World Happiness Report with visualizations

Sona Khloyan, Rafayel Mnatsakanyan, Pertchuhi Proshyan

2022-12-11

Contents

1 Overview

Happiness has been shown to predict positive outcomes in many different areas of life, including mental well-being, physical health, and overall longevity. Hence, the topic for the Data visualization course project has been analyzing the dependency between the happiness rate and other indicators (i.e. GDP per capita, Corruption rate, Freedom to make choices, etc.) by countries and continents.

This report includes the hypothesis and appropriate visualizations based on the dataset enhanced from the World Happiness report. The main findings and outcomes are visualized in an interactive dashboard (Appendix 1) created by R Shiny. The dashboard can be used to analyse patterns between any other variables included in the dataset.

2 Research methodology

In terms of this research, we have chosen the World Happiness report webpage as a data source (Helliwell, 2022). All the analysis and visualisation processes have been held in R. Initially, we have filtered the data for the time period from 2010 to 2020. After, we did data cleaning using dplyr library from RStudio and replaced all the Null values with the ten-year average of a certain variable for the particular country. For geographical visualisations, we have used a built-in World dataset from the SimpleFeatures R package and merged the datasets. After the data processing and cleaning phase, we created scatterplots, histograms, and boxplots using the ggplot2 R package and plotly package for the heatmap. To make our analysis more user-friendly, with the help of R Shiny we created an interactive application. Table 1 represents a little sample of our filtered dataset. The data contains 12 columns: year, country, happiness score, GDP per capita, social support, Healthy Life Expectancy, freedom to make life choices, generosity, corruption perception, positive affect, negative affect, and confidence in government.

Table 1: World Happiness Report dataset

country	year	happiness	gdp	soc_support	hle	freedom	generosity	corruption	positive	negative	conf_gov
Argentina	2010	6.441067	10.065660	0.9267986	66.30	0.7302582	-0.1275640	0.8546954	0.7654805	0.2109755	0.3518561
Argentina	2011	6.775805	10.112436	0.8890735	66.42	0.8158020	-0.1759191	0.7546457	0.7689425	0.2318552	0.6075377
Argentina	2012	6.468387	10.090750	0.9017764	66.54	0.7474984	-0.1495148	0.8165462	0.7440642	0.2722191	0.4182555
Argentina	2013	6.582260	10.103327	0.9098742	66.66	0.7372504	-0.1320352	0.8229003	0.7661568	0.2542047	0.4337493
Argentina	2014	6.671114	10.066885	0.9178704	66.78	0.7450578	-0.1660366	0.8541916	0.7694680	0.2379129	0.4089614
Argentina	2015	6.697131	10.083051	0.9264923	66.90	0.8812237	-0.1757464	0.8509062	0.7678453	0.3053549	0.3781692
Argentina	2016	6.427221	10.051457	0.8828191	66.95	0.8477022	-0.1936244	0.8509245	0.7317352	0.3116465	0.4195618
Argentina	2017	6.039330	10.068880	0.9066991	67.00	0.8319662	-0.1877939	0.8410525	0.7151824	0.2917173	0.3054303
Argentina	2018	5.792797	10.032199	0.8999116	67.05	0.8458947	-0.2123626	0.8552552	0.7320699	0.3205021	0.2613523
Argentina	2019	6.085561	10.001798	0.8963705	67.10	0.8170526	-0.2126453	0.8304598	0.7349058	0.3190548	0.2738530
Argentina	2020	5.900567	9.887899	0.8971038	67.15	0.8233916	-0.1267429	0.8157805	0.6793168	0.3424969	0.4307581

