

Analysis of Global Life Expectancy At Birth from 2013 - 2018

Haoda Li

February 7, 2019

Introduction

The purpose of this research is to analyze what influence do different factors have on global life expectancy at birth. This research will investigate the following questions.

1. Has global life expectancy at birth changed between 2013 to 2018? Which continents and countries have experienced the least and greatest changes?
2. Are a countries' health expenditures and carbon dioxide emissions from consumption of energy associated with life expectancy at birth?
3. Are a countries' military expenditures and birth rate associated with life expectancy at birth?

In this research, we will look at the following factors and variables of each country:

- Life expectancy at birth (years)
- Health expenditures (% of GDP)
- Carbon dioxide emissions from consumption of energy (MT)
- Population
- Military expenditures (% of GDP)
- Birth rate (Births per 1000 population)

All the data is obtained from CIA World Factbook country comparisons (<https://www.cia.gov/library/publications/download>).

Question 1

1.1 Has global life expectancy at birth changed between 2013 to 2018?

Investigation

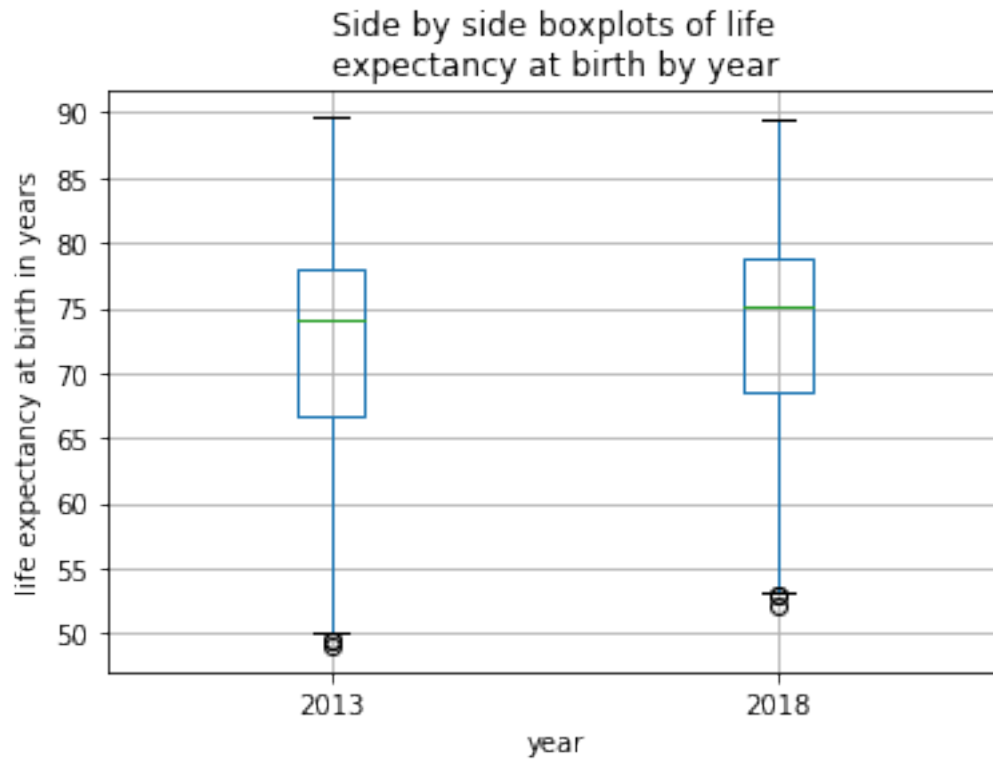
First, I examine the summary statistics of the two data sets.

```
count    228.000000
mean      71.581579
std        8.893017
min       49.070000
25%       66.717500
50%       74.155000
75%       78.075000
max       89.630000
Name: life expectancy, dtype: float64
```

```
count    228.000000
mean      73.301754
std        7.561915
min       52.100000
25%       68.550000
50%       75.150000
75%       78.900000
max       89.400000
Name: life expectancy, dtype: float64
```

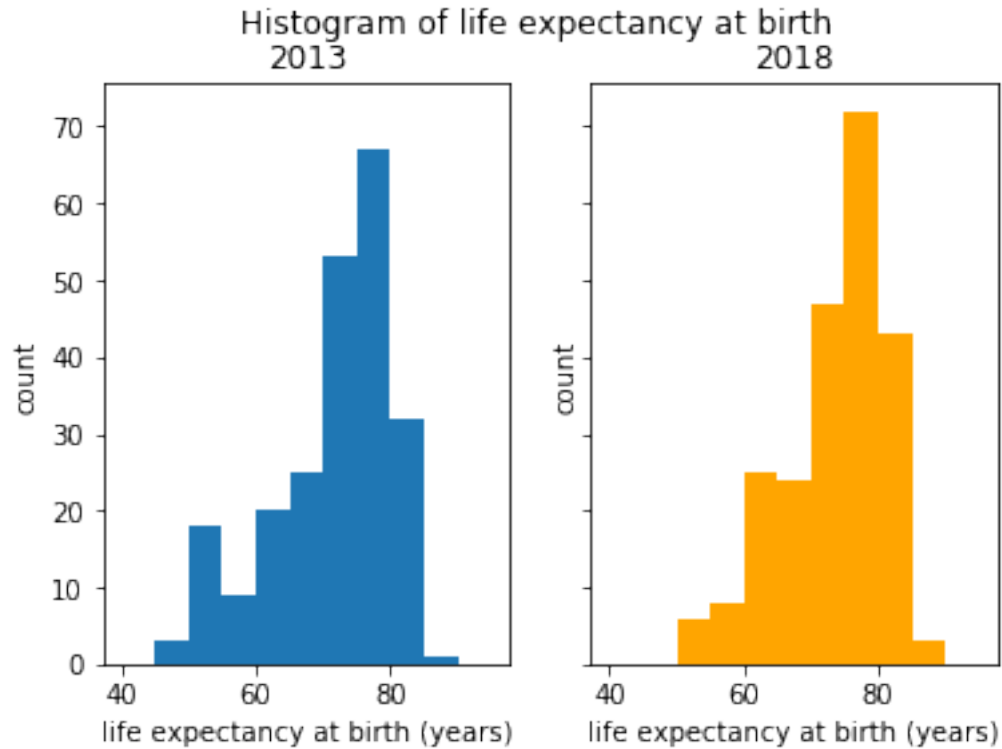
Looking at the summaries, there is a minor increase (1.73 years) in mean from 2013 to 2018. Also, although the maximum does not change much, the minimum increased from 49.07 to 52.10.

I use box plot and histogram to visualize how the distribution of life expectancy for 2013 and 2018.



Generally, we see that data points of 2018 is greater than data points of 2013, especially for countries in the lower quartile.

I further look the histogram, the overall distribution does not change much, while there are fewer countries with life expectancy in the range of 45-60 years.



The plots have evidence that there is some change in the distribution of life expectancy at birth, especially for data points in the first quartile. From the histogram, the number of countries with life expectancy at birth between 45-60 decreases.

Conclusion

From the observations, we can conclude that there is a change in global life expectancy at birth. From the side by side box plots, the change has more effect on countries with lower life expectancy, and has little effect on countries with high life expectancy.

Discussion

The methods used above also has limited considerations for the correlation between life expectancy and the country. For further analysis, I'd like to investigate how much change each country experienced in life expectancy over the 5 years, and measure the sum of such changes as a measurement of the changes of global life expectancy at birth.

