MSDS622 Data Visualization Final Project

World Happiness Score Analysis

by Zhe Yuan

I. Data Introduction

First I find this topic about world happiness on Kaggle, which includes data of world happiness score, ranking and each segment score from year 2015 to year 2017. To understand and supplement the data, I do some research online.

Gallup surveys people in more than 150 countries every year, and then the UN Sustainable Development Solutions Network compiles the data into World Happiness report. The reports are released on March 20th each year, which is the International Day of Happiness. The factors used to evaluate people's happiness are economy, family, health, social freedom, generosity and government corruption.

To make the data more integrated, I got data of 2018 and 2019 from the report's website and also downloaded the world GDP and population data of recent 5 years from World Bank open data.

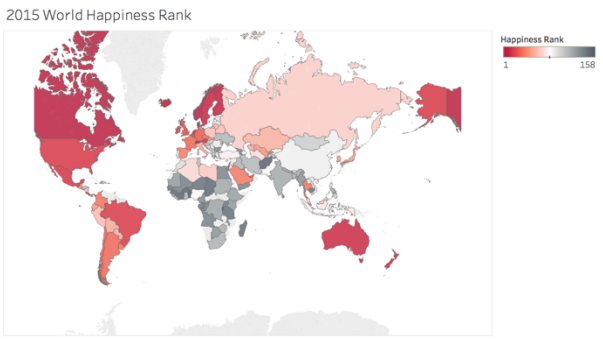
The data of world happiness score and ranking gives us plenty of information about countries' absolute scores and their orders on the leaderboard. Simultaneously, we are able to get more sense about the data if we analyze the time series data from year 2015 to year. After using various types of visualization to summarize the data, I determine to focus on the topic on the trends shown from the changing rankings and explore what caused significant changes in happiness rankings between 2015 and 2019 within countries.

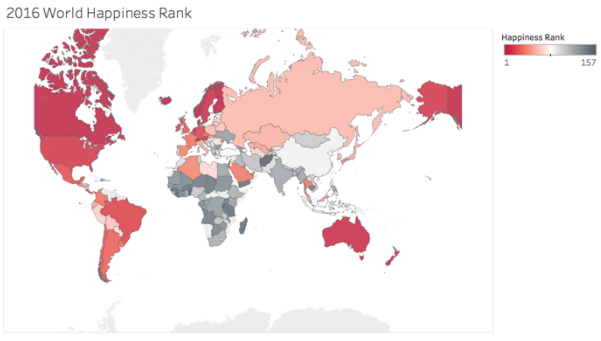
II. Summary of Data

1. Worldwide Attributes

1.1 Overall view of happiness score of each country

The chloropleth maps below show a first glance of the world happiness data. Each map is made by data of one year from 2015 to 2019. The countries are colored by happiness rankings where red color represents higher rankings. When it turns to gray, it reflects that such country stays at a lower location on the leaderboard.





