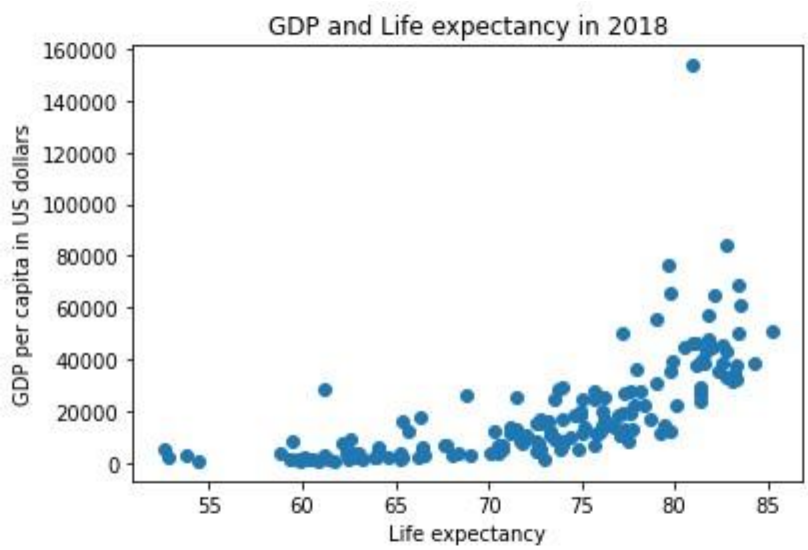
Assignment 1 Nikolay Shivarov

For this assignment I have used three tables with data from the website “Our world in data”. The first one contains the life-expectancy in all countries in the world for different years. The link for this data is: https://ourworldindata.org/grapher/life-expectancy?tab=table . The second one contains the GDP per capita in all countries in the world, again for different years. The link is https://ourworldindata.org/grapher/gdp-per-capita-maddison. Finally, the last one contains the overall GDP of the countries. The link is https://ourworldindata.org/grapher/national-gdp-wb. All the files have the format of “.csv”.

For this assignment I have decided that I should answer the questions, by only using the data from 2018. Firstly, it is not fair to compare countries from different periods. Secondly, what is happening now is more important than what was happening in the past. After doing that I am going to do absolutely the same with the year 1998. Before answering the questions, I had to remove a lot of rows and columns in the tables. From the three tables I removed the rows that had different year from 2018 and I also removed the rows, where the country code had a null value. In this way all the remaigning rows had the year 2018 and were showing a country not a region. From the new tables I only took the name of the country which had a column name Entity in all tables and the numerical values that were important: life-expectancy, GDP per capita, GDP. Then I did a natural join between table1 and table2 and between table1 and table3.

This is the scatter plot for task A:



For task B we are going to print the countries which satisfy our condition:

511	**Andorra**	**82.9923**
1136	**Australia**	**83.3871**

1214	Austria	81.6862
1835	Belgium	81.4787
2051	Bermuda	80.8253
3112	Canada	82.0508
4318	Cyprus	81.3820
4660	Denmark	80.9900
6106	Finland	81.6266
6325	France	82.5909
6469	French Polynesia	82.7834
6757	Germany	81.1717
6911	Gibraltar	82.1540
6983	Greece	81.3907
7214	Guadeloupe	82.5015
7436	Guernsey	82.3065
7946	Hong Kong	85.2456
8208	Iceland	82.7703
8584	Ireland	82.0863
8734	Israel	82.8232
8884	Italy	83.1847
9100	Japan	84.2971
9188	Jersey	80.8284
10642	Liechtenstein	83.1462
10932	Luxembourg	81.8035
11053	Macao	84.9285
11485	Malta	83.3448
11629	Martinique	82.5804
12257	Monaco	86.4643

13150	**Netherlands**	**81.7741**
13296	**New Zealand**	**82.3803**
14161	**Norway**	**82.7605**
15053	**Portugal**	**81.3947**
15197	**Qatar**	**80.8982**
15269	**Reunion**	**81.7661**
16160	**San Marino**	**82.9479**
16763	**Singapore**	**83.4576**
16982	**Slovenia**	**81.3581**
17342	**South Korea**	**83.3427**
17536	**Spain**	**83.1436**
18034	**Sweden**	**82.5326**
18180	**Switzerland**	**83.5615**
19535	**United Kingdom**	**81.1254**

For task C we are going to use GDP per capita. In order to satisfy the conditions a country must be poorer than 55% of countries and have a higher life-expectancy than 60% of countries. Initially, I was considering conditions that were stricter, but there were no countries that satisfied them. Finally, I decided that a country is enough to have a little above median life-expectancy and a little below median GDP per capita. In this way the list of countries is still small, but at least it is not empty.