1.2 Which continents have experienced the least and greatest changes?

Investigation

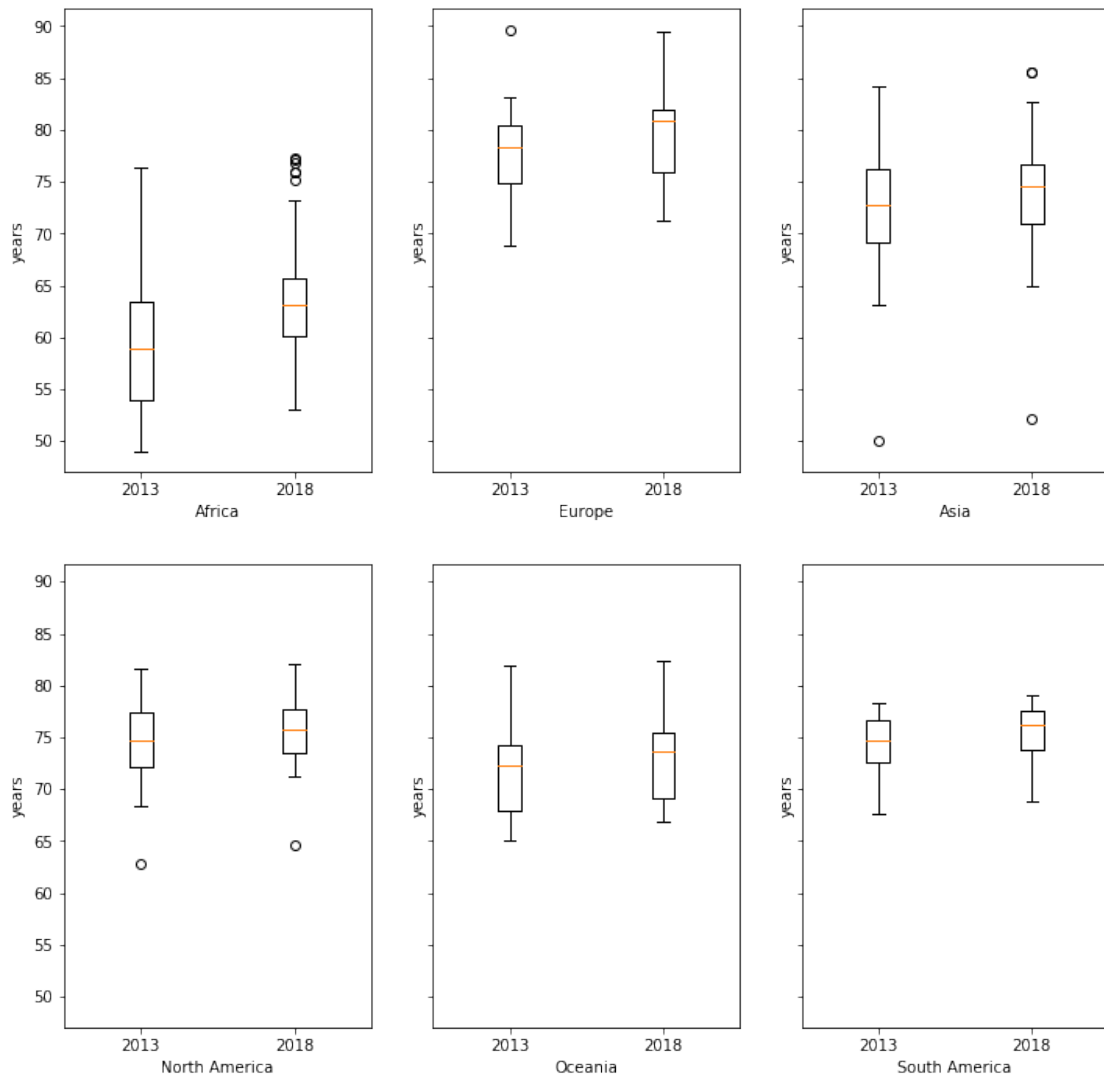
To examine the changes of life expectancy at birth among continents, I investigate the change of means, maximum, minimum, and median I put them into numerical summaries and side by side box plots for easier visualization.

continent	diff in Q1	diff in Q3	diff in max	diff in mean	diff in median	diff in min
Africa	6.1500	2.3275	0.99	3.715337	4.120	3.93
Europe	0.9975	1.5100	-0.23	1.171957	2.540	2.37
Asia	1.8175	0.4700	1.31	1.380833	1.760	1.99
North America	1.3450	0.3300	0.43	0.864783	0.950	1.75
Oceania	1.1875	1.1700	0.42	1.239286	1.280	1.79
South America	1.2625	0.9225	0.83	1.149167	1.575	1.22

Notice that Africa has the greatest changes in mean, and North America has the least changes in mean.

For a better visualization, we can see them in side by side box plots.

Side by Side Bxplots of Life Expectancy at Birth in 2013 and 2018 Among Continents



From the box plots above, we can observe that - Africa has the greatest changes in minimum and median among all continents. - North America has the least changes in median. South America has the least changes in minimum. Oceania has the least change in maximum. However, the difference of the changes in maximum is not significant between Oceania and North America.

Conclusion

From the observations, there is enough evidence suggesting that Africa experienced the greatest changes, and North America experienced the least changes.

Discussion

From the box plots above, I also observe that North America and Oceania both experienced little changes in life expectancy at birth. However, I notice that the number of countries that belong to North America or Oceania are significantly smaller than that of other continents. Since life expectancy at birth is closely associated with the country, this may results in less change in different statistical measurements we've done.

1.3 Which countries have experienced the least and greatest changes?

Methods

For the changes of each country, we can simply compare the difference of Life Expectancy at birth between 2013 and 2018.

country	change in years
Gaza Strip	0.00
Egypt	0.01
West Bank	0.06
Syria	0.06
Croatia	0.10
Virgin Islands	0.11
Macau	0.14
Argentina	0.18
Monaco	0.23
San Marino	0.28
country	change in years
Nigeria	6.84
Zimbabwe	7.24
Burkina Faso	7.37
Chad	8.43
Botswana	9.33
Malawi	9.42
Guinea-Bissau	11.90
Namibia	12.37
South Africa	14.62
Gabon	15.85

I display the 10 countries with the least changes and 10 countries with the greatest changes. I choose to display 10 countries for more information about the difference of change. Notice that the difference of change in years is much smaller for countries with less changes in life expectancy.

Conclusion

From the output above, we observe that Gaza Strip has the least change, while Egypt, West Bank, and Syria also has very little change in life expectancy at birth. Gabon has the most change in life expectancy

