

Assignment 1: Introduction to Data Science and Python

Authors: Kevin To, Filip Cederqvist

Work load: 10 h each

Note: Data sets has been gathered from <https://ourworldindata.org/>.

	Entity	Code	Year	Life expectancy at birth (historical)
0	Afghanistan	AFG	1950	27.7
1	Afghanistan	AFG	1951	28.0
2	Afghanistan	AFG	1952	28.4
3	Afghanistan	AFG	1953	28.9
4	Afghanistan	AFG	1954	29.2

	Entity	Code	Year	GDP per capita
0	Afghanistan	AFG	2002	1189.784668
1	Afghanistan	AFG	2003	1235.810059
2	Afghanistan	AFG	2004	1200.277954
3	Afghanistan	AFG	2005	1286.793701
4	Afghanistan	AFG	2006	1315.789062

	Entity	Code	Year	Population (historical estimates)
0	Afghanistan	AFG	-10000	14737
1	Afghanistan	AFG	-9000	20405
2	Afghanistan	AFG	-8000	28253
3	Afghanistan	AFG	-7000	39120
4	Afghanistan	AFG	-6000	54166

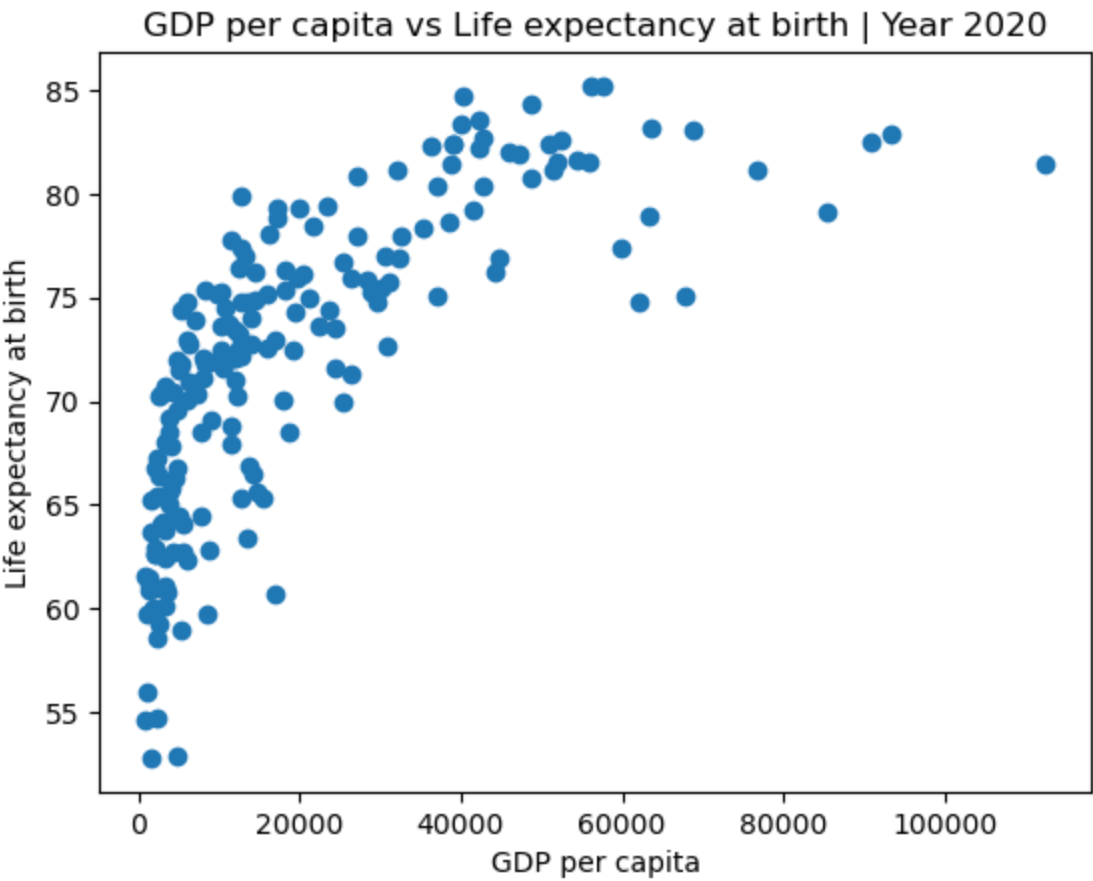
a) Write a Python program that draws a scatter plot of GDP per capita vs life expectancy. State any assumptions and motivate decisions that you make when selecting data to be plotted, and in combining data.

Assumption: We remove all countries in the dataset that begins with "OWID" as they are not real countries. We also remove "countries" that are defined as continents in the dataset. Further, we only gather data from the year 2020.

