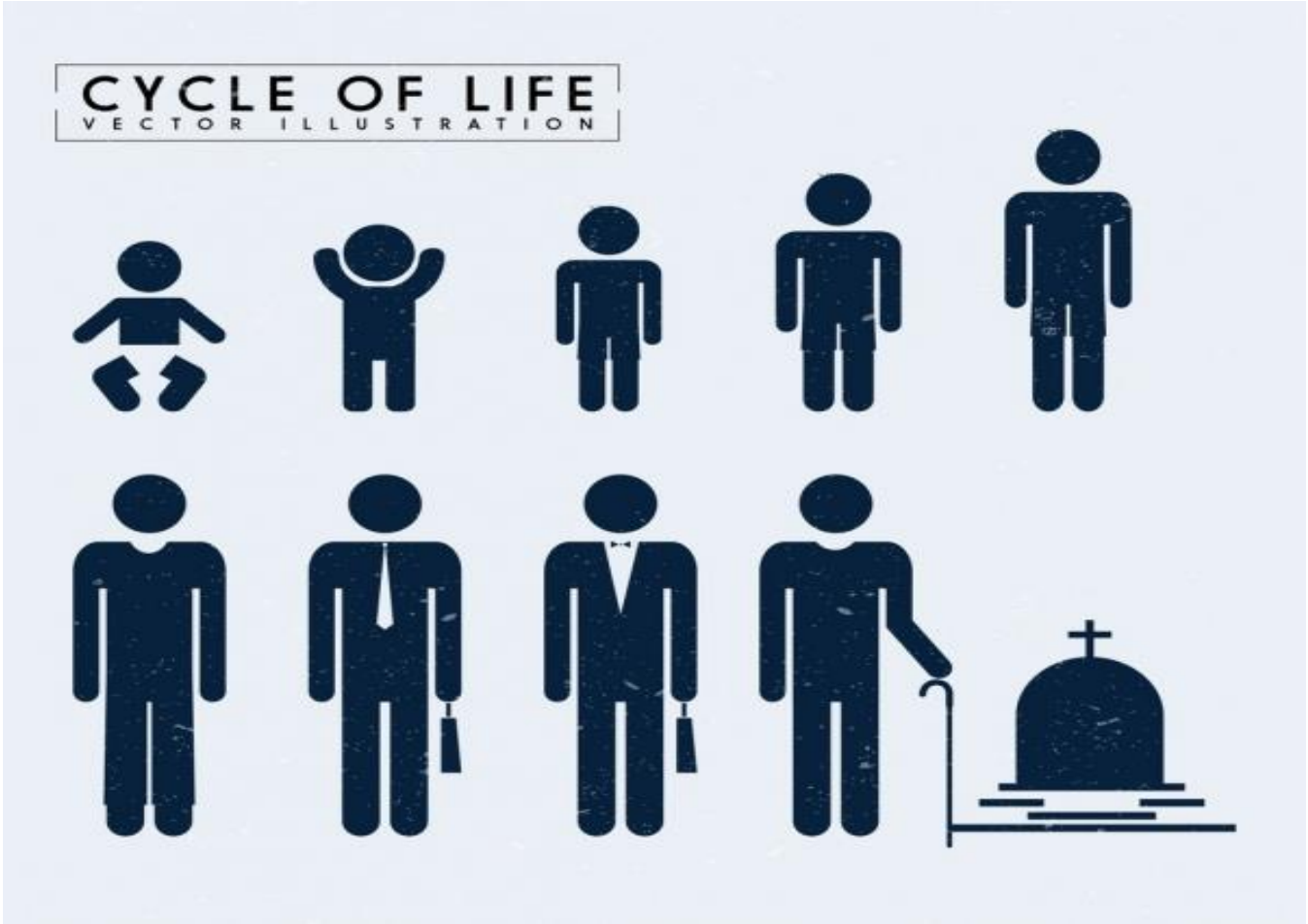
Effects Of Several Variables On Life Expectancy between 2000 and 2015



By Nigel Karikari

Goal Of This Project

- To discover how several variables effect life expectancy.
- To analyse the relationships between the variables.

Description Of The Data

The main variables of focus in the data will be Life expectancy, Adult Mortality, Infant deaths, Hepatitis B, Measles, BMI, under-five deaths, HIV/AIDS, GDP, Income composition of resources and Schooling. The timeframe of the data is between 2000 and 2015.

Introduction

R and Microsoft Excel were initially used to filter the data into developing countries, developed countries and then Ghana.

Analysis

Beginning the analysis on the main data set, We start out by looking at summary statistics and data types of the fields:

	Country	Year	...	Income composition of resources	Schooling
0	Afghanistan	2015	...	0.479	10.1
1	Afghanistan	2014	...	0.476	10.0
2	Afghanistan	2013	...	0.470	9.9
3	Afghanistan	2012	...	0.463	9.8
4	Afghanistan	2011	...	0.454	9.5

[5 rows x 14 columns]

	Country	Year	...	Income composition of resources	Schooling
2933	Zimbabwe	2004	...	0.407	9.2
2934	Zimbabwe	2003	...	0.418	9.5
2935	Zimbabwe	2002	...	0.427	10.0
2936	Zimbabwe	2001	...	0.427	9.8
2937	Zimbabwe	2000	...	0.434	9.8

[5 rows x 14 columns]

