PSTAT 100 Final Project: Exploring World Happiness Data

PSTAT 100: Spring 2024 (Instructor: Ethan P. Marzban)

Dulce Rodriguez, Meghna Chandrasekar, Jeffrey Fernandez, Yahir Moreno

2024-06-13

Abstract

This report aims to analyze World Happiness Report 2023 data to determine if the mean "happiness" of a country is signficantly different and changes over time differently in countries with a low versus high log GDP per capita. Linear regression models and exploratory data analysis techniques were applied to answer these questions. The analysis found that BLANK. The results suggests that countries with a high GDP have overall higher Life ladder scores and positive affect scores than countries with a low GDP, but the trends over time show unexpected peaks where this gap is bridged. Overall, analyzing world happiness report data can help identify important characteristics of a country that keep citizens happy.

Introduction

This project works with the World Happiness Report 2023 data set, which includes indices related to happiness and wellbeing by country, from 2008 to the present. The survey ranks 156 countries by how happy their citizens perceive themselves to be, and uses data primarily from the Gallup World Poll.

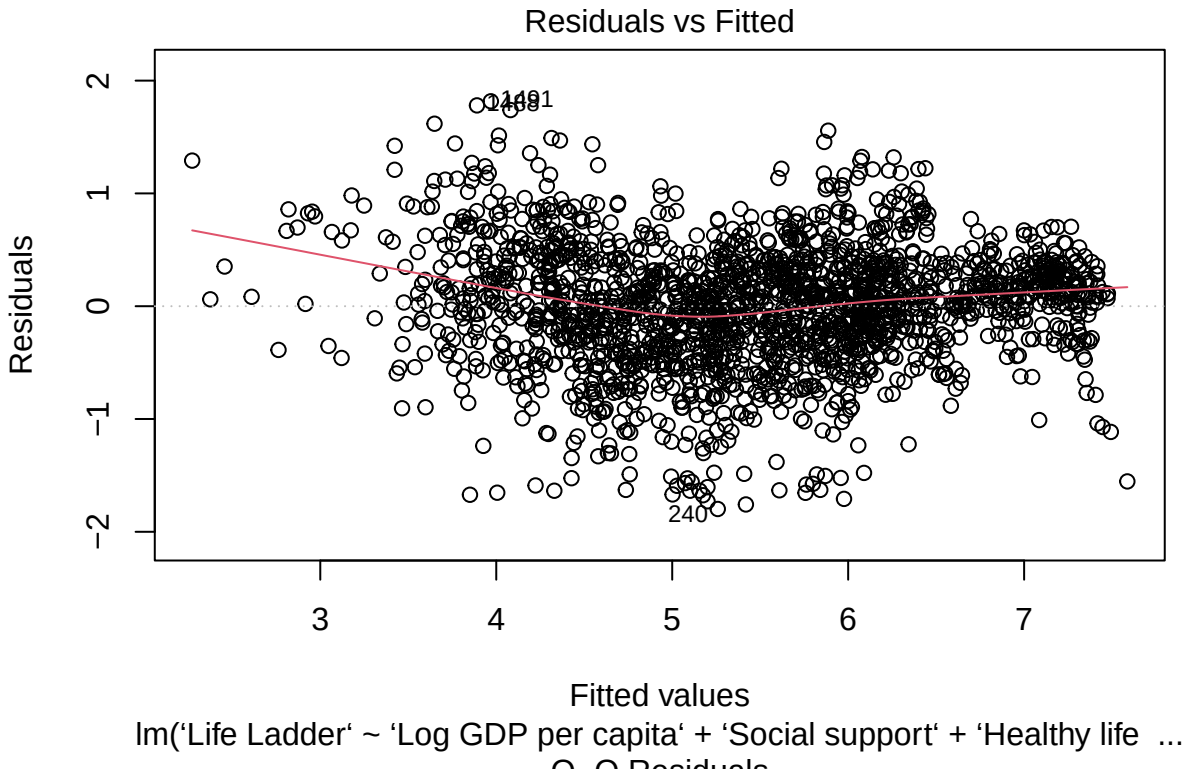
So, what makes a country's citizens "happy"? The "Positive affect" column has values ranging between 0 and 1, and represents the frequency of positive emotions such as laughter and enjoyment from respondents in a country. The "Life Ladder" metric measures overall life satisfaction. It is based on the Cantril Ladder, a scale where respondents are asked to imagine a ladder with steps numbered from 0 to 10, with the top (10) representing the best possible life for them, while the bottom (0) represents the worst possible life. Respondents evaluate their current life and indicate which step of the ladder they feel they currently stand.

