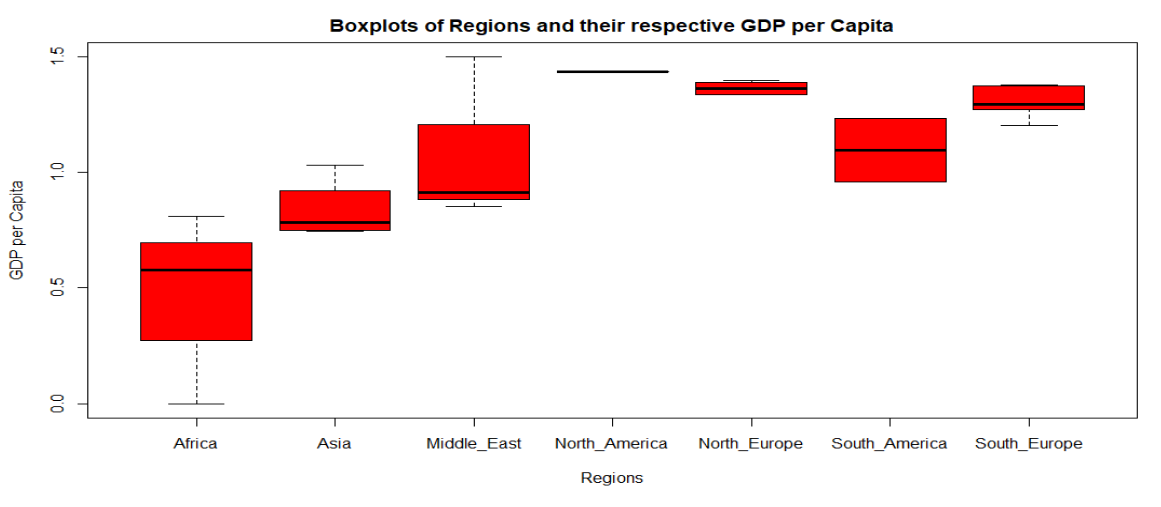
With the exploratory data analysis, we looked to find the answers to many of the questions that we posed. We started with a simple and yet important question, “What variables affect a country's happiness?” This question was answered by calculating the correlation between all variables and a country's happiness score. Interestingly enough, we found that a country's GDP correlated the most with happiness, while generosity correlated the least. Next, we aimed to take a closer look at social support and life expectancy. Specifically, how they relate to happiness. Unsurprisingly, we found that both social support and life expectancy are strongly, positively correlated with happiness with R values of 0.777 and 0.7798 respectively. Next, we sought out to compare GDP and life expectancy. After analyzing, the results were consistent with our predictions, we found a very strong correlation with an R-value of 0.8354.

Next, we took a look at how freedom and corruption are related to happiness, and how they are related to each other. We expected to find a strong negative correlation between corruption and happiness. Surprisingly, we found a 0.3856 correlation between the two. We also expected a strong negative correlation between corruption and freedom. Instead, we found a 0.4388 correlation. Finally, we expected a strong positive correlation between freedom and happiness. This was exhibited in the analysis with a R-value of 0.5667.

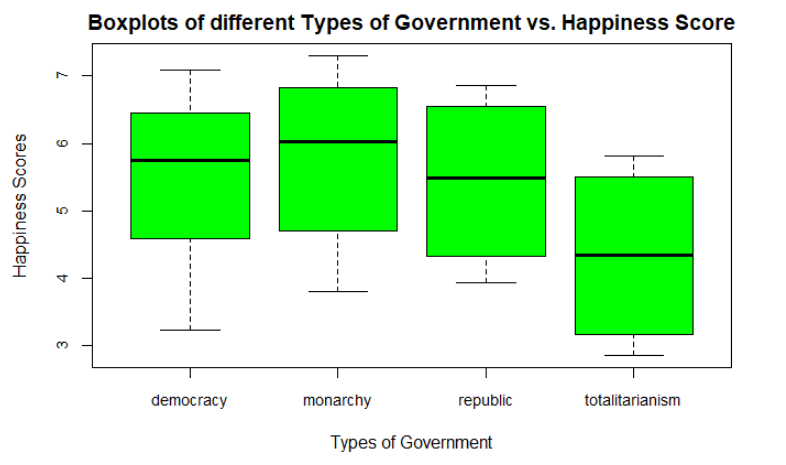
Then, we decided to add to the data frame to answer two intriguing questions. These questions were, “How does region affect a country’s happiness?” and “How does the type of Government affect a country’s happiness?” We will begin by discussing the first question. To solve this question we randomly selected 25 countries and then researched what region they are in. We then bound this data to the dataset. After making a side by side boxplot graph, we found that the happiest region was North Europe while the least happy region was Africa.



