Name: Liang Fang

Emplid: 23134567

Class: MAC 286 – Data Structure

Professor: Dr. Praveenkumar Khethavath

Global Happiness Analysis

What is happiness? Different people may hold different definitions on this word. Simply put, happiness is the state of being happy. According to the article “What is happiness?” by Daniel M Haybron, the author suggests that “Happiness can be divided into three elements: a state of attunement (feeling safe and secure), engagement with a situation (enjoying activities), and an endorsement of life (feeling that life is positively good). Tranquillity could be argued as the cornerstone of happiness. To complicate the picture, it seems that happiness includes not just experienced emotions and moods, but the nonconscious aspects of our emotional condition as well” (Hayborn, 2013). Many people are wondering what are the top 10 happiest countries in the world because we want to deeply understand and explore what happiness is and what are its essential components for happiness. While talking about the world happiness, it is more complicated than individual happiness because the attributes that count towards global happiness score are including economy, family, life expectancy, freedom, trust, generosity, etc. among different nations.

For this project, I am using the “World Happiness Report” open datasets downloaded from Kaggle.com. The datasets include happiness rank and scores by country in 2017 and 2016. According to the data publisher “Sustainable Development Solutions Network,” the World Happiness Report is a landmark survey of the state of global happiness. “The first report was published in 2012, the second in 2013, the third in 2015, and the fourth in the 2016 Update. The World Happiness 2017, which ranks 155 countries by their happiness levels, was released at the United Nations at an event celebrating International Day of Happiness on March 20th” (Sustainable Development Solutions Network, 2017). The latest World Happiness Report of 2018 was released in Rome, March 14, 2018. An overview of the article “Finland Wins Double Gold in World Happiness Report 2018” can be found in the World Happiness Report website: http://worldhappiness.report/news/2018/03/14/finland-wins-double-gold-in-2018-world-happiness-report/.

From the Udemy lecture “Python for data science and machine learning boot camp” instructed by Jose Portilla, I learned some basics of how to use Python programming language to read CSV file, analyze, and visualize the datasets. The numpy, pandas, matplotlib and seaborn are necessary libraries to use because numpy offers fast and efficient processing numerical data structures; while pandas offers the flexibility to read and write different data format, such as CSV, text files, it also has integrated plotting functionalities; matplotlib is a library that enables us to plot the graph and visualize the data and seaborn is also a Python data visualization library based on top of matplotlib for making statistical graphics. It is crucial to present the data visualized so that it makes us more accessible to the results.

The purpose of doing data analysis is to discover which factors that are more important to live a happier life so that people can focus more on those factors to improve their lives. The dataset contains 155 countries. This ranking could be considered as an indication of the country's policy-making decision skills.

Below are the variables presented in the data:

Happiness Rank: Determined by how high their happiness rank is. Range from 0-154.

Happiness Score: A score that measures the happiness of a country based on adding up the score of other factors.

Country: The total 155 countries were included in the report.

Economy: GDP per capita: GDP is a quarterly measurement that tells us how well we are and where we are regarding the national economy. With a higher GDP meaning a higher standard of living.

Family: Quality of family life and how important a family to a person is.

Health/Life expectancy: It ranks healthcare availability and average life expectancy in a country.

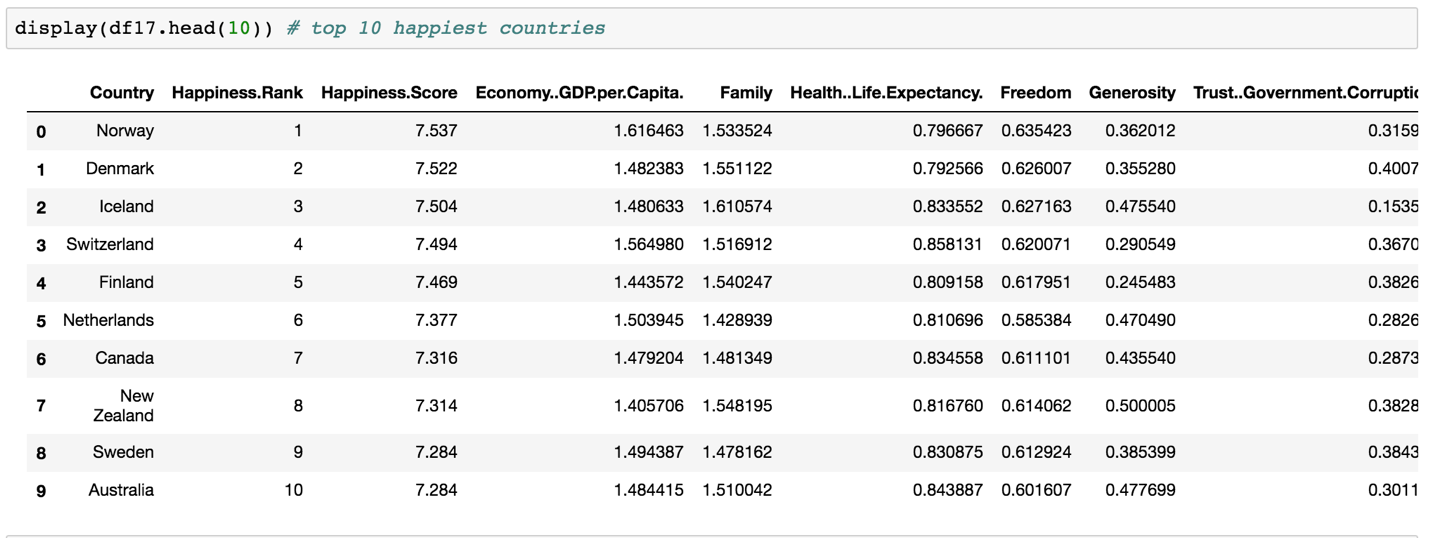
Freedom: It is a yearly survey that measures the degree of civil liberties and political rights in every nation.