<class 'pandas.core.frame.DataFrame'>

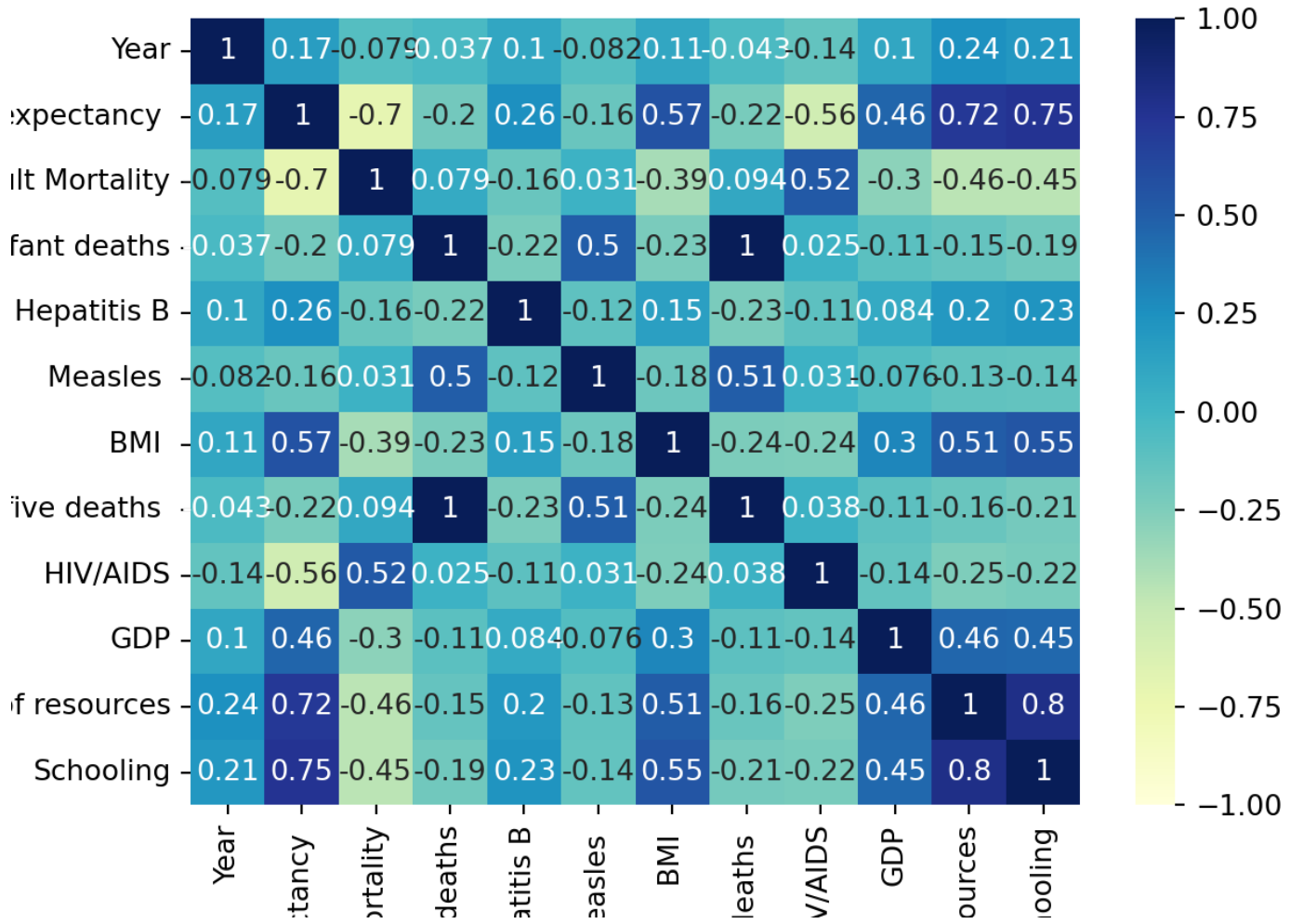
RangeIndex: 2938 entries, 0 to 2937

Data columns (total 14 columns):

#	Column	Non-Null Count	Dtype
0	Country	2938 non-null	object
1	Year	2938 non-null	int64
2	Status	2938 non-null	object
3	Life expectancy	2928 non-null	float64
4	Adult Mortality	2928 non-null	float64
5	infant deaths	2938 non-null	int64
6	Hepatitis B	2385 non-null	float64
7	Measles	2938 non-null	int64
8	BMI	2904 non-null	float64
9	under-five deaths	2938 non-null	int64
10	HIV/AIDS	2938 non-null	float64
11	GDP	2490 non-null	float64
12	Income composition of resources	2771 non-null	float64
13	Schooling	2775 non-null	float64

dtypes: float64(8), int64(4), object(2)
memory usage: 321.5+ KB

One way to analyse the relationships between variables is using correlation. The correlation heatmap will be utilized:



Looking at the figures, the following conclusions were made about relationships between the other variables and life expectancy (excluding year):

- First of all, a positive correlation between variables means that the variables increase at the same time, whereas a negative correlation means when one variable decreases, the other variable increases and vice-versa.
- A strong positive correlation (of 0.75) exists between schooling and life expectancy. This means when schooling increases, life expectancy would also increase in a similar fashion.
- A strong negative correlation (of -0.7) exists between the variables, adult mortality and life expectancy. This simply means when the value of one variable increases, the value of the other variable would decrease, and vice-versa.
- Again, a strong negative correlation exists between HIV/AIDS and life expectancy.
- Looking at the other variables, A positive correlation exists between the following variables and life expectancy in the descending order of magnitude: income composition of resources, BMI, GDP and Hepatitis B.
- A weak negative correlation also exists between life expectancy and the following variables in descending order of magnitude: Measles, infant deaths and under-five deaths.

Let us shift our attention to the developed countries, starting out with the summary statistics:

	Country	Year	...	Income.composition.of.resources	Schooling
0	Australia	2015	...	0.937	20.4
1	Australia	2014	...	0.936	20.4
2	Australia	2013	...	0.933	20.3
3	Australia	2012	...	0.930	20.1
4	Australia	2011	...	0.927	19.8

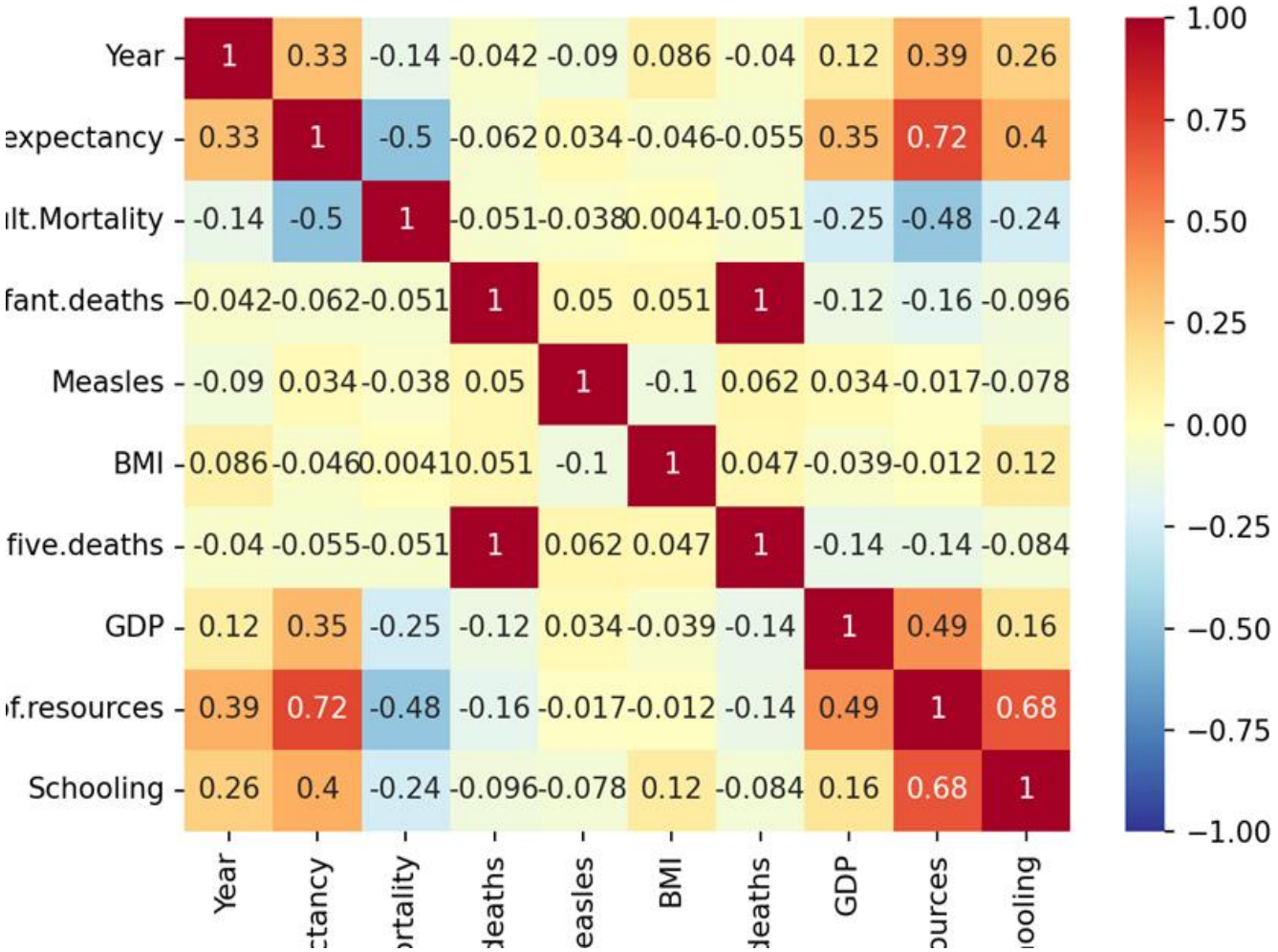
[5 rows x 12 columns]

