PSTAT 100 Final Project - World Happiness Report 2023 Analysis

Meghana Dhruv (mdhruv), Kira Jackson (kirajackson),
Hsinpin Wen (hcwen), Candis Wu (cwu383)

2024-06-08

Table of contents

Abstract

This report delves into the 2023 World Happiness Report dataset, examining how various factors influence happiness and well-being across countries from 2008 to 2023. The analysis explores relationships between Gross Domestic Product (GDP) per capita and several key factors: life ladder, social support, healthy life expectancy, freedom of choice, generosity, perceptions of corruption, and positive and negative affects. The primary focus is to investigate how economic wealth, as measured by GDP per capita, correlates with these happiness and well-being indicators over time. Through visualizations and statistical tests, the report identifies significant patterns and trends, providing a deeper understanding of the socio-economic factors that contribute to the quality of life in different nations. Additionally, the report includes a comparative study of life expectancy trends between China and the United States, bringing light to the complex interplay between economic conditions and health outcomes.

Question of Interest

In this report, we will seek to answer the question: How do socio-economic factors relate to the happiness and wellbeing indices reported in the World Happiness Report across different countries over time?

Sub-Questions

To explore the relationship between socio-economic factors and happiness across different countries over time, our group decided to dive into more specific questions in our data analysis before answering our main question of interest:

1. How does GDP affect happiness factors across countries?
2. How does life expectancy compare between China and the United States?
3. Is there a significant difference between positive/negative affect across South American countries over time?

Variable Descriptions

- `country_name`: Name of the country the data was collected in

- `year`: The year the data was collected in ranging from 2008-2023

- `life_ladder`: Aggregate measure of how people within a country perceive and evaluate their own well-being and quality of life from 0 to 10

- `gdp_per_capita`: A measure of economic prosperity and reflects the average income level of individuals in a country adjusted for population size

- `social_support`: The extent to which individuals feel they have someone to rely on in times of need, whether it be family, friends, or other forms of social support networks ranging from 0 to 1

- `healthy_life_expectancy`: The average number of years an individual can expect to live in good health at birth which indicates the overall health and well-being of the country's population

- `freedom_of_choice`: A measure of individual autonomy and personal freedoms in making life choices

- `genorosity`: The portion of the national average charitable donations that cannot be explained by GDP per capita

- `perceptions_of_corruption`: The extent of how much individuals feel their government and businesses are corrupted ranging from 0 to 1

- `positive_affect`: The average level of positive emotions in a country, based on responses to three questions in the Gallup World Poll: whether respondents smiled or laughed a lot yesterday, experienced enjoyment for much of the day yesterday, and learned or did something interesting yesterday.

- `negative_affect`: The negative affect variable represents the average level of negative emotions in a country, based on responses to three questions in the Gallup World Poll: whether respondents experienced worry, sadness, and anger for much of the day yesterday.

Country GDP vs Happiness Factors

We begin by tidying the 2023 World Happiness Report dataset with the `clean_names()` function in the Janitor library. This function makes the column names more predictable and easier to use in subsequent analysis by converting column names to lowercase, replacing spaces and special characters with underscores, and ensuring that the names are syntactically valid. We also decided to rename certain columns to a more succinct name that allows the variable to be displayed clearer in our plots.

A country's Gross Domestic Product, or GDP, gives information about the size of its economy and how an economy is performing. It translates to a measure of a country's overall wealth, where a higher number indicates more money, and vice versa. Naturally, the economy of a

country plays a large role in many aspects of how the country is doing. We will begin by exploring visualizations to examine the relationship between a country's GDP and a few of the factors contributing to the happiness and well-being of its people presented in the data set, namely life ladder, social support, healthy life expectancy, generosity, and perceptions of corruptions. The relationship between GDP and all factors can be further explored in the Shiny app.

To begin our exploration of the relationship between GDP and happiness factors by country, we first have to aggregate the data by country.