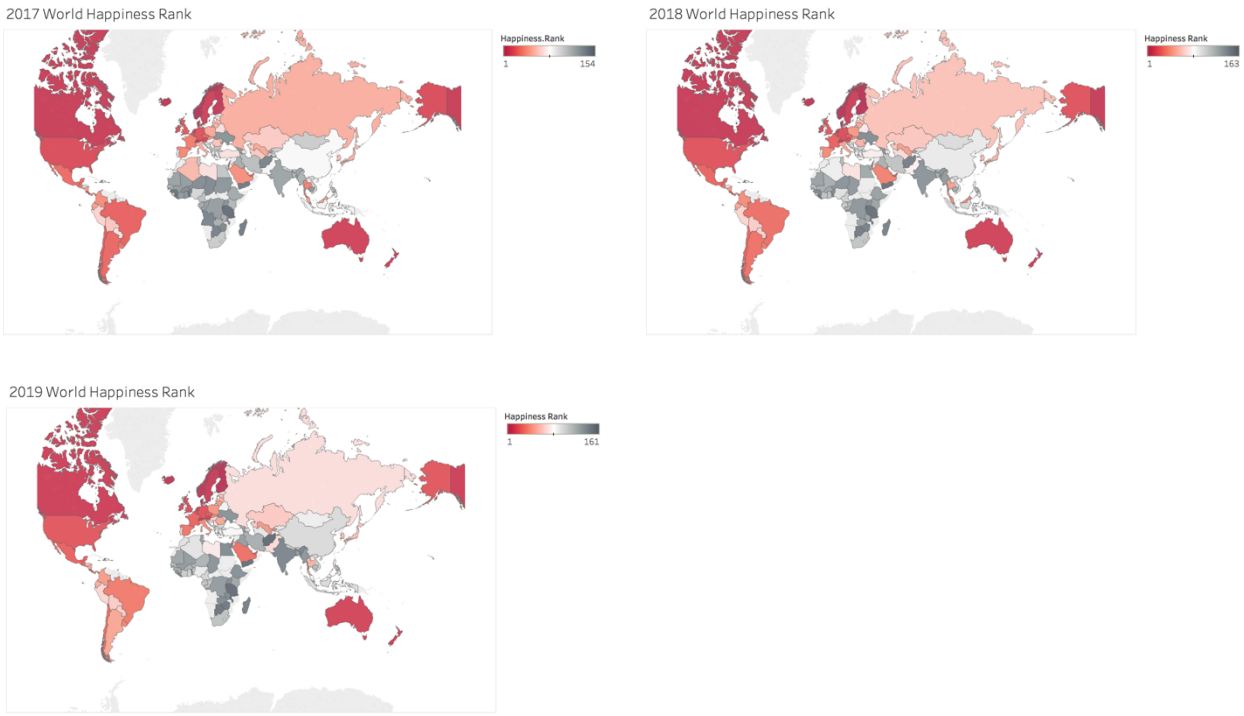


Fig 1 Chloropleth Map

1.2 Distribution of happiness score

The histogram below shows the distribution of average happiness score for each country during the period of year 2015 to 2019.

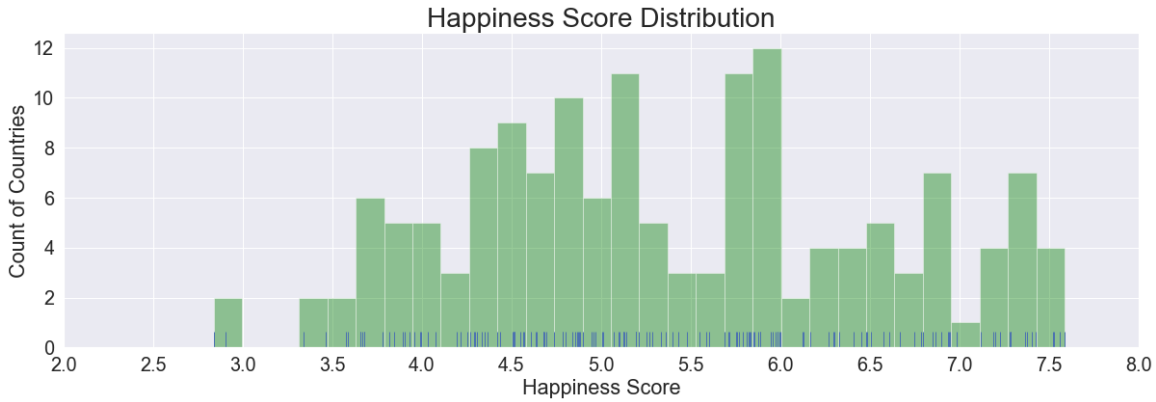


Fig 2 Histogram

1.3 Distribution of happiness score for each region

The boxplot below displays the distribution of average happiness score based on minimum, first quartile, median, third quartile and maximum. The score data is grouped by region all over the world. All countries are classified into totally ten regions, which are Southeastern Asia, Eastern Asia, Southern Asia, Central and Eastern Europe, Western Europe, North

America, Latin America and Caribbean, Middle East and Northern Africa, Sub-Saharan Africa and Australia and New Zealand. We can clearly find the difference in distribution of happiness score between different regions.

