

Yiwen Dai, Joanna Ho, Jasmine Lai, Melody Ma
Stat 20
Hank Ibser
12/07/2018

Final Project - Probable Factors of Happiness

Introduction

Background:

We have chosen the World Happiness data set to analyze for our final project. The data is provided by the World Happiness Report, including data points relating to the happiness score, log GDP per capita, social support, healthy life expectancy at birth, freedom to make life choices, generosity, perceptions of corruption, positive affect, negative affect, confidence in national government, democratic quality, delivery quality, Gini Index, and Gini of household income. These variables can be categorized into three types: wealth and economic health, social and cultural factors, and governmental factors.

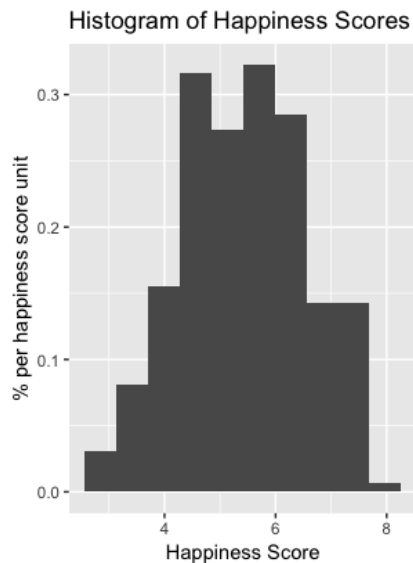
Questions:

- Which factors positively or negatively correlate with happiness? Which factors appear to not have correlation with happiness?
 - Is the phrase “money can’t buy happiness” true or false, based on the data we are given?
 - How do the wealth and economic factors correlate with the level of happiness?
 - Which factor correlates with happiness the most? And which factor correlates with happiness the least? Why?
 - Do all the cultural and social factors carry the same weight in terms of impacting the happiness score of the population? Or are there some that aren’t as impactful?
 - Are correlations by chance? Or are they actually suggestive of something?
- Are these correlations enough to predict a country’s happiness level? Are there any other factors?
- How does this data set inform the government/country of ways to improve the ‘happiness’ of its people?

Exploratory and Focused Data Analysis

We have conducted our data analysis in three main categories: wealth and economic health, social and cultural factors, and governmental factors. We intend to explore the effects of each of the categories mentioned on the happiness score. In particular, we are curious about whether variables included in the dataset correlate with the happiness score and if so, the extent to which they correlate with each other. In order to answer these two questions, we chose to calculate out the correlation coefficients and their regression lines. In addition, we want to know whether some correlations are due to chance, and thus hypothesis testings were utilized as well.

The happiness score is the national average response to the question “please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top. The top of the ladder represents the best possible life for you and the bottom of the ladder represents the worst possible life for you. On which step of the ladder would you say you personally feel you stand at this time?” According to the data we have, the average of happiness scores is 5.44; the standard deviation is 1.13; and the range is from 2.66 to 7.58.



(Histogram of the happiness scores which shows the density scale of each country's happiness scores, numbers provided by the data set)

According to the histogram of happiness scores above, we can tell that the average happiness score is just under 6.

Wealth and Economic Health

Our data for wealth and economic health include Log Gross Domestic Products(GDP) per capita, Gini Index and Gini of household income.

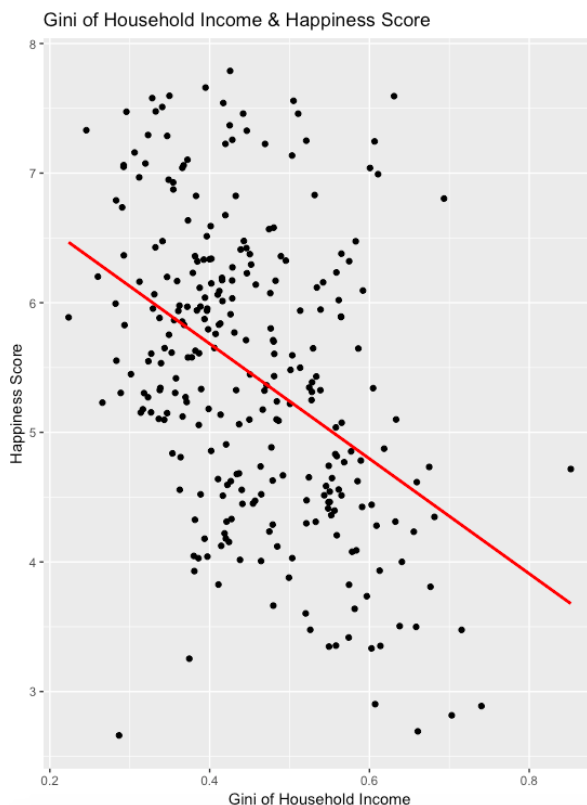
While GDP has been one of the crucial indicators for measuring the level of wealth within the country, Log GDP per capita measures the growth of capita. GDP measures the sum of the domestic consumption, investment, government spending and net export. Because GDP is usually a huge number, using Log GDP per capita would be easier for us to understand and observe the data set. The Log GDP per capita in the happiness data set ranges from 6.47 to 11.47, with an average of 9.3 and the median value of 9.53. The 25% quartile is at 8.46, while the 75% quartile is at 10.28. This indicator has a strong positive linear correlation with the happiness score at $r = 0.78$, proving that the GDP per capital truly influences the extent of happiness of a population.