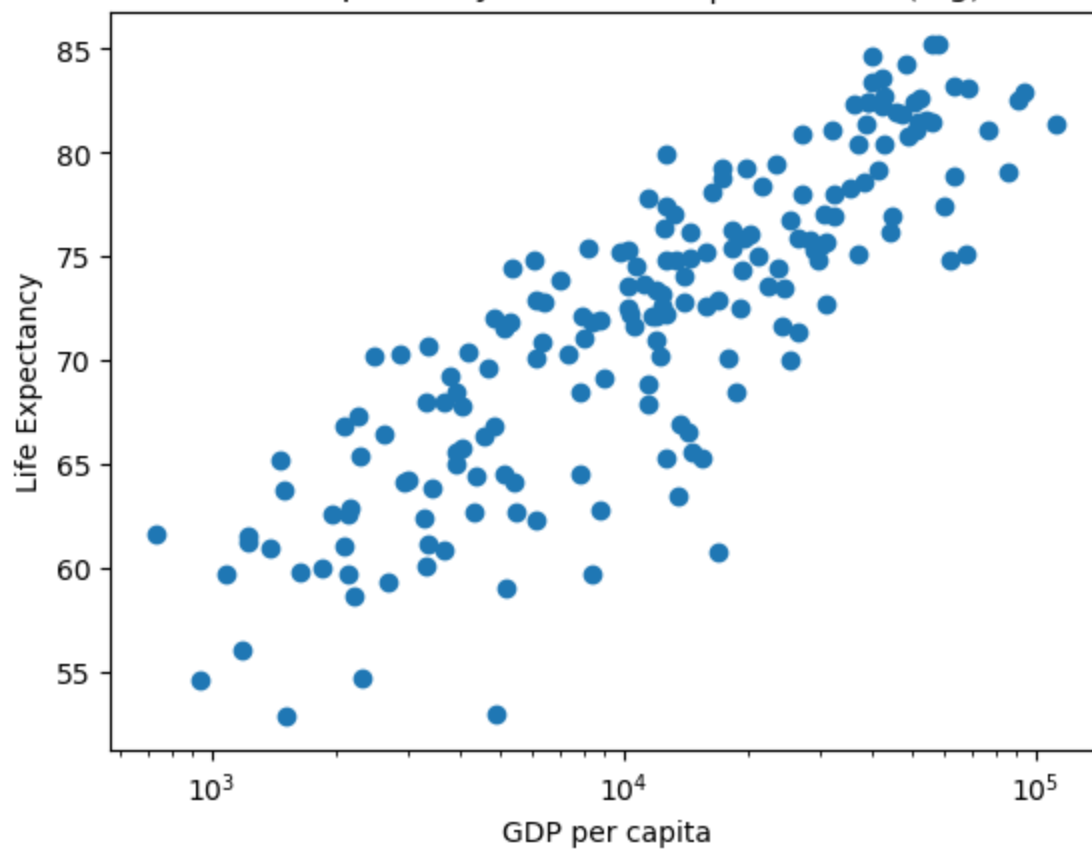
	Entity	Code	Year	Life expectancy at birth (historical)	GDP per capita	Population (historical estimates)
0	Afghanistan	AFG	2020	62.6	1970.560181	38972236
1	Albania	ALB	2020	77.0	13191.952148	2866850
2	Algeria	DZA	2020	74.5	10735.382812	43451668
3	Angola	AGO	2020	62.3	6110.274414	33428490

4	Antigua and Barbuda	ATG	2020	78.8	17293.302734	92672
...
185	Uzbekistan	UZB	2020	70.3	7331.859863	33526662
186	Vanuatu	VUT	2020	70.3	2854.292480	311694
187	Vietnam	VNM	2020	75.4	8200.332031	96648680
189	Zambia	ZMB	2020	62.4	3277.816406	18927716
190	Zimbabwe	ZWE	2020	61.1	3353.411377	15669663

189 rows x 6 columns



Life Expectancy versus GDP | Year 2020 (log)



b) Which countries have a life expectancy higher than one standard deviation above the mean?

Answer: See table below.

One standard deviation above the mean: 79.39052132447404

	Entity	Code	Year	Life expectancy at birth (historical)	GDP per capita	Population (historical estimates)
101	Macao	MAC	2020	85.2	57728.988281	676287
74	Hong Kong	HKG	2020	85.2	56153.972656	7500955
85	Japan	JPN	2020	84.7	40232.210938	125244760
7	Australia	AUS	2020	84.3	48678.683594	25670052
161	South Korea	KOR	2020	83.6	42335.867188	51844688
107	Malta	MLT	2020	83.4	39979.699219	515364
128	Norway	NOR	2020	83.2	63548.0	5379836
167	Switzerland	CHE	2020	83.1	68755.4375	8638609
155	Singapore	SGP	2020	82.9	93397.046875	5909874
123	New Zealand	NZL	2020	82.7	42774.691406	5061130
76	Iceland	ISL	2020	82.6	52375.679688	366682
81	Ireland	IRL	2020	82.5	90789.21875	4946123
166	Sweden	SWE	2020	82.4	50923.109375	10368968
82	Israel	ISR	2020	82.4	39056.417969	8757487
83	Italy	ITA	2020	82.4	39073.414062	59500576
162	Spain	ESP	2020	82.3	36210.875	47363808

