finalprojectcomp

August 6, 2023

0.1 PSTAT 100 Course Project - Can Money Buy Happiness?

Robert Oros, Mackenzie Sullivan, and Cameron Clark

Data Description: World Happiness Report 2005-2022 *Summary of the data:* The World Happiness Report dataset provides valuable insights into happiness and wellbeing across different countries over time. It comprises several key indices that offer a multidimensional view of factors contributing to the happiness and life satisfaction of individuals in different regions. By analyzing this dataset, we can gain a deeper understanding of the complex relationship between economic development, social support, individual freedom, perceived corruption, and happiness. This high-level exploration of the dataset allows us to investigate whether money can truly buy happiness and to explore the correlation between economic prosperity and life satisfaction.

Sampling: The dataset covers a wide temporal range, starting from 2005 and continuing until 2022. It includes data for various countries over these years. Each row in the dataset represents a specific country's data for a particular year. The data was likely collected through surveys and reports from multiple sources.

Data Semantics:

The dataset includes eleven variables. Here are their descriptions: 1. *Country*: Name of the country for which the data is recorded.

2. *Year*: The year for which the data is collected
3. *Life ladder*: A high value in the "Life ladder" column indicates that individuals in the country report a higher level of life satisfaction and happiness. This means that, on average, people in the country perceive their lives as more fulfilling, content, and positive. A low value indicates the opposite.
4. *Log GDP per capita*: A high value in the "Log GDP per capita" column indicates a higher level of economic prosperity and wealth in the country. This means that, on average, the Gross Domestic Product (GDP) per capita is higher, and the country's citizens may have access to better economic opportunities and a higher standard of living. A low value indicates the opposite.
5. *Social support*: A high value in the "Social support" column indicates a strong and robust social support system within the country. This means that individuals in the country feel supported by their family, friends, and community, leading to a sense of belonging and connection. A low value indicates the opposite.

6. *Healthy life expectancy at birth*: A high value in the "Healthy life expectancy at birth" column indicates a higher average life expectancy with good health at birth. This means that, on average, people in the country are expected to live longer lives and enjoy better health during their lifetime. A low value indicates the opposite.
7. *Freedom to make life choices*: A high value in the "Freedom to make life choices" column indicates a higher level of individual freedom and autonomy. This means that individuals in the country have more opportunities to make their life decisions, pursue their aspirations, and have control over their own lives. A low value indicates the opposite.
8. *Generosity*: A high value in the "Generosity" column suggests a higher level of generosity and willingness to help others within the country. This means that people in the country are more likely to engage in charitable acts and support their fellow citizens. A low value indicates the opposite.
9. *Perceptions of corruption*:A high value in the "Perceptions of corruption" column indicates a higher perceived level of corruption in the country. This means that people in the country believe that corruption is more prevalent in various institutions and public sectors. A low value indicates the opposite.
10. *Positive affect*: A high value in the "Positive affect" column indicates a higher frequency and intensity of positive emotions experienced by individuals in the country. This means that, on average, people in the country tend to experience more positive emotions such as happiness, joy, and satisfaction in their daily lives. A low value indicates the opposite.
11. *Negative affect*: Converse of positive affect.

Question of Interest: Can money buy happiness? Whether or not money can buy happiness for individuals is one thing, but this data allows us to see if the question can be answered on a country-wide scale. Our analysis will study just how predictive an objective measurement of a country's wealth (in our case, Log GDP per capita) is of its citizen's reported happiness. We focus on studying this variable's relationship with those in our dataset that relate to how satisfied a country's people are with their lives, rather than other objective measurements. A satisfactory answer to this question will address its complexity by explaining which variables have a strong relationship with GDP and what this implies for wealth's influence on a people's reported happiness.

Data Analysis First, we import libraries.

[94]:
```
import numpy as np
import pandas as pd
import altair as alt
from scipy import linalg
import matplotlib.pyplot as plt
import seaborn as sns
import statsmodels.api as sm
import seaborn as sns
import matplotlib.pyplot as plt
```

Now we import data and geta look at what we'll be working with.

[95]:
```
# import whr data
whr = pd.read_csv('data/whr-2023.csv')
```

[96]:
```
whr.head()
```

[96]:
```
  Country name  year  Life Ladder  Log GDP per capita  Social support  \
0  Afghanistan  2008        3.724               7.350           0.451
1  Afghanistan  2009        4.402               7.509           0.552
2  Afghanistan  2010        4.758               7.614           0.539
3  Afghanistan  2011        3.832               7.581           0.521
4  Afghanistan  2012        3.783               7.661           0.521

   Healthy life expectancy at birth  Freedom to make life choices  Generosity  \
0                              50.5                         0.718       0.168
1                              50.8                         0.679       0.191
2                              51.1                         0.600       0.121
3                              51.4                         0.496       0.164
4                              51.7                         0.531       0.238

   Perceptions of corruption  Positive affect  Negative affect
0                      0.882            0.414            0.258
1                      0.850            0.481            0.237
2                      0.707            0.517            0.275
3                      0.731            0.480            0.267
4                      0.776            0.614            0.268
```

First, we'll be looking at the 'Possitive affect' and 'Negative affect' varibles. Our plan is to used them to create a new variable called 'Net Positive Affect', which is simply the difference between the two, this will give us a sense of which country's residence have more positive feelings toward their home than negative ones. First, we check to see if the columns contain missing data.

[97]:
```
print(whr.isna().mean()) # missing data averages
```

```
Country name                        0.000000
year                                0.000000
Life Ladder                         0.000000
Log GDP per capita                  0.009095
Social support                      0.005912
Healthy life expectancy at birth    0.024557
Freedom to make life choices        0.015007
Generosity                          0.033197
Perceptions of corruption           0.052751
Positive affect                     0.010914
Negative affect                     0.007276
dtype: float64
```

They do, so we clean the data up by removing the rows that contain them.

[98]:
```
whr_clean = whr.dropna(subset=['Positive affect', 'Negative affect'])
```

Next, lets use the 'groupby' function to combine and average out the data for all the rows that share a country. This way we won't need to worry about the 'year' variable.