	Country	Year	...	Income.composition.of.resources	Schooling
491	United States of America	2004	...	NaN	NaN
492	United States of America	2003	...	NaN	NaN
493	United States of America	2002	...	NaN	NaN
494	United States of America	2001	...	NaN	NaN
495	United States of America	2000	...	NaN	NaN

[5 rows x 12 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 496 entries, 0 to 495

```
Data columns (total 12 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Country                         496 non-null    object
 1   Year                            496 non-null    int64
 2   Status                          496 non-null    object
 3   Life.expectancy                 496 non-null    float64
 4   Adult.Mortality                 496 non-null    int64
 5   infant.deaths                   496 non-null    int64
 6   Measles                         496 non-null    int64
 7   BMI                             496 non-null    float64
 8   under.five.deaths               496 non-null    int64
 9   GDP                             448 non-null    float64
 10  Income.composition.of.resources  464 non-null    float64
 11  Schooling                       464 non-null    float64
dtypes: float64(5), int64(5), object(2)
memory usage: 46.6+ KB
```

Just like the correlation heatmap for the entire data set was plotted, the same would be done for the developed countries:



Note: The Hepatitis B and HIV/AIDS variables were removed from the developed countries data set.

Conclusions on the correlation heatmap:

- A strong negative correlation (of -0.5) exists between adult mortality and life expectancy in developed countries. This means that, as adult mortality increases, life expectancy decreases and vice-versa.
- A strong positive correlation (of 0.72) exists between income composition of resources and life expectancy. This implies that income composition of resources and life expectancy increase simultaneously in a similar fashion.

Again, let us shift our focus to the developing countries.

	Country	Year	...	Income.composition.of.resources	Schooling
0	Afghanistan	2015	...	0.479	10.1
1	Afghanistan	2014	...	0.476	10.0
2	Afghanistan	2013	...	0.470	9.9
3	Afghanistan	2012	...	0.463	9.8
4	Afghanistan	2011	...	0.454	9.5

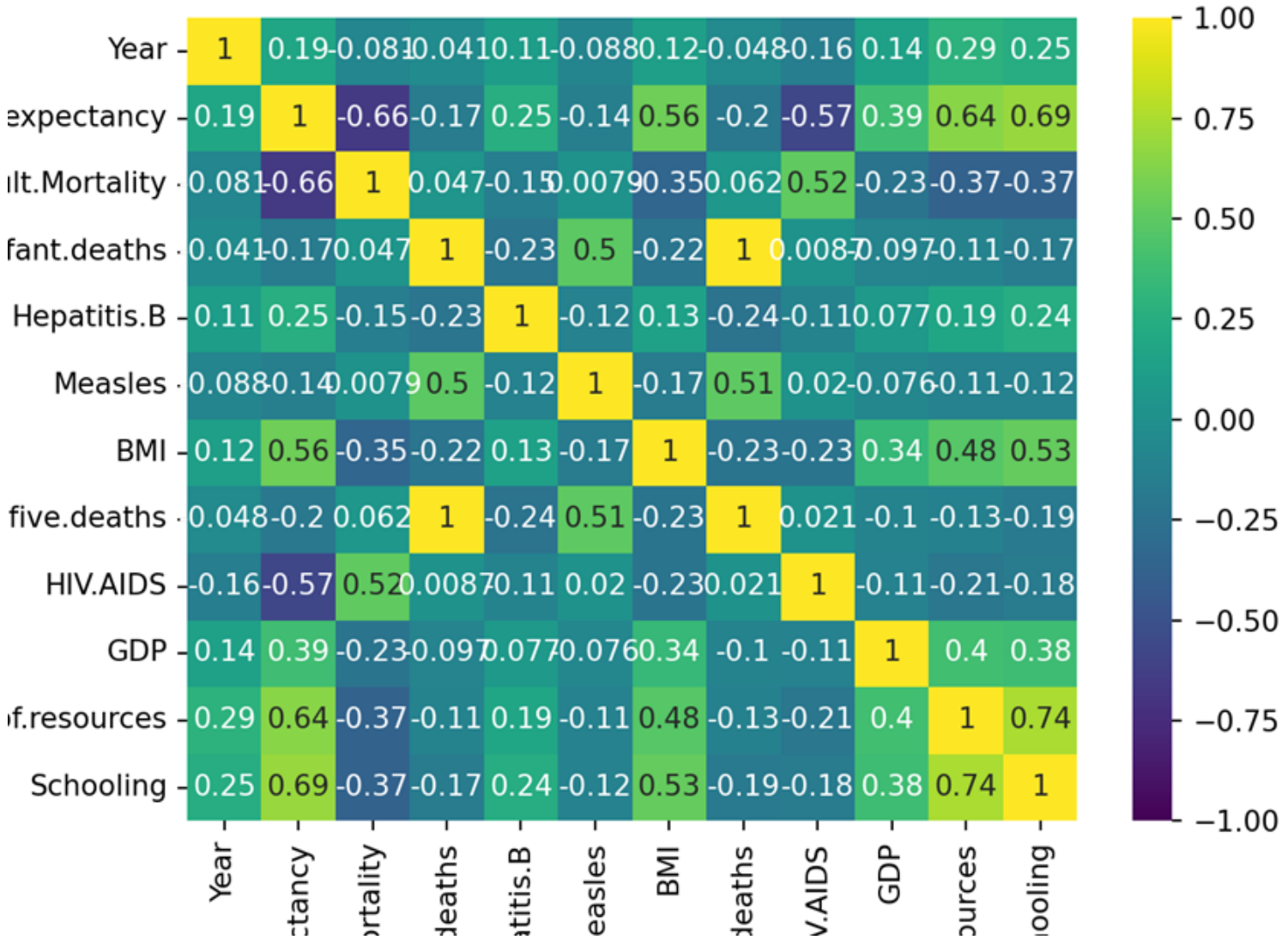
[5 rows x 14 columns]