Another important indicator for measuring the wealth distribution within the country turns out to be the Gini Index. Gini Index measures the extent of unequal income distribution within a country. The Gini Index is between 0 to 1. The score 0 represents perfect equality, while 1 represents perfect inequality. The Gini Index is normally calculated based on the Lorenz curve. The Lorenz curve plots the proportion of total income earned by a bottom X percentage of the total population. The y-axis is the cumulative share of income earned while the x-axis is the cumulative share of people from lowest to highest income. Without calculating the invalid answers, the average Gini Index for the data we have is 0.38. The median for all countries and years we have data on is 0.37, with the 25% quartile at 0.32, and the 75% quartile at 0.43. The data ranges from 0.23 to 0.63. This set of data has a relatively weak negative linear correlation with the happiness score at $r = -0.27$, demonstrating that the Gini Index does not affect happiness that much. Since the correlation is weak, we will do a hypothesis test on the regression line to check whether it can be explained by chance. For our data, we know that the average Gini Index for the data we have = 0.38, and the average happiness score = 5.44. The corresponding SDs are 0.084 and 1.13. From this information, we have that the slope of the regression line = -3.68. The SE corresponding to the slope of the regression line 0.82, where n = number of data points that are not NA = 250. The z-score is then = 4.8 and the p-value is <1%. This means that it is not possible that the correlation is due to chance.

Last but not least, the Gini of household income measures the distribution of household income within a nation. Similar to the Gini Index, the data will be between 0 to 1. While 0 represents the perfect equality for household income distribution, and 1 represents the perfect inequality for household income distribution. The average of Gini for household income from our data set is 0.45, and the median is 0.44. The 25% quartile for the data is 0.37, and the 75%

quartile for the data is 0.53. Overall, the data ranges from 0.22 to 0.85. This data set demonstrates a moderate linear negative linear correlation with the happiness score/life ladder of each country at $r = -0.42$, showing that a higher Gini of household income is correlated to a lower happiness score.

The below diagram represents the regression line between the Gini of household income and the score on the life ladder. As the diagram below indicates, there is a moderate negative correlation between the Gini of the household income and the score on the life ladder. The negative correlation between the two variables demonstrates a general trend that the score on the life ladder will be lower when the Gini of household income gets larger. In other words, people will be less happy when there is larger inequality of the national household income.



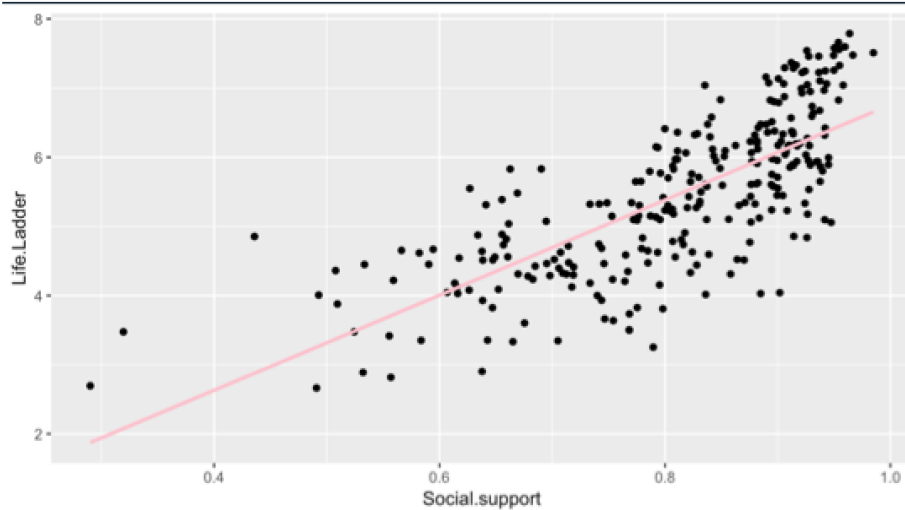
(Scatterplot of Gini of Household Income and Happiness Scoer, and its regression line)

Social and Cultural Factors

Our data for social and cultural factors comprises of social support, healthy life expectancy at birth, freedom to make life choices, generosity, as well as positive and negative affects.