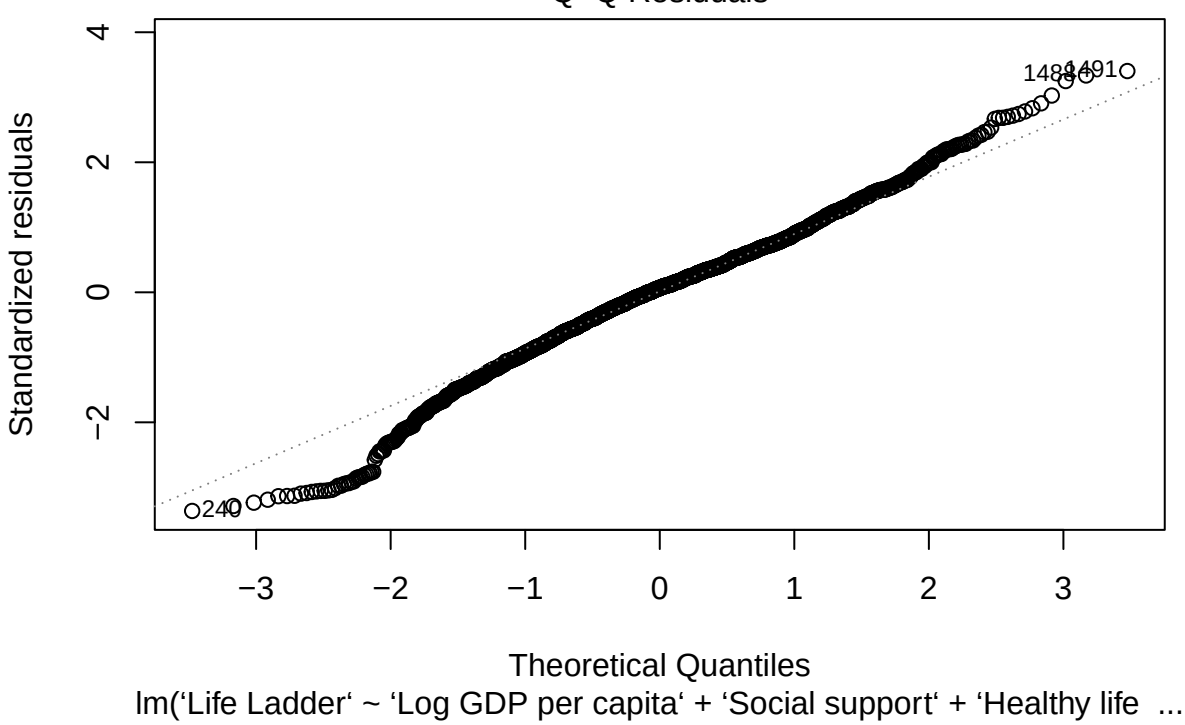
The questions this project aims to answer are: Does the mean "happiness" of a country change over time differently in countries with a log GDP per capita less than 7 versus higher than 10.5? Positive affect scores and Life ladder scores will be analyzed in comparison to log GDP. These questions about human fulfillment are answered through exploratory analysis of the data and linear models.