	Country	Year	...	Income.composition.of.resources	Schooling
2421	Zimbabwe	2004	...	0.407	9.2
2422	Zimbabwe	2003	...	0.418	9.5
2423	Zimbabwe	2002	...	0.427	10.0
2424	Zimbabwe	2001	...	0.427	9.8
2425	Zimbabwe	2000	...	0.434	9.8

[5 rows x 14 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2426 entries, 0 to 2425

```
Data columns (total 14 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Country                         2426 non-null   object
 1   Year                            2426 non-null   int64
 2   Status                          2426 non-null   object
 3   Life.expectancy                 2416 non-null   float64
 4   Adult.Mortality                 2416 non-null   float64
 5   infant.deaths                   2426 non-null   int64
 6   Hepatitis.B                     2046 non-null   float64
 7   Measles                         2426 non-null   int64
 8   BMI                             2392 non-null   float64
 9   under.five.deaths               2426 non-null   int64
 10  HIV.AIDS                        2426 non-null   float64
 11  GDP                             2042 non-null   float64
 12  Income.composition.of.resources  2307 non-null   float64
 13  Schooling                       2311 non-null   float64
dtypes: float64(8), int64(4), object(2)
memory usage: 265.5+ KB
```

Using the correlation heatmap again:



Conclusions on the correlation map of developing countries:

- A negative correlation exists between the variables HIV/AIDS and Life expectancy. This signifies an inverse relationship as one variable decreases when the other variable increases and vice-versa.
- Again, adult mortality has a negative correlation (-0.66) with life expectancy.
- A positive correlation (0.69), exists between schooling and life expectancy. A similar value (0.64) was observed between income composition of resources and life expectancy.

Finally, lets shift our attention to Ghana

	Country	Year	...	Income.composition.of.resources	Schooling
0	Ghana	2015	...	0.575	11.4
1	Ghana	2014	...	0.576	11.7
2	Ghana	2013	...	0.570	11.5
3	Ghana	2012	...	0.563	11.2
4	Ghana	2011	...	0.554	10.9

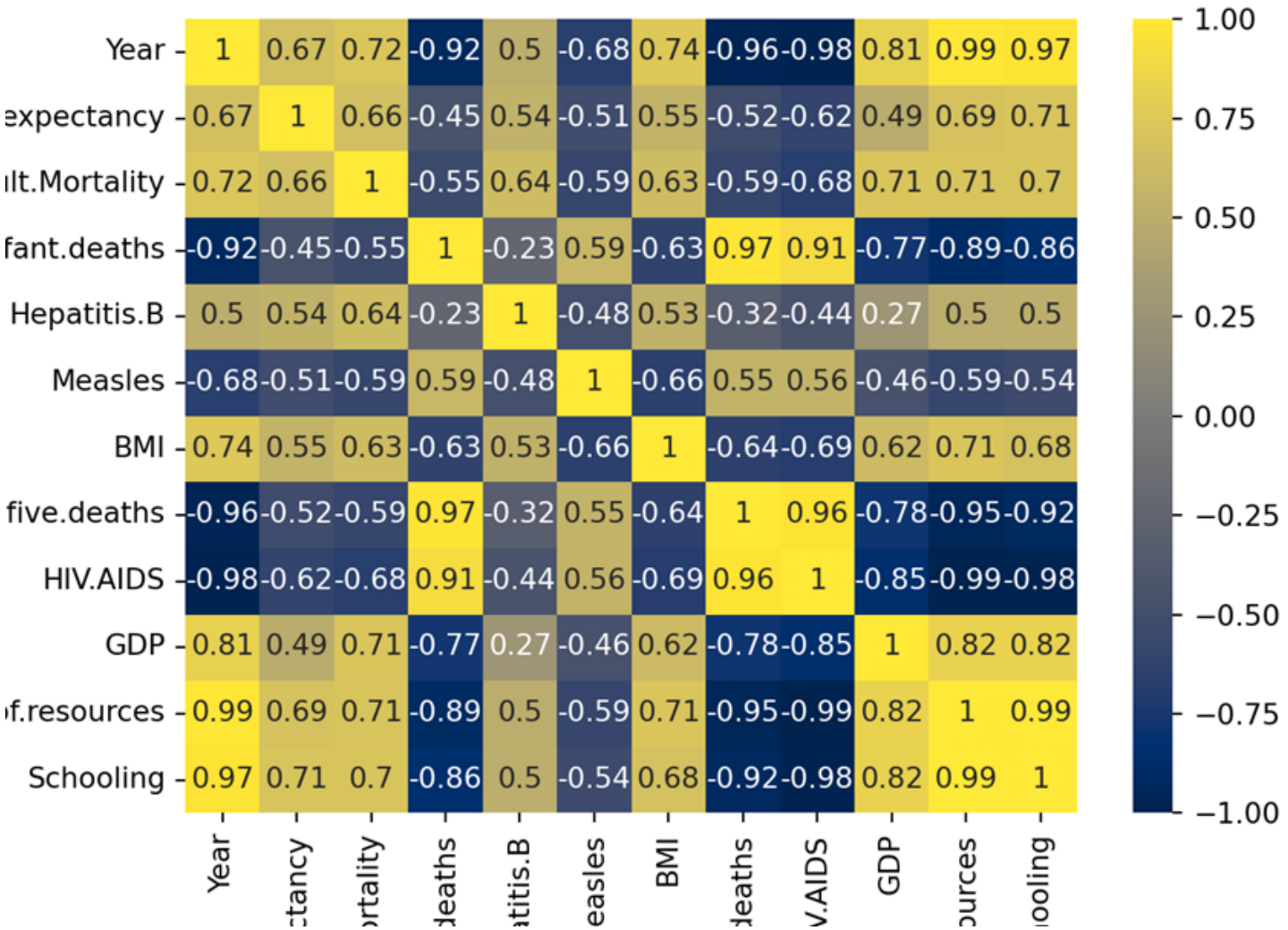
[5 rows x 14 columns]

	Country	Year	...	Income.composition.of.resources	Schooling
11	Ghana	2004	...	0.491	7.7
12	Ghana	2003	...	0.489	7.9
13	Ghana	2002	...	0.484	7.6
14	Ghana	2001	...	0.485	8.0
15	Ghana	2000	...	0.480	7.7