[99]:
```
whr_clean.groupby('Country name').mean()
```

[99]:
```
                     year  Life Ladder  Log GDP per capita  Social support  \
Country name
Afghanistan   2014.642857     3.346643            7.585615        0.484500
Albania       2014.933333     5.047933            9.396933        0.715800
Algeria       2016.444444     5.367778            9.343556        0.814889
Angola        2012.500000     4.420250            8.985750        0.738250
Argentina     2014.000000     6.283588           10.030412        0.902412
…                     …          …                  …              …
Venezuela     2013.882353     5.962059            8.588929        0.903706
Vietnam       2014.266667     5.396933            8.964333        0.824467
Yemen         2013.416667     3.912250            7.925250        0.739833
Zambia        2013.733333     4.453733            8.051533        0.729800
Zimbabwe      2014.000000     3.805294            7.611529        0.784824

              Healthy life expectancy at birth  Freedom to make life choices  \
Country name
Afghanistan                          52.533929                      0.498571
Albania                              68.505333                      0.683133
Algeria                              66.144444                      0.522000
Angola                               52.150000                      0.456250
Argentina                            66.664706                      0.774529
…                                          …                             …
Venezuela                            64.922353                      0.668353
Vietnam                              64.876000                      0.892769
Yemen                                58.309167                      0.622417
Zambia                               51.535000                      0.761867
Zimbabwe                             48.805882                      0.596412

              Generosity  Perceptions of corruption  Positive affect  \
Country name
Afghanistan     0.060000                   0.842786         0.433286
Albania        -0.074733                   0.869600         0.557267
Algeria        -0.131143                   0.709143         0.535667
Angola         -0.090500                   0.866750         0.625750
Argentina      -0.152471                   0.838647         0.739000
…                  …                         …              …
Venezuela      -0.098615                   0.797059         0.775118
Vietnam        -0.019200                   0.782083         0.617467
Yemen          -0.120917                   0.824667         0.458333
Zambia          0.021333                   0.828533         0.678867
```

```
Zimbabwe       -0.067588                   0.833706         0.651235

              Negative affect
Country name
Afghanistan          0.364357
Albania              0.293267
Algeria              0.267222
Angola               0.351250
Argentina            0.287588
…                       …
Venezuela            0.269529
Vietnam              0.210267
Yemen                0.293583
Zambia               0.298000
Zimbabwe             0.223471

[163 rows x 10 columns]
```

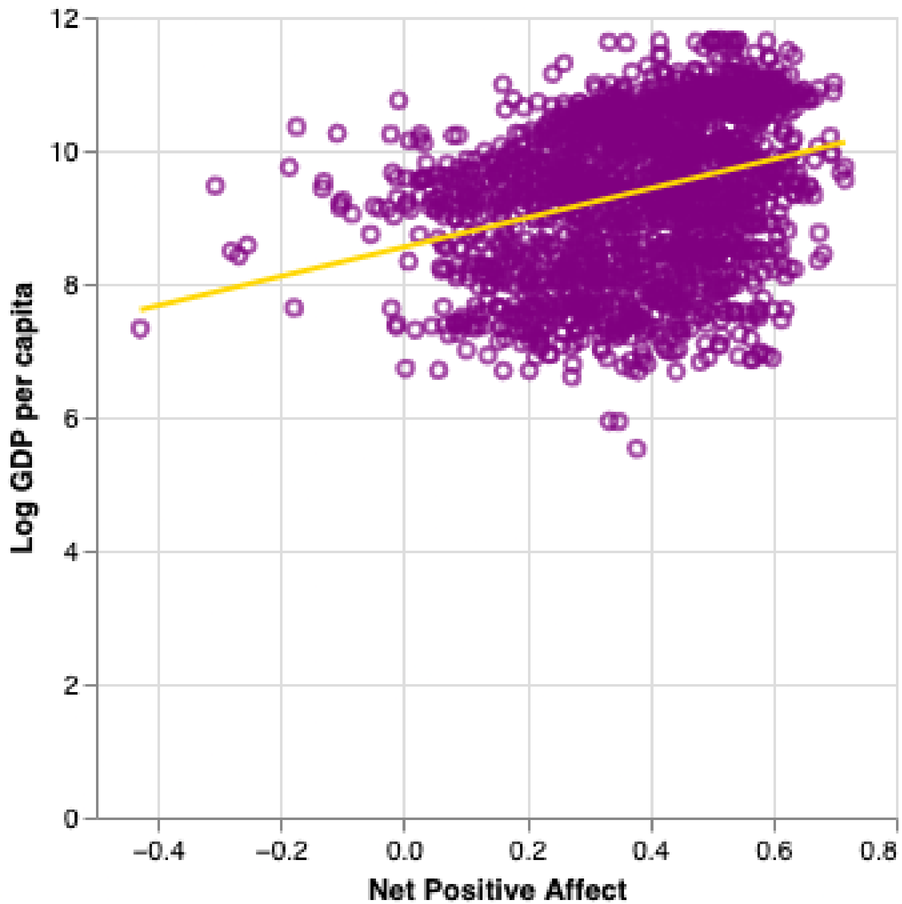
Now we create the new variable mentioned above.

```
[100]: whr_clean_plus = whr_clean.copy()
whr_clean_plus['Net Positive Affect'] = whr_clean['Positive affect'] -␣
 ↪whr_clean['Negative affect']
```

Lets plot our new variable's relationship with 'Log GDP per capita'. We'll be doing this on a scatter plot with a different-colored trend line that stands out.

```
[131]: figNPA = alt.Chart(whr_clean_plus).mark_point(color='purple').encode(
    x = 'Net Positive Affect',
    y = alt.Y('Log GDP per capita', scale = alt.Scale())
)

trendlineNPA = figNPA.transform_regression('Net Positive Affect', 'Log GDP per␣
 ↪capita').mark_line(color='gold')

display(figNPA + trendlineNPA)
#had problems displaying graph in PDF, saved image from Jupyter notebooks,␣
 ↪uploaded it, then posted it in markdown cell
```

```
alt.LayerChart(…)
```



