DAT405 Assignment 1 – Group 95

Martin Blom - (3 hrs)
Jakob Windt - (3 hrs)

January 18, 2023

Problem A

We started by acquiring two data-sets being: "GDP per capita"[1] and "Life expectancy"[2] for every country during every recorded year. We decided to drop all entries that were not from 2018, since it was the newest year with data from both data-sets. We then combined them by matching the country and appending the columns on the matching rows. We then renamed one unnecessarily long name and plotted a scatter-plot with life expectancy on the x-axis and GDP per capita on the y-axis.

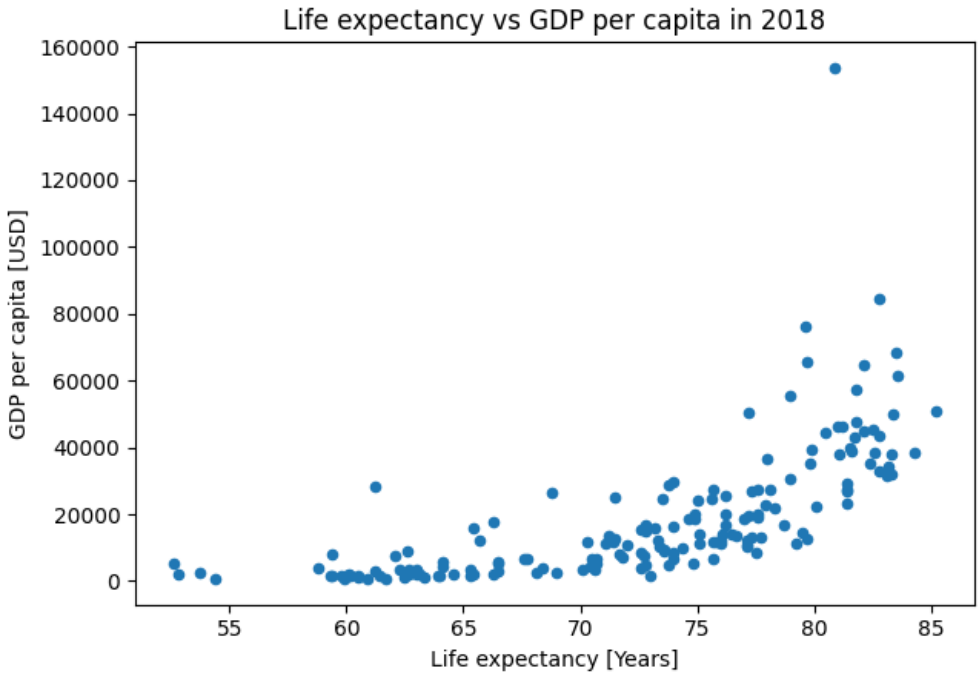

Figure 1: Scatter plot for Problem A

```python
import pandas as pd
import matplotlib.pyplot as plt

# Load data
life_exp = pd.read_csv('Lab_1/life_expectancy.csv')
gdppc = pd.read_csv('Lab_1/gdp_per_capita.csv')

# Sort for 2018
life_exp_2018 = life_exp[life_exp["Year"] == 2018]
gdppc_2018 = gdppc[gdppc["Year"] == 2018]

# Merge and sort data
data_2018 = pd.merge(life_exp_2018, gdppc_2018, how="left", on="Entity")
data_2018 = data_2018[data_2018["Year_y"] == 2018]
data_2018 = data_2018.drop(columns=["Year_x", "Year_y", "417485-annotations", "Code_x"
    , "Code_y"])
data_2018 = data_2018.rename(columns = {"Life expectancy - Sex: all - Age: at birth -
    Variant: estimates" : "Life expectancy"})

# Plot data
data_2018.plot.scatter(x="Life expectancy", y="GDP per capita", title="Life expectancy
     vs GDP per capita in 2018")

plt.xlabel("Life expectancy [Years]")
plt.ylabel("GDP per capita [USD]")
plt.show()
```

Listing 1: Python code – Scatter Plot.

Problem B

```
print("Statistics for 2018:")
print(data_2018.describe())
print("-"*50)

mean_le = 72.631928
std_le = 7.772033

countries = data_2018[data_2018["Life expectancy"] > mean_le + std_le]
print(countries)
```

Listing 2: Additional code

```
Statistics for 2018:
       Life expectancy  GDP per capita
count       166.000000      166.000000
mean         72.631928    19030.645795
std           7.772033    20286.783469
min          52.600000      623.488900
25%          66.350000     4568.405750
50%          74.000000    12237.940000
75%          78.600000    27324.020250
max          85.200000   153764.170000
--------------------------------------------------
            Entity  Life expectancy  GDP per capita
13       Australia             83.4       49830.800
14         Austria             81.7       42988.070
21         Belgium             81.5       39756.203
38          Canada             82.1       44868.742
54          Cyprus             81.4       27184.416
57         Denmark             81.0       46312.344
73         Finland             81.6       38896.700
74          France             82.6       38515.918
80         Germany             81.2       46177.617
83          Greece             81.4       23450.766
96       Hong Kong             85.2       50839.370
98         Iceland             82.8       43438.543
103        Ireland             82.1       64684.300
105         Israel             82.8       32954.770
106          Italy             83.2       34364.168
108          Japan             84.3       38673.810
133     Luxembourg             81.8       57427.500
140          Malta             83.3       32028.912
160    Netherlands             81.8       47474.110
162    New Zealand             82.4       35336.137
171         Norway             82.8       84580.130
183       Portugal             81.4       27035.600
185          Qatar             80.9      153764.170
205      Singapore             83.5       68402.340
208       Slovenia             81.4       29244.920
213    South Korea             83.3       37927.610
215          Spain             83.1       31496.520
219         Sweden             82.5       45541.890
220    Switzerland             83.6       61372.730
222         Taiwan             80.5       44663.863
239  United Kingdom             81.1       38058.086
```