Social support is determined by averaging the binary responses (either 0 or 1) from people who were asked "if [they] were in trouble, did [they] have relatives or friends [they could] count on to help [them] whenever [they] need them, or not?" The data ranges from 0.29 (Central African Republic) to 0.99 (Iceland), which evidently, is wide spread; the first quartile is 0.74 and the third quartile is 0.91. After ridding any non-applicable answers, the average of the data regarding social support is 0.81. Calculating the correlation between this factor and life ladder led to $r=0.75$, which could be concluded as a rather strong positive correlation. Hence, we assume that good social support is an important factor towards people's happiness.

Regression line of social support and happiness score



(Regression line and scatterplot diagram pertaining to social support and happiness scores, as provided by the data set)

As seen from above, there is a positive linear correlation between social support and the happiness score on the life ladder, indicating the higher people rate the support from their peers, family etc., the happier they perceive themselves to be. Hence, we assume that societal factors are of significance when it comes to its population's happiness and wellbeing.

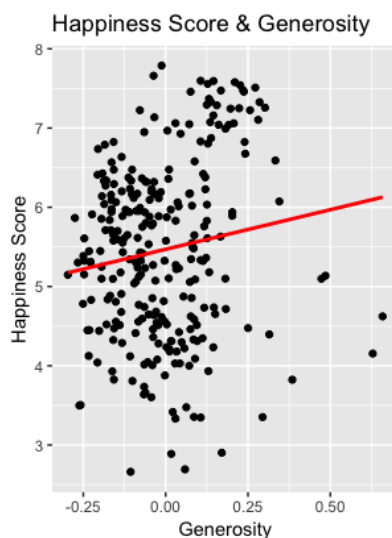
Healthy life expectancy at birth simply refers to the average time span of a human life. In the data, the average is 63.27 years, stretching from 43.98 years (Central African Republic) to 76.54 years (Hong Kong). The correlation for this in regards to the happiness score is $r=0.76$. Once again, there is a strong indication that a longer and healthier life makes people rate their happiness higher. This may be because people value more opportunities in life since they live longer and healthier.

Freedom to make life choices is data also gathered from binary responses (0 or 1), this time based off the question "are you satisfied or dissatisfied with your freedom to choose what you do with your life?" Here, the average is 0.77, the lowest being 0.30 (Haiti) and the highest being 0.99 (Uzbekistan). However, in contrast to the previous two factors, the correlation between the freedom to make life choices and the happiness score turned out weaker, with $r=0.55$. Therefore, one can assume that this does not play big enough of a role for a presumption that an individual rates their happiness according to this factor. Some may interpret this as strange, as one supposes that freedom to do as they please should count towards happiness, but this data refutes this assumption since it has a weaker correlation. Since this is a weaker correlation compared to the previous two factors, we will do a hypothesis test to determine whether chance plays a role. For our data, we know that the average freedom to make life choices score for the data we have = 0.77 and the average happiness score = 5.44. The corresponding SD for freedom to make life choices is 0.13 while the SD for the happiness score is 1.13. From this information, we know that the slope of the regression line = 4.81. The SE corresponding to the slope of the regression line is 0.44, where n = number of data points that are not NA = 280. The z-score is then = 10.83 and the p-value is nearly 0%. This means that it is highly likely not due to chance variation.

Generosity measures the difference between the actual amount of money that each country donated to a charity in the past month and the expected amount that is supposed to be donated, estimated from the regression line which tells the relationship between each country's GDP per capita and the money it donated in the past month. The numbers measured for this variable are all between -1 and 1, with a range from -0.29 to 0.66. The average value is -0.0075; the standard deviation is 0.16. If the number is positive, it means that the corresponding country donates more money than expected; and vice versa. In addition, the greater the magnitude the number is, the larger the difference between the actual value and the

expected value is. To find out if happiness score and generosity are correlated, we use RStudio to calculate out the correlation coefficient between these two variables. After calculations, we come to the conclusion that $r=0.14$, which means that the correlation between these two variables are weak. Since the correlation is weak, we will do a hypothesis test on the regression line to check whether it can be explained by chance. From our data, we know that the SD of happiness score and generosity are 1.13 and 0.13 respectively. With this information, we have that the slope of the regression line = 1.24. The SE corresponding to the slope of the regression line = 0.53, where n = number of data points that are not NA = 271. The z-score is then $(1.24-0)/0.53=2.34$ and the p-value is $<5\%$. This means that it is not possible that the correlation is due to chance.

