Happiness

AN EXPLORATORY DATA ANALYSIS BY ZACH MUMBAUER

Statistical Question

What are the primary factors that significantly influence the happiness score (Life Ladder) of countries, and how do these factors compare in terms of impact?

Hypothesis

Higher GDP per capita, greater social support, higher healthy life expectancy, and more freedom to make life choices are positively correlated with higher happiness scores.

Key Variables

- **Life Ladder (Happiness Score) => Dependent Variable**

- Log GDP per Capita

- Social Support

- Healthy Life Expectancy

- Freedom to Make Life Choices

- Generosity

- Perceptions of Corruption
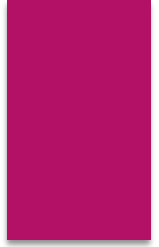
Descriptions of Selected Variables:

1. **Ladder score**:
   - Based on the Cantril Ladder question, where respondents rate their lives on a scale of 0 (worst possible life) to 10 (best possible life).
2. **Log GDP per capita**:
   - The natural logarithm of the Gross Domestic Product per capita adjusted for purchasing power parity (PPP) to reflect differences in living costs between countries.
3. **Social support**:
   - A measure of the average response to whether individuals have relatives or friends they can count on in times of trouble.
4. **Healthy life expectancy**:
   - The average number of years someone is expected to live in good health.
5. **Freedom to make life choices**:
   - Represents the average level of satisfaction individuals report about their ability to make choices about their own lives.
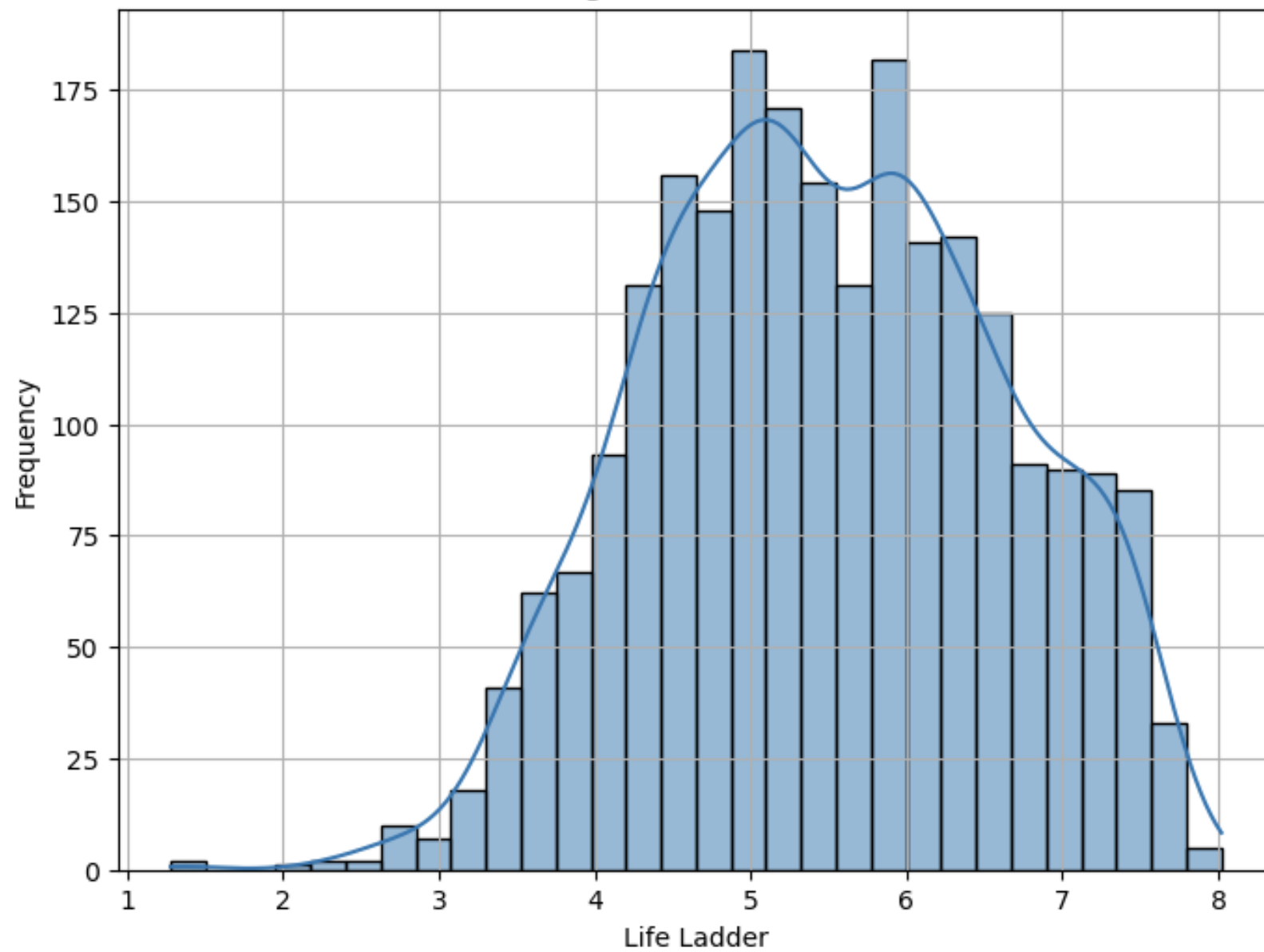6. **Generosity**:
   - Reflects altruistic behavior and its potential impact on happiness.
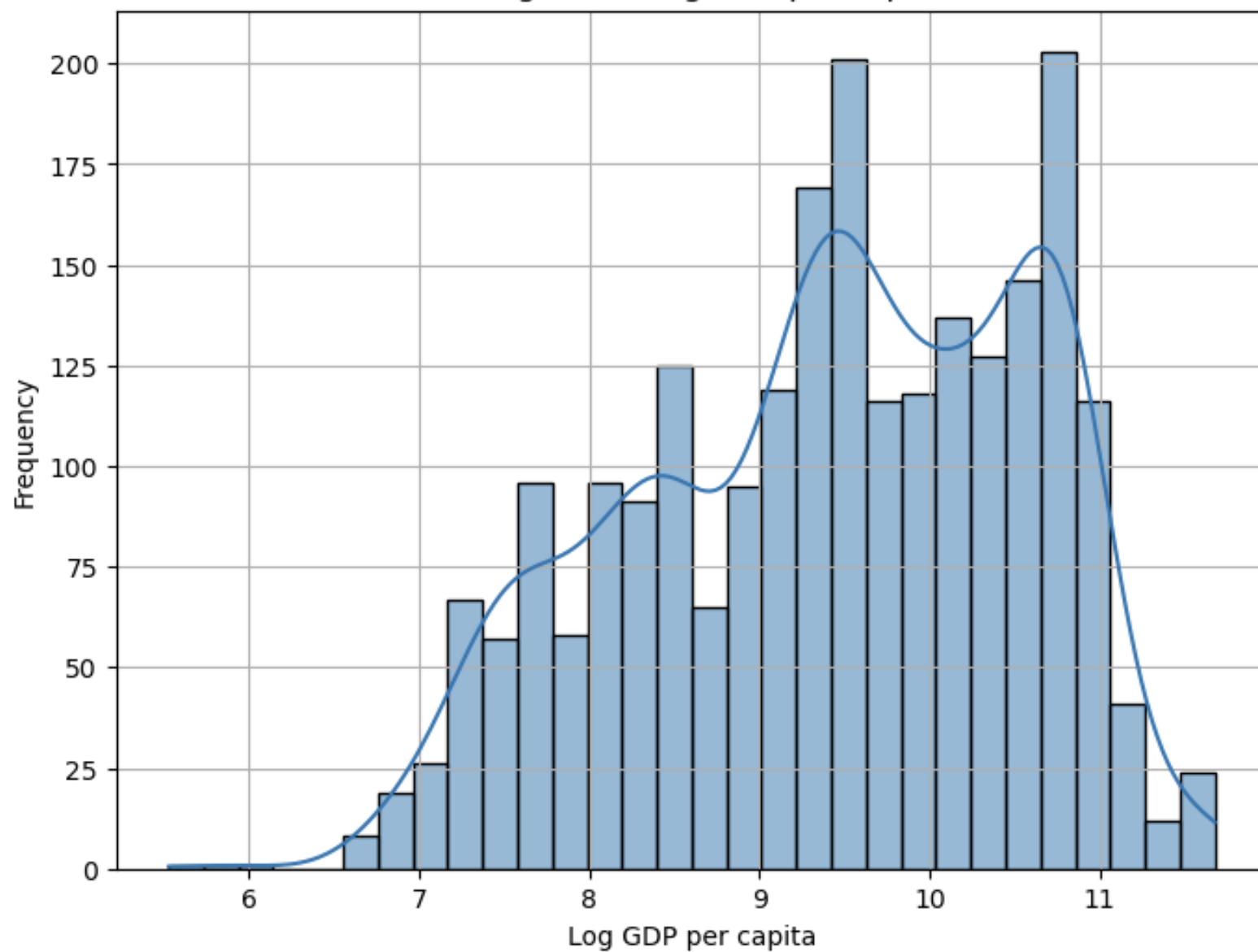7. **Perceptions of corruption**:
   - The national average of survey responses about the perceived extent of corruption in government and businesses.
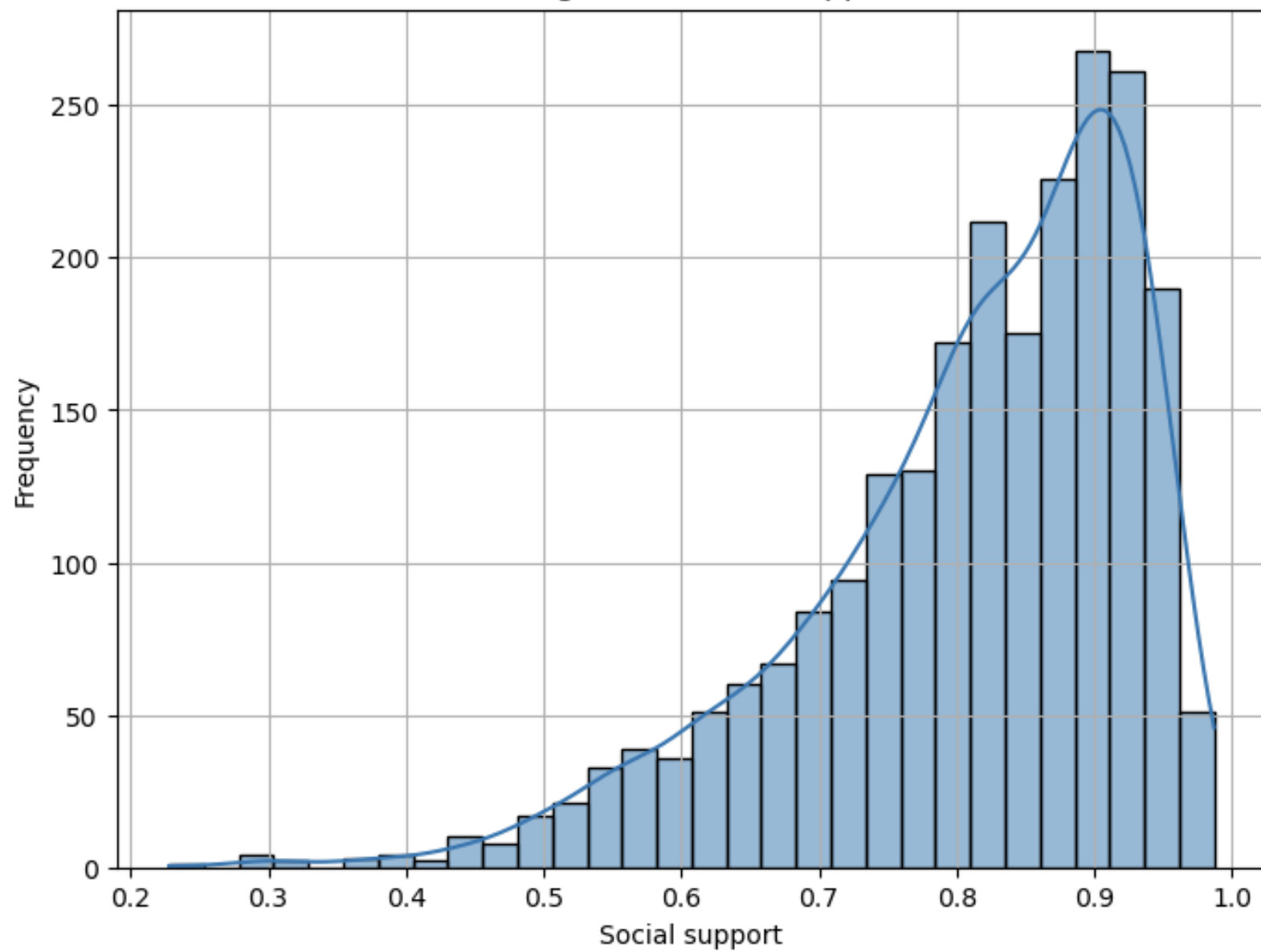
Histograms
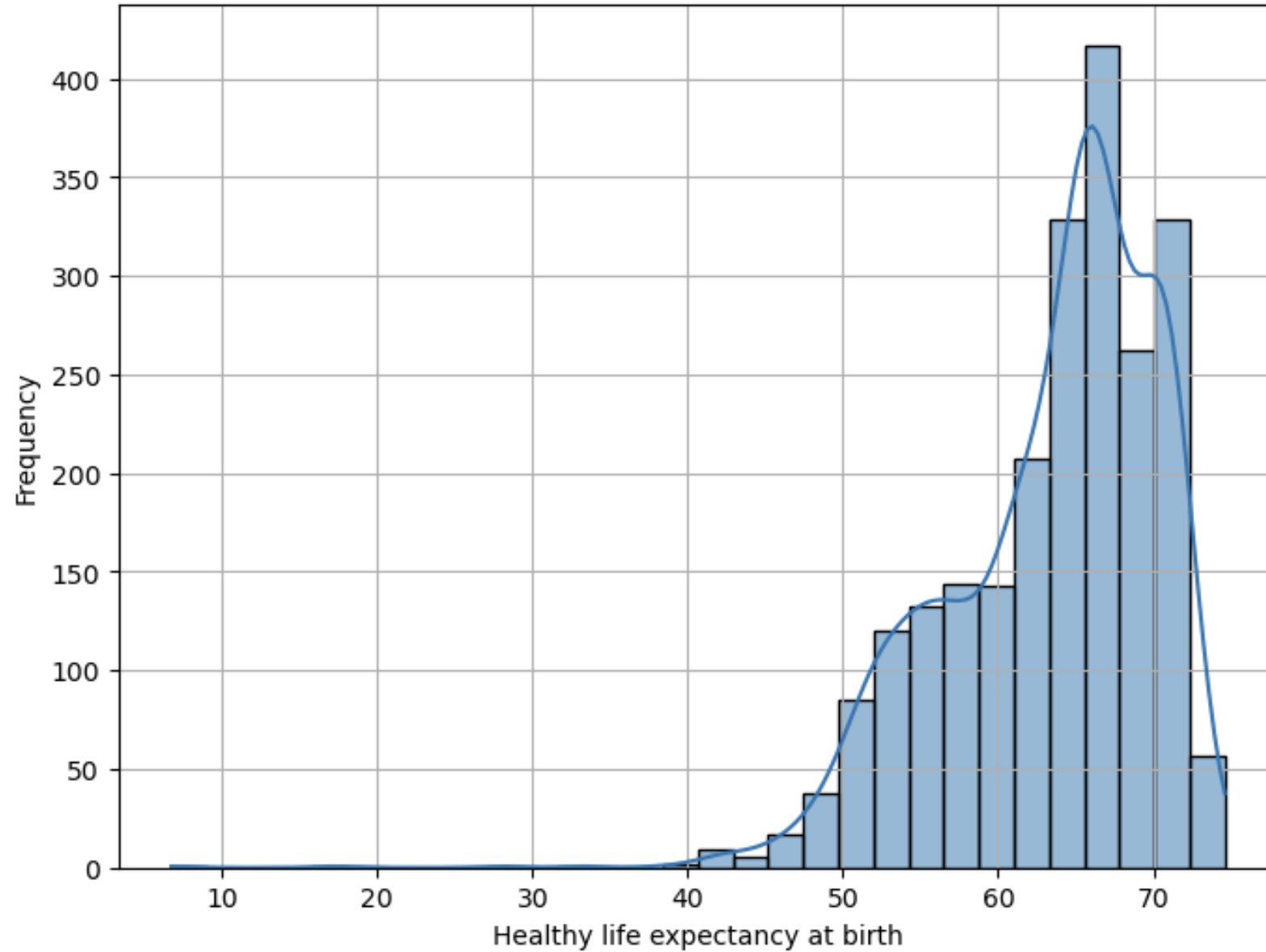
Histogram of Life Ladder
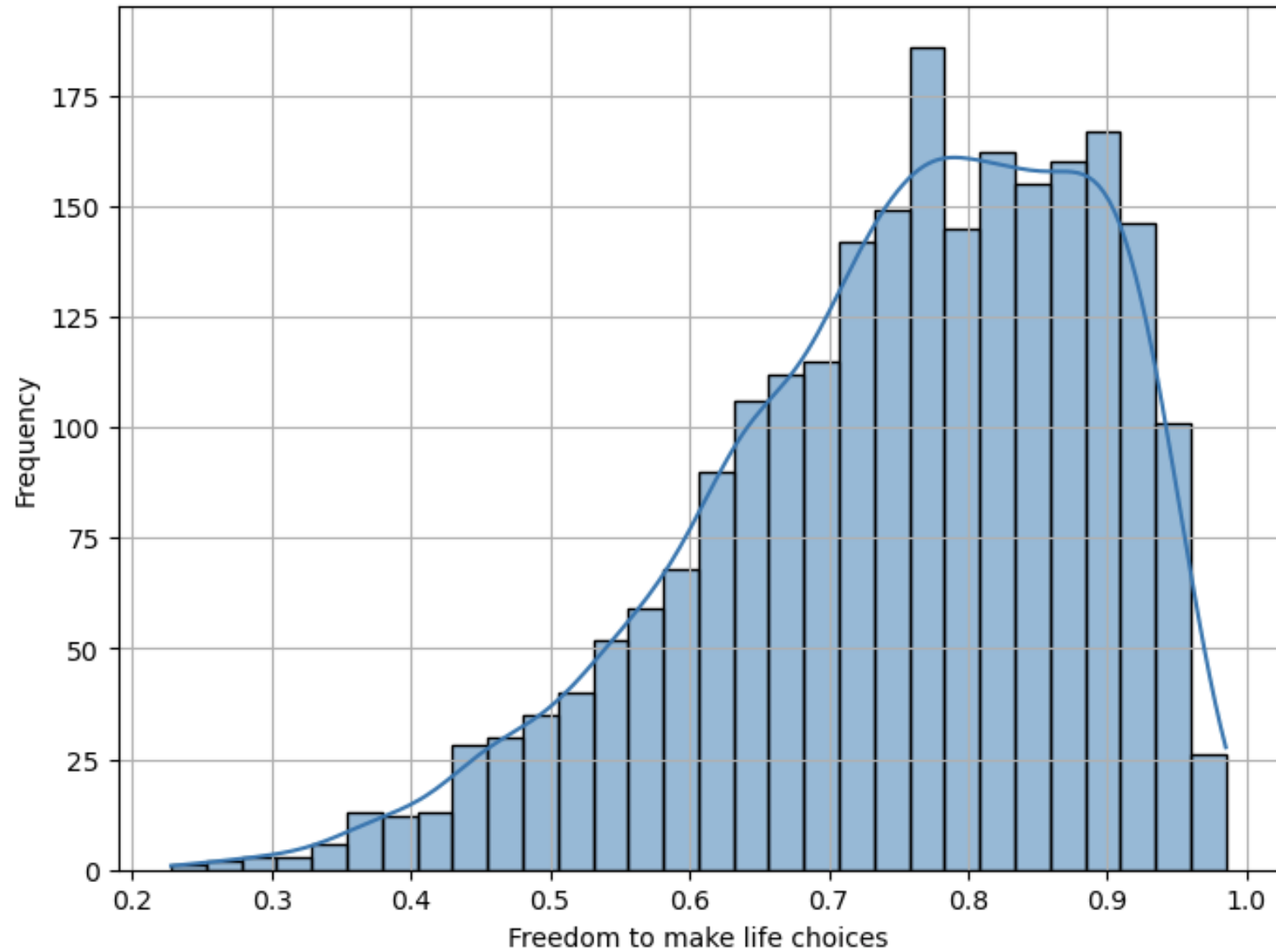
Histogram of Log GDP per capita
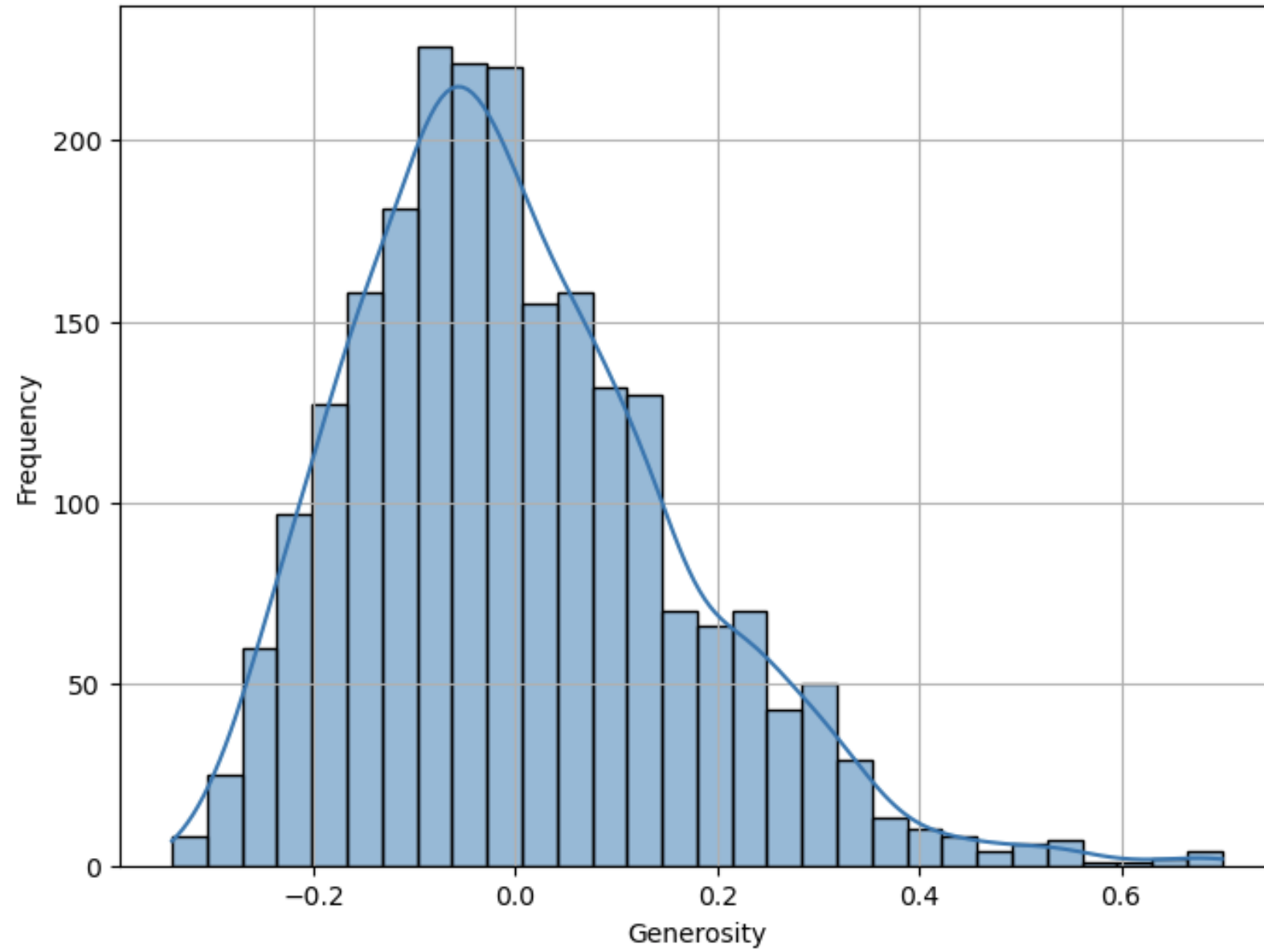
Histogram of Social support