Discussion

Notice that there are many countries with a small change in life expectancy at birth. However, their population may differ a lot. Simply report the change in life expectancy by the difference may not be informative enough. For example, consider Egypt (the second smallest change in life expectancy) and Gaza Strip (the smallest change in life expectancy)

population of Egypt: 85294388.0

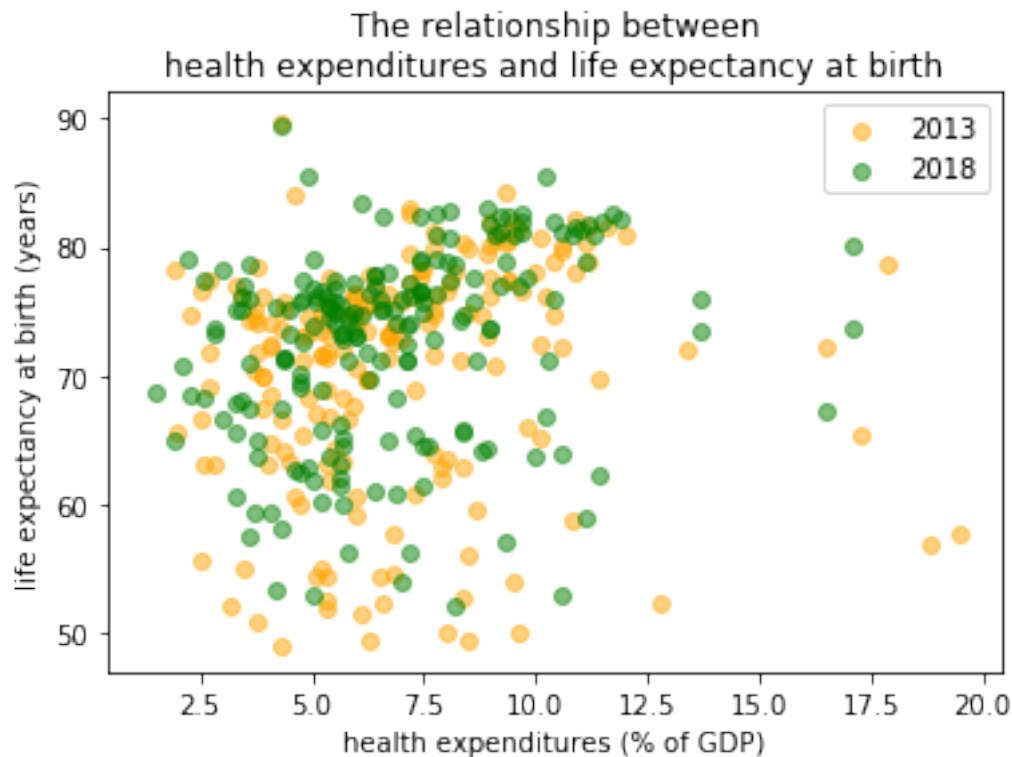
population of Gaza Strip: 1763387.0

Question 2

2.1 Is a countries' health expenditures associated with life expectancy at birth?

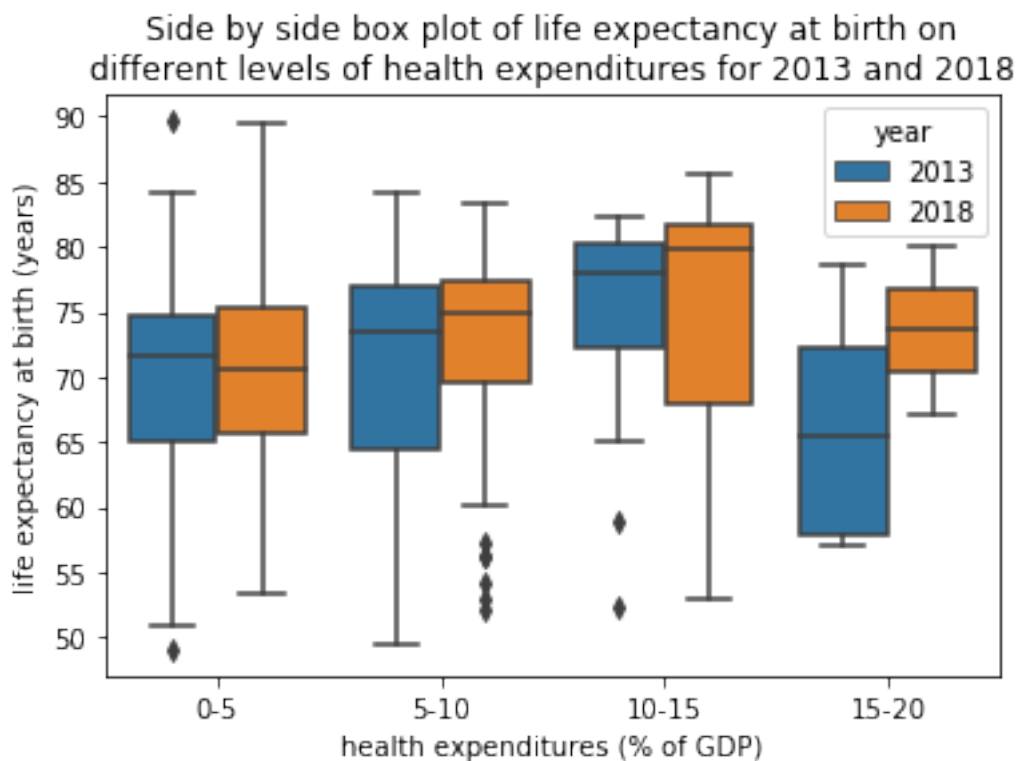
Investigation

First, we will do a scatter plot of the two variables to visually examine the data.

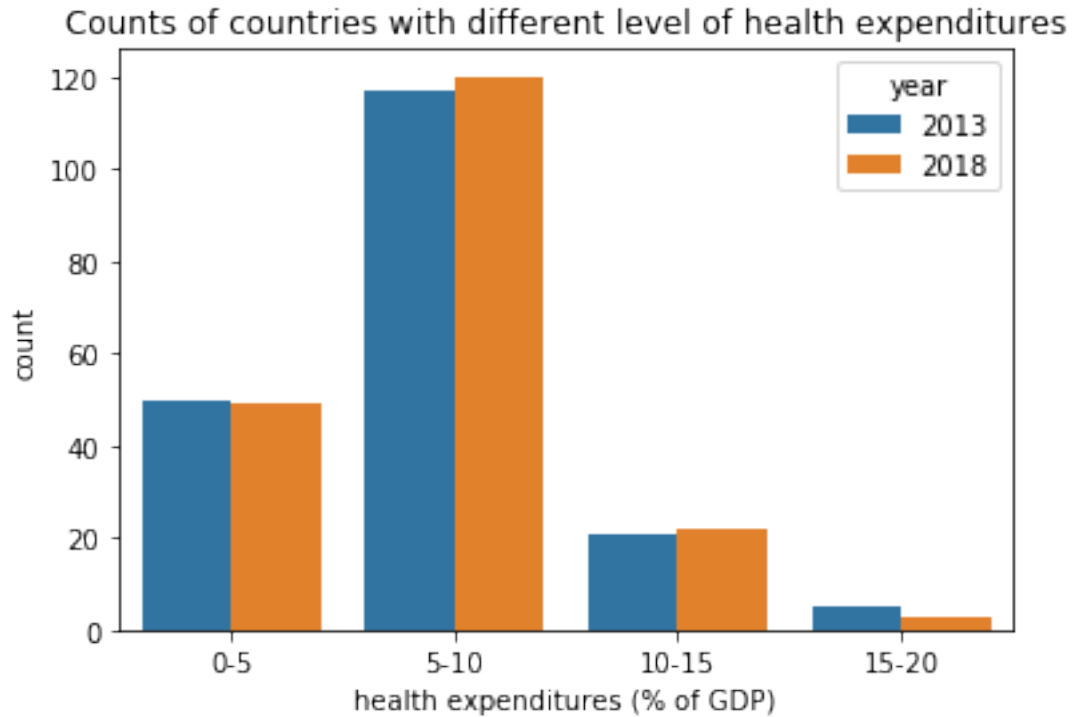


Notice that most countries have life expenditures between 5% – 12.5%, and have life expectancy at birth in between 70 – 85 years. Also, as health expenditures increase, life expectancy at birth generally increases. The plot shows some evidence that there is a positive linear relationship between health expenditures and life expectancy at birth.

For further investigation, we encode the health expenditure into a interval variable. Because I expect that the relationship between life expectancy and health expenditures is linear, I group the data points by health expenditures into 4 even intervals: 0% – 5%, 5% – 10%, 10% – 15%, 15% – 20%.



