

The Global Nature of Happiness

About Us

Salim Bhana: xSalim1

Christopher Johnson: ScholarChrispy

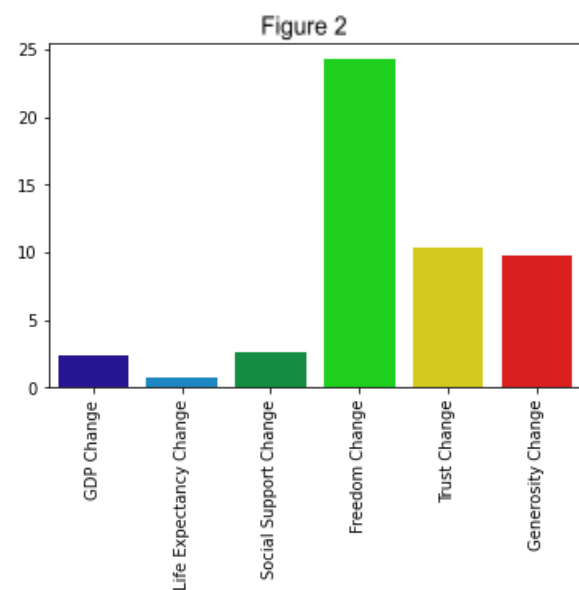
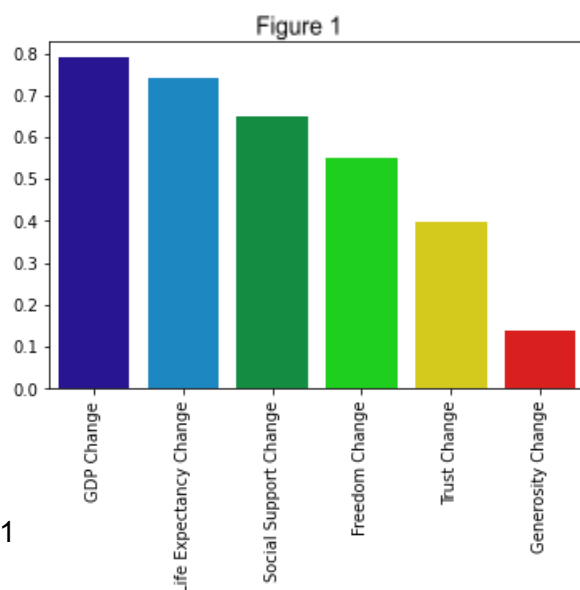
James Currie: JamesCurrie1

Introduction

World Happiness is a very interesting topic because it attempts to take a ton of different factors about a person's or population's general happiness, and boil it down to a simple score. The dataset we have chosen does exactly that. Through various attributes, such as GDP per capita, life expectancy, social support, and more, it tries to quantify the happiness of a specific country or region, and then organize countries based upon that score. This dataset, in particular, spans 4 years (2015 to 2019) This allows for a deeper insight as to how the happiness of every individual country changes over time. The dataset was found on *Kaggle*, and is called the “*World Happiness Report*”. We give full credit to the creators of this dataset for any work to collect and aggregate this data. We simply believe that it is a cool and interesting topic to look into, and to answer some questions about the nature of happiness around the world.