3 Literature review

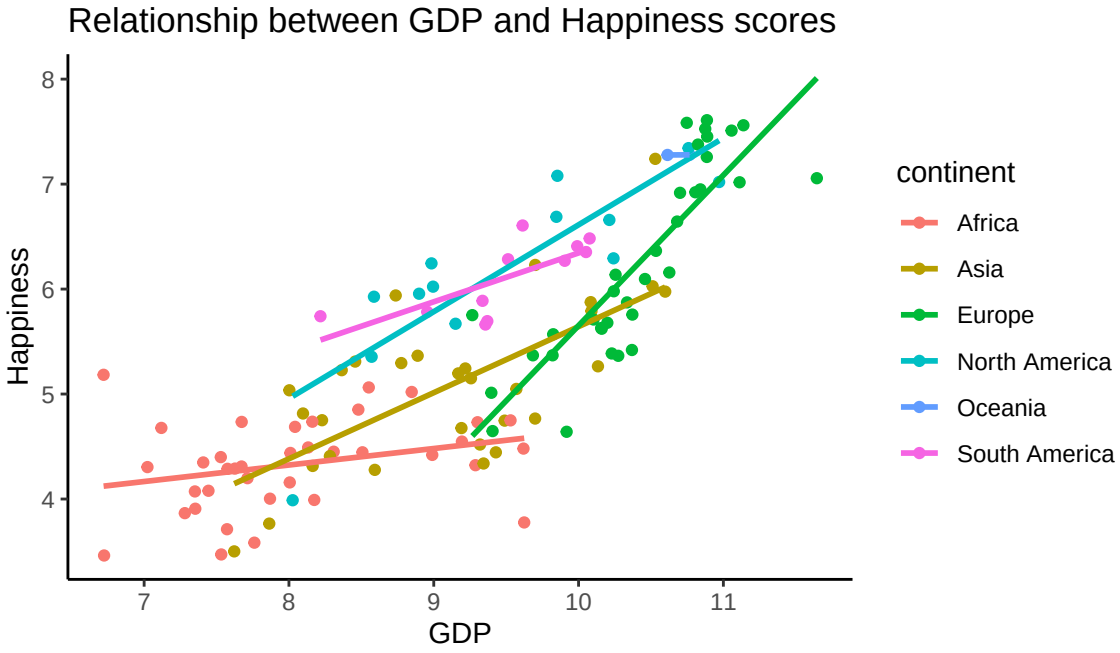
In the "Happiness Score Identification: A Regression Approach" paper, Yichen Ma et al. presented the regression charts of economy, family, health, freedom, perception of corruption, and generosity with happiness from 2015 and 2019. They used the regression approach, which features an interrelation of several independent variables, commonly known as x, and one dependent variable, commonly known as y, and offers an insight into the future trend. The performance of the regression lines of all the factors against the happiness score discovered which factor contributes the most to the happiness score. The strength of correlation (a good, moderate, or bad correlation), the slope, and the Rsquare furthered the understanding of the factors. Both the linear regression and multiple regression with the factors in the Countries of the World and the residuals in the World Happiness Report, bad correlations, and very low R-squares, emerged on regression graphs. Following the findings that a relationship between last year's happiness score and this year's residual exists, the 2020 residual values were predicted. They fitted each year's residual happiness score data into a linear regression model and used that model to predict the 2020 residual value. (Ma et al., 2020)

4 Analysis

After data collection, processing, cleaning and merging, we have formulated a number of hypotheses.