GDP vs Life Ladder

GDP per Capita vs Life Ladder
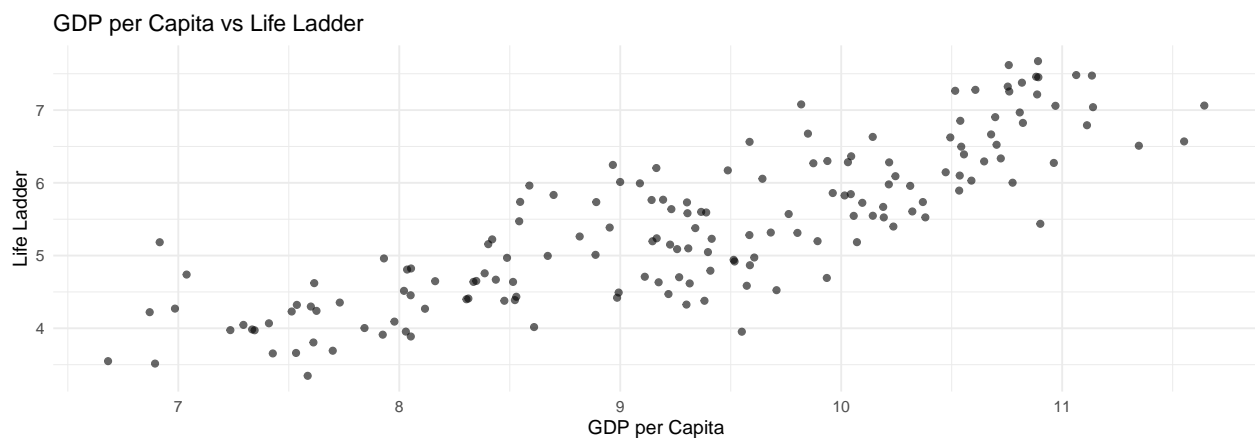
From the graph above we can see that as gdp_per_capita increases, so does life_ladder. There seems to be a clear positive correlation between the two. Since life ladder represents how people feel about their quality of life, it makes sense that this number would increase as average economic prosperity does as well. In other words, more money equals more happiness and better overall well-being.

## GDP vs Social Support

GDP per Capita vs Social Support
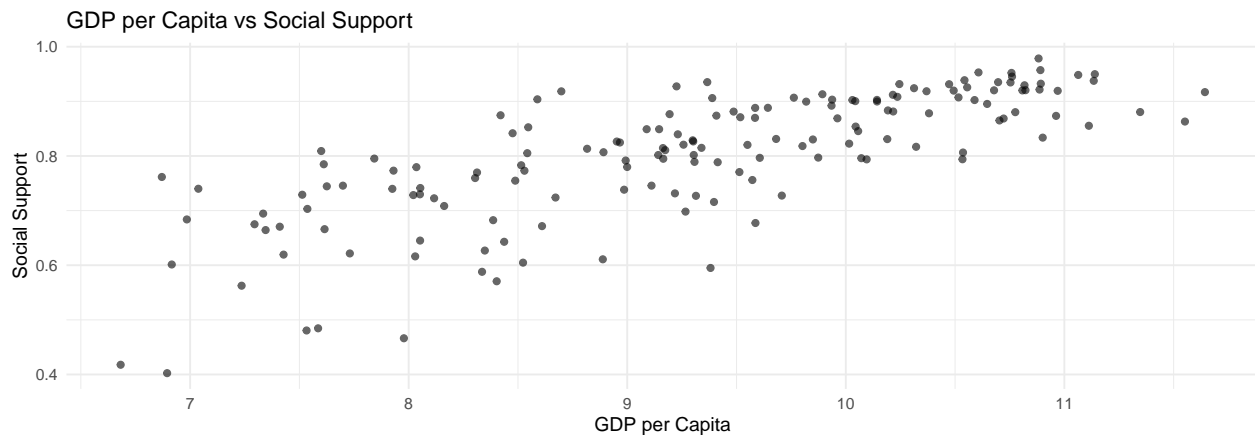
Similarly, as gdp_per_capita increases, so does social_support. Being that social support is the extent that individuals have people they feel they can rely on, it also makes sense that this would have a positive correlation with economic prosperity. This correlation is not nearly as strong as the correlation from the previous graph with life ladder, but it is still apparent. There is also a larger increase at the beginning and then the amount of increase eventually tapers out.

GDP vs Healthy Life Expectancy

GDP per Capita vs Healthy Life Expectancy
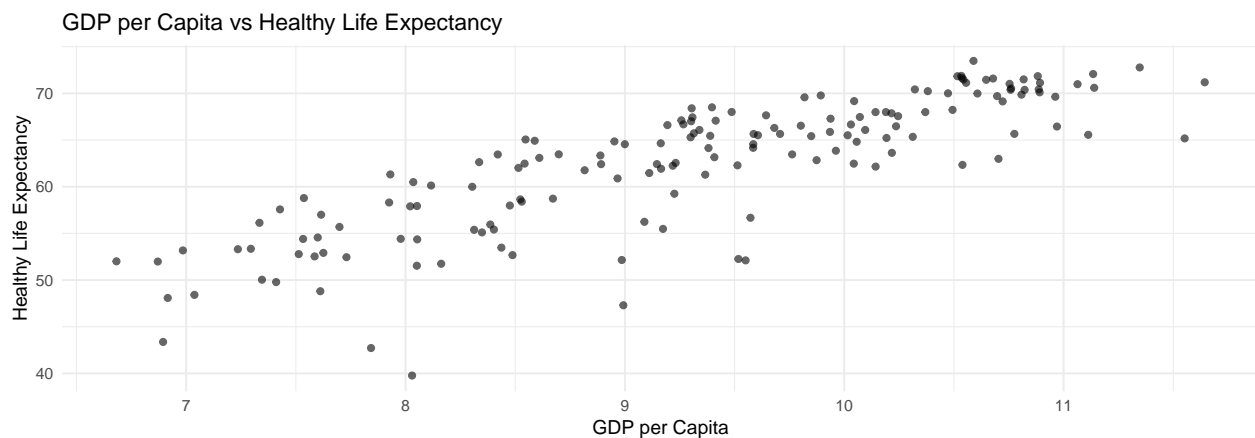
Similarly, as gdp_per_capita increases, so does healthy_life_expectancy. As seen by the graph above, the two variables have a clear positive correlation. This also makes sense because higher economic prosperity could translate to better living conditions, thus meaning better life expectancy.