60	France	FRA	2020	82.2	42320.523438	64480052
30	Canada	CAN	2020	82.0	46064.253906	37888704
59	Finland	FIN	2020	81.9	47154.425781	5529468
122	Netherlands	NLD	2020	81.6	54324.382812	17434562
47	Denmark	DNK	2020	81.5	55819.910156	5825638
8	Austria	AUT	2020	81.5	51857.691406	8907780
44	Cyprus	CYP	2020	81.4	38816.09375	1237540
100	Luxembourg	LUX	2020	81.4	112557.3125	630401
64	Germany	DEU	2020	81.1	51423.234375	83328992
18	Bermuda	BMU	2020	81.1	76626.523438	64051
139	Portugal	PRT	2020	81.1	31961.775391	10298186
66	Greece	GRC	2020	80.9	27072.619141	10512232
15	Belgium	BEL	2020	80.8	48769.726562	11561716
157	Slovenia	SVN	2020	80.4	37050.546875	2117648
182	United Kingdom	GBR	2020	80.4	42675.78125	67059472
105	Maldives	MDV	2020	79.9	12744.429688	514451
35	Chile	CHL	2020	79.4	23324.525391	19300318

c) Which countries have high life expectancy but have low GDP? (note: GDP and not GDP per capita in question c and d) Motivate how you have chosen to define "high" and "low".

We define "low" GDP to be below the **median** of all countries GDP since the distrubution of countries' GDP is skewed. This is illustrated in the plot above, where you can see that the values vary a lot. On the other hand, we define "high" life expectancy to be above the **mean** of world's life expectancy. This is because this distrubition is more like a normal distrubtion.

Answer: See table below.

Median GDP: 70956242357.3125
Mean Life Expectancy: 71.93809523809517

	Entity	Life expectancy at birth (historical)	GDP
101	Macao	85.2	3.904136e+10
107	Malta	83.4	2.060410e+10
76	Iceland	82.6	1.920522e+10
44	Cyprus	81.4	4.803647e+10
18	Bermuda	81.1	4.908005e+09
105	Maldives	79.9	6.556385e+09
11	Bahrain	79.2	6.128664e+10
4	Antigua and Barbuda	78.8	1.602605e+09
55	Estonia	78.3	4.687260e+10
95	Lebanon	77.8	6.505802e+10
13	Barbados	77.4	3.552422e+09

1	Albania	77.0	3.781935e+10
115	Montenegro	76.3	1.148594e+10
21	Bosnia and Herzegovina	76.2	4.814769e+10
43	Curacao	76.1	3.844472e+09
94	Latvia	75.5	5.710193e+10
127	North Macedonia	75.2	3.363235e+10
32	Cayman Islands	75.1	4.569564e+09
177	Turks and Caicos Islands	75.0	9.358605e+08
67	Grenada	74.9	1.791826e+09
24	Brunei	74.8	2.747637e+10
31	Cape Verde	74.8	3.522131e+09
174	Trinidad and Tobago	74.4	3.601409e+10
132	Palestine	74.4	2.707504e+10
110	Mauritius	74.3	2.525958e+10
49	Dominica	73.6	7.409116e+08
153	Seychelles	73.5	2.578223e+09
146	Saint Lucia	73.4	2.159710e+09
16	Belize	72.9	2.417706e+09
63	Georgia	72.8	5.259596e+10
148	Samoa	72.8	1.379184e+09
10	Bahamas	72.7	1.253846e+10
165	Suriname	72.6	9.630861e+09
98	Libya	72.5	6.841781e+10
6	Armenia	72.2	3.540650e+10
114	Mongolia	72.1	3.862242e+10
147	Saint Vincent and the Grenadines	72.1	1.260371e+09

d) Does every strong economy (normally indicated by GDP) have high life expectancy?

Answer: No. As can be seen below, there are some countries that have a high GDP, but still a low life expentancy.

