

Analyzing GDP, Happiness and Life satisfaction among countries

Elmar Hajizada

1 Task 1.1

In this section, we will analyze the life expectancy and real GDP of countries using data brought from the website "The world in data". To start, we downloaded the data as a text file with comma separated values. Then, we used the method `read_csv` from the Python package Pandas to turn the data into a data frame called `df`. Before we started with any other parts of the assignment, we first cleaned `df` by first using filter in Pandas to only keep the columns containing the country, year, life expectancy and GDP. Then, we removed all rows where either the life expectancy or the GDP did not have any data.

After we've cleaned the data, we started to look at all countries with life expectancy one standard deviation above the mean. We first assumed that we would only look at the data from one year, in our case 2016. Therefore, we made a method to filter the data frame using a boolean expression to only keep entries from one year and had it keep the entries from the year 2016, the highest year that contained any data using the following code:

```
def filterYear(frame, year):  
    return frame[frame['Year']==str(year)]
```

Then, we used the package numpy to calculate the mean and standard deviation of the life expectancy and used another boolean expression similar to the filter from before to keep all entries with life expectancy above the mean plus standard deviation, and plotted the life expectancy using a bar graph.

Here we can see that a fairly large number of countries have a high life expectancy, especially from european countries, but also others like Israel, Australia, and Singapore.

In the next task, we will analyze and determine which countries have high life expectancy but low GDP. So while analyzing data as high life expectancy, we focused on the countries whose life expectancy is above one standard deviation from mean, on the other hand, by saying low GDP we tried to focus on countries whose GDP is under average. But after trying to apply this conditions on our dataframe we found out that, there is no such a country whose life expectancy is above one standard deviation from mean and GDP is below average. So that's why by saying low GDP we assumed that we can look for the countries whose GDP is the range between mean and one standard deviation from mean. We used scatter plot to show the countries that fulfilled these both conditions.

```
def question_1b(df):  
    lf_expect_mean = df.life_expectancy.mean()  
    lf_expect_std = df.life_expectancy.std()  
    gdp_mean = df.gdp.mean()  
    gdp_std = df.gdp.std()  
    df = df.loc[(new_df['life_expectancy'] > (lf_expect_mean + lf_expect_std))  
                & (df['gdp'] < (gdp_mean + gdp_std))]
```

So as a result of scatter plot we can definitely say that for example Japan, whose life expectancy is quite high, has GDP which is around mean. So its GDP is in between mean and standard deviation + mean. So all of them approximately has a life expectancy which is higher than 80, however we can say from this

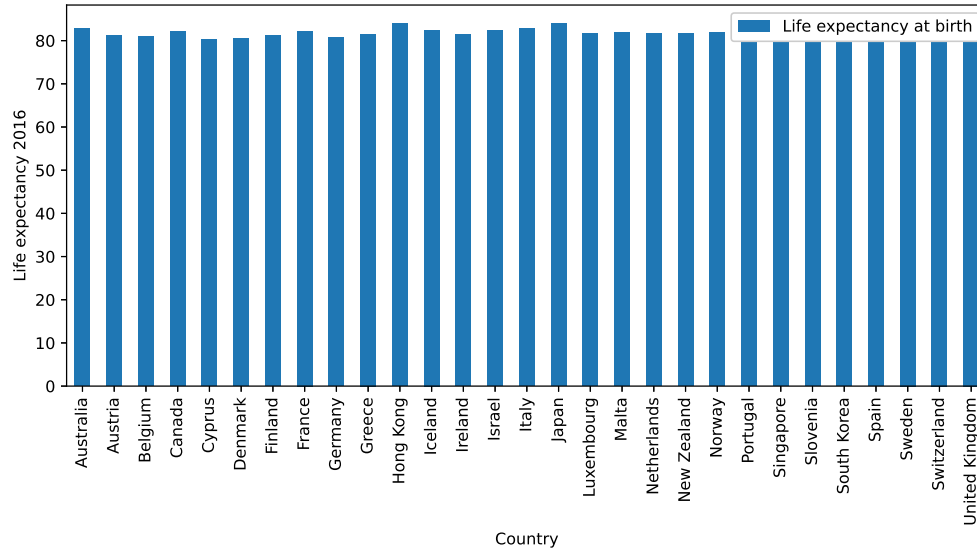


Figure 1: Bar graph of all countries with life expectancy one standard deviation above the mean

scatter plot that having high life expectancy does not always mean that GDP of that country will also be higher.