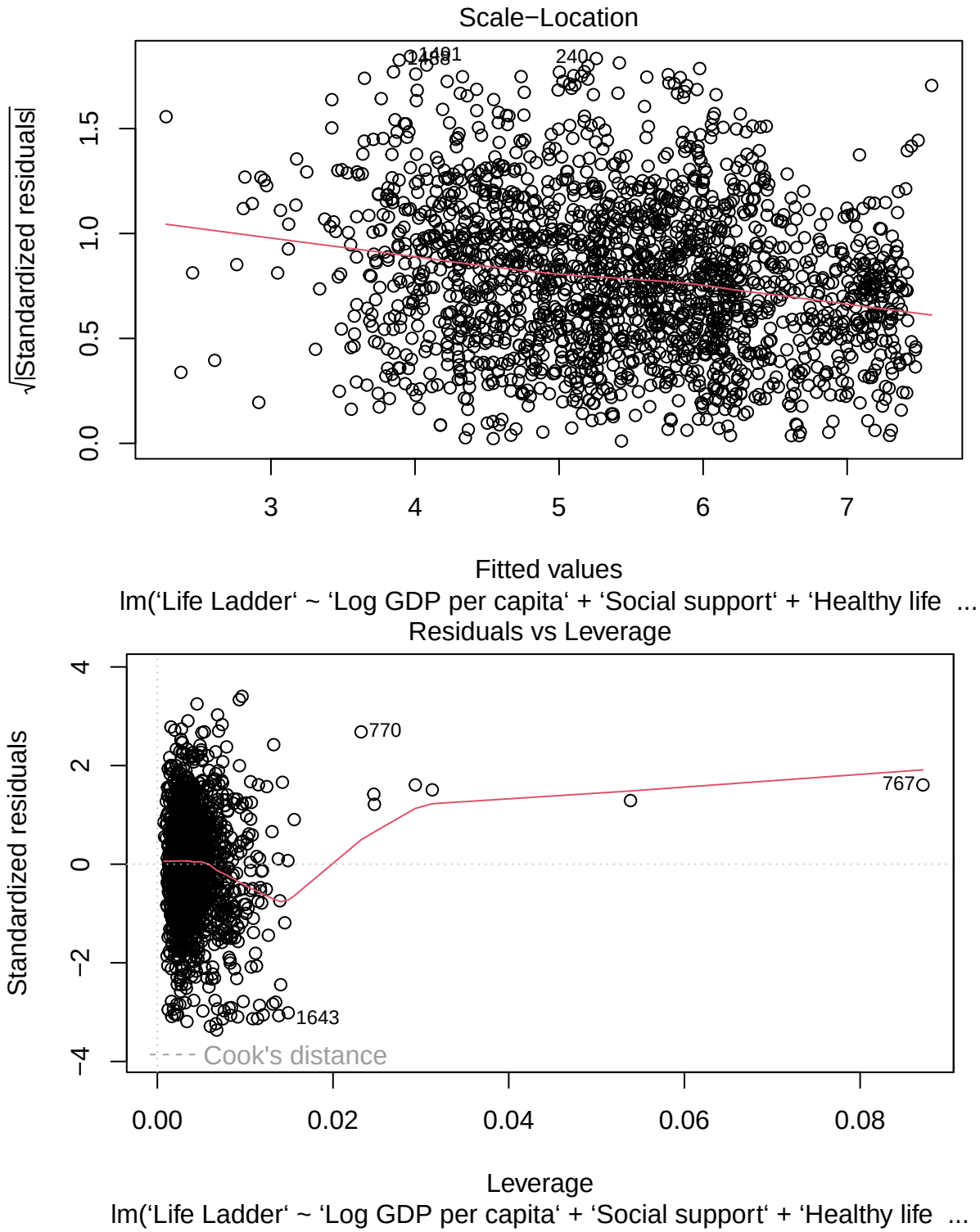
Exploration and Analysis of the Data

First, in order to explore what factors affect the Life Ladder metric the most, the following linear model was created: with the response variable being the Life Ladder metric, and predictors being Log GDP per capita, social support, healthy life expectancy at birth, freedom to make life choices, generosity, perceptions of corruption, and positive affect as predictors.

$$Y(\text{Life Ladder}) = \beta_0 + \beta_1(\text{logGDP}) + \beta_2(\text{social supp.}) + \beta_3(\text{healthy life expect.}) + \beta_4(\text{freedom})$$
$$+\beta_5(\text{generosity}) + \beta_6(\text{perceptions of corruption}) + \beta_7(\text{positive affect})$$


Residuals vs Fitted
Residuals
Fitted values
lm(‘Life Ladder‘ ~ ‘Log GDP per capita‘ + ‘Social support‘ + ‘Healthy life ...