	Entity	Life expectancy at birth (historical)	GDP
126	Nigeria	52.9	1.024288e+12
41	Cote d'Ivoire	59.0	1.389202e+11
46	Democratic Republic of Congo	59.7	1.005085e+11
29	Cameroon	60.8	9.712721e+10
3	Angola	62.3	2.042572e+11
0	Afghanistan	62.6	7.679714e+10
88	Kenya	62.7	2.256108e+11
179	Uganda	62.9	9.658140e+10

65	Ghana	64.1	1.752456e+11
20	Bolivia	64.5	9.364129e+10
160	South Africa	65.3	7.447753e+11
57	Ethiopia	65.4	2.691673e+11
164	Sudan	65.6	1.745197e+11
130	Pakistan	66.3	1.036613e+12
169	Tanzania	66.4	1.626134e+11
118	Myanmar	66.8	2.594973e+11
9	Azerbaijan	66.9	1.411792e+11
78	Indonesia	68.8	3.111404e+12
80	Iraq	69.1	3.835398e+11
121	Nepal	69.2	1.115267e+11
87	Kazakhstan	70.0	4.813739e+11
111	Mexico	70.1	2.249367e+12
77	India	70.1	8.609769e+12
185	Uzbekistan	70.3	2.458128e+11
52	Egypt	71.0	1.284364e+12
143	Russia	71.3	3.852509e+12
68	Guatemala	71.8	1.457302e+11

e) Related to question d, what would happen if you use GDP per capita as an indicator of a strong economy? Explain the results you obtained, and discuss any insights you get from comparing the results of d and e.

Answer: From the tables below, it is clear that GDP per capita corrolates better with life expectancy than GDP. When GDP is the indicator of a strong economy, a lot more countries with below average life expenctancy is included. As such, GDP per capita is a better indicator for a countries wellness.

Moreover, we could also see that a high GDP don't always corrolate to a high GDP per capita. Examples of this is India and Nigeria. Both countries places high when sorting on GDP, but their GDP per capita is low. One take-away from this is that GDP doesn't really indicate anything, since it must be taken into proportion to its population. Countries with larger population generally have a larger GDP, but this does not mean that they have a high GDP per capita.

	Entity	Life expectancy at birth (historical)	GDP per capita	GDP	Population (historical estimates)
143	Russia	71.3	26456.388672	3.852509e+12	145617328
87	Kazakhstan	70.0	25363.166016	4.813739e+11	18979250
145	Saint Kitts and Nevis	71.6	24319.363281	1.159207e+09	47666
71	Guyana	68.5	18679.955078	1.489179e+10	797207
111	Mexico	70.1	17852.359375	2.249367e+12	125998296
54	Equatorial Guinea	60.7	17007.625000	2.714514e+10	1596057
131	Palau	65.3	15473.251953	2.785031e+08	17999

22	Botswana	65.6	14654.651367	3.731666e+10	2546404
61	Gabon	66.5	14320.844727	3.283173e+10	2292583
9	Azerbaijan	66.9	13726.769531	1.411792e+11	10284952
120	Nauru	63.4	13594.492188	1.677424e+08	12339
160	South Africa	65.3	12665.832031	7.447753e+11	58801928

	Entity	Life expectancy at birth (historical)	GDP per capita	GDP	Population (historical estimates)
77	India	70.1	6165.746582	8.609769e+12	1396387072
143	Russia	71.3	26456.388672	3.852509e+12	145617328
78	Indonesia	68.8	11444.960938	3.111404e+12	271857984
111	Mexico	70.1	17852.359375	2.249367e+12	125998296
52	Egypt	71.0	11951.447266	1.284364e+12	107465128
130	Pakistan	66.3	4562.622559	1.036613e+12	227196736
126	Nigeria	52.9	4916.721191	1.024288e+12	208327408
160	South Africa	65.3	12665.832031	7.447753e+11	58801928
87	Kazakhstan	70.0	25363.166016	4.813739e+11	18979250
80	Iraq	69.1	9012.381836	3.835398e+11	42556988
57	Ethiopia	65.4	2296.827393	2.691673e+11	117190920
118	Myanmar	66.8	4857.389648	2.594973e+11	53423200
185	Uzbekistan	70.3	7331.859863	2.458128e+11	33526662
88	Kenya	62.7	4339.856934	2.256108e+11	51985780
3	Angola	62.3	6110.274414	2.042572e+11	33428490
65	Ghana	64.1	5445.725586	1.752456e+11	32180402
164	Sudan	65.6	3927.043945	1.745197e+11	44440484
169	Tanzania	66.4	2635.356689	1.626134e+11	61704520
68	Guatemala	71.8	8393.284180	1.457302e+11	17362716
9	Azerbaijan	66.9	13726.769531	1.411792e+11	10284952
41	Cote d'Ivoire	59.0	5181.311035	1.389202e+11	26811792
121	Nepal	69.2	3800.065674	1.115267e+11	29348626
46	Democratic Republic of Congo	59.7	1082.445190	1.005085e+11	92853168
29	Cameroon	60.8	3666.410645	9.712721e+10	26491090
179	Uganda	62.9	2175.031006	9.658140e+10	44404608
20	Bolivia	64.5	7845.170898	9.364129e+10	11936169
0	Afghanistan	62.6	1970.560181	7.679714e+10	38972236