In the next task, we are supposed to show that if every strong economy has high life expectancy. By saying strong economy we assume that GDP of the country should be one standard deviation above mean. In this task we will find out if there are countries that has high GDP but low life expectancy, so that we can say that the assumption "every strong economy have high life expectancy" is not always right. So we focused on the countries which has high GDP means one standard deviation above from mean and also low life expectancy meaning under average. But again while analyzing the data we found out that there is no such country whose GDP can be one standard deviation above from mean but having life expectancy below average. That's why by saying low life expectancy we focused on countries whose life expectancy is in between mean and standard deviation + mean. Again we have used scatter plot to show the countries that fulfilled these both conditions.

```
def question_1c(df):
    lf_expect_mean = df.life_expectancy.mean()
    lf_expect_std = df.life_expectancy.std()
    gdp_mean = df.gdp.mean()
    gdp_std = df.gdp.std()
    df = df.loc[(df['life_expectancy'] < (lf_expect_mean + lf_expect_std))
                & (df['gdp'] > (gdp_mean + gdp_std))]
```

As we can see from the scatter plot there are 7 countries whose GDP is above std + mean but life expectancy is in between mean and std + mean. We can say that having a strong economy does not always mean that there will be high life expectancy in that country. Life expectancy in these countries which are shown in scatter plot, is in between mean and one standard deviation above from mean.