GDP vs Generosity
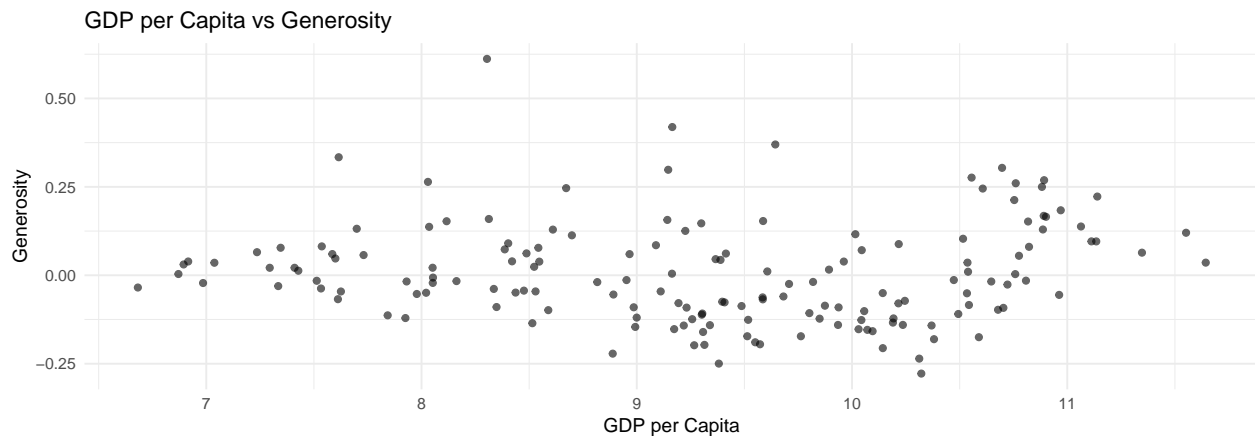
As gdp_per_capita increases, it seems that generosity slightly decreases. Though not completely apparent, there does seem to be a slight negative correlation between the two. However, for the most part, there seems to be no relationship at all. This makes sense because generosity, the portion of the national average charitable donations, is not necessarily dependent on only GDP per capita.

GDP vs Perceptions of Corruption
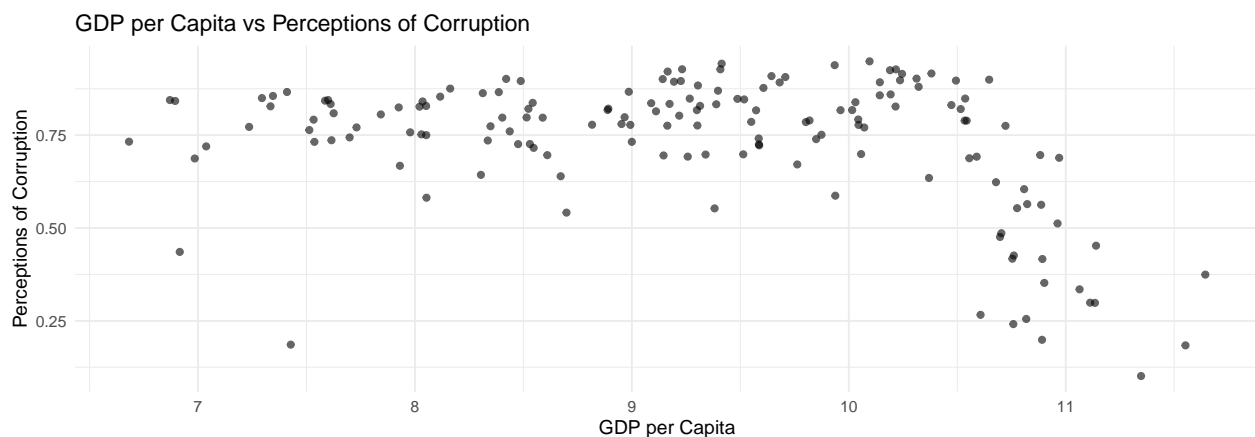
This scatterplot is noticably different compared to the ones shown above. We notice that as gdp_per_capita increases, perceptions of corruption eventually decreases. At first, there seems to be a stagnant trend in gdp_per_capita, but as it reaches 10, there is a steep decline and perception of corruption noticeably falls. This makes sense because increased GDP per capita is often associated with higher standards of living. As citizens' basic needs are better met, they may have greater capacity to demand accountability and transparency from their governments. With this the general public will be more satisfied with their government as there is more communication between the two parties causing the perception

of corruption to decrease.