[5 rows x 14 columns]
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16 entries, 0 to 15

```
Data columns (total 14 columns):
 #   Column                          Non-Null Count  Dtype
---  ------                          --------------  -----
 0   Country                         16 non-null     object
 1   Year                            16 non-null     int64
 2   Status                          16 non-null     object
 3   Life.expectancy                 16 non-null     float64
 4   Adult.Mortality                 16 non-null     int64
 5   infant.deaths                   16 non-null     int64
 6   Hepatitis.B                     14 non-null     float64
 7   Measles                         16 non-null     int64
 8   BMI                             16 non-null     float64
 9   under.five.deaths               16 non-null     int64
 10  HIV.AIDS                        16 non-null     float64
 11  GDP                             16 non-null     float64
 12  Income.composition.of.resources  16 non-null     float64
 13  Schooling                       16 non-null     float64
dtypes: float64(7), int64(5), object(2)
memory usage: 1.9+ KB
```

The correlation heatmap will be employed for one final time:



Observing the correlation heatmap again:

- Schooling, income composition of resources and adult mortality (in descending order of magnitude) have a strong positive correlation with life expectancy in Ghana. The above mentioned variables increase as life expectancy also increases.
- In the descending order of magnitude, the variables HIV/AIDS, under-5 deaths, measles and infant deaths are negatively correlated with life expectancy. This means that when a variable increases, the other variable will decrease and vice-versa.
- The remaining variables have a relatively less strong positive correlation with life expectancy: BMI, Hepatitis B, and GDP.

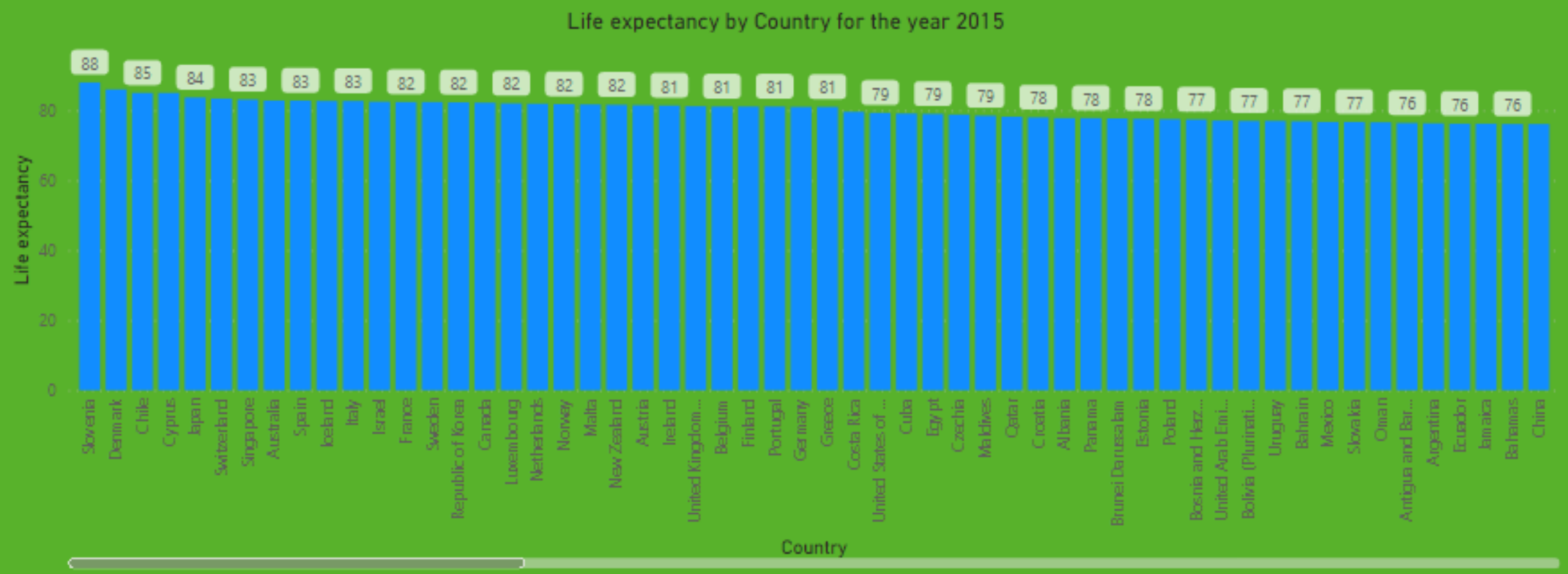
The next few pages show the visualizations of the data in Microsoft Power BI.
The last page after the visualizations will be a summary of the insights observed in the used data.

The following is a visualization of the life expectancy in 2015 :

Slovenia had the highest life expectancy in 2015 at 88 years, whereas Sierra Leone had the lowest life expectancy at 51 years.

Five of the least life expectancies were recorded in African countries, namely : Cote d'Ivoire, Chad, Central African Republic, Angola and Sierra Leone.

Egypt had the highest life expectancy amongst the African countries, at 79 years.

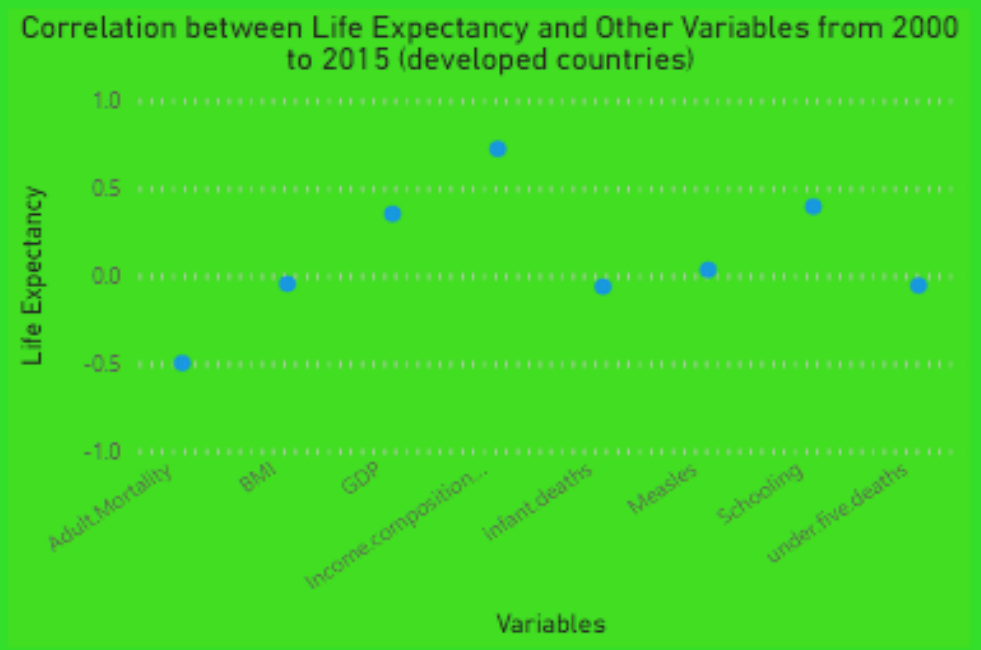


The next couple of pages displays the correlation between life expectancy and other variables.