From the scatterplot below, we can tell that the points are roughly spread in a circle, which suggests that the correlation between these two variables are not strong. In addition, the slope of the regression line is relatively flat, which also indicates the weak correlation between these two variables.



(Scatterplot of generosity index and happiness scores and its regression line, numbers provided by the dataset)

Positive affect reflects people's level of laughter and enjoyment. The numbers measured for this variable are determined by two questions: "did you smile or laugh a lot yesterday?" and "did you experience the feelings of enjoyment a lot yesterday?" The numbers are all between 0 and 1, with bigger number representing greater extent of laugh and enjoyment, and vice versa. The range of these numbers is from 0.42 to 0.80. The average value is 0.71; the standard deviation is 0.10. To find out if the happiness score and positive affect are correlated, we use RStudio to calculate out the correlation coefficient between these two variables. After calculations, we get $r=0.53$. Therefore, we can conclude that the correlation between these two variables is moderate, and they are positively correlated. Generally speaking, the higher the positive affect index is, the greater the happiness score will be.

Negative affect shows people's level of worry, sadness, and anger. The numbers measured for this variable are derived by asking people the following questions: "did you experience the feelings of worry a lot yesterday? how about sadness? how about anger?" The numbers are all between 0 and 1, with bigger number reflecting greater level of worry, sadness, and anger. The range of these numbers is from 0.11 to 0.60. The average value is 0.29; the standard deviation is 0.091. In order to find out if the happiness score and negative affect are correlated, we use RStudio to calculate out the correlation coefficient. After calculations, we get $r=-0.52$. Thereby, it is proper to conclude that the correlation between these two variables is moderate, and they are negatively correlated. Generally speaking, the higher the negative affect index is, the lower the happiness score will be.

Governmental Factors

Our data for governmental factors include perceptions of corruption, confidence in national government, democratic quality, and delivery quality.

Perceptions of corruption is calculated via a survey of two questions: “is corruption widespread throughout the government or not?” and “is corruption widespread within businesses or not?” The individual overall perception is the average of the 0, 1 responses to the above questions, excluding non-responses, and the national overall perception is the average of all the individuals’ overall perceptions for a country. The median for all countries and years we have data on is 0.80, with a 25% quartile at 0.70, a 75% quartile at 0.86, and a range from 0.047 to 0.97. This has a moderate negative linear correlation with the happiness score/ life ladder of each country at $r = -0.45$, suggesting that a greater perception of corruption relates somewhat to a lower happiness score.

Confidence in national government is the average of responses to “do you have confidence in each of the following, or not? How about the national government?” The median for all countries and years we have data on is 0.46, with a 25% quartile at 0.33, a 75% quartile at 0.62, and a range from 0.084 to 0.99. This has a fairly weak negative linear correlation with the happiness score at $r = -0.16$, suggesting that confidence in national government does not affect happiness that much. Since the correlation is weak, we will do a hypothesis test on the regression line to check whether it can be explained by chance. For our data, we know that the average response to confidence in national government = 0.49 and the average happiness score = 5.44. The corresponding SDs are 1.13 and 0.20. From this information, we have the slope of the regression line = -0.028. The SE corresponding to the slope of the regression line = 0.014, where n = number of data points that are not NA = 259. The z-score is then = 2.00 and the p-value is $< 5\%$. This means that even though the correlation is weak, it is most likely not due to chance variation.

Democratic quality averages the voice and accountability of the country along with political stability and absence of violence, based on the Worldwide Governance Indicators that are on a scale roughly with mean 0 and a standard deviation of 1. The median for all countries and years we have data on is -0.17, with a 25% quartile at -0.74, a 75% quartile at 0.66, and a range from -2.22 to 1.47. This has a moderate linear correlation with the happiness score at $r = 0.64$, suggesting that higher democratic quality coincides with a higher happiness score.

Delivery quality averages scores for government effectiveness, regulatory quality, rule of law, and control of corruption, based on the Worldwide Governance Indicators that are on a scale roughly with mean 0 and a standard deviation of 1. The median for all countries and years we have data on is -0.20, with a 25% quartile at -0.72, a 75% quartile at 0.83, and a range from -2.22 to 2.07. The mean differs significantly from the median at 0.005, indicating that there are more countries below the mean than above. This has a strong positive linear correlation with the happiness score at $r = 0.75$, suggesting that better delivery quality is associated with a higher happiness score.