	Entity \
1	Albania
16	Bosnia and Herzegovina
36	Cuba
44	Ecuador

	Life expectancy (period) at birth - Sex: all - Age: 0	GDP per capita
1	79.1838	11104.1660

16	77.0926	10460.5205
36	77.4962	8325.6310
44	77.0938	10638.8250

For task D we are going to count the number of countries that are in the top 20% for overall GDP. And then we are going to see how many of them are in the top 40% for life expectancy.

The results showed that 39 countries are in the top 20% for overall GDP, but only 25 of them have a high life expectancy. So having a high overall GDP does not guarantee high life expectancy.

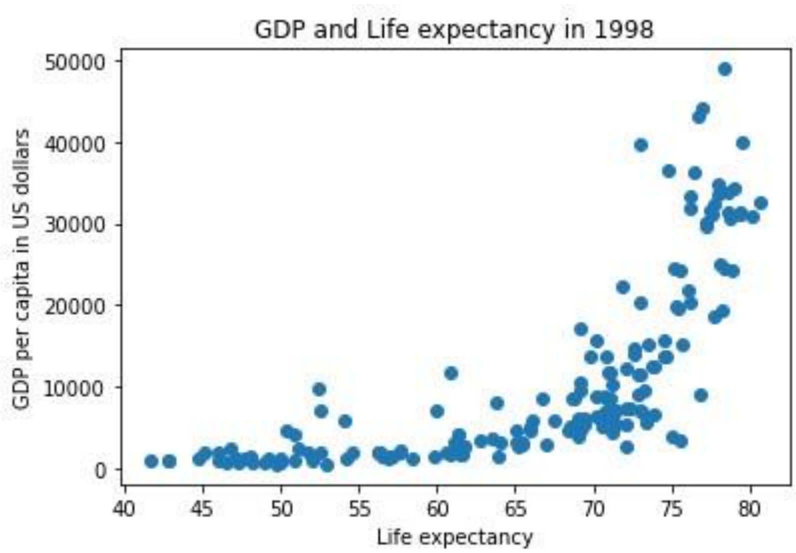
For task E we are going to do the same as in task D, but we are going to use GDP per capita instead of GDP.

This time the results are very different every country that has a high GDP per capita has a high life expectancy.

Obviously, there are countries with big populations that have big economies, but the people living there are not as rich as in other countries. For this reason, we see such different results in the last two tasks. We can conclude that nowadays, generally countries where the average person is rich have a high life expectancy.

Now we will do the same for 1998

This is the scatter plot for task A:



For task B we are going to print the countries which satisfy our condition:

491	Andorra	79.4259
1116	Australia	79.0275
1194	Austria	77.7833
2031	Bermuda	79.9490
3092	Canada	78.6739
6305	France	78.5885
6449	French Polynesia	78.3862
6737	Germany	77.5768
6891	Gibraltar	80.6635
6963	Greece	78.2311
7416	Guernsey	79.4038
7926	Hong Kong	80.1369
8188	Iceland	79.4529
8714	Israel	78.4402
8864	Italy	78.8062
9080	Japan	80.6402
9168	Jersey	78.5510
10622	Liechtenstein	77.8321
11033	Macao	80.3522
11465	Malta	77.7172
12237	Monaco	80.5857
13130	Netherlands	77.9971
13276	New Zealand	78.1725
14141	Norway	78.4083
16140	San Marino	80.0336

16743	**Singapore**	**77.9875**
17516	**Spain**	**78.9428**
18014	**Sweden**	**79.4155**
18160	**Switzerland**	**79.4749**

For task C we are going to use GDP per capita. In order to satisfy the conditions a country must be poorer than 55% of countries and have a higher life-expectancy than 60% of countries. Initially, I was considering conditions that were stricter, but there were no countries that satisfied them. Finally, I decided that a country is enough to have a little above median life-expectancy and a little below median GDP per capita. In this way the list of countries is still small, but at least it is not empty as it was in the 2018 list. Albania and Cuba are in the list again.