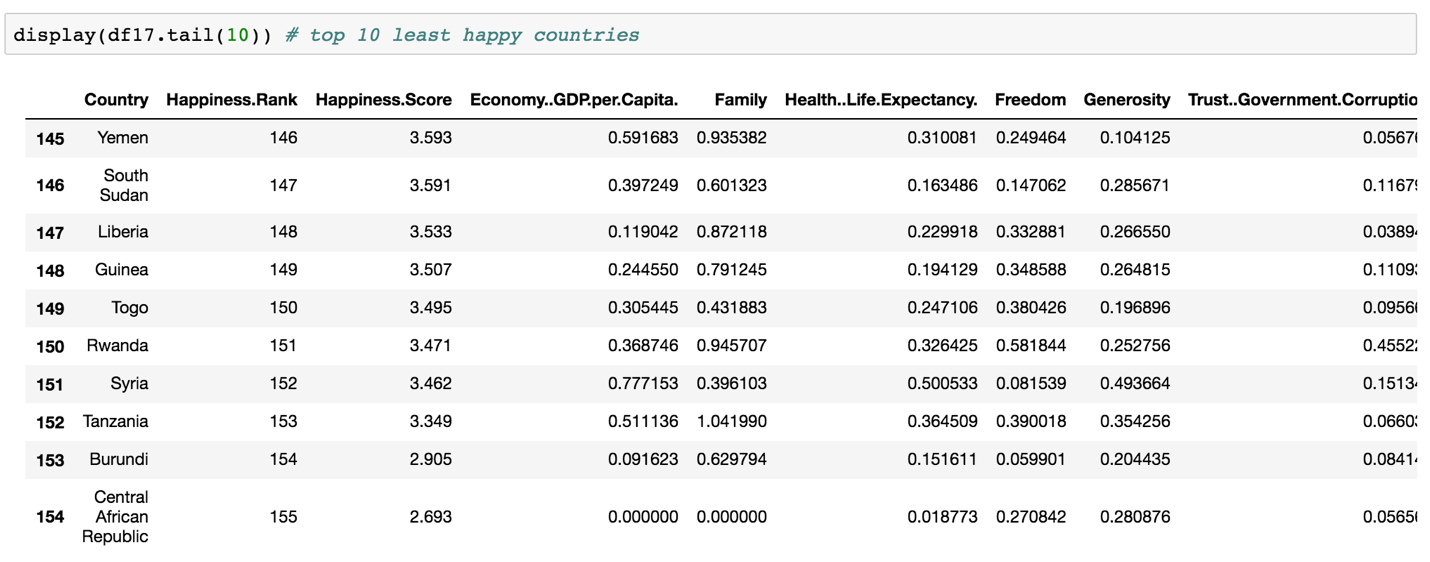
Trust: A measurement of citizens trusts in the government to not be corrupted.

Generosity: A measurement of how charitable countries are.

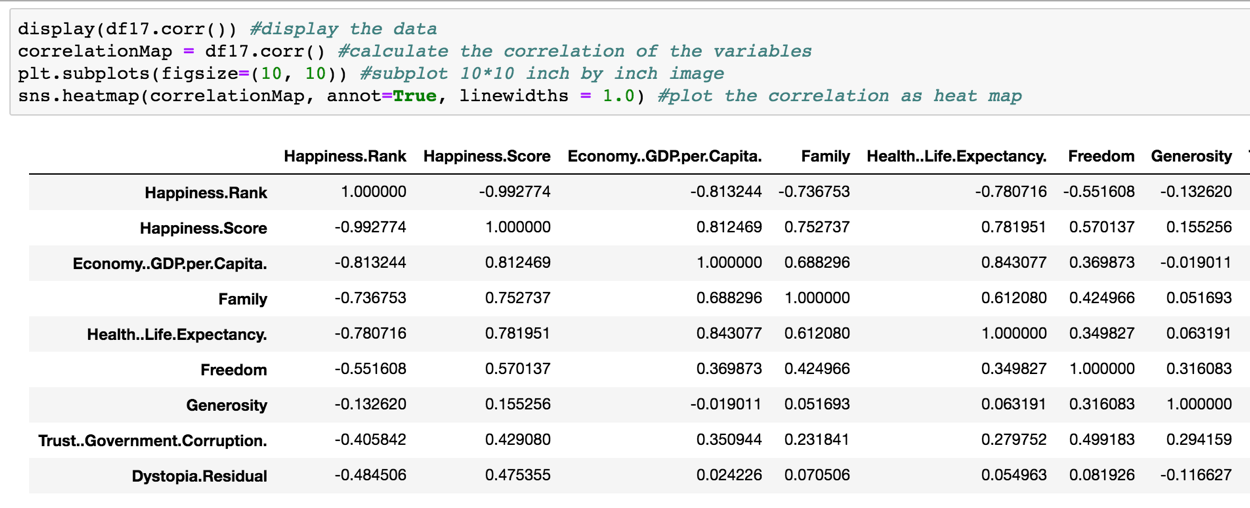
Dystopia Residual: Dystopia is an imaginary country that has the world’s least-happy people. In other words, we can say how bad that country is.

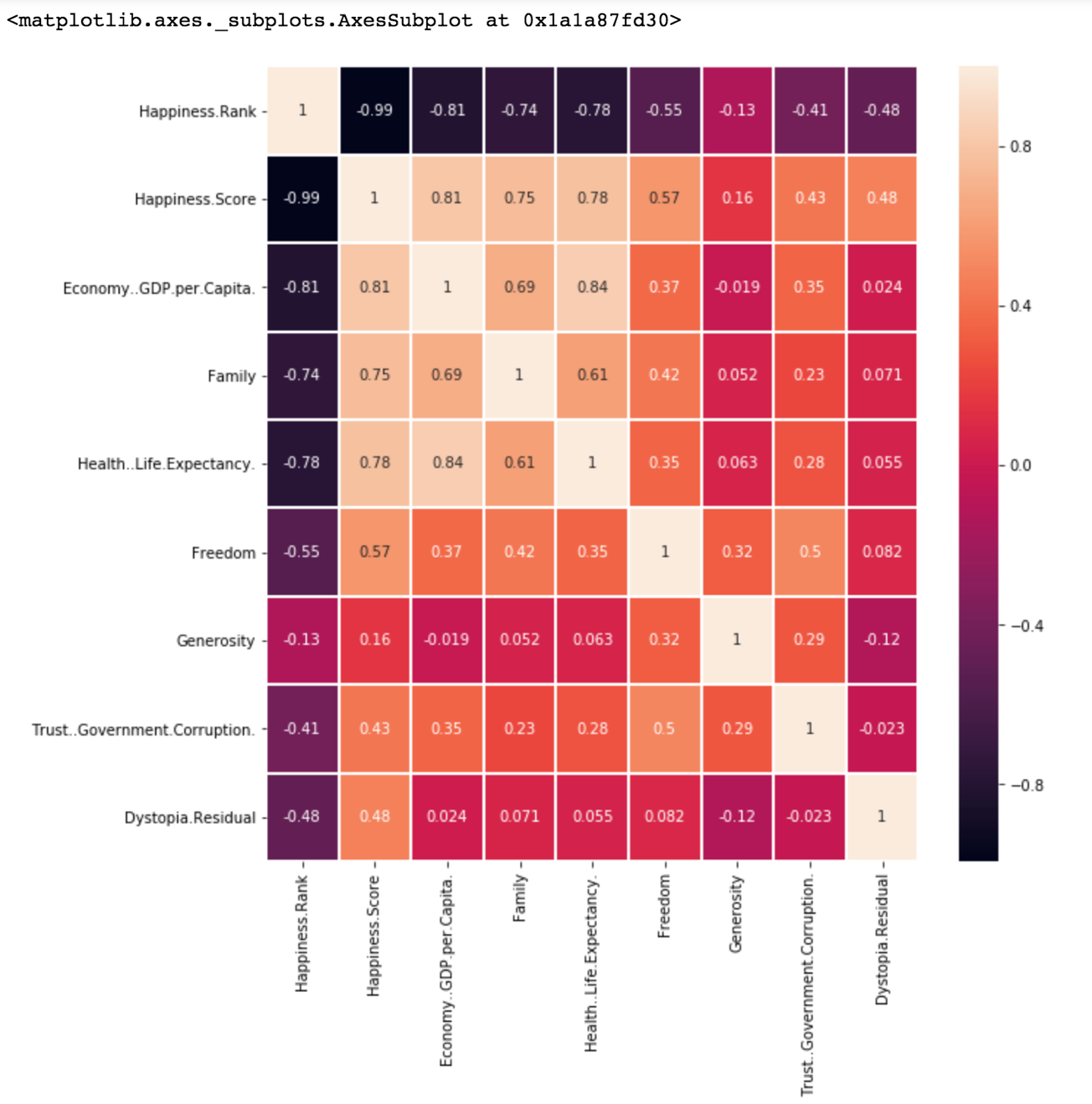
1, By displaying head top 10 countries and tail top 10 countries, we can easily see that where are the best and worst countries to live. Norway ranked first, and the Central African Republic ranked last.