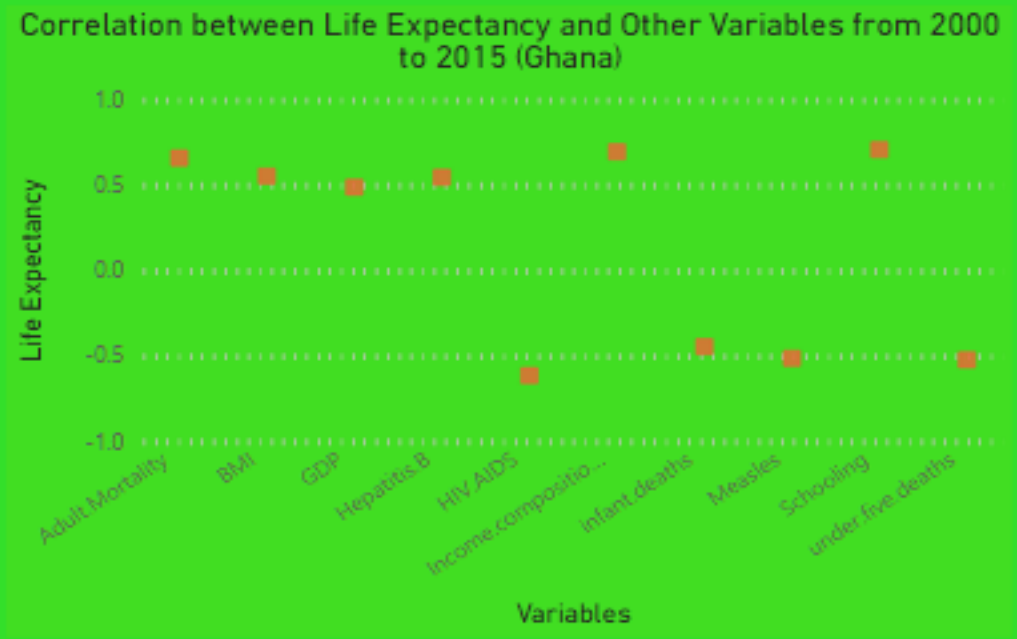
As adult mortality increases, life expectancy decreases and vice-versa, in a similar fashion.

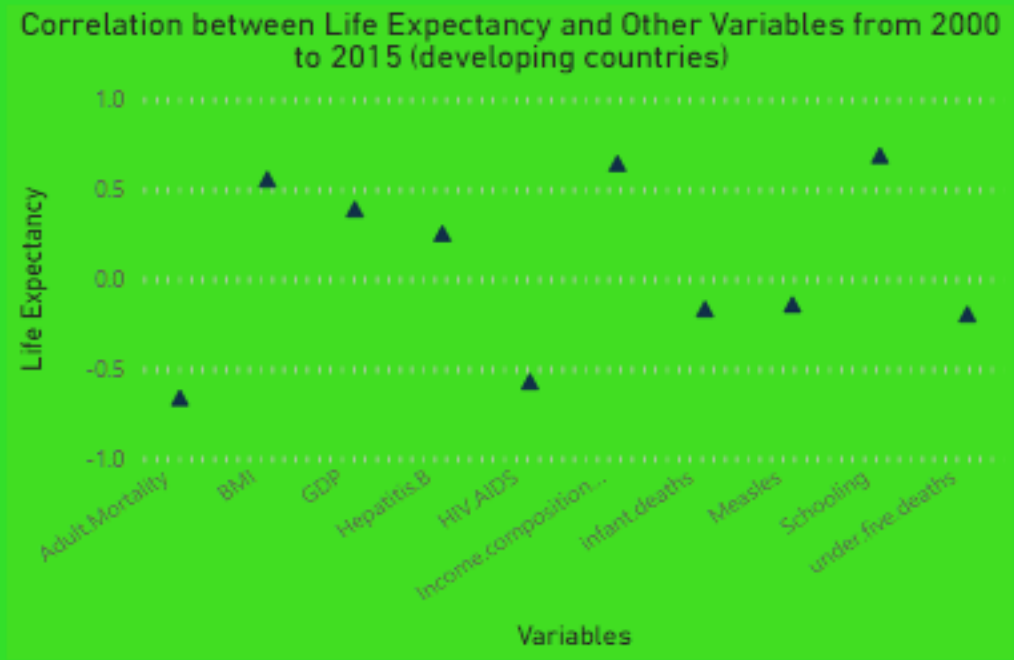
As adult mortality decreases, an increase in life expectancy is recorded and vice-versa, in the developed world.

As income composition of resources increase, life expectancy also increases and vice-versa.



Life expectancy increases as schooling increases and vice-versa, both in Ghana and developing countries.