4.1 Hypothesis №1

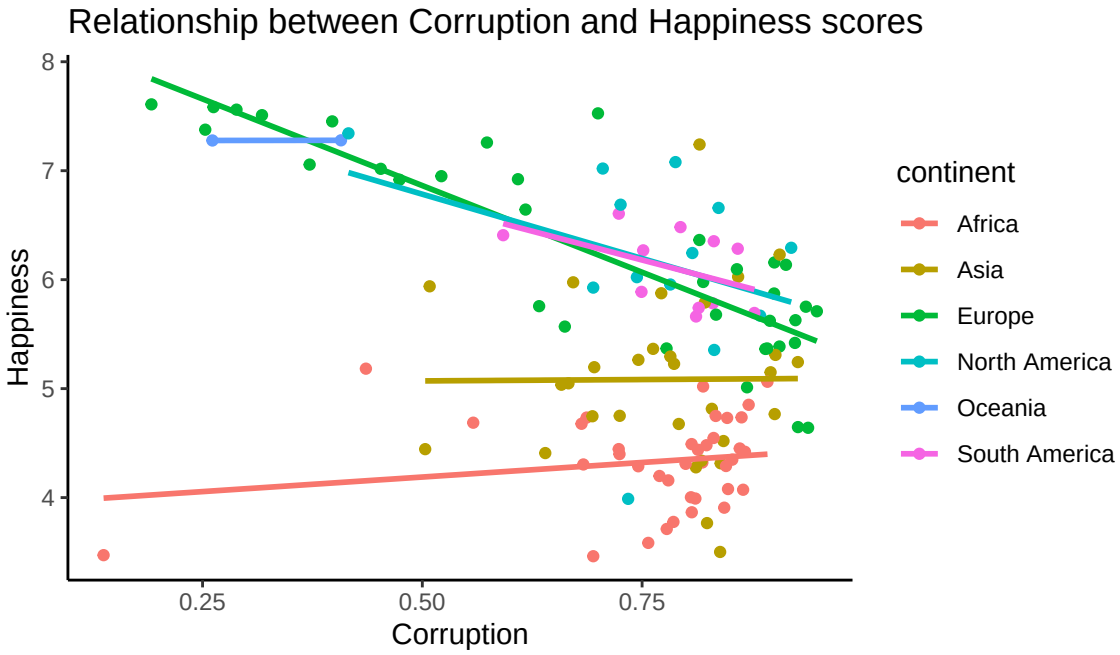
Our first hypothesis is that a higher GDP per capita in a particular country positively affects people's happiness.



There is a slightly increasing pattern in the case of Europe, North America and Oceania. GDP per capita affects positively on happiness rate in Africa and South America as well. So far, we can conclude from the scatterplot above that our hypothesis is accepted.

4.2 Hypothesis №2

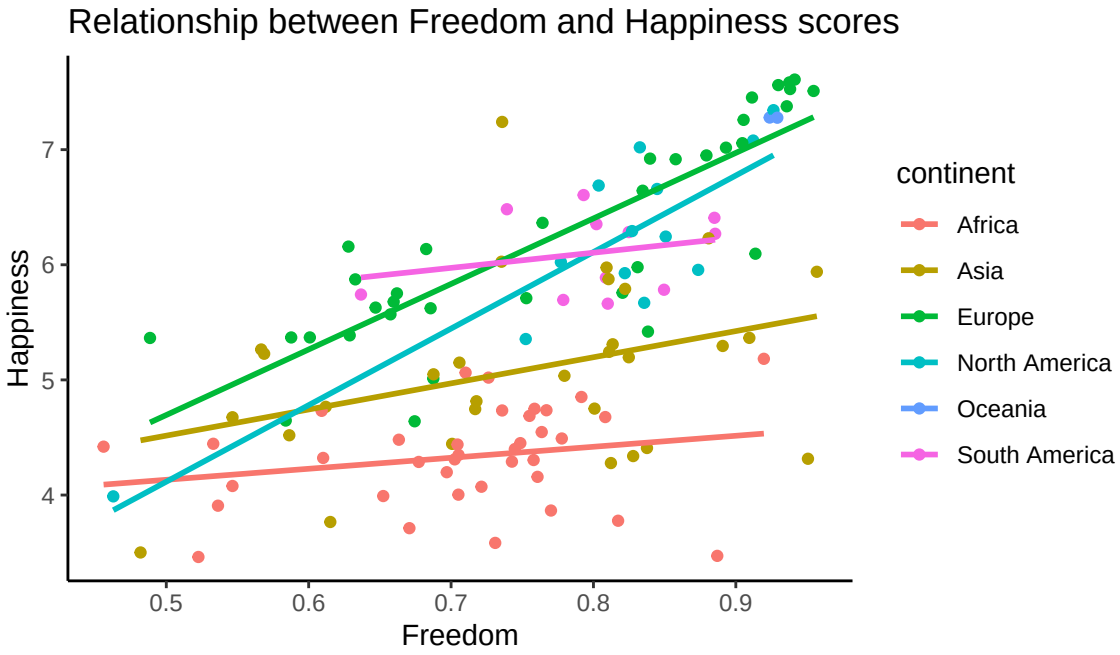
Our second hypothesis is that a higher corruption rate affects people's happiness negatively anywhere.



