Global Happiness Analytics: A Data-Driven Visualization of Well-Being Across Nations Report

Group 25 Member:

Chen Zixiao(3036410316)

Hu Hongyi(3036410029)

Li Dayu(3036411152)

Li Xinlei(3036410720)

Xu Yue(3036410770)

1. Foreword

Global happiness analysis is a complex and important research area, vital for understanding the quality of life, social welfare, and the effects of policies across different countries and regions. With the rapid development of globalization and information technology, happiness has become one of the key indicators for measuring social progress and public satisfaction. By utilizing data visualization methods to analyze global happiness, this project aims to reveal regional disparities in happiness levels and the socio-economic factors behind them, providing scientific evidence for policymakers and promoting global well-being.

2. Our work in bullet points

Multidimensional Visualization Analysis: A global happiness analysis dashboard was created, integrating multiple charts to provide coordinated multi-chart interactions. This allows users to explore happiness data from multiple dimensions, offering a deeper understanding of the data.

Exploring Regional Happiness Distribution: Using visualizations such as maps and bar charts, the distribution of happiness across different regions was presented, revealing significant differences between regions.

Study of Key Factors and Their Relationship with Happiness: The dataset was cleaned and linked with multiple related datasets. Using visualization tools such as Tableau and Plotly, along with regional happiness distribution data, a deeper analysis was conducted on the relationship between factors such as GDP per capita, educational inequality, colonial history, and national health levels with happiness.

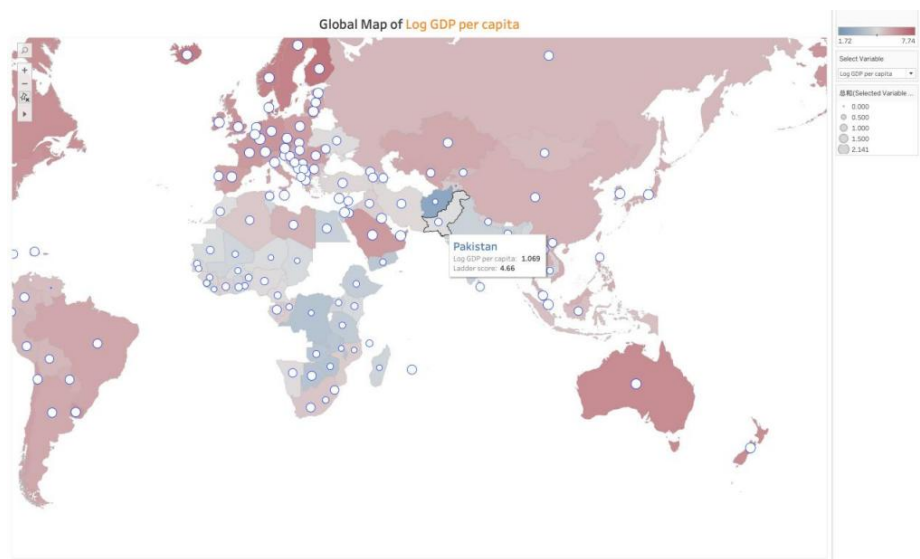## 3. Exploring Key Factors and Regional Disparities

The study of global happiness is an important field that has drawn significant public attention, as it helps us understand the differences in happiness levels across regions and the driving factors behind them. In this analysis, we focused on five key parameters from the 2024 World Happiness Report dataset: generosity, healthy life expectancy, GDP per capita, perceptions of corruption, and social support. Using data visualization methods, we aimed to reveal their distribution characteristics on a global scale and the disparities between regions. Below is the visualization plan we developed based on the data characteristics.
Data Preparation

Firstly, we divided the analysis variables into three categories: first, the happiness index (Ladder Score) was selected as the core research subject; second, five key socio-economic variables were chosen for correlation analysis, including generosity, healthy life expectancy, logarithm of GDP per capita, perceptions of corruption, and social support, to explore their impact on the happiness index; moreover, national and regional indicators, as well as other auxiliary variables, were incorporated to facilitate subsequent geographical grouping analysis and to provide supplementary explanations for the model. To study the yearly changes in the data, we incorporated the time variable (Year) from the dataset and combined it with other indicators, such as Country Name and Life Ladder, to analyze cross-year trends. Additionally, we created calculated fields to extract the maximum (MAX) and minimum (MIN) values of variables during the change process, such as Generosity and Healthy Life Expectancy at Birth, to comprehensively present the dynamic changes in the happiness index and its driving factors.


Visualization Process

This map visualizes the global distribution of the happiness index (Ladder score) and a selected parameter, with color and circle size representing the values of the happiness index and the selected parameter, respectively. The color gradient from red to blue reflects the transition of the happiness index from high to low, while the size of the circles represents the magnitude of the selected parameter (e.g., social support). The choice of red and blue as the color gradient is not only because red is often associated with happiness, warmth, and positive emotions, while blue is linked to calmness or negative emotions, but also because the strong contrast between these two colors makes it easier to distinguish data levels. Additionally, the red-to-blue gradient is particularly suitable for representing continuous data trends from high to low. Regions with higher happiness levels, such as Northern Europe, North America, and Australia, are typically accompanied by larger circles, indicating higher values for social support or other selected variables. Conversely, regions with lower happiness levels, such as Sub-Saharan Africa and parts of South Asia, have smaller circles, reflecting weaker performance in these parameters. This visualization highlights the potential relationship between the selected parameter and the happiness index.

In the exploratory analysis, we selected five parameters sequentially to observe their relationship with the happiness index. However, relying solely on circle size and color does not provide a clear observation of specific correlations, prompting us to conduct further analysis. We used a line chart combined with a timeline to visually present the changing trends of different parameters across different years. To clearly highlight key points, we created new calculated fields, Max and Min, to identify the maximum and minimum values of the variables, making it easier to pinpoint critical time periods. As shown in the figure, this method not only reveals the overall trends of the variables but also highlights key years, providing a clear perspective for analyzing the dynamic changes of the variables.