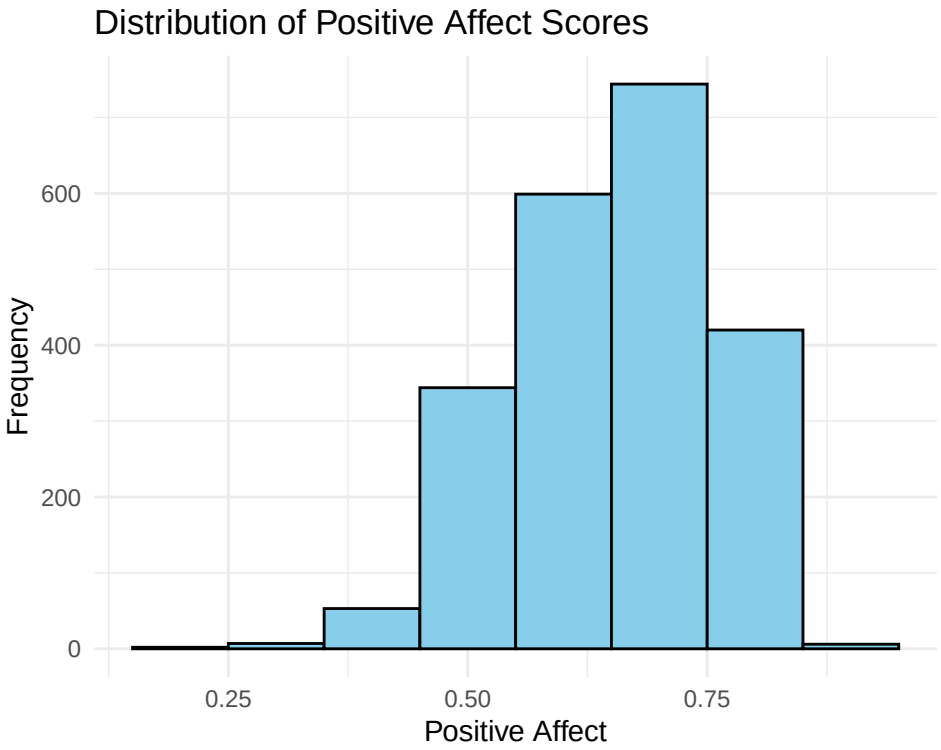

Q−Q Residuals
Standardized residuals
Theoretical Quantiles
lm(‘Life Ladder‘ ~ ‘Log GDP per capita‘ + ‘Social support‘ + ‘Healthy life ...



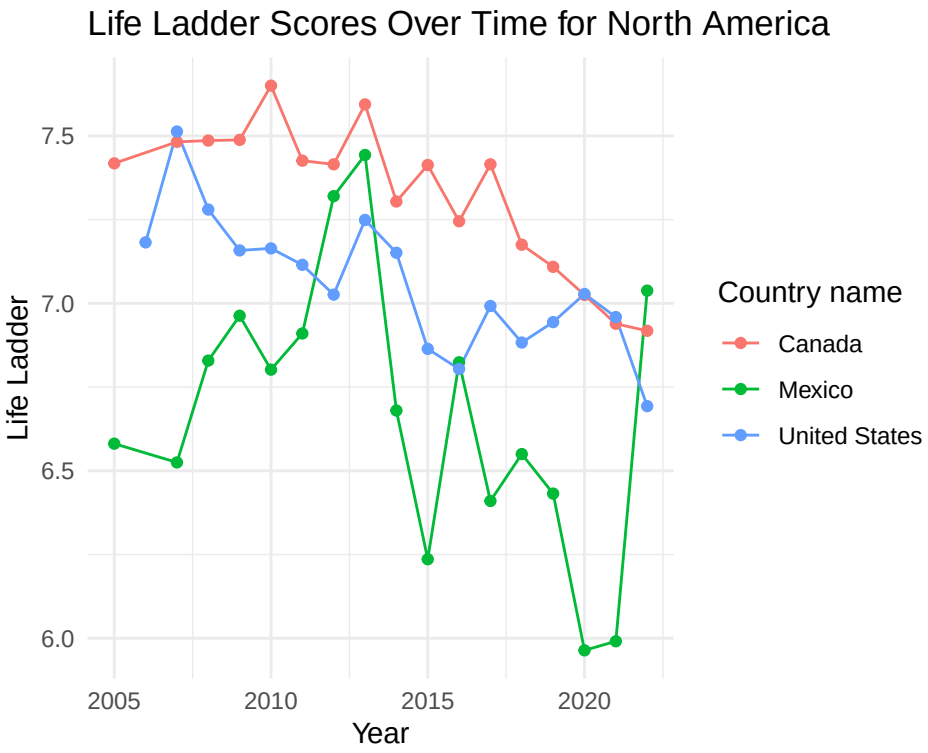
All predictors in this model are significant; interestingly “Freedom to make life choices” and “Generosity” are the least significant. The R^2 value is 0.5364 and adjusted R^2 is 0.779, on 1951 degrees of freedom.