In Europe, North America and South America it is obvious that the higher corruption rate in the country makes people feel quite unhappy. In the case of Asia, we can conclude that the corruption rate does not affect the happiness rate of people significantly. The surprising facts are that in Oceania and Africa the correlation between the corruption rate and happiness rate is positive: a higher corruption rate makes people feel happier. So by these observations, we reject our hypothesis about the higher corruption rate affecting people's happiness negatively anywhere in the World.

4.3 Hypothesis №3

Our third hypothesis is that freedom to make life choices in a particular country positively affects people's happiness.



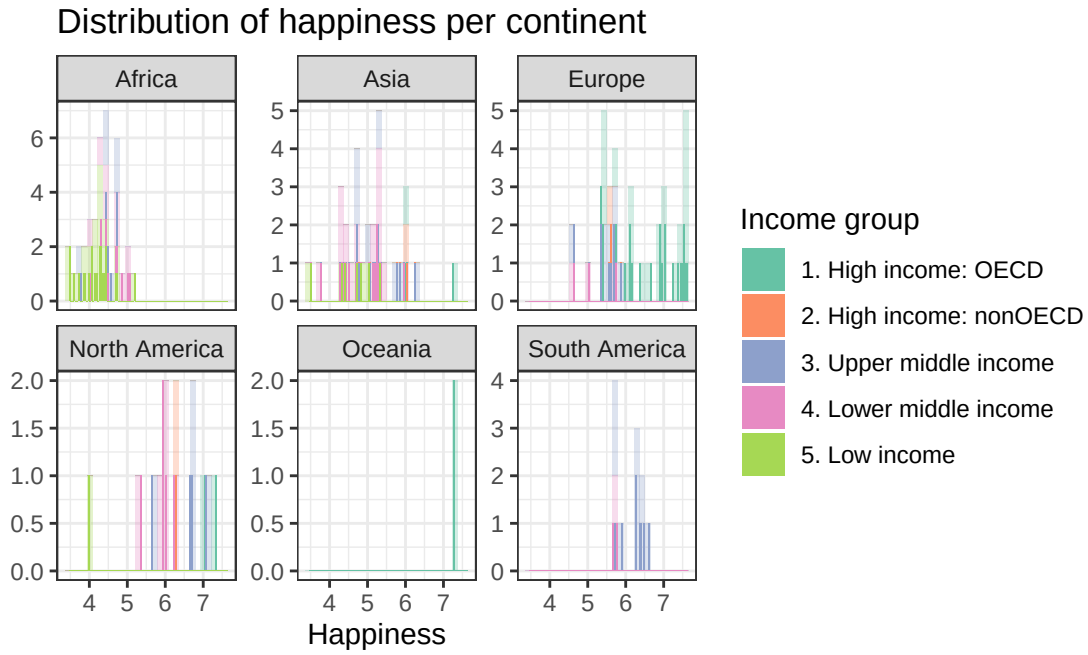
It is obvious that in every continent, freedom to make life choices makes people happier and increases the happiness rate. So, our hypothesis that the freedom to make life choices positively affects people's happiness is accepted.

4.4 Hypothesis №4

Our fourth hypothesis is that the continent's lowest happiness rate belongs to Africa.

