Change in Global Life Expectancy and Possible Correlating Factors

February 7, 2019

JSC270 Assignment 1: Change in Global Life Expectancy and Possible Correlating Factors

Shirley Wang, 1003946509

Feb 8 2019

1 Introduction

In this report, we will investigate if global life expectancy has changed between 2013 and 2018, and if there are specific factors that are associated with these results. We are using the CIA Factbook data available on the CIA Factbook website in this report. We will answer these questions:

1.1 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. Use data from the 2013 and 2018 World Factbook to investigate if a countries' health expenditures and carbon dioxide emissions from consumption of energy are associated with life expectancy at birth. Create categorical variables using health expenditures and carbon dioxide emissions. Do your results change?
3. Answer 2, except this time choose two other variables (i.e., not health expenditure or carbon dioxide emissions) from the CIA World Factbook to investigate if these variables are associated with life expectancy at birth. We will use Infant Mortality and External Debt.

2 Method

We will use the data from the CIA Factbook website detailing Life Expectancy at Birth in the years 2013 and 2018. We will also be looking at the variables: Health Expenditures, CO2 Emissions, Infant Mortality, and External Debt.

2.1 Question 1

We will analyze if global life expectancy at birth has changed between 2013 and 2018 by creating a new variable that represents the difference in life expectancy from 2013 to 2018, then observe

its distribution and see if it's majority is significantly different from zero. We will also plot the differences in every continent, to see if the distribution differs from region to region.

2.2 Question 2

We will investigate if health expenditures and CO2 emissions are associated with life expectancy at birth by observing the scatterplot of these variables and the corresponding regression line. Then we will create categorical variables corresponding to health expenditures and CO2 emissions, by splitting the variable into the categories 'high' and 'low', where values that are 'high' are above the median of all values in that dataset, and 'low' if below the median. Then we will observe the distribution of life expectancy in each category.

2.3 Question 3

We will investigate if infant mortality and external debt are associated with life expectancy at birth by observing the scatterplot of these variables and the corresponding regression line. Then we will create categorical variables corresponding to these variables by splitting the variable into the categories 'high' and 'low', where values that are 'high' are above the median of all values in that dataset, and 'low' if below the median. Then we will observe the distribution of life expectancy in each category.

2.3.1 2013 Dataframe

`Out[6]:`

	Country	Life Expectancy at Birth	Health Expenditures	CO2 Emissions
0	Monaco	89.63	4.3	NaN
1	Macau	84.46	NaN	1.805000e+06
2	Japan	84.19	9.3	1.181000e+09
3	Singapore	84.07	4.6	2.124000e+08
4	San Marino	83.12	7.2	NaN

`Out[7]:`

	Infant Mortality	External Debt	Continent
0	1.81	NaN	Europe
1	3.15	0.000000e+00	NaN
2	2.17	3.024000e+12	Asia
3	2.59	1.174000e+12	Asia
4	4.58	NaN	Europe

2.3.2 2018 Dataframe

`Out[9]:`

	Country	Life Expectancy at Birth	Health Expenditures	CO2 Emissions
0	Monaco	89.4	4.3	NaN
1	Japan	85.5	10.2	1.268000e+09
2	Singapore	85.5	4.9	2.495000e+08
3	Macau	84.6	NaN	2.563000e+06
4	San Marino	83.4	6.1	NaN

	Infant Mortality	External Debt	Continent
0	1.8	NaN	Europe
1	2.0	3.240000e+12	Asia
2	2.3	5.661000e+11	Asia
3	3.1	0.000000e+00	NaN
4	4.3	NaN	Europe