Fig 3 Boxplot

2. From the aspect of regional data

2.1 Number of countries (observations)

The barplot below displays how many countries there are in each region.

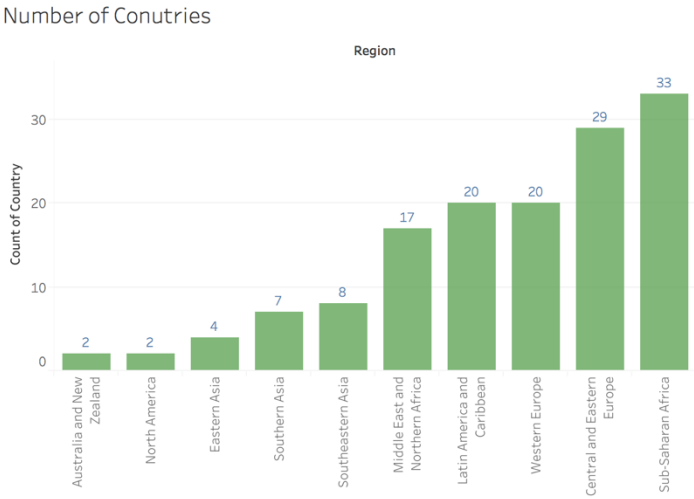


Fig 4 Barplot

2.2 Distribution of population and GDP of each region

The bubble maps below are made based on average population and GDP per region from 2015 to 2019. The size of bubble reflects the amount of population or GDP, which will give us a direct cognition of distinction in economy and population between different regions in the world. Understanding of population and economy will serve as a background knowledge and help us analyze people's happiness.

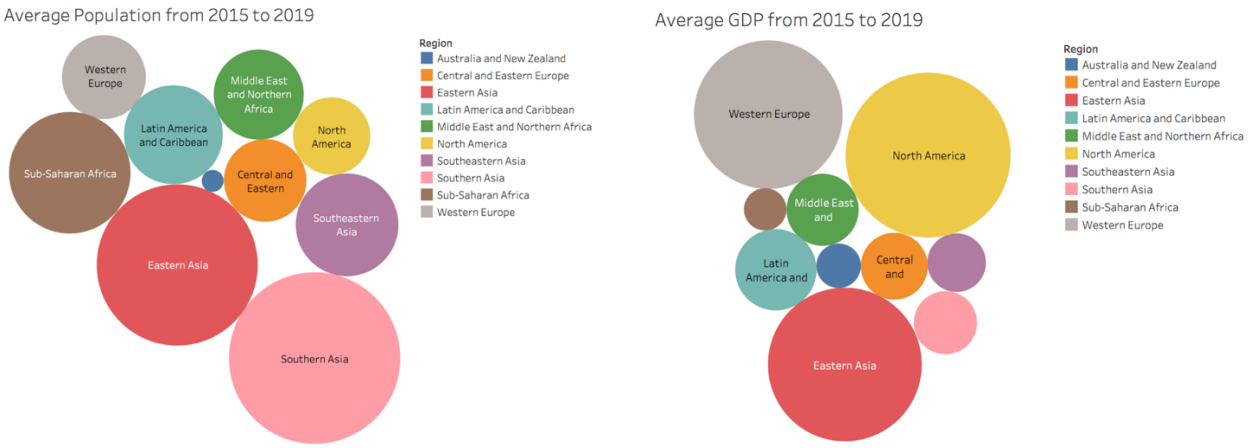
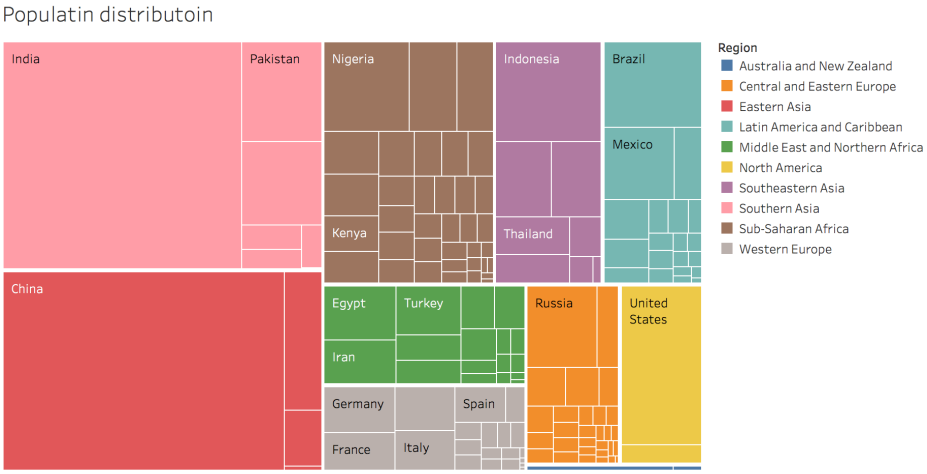