When we visualize social support as shown in the chart, it reveals its changing trend from 2004 to 2024. Overall, the variable shows a significant increase from 2004 to 2010, followed by a slower growth rate and slight fluctuations between 2015 and 2018. However, a noticeable decline occurs after 2019, potentially influenced by external factors. By marking the maximum and minimum values, the chart clearly highlights the variable's peak (around 2017) and lowest point

(2004). By visualizing five different parameters, We observed that significant fluctuations occur during major global events such as the COVID-19 pandemic, particularly in GDP. To further validate my observations, We plan to apply new visualization methods in the next step. To further explore the factors influencing the happiness index, based on the dataset, We decided to combine the Ladder Score with regional information to analyze countries within the same region that exhibit significant differences in their happiness index. This approach can help identify which parameters are responsible for the noticeable variations in happiness levels among countries sharing similar geographical, cultural, or economic contexts.
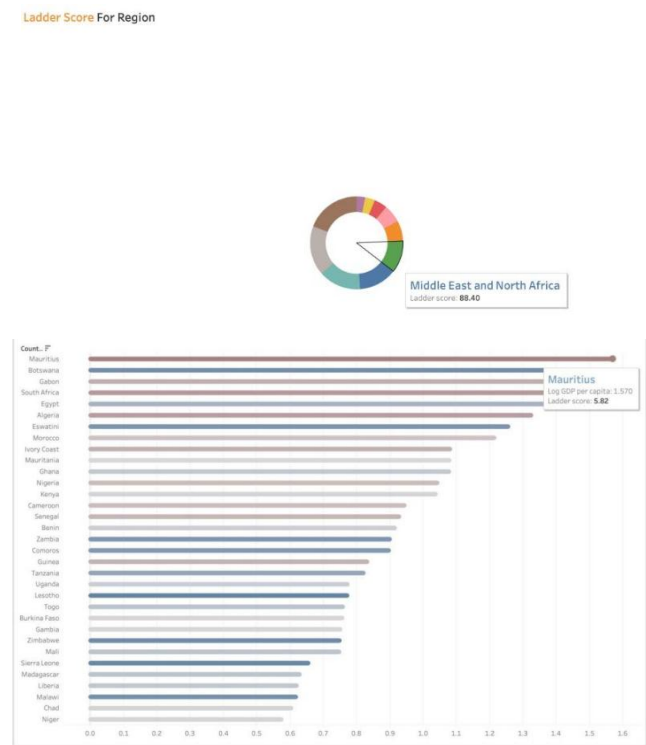
To further explore the factors influencing the happiness index, based on the dataset, We decided to combine the Ladder Score with regional information to analyze countries within the same region that exhibit significant differences in their happiness index. For example, as shown in the chart, the Middle East and North Africa region has a Ladder Score of 88.40, while the lowest region is South Asia. From the visualization, We observed that regions with lower happiness indices and those with higher indices often include countries that were historically colonies, particularly under British and French rule. Since the happiness levels of such colonies have been studied in the past, We plan to filter out countries that were former British and French colonies to specifically analyze the distribution of the five parameters within these countries. Therefore, for the second chart, We chose a bar chart to visualize the selected countries. The chart uses color to represent the happiness index levels and bar length to display the magnitude of the parameters. Next, We will conduct a more detailed analysis of the happiness index.
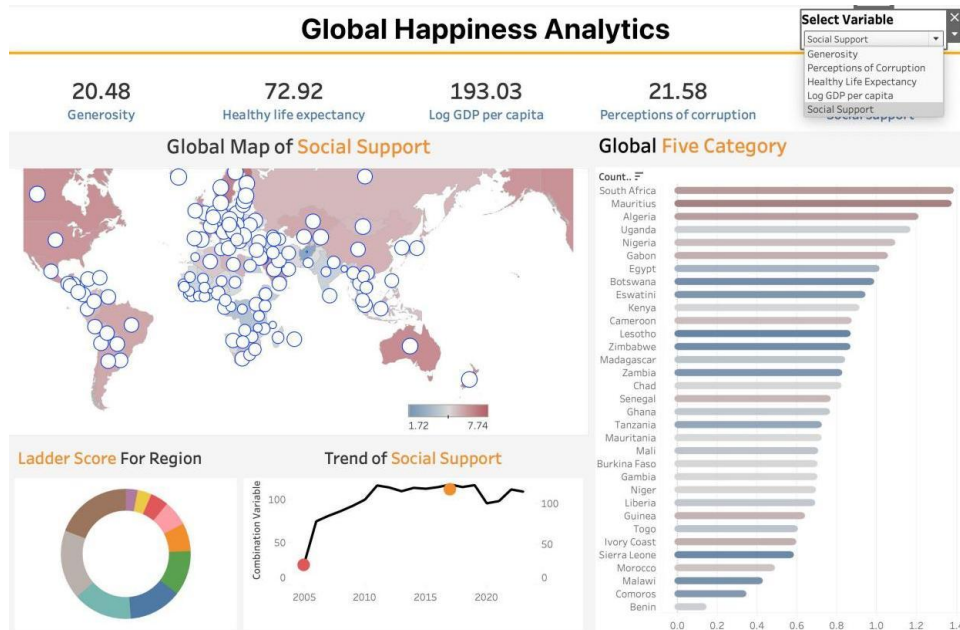
The "Global Happiness Analytics" dashboard integrates the aforementioned charts and achieves coordinated multi-chart interactions, enabling multi-dimensional visualization. The top section provides an overview of key statistical indicators, including global averages of Generosity, Healthy Life Expectancy, Log GDP per Capita, and Perceptions of Corruption, offering a reference for overall trends. The global map uses a color gradient to represent happiness index levels, with circle size reflecting the magnitude of social support, making it easier to identify geographical distributions and outliers. The pie chart displays the happiness index distribution across different regions, facilitating regional comparisons. The line chart illustrates the trend of social support from 2004 to 2024, with clearly marked peaks and troughs highlighting critical turning points. The bar chart on the right represents parameter magnitudes through bar lengths and happiness index levels through colors, providing a detailed perspective for comparing parameters across countries. Users can also use the dropdown menu in the top-right corner to switch parameters, enabling dynamic exploration of the relationships between different variables and the happiness index. After conducting cross-dimensional analysis of the dashboard by region, country, and time, We found that happiness index is strongly associated with GDP and social support. Additionally, global happiness levels are not only influenced by historical and geographical factors but also shaped by cultural and social systems.

## 4. Regional Distribution of World Happiness

One of the important tasks of world happiness visualization is to present the regional distribution of world happiness. By visualizing the happiness comparison of different regions or countries, we can intuitively feel the level of world development. This also helps us further highlight the successful cases of some countries or regions and explain their measures to improve happiness.