Overall, some notable trends between the happiness indicators and GDP are as gdp_per_capita increases, so do these factors: healthy life expectancy, social support, and life ladder. On the other hands, the happiness factors that decrease include perceptions of corruption and generosity.

Life Expectancy

Life expectancy is a crucial indicator of a country's overall health and well-being. It reflects the average number of years a person can expect to live, given current mortality rates. This metric is important because it provides a comprehensive snapshot of the living conditions, health care quality, lifestyle choices, and socio-economic environment within a country. An increase in life expectancy generally signifies improvements in various aspects of life, including better living conditions, enhanced nutrition, improved education, and greater access to health services that can collectively contribute to a healthier and longer life for the population. Countries with a higher Gross Domestic Product (GDP) often tend to have a longer life expectancy. Wealthier nations can afford more robust public health initiatives and have the resources to combat diseases more effectively, leading to a healthier population and extended life expectancy.

In the plot below, we will start by examining the average life expectancy of the world over the years. To begin we manipulate the data to get the average life expectancy per year. This analysis will provide insights into global health trends and highlight the progress made in improving the quality of life across different regions.

Average World Life Expectancy Over the Years
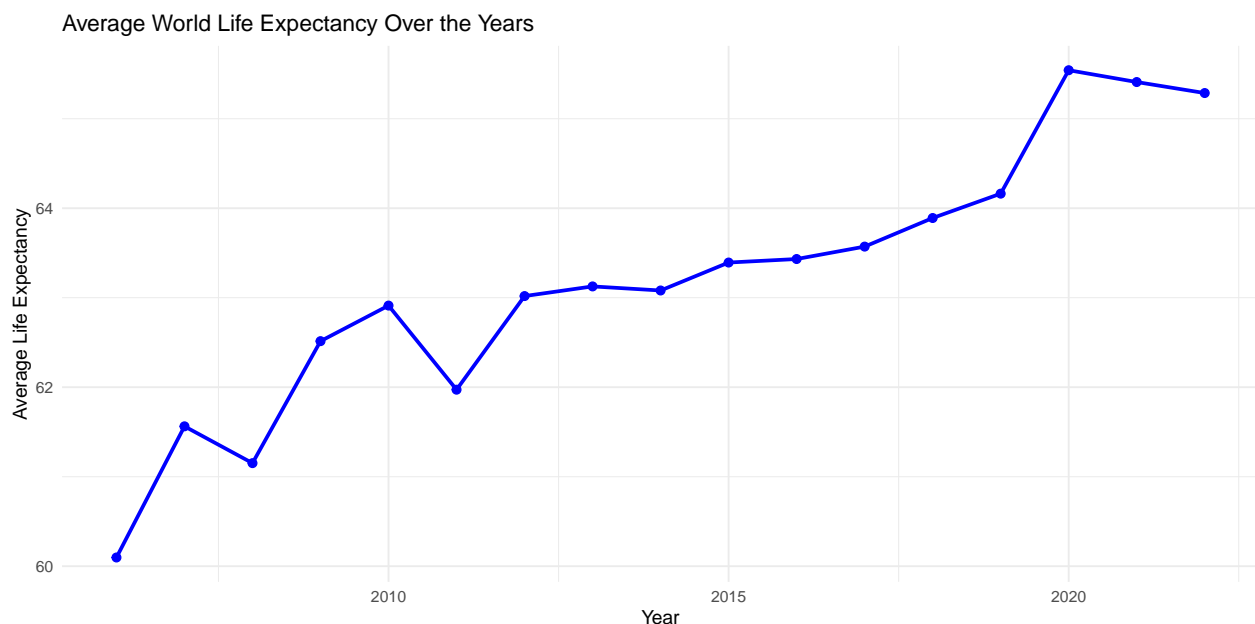
The average life expectancy in the world from 2006 to 2020 a somewhat steady increase. Fluctuations see in the graph can be attributed to various global events and health crises. Steep declines throughout in 2007 and 2010 may reflect the impact of regional conflicts, economic downturns, or disease outbreaks that disproportionately affected certain populations. However, the moderate steady increase in life expectancy up to 2020 suggests overall improvements in global health, healthcare access, and socio-economic conditions. We are all well aware that in 2020, the COVID-19 pandemic had a significant global impact, leading to increased mortality rates and a subsequent decrease in life expectancy, which is why we see a drop in the average life expectancy as this catastrophic event was something that impacted everyone in the world no matter where you lived. The pandemic strained healthcare systems worldwide and highlighted disparities in health infrastructure and access to medical care, underscoring the vulnerability of global health to widespread infectious diseases.
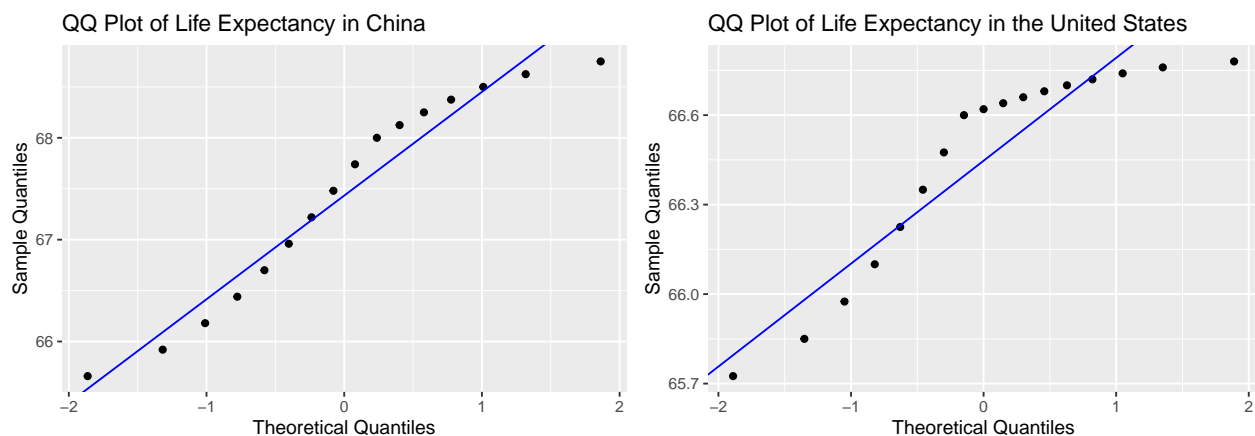
China and United States Life Expectancy Comparison