The mean difference of countries in 15% – 20% interval between 2013 and 2018 is significantly larger than the rest intervals. Also, the mean is significantly smaller than those of 10-15%. This result is unexpected. I examine these countries closely.



country	health expenditures	life expectancy	year	level
Liberia	19.5	57.81	2013	3
Sierra Leone	18.8	56.98	2013	3
United States	17.9	78.62	2013	3
Marshall Islands	16.5	72.31	2013	3
Tuvalu	17.3	65.47	2013	3
United States	17.1	80.10	2018	3
Marshall Islands	17.1	73.60	2018	3
Tuvalu	16.5	67.20	2018	3

Notice that the number of counties that has 15% – 20% health expenditures are significantly smaller than the other 3 levels. Also, since the sample size is too small, the box plot for 15-20% health expenditures are not very informative.

For the rest of countries (with health expenditures < 15%), the boxplot shows a constant increase of mean health expectancy when health expenditure increases. I will do an One-way ANOVA to further test whether the mean changes.

One-way Analysis of variance (ANOVA)

Model: $Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \epsilon$ where

- Y_i is the life expectancy at birth of the i th country,
- $X_{i,1}$ is the indicator that the i th country has health expenditures in between 5 – 10%,
- $X_{i,2}$ is the indicator that the i th country has health expenditures in between 10 – 15%

Assumptions: the error term follows that $e_i \sim N(0, \sigma^2)$

Null hypothesis: $H_0 : \beta_1 = \beta_2 = 0, H_a : \beta_1 \neq 0 \vee \beta_2 \neq 0$

Test statistics: $F = \frac{MS_{Reg}}{MSE} = \frac{\sum_i (\hat{Y}_i - \bar{Y})^2 / 2}{\sum_i (Y_i - \hat{Y}_i)^2 / 452} \sim F_{2,452}$
 Result of one-way ANOVA on 2013 data

Test statistics: 2.9511006293486997
 p-value: 0.054748632934573105

Result of one-way ANOVA on 2018 data

Test statistics: 3.4662253842039736
 p-value: 0.03324515741607386

Since p-value is small (0.055 for 2013, 0.033 for 2018), we can reject the null hypothesis, there is some evidence that the means of life expectancy aren't equal.

Conclusion

From all the methods, we can conclude that there is an association between health expenditures and life expectancy. There is some evidence that more life expenditures is associated with higher life expectancy when life expenditures is below 15%.

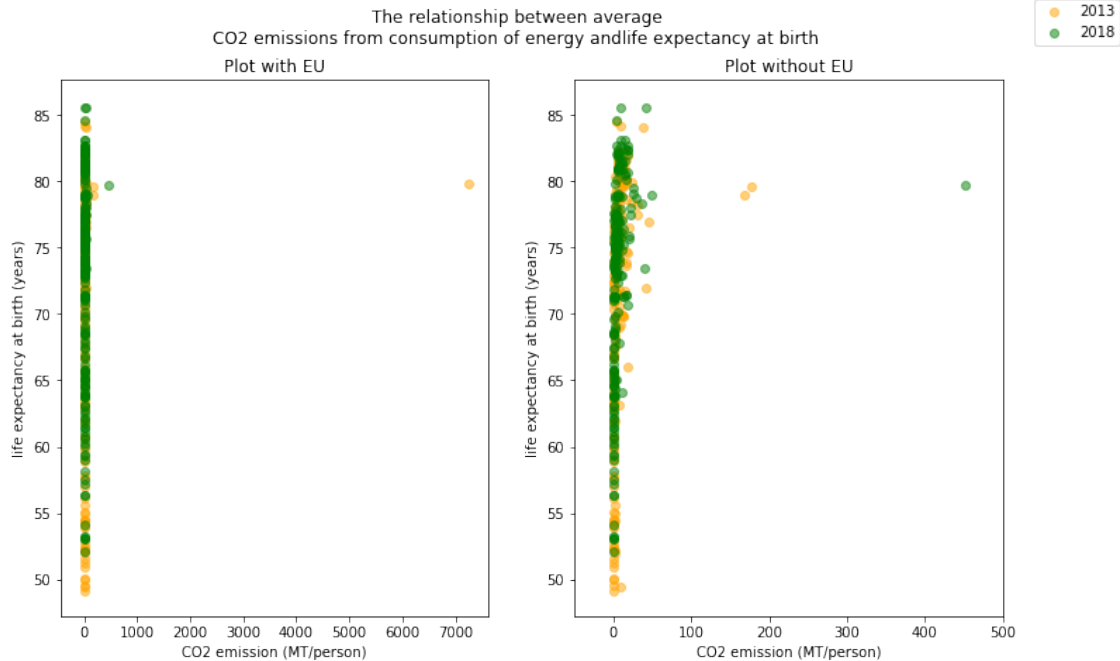
2.2 Is a countries' carbon dioxide emissions from consumption of energy associated with life expectancy at birth?

Notice that the data set only provides the total carbon dioxide emissions from consumption of energy of a country, while life expectancy at birth is the average measurement of individuals. Therefore, it's less meaningful to investigate the relationship between the two data sets. Instead, we will calculate the average carbon dioxide emissions from consumption of energy of one individual of a country (MT/person), the average is calculated

$$\text{Average carbon dioxide emission} = \frac{\text{Carbon dioxide emission}}{\text{population}}$$

Methods

I first use a scatter plot to visually investigate the dataset.



There is one data point that is extremely distant from the rest and results in the very unexpected plot (plot 1). Look closely on the abnormal data point.

country	carbon dioxide emission	life expectancy	population	year	average co2 emission
European Union	3.700000e+12	79.86	509365627.0	2013	7263.937345

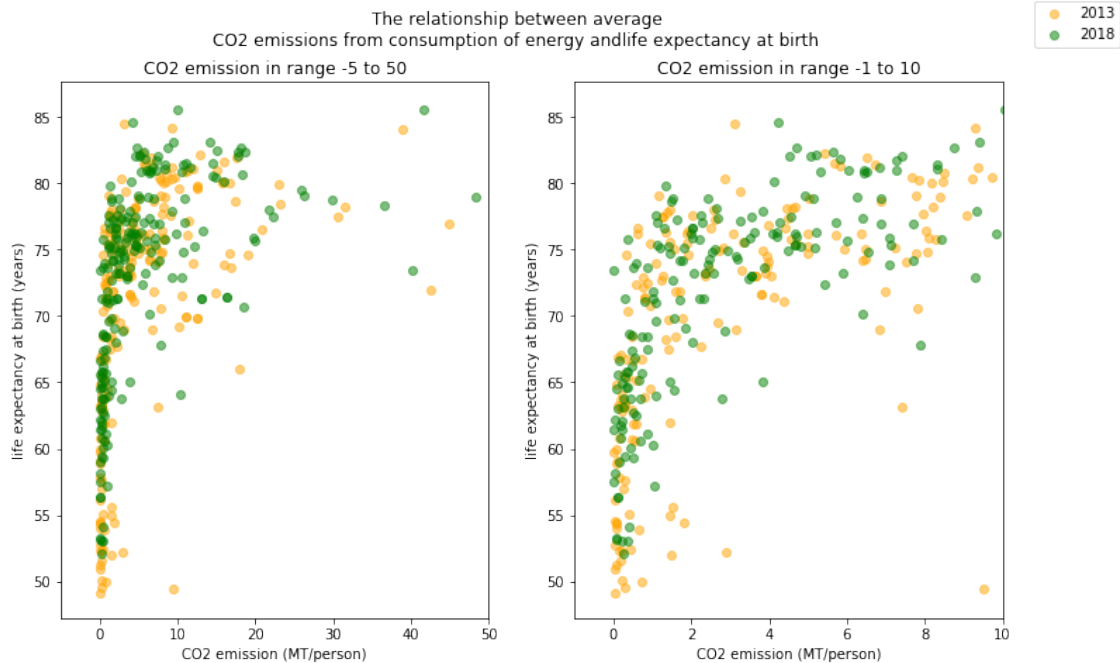
sum of CO2 emission of all countries except EU: 74991386165.0