Fig 5 Bubble Map

The interactive tree maps below are made by Tableau and colored by regions. Countries are set as filter so that we will be able to check average GDP and population for each countries from 2015 to 2019.



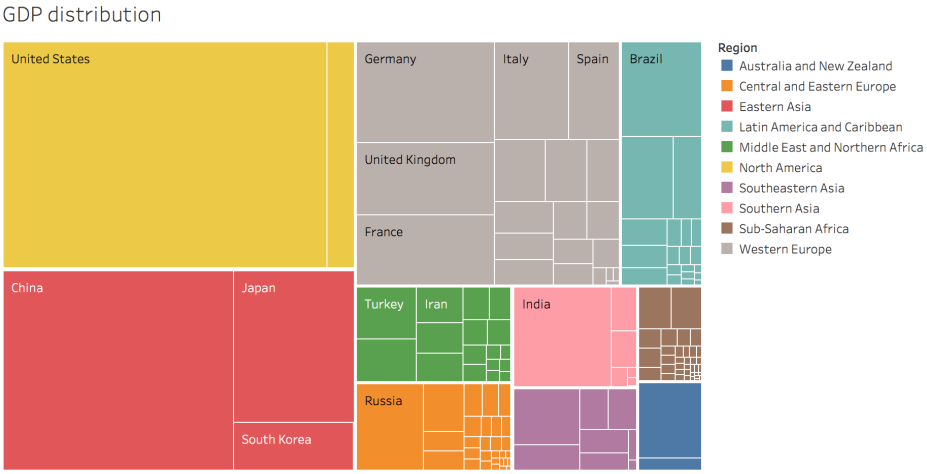


Fig 6 Interactive Tree Map

2.3 Average score and rankings for each region

The line plot below shows trends of average ranking got by countries in each region from 2015 to 2019. The regions which contain a small number of countries, such as Australia and New Zealand and North America stay at a high level. Besides that, countries from Western Europe have always received high rankings.