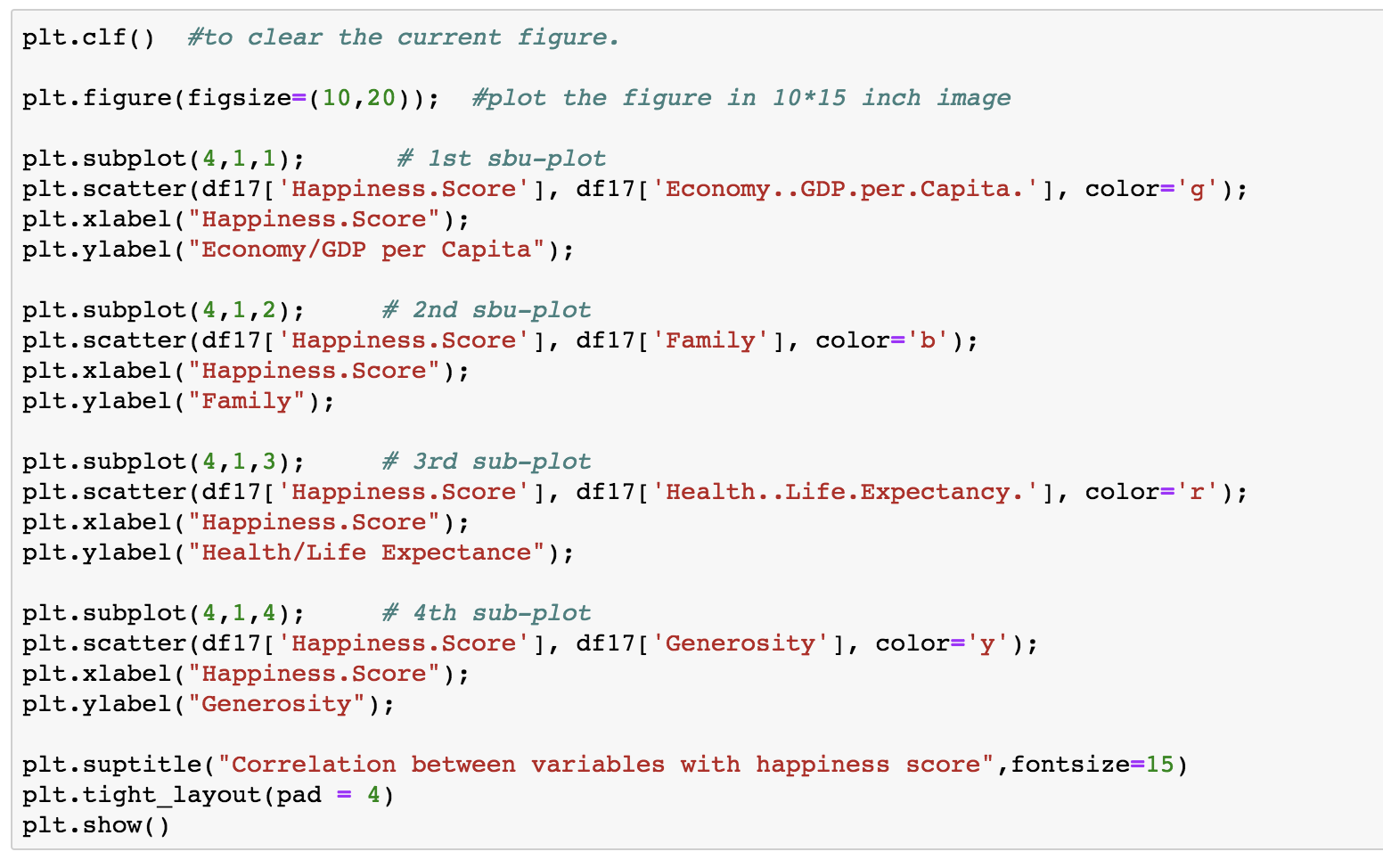


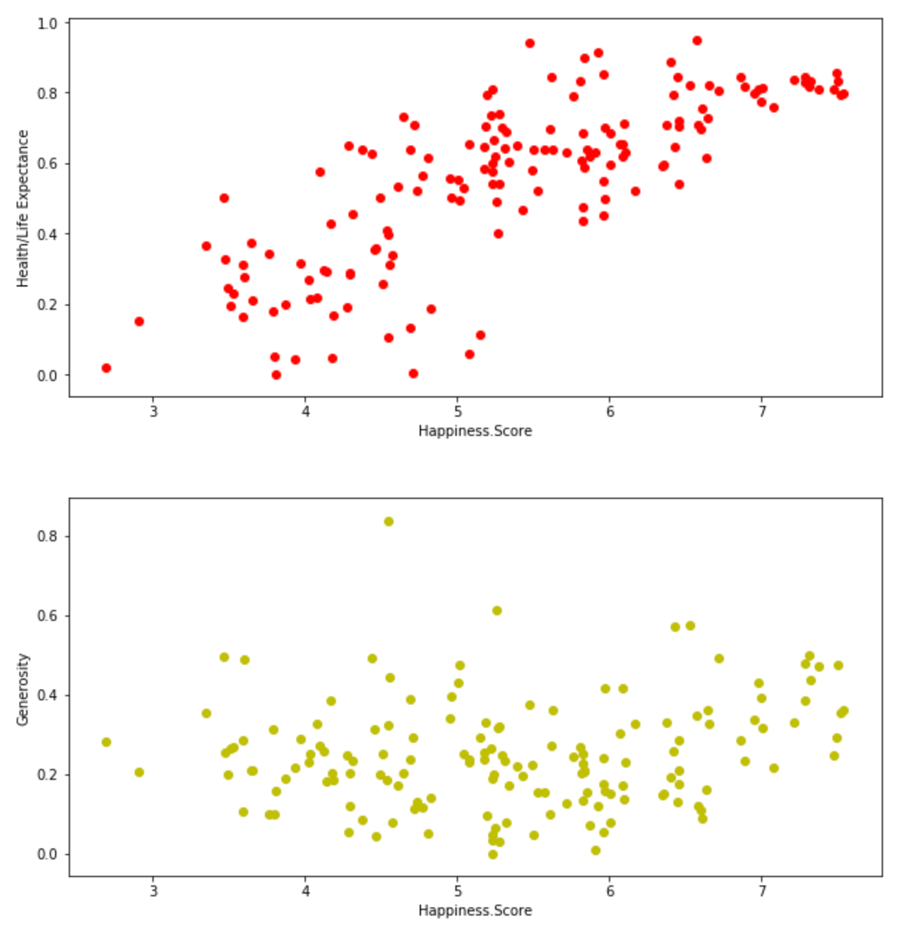
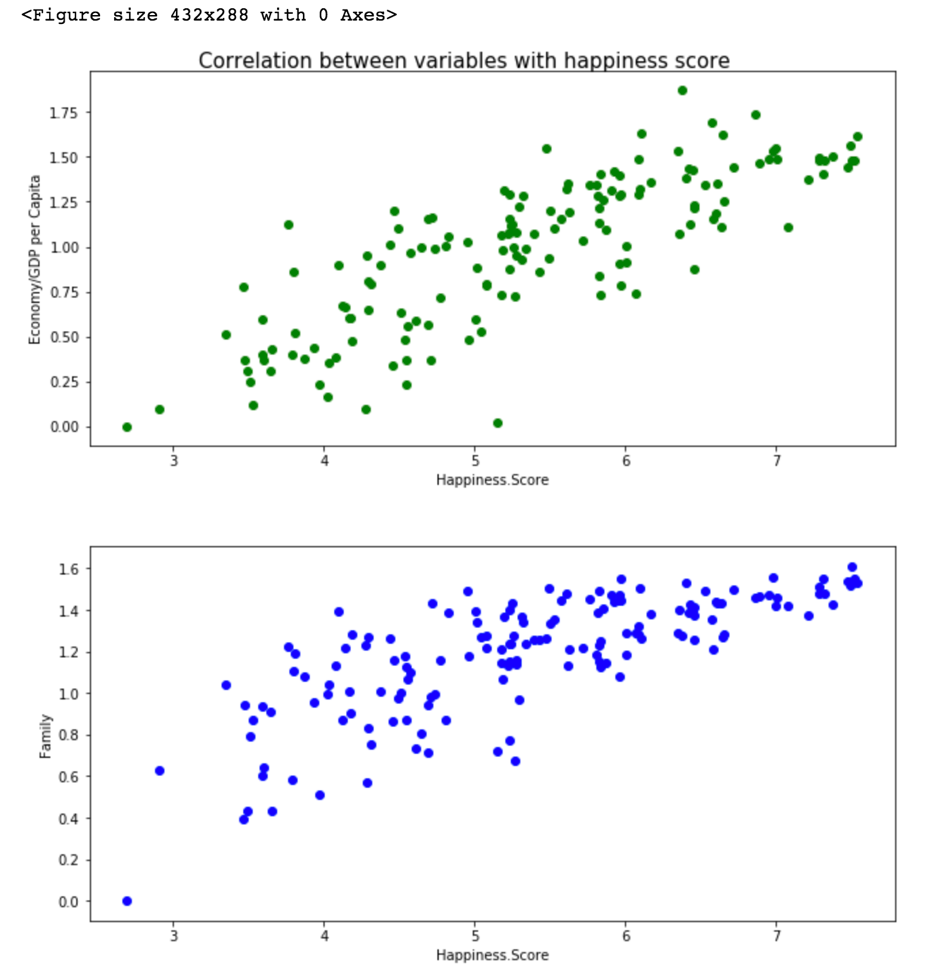
2, Correlation between the variables and happiness. To understand the correlation between the given data, we need to compute the numerical variables in our dataset and plot the table as a heat map. By obtaining the heat map of the correlation among the variables, the color square represents the high or low of correlation variables. The lighter the shade means, the higher the correlation. We can observe that the happiness score is highly correlated with the economy (GDP per Capita), family and life expectancy; while generosity has the lowest impact on the happiness score.