2.4   Exploring the Data

There are a number of missing values for each variable in both dataframes. As well, there are 16 more countries in the 2018 dataset, but also more missing values overall for it.
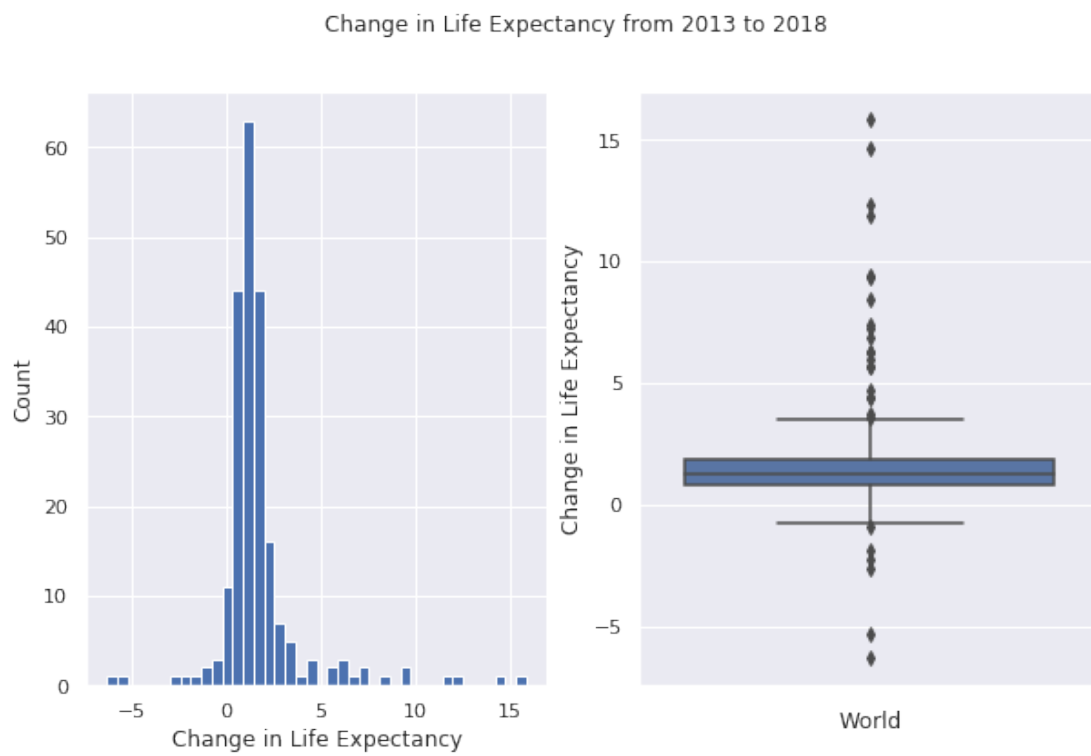
2.5   Change in Life Expectancy Data

	Country	Life Expectancy 2013	Life Expectancy 2018	Continent	Change
0	Monaco	89.63	89.4	Europe	-0.23
1	Macau	84.46	84.6	NaN	0.14
2	Japan	84.19	85.5	Asia	1.31
3	Singapore	84.07	85.5	Asia	1.43
4	San Marino	83.12	83.4	Europe	0.28

2.6 Question 1 Graphs

2.6.1 Figure 1.1

Change in Life Expectancy from 2013 to 2018

2.6.2 Figure 1.2

Change in Life Expectancy by Continent
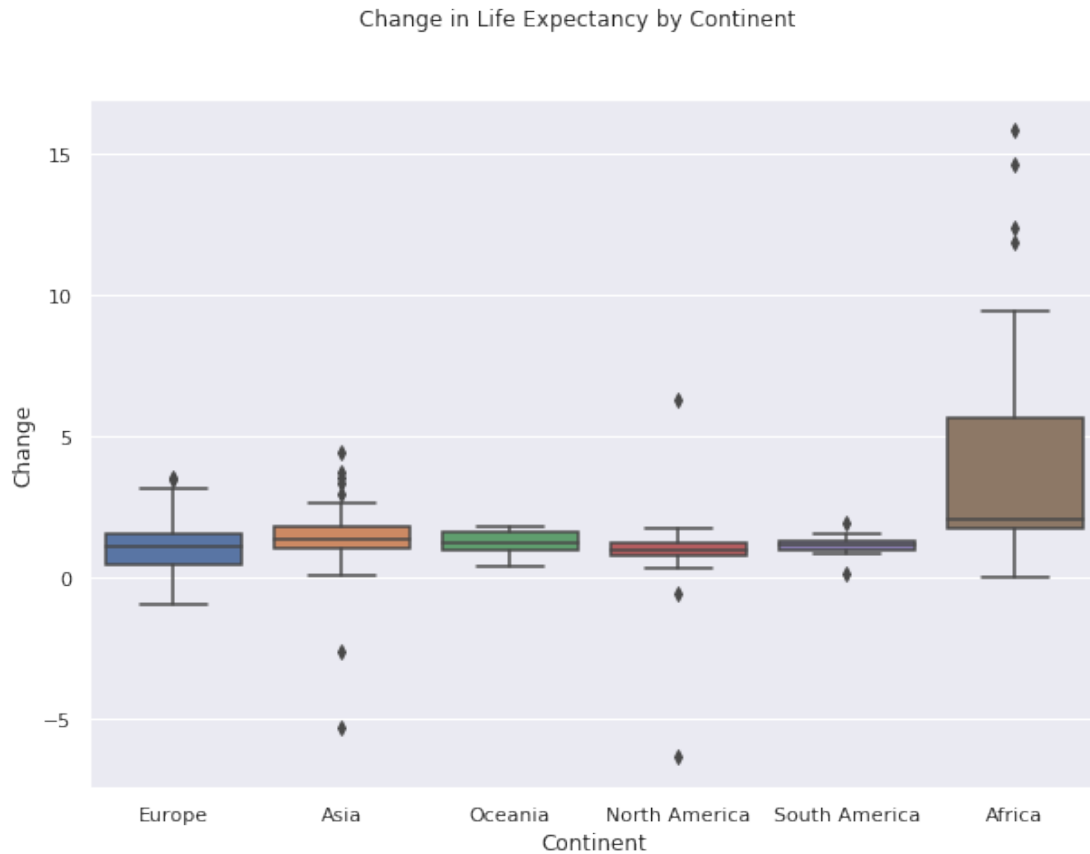
2.6.3 Figure 1.3