Histogram of Healthy life expectancy at birth

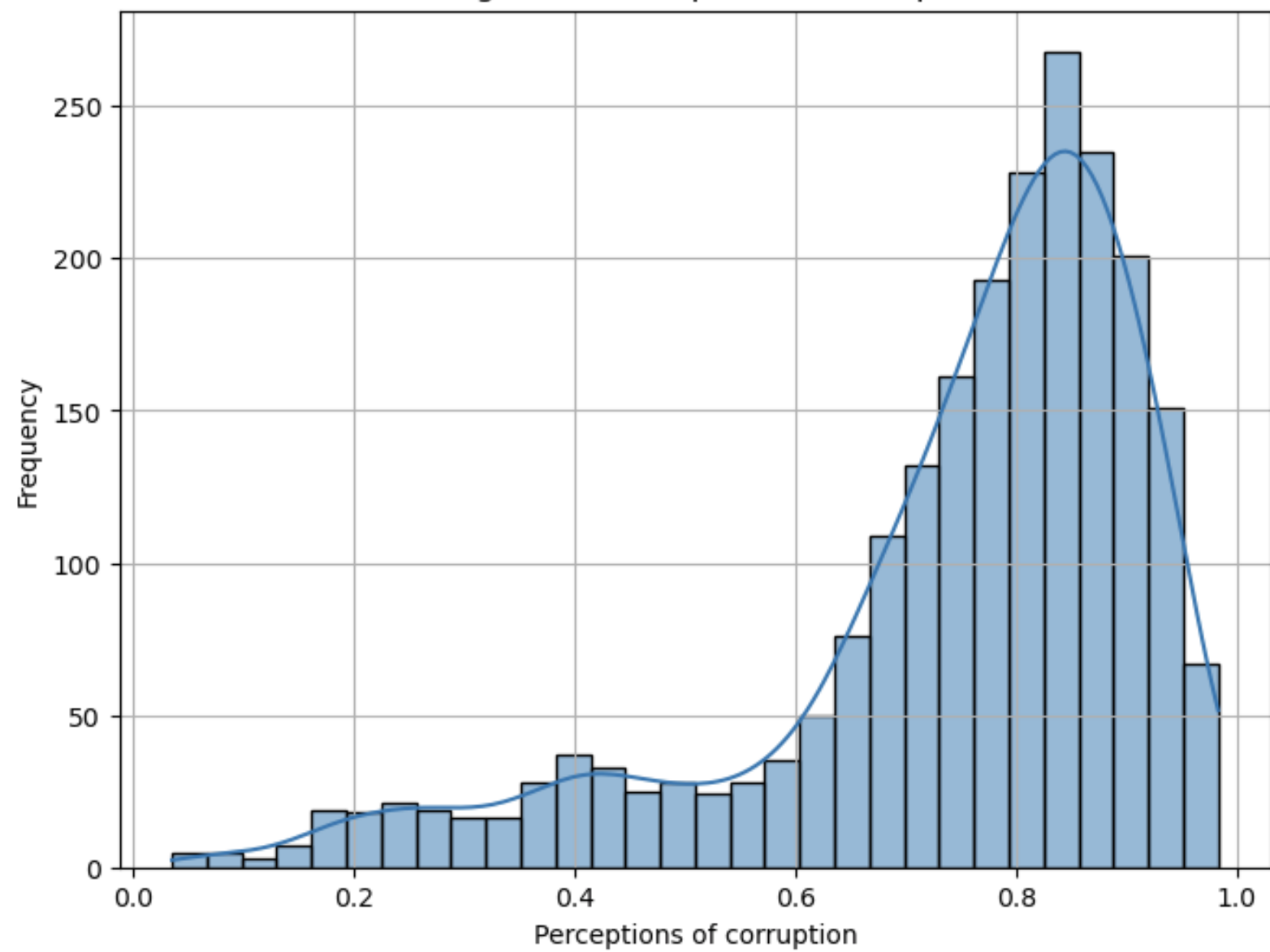Histogram of Freedom to make life choices