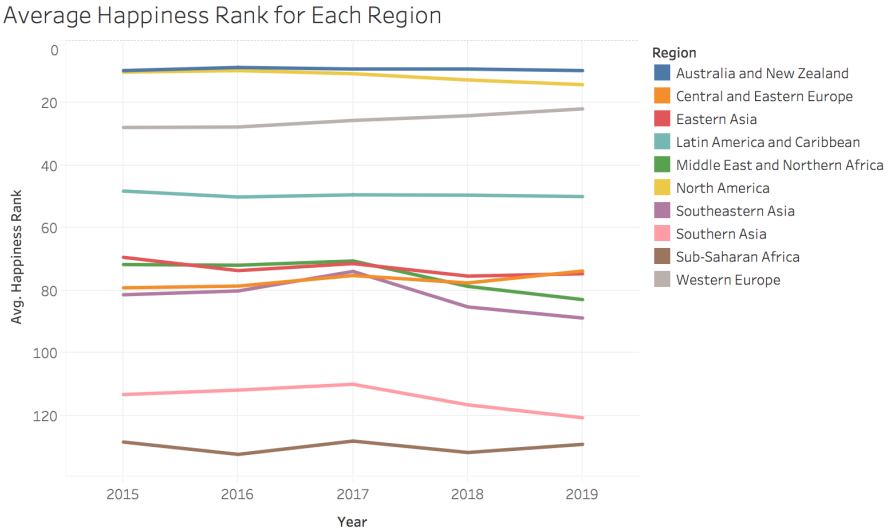


Fig 7 Line Plot

3. Components of happiness score

There are seven components of happiness score given in the UN's happiness score report. Here are the explanation of each component.

Component Name	Explanation
Economy	Score explained by GDP per capita
Family	Score explained by social support
Health	Score explained by healthy life expectancy
Freedom	Score explained by freedom to make life choice
Generosity	Score explained by generosity
Trust	Score explained by perceptions of corruption
Dystopia Residual	Score explained by variation from dystopia

Table 1 Component Explanation

3.1 Changing component score from 2015 to 2019

In the UN's world happiness score report, economy, family, health, social freedom, generosity, government corruption and dystopia residual are the factors used to evaluate

people's happiness. Stacked area graphs are used here to display the average changing of each component from 2015 to 2019 within the scope of total score. Correspondingly, trends for each components are shown in the line plot.

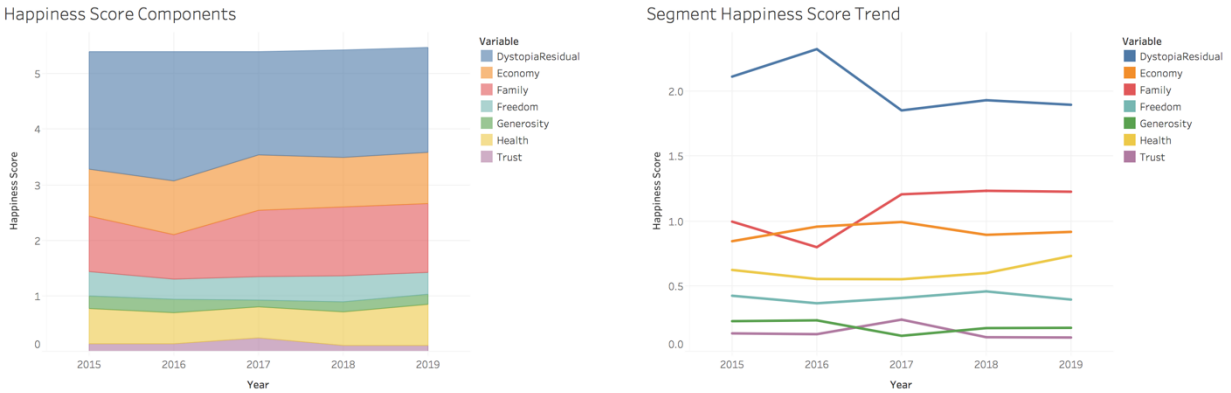


Fig 8 Stacked Area and Line Plot

3.2 Correlation between different features

The heatmap below reflects the correlation between different features, including both positive and negative relationship.