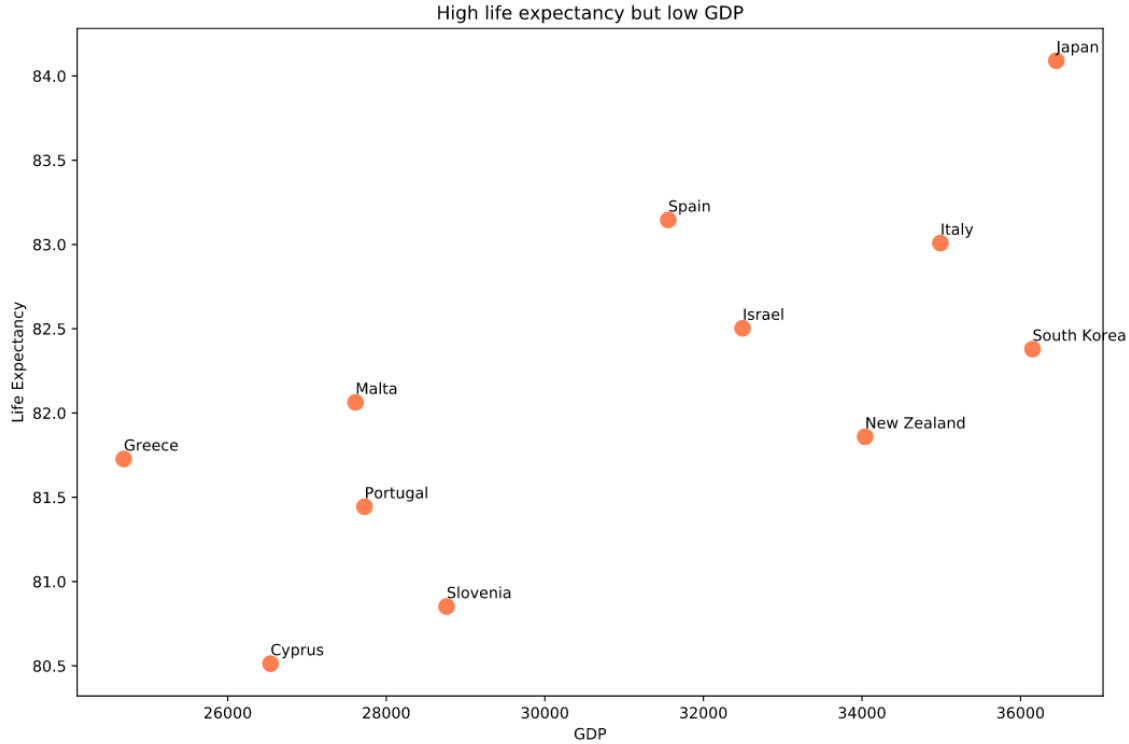


Figure 2: Countries that has high Life expectancy but low GDP

2 Task 1.2

For the second part of the assignment, we downloaded new data from "The world in data", this time a comparison between happiness and life satisfaction. Much like the previous task, we turned the data into a data frame and cleaned it on unnecessary data and values without numbers. First, we looked at the data from the year 2014 to see if there was anything interesting in the data. We first looked at the countries with high happiness, once again using the assumption that having high happiness means being one standard deviation above the mean, using similar code to the previous task. The results can be found in figure 4, where we can see four countries having high happiness, where over 90% of the population are happy for all countries.

Similarly, we can look at countries which have lower happiness than one standard deviation below the mean. These countries can be found in figure 5 where we see that most are slightly below 70%, except Egypt which only has a happiness of 30%. This means that Egypt is an outlier compared to the rest of the countries.

After looking at which countries had the highest and the lowest happiness, we decided to see if high satisfaction would lead to high happiness and vice versa, once again using similar methods as in Task 1.1. Through this, we found that there while there were no countries with high satisfaction that had happiness below average, there were countries that did not have a high degree of happiness. These countries can be found in figure 6.

When we tried to look at countries with high happiness but without high satisfaction, we get the results in figure 7. There, we see that of the four countries with high happiness from before, only Sweden also has a high life satisfaction. In fact, Kyrgyzstan has a life satisfaction below average when we tried to see if one of the happy countries was below the mean in satisfaction.

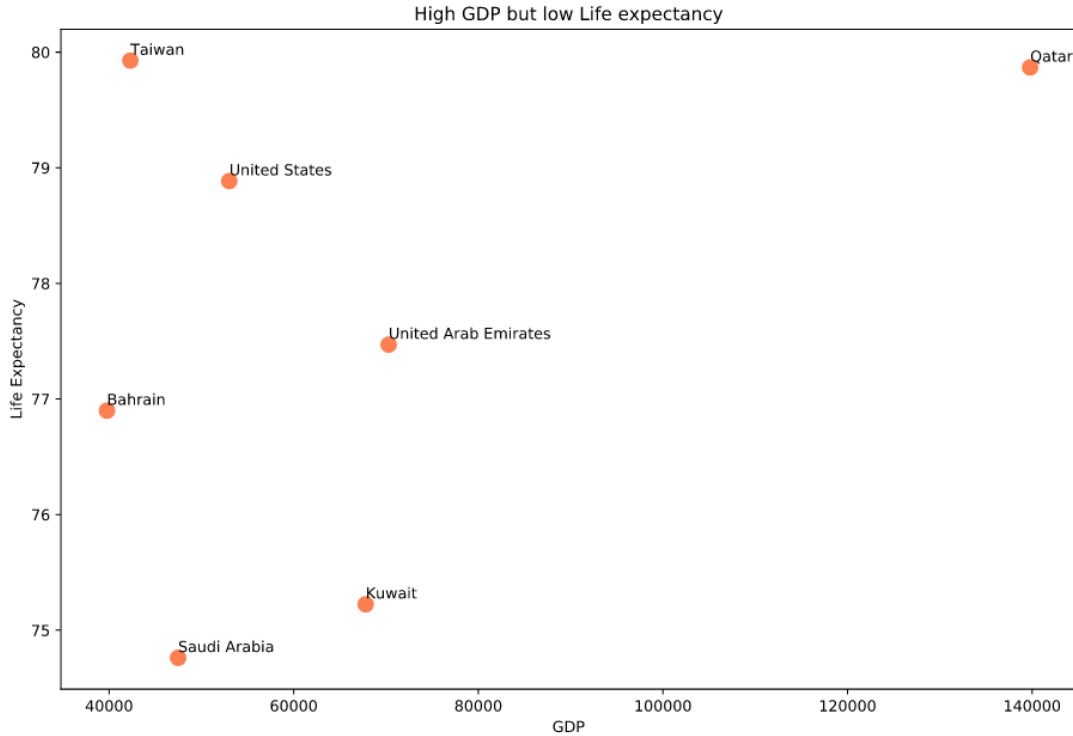


Figure 3: Countries that has high GDP but low Life expectancy

In addition to this analyses, we also tried to find out which countries are in top 10 according to life satisfaction both in year 2009 and 2014. So by this means, we can say that which countries kept their life satisfaction from the year 2009 till 2014.