4.5 Hypothesis №5

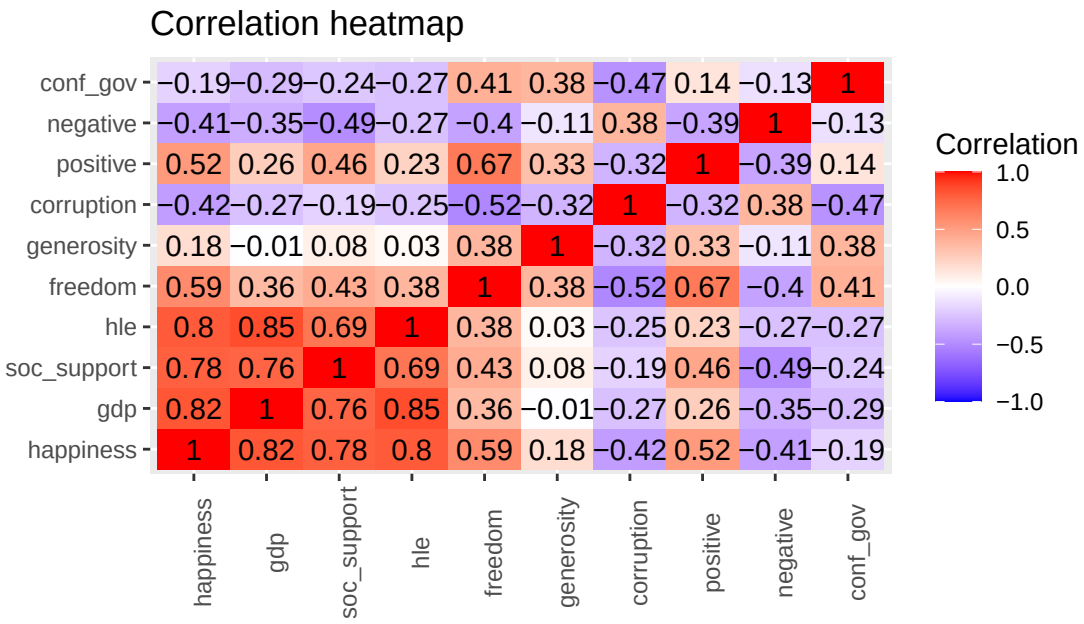
The most differentiated happiness rates among the continents belong to Asia.



Concluding from the histograms plotted above, the highest happiness rate in African countries does not exceed 5.2, while in the rest of the continents, it's relatively higher. Also, happiness rates are mostly differentiated in Asia. In Europe, Oceania, North America and South America, the situation is relatively oriented and constant. Hence, both hypotheses about Africa having the lowest happiness rate and Asia having the most differentiated happiness rates among the continents are accepted.

4.6 Hypothesis №6

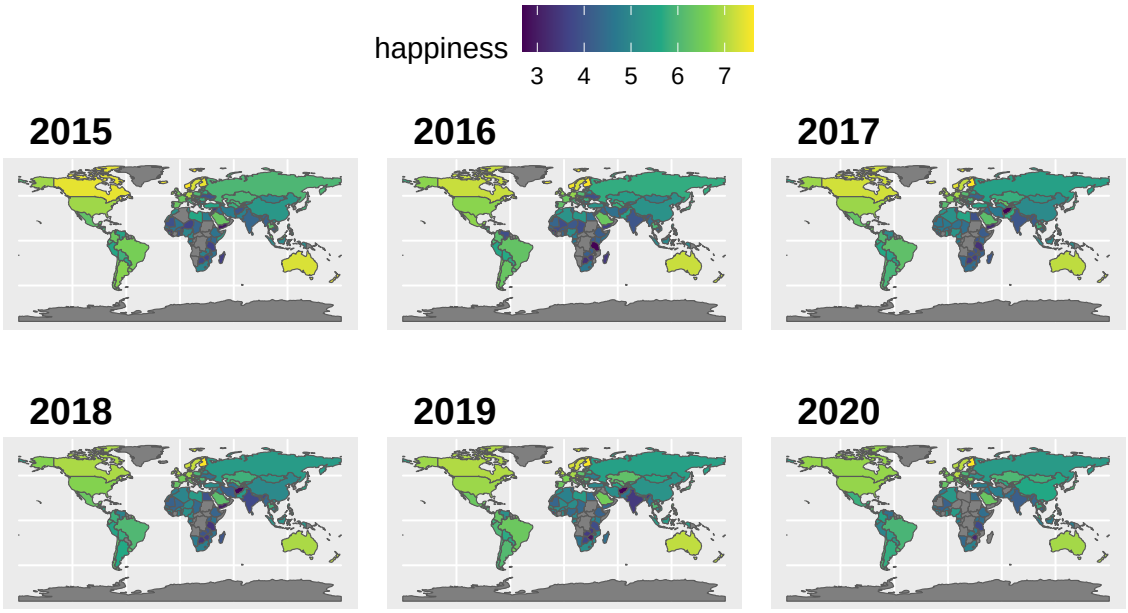
The happiness index is positively correlated with GDP per Capita, Freedom to make life choices, Healthy life expectancy, and Social support.