	Entity	Life expectancy (period) at birth - Sex: all - Age: 0 \
1	Albania	74.9899
36	Cuba	75.5639
99	Montenegro	73.3662

	GDP per capita
1	3873.3733
36	3415.5452
99	5522.1045

For task D we are going to count the number of countries that are in the top 20% for overall GDP. And then we are going to see how many of them are in the top 40% for life expectancy.

The results showed that 36 countries were in the top 20% for overall GDP, but only 19 of them had a high life expectancy. So having a high overall GDP did not guarantee high life expectancy in 1998.

For task E we are going to do the same as in task D, but we are going to use GDP per capita instead of GDP.

This time the results are very different. Only one country that had a high GDP per capita did not have high life expectancy.

We can conclude that for all the tasks the results in 2018 were very similar to those in 1998.

Here is the python code for 2018:

```
import pandas as pd

import numpy

import matplotlib.pyplot as plt

from pandas.plotting import table


df1 = pd.read_csv("life-expectancy.csv")

df2 = pd.read_csv("gdp-per-capita-maddison.csv")

df3 = pd.read_csv("national-gdp-wb.csv")


pd.set_option("display.max_rows", None, "display.max_columns", None)


df1_2018 = df1[(df1["Code"].notna()) & (df1["Year"] == 2018)]

countryLifeExp2018 = df1_2018[["Entity", "Life expectancy (period) at birth - Sex: all - Age: 0"]]


df2_2018 = df2[(df2["Code"].notna()) & (df2["Year"] == 2018)]

countryGdp2018 = df2_2018 [["Entity", "GDP per capita"]]


df3_2018 = df3[(df3["Code"].notna()) & (df3["Year"] == 2018)]

Gdp2018 = df3_2018 [["Entity", "GDP, PPP (constant 2017 international $)"]]


gdpLe2018 = pd.merge(countryLifeExp2018, countryGdp2018, on="Entity")

gdpLe2018overall = pd.merge(countryLifeExp2018, Gdp2018, on="Entity")


##A

plt.scatter(gdpLe2018["Life expectancy (period) at birth - Sex: all - Age: 0"], gdpLe2018["GDP
per capita"])

plt.title("GDP and Life expectancy in 2018")
```

```
plt.xlabel("Life expectancy")

plt.ylabel("GDP per capita in US dollars")

##B

averageAge = countryLifeExp2018["Life expectancy (period) at birth - Sex: all - Age: 0"].mean()

standardDeviationAge = countryLifeExp2018["Life expectancy (period) at birth - Sex: all - Age: 0"].std()

highAge = averageAge + standardDeviationAge

countriesHigherLifeExpectancy = countryLifeExp2018[countryLifeExp2018["Life expectancy (period) at birth - Sex: all - Age: 0"] > highAge ]

print(countriesHigherLifeExpectancy[["Entity","Life expectancy (period) at birth - Sex: all - Age: 0"]])

##C

poorCountriesGdp = countryGdp2018["GDP per capita"].quantile(0.45)

highLifeExpectancy = countryLifeExp2018["Life expectancy (period) at birth - Sex: all - Age: 0"].quantile(0.6)

highLifePoor = gdpLe2018[(gdpLe2018["GDP per capita"] <= poorCountriesGdp) & (gdpLe2018["Life expectancy (period) at birth - Sex: all - Age: 0"] >= highLifeExpectancy)]

print(highLifePoor[["Entity","Life expectancy (period) at birth - Sex: all - Age: 0","GDP per capita"]])

##D

richCountriesOverallGdp = Gdp2018["GDP, PPP (constant 2017 international $)"].quantile(0.8)

richCountriesOverall = gdpLe2018overall[gdpLe2018overall["GDP, PPP (constant 2017 international $)"] > richCountriesOverallGdp]

richCountriesOverallHighLife = richCountriesOverall[richCountriesOverall["Life expectancy (period) at birth - Sex: all - Age: 0"] > highLifeExpectancy]

##E

richCountriesGdp = countryGdp2018["GDP per capita"].quantile(0.8)

richCountries = gdpLe2018[gdpLe2018["GDP per capita"] > richCountriesGdp]

richCountriesHighLife = richCountries[richCountries["Life expectancy (period) at birth - Sex: all - Age: 0"] > highLifeExpectancy]
```