Notice that the sum of carbon dioxide emission of all the rest countries are about 7.5×10^{10} MT. I suspect that there is some error in the collected data and decided to remove this data point.

country	carbon dioxide emission	life expectancy	population	year	average co2 emission
Virgin Islands	18590000.0	79.61	104737.0	2013	177.492195
Gibraltar	4905000.0	78.98	29111.0	2013	168.493010
Gibraltar	13340000.0	79.70	29461.0	2018	452.802009

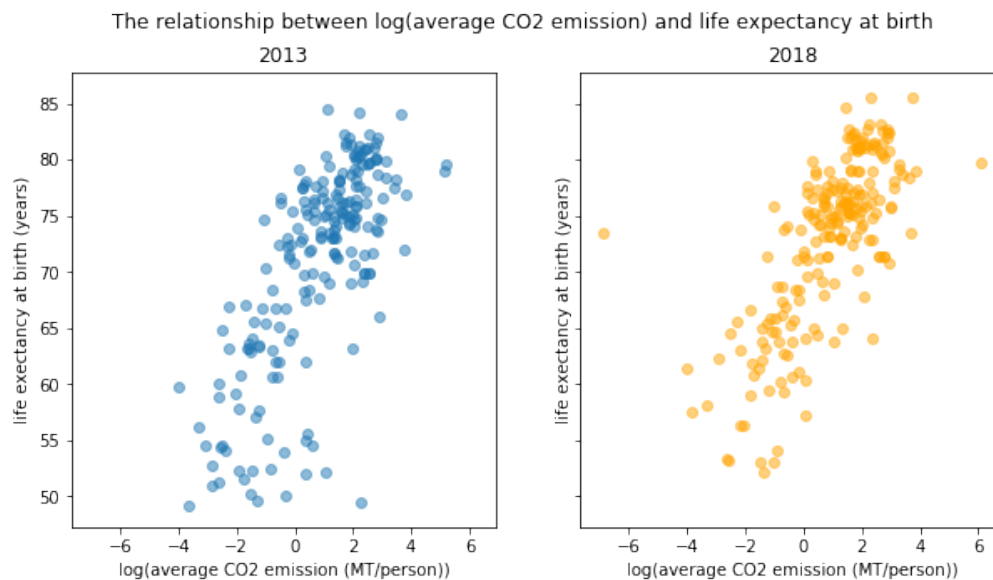
There are still a few (3) influential points that are so far from the majority of the data points.

Because we want to investigate the relationship, and the abnormal points is so few and are not very significant. We will try to “room in” and see how the most data points distributed.



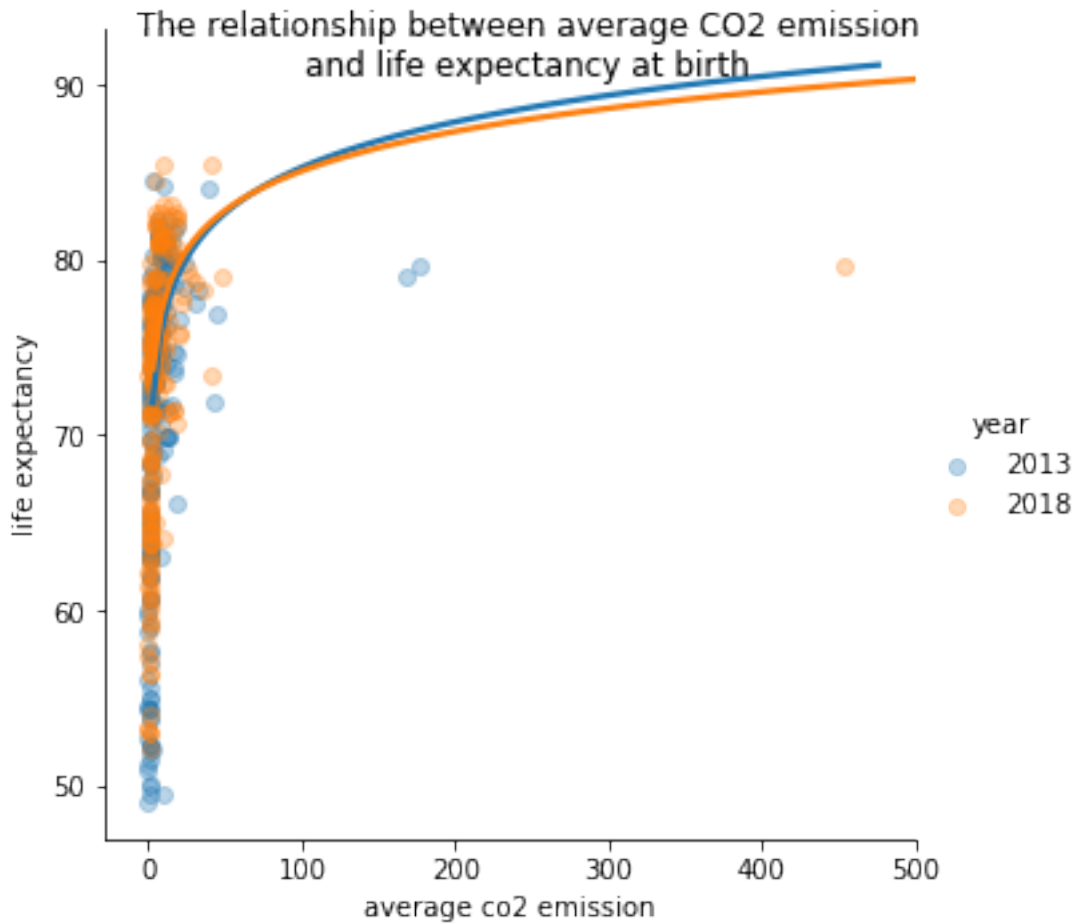
Generally the increase in CO2 emission is associated with the increase in life expectancy at birth. However, the relationship between the two variables obviously does not follow a linear line.

We can do a transformation on average CO2 emission and further investigate the relationship.

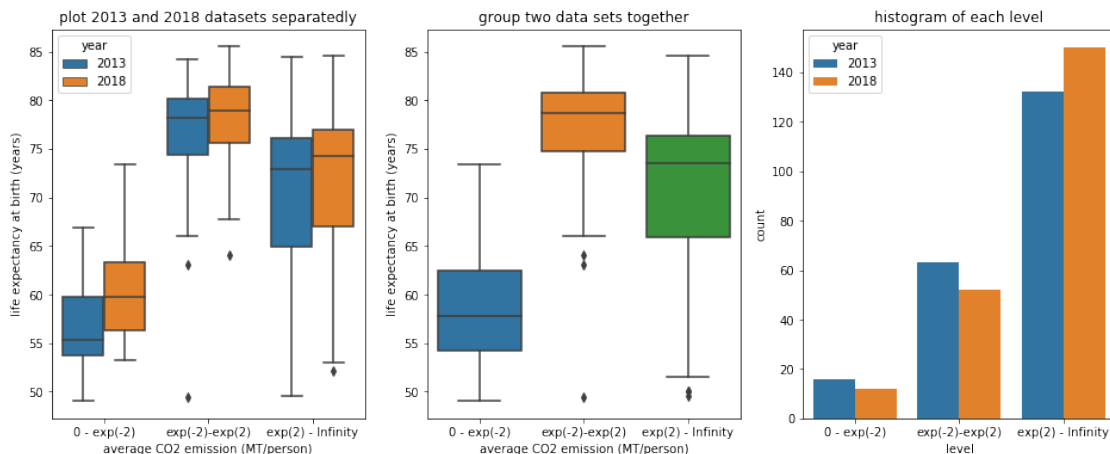


I try a log transformation on average CO2 emission, the resulted scatter plot is above. It seems that after the transformation, there is a positive linear relationship between the natural log of average CO2 emission and life expectancy at birth.