|	Country	Life Expectancy 2013	Life Expectancy 2018	Continent	Change
213	Gabon	52.15	68.0	Africa	15.85
214	Namibia	52.03	64.4	Africa	12.37
220	Guinea-Bissau	49.50	61.4	Africa	11.90
221	South Africa	49.48	64.1	Africa	14.62

2.6.4 Figure 1.4

	Country	Life Expectancy 2013	Life Expectancy 2018	Continent	Change
28	Jordan	80.30	75.0	Asia	-5.30
61	Dominican Republic	77.62	71.3	North America	-6.32

2.7 Question 2 Graphs

2.7.1 Figure 2.1

Life Expectancy vs Health Expenditures

2.7.2 Figure 2.2

Life Expectancy vs CO2 Emissions
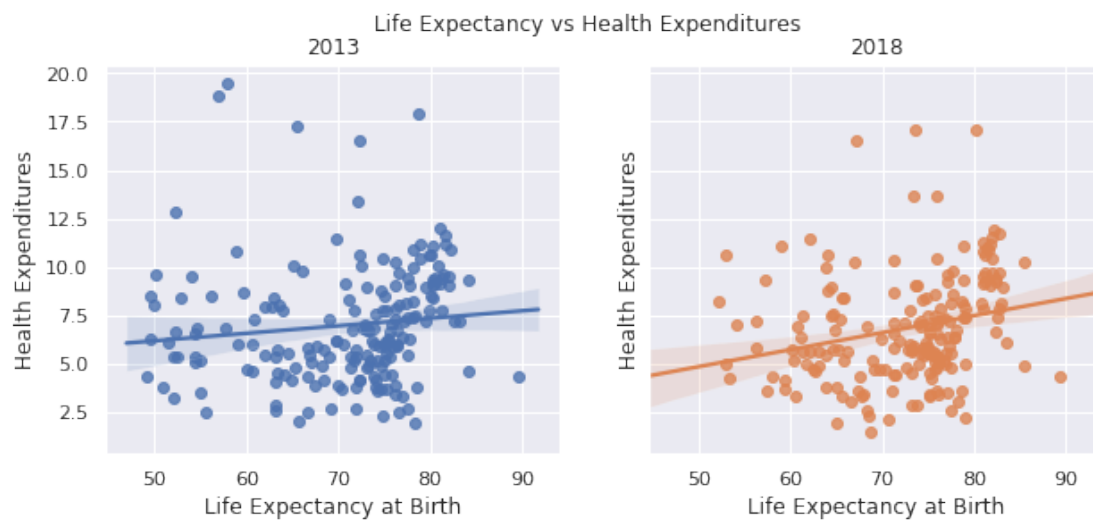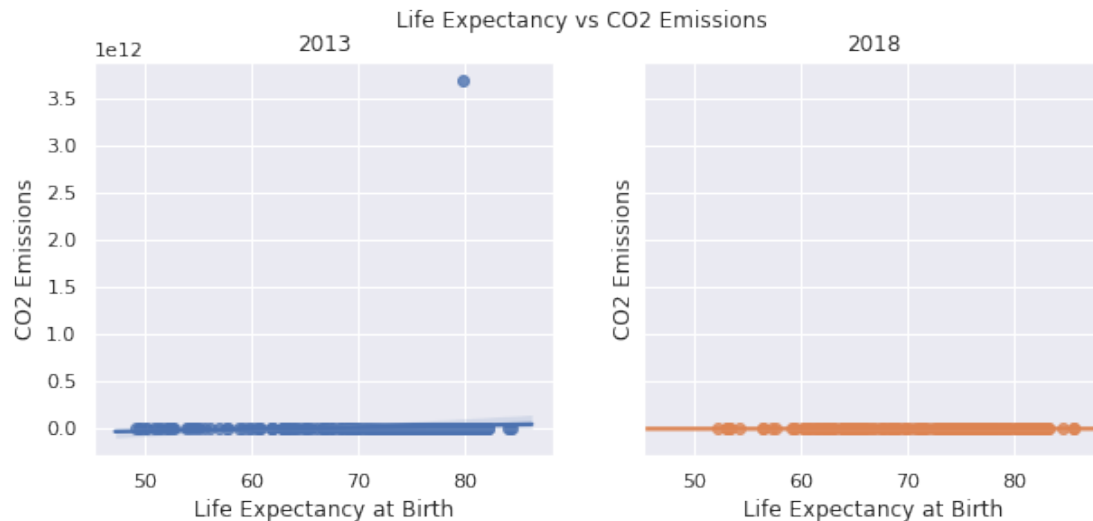
We appear to have an outlier in the 2013 dataset. Investigating this further: `Out[19]:`