Histogram of Generosity

Histogram of Perceptions of corruption

```
                                        Mean    Mode Spread (Std Dev)    Min  \
Life Ladder                         5.483566   5.252           1.125522  1.281
Log GDP per capita                  9.399671  10.878           1.152069  5.527
Social support                      0.809369   0.937           0.121212  0.228
Healthy life expectancy at birth   63.401828    66.6           6.842644   6.72
Freedom to make life choices        0.750282   0.838           0.139357  0.228
Generosity                          0.000098   0.068           0.161388  -0.34
Perceptions of corruption           0.743971   0.844           0.184865  0.035


                                      Max Lower Bound Upper Bound  \
Life Ladder                         8.019     2.13225     8.83825
Log GDP per capita                 11.676      5.6775     13.2215
Social support                      0.987       0.504       1.144
Healthy life expectancy at birth    74.6    45.15875    82.58875
Freedom to make life choices        0.985      0.3595      1.1635
Generosity                            0.7   -0.420625    0.402375
Perceptions of corruption           0.983    0.415875    1.138875
```

Possible Explanations for the Outliers

1. Life Ladder (Happiness Score):

Both values map to Afghanistan. The U.S. withdrew from the country a few years ago, and the Taliban took over, which tracks with the years of 2022 and 2023.

2. Log GDP per capita:

There is only one outlier here, and it is Venezuela. This is due to their ongoing economic problems.

3. Social Support:

The outliers might correspond to countries with weaker community ties and more individualistic cultures.

4. Healthy Life Expectancy at Birth:

The very low outliers represent countries with poor healthcare systems, malnutrition, or war-related mortality.

5. Freedom to Make Life Choices:

Countries with very low outliers in this category often experience governance or cultural restrictions that severely limit individual autonomy and decision-making.

6. Generosity:

High values might reflect cultural or religious norms and low values could be linked to economic hardship.

7. Perceptions of Corruption:

I think this has to do with a more binary response to this variable.rs.