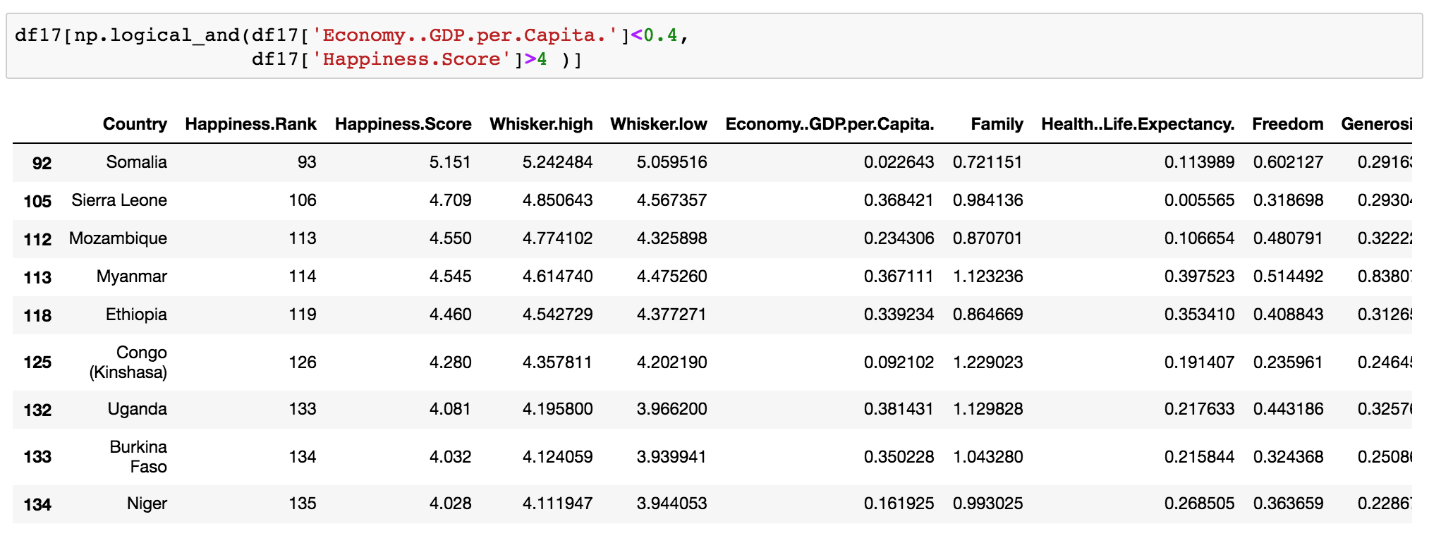


3, We can also observe that in the “Correlation between variables with happiness score” figure, the relationship between “Economy(GDP per Capita) with Happiness score”, “Family with Happiness score”, and “Health/Life Expectancy with Happiness score” are presenting positive slops, which mean those variables do have a big impact on happiness. In contrast, “Generosity with Happiness score” presents a linerly slopped scattered plot, which means it does not have too much impact on happiness.

