

Assignment 1
DIT852: Introduction to Data Science

1.

- A) To draw a scatter plot that illustrates if there is a correlation between GDP per capita and life expectancy, world data is downloaded from www.ourworldindata.org. The data of interest is, except from GDP and life expectancy, also year and countries since one would like to know if a country with a certain GDP value can indicate something else. Before observing the data, assumptions of that there is a relationship between rich countries and life expectancy were taken, the higher GDP per capita in a country the higher is the life expectancy there. However, when selecting the data, it was of high importance that all of the columns for each variable were not missing any values in order to take fair conclusions of the data when comparing. Moreover, extra data that is not relevant for the analysis such as information on world continents was excluded due to the purpose of narrowing the data as much as required.

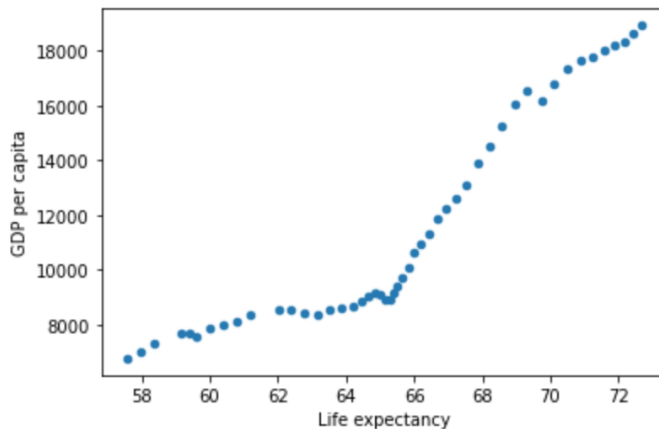


Figure 1: scatter plot of GDP per capita and life expectancy

Figure 1 illustrates the relationship between GDP per capita and life expectancy throughout the data for all the countries based on the yearly mean values. The observation from this is that lower GDP per capita generates low life expectancy in a country. Also, it is observed that these values increase over the years as GDP per capita and life expectancy becomes higher with time.

- B) It is shown on Figure 1 that there exists a positive correlation between GDP per capita and life expectancy which was assumed in the beginning, before executing the scatter plot. The relationship between these two variables makes sense since a country that has higher GDP per capita usually has a greater availability of medical help, and if not, the majority of the citizens in that country could afford to buy it which is a factor that directly affects life expectancy. Meanwhile, a poorer country with a low GDP per capita usually has a lower education level and thereby a less successful health care system.
- C) Values from 1969 and before were decided to be dropped as latter data is more interesting but also, to avoid missing data that is not measured. Moreover, extra data that is not relevant for the analysis such as information on world continents was excluded due to the purpose of narrowing the data as much as required. Hence, the data is based on values from 1970 to 2018 in all of the different world countries.

D)

Country	Life expectancy	GDP per capita
Australia	77,82	32 669
Canada	77,89	32 684
Cyprus	76,98	17 548
France	77,56	29 356
Greece	77,02	18 572
Hong Kong	78,61	29 496
Iceland	78,79	29 945
Israel	77,51	22 620
Italy	77,84	27 248
Japan	79,40	28 863
Malta	77,18	16 027
Netherlands	77,77	32 233
New Zealand	76,71	25 247
Norway	77,96	46 060
Spain	78,12	21 891
Sweden	78,51	31 358
Switzerland	78,58	39 590
United Kingdom	76,62	27 933

Table 1

The countries stated in Table 1 have a life expectancy higher than one standard deviation above the mean.

- E) For this question, we defined that a high life expectancy is 70 years and above. In Table 1 it is presented that the following countries: Cyprus, Greece and Malta has a low GDP per capita but yet, a high life expectancy.

F)

Country	Life expectancy	GDP per capita
Oman	67,38	20 285
Trinidad and Tobago	68,85	20 346
Slovenia	74,76	21 168
Taiwan	74,73	21 477
Spain	78,12	21 891
Puerto Rico	75,50	22 169
Israel	77,51	22 620
New Zealand	76,71	25 247
Saudi Arabia	68,27	26 296
Italy	77,84	27 248
United Kingdom	76,62	27 933
Finland	76,37	28 065
Japan	79,40	28 863
Belgium	76,49	29 350
France	77,56	29 356
Hong Kong	78,61	29 496
Ireland	75,98	29 784
Iceland	78,79	29 945
Austria	76,42	29 950
Germany	76,33	30 281
Sweden	78,51	31 358
Netherlands	77,77	32 233
Singapore	76,47	32 246
Australia	77,82	32 669
Canada	77,89	32 684
Denmark	76,35	33 404
United States	75,77	39 570
Switzerland	78,58	39 590
Kuwait	71,89	39 787
Luxembourg	76,03	39 946
United Arab Emirates	71,93	46 023
Norway	77,96	46 060
Qatar	75,76	64 226

Table 2

Strong economies, in this assignment we choose a GDP per capita at 20 000 USD and above, tend to have a population with a high life expectancy however there is still some exceptions. Interestingly, Oman, Trinidad and Tobago and Saudi Arabia have high values on their GDP per capita and still a low life expectancy (Table 2). What these

three countries have in common is that oil and natural gas are their largest export revenues which explains their high GDP values. Therefore, not every strong economy has a high life expectancy.