Discussion

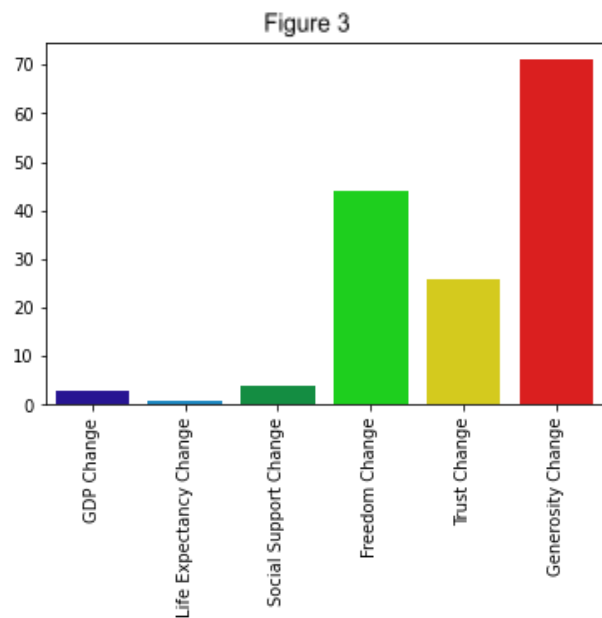
One of our most interesting findings was what factors are the most important to a country's happiness score. This graph we made shows the significance of each of the factors in the dataset. We found these values by measuring the rate of change on a



scatterplot of the factor by the happiness score. The steeper the rate of change slope, the more impactful it is to the happiness score.

This can be seen in *Figure 1* where we found that GDP per capita had the largest influence over a country's happiness score. GDP means 'gross domestic product' and this means the total economic profit, per capita then means 'per person'. GDP having the largest influence on happiness shows that the financial state of a country has the greatest impact on happiness. This is followed by Life Expectancy Change, then Social Support Change, Freedom Change, Trust Change, and finally Generosity Change.

We then took this correlation data, and used it to get a better understanding of the real-world importance of each attribute. We did this by finding how each country had changed over time covered by the dataset. We then isolated the biggest improvement for each country; then counted the amount of other countries which also made this improvement. This can be seen in *Figure 3*, where it appears that an increase in Generosity is the most important factor in a country's happiness. That conclusion would be wrong, since it does not take into account the importance of each attribute. So we must now weigh the data, to make it more accurate.



We did this by using the correlation data to weight the count for each improved attribute, which would give a more accurate representation of the real-world impact of the improvements. This gave us the data represented in *Figure 2*, which suggests that an increase in Freedom is the most likely factor to increase a country's happiness. This is followed by an increase in Trust, then an increase in Generosity.

Funnily enough, an increase in GDP is the second least significant factor, even though it has the highest correlation with a country's happiness score. This is likely due to the abstract nature of a country's GDP, and that it is not noticed on a personal timescale. It is only important on the scale of an entire economy, over a period of months.

Not only that, but the importance of Generosity was cut in half between its initial count, and after weighting. This completely changes the conclusions made based upon this

data, and thus makes weighing the data a very important step in the process of interpreting the data.

This finding was very interesting because it is simply not what you would expect. You would think that an increase in something like Social Support or GDP per capita would play a much bigger role in a country's happiness, but after doing the work, it is clear that is not the case. Not only that, but it is interesting to see the importance of context when working with this data. Not weighing the data would completely skew the results and the findings, thus making our conclusions completely off-base.

Conclusion

In conclusion, the data we analyzed allows us to see how happiness can be influenced by several factors relating to a country. The most important factor is GDP. These findings may be important to countries around the world who are looking to improve the overall happiness of their citizens. Some things we learned as a group while working on this project were how to read datasets and understand what we want to get out of the data. We also learned how to work with multiple .csv files at one time, this is important for working with multiple datasets because it is unlikely that everything we work with in the future will be in only one .csv file. Some things we were able to improve upon were making coherent charts and graphs to represent the data we found.

Acknowledgements

This project was submitted as the final course project for CSCI 2000U "Scientific Data Analysis" during Fall 2021. The authors certify that the work in this repository is original and that all appropriate resources are rightfully cited."

README

How to use our work:

1. Download the main Jupyter notebook from the repository
2. Download the data from the repository, and ensure that it is in a folder called "world_happiness_data", and the folder is in the same directory as the notebook.
 - **Note:** You **MUST** use the data from the repository, as it is a modified version of the data from the original source. Using the data directly from the *Kaggle* site will cause errors.
3. Ensure you are running the code in the notebook with a moderately new version of Python (Python 3 or above was used for our work).