We can also see the plot with the fitted line below.

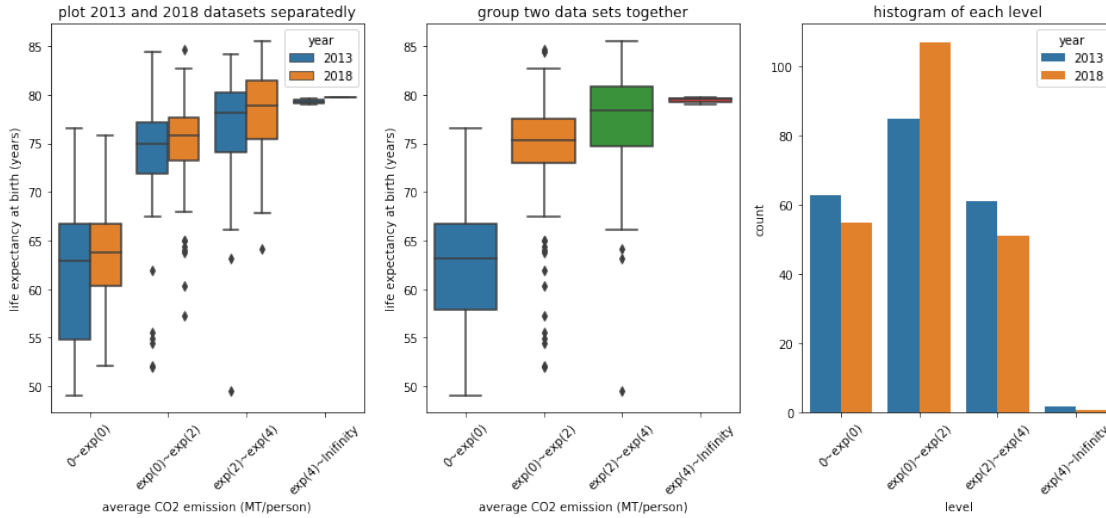


Because the result above suggest that there is a linear positive relationship after a natural log transformation. I categorize the data points with average CO2 emission into 3 intervals $0 - e^{-2}, e^{-2} - e^2, e^2 - \infty$ where e is the base of natural log, $e \approx 2.72$.



Unexpectedly, we observe that countries in interval $e^2 - \infty$ has lower mean of life expectancy than those of in level $e^{-2} - e^2$, and the histogram shows that the sample size is large. The side by side box plots seems to suggest that the increase in average CO2 emission will not associate with the increase in health expenditure. However, we also observe that there is a large difference between sample size in each level, we can try a different categorization.

Try the new categorization by the data points with average CO2 emission in between $0 - 1, 1 - e^2, e^2 - e^4, e^4 - \infty$.



country	average co2 emission	life expectancy
Virgin Islands	177.492195	79.61
Gibraltar	168.493010	78.98
Gibraltar	452.802009	79.70

Notice that for the first three levels, the sample size are close enough, and since the sample size of the last level is so small, they are not overall significant. Also, the mean of average CO2 emission in each level is positively associated with life expectancy at birth. As previous investigation suggests, the relationship is not linear.

Conclusion

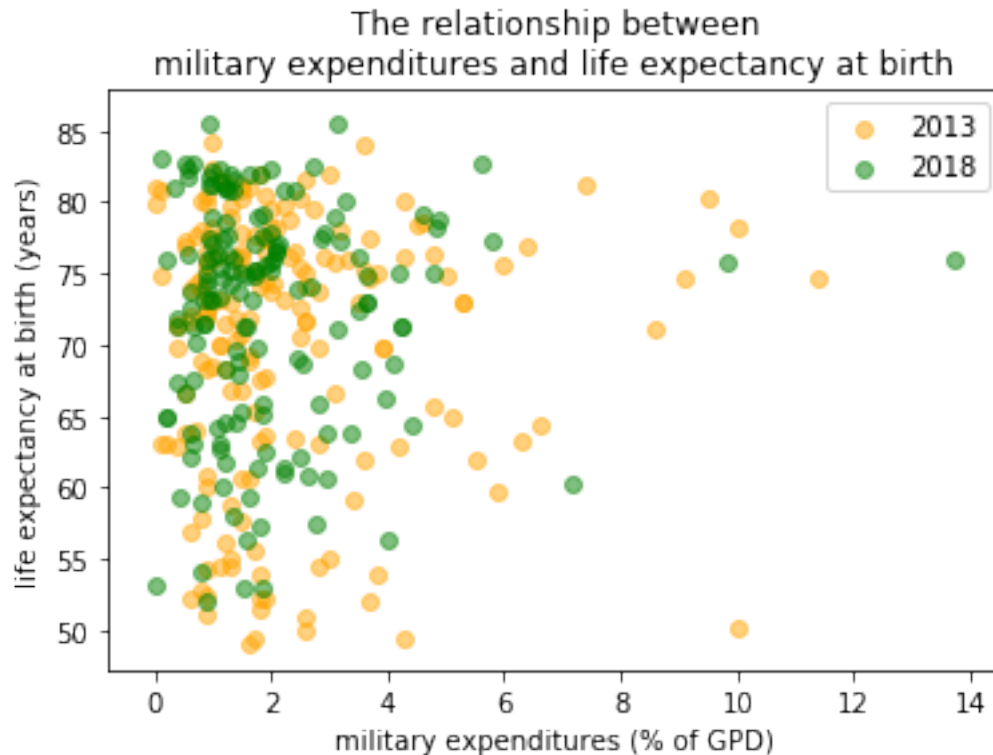
For the investigations above, we have some evidence that there is an positive association between average carbon dioxide emission from consumption of energy and life expectancy at birth. This relationship is likely to be linear between life expectancy at birth and the natural log of average CO2 emission.

Question 3

3.1 Is a countries' military expenditures associated with life expectancy at birth?

Investigation

First, we use a scatter plot to visually examine all the data points.