	Country	CO2 Emissions
36	European Union	3.700000e+12

The European Union is not a country, it's a union of European countries. Then there are two possibilities. One, the CO2 Emissions here is actually the sum of all CO2 emissions of all the

European countries. Two, this is a typo. Either way, since the European Union is technically not a country, we should be fine excluding it from our analysis.
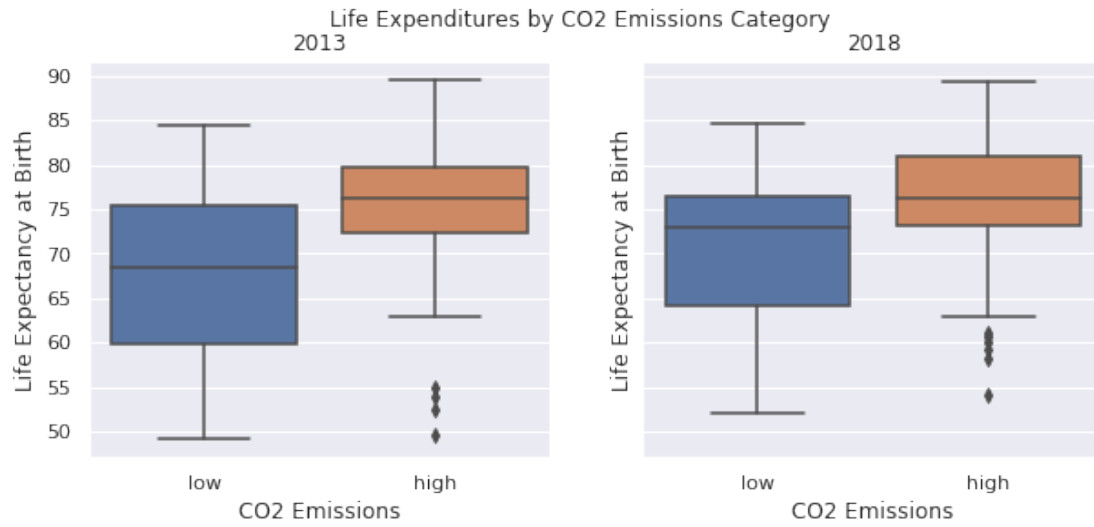
2.7.3   Figure 2.3