Data Preparation

According to the dataset, all countries in the world are divided into the following regions: **Western Europe, North America and ANZ, Middle East and North Africa, Latin America and Caribbean, East Asia, Southeast Asia, Central and Eastern Europe, Commonwealth of Independent States, Sub-Saharan Africa, South Asia.** Based on the above division, we first counted the average happiness score of each country from 2019 to 2024, and then distributed it according to the regional attributes provided by the dataset, so that we got the sum of the happiness scores of each region. Then, for each region, we calculated its mean value according to the number of countries and we can get the happiness score of each region. From the final effect, the happiness score divided by region is also in line with our common understanding. Countries with high happiness index are concentrated in Western Europe, North America, Australia and other regions, while countries with low happiness index are concentrated in Africa.

Visualization Process

Horizontal Bar to visualize the average happiness score in recent five years
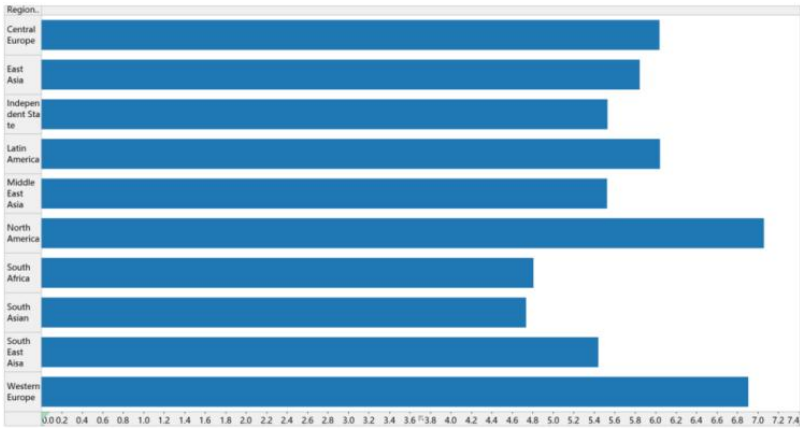
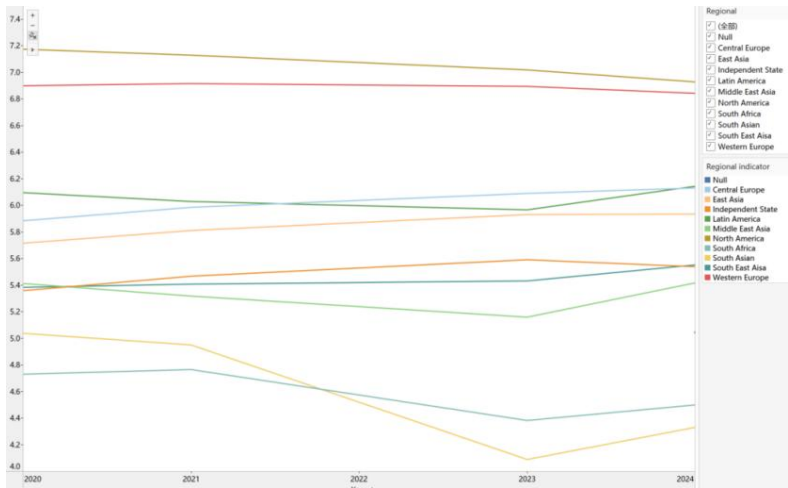Line chart to visualize the country happiness score year by year
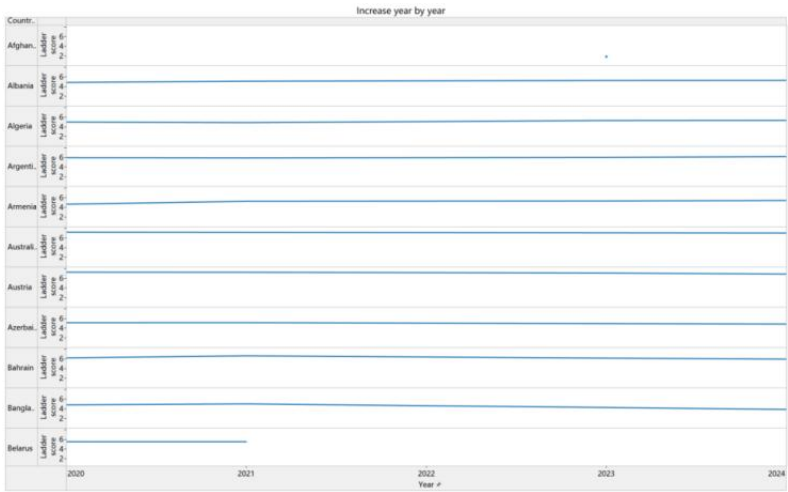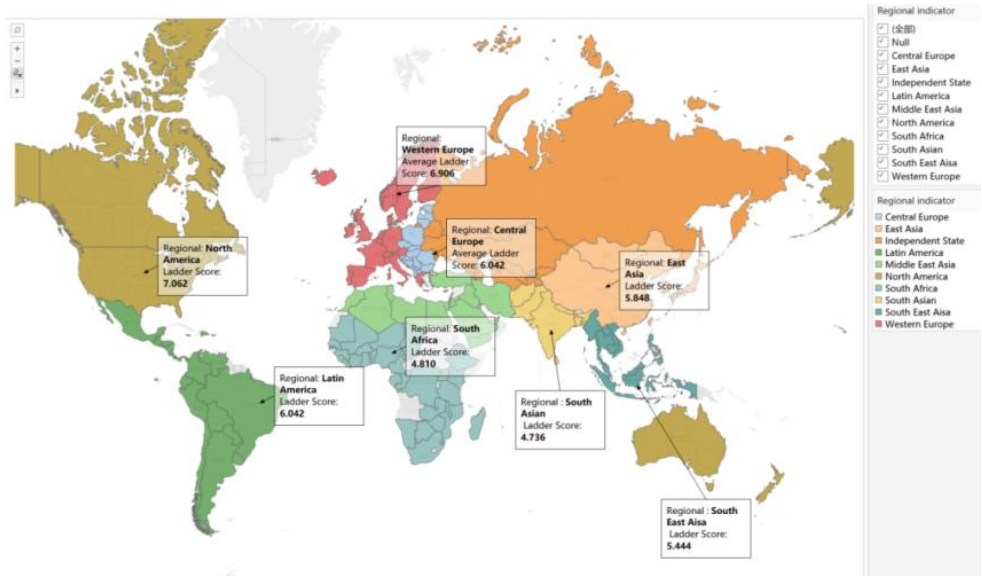
Use Score range table as selection

Table1. Score Range Table

Score Range	Country Nums
3.0-4.0	11
4.0-5.0	29
5.0-6.0	41
6.0-7.0	45
7.0 or higher	11