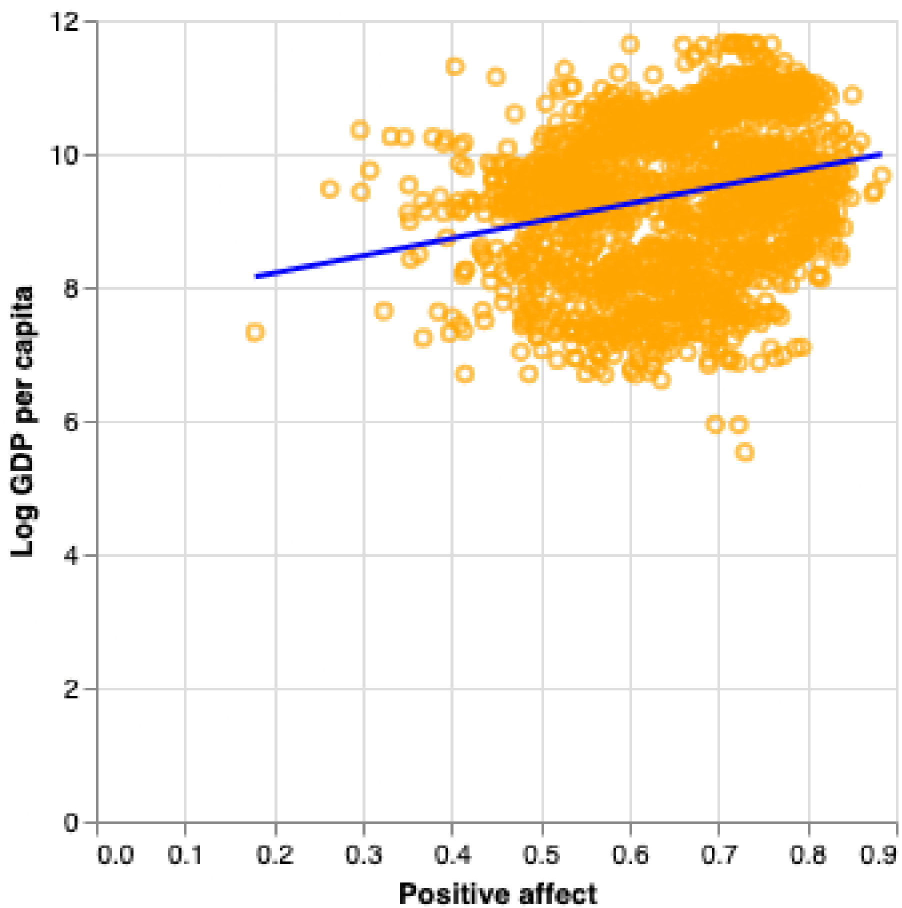
This display reveals that this question is going to need some further analysis to really answer. While the trend line does show a noticeable (though not particularly large) positive correlation between the two variables, the data is so varied that it isn't clear whether or not they have a true relationship. This is echoed when we graph the positive and negative affect individually.

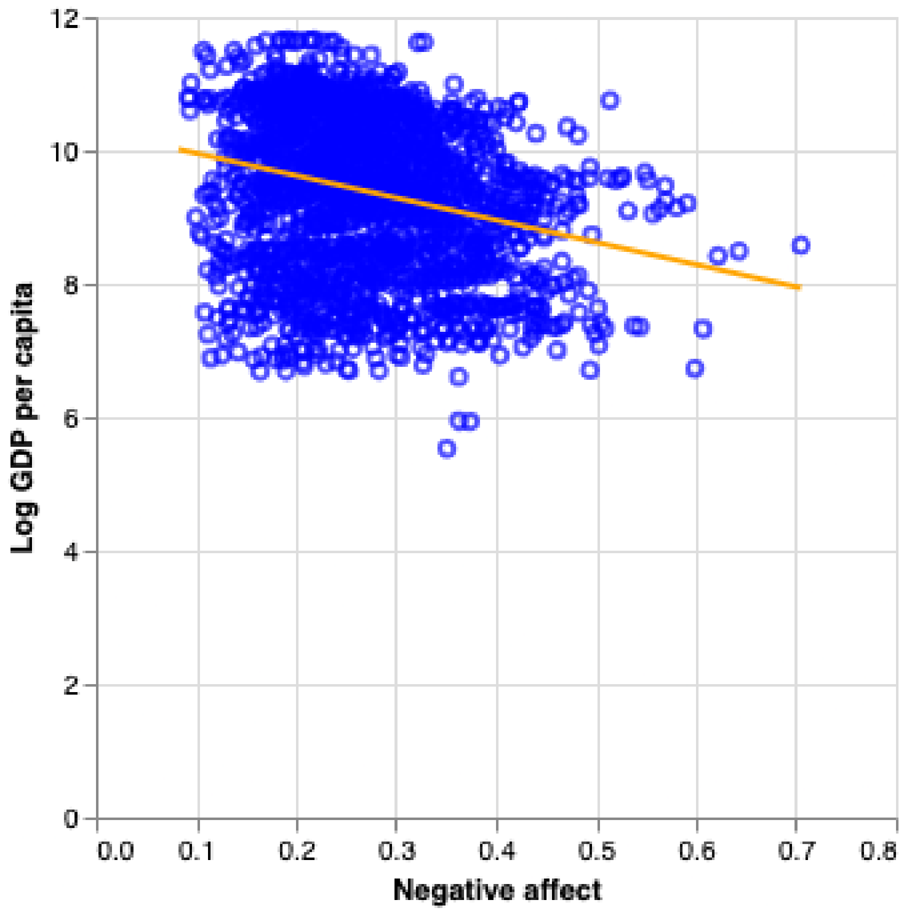
[145]:
```
figPA = alt.Chart(whr_clean_plus).mark_point(color='orange').encode(
    x = 'Positive affect',
    y = alt.Y('Log GDP per capita', scale = alt.Scale())
)

trendlinePA = figPA.transform_regression('Positive affect', 'Log GDP per␣
 ↪capita').mark_line(color='blue')

display(figPA + trendlinePA)

figNA = alt.Chart(whr_clean_plus).mark_point(color='blue').encode(
    x = 'Negative affect',
    y = alt.Y('Log GDP per capita', scale = alt.Scale())
)
```

```
trendlineNA = figNA.transform_regression('Negative affect', 'Log GDP per␣
↪capita').mark_line(color='orange')

display(figNA + trendlineNA)
#had problems displaying charts in PDF, saved images from Jupyter notebooks,␣
↪uploaded them, then posted them in markdown cell
```

alt.LayerChart(…)

alt.LayerChart(…)





We will need to go deeper to get a better sense of how GDP relates to some of our other variable. Lets try something else with the 'Freedom to make life choices' score.

Which countries have the most freedom to make life choices?

A higher freedom score indicates that the members of that country are more satisfied with their freedom to choose what they do with their life than those in lower scoring countries.

[103]:
```
gdp_low = whr['Log GDP per capita'].quantile(0.25)
gdp_medium = whr['Log GDP per capita'].quantile(0.5)
gdp_high = whr['Log GDP per capita'].quantile(0.75)

# This Creates a new column to categorize countries based on GDP groups
whr['GDP Group'] = pd.cut(whr['Log GDP per capita'], bins=[float('-inf'),␣
↪gdp_low, gdp_medium, float('inf')], labels=['Low', 'Medium', 'High'])
```

[104]:
```
whr_time_mean = whr.drop(columns = [ 'year', 'GDP Group']).groupby(['Country␣
↪name'],
                        as_index = False).mean().sort_values(by='Freedom to make␣
↪life choices', ascending=False)
```

```
top_countries_df = whr_time_mean.head(5)[['Country name', 'Freedom to make life␣
↪choices', 'Log GDP per capita']]
top_countries_df
```

[104]:

	Country name	Freedom to make life choices	Log GDP per capita
111	Norway	0.951583	11.063583
39	Denmark	0.943471	10.890588
48	Finland	0.942533	10.758267
23	Cambodia	0.939937	8.116471
159	Uzbekistan	0.933571	8.698500

These scores indicate that Norway, Denmark, Finland, Cambodia and Uzbekistan have citizens who are satisfied with the level of freedom and autonomy that are provided to them in their country. These citizens are able to have more control over what they do and what goals they aim to pursue.