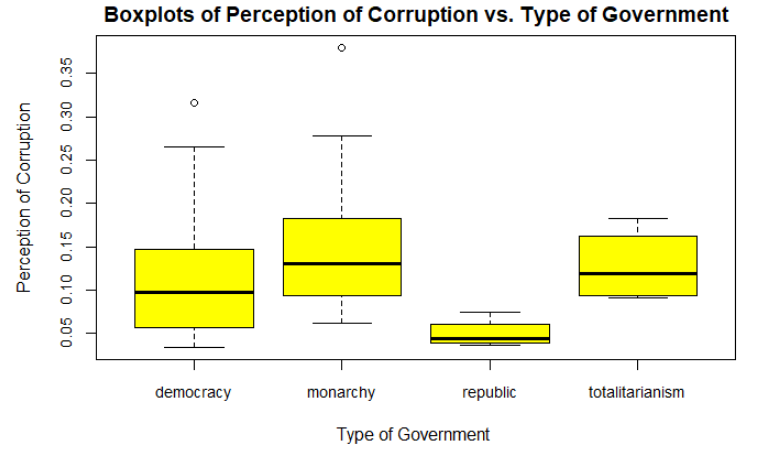
Next, we also looked at how a country's region relates to its GDP per capita. We found that North America had the highest average GDP per capita while Africa had the lowest GDP per capita.



Furthermore, we took another random sample of 30 countries and identified their type of government, and bound that data to our dataset. We found that on average, countries with a monarchy system of government were the happiest while countries with a totalistic system of government were the least happy.

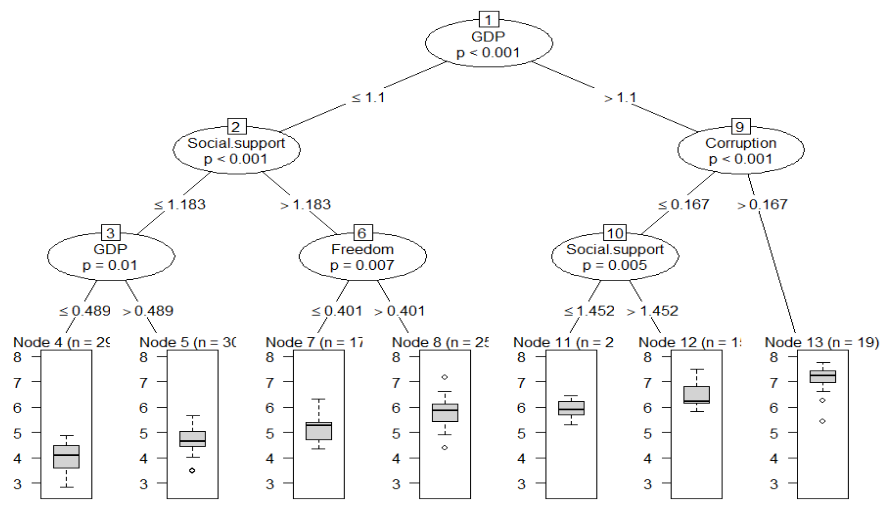


Finally, we compared the type of government and the perception of corruption. Surprisingly enough, monarchy had the highest perception of corruption, and republic had the lowest.