Overall, there appears to be some correlation between a nation’s government and its happiness score, according to the data we were provided. As a specific example, we have provided the scatter plot for “Delivery Quality vs. Happiness Score” for all available data points, along with the corresponding regression line, in red.



(scatterplot of delivery quality vs. happiness scores and its regression line, as from the provided data)

Conclusion

As concluded from the results above, factors that may affect a country's happiness score include the GDP, household income, life expectancy, social support, freedom, positive and negative effects, perception of corruption, democratic quality, and delivery quality. Factors that did not appear to correlate include the Gini index, generosity, and confidence in national government. The factors that had the strongest and weakest correlation with happiness are GDP and generosity, respectively.

[Wealth and Economic Health Effects on the Happiness Score]

As the GDP is highly correlated with the happiness level of a country, according to our data, the strongest correlation between GDP and the happiness level suggests that money could buy happiness. However, the Gini Index suggests a different conclusion. We are especially surprised that the Gini Index did not have a strong correlation with happiness because we generally assume that a higher level of happiness will be achieved if the Gini index is low within a country. However, according to our data, the correlation coefficient is weak between the Gini index and the happiness level of a country. We theorize that the Gini index may have not had as strong an association with happiness because of the different purchasing power in different areas within a nation. For some regions, the sales and income tax may not be as high as other regions, and the price level of products may not be as high as other regions as well. Therefore, even if the income level in those regions is relatively lower than other regions, people in those regions could still purchase sufficient items to satisfy their needs and wants, reaching a satisfactory level of happiness. From these results, the Gini Index and the GDP appears to be contradicting answers to the question "can money buy happiness?".

In addition, the regression line between the Gini of household income and the level of happiness demonstrate the negative correlation between these two variables. In other words, a lower Gini of household income will generally lead to a high level of national happiness.

[Social and Cultural Effects on the Happiness Score]

Overall, the social and cultural aspects of a country seem to support the idea that it does impact the happiness that its population experiences. In both the factors of social support and healthy life expectancy, there are strong positive linear correlations between them and the happiness score, evident from the regression line and scatter plot diagram. In both positive affect and negative affect, there are moderate correlations between them and the happiness score. However, regarding the freedom to make life choices and generosity, it is interesting how these two factors are rather weakly correlated to the happiness score compared to the other two mentioned earlier on as one would suppose it would carry similar weight. However, as the hypothesis test suggests, these two correlations are not due to

chance variation. Therefore, we can conclude that freedom to make life choices does not impact an individual's happiness as much as social support and life expectancy given the data set.

[Government Effects on the Happiness Score]

The government of a country appears to have an effect on the happiness of its citizens. For example, as seen on the graph of the regression line, there is a strong positive linear correlation between delivery quality and happiness score, indicating that there may exist some relation between the two. More specifically, lower perception of corruption within the government, higher democratic quality, and better delivery quality all definitely correlate with a higher happiness score and a happier population. The only exception is confidence in national government; however, the hypothesis test on the regression line shows that even though the regression is weak, it is most likely not due to chance.

References

- As all the happiness scores we have in the dataset are from 2016 and 2017, we check online to see which countries are the happiest according to data from 2018. In 2016, the top ten happiest countries are Finland, Norway, Denmark, Netherlands, Iceland, Switzerland, Sweden, Australia, Canada, and New Zealand. In 2017, the top ten happiest countries are Finland, Denmark, Norway, Iceland, Switzerland, Netherlands, Israel, New Zealand, Austria, and Sweden. In 2018, the top ten happiest countries are Finland, Norway, Denmark, Iceland, Switzerland, Netherlands, Canada, New Zealand, Sweden, and Australia. We find that Finland is the happiest country in 2016, 2017, and 2018. In addition, Finland, Norway, Denmark are always the top three happiest countries in these three years.

Citation: World Happiness Report 2018: See the world's 18 happiest countries, Newsday, 15 Oct. 2018, www.newsday.com/news/world/world-happiness-report-2018-1.17466485. Accessed 6 Dec. 2018.

- Although the dataset suggests that there are lots of factors that can affect a country's happiness level, according to our research, there are some other factors. For example, a country's environmental condition as well as people's life-work balance can also affect a country's happiness level.

Citation: Lobello, Carmel. How do you measure a country's happiness?, 28 May 2013, theweek.com/articles/463919/how-measure-countrys-happiness. Accessed 7 Dec. 2018.