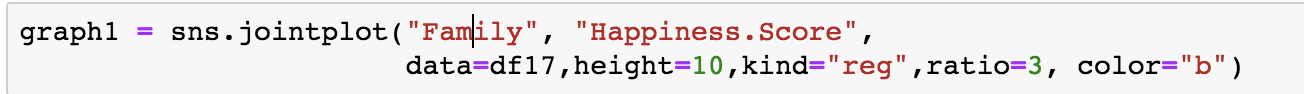


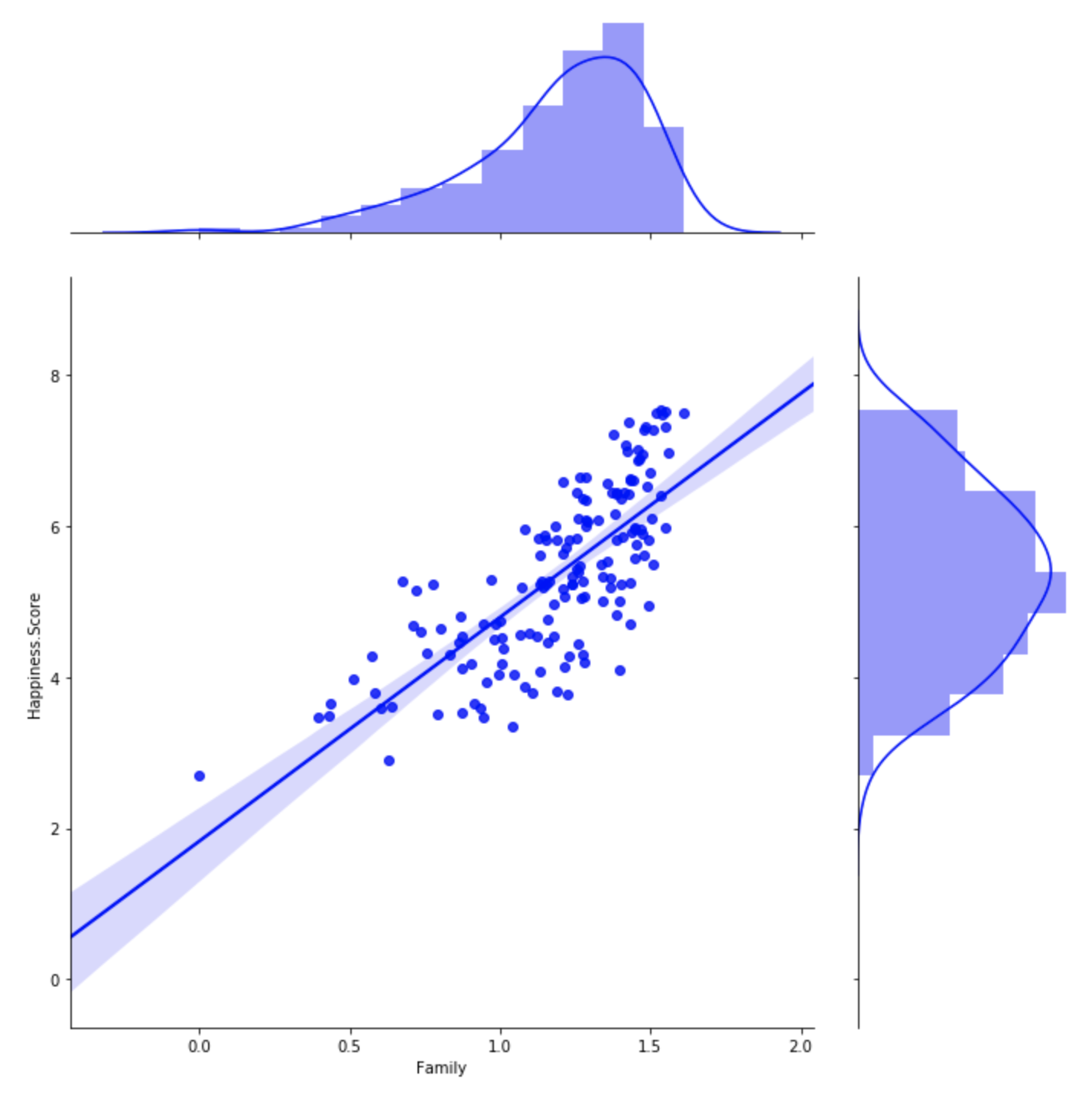


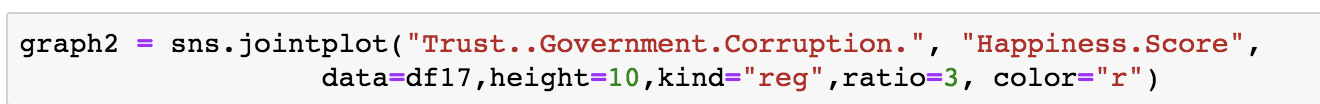
4, By using numpy logical operation, we can filter the “Economy/GDP per Capita” < 0.4 and “Happiness score” > 4.0, the result shows that only 9 among 155 countries that with the low economic performance but still have relatively high happiness score. In other words, there are 9 countries which have higher happiness score value than 4.0 although they have GDP value lower than 0.4. So we can conclude that GDP is a strong indicator of happiness.