Next, we will narrow down our analysis to two economic global powerhouses: China and the United States. According to the US Department of State, "China is currently the third-largest export market for U.S. goods (after Canada and Mexico), and the United States is China's largest export market." In examining global health metrics, comparing the life expectancy between the United States and China offers valuable insights into the health outcomes of two of the world's largest economies. Life expectancy, a key indicator of a population's overall health and well-being, reflects the average number of years a person can expect to live based on current mortality rates. By analyzing this metric, we can uncover the impact of various health behaviors, socio-economic conditions, and healthcare systems on the longevity of individuals in these countries. This comparison aims to highlight significant differences and underlying factors contributing to the health disparities between the United States and China.
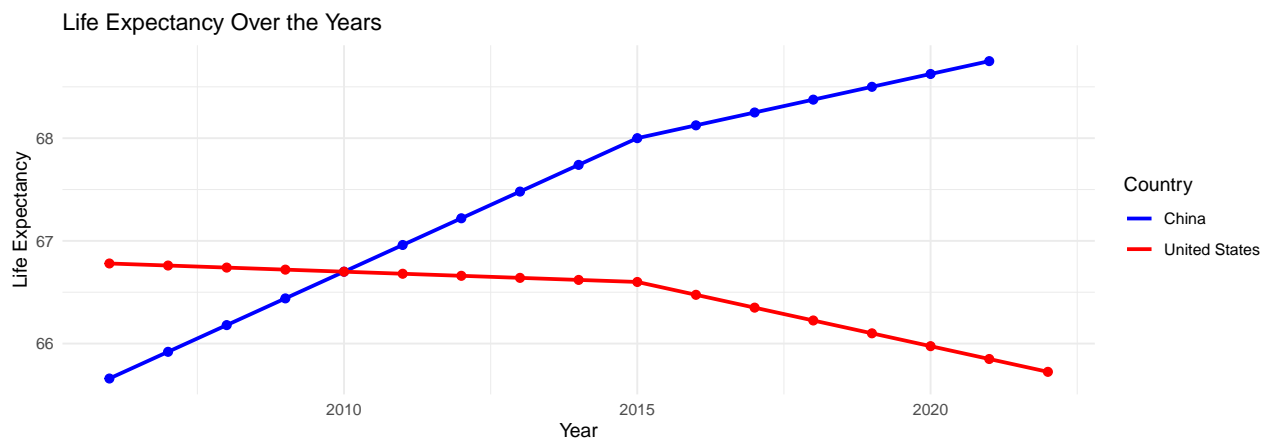
In comparison to the United States, China experienced a steady incline in life expectancy from around 2005 to 2020. The United States, on the other hand, has been experiencing a decline in life expectancy with the rate of decline increasing in 2015.

Interestingly enough, we can see that while China's life expectancy has steadily increased over the years, the United States' life expectancy has been decreasing. Before we dive into this difference, we need to answer the following question: Is there a significant difference between the average life expectancy in the United States and China?

To answer this question, we will perform a t-test. A t-test is a statistical test used to compare the means of two groups and determine if they are significantly different from each other. This test assumes that the data from both groups are normally distributed and have similar variances. If these assumptions are met, the t-test can reliably indicate whether the observed differences between group means are due to random chance or if they reflect true underlying differences. We will begin by plotting QQ Plots of the life expectancies of China and the US to check for normality.