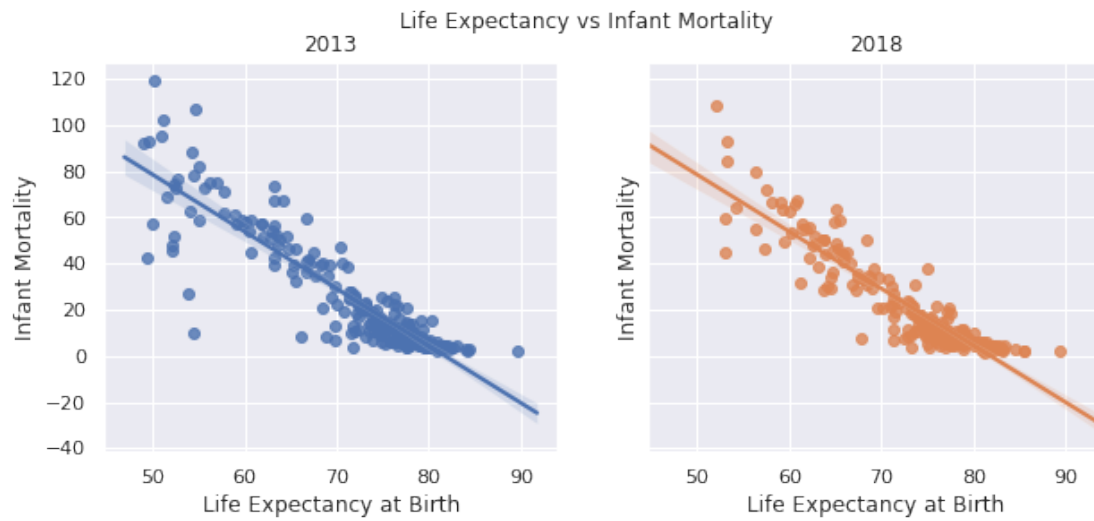Life Expectancy vs CO2 Emissions
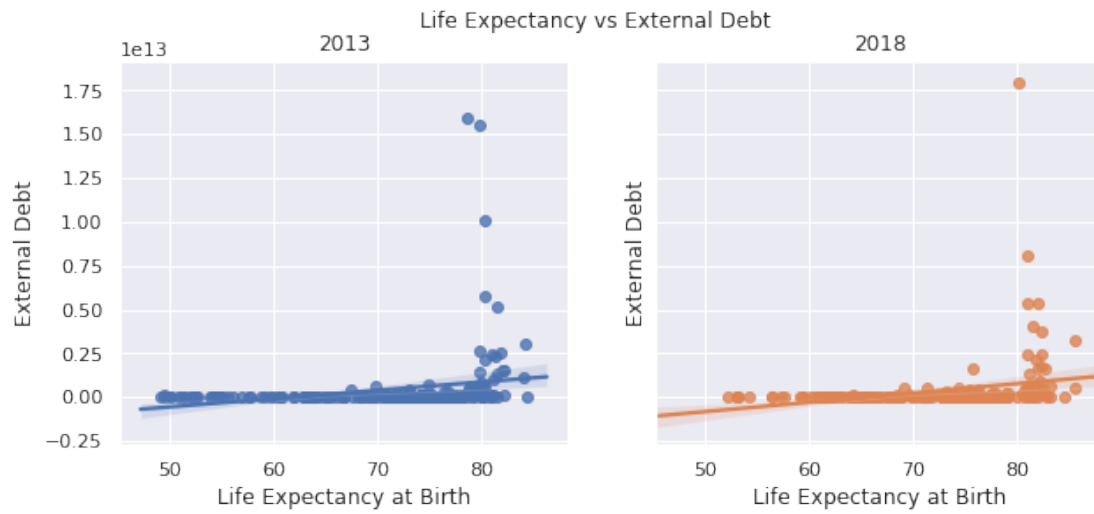
2.7.4   Figure 2.4