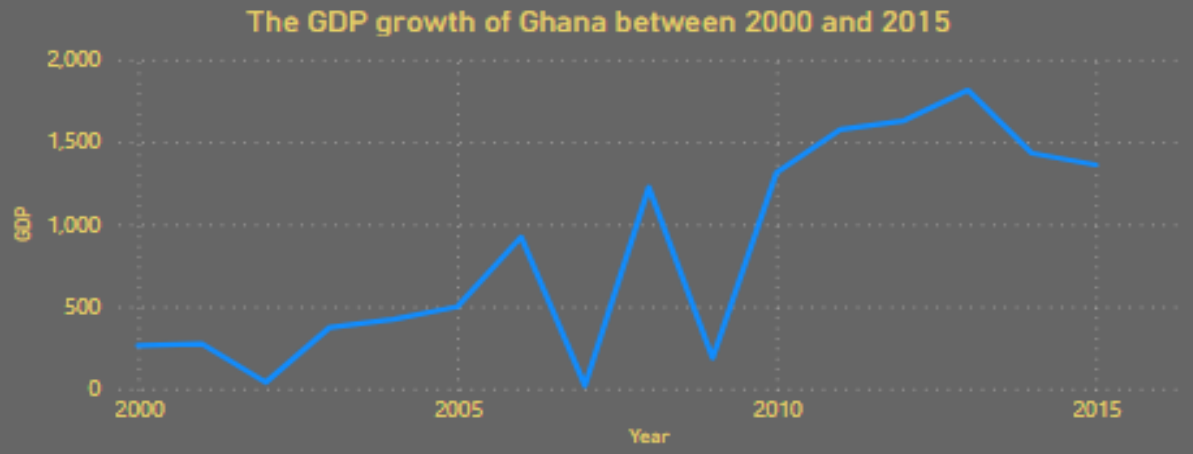




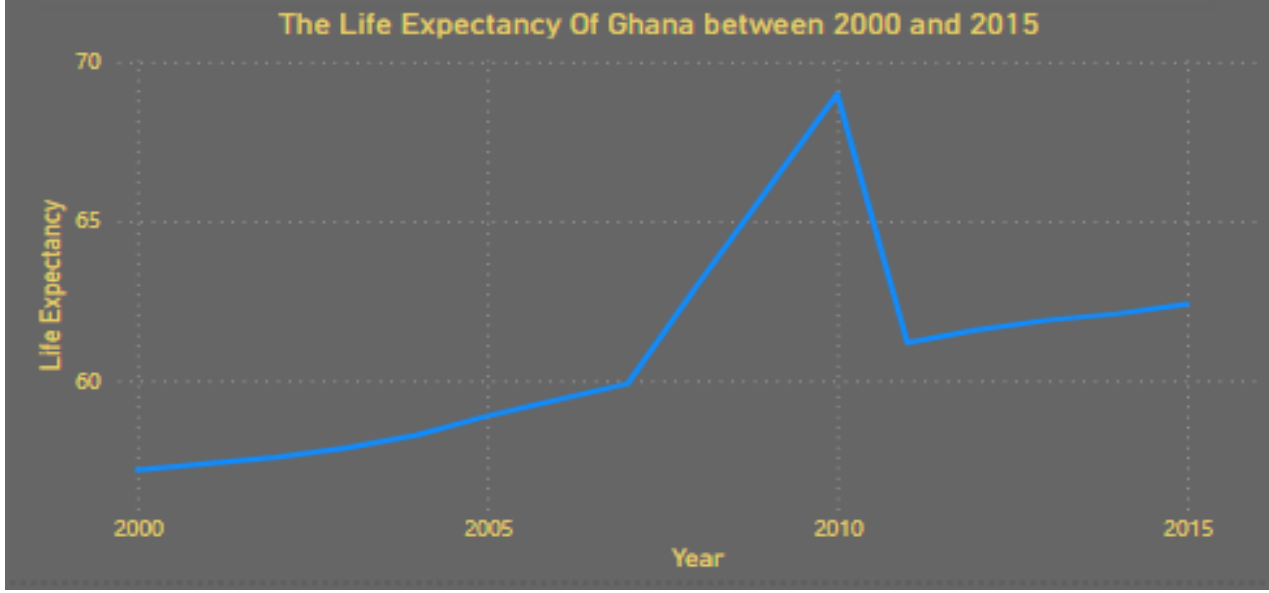
Life expectancy increases, as adult mortality decreases and vice-versa, in developing countries.

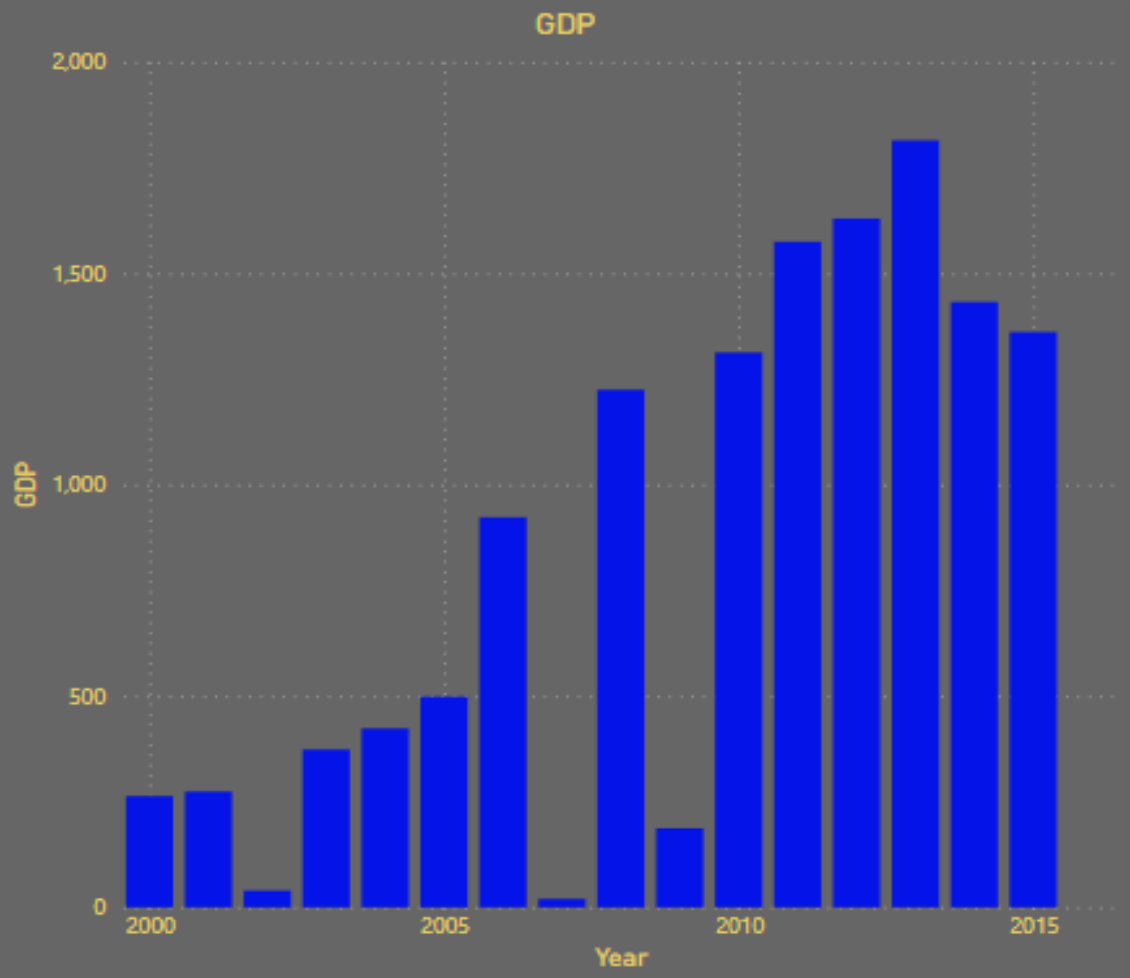
As HIV/AIDS decreases, life expectancy increases and vice-versa, in both Ghana and the developing countries.

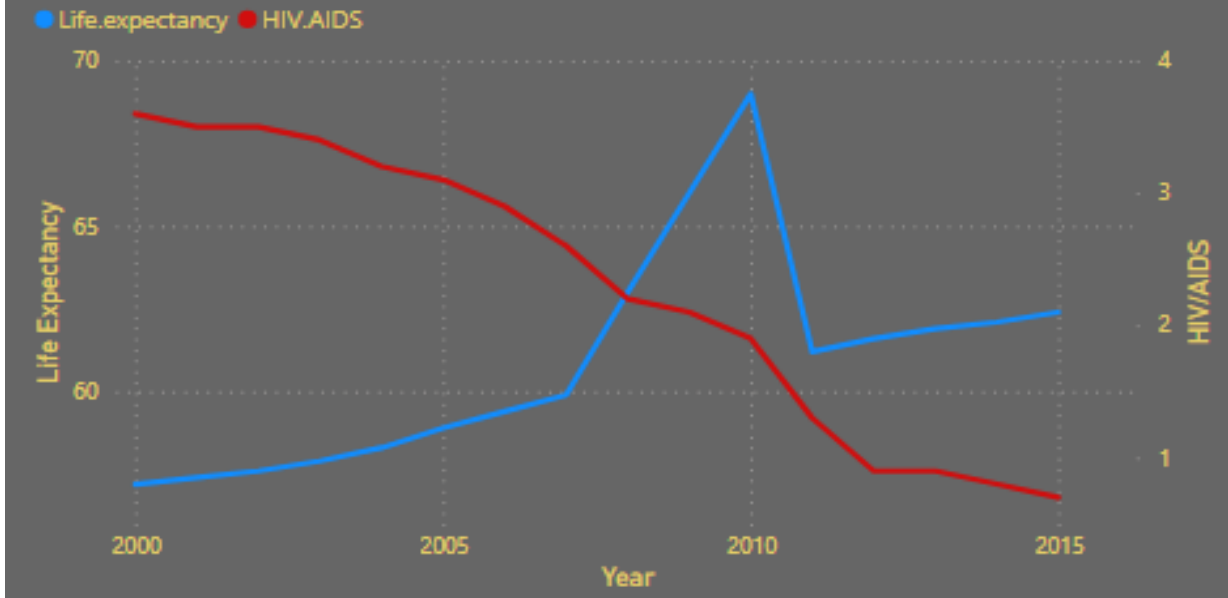
Let us now shift our focus to Ghana.