Now, looking at the distribution of the life ladder scores, the distribution assumes a bell shape with a slight skew to the left. The median response for the life ladder score is 5.432, with standard deviation of 1.125527.

The distribution of the positive affect scores is below, which assumes a curve with a stronger skew to the left. The median score is 0.6630, with standard deviation of 0.1059126.

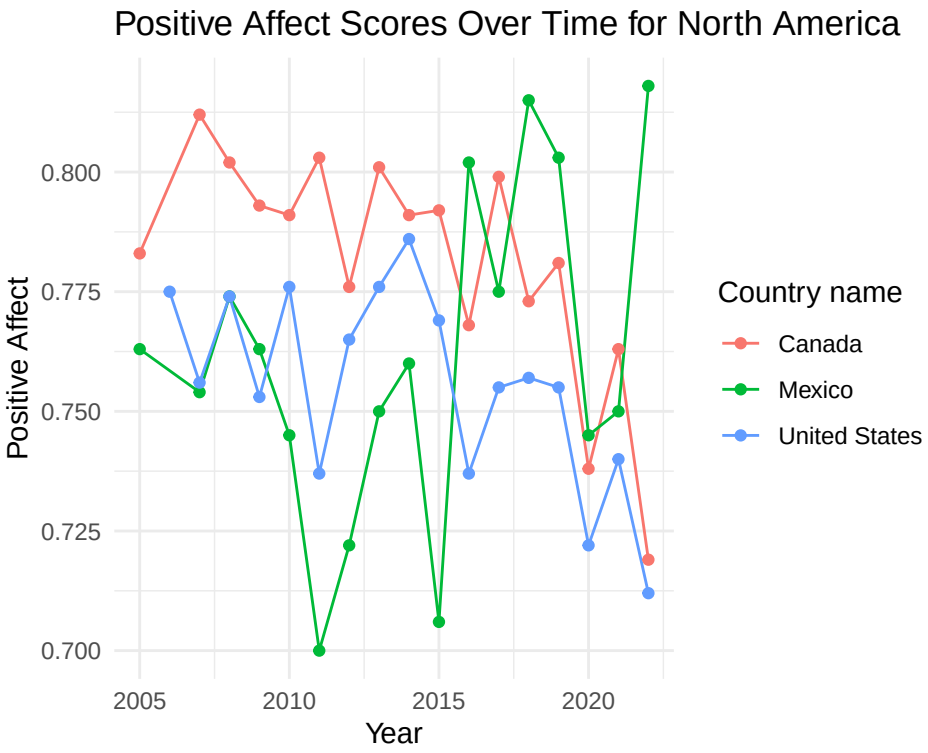


In order to see how life ladder scores change over time, we take a look at the life ladder scores over the years from 2005 to 2028. It appears as if Canada and United States have a similar trend in these scores, which is actually decreasing from 2005 to 2008, while Mexico has more variability, with a huge dip in 2020.



The United States has a mean log GDP of 10.96965, Canada has an average logGDP of 10.75376, and Mexico has a mean log GDP of 9.849412 per capita.

Similarily, we will see the trend in positive affect over time.



Interestingly, the positive affect scores over the years from 2005 to 2028 between North American countries show a different trend. It again appears as if Canada and United States have a similar decreasing trend,

while Mexico again has more variability; however this time there is a large *peak* near 2020.