Life Expenditures by Health Expenditures Category

2.7.5 Figure 2.5

Life Expenditures by CO2 Emissions Category
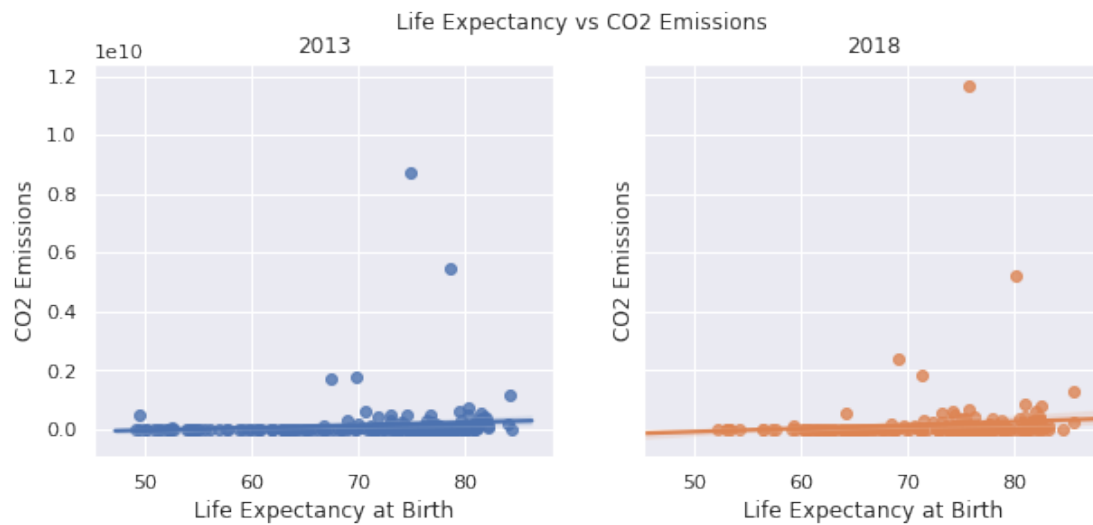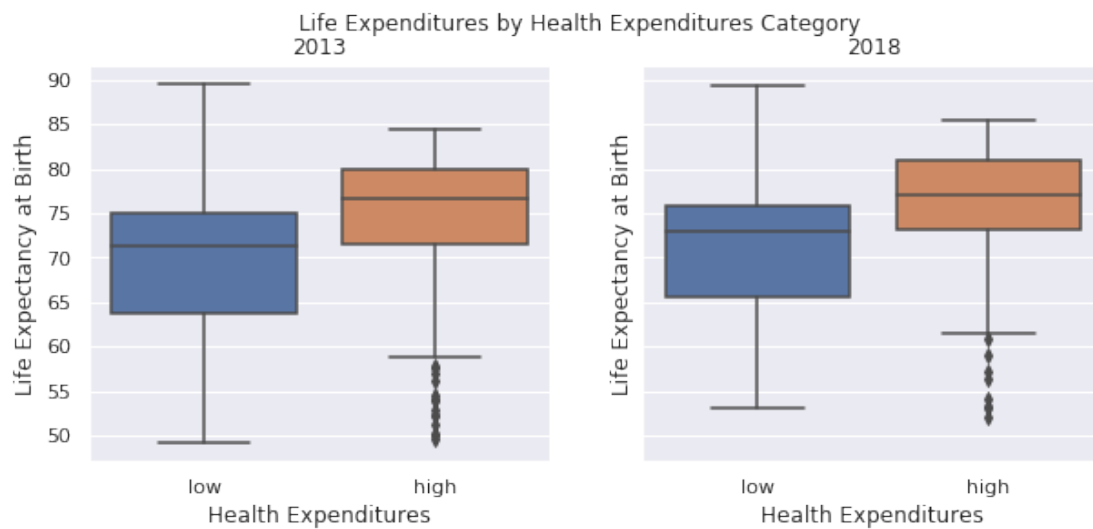
2.8 Question 3 Graphs