[105]:
```
whr_time_mean = whr.drop(columns = [ 'year', 'GDP Group']).groupby(['Country␣
↪name'],
                         as_index = False).mean().sort_values(by='Freedom to make␣
↪life choices')
bot_countries_df = whr_time_mean.head(5)[['Country name', 'Freedom to make life␣
↪choices', 'Log GDP per capita']]
bot_countries_df
```

[105]:

	Country name	Freedom to make life choices	Log GDP per capita
36	Cuba	0.281000	NaN
22	Burundi	0.450800	6.682200
3	Angola	0.456250	8.985750
59	Haiti	0.461636	8.029818
135	South Sudan	0.493750	NaN

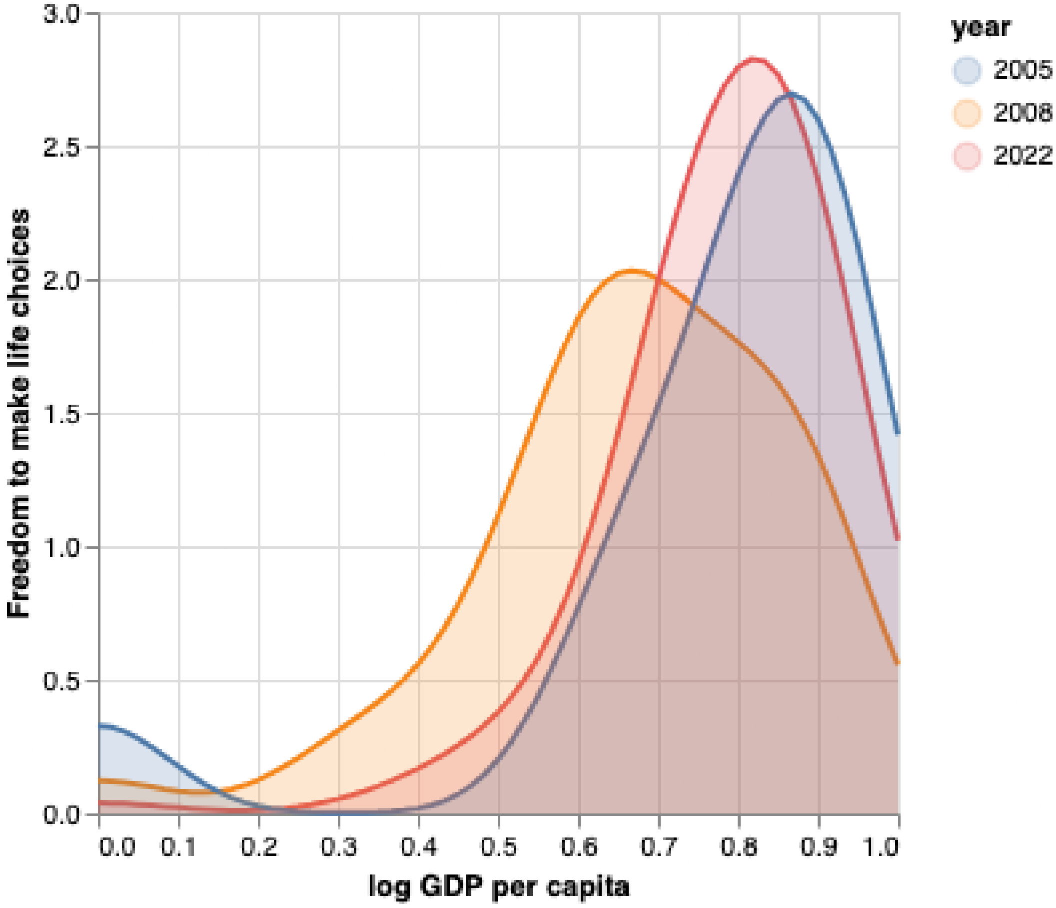
These scores indicate that Cuba, Burundi, Angola, Haiti and South Sudan have citizens who are dissatisfied with the level of freedom and autonomy that are provided to them in their country. These citizens have limited control over what they do or what goals they aim to pursue.

Histogram comparing the freedom to make life choices at the first year studied, recession year and last year studied:

[130]:
```
p = alt.Chart(whr).transform_filter(
    alt.FieldOneOfPredicate(field = 'year',
                            oneOf = [2005, 2008, 2022]) #first, recession, last
).transform_density(
    density = 'Freedom to make life choices', #variable being compared
    groupby = ['year'],
    as_ = ['log GDP per capita', 'Freedom to make life choices'], #y-axis,␣
↪x-axis
    bandwidth = 0.09,
    extent = [0, 1],
    steps = 60
```

```
).mark_line().encode(
    x = 'log GDP per capita:Q',
    y = 'Freedom to make life choices:Q',
    color = 'year:N'
)

display(p + p.mark_area(opacity = 0.2))
#had problems displaying chart in PDF, saved image from Jupyter notebooks,␣
↪uploaded it, then posted it in markdown cell
```

alt.LayerChart(…)