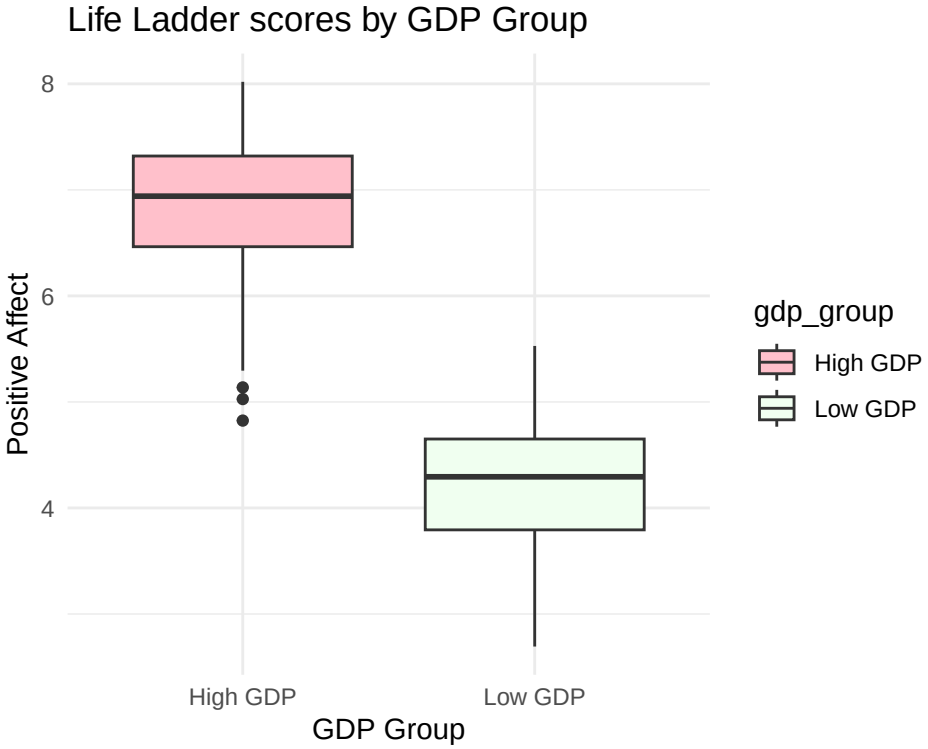
These results bring up the question: Do countries with starkly different log GDPs per capita show different trends in metric of "happiness": Life Ladder and positive affect? Additionally, does the relationship between log GDP per capita of a country and its Positive affect differ than between GDP and Life Ladder scores?

In order to show a strong dichotomy, we decided to split countries with a log GDP of less than 7 compared to countries with a logGDP of greater than 10.5 and compare their Life ladder and positive affect scores.

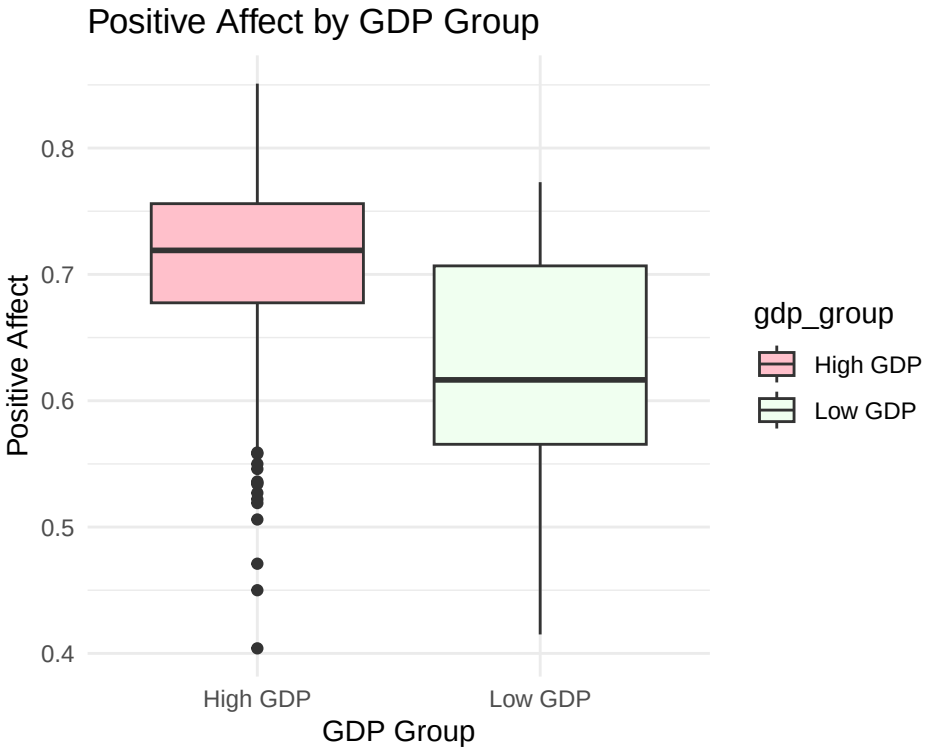
Comparing high and low GDP countries for Life Ladder and Positive affect scores.