Furthermore, we also decided to look at the countries which had dramatic changes based on life satisfaction from 2009 to 2014. So we extracted the countries whose life satisfaction was above one standard from mean in 2009 and became under average in 2014. On the other hand, countries, whose life satisfaction was under average in 2009 but increased dramatically in till 2014, by meaning that it became above one standard deviation from the mean, have been extracted from the data.

```
def df1 = df1.sort_values('Life_satisfaction', ascending=False)
df2 = df2.sort_values('Life_satisfaction', ascending=False)
mean = df1['Life_satisfaction'].mean()
std = df2['Life_satisfaction'].std()
increased_life_satisfaction = pd.merge(df1.loc[(df1['Life_satisfaction'] < mean)], df2.loc[(df2 <=
    ['Life_satisfaction'] > mean + std)], how='inner', on=['Country'])
decreased_life_satisfaction = pd.merge(df1.loc[(df1['Life_satisfaction'] > mean + std)], df2.loc[(df2 <=
    df2['Life_satisfaction'] < mean)], how='inner', on=['Country'])
df = pd.concat([increased_life_satisfaction, decreased_life_satisfaction])
```

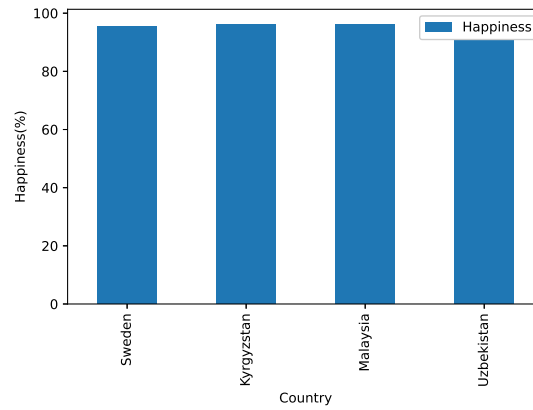


Figure 4: Bar graph of all countries in the year 2014 with high happiness sorted from lowest to highest

As can be seen from bar plot, we can say that Life satisfaction in Thailand peaked up from 5.476 to 6.985 between the years 2009 and 2014. However, in Cyprus it became worse, since life satisfaction dramatically decreased from 6.833 to 5.627 in the years between 2009 and 2014.

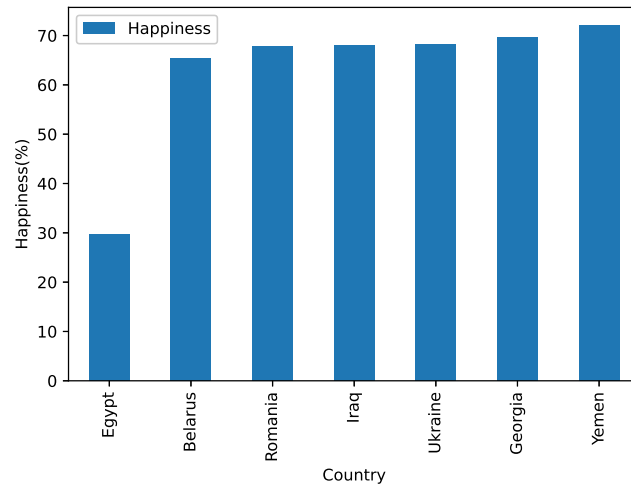


Figure 5: Bar graph of all countries in the year 2014 with really low happiness sorted from lowest to highest