G) GDP per capita is a global measurement of a country's economic situation however there are some weaknesses with this metric. Only products and services on the official market are counted and for instance, not work that is done in homes nor trades on the black market (which could be dominating in some countries). Hence, the measurement can give a skewed image of a nation's economy. Additionally, the amount of population in a country has an impact on the value since a higher number of inhabitants results in a lower GDP per capita. This is why smaller countries such as Switzerland and Luxembourg, usually have a high GDP per capita. Oman and Trinidad and Tobago have also lower amounts of inhabitants which, likewise, is beneficial for them when counting their GDP per capita. Conversely, Saudi Arabia has 34 million inhabitants but yet, a high GDP per capita due to the successful business within their oil industry. As a conclusion of what's been discussed and seen from Table 2, one should not use GDP per capita to indicate if a country has a strong economy without considering the underlying factors of what can be the reason behind their high or low GDP value.

2.

A) For this analysis data of life satisfaction, corruption and GDP per capita will be used from the years 2012 to 2018. The questions that will be examined is if there is any correlation between:

- Life satisfaction and corruption
- Life satisfaction and GDP per capita
- Corruption and GDP per capita

It is interesting to investigate if high corruption indicates low life satisfaction as corruption could have an impact on a country's stability which tends to be related to economic growth. Since analyzing the effects on GDP per capita in the previous questions about life expectancy it would be of interest to also compare how GDP and life satisfaction is related. Furthermore, GDP per capita and the level of a country's corruption would also be interesting to measure in order to observe if a nation's economic growth is affected by how high or low the corruption is perceived to be.

B)

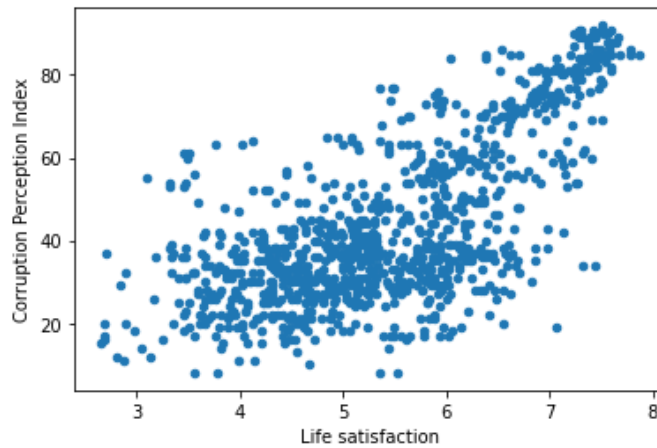


Figure 2: scatter plot of corruption and life satisfaction

Figure 2 illustrates the relationship between corruption and life satisfaction. Corruption is measured as an index between 0 to 100 where 0 is highly corrupt. From the figure it can be assumed that a country with low corruption also indicates high life satisfaction while high corruption gives low life satisfaction. No extreme outliers are found but the observations are widely spread throughout the scatter plot.

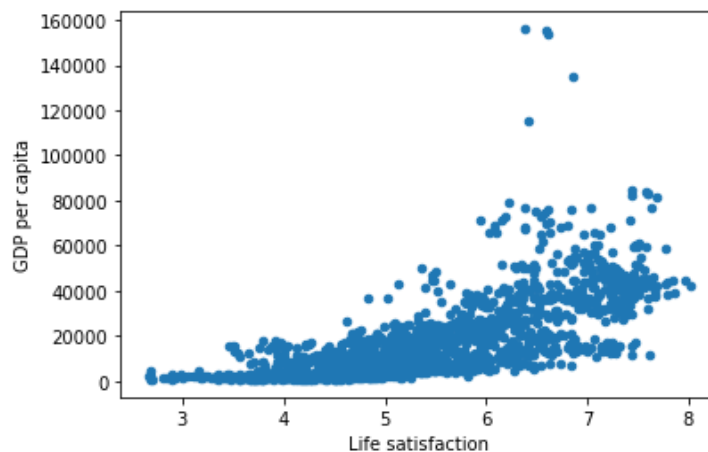


Figure 3: scatter plot of GDP per capita and life satisfaction

When observing Figure 3, one can conclude that a high value of GDP per capita implies a higher life satisfaction which is similar to the relationship between GDP per capita and life expectancy that was analyzed in problem 1. It is plausible that the data indicates that high GDP per capita will improve the economic conditions for people resulting in a better life both in terms of their perceived life satisfaction, and their actual life expectancy.

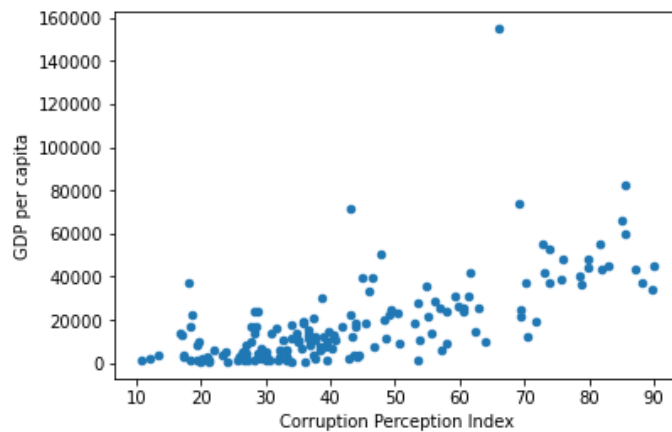


Figure 4: scatter plot of GDP per capita and corruption

When making the plot for Figure 4 mean values for the countries are used instead of presenting all individual observations, and as in Figure 2 corruption is measured as an index between 0 to 100, where 0 is highly corrupt. The scatter plot above illustrates that low corruption usually gives a higher GDP per capita. It is not surprising that a country with high corruption has a lower GDP value as economic growth could be influenced by governance stability. However, there are outliers in Figure 4 that indicates high corruption but yet a high GDP per capita. This can be explained by the fact that there are underlying factors not accounted for in this plot since the GDP measurement has its weaknesses as discussed in question G problem 1.