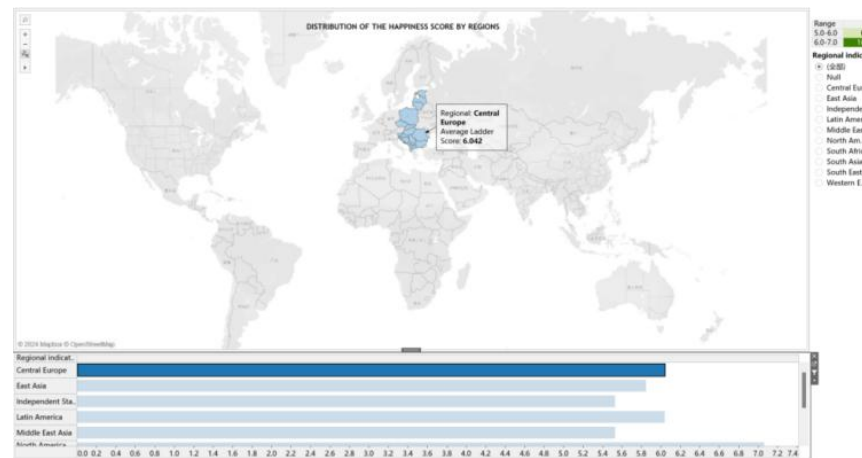Line chart to visualize the region happiness score year by year
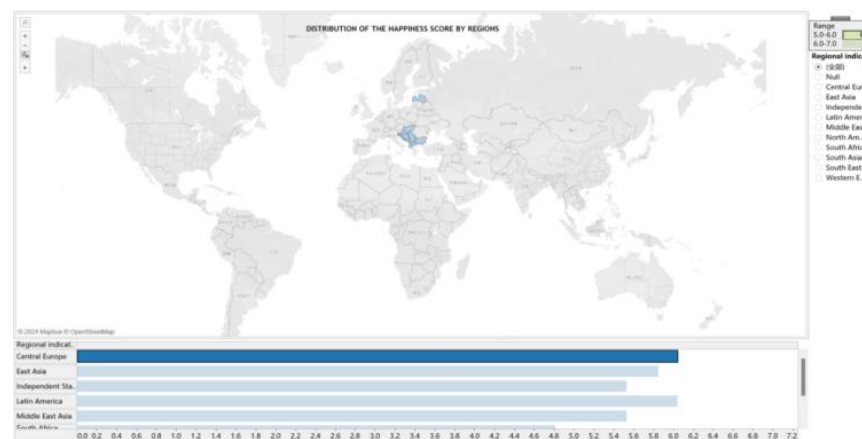

Integrate into map visualization

dashboard

Regional Selection：We could select the region in the line chart to view the regional countries distribution.

Score Range Selection：We could select the score range table to further filter the countries.

Limitations

Missed Values

The happiness score of many countries over the years is missing in the source data, which means that some countries are not included. This missing data will affect the accurate assessment of the happiness of each region, which means that the conclusions we draw when analyzing a country through happiness scores in different years are not sufficient.

The algorithm needs to adjust the weight

In the original algorithm, we simply added together the happiness index of all countries in each region and took the average. This algorithm assumed that all countries in the same region were equally important, but this is not the case in reality. The influence of some countries in the region may be greater than the sum of the other countries. For example, in East Asia, China's influence is obviously greater than the sum of the other countries in the region. Simply taking the average may not be very reasonable. We should also assign weights to each country based on its

economic level and population size.

The information should be more specific

The current visualization may lack specific information about the factors contributing to the happiness scores in different regions. For instance, breaking down the happiness index into sub-components like income, social support, life expectancy, freedom, generosity, and perceptions of corruption can provide deeper insights. Moreover, including historical context or socio-economic data alongside the happiness scores would enhance understanding. This specificity would enable policymakers and researchers to identify targeted interventions and better understand the dynamics of happiness in various cultural and economic contexts.
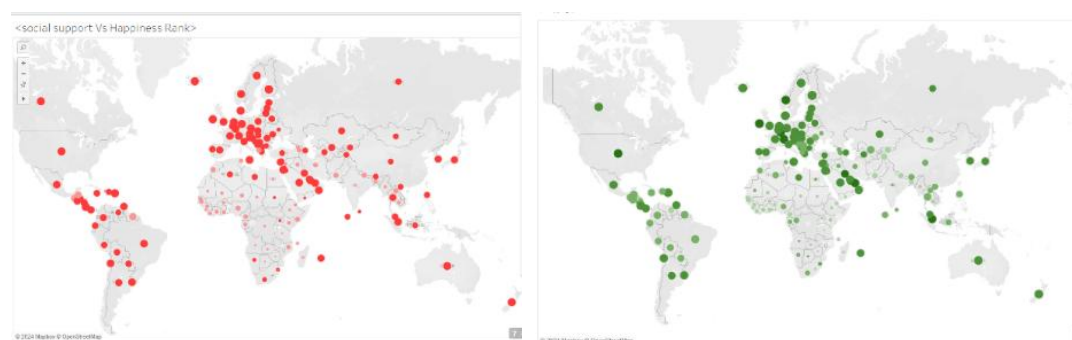
5. Social Support

Data Preparation

In this module we used nearly ten years of data, from 2015-2024. We added a list of attributes to all the data, indicating that the area had been colonized or heavily influenced, with three values: British, French, and other.

We also calculated the average happiness score across all regions, then subtracted this average from each region's happiness score and applied a logarithmic transformation to enlarge the distances of happiness scores among different regions.

Visualization Process

In this part, we take social support as the main entry point and analyze its impact on happiness score combined with other related attributes. And some potential data characteristics of the adopted world happiness report dataset. Visualize each attribute and happiness sore on the world map: The darker the color, the larger the attribute value, and the higher the happiness score, the larger the circle .And we try to find attributes similar to the distribution trend of social support (red in figure   below) and combine them to analyze the impact on happiness score. It is found that the distribution trend of the influence of GDP on happiness (green in figure below) is similar to that of social support: the larger the dot, the darker the color.