Based on the source of this dataset being the opinions of people on their happiness I think all of these outliers should be included in the analysis for the time being
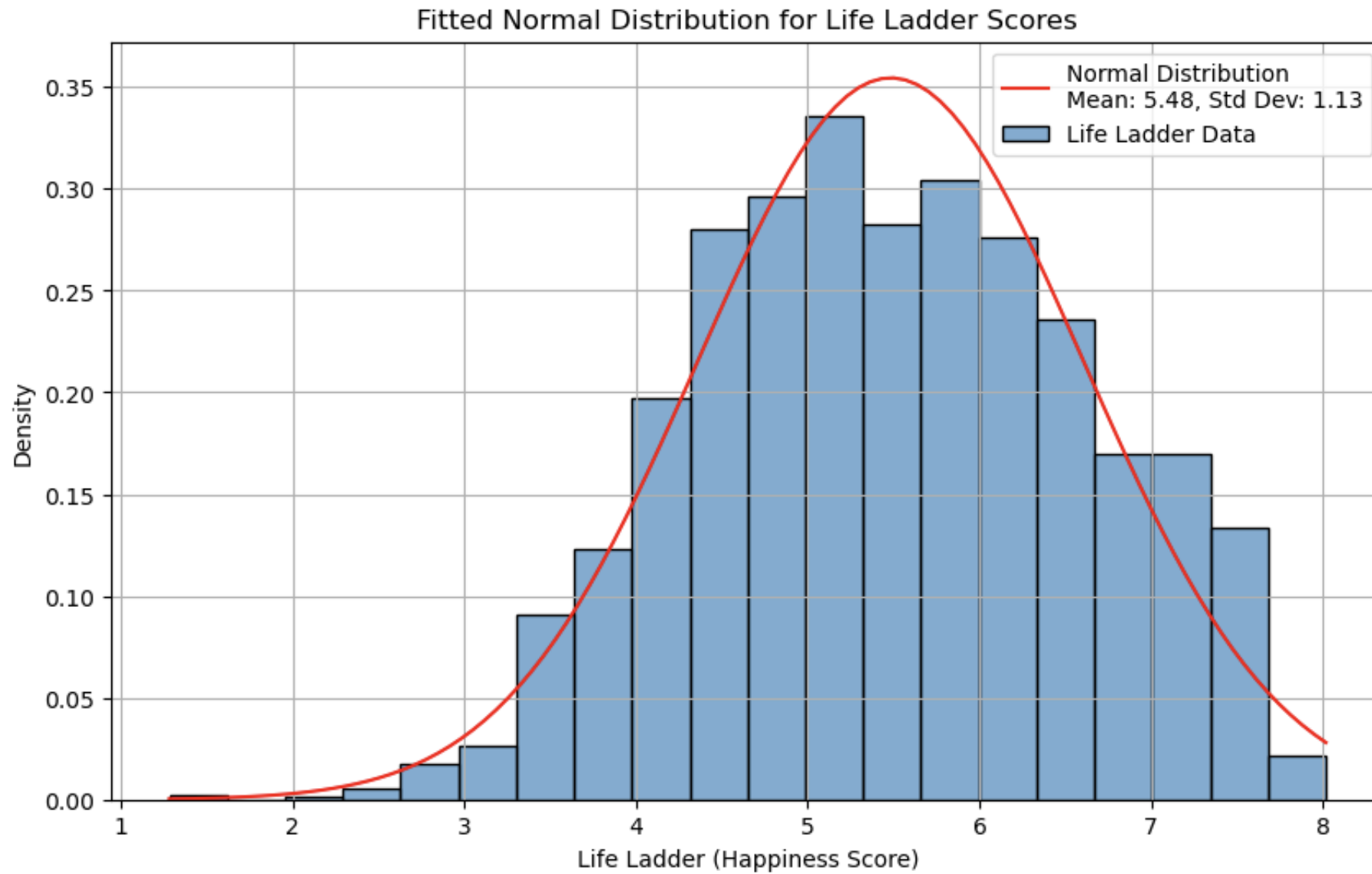
PMF of Life Ladder: High vs Low GDP per Capita
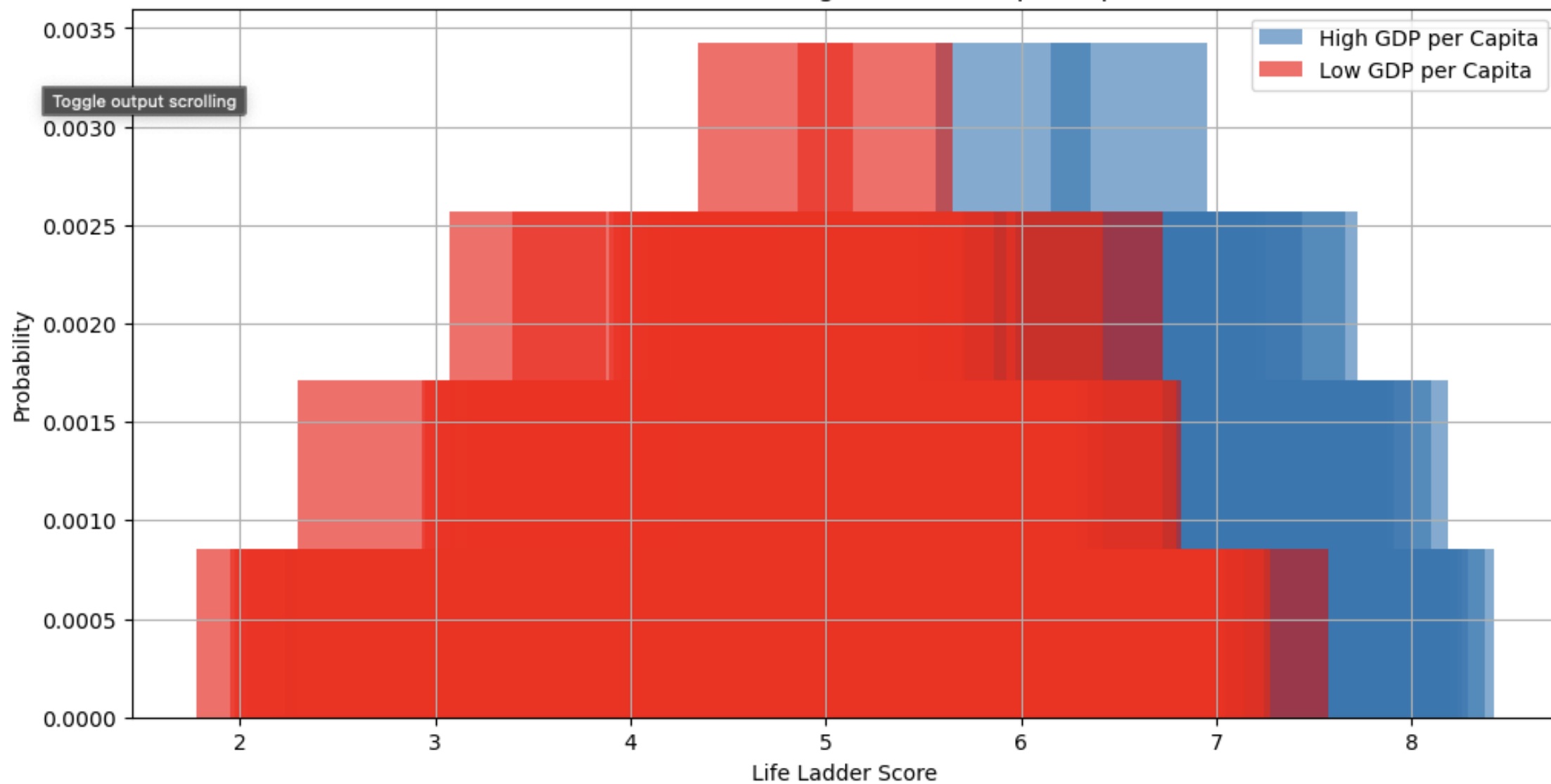
The PMF shows that countries with higher GDP per capita have a higher concentration of happiness scores in the upper range while lower GDP countries are more distributed toward moderate and low happiness levels, highlighting GDP's significant role in happiness distribution.

Cumulative Distribution Function (CDF) of Life Ladder Scores
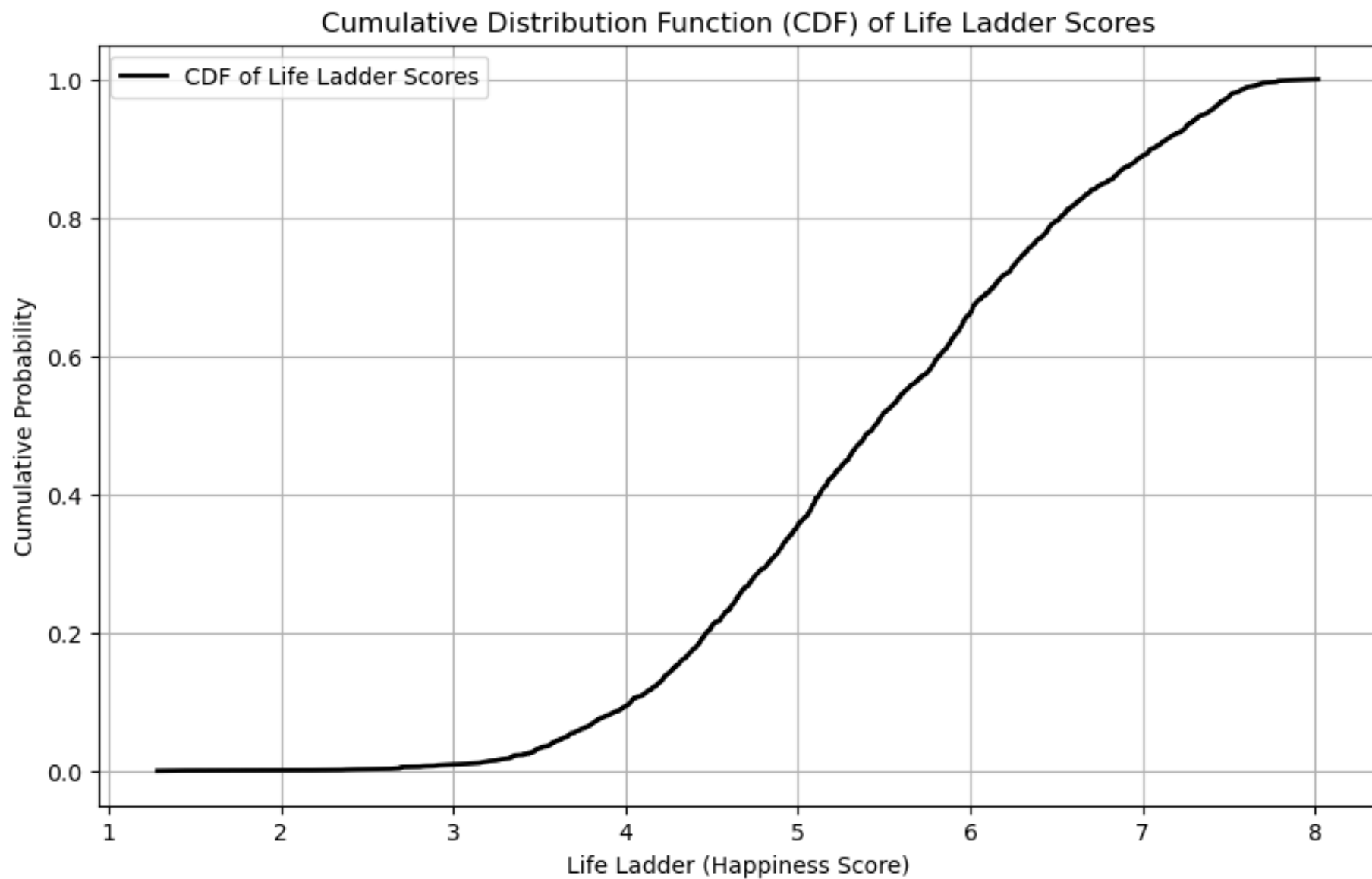