2.8.1 Figure 3.1

Life Expectancy vs Infant Mortality

2.8.2 Figure 3.2

Life Expectancy vs External Debt

2.8.3 Figure 3.3

Life Expenditures by Infant Mortality Category

Life Expenditures by External Debt Category
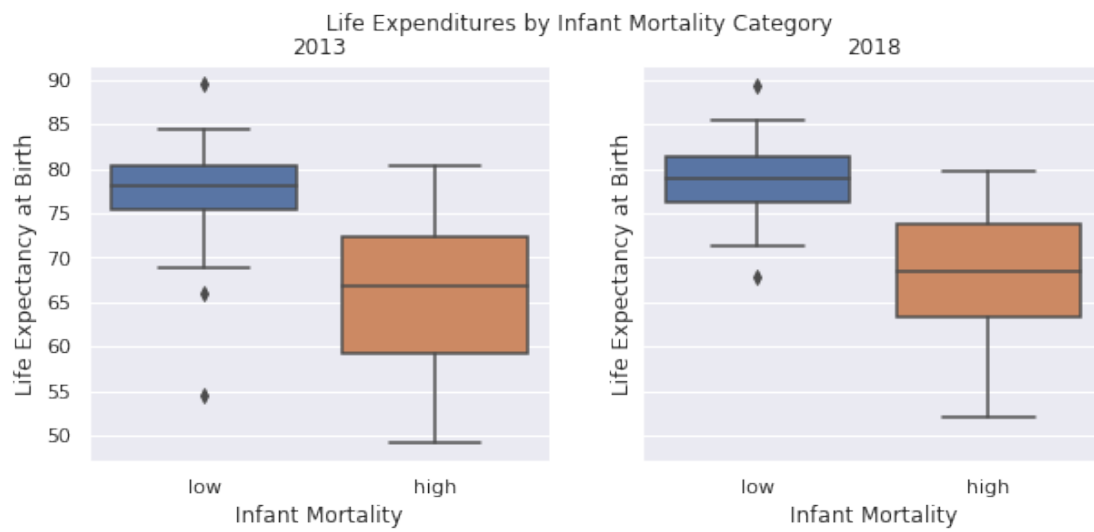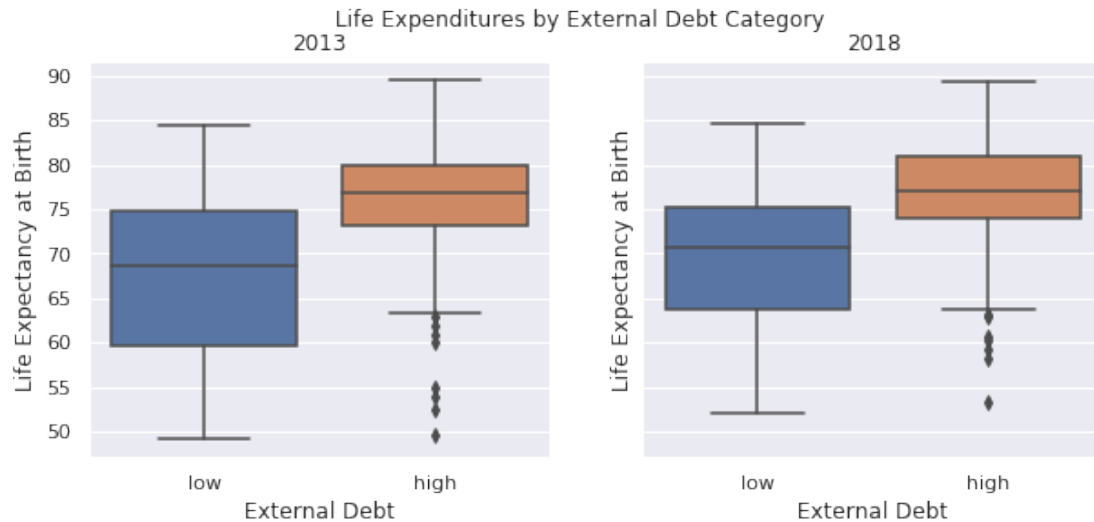
3   Results

3.1   Question 1

From Figure 1.1, looking at the histogram, we can see that the mode of values is above zero, and the distribution is slightly skewed right. Looking at the boxplot, the interquartile range of the boxplot is entirely above zero, meaning at least 75% of countries had a positive change in life expectancy. This is significantly different from zero. Both of these graphs imply that global life expectancy at birth has slightly increased from 2013 to 2018.

From Figure 1.2, the boxplot of changes in life expectancy for each continent, we can see that Africa had overall the largest positive changes in life expectancy, with all its countries having a positive life expectancy and many having very large changes. Europe has the lowest overall change in life expectancy, with a very low median and whiskers that expended below 0. But Asia and North America had the countries that experienced the largest negative change in life expectancy.

From Figure 1.3 and 1.4, looking at the extreme data values, the countries with the largest positive and negative changes also originate from the continents that experienced these large changes, with the largest positive change in Gabon in Africa at 15.85 years and the largest negative change in the Dominican Republic in North America at -6.32 years.

3.2   Question 2

From Figure 2.1, looking at the scatterplot of Health Expenditures vs Life Expectancy, for 2013, it appears that there might be a positive correlation between Life Expectancy and Health Expenditures. However, the confidence interval around the regression line also has the possibility of being horizontal, so we can't be sure. For 2018, a positive correlation seems more obvious and probable, since the confidence interval doesn't include the possibility of the regression line being horizontal.

From Figure 2.4, looking at Health Expenditures from the perspective of the categories high and low, for 2013, the median of the high category is higher than the low, as well as skinner. Its median is higher than the low category, however the low category's distribution is more spread out than the high. Since the low category's distribution is wide enough to cover all the high values, we still can't make any conclusions here. For 2018, it has the same problem, where the high category's interquartile range is somewhat higher than the low category's, but the range of the low category also covers the range of the high category.