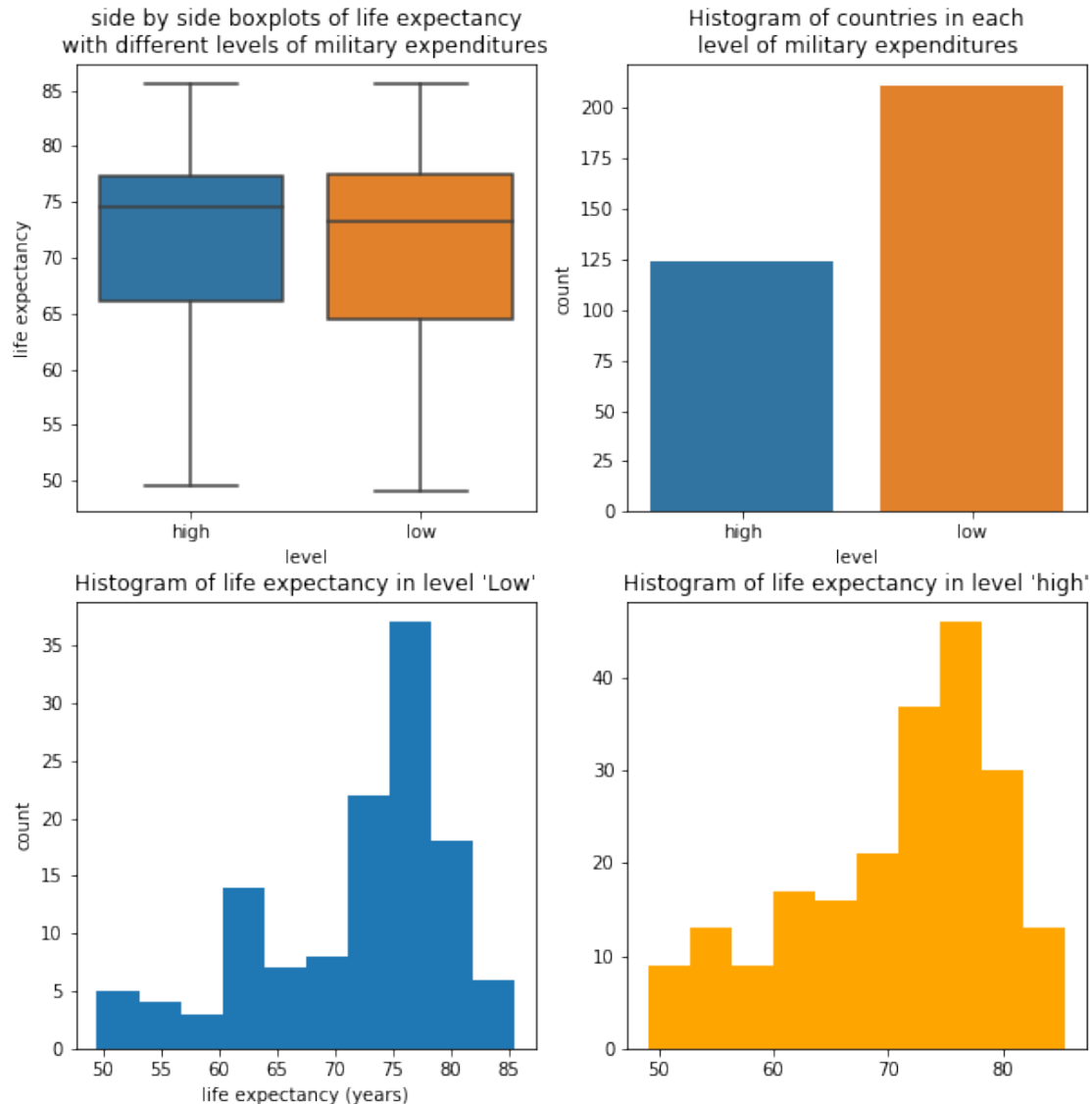


Looking at the scatter plot, we do not see any significant relationship between the two variables. Using Pearson correlation coefficient to evaluate the correlation between the two variables.

The Pearson correlation coefficient is: 0.016556740216071866

The Pearson correlation coefficient is very close to 0, therefore we have evidence that there is no evidence that life expectancy at birth is associated with military expenditures.

I divide the data points into two levels, the countries with military expenditures below 2% are in 'low' and the rest are 'high'. By this categorization, I expect that the two groups have a relatively equal size. Then, I further investigate the mean difference of life expectancy of the two groups.



From the side by side box plot, the mean of life expectancy of two groups are very close. Also, from the pair of histograms, we notice that the distribution of two groups are very similar.

We will further run a two sample t-test to investigate whether the means are equal.

The result of two sample t-test:

Test statistic: -0.958064378732529

p-value: 0.3387253529228582

Two sample t-test Null hypothesis: $H_0 : \mu_a = \mu_b$ Alternative hypothesis: $H_a : \mu_a \neq \mu_b$
 Where μ_x is the mean life expectancy of countries in 'low' group, μ_y is the mean life expectancy of countries in 'high' group. Test statistic: 0.96 p-value: 0.34

Since the p-value is large, we cannot reject the null hypothesis.

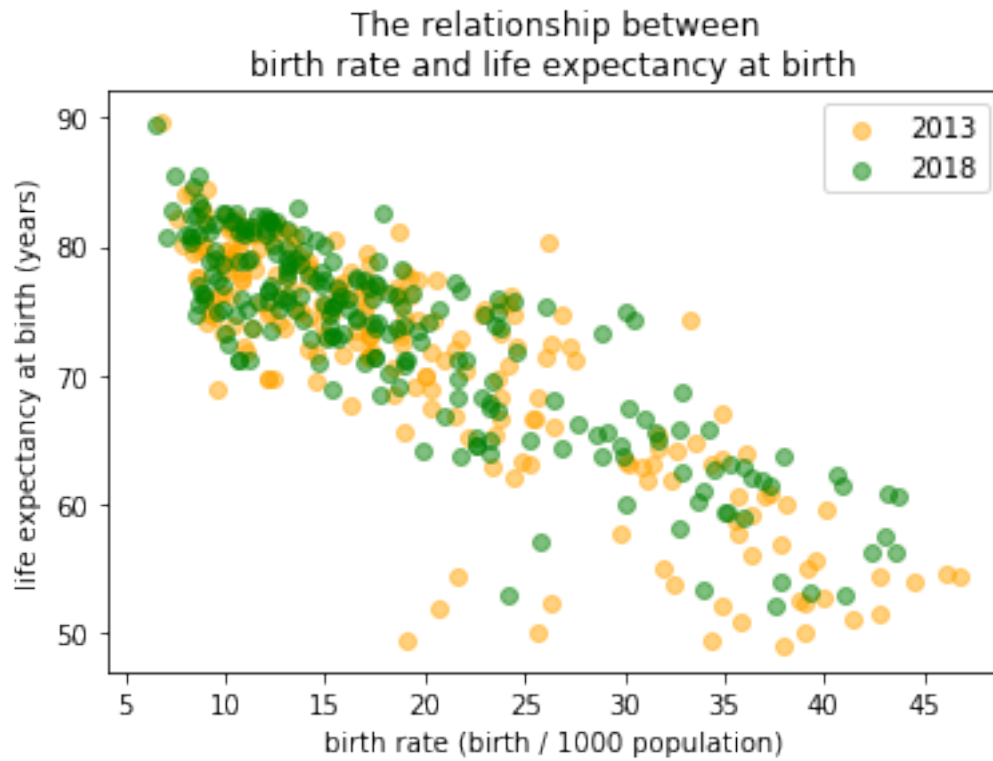
Conclusion

By the investigation above, we have some evidence that a country's military expenditures is not associated with life expectancy at birth.

3.2 Is a countries' birth rate associated with life expectancy at birth?

Investigation

First, we use a scatter plot to visually examine all the data points.



The scatter plot shows some evidence that there is a negative linear relationship between birth rate and life expectancy. We can further do a linear regression.