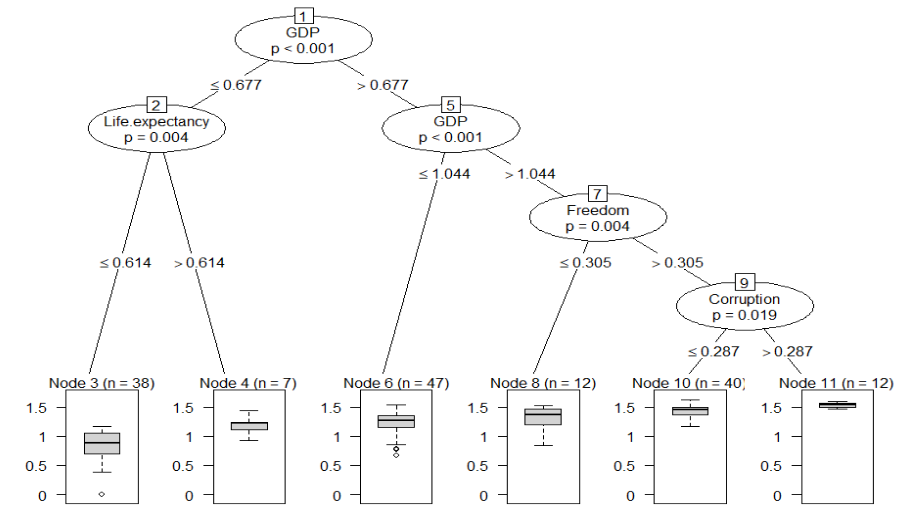
****

**Model and interpretation**

The first model we made was a regression tree with the happiness score as the predictor variable and all the other factors as explanatory variables to determine which factor was the most and least influential. As shown in the tree, GDP per capita proves to be the most important factor contributing to a country’s happiness.

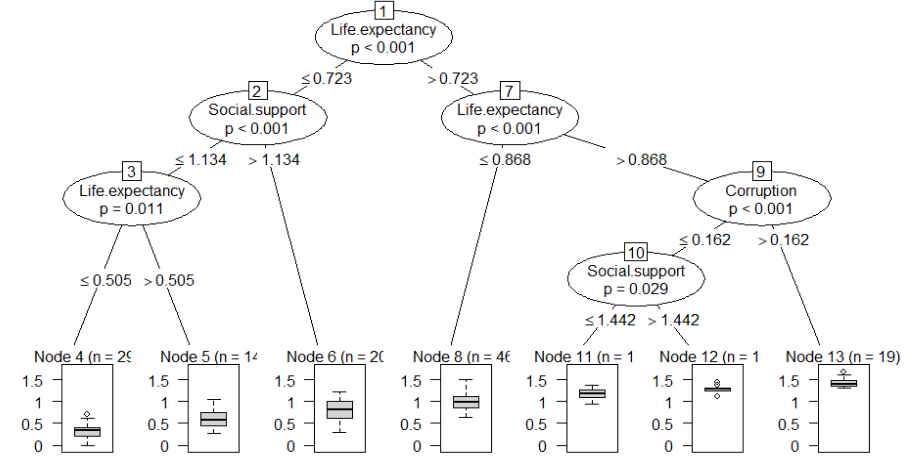


This was expected since money plays an important role in one’s social life. We first calculated the R-values then proved that GDP has a strong effect on the happiness score by creating a regression tree. On the contrary, generosity was not shown on the regression tree proving its insignificance in relation to a country’s overall happiness. The second question regarded the influence of social support and healthy life expectancy. Not only did we find that they are two of the most influential factors relating to happiness, but we also found that they have a strong correlation with one another, as shown in the regression tree of social support below.