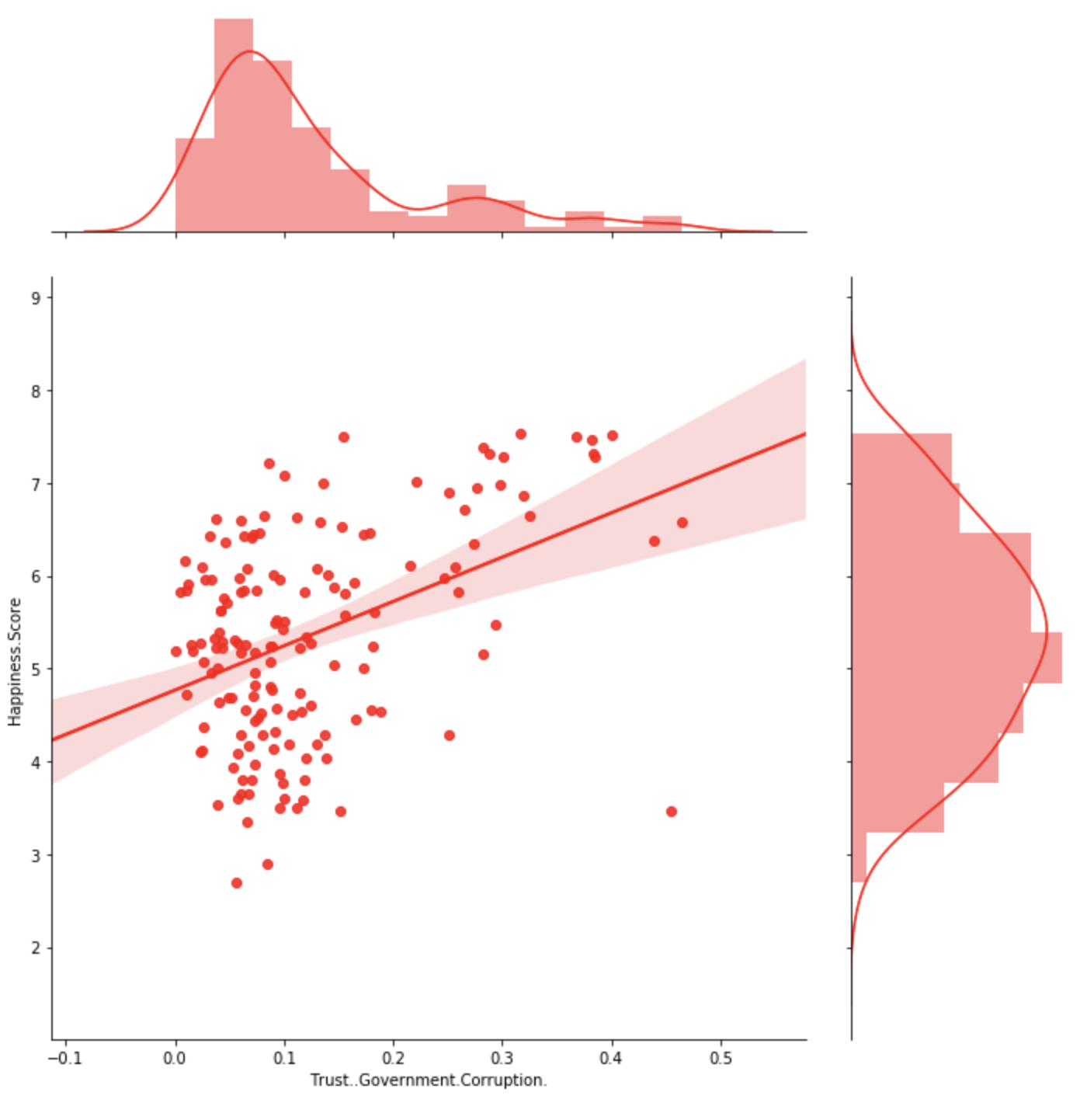


5, By using seaborn.jointplot to plot the linear regression line with normal distributions, we can validate the strong correlation between “Family” and “Happiness score” as we can observe a steep slope. In contrast, in the graph of “Trust/Government corruption” with “Happiness score”, the slope of the linear regression line is not as steep as “Family”, it shows a moderate slope, because Trust (Government corruption) is less correlated with happiness.