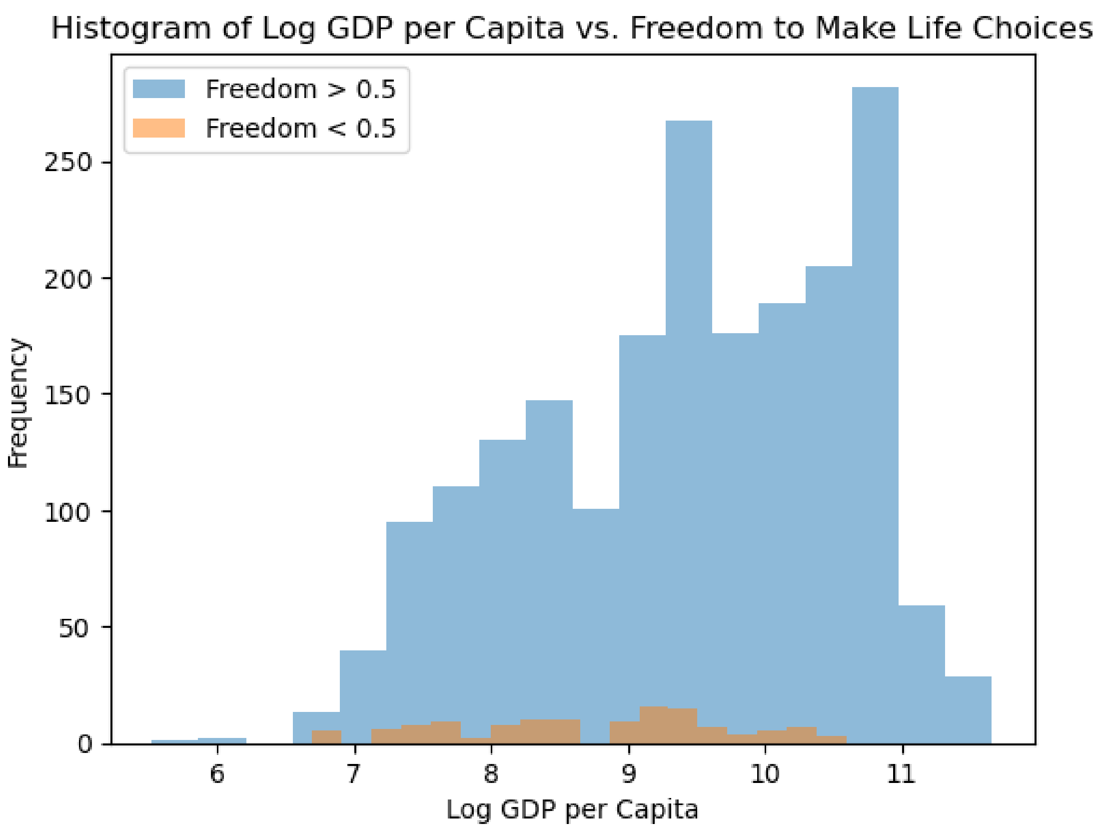


How many countries are more satisfied with their ability to choose than dissatisfied?

[123]:
```
whr_data = whr[(whr['Freedom to make life choices'] > 0.5) | (whr['Freedom to␣
↪make life choices'] < 0.5)]
# Data transformation
whr_data['log GDP per capita'] = whr.loc[:,'Log GDP per capita']
# Number of bins for the histogram
num_bins = 18
```

```
# Plotting the histogram
plt.hist(whr_data[whr_data['Freedom to make life choices'] > 0.5]['log GDP per␣
↪capita'], bins=num_bins, alpha=0.5, label='Freedom > 0.5')
plt.hist(whr_data[whr_data['Freedom to make life choices'] < 0.5]['log GDP per␣
↪capita'], bins=num_bins, alpha=0.5, label='Freedom < 0.5')
plt.xlabel('Log GDP per Capita')
plt.ylabel('Frequency')
plt.legend()
plt.title('Histogram of Log GDP per Capita vs. Freedom to Make Life Choices')
plt.show()
```

```
/tmp/ipykernel_231/1359861103.py:3: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-
docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  whr_data['log GDP per capita'] = whr.loc[:,'Log GDP per capita']
```



From this histogram we can see that the amount of countries that are more disatisfied with their ability to make life choices in significantly less than countries whose citizens are more satisfied with their ability to make life choices. There are 42 countries that are more dissatisfied, including Afghanistan, Cuba and Greece. There are 165 countries included in this data set, meaning that 25.45% of all countries included in the study where found to have at least one year from 2005 to 2022 where the citizens believed they couldn't make choices regarding their future. The GDP per capita does not have a significant effect on whether a countries citizens believe they have the Freedom to make life choices, as seen by the graph. The freedom level is spread along the axis pertaining to GDP per capita, indicating that GDP per capita is not the main reason citizens believe they do not have autonomy.

[124]:
```
sub = whr[whr['Freedom to make life choices'] < 0.5]['Country name'].nunique()
tot = whr['Country name'].nunique()
subtot = sub/tot
subtot
```