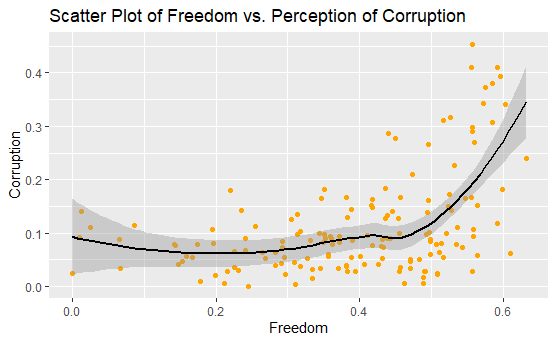


We chose to look at these two factors together because one would expect that a strong sense of social support would be closely related to an expectation of a long and healthy life. This is also why we chose to plot social support vs. life expectancy as well as both factors with happiness score. We realized that social support and life expectancy had similar effects on happiness score.

The next question asked about the correlation between GDP and healthy life expectancy. We found that GDP and healthy life expectancy have the strongest correlation of any factors, as shown in the GDP regression tree below.



One would expect a high positive correlation between GDP and life expectancy because a country with a high GDP is more likely to have more advanced medical technology than a country with a low GDP. People in countries that generally have more money are able to live longer and sustain a healthy lifestyle. These results resonate with the fact that healthcare is still difficult to obtain in poorer countries. On a different note, we found an ironic positive correlation between freedom to make choices and perception of corruption. This was ironic because one would expect freedom and corruption to have a strong negative correlation however, we’ve proved they in fact have a weak positive correlation.



As shown in the plot above, the countries with high scores for corruption unexpectedly also have a high score for freedom to make choices. On the other hand, there are some countries that expectedly have a high score for freedom and a low score for corruption. This creates a swayed correlation.

The next question takes a look at regions and compares them with the happiness score. We found that North Europe has the highest average happiness score while Africa has the lowest. These results were certainly expected because Africa is filled with countries that also have lower scores in factors such as GDP per capita. Whereas regions like North America, North Europe, and South Europe have high average values for GDP and happiness. This can be explained since a country with more money provides more opportunities for its citizens than a country with less.

Our approach to the final question was similar to that of the previous question. It was not surprising to see that Totalitarian governments had the lowest average happiness score whereas Democracy and Monarchy had the highest. While we expected totalitarian governments to have the highest perception of corruption, monarchy ironically had the highest perception of corruption.

**Conclusion**

After our preliminary presentation, we took Dr. Banerjee’s suggestions and decided to put some of them into our code for the final project presentation and report. We first got rid of the zero outliers in our correlation graphs in question #2. Then we decided to take Dr. Banerjee’s suggestion of changing our strip charts with the different types of governments into boxplots in questions five and six. We also took Dr. Bannerjee’s suggestion of putting our models at the end of our presentation instead of at the beginning. We also decided to look at our future goals from the preliminary presentation and make some of those adjustments in our final presentation and report. We started off by converting all of our graphs from the built-in R graphs to the ggplot function. We then removed the NA variables in questions 5 & 6 to make our plots cleaner and easier to read.

After all of these things were cleaned up and fixed, our project was finished. We had done all the necessary coding that we needed to do in order to answer our 6 questions from our project proposal. We learned and observed many things from the Happiness data set that we used for our project. We learned that a happiness score doesn’t just depend on one factor, it depends on a variety of variables. All of these variables are combined together to get the overall happiness score for a country. We broke down certain variables to figure out if some variables were more impactful than others. We even decided to add our own variables into the data set to find out more about a country’s happiness. With our work and research, we found out that there wasn’t a single variable solely responsible for a country’s happiness score. After unveiling the secrets behind happiness, we now feel confident that we understand the complex and often paradoxical nature of it.

**Contributions**

Luke’s role in this project was writing the R code for the group while sharing his screen and getting input from the other groupmates. He also helped solve some problems with questions 5 and 6 involving the box plot and bind function. Sam’s role in the project was co-writing the project proposal with Fadi through a google document. He also provided input on coding for all questions and helped particularly on questions 5 and 6 which posed as the most difficult. Fadi’s role in the project was also helping to co-write the project proposal while sharing his screen with the other groupmates, and also taking some input from them as well. He also led the creation of the regions and government variables to bind to the dataset for both questions 5 and 6. All of the work for the project was done through zoom as a group.