The mean life ladder metric for countries with a log GDP of less than 7 was found to be 4.273706 with a standard deviation of 0.6414622. On the other hand, the mean positive affect for countries with a log GDP of more than 10.5 was found to be 6.858506 with a standard deviation of 0.5772171.

The mean positive affect for countries with a log GDP of less than 7 was found to be 0.6268529 with a standard deviation of 0.08513. On the other hand, the mean positive affect for countries with a log GDP of more than 10.5 was found to be 0.709 with a standard deviation of 0.06813.



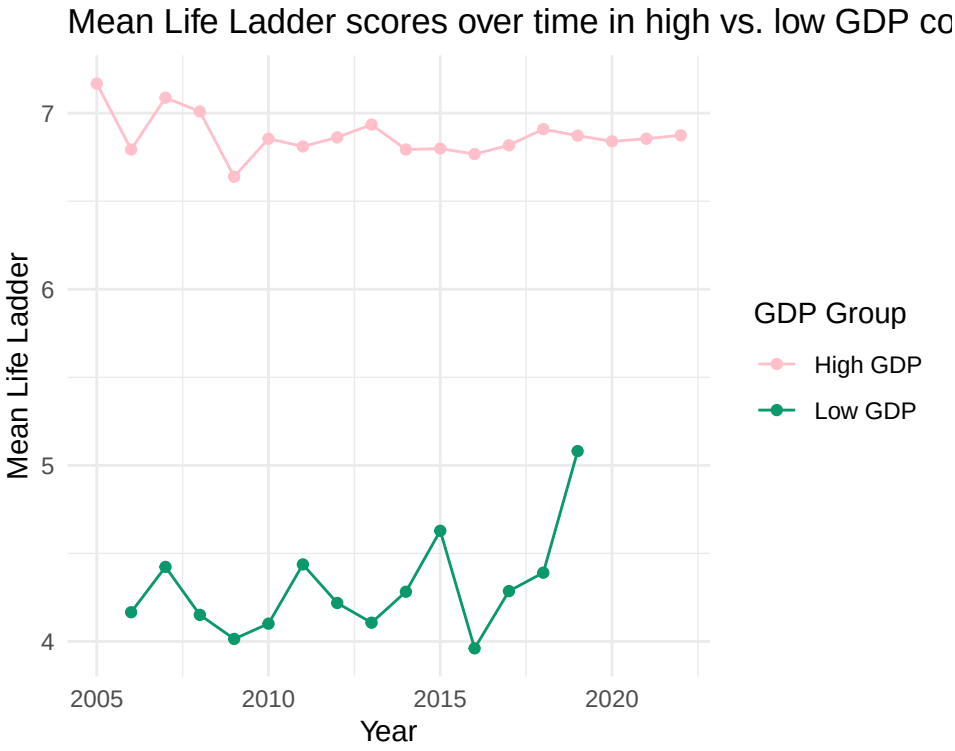
Here the distribution of the Life Ladder scores are shown between high GDP (log GDP > 10.5) and low GDP (log GDP < 7) countries. There is an even clearer difference in the overall distribution between the groups, with the High GDP group is concentrated towards much higher values. It was found that the low gdp group had a significantly lower average life ladder score than the high gdp group with a p value of <2.2e-16.