According to the above visual exploration, we assume that both social support and GDP have a strong positive correlation with happiness score, and we use linear regression to calculate its correlation coefficient to verify our conjecture:
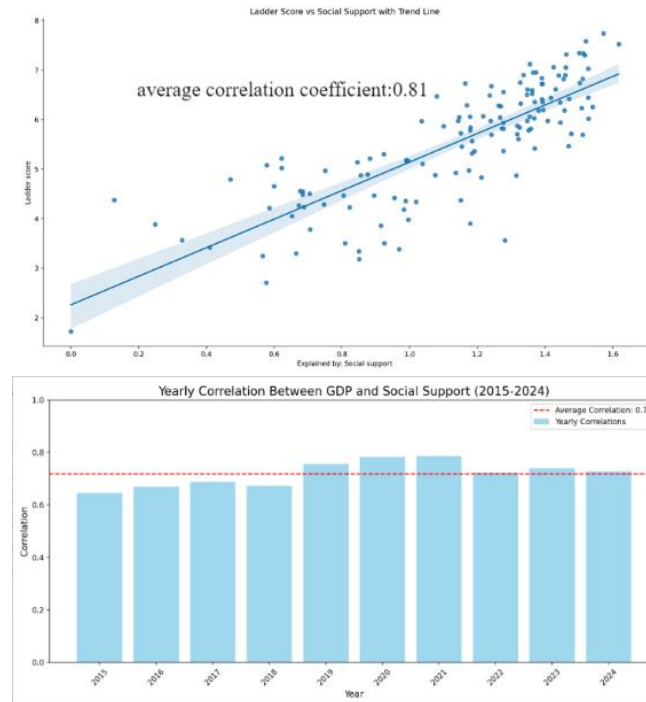
It can be seen that social support has a strong positive correlation with GDP and happiness socre. Therefore, it can be concluded that regions with higher GDP generally have higher investment in social support, which will improve the happiness level of the region.

Combined with the visualization of the world map, it is found that the happiness score of several former British colonies is obviously higher than that of former French colonies. Therefore, it is assumed that this pattern is also applicable worldwide. The data was processed to add British or French attributes to some areas (those that were colonized by both countries or had a history that was heavily influenced by both countries). Here comes a difficulty: Because the higher value of happiness score in the dataset is about 7 and the lower value is about 5, there is no significant difference between the two numerically, and it is difficult to distinguish them directly for visualization. Even though happiness in the British and French regions has distinguishable distribution features, we tried PCA to extract the features here. However, the effect is not obvious (figure below):
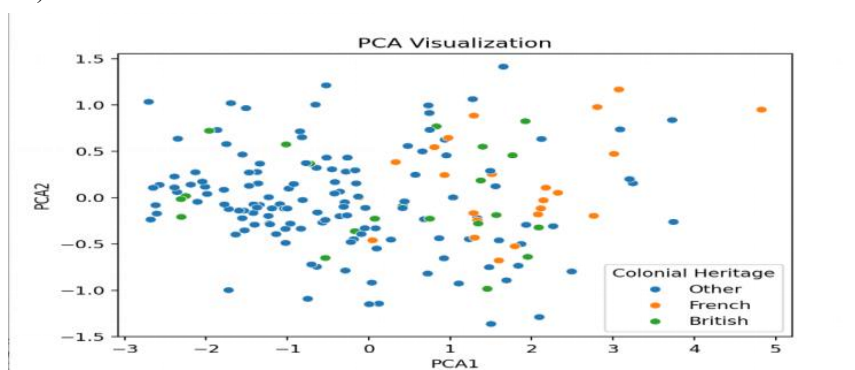

Then we use the density distribution map for observation, and we can see that the happiness score in the British area has a better distribution trend than that in the French area, and it also has better GDP and social support.

Distribution of Happiness Score

Distribution of GDP

Distribution of Social support
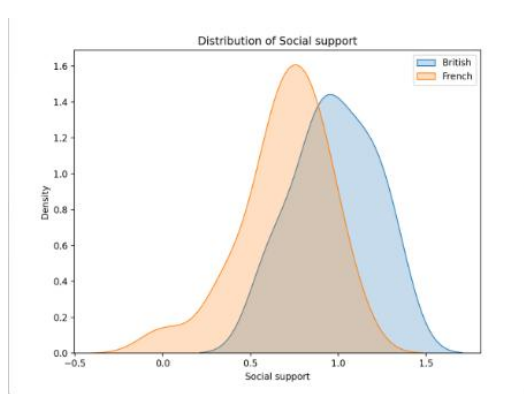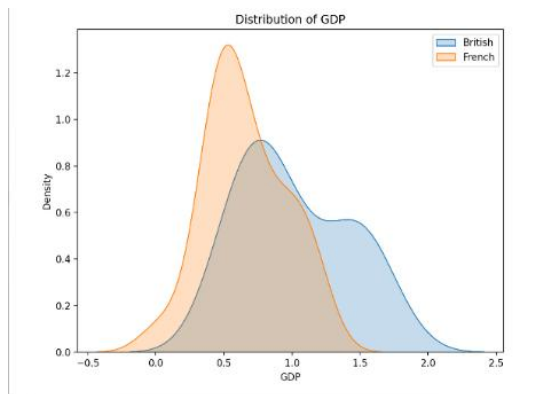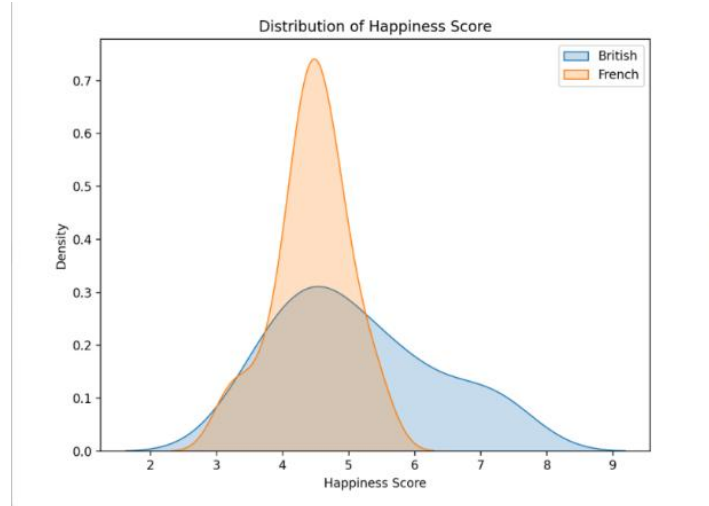
However, while this visualization method allows us to observe general trends, it falls short in clearly distinguishing the data characteristics of the two types of regions. Additionally, it is not particularly effective for identifying anomalies (regions that deviate from the expected patterns based on our hypotheses), should they exist. Therefore, we further processed the happiness scores with the goal of amplifying the differences between them. The method applied was as follows: we calculated the average happiness score across all regions, then subtracted this average from each region's happiness score and applied a logarithmic transformation.