Figure 2: List of Countries for Problem B and Additional Statistics

The picture above shows a list of countries where the life expectancy is higher then the mean plus one standard deviation. The mean is equal to 72.63 years, and one standard deviation is 7.77 years.

Problem C

For this question we had to decide for ourselves what we defined as a high life expectancy and low GDP. We believe that a high life expectancy is when a countries average life expectancy is greater than the mean. To decide a low GDP we chose countries with a GDP lower then the mean.

```
mean_gdp = 19030.645795
std_gdp = 20286.783469

countries_highLE = data_2018[data_2018["Life expectancy"] > mean_le ]
countries_highLE_lowGDP = countries_highLE[countries_highLE["GDP per capita"] <
    mean_gdp]
print("Countries with high life expectancy and low GDP per capita:")
print(countries_highLE_lowGDP)
```

Listing 3: Additional code

Countries with high life expectancy and low GDP per capita:

	Entity	Life expectancy	GDP per capita
2	Albania	79.2	11104.1660
3	Algeria	76.1	14228.0250
9	Argentina	77.0	18556.3830
10	Armenia	75.1	11454.4250
15	Azerbaijan	72.8	16628.0550
19	Barbados	77.1	11995.1870
20	Belarus	74.6	18727.3180
28	Bosnia and Herzegovina	77.1	10460.5205
30	Brazil	75.1	14033.5650
33	Bulgaria	74.9	18444.2600
39	Cape Verde	75.7	6831.2160
44	China	77.7	13101.7060
45	Colombia	76.7	13545.0500
49	Costa Rica	79.5	14686.2540
52	Cuba	77.5	8325.6310
59	Dominica	73.6	9021.1740
60	Dominican Republic	73.2	15912.3990
61	Ecuador	77.1	10638.8250
79	Georgia	73.3	11984.9050
88	Guatemala	72.7	7402.1147
95	Honduras	72.8	5041.6353
101	Iran	76.2	17011.3050
110	Jordan	75.8	11506.3380
122	Lebanon	79.7	12558.9670
128	Libya	72.8	15013.3125
146	Mexico	74.0	16494.0780
154	Morocco	74.0	8451.1360
163	Nicaragua	73.8	4952.4770
167	North Korea	73.0	1596.3517
168	North Macedonia	77.3	13074.2310
176	Palestine	74.8	5207.7570
179	Paraguay	73.6	9338.9480
180	Peru	76.0	12310.0850
193	Saint Lucia	73.4	10475.3690
202	Serbia	76.5	14124.1180
216	Sri Lanka	75.7	11662.9060
225	Thailand	78.7	16648.6230
231	Tunisia	76.0	11353.8870
237	Ukraine	74.4	9813.3630
248	Vietnam	74.0	6814.1420

Figure 3: List of Countries for Problem C

Problem D

No, during 2018 there were three countries with a low life expectancy while having a high GDP per capita. This means that not all strong economies (by our definition) have a high life expectancy.

```
countries_highGDP = data_2018[data_2018["GDP per capita"] > mean_gdp]
countries_highGDP_lowLE = countries_highGDP[countries_highGDP["Life expectancy"] <
    mean_le]
print("Countries with high GDP per capita and low life expectancy:")
print(countries_highGDP_lowLE)
```

Listing 4: Additional code

```
Countries with high GDP per capita and low life expectancy:
                Entity  Life expectancy  GDP per capita
64   Equatorial Guinea             61.2       28528.953
111         Kazakhstan             71.5       25307.555
233       Turkmenistan             68.8       26318.365
```

Figure 4: List of Countries for Problem C

Problem E

Since we defined that countries with strong economies must have a GDP per capita higher then the mean, that means that we would get same results on this question. We researched some data on countries with a high GDP[3]. We can see that some of these countries, like for example India, which has a GDP of 2.6 trillion dollars but still has a low life expectancy of 70.7 years.

As shown, we can conclude that the statistics differ between looking at a countries GDP and GDP per capita because GDP doesn't take the countries population into account.

References

[1] "GDP per Capita." Our World in Data, https://ourworldindata.org/grapher/maddison-data-gdp-per-capita-in-2011us.

[2] "Life Expectancy." Our World in Data, https://ourworldindata.org/grapher/life-expectancy-at-birth-total-years?time=2018.

[3] "Gross Domestic Product (GDP)." Our World in Data, https://ourworldindata.org/grapher/gross-domestic-product?time=2018.