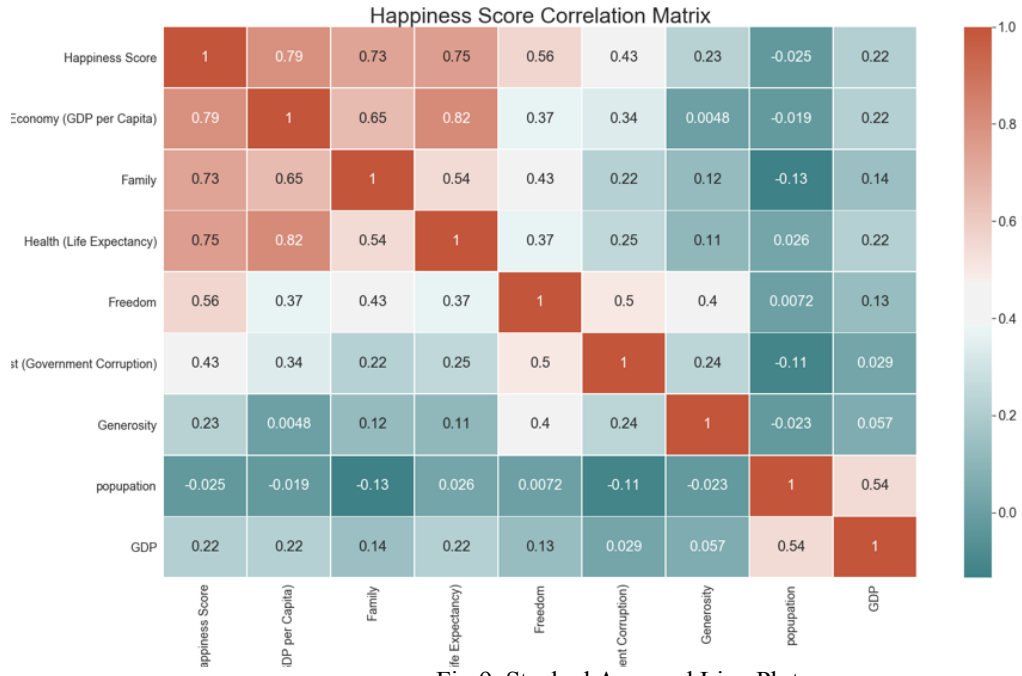


Fig 9 Stacked Area and Line Plot

3.3 Correlation between happiness score and its components

From the heatmap above, we can find that there exists high correlation between happiness score and economy, family and health. Therefore, scatter plots which are colored by regions are used here to display specific relationship.