If British and French colonial regions exhibit distinct characteristics in terms of happiness, the visualization of the processed data with enhanced differences should allow for a clear distinction between the two types of regions. It can be seen that the happiness score of the former British countries in the world is indeed higher than that of France, and the GDP and social support are also at a higher level.

Therefore, it is speculated that the former British colonists paid more attention to the economic construction of the colonies than the French colonists (either social support), so that 100 years later, the once left on today's happiness score can still distinguish the impact of discovery rules. But in the figure 2.8 we can still find two obvious anomalies: India and Botswana, which were once part of Britain, are very unhappy.
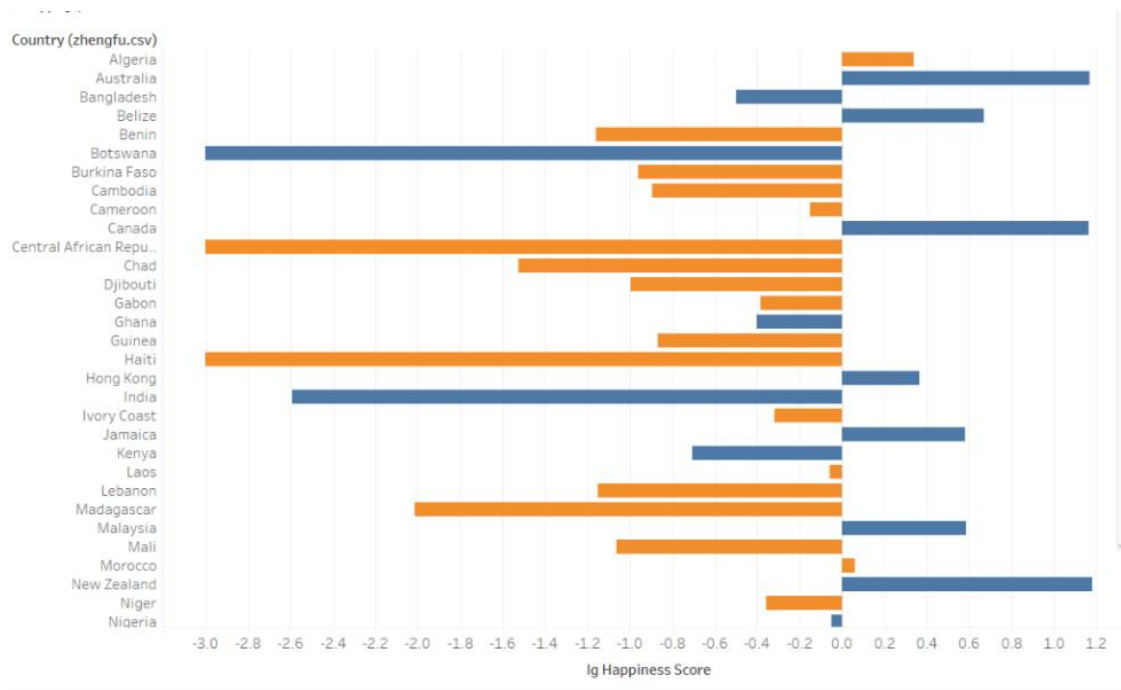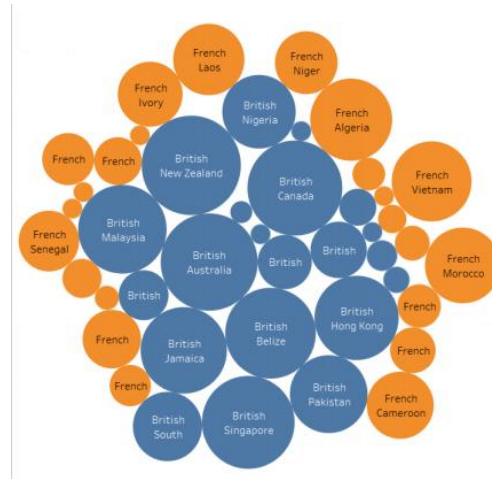
6. The Correlation Between Alcohol and Happiness

In this observation, We used bubble plot, scatter plot, line chart and other types of classical plots to prove my conclusion on the correlation between alcohol consumption and happiness score.

Visualization process

For providing the assumption, We used a bubble plot to reveal the correlation between the alcohol and happiness as follow:
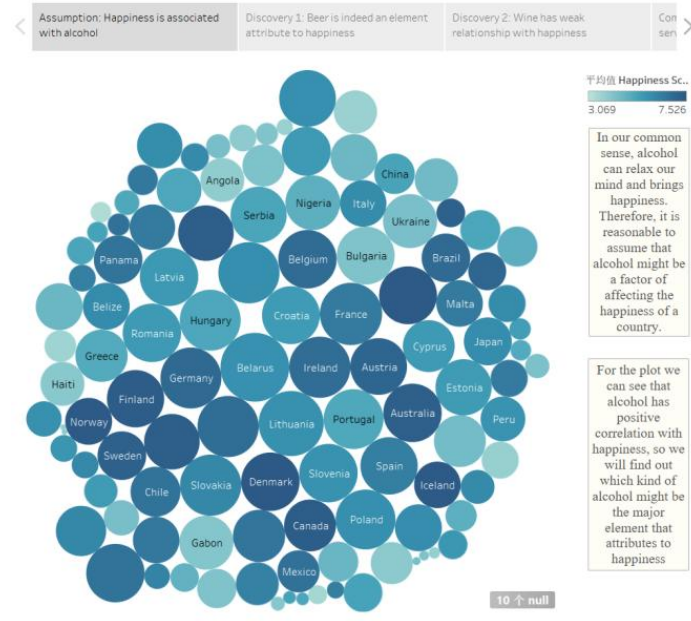
The color of this plot represents the degree of *Happiness Score* of a country while the size of the circle means the alcohol consumption situation of this country. Also, the text at the side elaborates the idea of discovering the relationship between alcohol and happiness.

During the process of revealing the correlation of these two attributes, We firstly tried to use scatter plot to show the correlation. However, We found that scatter plot has the problem in directly demonstrate correlation. Therefore, We used bubble plot as it can match my demand.