The correlation scores among happiness rate and GDP per Capita, Freedom to make life choices, Healthy life expectancy, and Social support are respectively 0.82, 0.59, 0.8 and 0.78. As all the correlation values are positive, our hypothesis about the positive correlation between happiness rate and GDP per Capita, Freedom to make life choices, Healthy life expectancy, and Social support is accepted.

4.7 Hypothesis №7

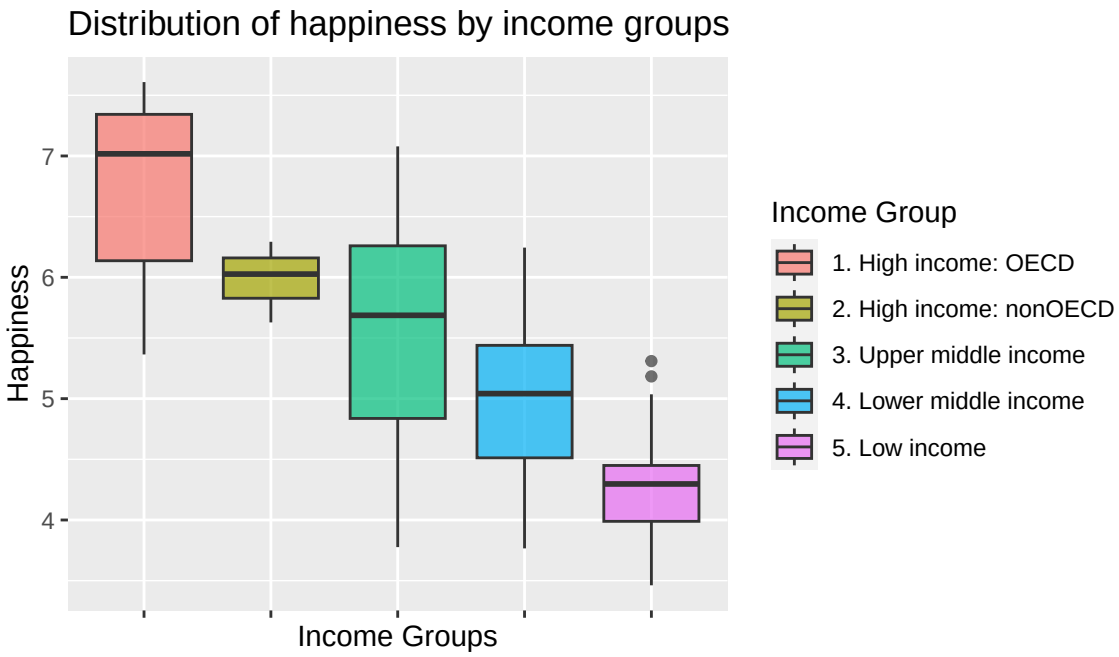
Our hypothesis is that in the period of 2015 to 2020, the happiness index over the World decreased.