The curve shows a steep rise between scores of 4 and 6, indicating that most countries have moderate happiness levels. The lower tail shows a small group of countries with low happiness. There is an even smaller upper tail showing the rarity of extremely high happiness levels

Fitted Normal Distribution for Life Ladder Scores

Legend:
Normal Distribution
Mean: 5.48, Std Dev: 1.13
Life Ladder Data

X-axis: Life Ladder (Happiness Score)
Y-axis: Density

The fitted normal distribution confirms that most countries experience moderate happiness. From the longer tail on the left, we can estimate that it is more common to be less happy than more happy.

Scatter Plot: Log GDP per Capita vs Life Ladder
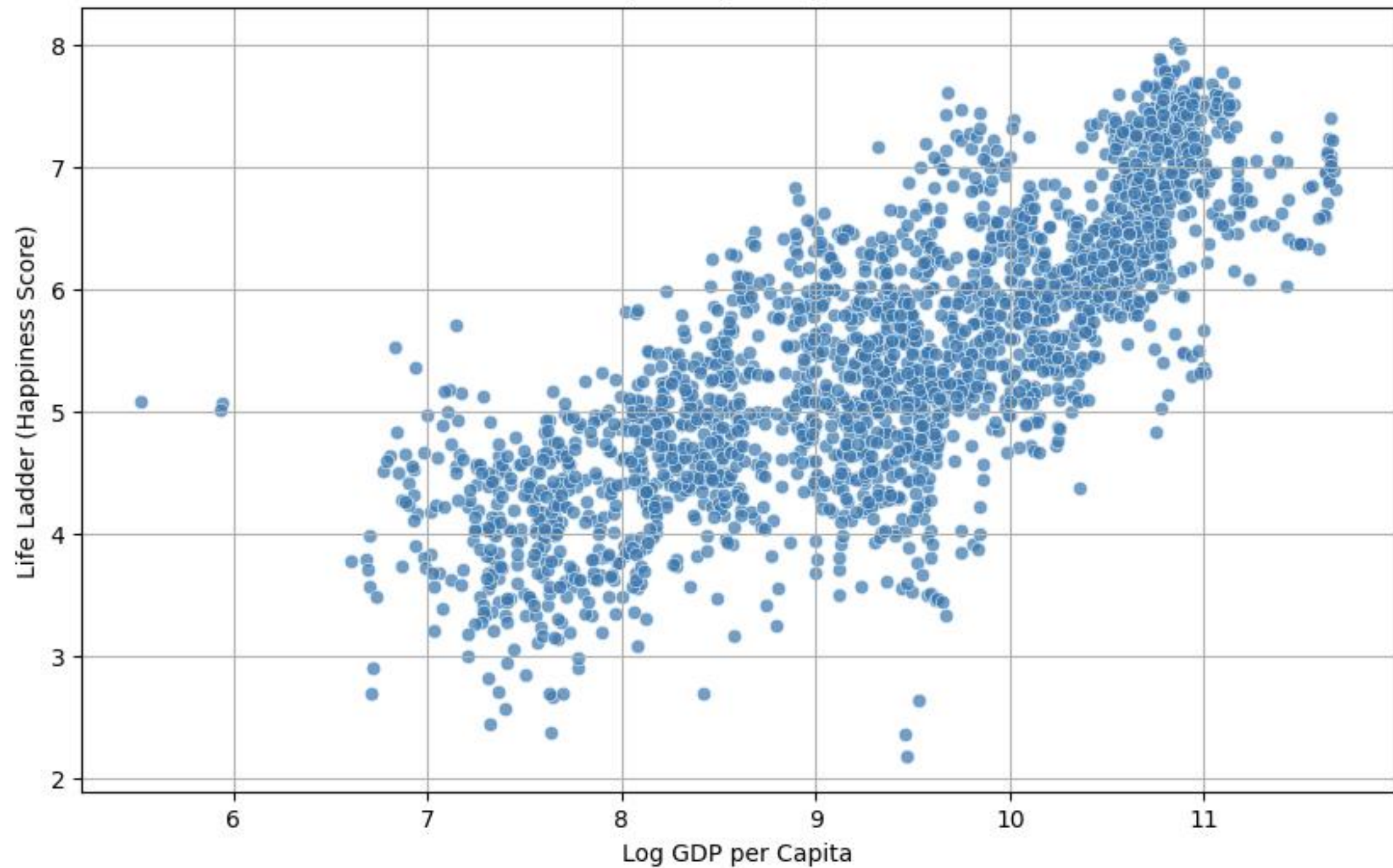
Scatter Plot: Social Support vs Life Ladder
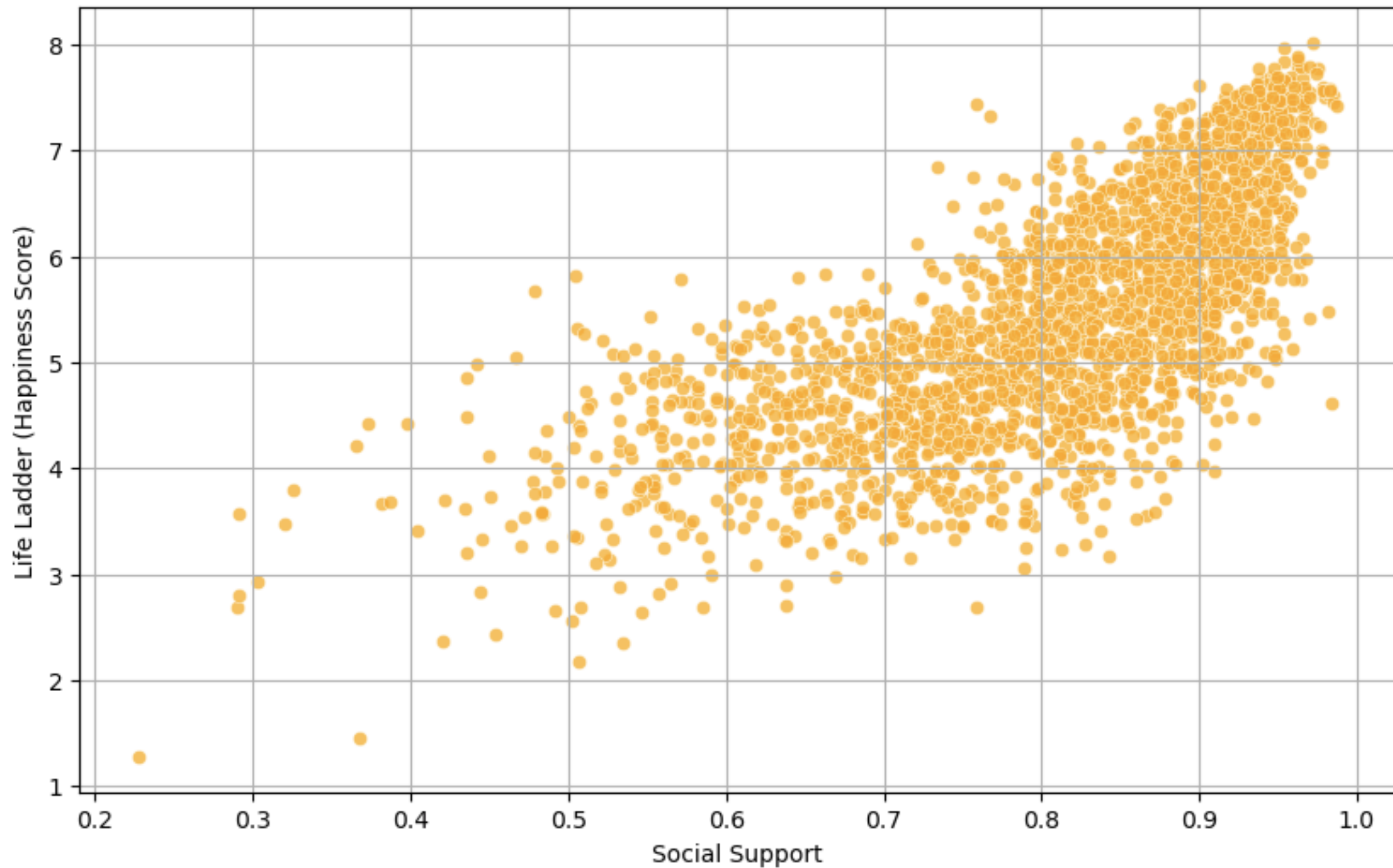
```
Correlation Results:
                Variable Comparison  Pearson Correlation  P-Value
0  Log GDP per Capita vs Life Ladder             0.783556      0.0
1      Social Support vs Life Ladder             0.722738      0.0
```

```
Covariance between Log GDP per Capita and Life Ladder: 1.0098891816427487
Covariance between Social Support and Life Ladder: 0.09860034315371864
```