From this plot we can obviously see that **Alcohol Consumption** indeed has the positive correlation with **Happiness Score** of a country. However, we can acquire from the plot that their correlation is not strong enough. Thus, we can assume that happiness might only associated with some specific type of alcohol. Therefore, We made a further observation on it.
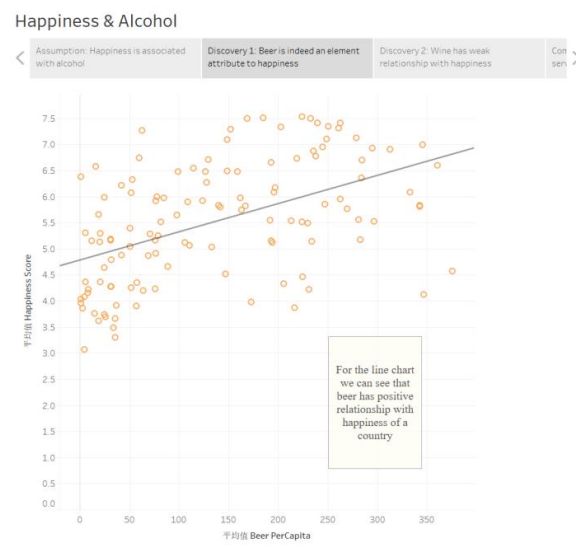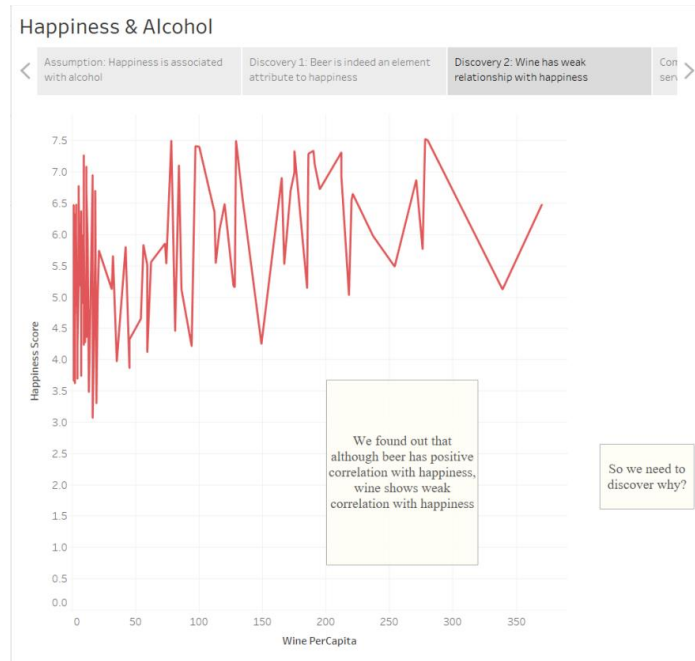
To demonstrate the correlation between **Beer Consumption** and **Happiness Score**, We used a scatter plot with a trend line to reveal it. The **X axis** represents the **Beer Consumption** while **Y axis** represents the **Happiness Score**.

With the second plot, We used a line chart to reveal the correlation between the **Wine Consumption** and **Happiness Score**. The **X axis** represents the **Wine Consumption** while the **Y axis** represents the **Happiness Score**.

At first, We used the average consumption to form the plot. However, We found that using the average value to represent the correlation might encountered the problem that it is hard to see the correlation. Thus, We changed to use the sum value.

Acquire from these two plots, we can conclude that **Happiness Score** indeed affected by the **Beer Consumption**. However, it might not be associated with the **Wine Consumption**. Thus, We will explore it in the next few plots.

In this dashboard, We combined the form of **Wine Average Servings** and **Beer Average Servings** together for a more directly demonstration. Plot on the right side represents the **Average Servings of Beer** while the left one represents the **Wines'**. The color and the number both represent the average value of corresponding alcohol servings.

At first, We separated the **Average Servings of Beer** and **Wine** into two plots. However, We found that it is not directly enough to show the comparison between these two attributes. Therefore, We used dashboard in Tableau to solve this problem.

From the dashboard we can see that the **Average Servings of Beer** is way higher than the **Average Servings in Wine**. We can make a brief conclusion that due to the **Consumption on Beer** is much higher, which might lead to more production and more jobs, that affect the **GDP** of a country. Finally, affecting the happiness of the residents.

We used two plots to prove my points. The first one shows the correlation between **GDP** and **Alcohol Consumption** while the second one shows correlation between the **Happiness Score** and **GDP**.

From the plot we can conclude that alcohol indeed affect the **GDP** of a country. Also, happiness score is associated with **GDP** of a country.

7. Analyzing Global Happiness: The Role of GDP, Education Inequality, and Colonial Background

In this part, We mainly want to explore the relationship between the global happiness index and factors such as GDP per capita, educational inequality and historical colonial background through data visualization. We will show and analyze a series of charts to reveal how these key factors affect the happiness index of various countries.

Data preparation

The data set used in this part contains the happiness index, GDP per capita, education inequality index, HIV/AIDS and life expectancy of different countries in the same period after data processing, and the data of this period is further averaged.

Visualization process

As is shown in figure bel, a scatter plot is used to show the relationship between GDP per capita and happiness index. The scatter color represents the data of different years to form an intuitive visual effect.
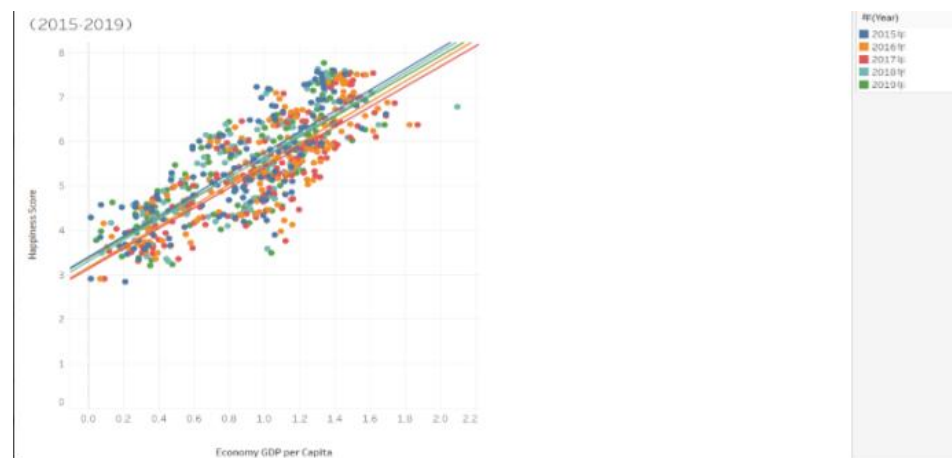
In figure below,the relationship between GDP per capita, education imbalance index and happiness index is shown through the scatter plot. The size of the bubble represents the happiness index, and the color also changes with the happiness index. The X-axis and Y-axis are GDP per capita and education imbalance index, respectively, to form a more intuitive visual effect.
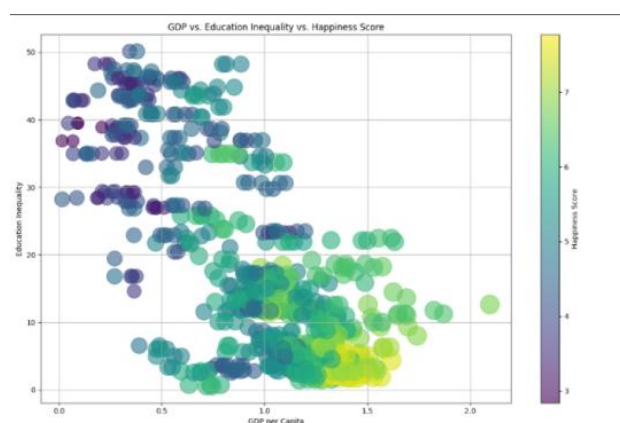