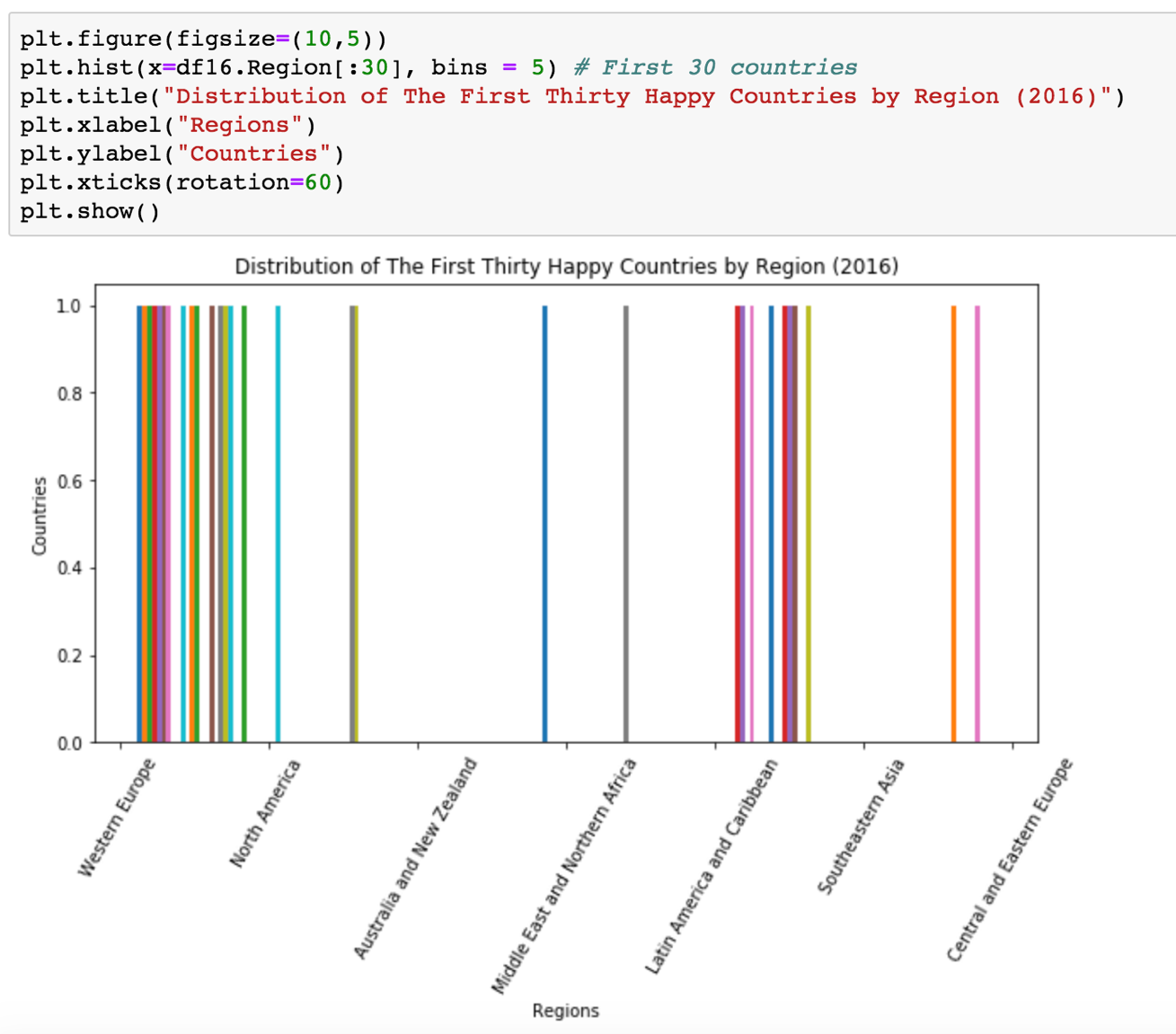






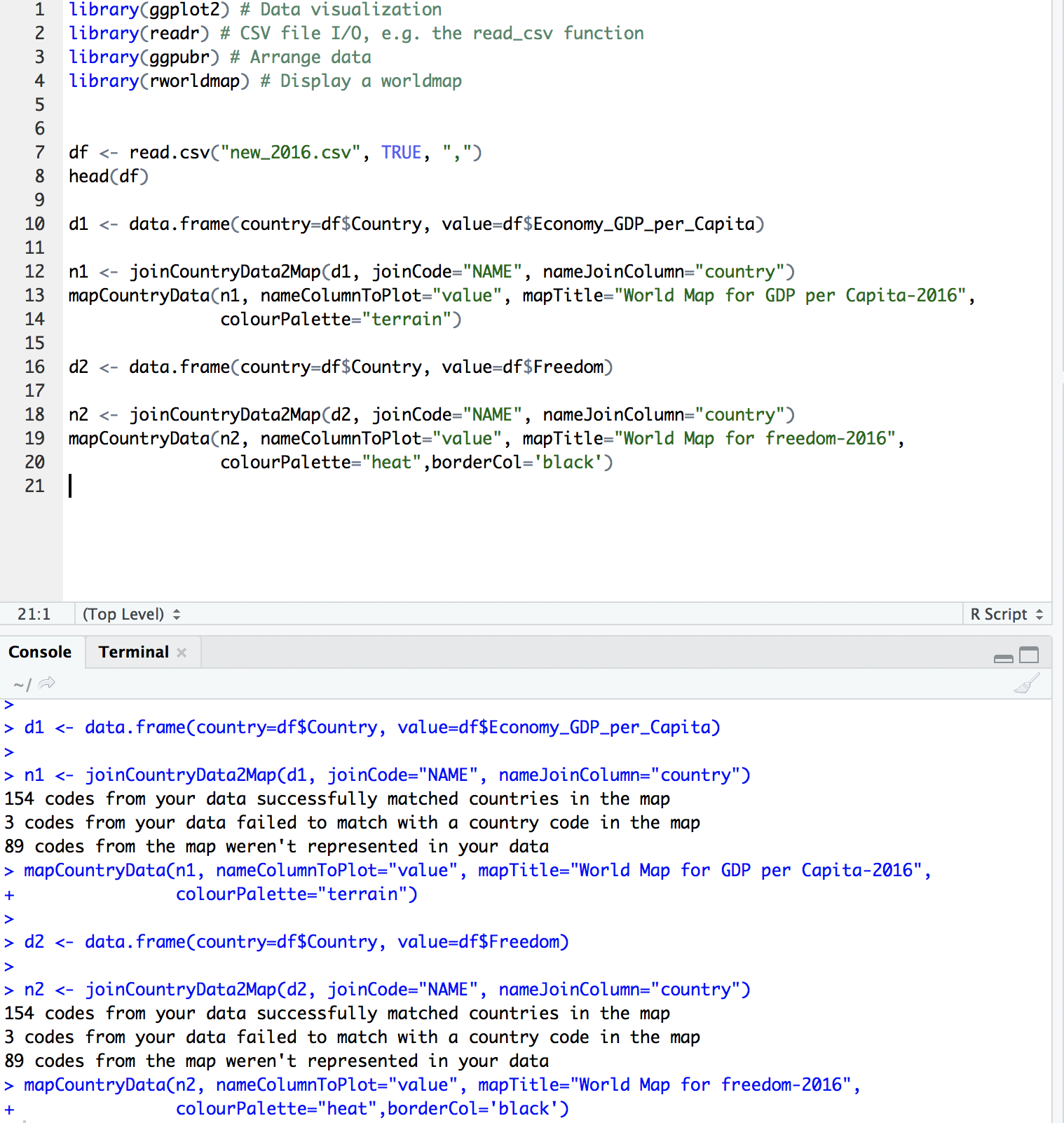


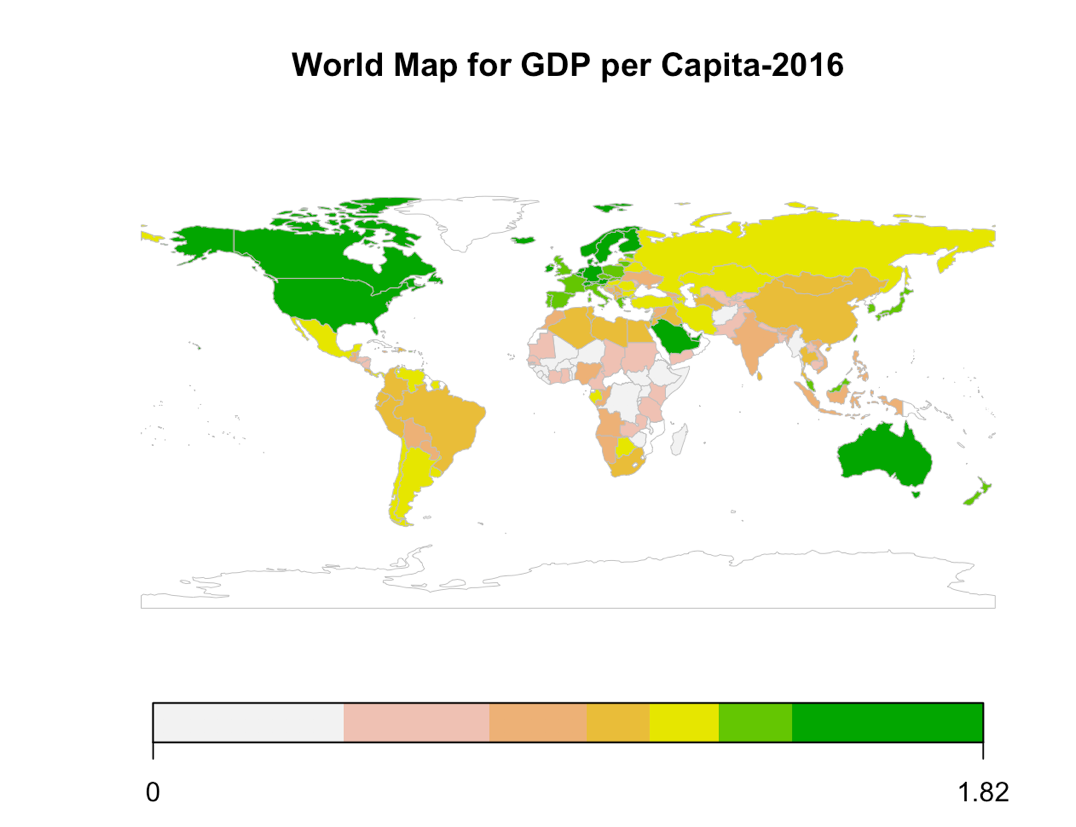
6, By drawing the histogram which consists of top 30 happiest countries by region, we can observe that most countries are located in Western Europe, the following region is Latin American and Caribbean.