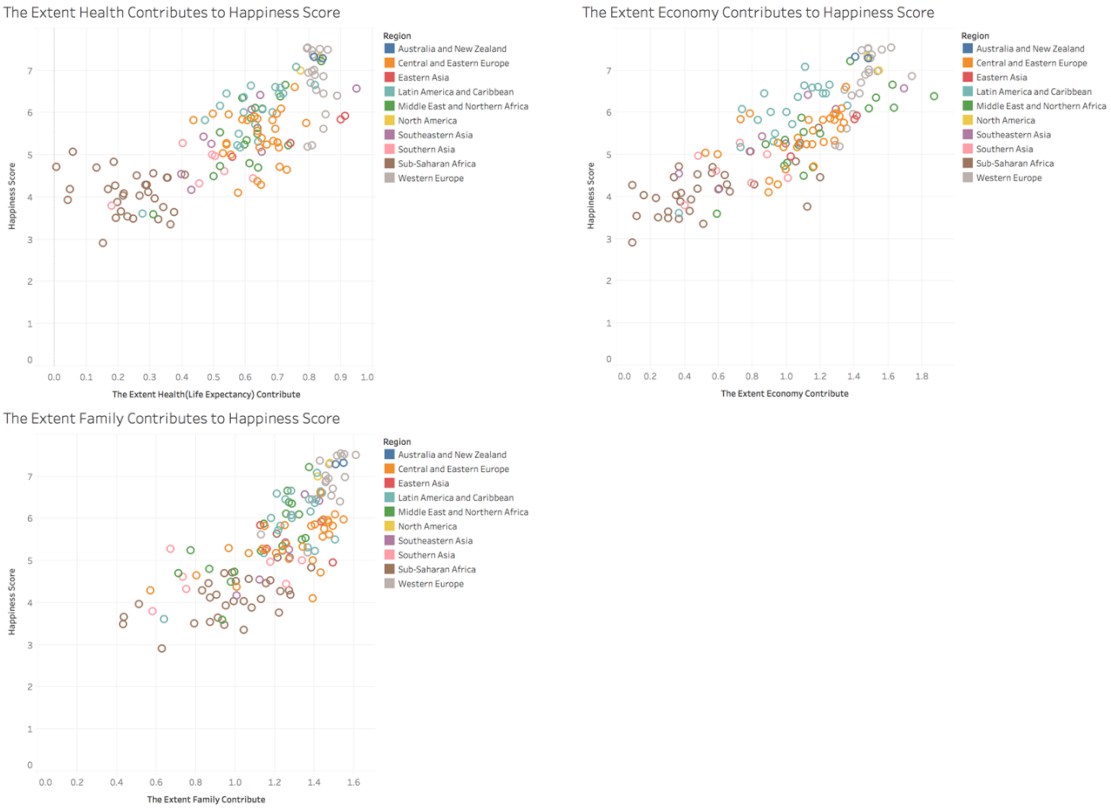


Fig 10 Scatter Plot

3.4 Connection between happiest countries

Since the data used in this analysis is not transportation-related, connection map cannot be used to show the shortest distance between different location. Here is a simple instance of connection map which links the top 5 happy countries in the world.

Fig 11 Connection Map

III. Storyline

The countries are ranked again based on their progress in happiness ranking, and the barplots below shows the top 20 countries who made greatest progress in happiness ranking, which means people from these countries became much happier than they were 5 years ago, as well as the bottom 20 countries who had the most obvious regressing in happiness ranking, which may reflect people are facing some issues in certain aspects.

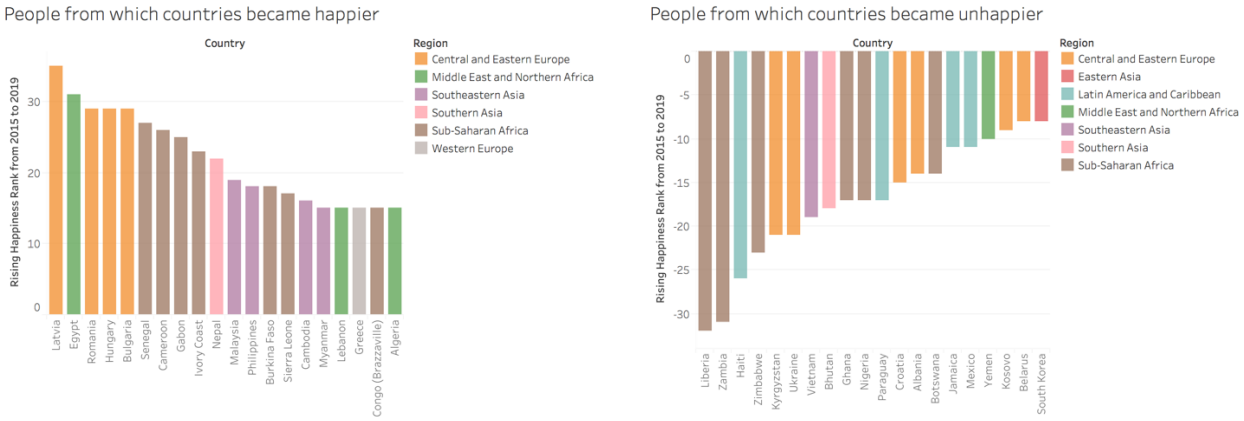
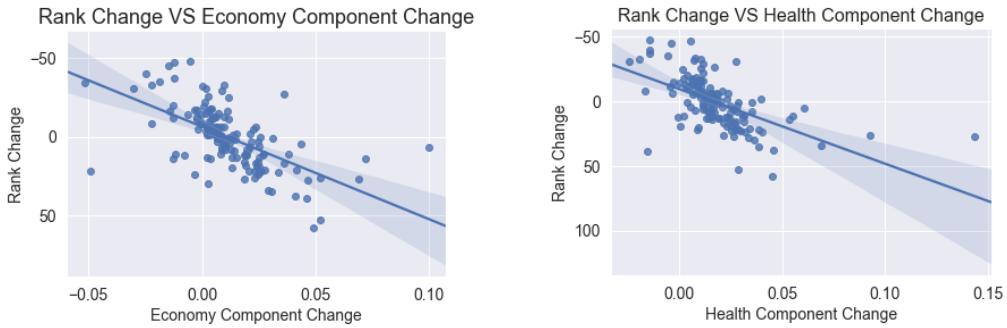