We can see from the QQ plots above that the life expectancy data for both the United States and China are not quite normally distributed. However, according to the World Happiness Report 2023, our sample size is larger than 30, thus it is generally considered large enough to rely on the central limit theorem, which suggests that the sampling distribution of the mean approaches normality as the sample size increases. This indicates we can still perform a t-test despite the non-normal distribution.

However, it is important to note that using t-tests and ANOVA tests with a large sample size can give meaningful results even if the data is not perfectly normal, but it is not immune to errors, so we will be interpreting the results with caution.

```
    Welch Two Sample t-test

data:  china_data$healthy_life_expectancy and us_data$healthy_life_expectancy
t = 3.68, df = 18.219, p-value = 0.001685
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.4234717 1.5480356
sample estimates:
```

```
mean of x mean of y
 67.43281  66.44706
```

Upon conducting the above t-test, the resulting p-value was found to be 0.001685, which indicates a statistically significant difference in healthy life expectancy between the United States and China, as it is less than the commonly used alpha value of 0.05. This low p-value suggests that the observed difference in life expectancies is unlikely to have occurred by random chance. In practical terms, this means there is strong evidence to conclude that the average life expectancy in China and the United States are not the same, with China's life expectancy showing a steady increase over the years while the United States has experienced a decline. This significant difference could be attributed to various factors, including healthcare systems, lifestyle choices, socioeconomic conditions, and public health policies in each country. The results prompt further investigation into the underlying causes and potential policy implications to address the disparities in life expectancy between the two nations.

From the earlier plot of GDP versus life expectancy, there is a positive linear trend between a country's GDP and healthy life expectancy. However, we are not seeing that in the case of the United States. Taking a look at the plot of China and the United States' GDP below, we can see that the US has had a consistently high GDP over the years. So, why has the United States' life expectancy been decreasing despite its high GDP?
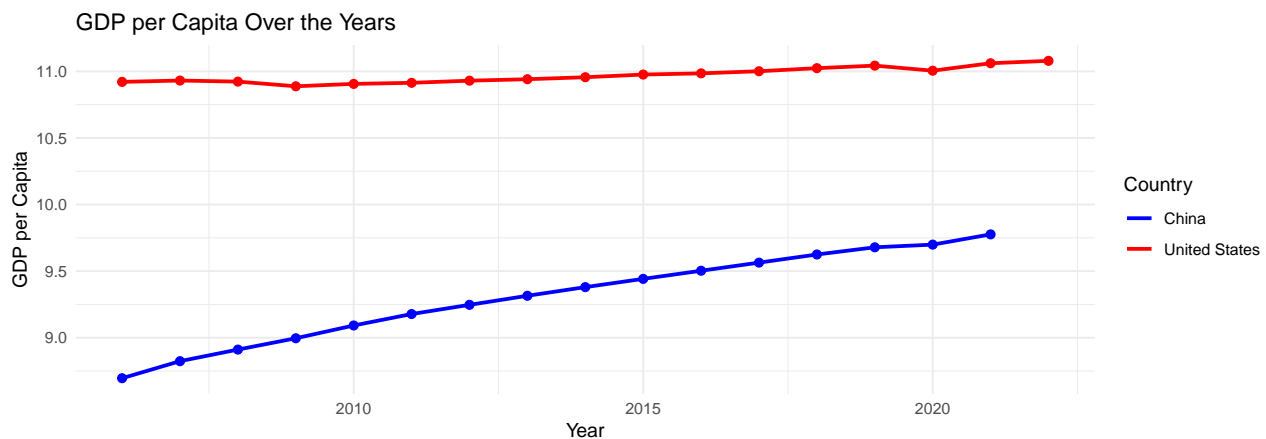
In terms of GDP per Capita over the years, China has started much lower than the United States but has experienced a steady incline till 2020. Though the GDP per capita in the United States is much higher, it has stayed relatively stagnant with an extremely slight rate of increase till 2020.

Upon further exploration, we discovered that Americans have a lower life expectancy than other rich countries in the world. Below is a table showing the top 10 countries with the highest GDP's and their corresponding life expectancies:

Country Name	Average GDP per Capita	Average Healthy Life Expectancy
Luxembourg	11.64	71.19
Qatar	11.55	65.18
Singapore	11.35	72.77
Ireland	11.14	70.6
Switzerland	11.13	72.07
United Arab Emirates	11.11	65.56
Norway	11.06	70.98
United States	10.97	66.45
Kuwait	10.96	69.64
Netherlands	10.89	71.13

We can see from this table that apart from the Middle Eastern countries (Qatar, UAE, and Kuwait), the United States has the lowest average healthy life expectancy. We've excluded Middle Eastern countries from this observation because of the difference in sociopolitical, demographic, and economic environments compared to traditionally "Western" countries, since these countries' economies are driven by oil and other natural resources and an expansive migrant worker population. More information here.