Linear regression for Year 2013:

slope: -0.7724902036388788

intercept: 86.78089698508221

rvalue: -0.8411658277065387

pvalue: 5.3560305325866175e-62

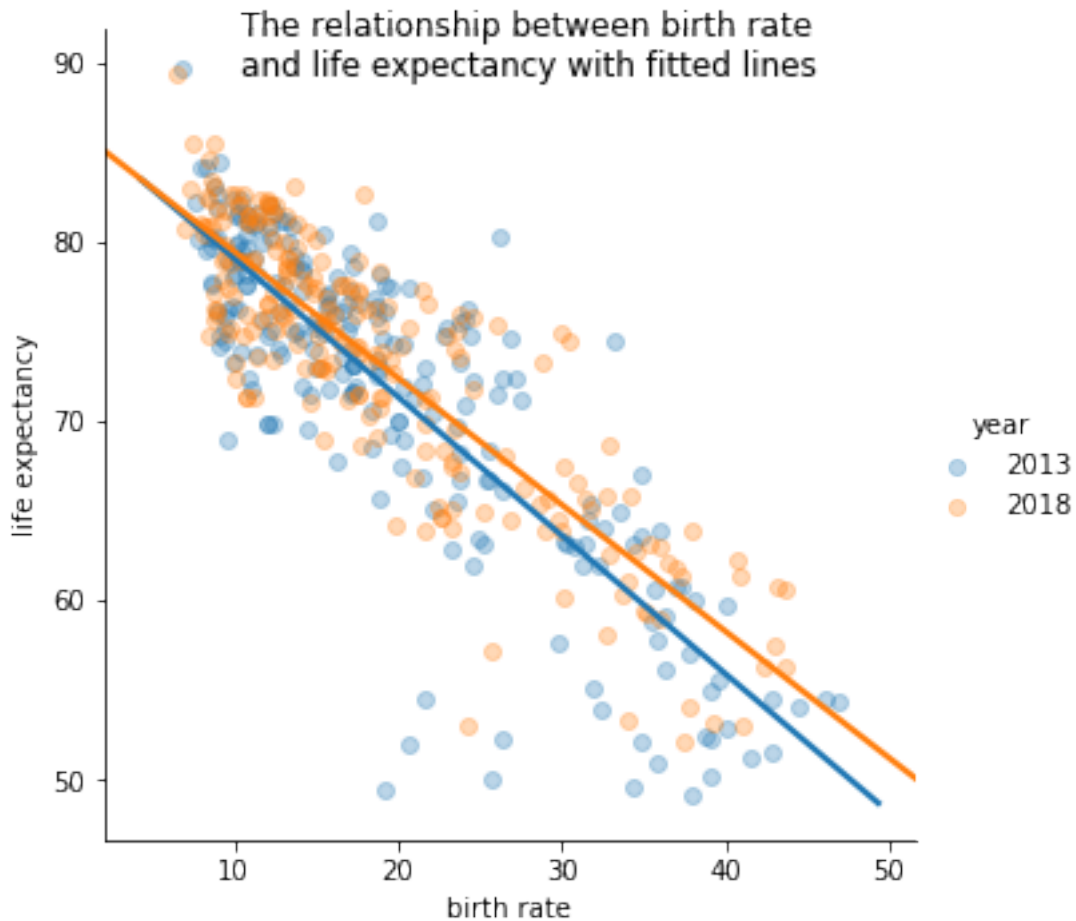
Linear regression for Year 2018:

slope: -0.7065512392511272

intercept: 86.51488234174981

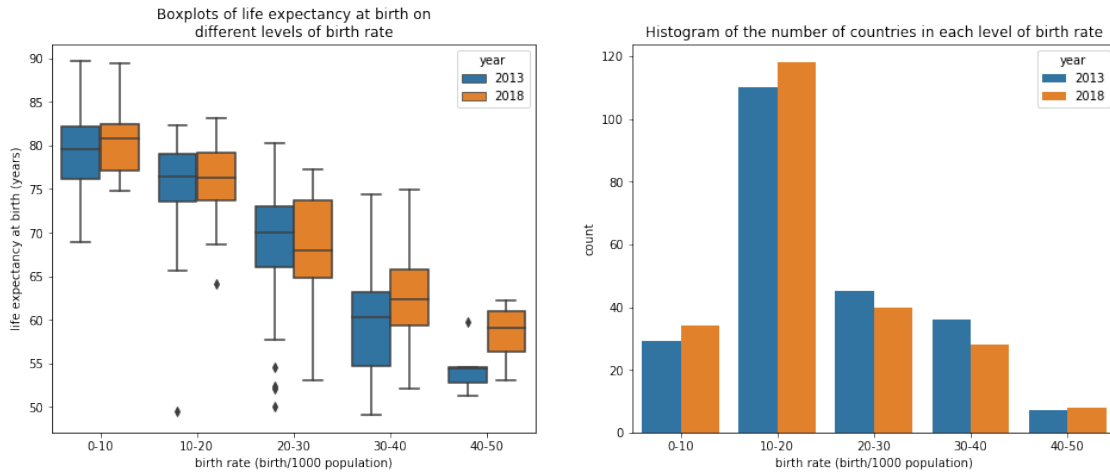
rvalue: -0.8625574819675528

pvalue: 8.280249892253233e-69



The fitted line has a negative slope. The hypothesis test gives p-value very close to 0 (5.35×10^{-62} , 8.28×10^{-69}), and the correlation coefficient are large (0.86, 0.86). Therefore, there is strong evidence that birth rate and life expectancy at birth are negatively associated.

Then, I will categorize data points using birth rate into 5 intervals: 0 – 10, 10 – 20, 20 – 30, 30 – 40, 40 – 50 because I suspect that the relationship is linear. This categorization is expected to give a relatively equal size for each level and a constant decrease in mean. Then, I do the side by side box-plots based on the categorization.



The side by side box plot shows that the relationship is negative, and the relationship is likely to be linear.

Conclusion

By the investigations, we have strong evidence that there is a negative linear relationship between birth rate and life expectancy at birth.

Conclusion and Discussion

In this report, I investigate the questions and conclude that:

1. Has global life expectancy at birth changed between 2013 to 2018? Which continents and countries have experienced the least and greatest changes?

Global life expectancy at birth improved between 2013 and 2018. The change had more influence on the countries of low life expectancy, and had little influence on the countries of high life expectancy. More specifically, the mean life expectancy had only a little improvement, but the lower quartile improved significantly.

Among the continents, Africa experienced the greatest changes, and North America experienced the least changes.

Among all countries, Gaza Strip experienced the least changes, and Gabon experienced the most changes.

2. Are a countries' health expenditures and carbon dioxide emissions from consumption of energy associated with life expectancy at birth?

A countries' health expenditures and carbon dioxide emissions from consumption of energy are likely to be associated with life expectancy at birth. This association is likely to be positive.

For the association between health expenditures and life expectancy at birth, the relationship is likely to be positive linear, the association is stronger for countries with less than 15% health expenditures.

For the association between carbon dioxide emission from consumption of energy, notice that we measure average carbon dioxide emission per person, and the relationship is likely to be positive and linear after a log transformation to average carbon dioxide emission.

3. Are a countries' military expenditures and birth rate associated with life expectancy at birth?

There is no evidence suggesting that a countries' military expenditures is associated with life expectancy at birth, while there is strong evidence suggesting that a countries' birth rate is associated with life expectancy at birth. The relationship between life expectancy at birth and birth rate is likely to be negative linear.

For the limitations and further improvements: - Although the collected datasets are assumed to be data from 2013 and 2018, the date of information shows that the assumption is not true. The problem is significant when I measure the average CO2 emission for Question 2.2, I suspect that the different date of information for CO2 emission and population may result in errors. - Further research should be done on the wars and/or natural disasters happened from 2013 to 2018. These incidents will have great influence on individual countries' life expectancy at birth. We may consider to remove them since they are not very significant in investigating relationships with other factors.