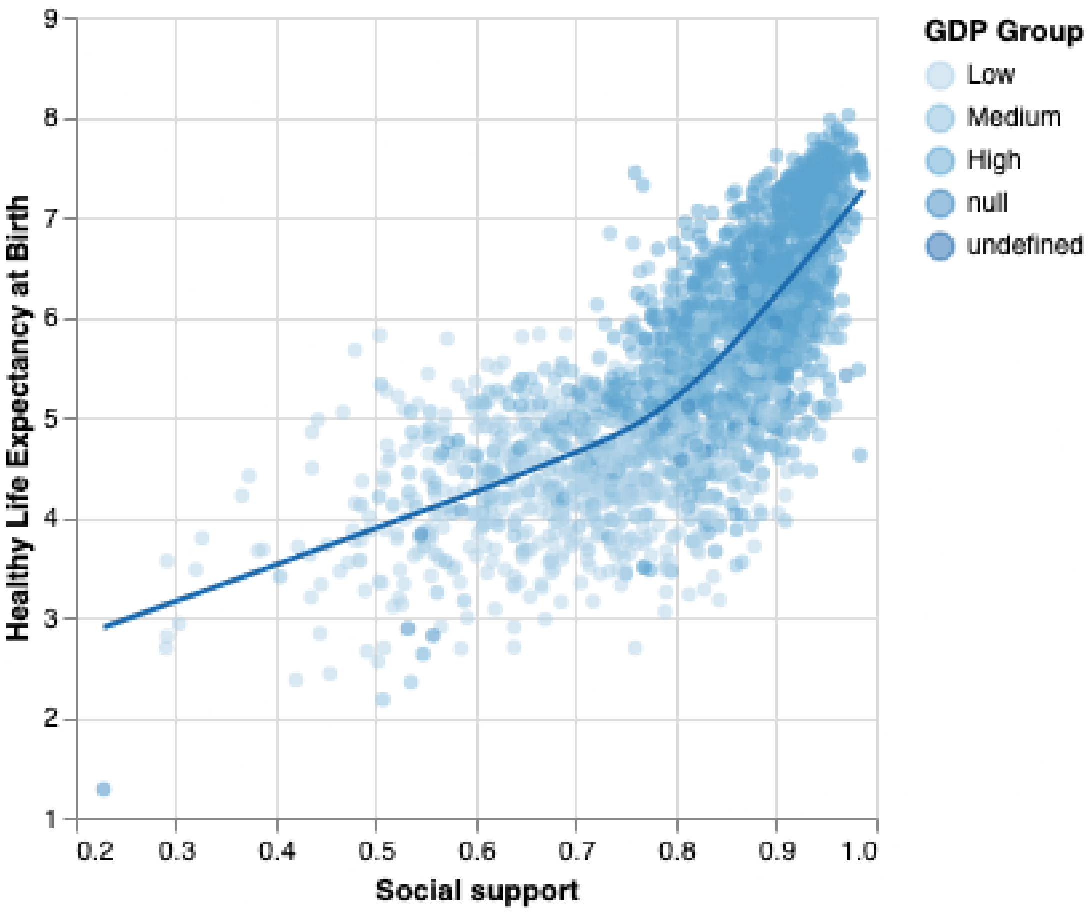
[124]: 0.2545454545454545

The freedom to make life choices is significantly less during the year 2008, since the recession made it difficult for people to climb the ladder and make a better path for themselves. The score is less in every country studied. The recession took away an individual's ability to make life choices and improve their situations. People felt trapped in their lives, indicating that the fall in GDP per capita during the recession had a negative influence on citizens.

How does Social Support correlate with Healthy Life Expectancy at Birth scores? Here is a scatterplot to analyze the two variables:

[135]:
```
whr['GDP Group'] = whr['GDP Group'].dropna()
scatter = alt.Chart(whr).mark_circle(opacity = 0.5).encode(
    x = alt.X('Social support', scale = alt.Scale(zero = False)),
    y = alt.Y('Life Ladder', title = 'Healthy Life Expectancy at Birth', scale␣
 ↪= alt.Scale(zero = False)),
    color = 'GDP Group',
)

# compute smooth
smooth = scatter.transform_loess(
    on = 'Social support',
    loess = 'Life Ladder',
    bandwidth = 0.8
).mark_line(color = 'black')

display(scatter + smooth)
#had problems displaying chart in PDF, saved image from Jupyter notebooks,␣
 ↪uploaded it, then posted it in markdown cell
```

alt.LayerChart(…)



Lower social support values are recorded with lower GDP per capita values, indicating that there is a weak social support system in these countries. Citizens of these countries feel more separated, making them less likely to lean on each other for help. As the GDP per capita value increases, the social support value and healthy life expectancy at birth tends to increase. When there is more support from citizens in these countries, people tend to live longer, healthier lives. This is caused by an increase in GDP. With higher social support values we can deduce that the individuals in these countries feel supported by their family, friends and community. With stronger social relationships and bonds the individuals in these countries are happier on average. The happiest countries are those that are in the upper quantile according to GDP per capita.

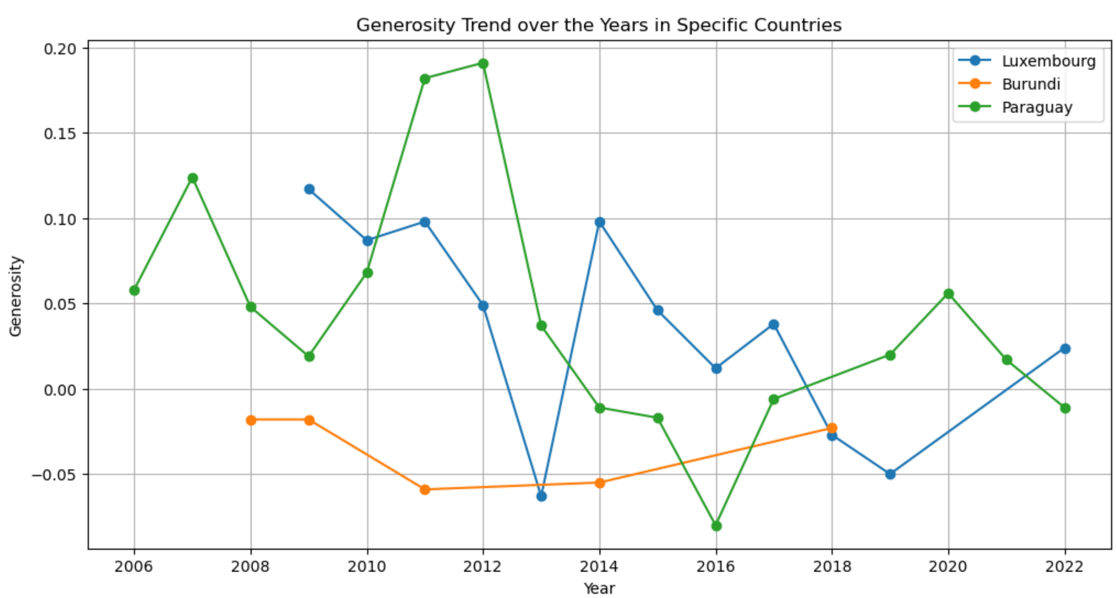
How has Generosity changed over time in specific countries? Lets create a line plot or bar chart to visualize the trend of generosity over the years for selected countries.

[110]:
```
whr_time_mean = whr.drop(columns = [ 'year', 'GDP Group']).groupby(['Country␣
↪name'],
                        as_index = False).mean().sort_values(by='Log GDP per␣
↪capita', ascending=False)
top_gen_df = whr_time_mean.head(1)[['Country name', 'Generosity']]

whr_time_mean = whr.drop(columns = [ 'year', 'GDP Group']).groupby(['Country␣
↪name'],
```

```
                              as_index = False).mean().sort_values(by='Log GDP per␣
 ↪capita')
bot_gen_df = whr_time_mean.head(1)[['Country name', 'Generosity']]

length = len(whr_time_mean)
mid_ind = length // 2
mid = [whr_time_mean['Country name'].iloc[mid_ind]]
top = top_gen_df['Country name'].tolist()
bot = bot_gen_df['Country name'].tolist()
```

How has Generosity changed over time in specific countries? Create a line plot or bar chart to visualize the trend of generosity over the years for selected countries.

[111]:
```
countries = top + bot + mid
country_data = whr[whr['Country name'].isin(countries)]

# Create the line plot
plt.figure(figsize=(12, 6))  # Set the size of the plot

# Loop through each country and plot its Generosity over the years
for country in countries:
    NEWcountry_data = country_data[country_data['Country name'] == country]
    plt.plot(NEWcountry_data['year'], NEWcountry_data['Generosity'],␣
 ↪marker='o', label=country)
plt.xlabel('Year')
plt.ylabel('Generosity')
plt.title('Generosity Trend over the Years in Specific Countries')
plt.legend()
plt.grid(True)
plt.show()
```



We selected a country with the highest income, middle income and lowest income from the data. Luxembourg has the highest income on average over the years and according to the graph, we can see that there is significant variability from year to year. Similarly, Paraguay, the middle tier country has notable variability. Both of these countries have sufficient income in order to be generous and choose how much they should give. The lowest income country, Burundi, has restricted variability as indicated by the relatively flat generosity line. This country cannot choose to be generous, since the citizens have a low average GDP per capita. A higher GDP per capita allows citizens of different countries to choose their level of generosity. With a higher generosity score, citizens of these countries can live happier knowing they have enough that they can willingly support each other as well as themselves comfortably.