Further research on the general health of Americans reveals that the United States has a higher rate of smoking, obesity, homicides, opioid overdoses, suicides, road accidents, poverty, economic inequality, and lack of healthcare access compared to other rich countries. These factors contribute significantly to the nation's overall health outcomes and life expectancy. Importantly, many of these causes disproportionately affect young people which has a notable impact on the average life expectancy.

In conclusion, these findings reveal the relationship between health behaviors, societal challenges, and life expectancy in the United States. While the nation boasts considerable advancements in healthcare and medical technology, disparities in health outcomes persist due to many of the country's prevalent social, economic, and environmental issues. While these disparities present significant challenges, they also underscore opportunities for targeted interventions and policy initiatives aimed at improving public health and enhancing overall well-being in the United States. By addressing underlying social determinants of health there is potential to mitigate the factors contributing to lower life expectancy and create a healthier, more resilient society. Ultimately, by fostering collaboration across sectors and implementing evidence-based strategies, we can work towards a future where all individuals have the opportunity to lead long, fulfilling lives, regardless of their socio-economic circumstances.

Comparing Positive and Negative Affect Levels Across South America

Understanding the emotional well-being of populations is crucial for gauging the overall mental health and happiness of societies. Positive affect, which encompasses feelings such as joy, contentment, and enthusiasm, can provide valuable insights into the quality of life and social dynamics within a nation. Conversely, negative affect, including feelings of sadness, anger, and stress, can highlight areas needing improvement in public health and social policies.

## Difference Between Positive Affect Levels in South America Over Time

In this section, we aim to determine if there is a statistically significant difference in positive affect ratings in South America over time. Analyzing these trends can reveal how economic, social, and political changes impact the emotional well-being of a population. From our data set we will filter to only select data from the countries in South America. This includes the countries: "Argentina", "Bolivia", "Brazil", "Chile", "Colombia", "Ecuador", "Guyana", "Paraguay", "Peru", "Suriname", "Uruguay", and "Venezuela".

An ANOVA (Analysis of Variance) test is a statistical method used to compare the means of three or more groups to determine if there are any statistically significant differences among them. In our context, the ANOVA test will help us assess whether the mean positive affect scores differ significantly across different years within South America. Before proceeding with the ANOVA test, it is essential to ensure that the sample sizes are sufficiently large.

```
Number of observations:  171
```

As we can see there are 171 observations. Therefore since the sample size is large we can utilize the Anova test.

First we will focus on South America's average positive affect ratings over the years by producing a scatterplot to visualize the trend.

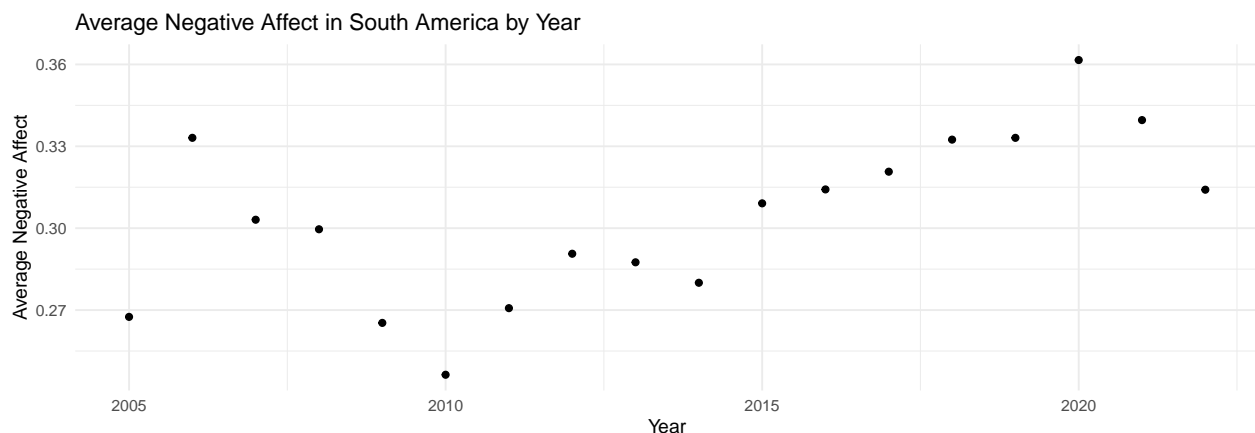Average Positive Affect in South America by Year
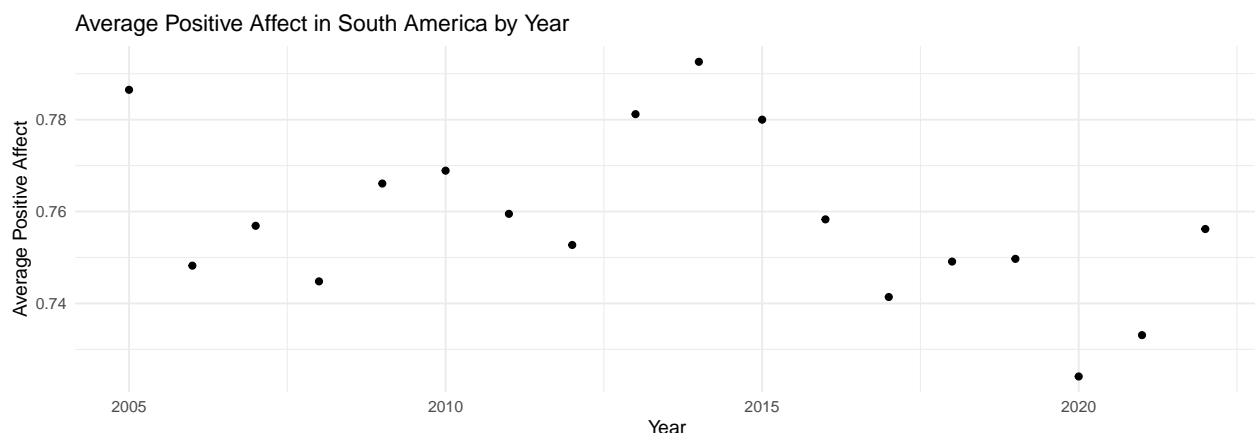
Right off the bat, there seems to be fluctuations in average positive ratings over time. This might occur due to certain events that happen during a year which can cause positive emotions and wellbeing to change. Let's proceed with our ANOVA test.

```
          Df Sum Sq  Mean Sq F value Pr(>F)
year       1 0.0057 0.005730   2.674  0.104
Residuals 169 0.3622 0.002143
```