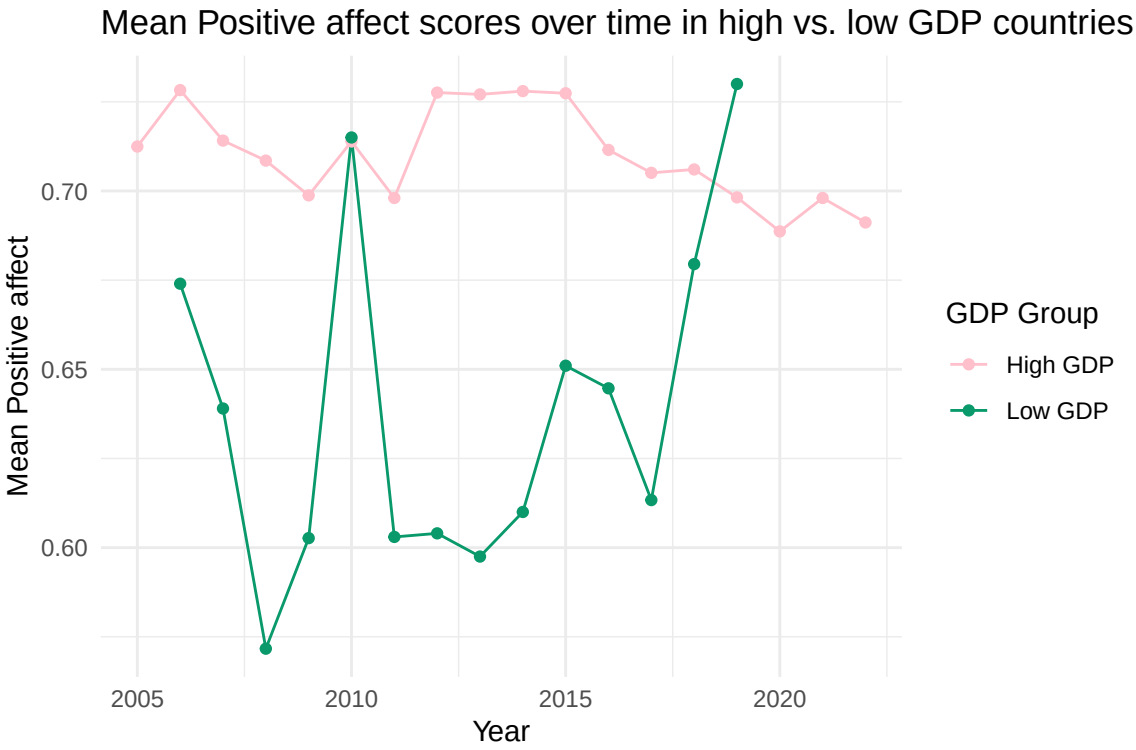
Above plots the distribution of the positive affect scores between high GDP (log GDP > 10.5) and low GDP (log GDP < 7) countries in the data set. While there are outlier points in the high GDP group, it is clear the overall distribution of the High GDP group is concentrated towards higher values. It was found that the low gdp group had a significantly lower positive affect score than the high gdp group with a p value 1.602e-06.

Plotting Life Ladder and Positive Affect in Low GDP vs High GDP countries over time.

Finally, we can plot the trend of average Life Ladder and positive affect over time between high GDP and low GDP countries.



The gap between life ladder scores in high gdp countries versus low gdp countries is stark, but that gap is showing a trend of becoming smaller in recent years. There appears to be a lack of data for low GDP countries after the year 2019.



This plot shows that gap between positive affect scores in high gdp countries versus low GDP countries is also large, but that gap appears to be interestingly closed in 2010 and as the year approaches 2020.

Conclusion

Overall, the mean "happiness" of a country measured by Life Ladder Scores and Positive Affect scores does change over time differently in countries with a low log GDP per capita (less than 7) versus a high log GDP per capita (higher than 10.5). The difference in these scores are significant, but the trend shows that there are years where the gap between these two groups get smaller, which can be further studied in the context of the countries. One potential cause of the high variability in low GDP countries like Mexico could be due to high rates of natural disasters that affect the country periodically. Overall, analyzing world happiness report data can help identify important characteristics of a country that keep citizens happy.