Is there a correlation between a country's GDP and its life ladder score (life satisfaction or happiness)?

[112]:
```
# Calculate the correlation coefficient between "Log GDP per capita" and "Life␣
↪Ladder" scores
correlation_coefficient = whr['Log GDP per capita'].corr(whr['Life Ladder'])

# Check the correlation coefficient and print the result
print("The Correlation coefficient between Log GDP per capita and Life Ladder␣
↪scores is ", correlation_coefficient)

# Determine if wealthier countries are generally happier based on the␣
↪correlation
if correlation_coefficient > 0:
    print("This shows that there is a positive correlation between Log GDP per␣
↪capita and Life Ladder scores.")
    print("This indicates that Wealthier countries tend to be happier.")
elif correlation_coefficient < 0:
    print("There is a negative correlation between Log GDP per capita and Life␣
↪Ladder scores.")
    print("Wealthier countries tend to be less happy.")
else:
    print("There is no significant correlation between Log GDP per capita and␣
↪Life Ladder scores.")
    print("Wealthier countries do not necessarily have higher or lower Life␣
↪Ladder scores.")


# Define the GDP groups based on quartiles
gdp_low = whr['Log GDP per capita'].quantile(0.25)
gdp_medium = whr['Log GDP per capita'].quantile(0.5)
gdp_high = whr['Log GDP per capita'].quantile(0.75)

# This Creates a new column to categorize countries based on GDP groups
```

```
whr['GDP Group'] = pd.cut(whr['Log GDP per capita'], bins=[float('-inf'),␣
↪gdp_low, gdp_medium, float('inf')], labels=['Low', 'Medium', 'High'])

# Calculate the average life ladder score for each GDP group
average_life_ladder_by_gdp = whr.groupby('GDP Group')['Life Ladder'].mean()

# Print the result
print(average_life_ladder_by_gdp)
```

```
The Correlation coefficient between Log GDP per capita and Life Ladder scores is
0.7848684422556059
This shows that there is a positive correlation between Log GDP per capita and
Life Ladder scores.
This indicates that Wealthier countries tend to be happier.
GDP Group
Low       4.335321
Medium    5.177165
High      6.214879
Name: Life Ladder, dtype: float64
```

Which countries have the highest life ladder scores, and do they correlate with high GDP values?

[113]:
```
# Sort the DataFrame by "Life Ladder" scores in descending order
whr_time_mean = whr.drop(columns = [ 'year', 'GDP Group']).groupby(['Country␣
↪name'],
                        as_index = False).mean().sort_values(by='Life Ladder',␣
↪ascending=False)

#print(whr_time_mean.shape)
#whr_time_mean.dropna(subset = ['Life Ladder', 'Log GDP per capita'], inplace =␣
↪True)
#print(whr_time_mean.shape)

 #Display the top countries with the highest "Life Ladder" scores and their␣
↪corresponding "Log GDP per capita" values
top_countries = 5
top_countries_df = whr_time_mean.head(top_countries)[['Country name', 'Life␣
↪Ladder', 'Log GDP per capita']]
print("Countries with the highest Life Ladder scores:")
print(top_countries_df)
```

```
Countries with the highest Life Ladder scores:
    Country name  Life Ladder  Log GDP per capita
39       Denmark     7.673529           10.890588
48       Finland     7.619067           10.758267
111       Norway     7.481750           11.063583
142  Switzerland     7.474583           11.134583
63       Iceland     7.458600           10.882200
```

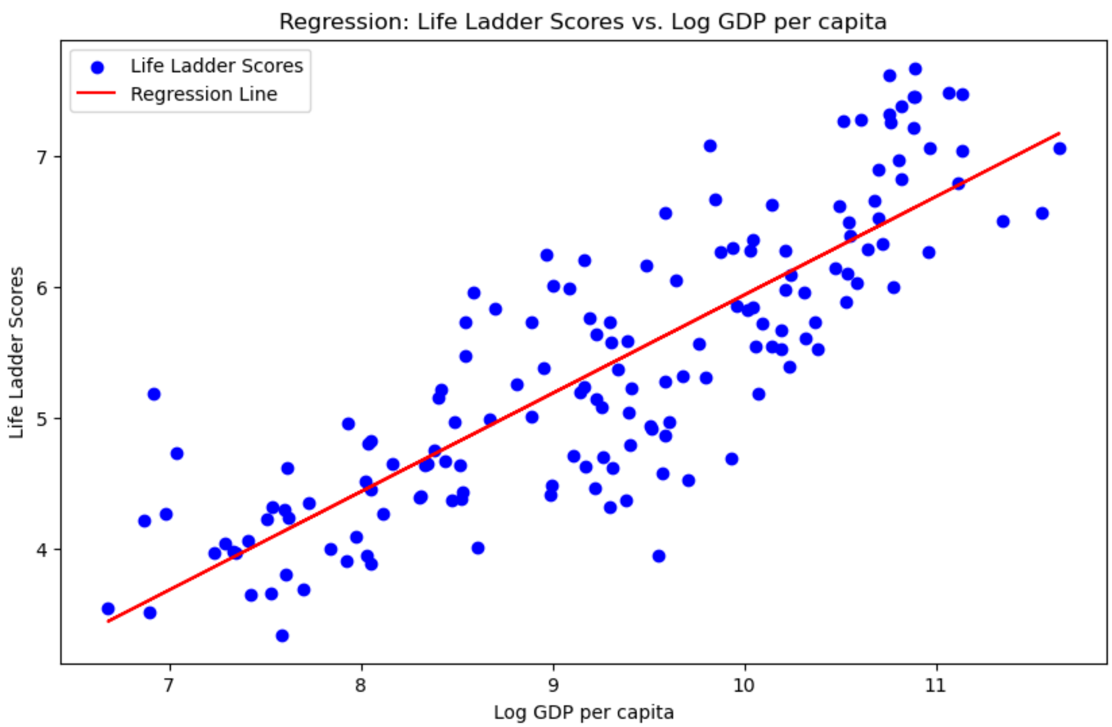
Yes using the defintion of high gdp from the previous question, the countries with the top 5 highest life ladder scores have GDPs that well exceed the high GDP

Is there any evidence of an inflection point or a threshold beyond which increasing GDP does not significantly impact life ladder scores?