The analysis highlights significant positive relationships between key predictors and happiness scores. Log GDP per capita exhibits a strong correlation with Life Ladder scores ($\rho = 0.78$, $p < 0.001$) and a covariance of 0.87, suggesting that economic prosperity plays a critical role in shaping happiness levels. Similarly, Social Support shows a strong correlation ($\rho = 0.72$, $p < 0.001$) and a covariance of 0.68, emphasizing the importance of strong social networks in enhancing well-being. Both variables display linear trends, which supports the hypothesis that higher GDP and greater social support contribute to increased happiness.

```
T-Test Results for High vs Low GDP Groups
 T-Statistic        P-Value  High GDP Mean   Low GDP Mean   High GDP Standard Deviation   Low GDP Standard Deviation
   41.594706 5.362231e-283       6.221256       4.761507                      0.879641                      0.814932
```

I used a two-sample t-test to compare happiness scores for countries with high vs. low GDP per capita. High-GDP countries had a mean score of 6.22 (SD = 0.88) and low-GDP countries scored 4.76 (SD = 0.81). The test showed a significant difference (t = 41.59, p < 0.001), supporting the hypothesis that higher GDP correlates with greater happiness.

OLS Regression Results

Dep. Variable:	Life Ladder	**R-squared:**	0.734
Model:	OLS	**Adj. R-squared:**	0.733
Method:	Least Squares	**F-statistic:**	1540.
Date:	Sat, 16 Nov 2024	**Prob (F-statistic):**	0.00
Time:	22:50:03	**Log-Likelihood:**	-1970.6
No. Observations:	2241	**AIC:**	3951.
Df Residuals:	2236	**BIC:**	3980.
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-3.0900	0.118	-26.177	0.000	-3.321	-2.858
Log GDP per capita	0.4016	0.021	19.230	0.000	0.361	0.443
Social support	2.4684	0.144	17.167	0.000	2.186	2.750
Healthy life expectancy at birth	0.0221	0.003	6.945	0.000	0.016	0.028
Freedom to make life choices	1.8665	0.099	18.788	0.000	1.672	2.061

Omnibus:	26.485	**Durbin-Watson:**	0.525
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	38.629
Skew:	-0.121	**Prob(JB):**	4.09e-09
Kurtosis:	3.596	**Cond. No.**	768.
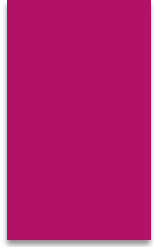
The OLS regression evaluated the impact of Log GDP per capita, Social Support, Healthy Life Expectancy, and Freedom to Make Life Choices on happiness scores. The model explained 73.4% of the variance (R-squared = 0.734), with all predictors showing significant positive relationships:

- **Log GDP per capita**: Coefficient = 0.4016, $p < 0.001$
- **Social Support**: Coefficient = 2.4684, $p < 0.001$
- **Healthy Life Expectancy**: Coefficient = 0.0221, $p < 0.001$
- **Freedom to Make Life Choices**: Coefficient = 1.8665, $p < 0.001$