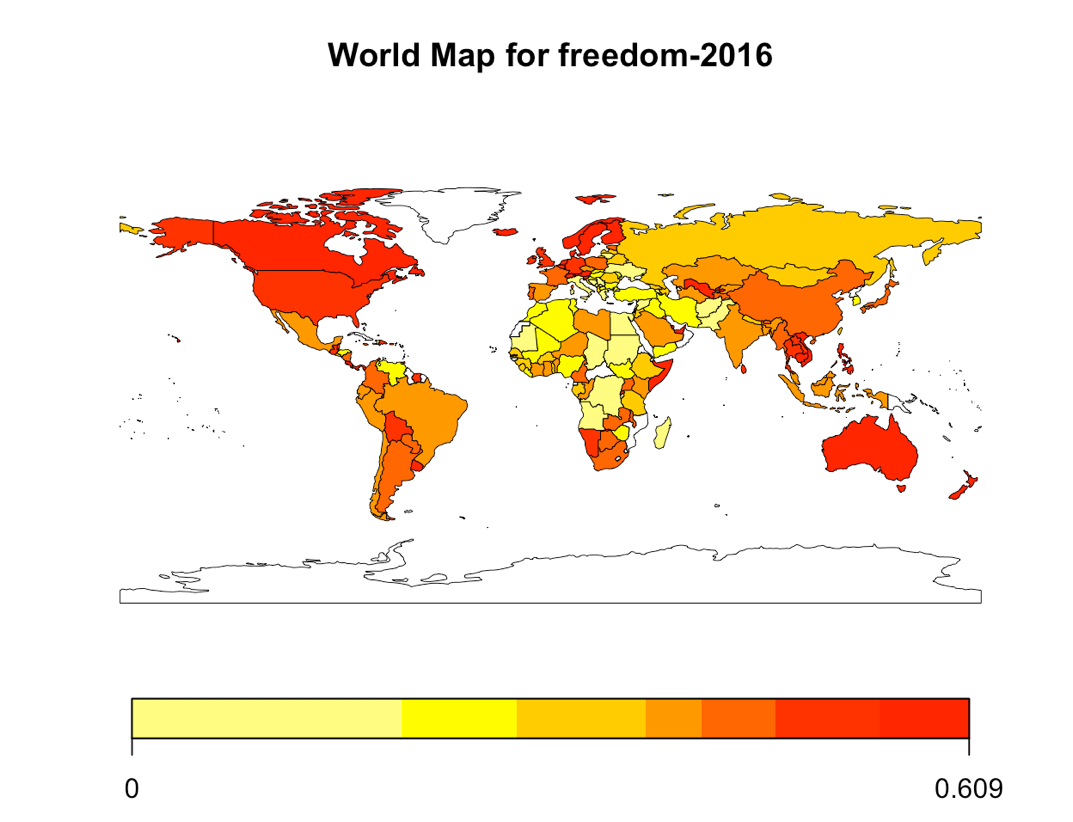


In order to further understand the World Happiness Report, I decided to use R programming language to do further data visualization. Many plotting features may be also available in Python, but some are more complicated to use Python libraries to achieve. The dataset I am using is World Happiness Report 2016.

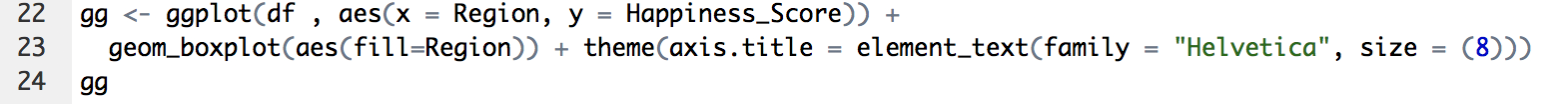
By using the rworldmap library, the R programming provides us a World map to visually view the economy and freedom in each country. The darker the colors, the better the performances are.

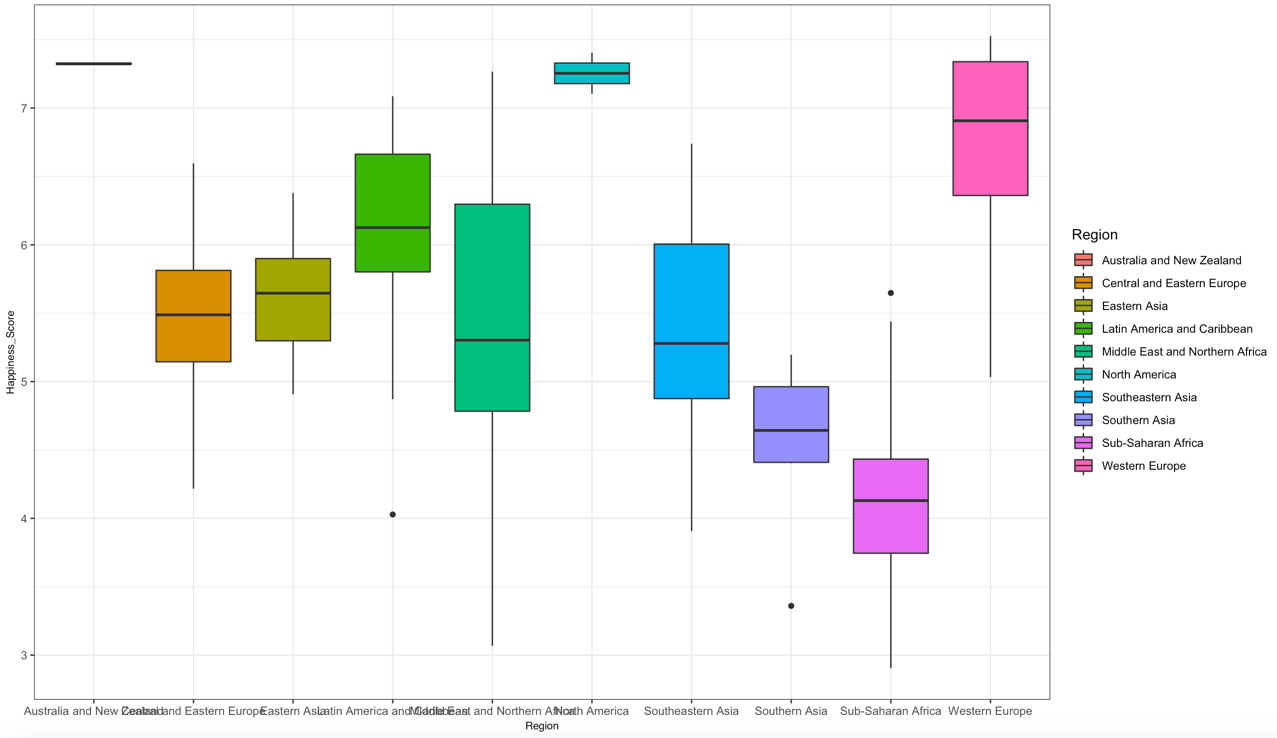






Another useful method to do is showing the box plot. We can analyze the Happiness Score by region. Each box contains 1st quartile, 2nd quartile, 3rd quartile, 4th quartile, whiskers, lower bond, upper bond, median, and outliers. By analyzing, we can see that Australia has the highest median happiness score. (Australia should not be considered as the first here, because there are only two countries in Australia and New Zealand). North America is in the second place, and Western Europe is in the third place regarding median happiness score. Sub-Africa is in the last place. We can see the range of happiness score concentrations for different continents.





In conclusion, by analyzing and visualizing the World Happiness Report datasets, we can have a general idea about how happiness is related with other factors, and we can use this as a guild line to the quality of life. We can see that happiness and economy support each other, and the top happiness countries are mainly located in Western Europe. There always be more advanced tools and analysis that we can do further. Therefore, to learn more about the statistics and data analysis will be a big trend.

Reference

Haybron, Daniel M., 'What is happiness?' in *Happiness: A Very Short Introduction*(Oxford, 2013; online edn, Very Short Introductions online, Sept. 2013), http://dx.doi.org/10.1093/actrade/9780199590605.003.0002, accessed 25 Nov. 2018.

Portilla, J. (n.d.). Learn how to use NumPy, Pandas, Seaborn , Matplotlib , Plotly , Scikit-Learn , Machine Learning, Tensorflow , and more! Lecture. Retrieved from https://www.udemy.com/python-for-data-science-and-machine-learning-bootcamp/learn/v4/overview

Sustainable Development Solutions Network. (2017, June 14). World Happiness Report. Retrieved November 25, 2018, from https://www.kaggle.com/unsdsn/world-happiness/home