[114]:
```python
whr_time_mean = whr_time_mean.dropna()

X = sm.add_constant(whr_time_mean['Log GDP per capita'])
y = whr_time_mean['Life Ladder']

# Fit the regression model
model = sm.OLS(y, X).fit()

# Print the summary of the regression results
print(model.summary())

# Plot the regression line
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 6))
#plt.scatter(whr_cleaned['Log GDP per capita'], y, label='Life Ladder Scores',␣
↪color='blue')
#plt.plot(whr_cleaned['Log GDP per capita'], model.fittedvalues,␣
↪label='Regression Line', color='red')

plt.scatter(whr_time_mean['Log GDP per capita'], y, label='Life Ladder Scores',␣
↪color='blue')
plt.plot(whr_time_mean['Log GDP per capita'], model.fittedvalues,␣
↪label='Regression Line', color='red')

plt.xlabel('Log GDP per capita')
plt.ylabel('Life Ladder Scores')
plt.legend()
plt.title('Regression: Life Ladder Scores vs. Log GDP per capita')
plt.show()
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:            Life Ladder   R-squared:                       0.707
Model:                            OLS   Adj. R-squared:                  0.705
Method:                 Least Squares   F-statistic:                     371.1
Date:                Sun, 06 Aug 2023   Prob (F-statistic):           7.32e-43
Time:                        01:27:34   Log-Likelihood:                -135.42
No. Observations:                 156   AIC:                             274.8
Df Residuals:                     154   BIC:                             280.9
Df Model:                           1
Covariance Type:            nonrobust
```

```
==============================================================================
======
                        coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
------
const                  -1.5681      0.365     -4.294      0.000      -2.290
-0.847
Log GDP per capita      0.7508      0.039     19.263      0.000       0.674
0.828
==============================================================================
Omnibus:                        0.673   Durbin-Watson:                   1.419
Prob(Omnibus):                  0.714   Jarque-Bera (JB):                0.803
Skew:                           0.103   Prob(JB):                        0.669
Kurtosis:                       2.715   Cond. No.                         74.5
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```



Here We calculated the Residual Sum of Squares (RSS) for each model and select the breakpoint that minimizes the RSS. Finally, we ploted the segmented regression lines along with the data to visualize the relationship between GDP and life ladder scores in different segments.

Based on the OLS regression results provided, we can analyze the relationship between the dependent variable "Life Ladder" and the independent variable "Log GDP per capita." Here are some key observations:

R-squared: The R-squared value is 0.707, which indicates that approximately 70.7% of the variation in the 'Life Ladder' scores can be explained by the variation in the 'Log GDP per capita'. A higher R-squared value suggests a better fit of the model to the data.

Adjusted R-squared: The adjusted R-squared value is 0.705. It is similar to the R-squared but takes into account the number of predictors in the model. It penalizes the R-squared value for having more predictors and provides a more reliable measure of the model's goodness of fit.

F-statistic: The F-statistic is 371.1, and its associated p-value (Prob (F-statistic)) is very close to zero (7.32e-43). This suggests that the overall model is statistically significant, meaning that there is a relationship between the dependent and independent variables.

Coefficients: The coefficients of the regression model represent the estimated impact of the independent variable 'Log GDP per capita' on the dependent variable 'Life Ladder'. The constant term (intercept) is -1.5681, and the coefficient for 'Log GDP per capita' is 0.7508. The intercept is the expected 'Life Ladder' score when 'Log GDP per capita' is zero.

Standard Errors: The standard errors associated with the coefficients measure the variability of the estimates. Smaller standard errors indicate more precise estimates.

t-statistics and P-values: The t-statistics measure the significance of each coefficient. The P-values (P>|t|) provide the probability of observing a t-statistic as extreme as the one computed if the null hypothesis (no relationship between the variables) were true. In this case, both the constant and 'Log GDP per capita' have extremely low P-values, indicating that they are highly significant in predicting 'Life Ladder' scores.

Confidence Intervals: The 95% confidence intervals provide a range of values within which we can be 95% confident that the true population coefficients lie.

The positive coefficient of 0.7508 for 'Log GDP per capita' indicates that as the 'Log GDP per capita' increases, the 'Life Ladder' scores also tend to increase. This suggests a positive relationship between a country's economic prosperity (GDP per capita) and the subjective well-being or happiness (Life Ladder scores) of its citizens.

Are there any outliers where countries with lower GDP levels have unexpectedly high life ladder scores or vice versa?

[115]:
```
low_gdp_countries = whr_time_mean.loc[whr_time_mean['Log GDP per capita'] <=␣
  ↪gdp_low]
low_gdp_life_ladder_outliers = low_gdp_countries.loc[low_gdp_countries['Life␣
  ↪Ladder'] > average_life_ladder_by_gdp[1]]['Country name']
#there are only 2 countries that are the outliers
print("The only 2 countries that are the outliers are :",␣
  ↪low_gdp_life_ladder_outliers)



high_gdp_countries = whr_time_mean.loc[whr_time_mean['Log GDP per capita'] >=␣
  ↪gdp_high]
```

```
high_gdp_life_ladder_outliers = high_gdp_countries.loc[high_gdp_countries['Life␣
↪Ladder'] < average_life_ladder_by_gdp[1]]
#there are no high or very high gdp countries that are above the 50th percentile
print("There are no high or very high GDP countries that are above the 50th␣
↪percentile")
```

```
The only 2 countries that are the outliers are : 79     Kyrgyzstan
131       Somalia
Name: Country name, dtype: object
There are no high or very high GDP countries that are above the 50th percentile
```

Conclusion A complex question yields a complex answer. Our study sought to explore the relationship between a nation's wealth (Log GDP per capita) and the reported happiness of its citizens. The analysis revealed a complex: while there was a positive correlation between GDP and our own Net Positive Affect variable, the significant variance suggested other factors at play. Economic conditions, exemplified by the global downturn in 2008, impacted 'Freedom to make life choices', indicating the indirect influence of wealth on happiness. A trend was observed with higher GDP yielding increased social support and healthier life expectancy, signifying economic prosperity's contribution to social cohesion and improved health. Moreover, a nation's capacity for generosity was seen to correlate with its economic standing. The life ladder variable also had a noticable correlation, with a positive regression coefficient of 0.7508 for 'Log GDP per capita'. In conclusion, wealth, while not a standalone determinant, appears to influence conditions conducive to happiness, requiring further research to isolate its exact impact and interplay with other factors.