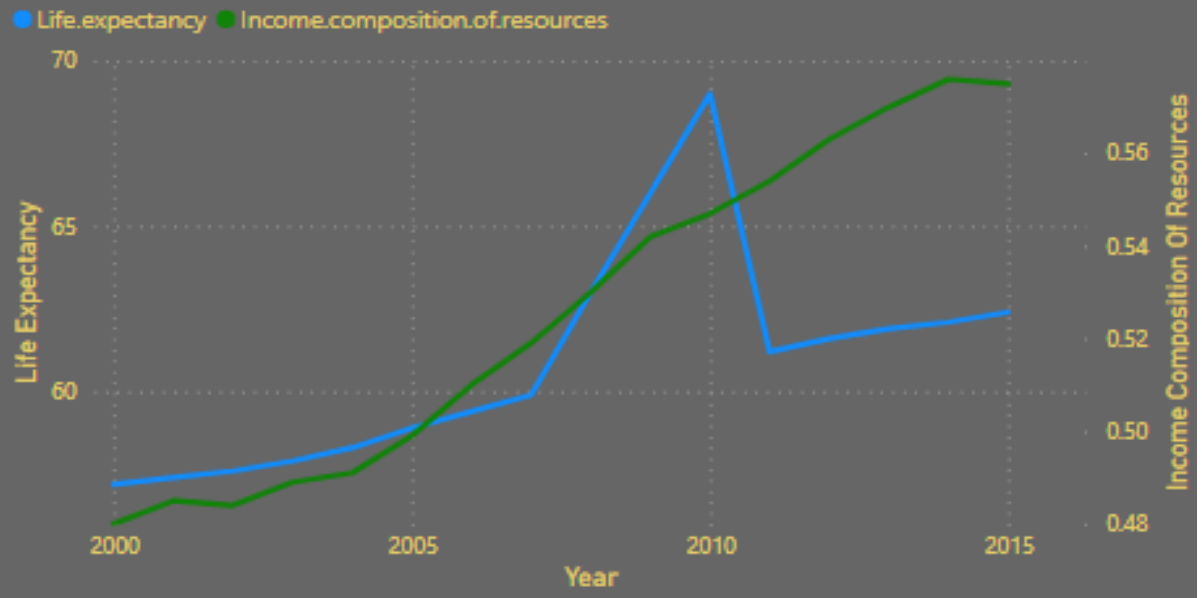


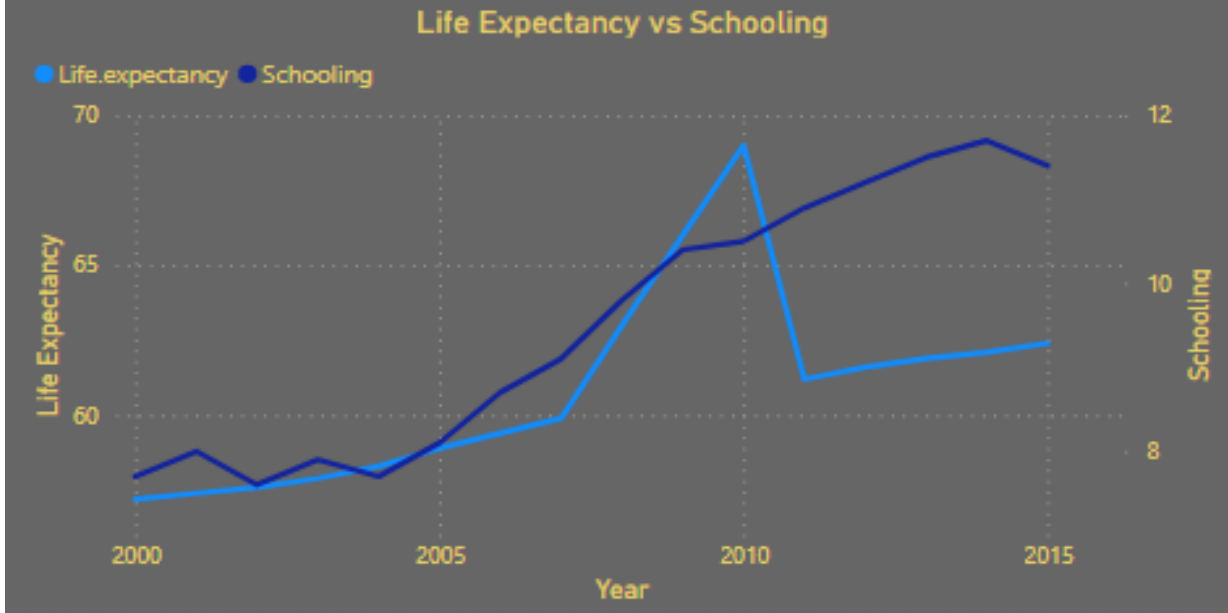
Ghana's life expectancy dropped sharply from 2010 to 2011 but was rising quite steadily from 2011 to 2015 however, the GDP also fell within the period 2013 to 2015. The highest GDP was 1814.49, in 2013.











As Schooling increases, life expectancy also increases and vice-versa.

As income composition of resources increases, life expectancy increases and vice-versa.

As HIV/AIDS increases, life expectancy decreases, and vice-versa.

Summary Of Key Insights

- Basically, efforts can be made to improve schooling in order to increase life expectancy. Similar values were observed in Ghana and developing countries as a whole.
- As adult mortality increases, life expectancy decreases and vice-versa, in a similar fashion.
- As HIV/AIDS increases, life expectancy diminishes, the reverse also holds. Similar values were observed in both Ghana and developing countries.
- As income composition of resources increase, life expectancy also increases and vice-versa.
- Slovenia had the highest life expectancy in 2015, with 88 years, whereas Sierra Leone had the least, 51 years.
- On the African continent, Egypt had the highest life expectancy, with 79 years.

Thanks for reading,

Visit kacelykodes.github.io to view similar projects.