Here is the python code for 1998:

```
import pandas as pd

import numpy

import matplotlib.pyplot as plt

from pandas.plotting import table


df1 = pd.read_csv("life-expectancy.csv")

df2 = pd.read_csv("gdp-per-capita-maddison.csv")

df3 = pd.read_csv("national-gdp-wb.csv")


pd.set_option("display.max_rows", None, "display.max_columns", None)


df1_1998 = df1[(df1["Code"].notna()) & (df1["Year"] == 1998)]

countryLifeExp1998 = df1_1998[["Entity", "Life expectancy (period) at birth - Sex: all - Age: 0"]]


df2_1998 = df2[(df2["Code"].notna()) & (df2["Year"] == 1998)]

countryGdp1998 = df2_1998 [["Entity", "GDP per capita"]]


df3_1998 = df3[(df3["Code"].notna()) & (df3["Year"] == 1998)]

Gdp1998 = df3_1998 [["Entity", "GDP, PPP (constant 2017 international $)"]]


gdpLe1998 = pd.merge(countryLifeExp1998, countryGdp1998, on="Entity")

gdpLe1998overall = pd.merge(countryLifeExp1998, Gdp1998, on="Entity")


##A
```

```
plt.scatter(gdpLe1998["Life expectancy (period) at birth - Sex: all - Age: 0"], gdpLe1998["GDP per capita"])

plt.title("GDP and Life expectancy in 1998")

plt.xlabel("Life expectancy")

plt.ylabel("GDP per capita in US dollars")

##B

averageAge = countryLifeExp1998["Life expectancy (period) at birth - Sex: all - Age: 0"].mean()

standardDeviationAge = countryLifeExp1998["Life expectancy (period) at birth - Sex: all - Age: 0"].std()

highAge = averageAge + standardDeviationAge

countriesHigherLifeExpectancy = countryLifeExp1998[countryLifeExp1998["Life expectancy (period) at birth - Sex: all - Age: 0"] > highAge ]

print(countriesHigherLifeExpectancy[["Entity","Life expectancy (period) at birth - Sex: all - Age: 0"]])

##C

poorCountriesGdp = countryGdp1998["GDP per capita"].quantile(0.45)

highLifeExpectancy = countryLifeExp1998["Life expectancy (period) at birth - Sex: all - Age: 0"].quantile(0.6)

highLifePoor = gdpLe1998[(gdpLe1998["GDP per capita"] <= poorCountriesGdp) & (gdpLe1998["Life expectancy (period) at birth - Sex: all - Age: 0"] >= highLifeExpectancy)]

print(highLifePoor[["Entity","Life expectancy (period) at birth - Sex: all - Age: 0","GDP per capita"]])

##D

richCountriesOverallGdp = Gdp1998["GDP, PPP (constant 2017 international $)"].quantile(0.8)

richCountriesOverall = gdpLe1998overall[gdpLe1998overall["GDP, PPP (constant 2017 international $)"] > richCountriesOverallGdp]

richCountriesOverallHighLife = richCountriesOverall[richCountriesOverall["Life expectancy (period) at birth - Sex: all - Age: 0"] > highLifeExpectancy]

##E

richCountriesGdp = countryGdp1998["GDP per capita"].quantile(0.8)
```

richCountries = gdpLe1998[gdpLe1998["GDP per capita"] > richCountriesGdp]

richCountriesHighLife = richCountries[richCountries["Life expectancy (period) at birth - Sex: all - Age: 0"] > highLifeExpectancy]