The above maps show that the yellow-colored countries have a high happiness rate and the blue-colored countries have a lower happiness rate. So starting from 2015, the amount of yellow-colored countries is becoming green. Hence the happiness rate is decreasing. And visually, each year, the map is getting greener. Hence, the hypothesis that in the period of 2015 to 2020, the happiness index over the World decreased is approved.

4.8 Hypothesis №8

The wealthier the country, the happier people live there.



As we can see from the boxplot above, the higher-income countries have a higher average happiness rate. This outcome was quite expected as we already analysed that the happiness rate is positively correlated with GDP. Hence, it is natural for higher-income countries to have higher happiness rates.

5 Conclusions and recommendations

In this paper, we presented the relationship between the happiness rate and other indicators, such as GDP per Capita, Corruption rate, etc. Our visualisations show that GDP per Capita, Corruption rate, Freedom to make life choices, Trust towards the Government and the rest of the factors somehow affect the happiness rate. Our interesting findings are that in some parts of the World, these factors can perform in different directions. To demonstrate, corruption rate can surprisingly increase the happiness rate in some parts of the World and, more interestingly, contradictorily, the trust rate towards the local Government sometimes makes the happiness level go down. Expectedly, the happiness rate has gradually decreased over the 2015-2020 time period, and the World has gone sadder. We are positive to think that the same analysis to be conducted for the next decade will demonstrate completely reversed results, and we'll see happier results.

6 References

Ma, Y., Liu, A., Hu, X., & Shao, Y. (2020). HAPPINESS SCORE IDENTIFICATION: A REGRESSION APPROACH. E3S Web of Conferences, 218, 01051. Helliwell, J. F. (2022, March 18). World Happiness Report 2022 | The World Happiness Report. https://worldhappiness.report/ed/2022/

7 Appendices

7.1 Appendix 1

Interactive dashboard created with R Shiny.

7.2 Appendix 2

The table below represents the metadata of our dataset:

Variable name	Description
Happiness	The national average response to the question of life evaluations. The English wording of the question is "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 represents your worst possible life.
GDP	The Gross Domestic Product per capita, or GDP per capita, is a measure of a country's economic output that accounts for its number of people
HLE	Healthy life expectancies at birth are based on the data extracted from the World Health Organization's Global Health Observatory data repository.
Soc_support	Binary responses to the GWP question "If you were in trouble, do you have relatives or friends you can count on to help you whenever you need them, or not?"
Freedom	Freedom to make life choices. The national average of responses to the GWP question "Are you satisfied or dissatisfied with your freedom to choose what you do with your life?"
Generosity	Generosity is the residual of regressing the national average of response to the GWP question "Have you donated money to a charity in the past month?" on GDP per capita.
Corruption	The measure is the national average of the survey responses to two questions in the GWP: "Is corruption widespread throughout the government or not" and "Is corruption widespread within businesses or not?"
Positive	The average of three positive affect measures: laughing, enjoyment and doing interesting things. These measures are the responses to the following three questions, respectively: "Did you smile or laugh a lot yesterday?" and "Did you experience the following feelings during A LOT OF THE DAY yesterday? How about Enjoyment?", "Did you learn or do something interesting yesterday?"
Negative	The average of three positive affect measures: worry, sadness and anger. These measures are the responses to the following questions3: "Did you experience the following feelings during A LOT OF THE DAY yesterday? How about Worry?", "Did you experience the following feelings during A LOT OF THE DAY yesterday? How about Sadness?", and "Did you experience the following feelings during A LOT OF THE DAY yesterday? How about Anger?"
Conf_gov	Principal component of the following five measures: confidence in the national government, confidence in the judicial system and courts, confidence in the honesty of elections, confidence in the local police force, and perceived corruption in business.

Figure 1: Metadata