Figure below presents a 3D Scatter Plot that uses colors to highlight the dimensions of the happiness index and shows the relationship between the three in three-dimensional space. It can be found from the figure that countries with higher per capita GDP tend to have lower education inequality index and higher happiness index. This shows that economic development and education equity play an important role in improving happiness. At the same time, although some countries have lower per capita GDP, they have higher happiness indexes due to lower education inequality index, which shows that education equity also plays a key role in national happiness.

3D Scatter Plot: GDP, Education Inequality, and Happiness Score
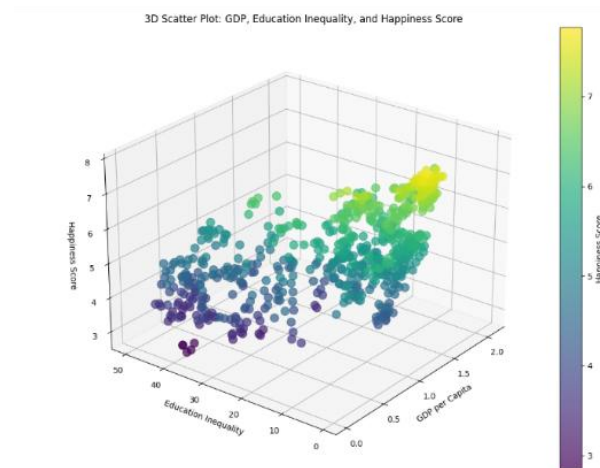
The distribution of GDP per capita in various countries with British and French colonial backgrounds is shown through density curves(figure below). In the peak comparison, the red (British colonies) has a wider peak and is distributed at higher GDP values. The blue curve (French colonies) shows a sharp peak near the lower per capita GDP value, indicating that most countries are concentrated at lower per capita GDP levels. The red curve is more scattered and extends to higher per capita GDP levels, which means that the variation of per capita GDP in this group is greater and contains more high-GDP countries. The blue curve is narrower, indicating that the variation of per capita GDP in this colonial background group is smaller.

Distribution of Economy GDP per Capita by Colonial Background
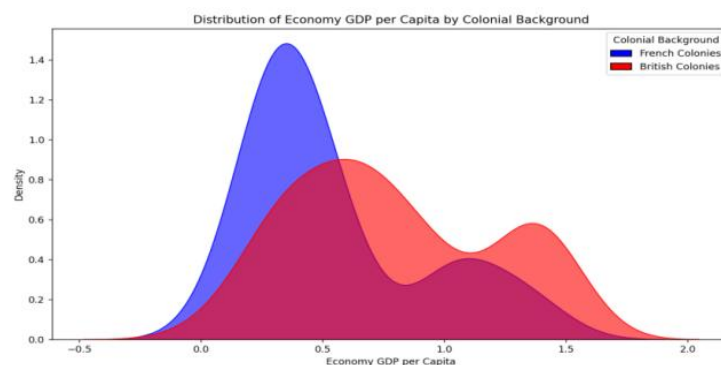
The differences in the distribution of happiness indexes in countries with different historical colonial backgrounds are compared through box plots(figure below). The median happiness index of the former British colonies (about 5) is significantly higher than that of the former French colonies (about 4). The happiness index of the British colonies is widely distributed, with some countries having a happiness index close to 7, while the happiness index of the French colonies is generally concentrated between 3 and 5, with some low outliers. It can be found that there is a certain similarity with the GDP distribution curve above.

Happiness Score Distribution by Colonial Background
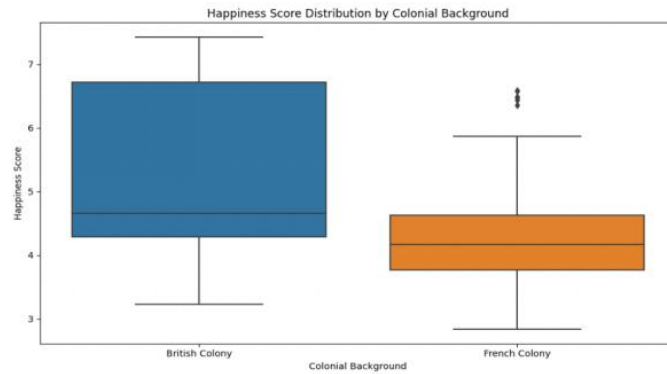
Finally, figure below presents a correlation matrix.The correlation matrix is used to analyze the correlation between the happiness index and other variables. Happiness index: It is highly positively correlated with GDP per capita (0.79), indicating that economic development is one of the key factors affecting the happiness index. It is also strongly positively correlated with life expectancy (0.75), indicating that health and quality of life play an important role in happiness. It is moderately negatively correlated with educational inequality (-0.67), indicating that educational equity has a certain impact on improving happiness. It is negatively correlated with the prevalence of HIV/AIDS (-0.44), indicating that high disease burden has a certain negative impact on the happiness index, but the degree is weaker than other. GDP per capita has a significant positive impact on happiness, and economic development is an important way to improve happiness. Educational equity is also one of the key factors in improving happiness, and reducing educational inequality helps to improve overall happiness. Due to structural differences such as economy and education, different historical colonial backgrounds also have an impact on the happiness index. At the same time, it is necessary to strengthen public health intervention to reduce diseases and improve the physical health of the people, which can also indirectly improve the level of happiness.

Correlation Matrix of Key Variables