From Figure 2,2, looking at the scatterplot of CO2 Emissions vs Life Expectancy, for 2013, it appears that there might be a positive correlation between Life Expectancy and CO2 Emissions, although the few countries with very high values makes it hard to be sure. For 2018, the same holds. There appears to be a positive correlation, although the positive slope in the regression line might be due to the very high outliers.

From Figure 2.5, looking at CO2 Emissions from the perspective of the categories high and low, for 2013, since the boxplot's entire distribution of the CO2 Emissions high category is shifted higher than the low category, it does seem that countries with high CO2 emissions also have higher life expectancy at birth. For 2018, the same holds. So there is some evidence that there is a positive correlation between CO2 Emissions and life expectancy at birth.

3.3 Question 3

From 3.1, looking at the scatterplot of Infant Mortality vs Life Expectancy, for 2013, we have a clear negative correlation between them. The regression line is clearly negatively sloped with a small confidence interval. For 2018, the confidence interval is even narrower, so still have its clear negative correlation.

From 3.3, looking at Infant Mortality from the perspective of the categories high and low, for 2013, we see more support for the idea that infant mortality and life expectancy are negatively correlated, since the interquartile range of the high infant mortality category is entirely below the low infant mortality category. For 2018, the same is true.

From 3.2, looking at the scatterplot of External Debt vs Life Expectancy, for 2013, there appears to be a positive correlation between them. Even the very high values occur with high life expectancies, so we can't be sure that they're typos. For 2018, the same holds. There appears to be a positive correlation, but it could be due to the very high positive outliers affecting the regression line's slope.

From 3.4, looking at External Debt from the perspective of the categories high and low, for 2013, we see support for the idea that there is a positive correlation between external debt and life expectancy, since the distribution of the high category is clustered higher than majority of the low category. But there is still a lot of overlap between the two distributions, and lots of low value outliers in the high category that would be explained if there was no difference in the two categories, so we can't be sure. For 2018, we get the same result.

4 Conclusion

We conclude that global life expectancy at birth has experienced a positive change from 2013 to 2018, since the distribution of the change in life expectancy is significantly above zero.

Africa had the overall largest positive changes in life expectancy, and Europe had the overall smallest change in life expectancy, but there were certain countries in North America and Asia that had very large negative changes in life expectancy, but those could be considered outliers, since

the other countries in these continents all experienced a small positive change. The countries with the largest positive and negative changes also originate from these continents respectively, with the largest positive change in Gabon in Africa at 15.85 years and the largest negative change in the Dominican Republic in North America at -6.32 years. Interestingly, the countries that experienced the largest positive changes in life expectancy all originally had low life expectancies. This could be a sign that living conditions are steadily improving in poorer countries.

For Health Expenditures, the regression line and boxplots suggest a possible positive correlation, but it isn't very strong. We conclude that there is some evidence suggesting that Health Expenditures could be positively correlated with life expectancy, but we can't make a solid conclusion about it. We can think of this result as possibly being the related to countries with free healthcare and countries without. Investigating if countries with public healthcare systems spend more or less on health expenditures, and if that in turn affects life expectancy if healthcare is available to all is something that could be studied in the future.

For CO2 Emissions, the regression line and boxplots suggest a positive correlation. We conclude that there is evidence suggesting that CO2 Emissions and life expectancy are positively correlated. For CO2 Emissions in the scatter plot, we see a general positive trend, and very high outliers for countries with high life expectancy. This could be related to GDP, where countries with high GDP and manufactures lots of things as a result have better living conditions and higher CO2 Emissions.

For Infant Mortality, the regression line and boxplots suggest a strong negative correlation. We conclude that there is very strong evidence suggesting that Infant Mortality and life expectancy are negatively correlated. This makes sense, since countries whose healthcare allows less babies to die at childbirth in turn have better healthcare overall, and people probably live longer with better healthcare.

For External Debt, the regression line and boxplots suggest a positive correlation. We conclude that there is evidence suggesting a positive correlation between External Debt and life expectancy. This might be due to countries that have higher GDP will have a higher external debt, and since there is more money more can be invested into better healthcare. This could be something to investigate in the future.

Our data is entirely dependent on the accuracy of the CIA Factbook. The factbook tries its best to update with each year, but for some countries it has missing values, and for some countries it can't properly update the values every year, so they might be outdated. This could have lead to mistakes in our analysis.