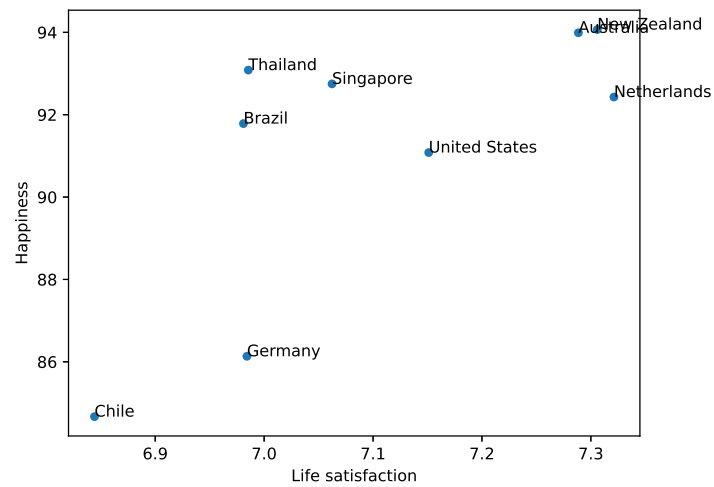


Figure 6: Scatter plot of all countries that have high satisfaction but do not have high happiness

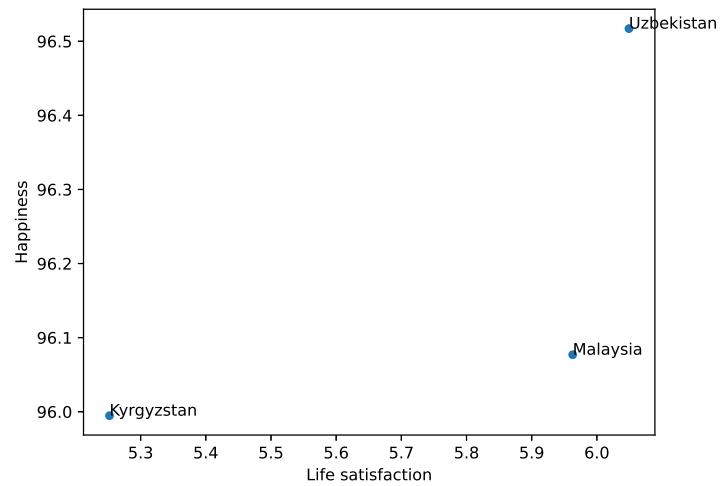


Figure 7: Scatter plot of all countries that have high happiness but do not have high satisfaction

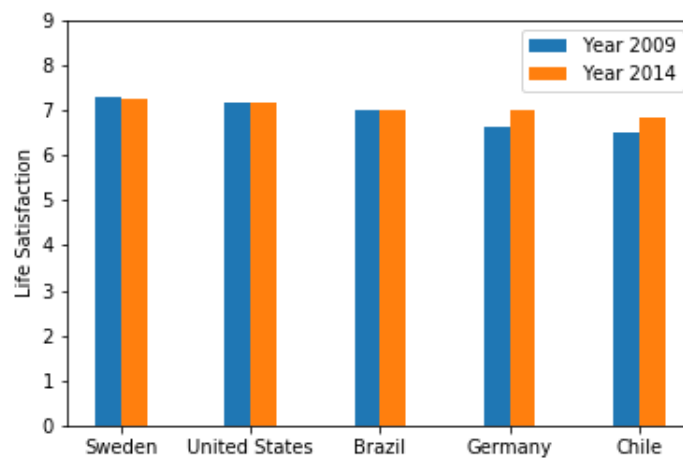


Figure 8: Bar graph of countries which is in top 10 among all countries both in year 2009 and 2014

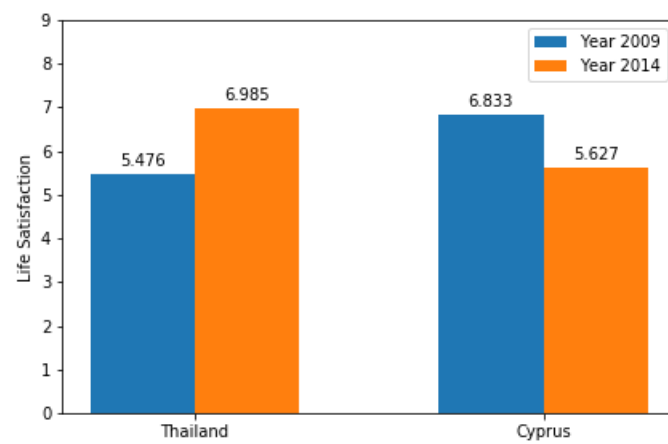


Figure 9: Countries where life satisfaction changed dramatically by comparing year 2009 and 2014