Fig 12 Barplot

To explore what caused significant changes in happiness rankings between 2015 and 2019 within countries, scatter plots and regression lines are used to display the relationship. Changing rankings are shown by y-axis. For each component score of happiness, I calculate what percentage of the total score it made up for year 2015 and 2019, calculate the variation between these two percentages, and use x-axis to show the variation. Therefore, the relationship of changing rankings versus changing component score percentage can be detected by the plots below.



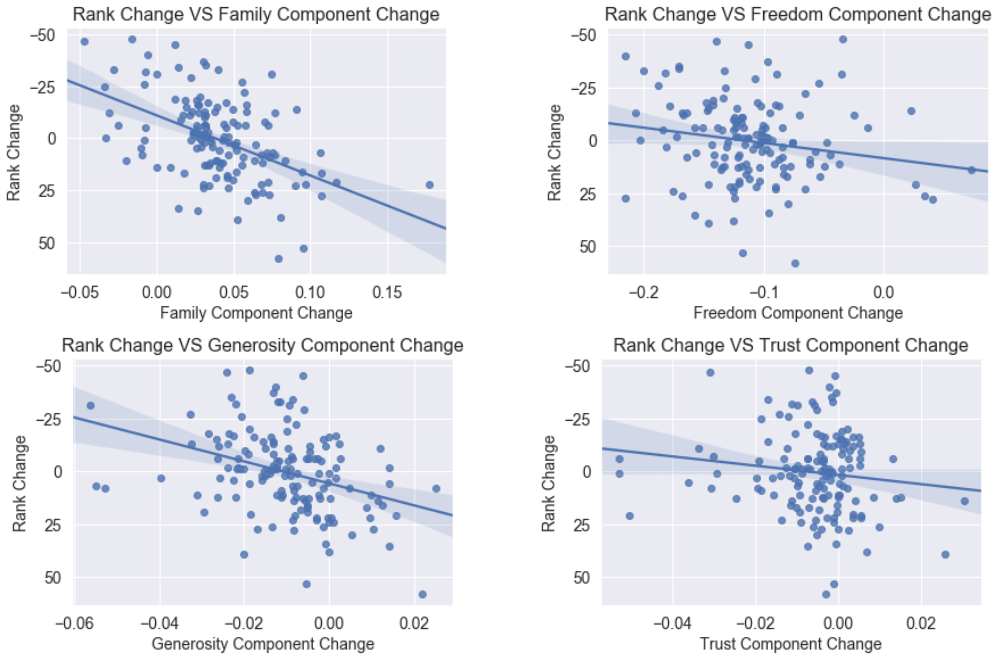


Fig 13 Scatter Plot and Regression Line

Conclusion

Generally speaking, people from Oceania, North America and Western Europe are always at a high level on the leaderboard of happiness score due to these countries' economy, politics, geographical location, natural conditions and many other factors. However, if we focus on the changing trend instead of the absolute values, we will find more information and become more confident about the countries who made greater progress on the happiness rankings. When talking about each component of happiness score, including economy, health, family, freedom, generosity, trust to government, the plots above show a surprisingly negative correlation. If the percentages of happiness score explained by economy, health and family increase, this country's ranking will decrease. The relationship between happiness score and freedom, generosity and trust do not show obvious negative correlation.

IV. Github link

https://github.com/zheyuan15/data_viz