However, despite these observed fluctuations, the ANOVA test results indicate that the overall trend in positive affect ratings across different years is not statistically significant as the p-value from the ANOVA test is 0.104, which is greater than the chosen significant level of 0.05. This lack of statistical significance suggests that the observed variations in positive affect scores could be due to random chance rather than meaningful differences over time. Therefore, while there may be fluctuations in positive affect ratings over time within South America, these fluctuations do not appear to follow a consistent pattern or demonstrate a statistically significant trend.

Difference Between Negative Affect Levels Within South America Over Time

In this section, our objective is to explore whether there exists a statistically significant difference in negative affect levels in South America to analyze how positive and negative affects differs over time. To conduct a comprehensive analysis, we will leverage another ANOVA test. From above we have already shown that our sample size is large, thus we will begin by producing a scatterplot to visualize the trend of mean negative affect scores in South America over time.

Similar to our scatterplot displaying the average positive affect levels, the average negative affect levels also seem to fluctuate depending on the year. The average negative affect levels in South America tend to be lower than the average positive affect levels which could mean that the overall emotional well-being of the population, on average, leans more towards positive emotions than negative ones.

```
          Df Sum Sq Mean Sq F value   Pr(>F)
year        1 0.0512 0.05117    13.3 0.000353 ***
Residuals 169 0.6502 0.00385
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Based on the ANOVA test above we can conclude that unlike positive affect levels, negative affect levels in South America have a statistically significant difference over time. The p-value is 0.000353, which is notably lower than the chose significance level of 0.05. The statistically significant difference observed in negative affect levels across different years in South America suggests that there are underlying factors influencing the population's emotional experiences over time. One possible explanation for this finding could be the impact of socio-economic and environmental changes on individuals' psychological well-being. During periods of economic downturns or political unrest, individuals may experience heightened levels of stress, anxiety, or frustration, leading to an increase in negative affect ratings.

The lack of a statistically significant difference in positive affect levels across different years in South America, contrasted with the significant difference observed in negative affect levels, could stem from several factors. One possible explanation is the resilience of positive emotions to external fluctuations compared to negative emotions. Positive affect levels may be influenced more by individual traits, coping mechanisms, and daily experiences rather than broader socio-economic trends. In contrast, negative affect levels may be more sensitive to economic disparities and societal challenges, leading to more noticeable variations across different periods.

Conclusion

The analysis reveals several noteworthy findings. Generally, indicators such as healthy life expectancy, social support, and life ladder exhibit positive correlations with GDP per capita, while perceptions of corruption and generosity tend to decrease as GDP per capita increases. Global life expectancy trends show fluctuations over the years, with a couple decline around in the early 2000s followed by a steady increase until 2020, interrupted by the COVID-19 pandemic. China demonstrates a consistent increase in life expectancy, contrasting with a declining trend in the United States, leading to a significant disparity between the two countries. Health disparities in the United States, including higher rates of smoking, obesity, homicides, and lack of healthcare access, contribute to its lower average healthy life expectancy compared to other wealthy nations. Despite these challenges, there are opportunities for targeted interventions and policy initiatives to address health disparities and improve overall well-being. Additionally, ANOVA test results indicate significant differences in negative affect levels in South America over time, suggesting varying cultural, economic, and political influences on emotional well-being.