|  |
| --- |
|  |
| **Project: Life Expectancy and Happiness**  Professor: Shakeel Khan  Students: Hien Do & Khanh Dang |
|  |

**Table of Contents**

[Life Expectancy 1](#_Toc65793806)

[Relation between Life Expectancy and factors 5](#_Toc65793807)

[Relation between Happiness and factors 8](#_Toc65793808)

[Conclusion 11](#_Toc65793809)

[Appendix 12](#_Toc65793810)

[Use of charts 12](#_Toc65793811)

[Data Preparation: 13](#_Toc65793812)

[Data Source: 13](#_Toc65793813)

[Data Cleaning: 13](#_Toc65793814)

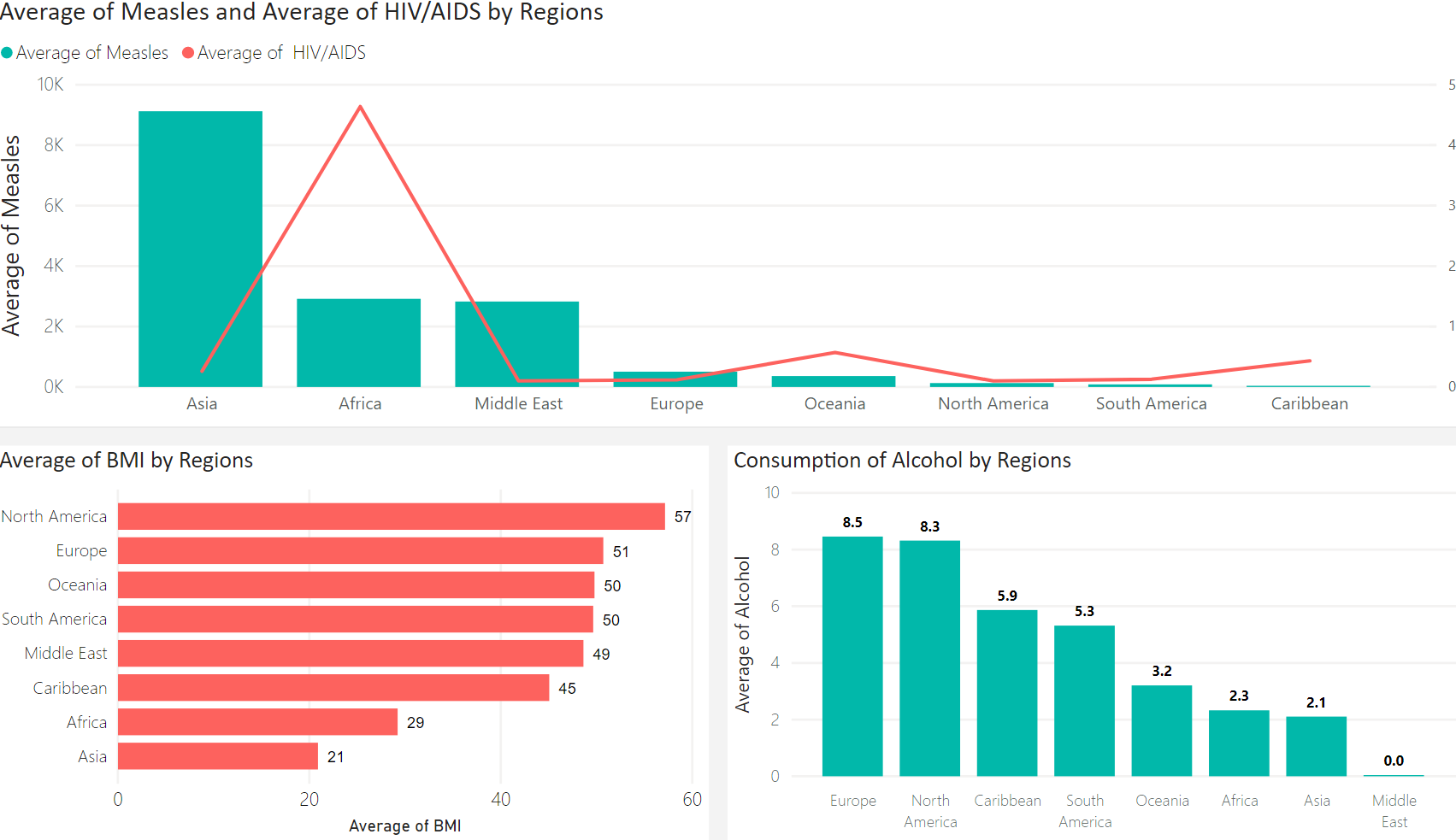
[Data Dictionary: 14](#_Toc65793815)

# Life Expectancy

There are several factors influence on life expectancy in different levels which including health, immunization and mortality, and economy dimensions

**Health Factor**

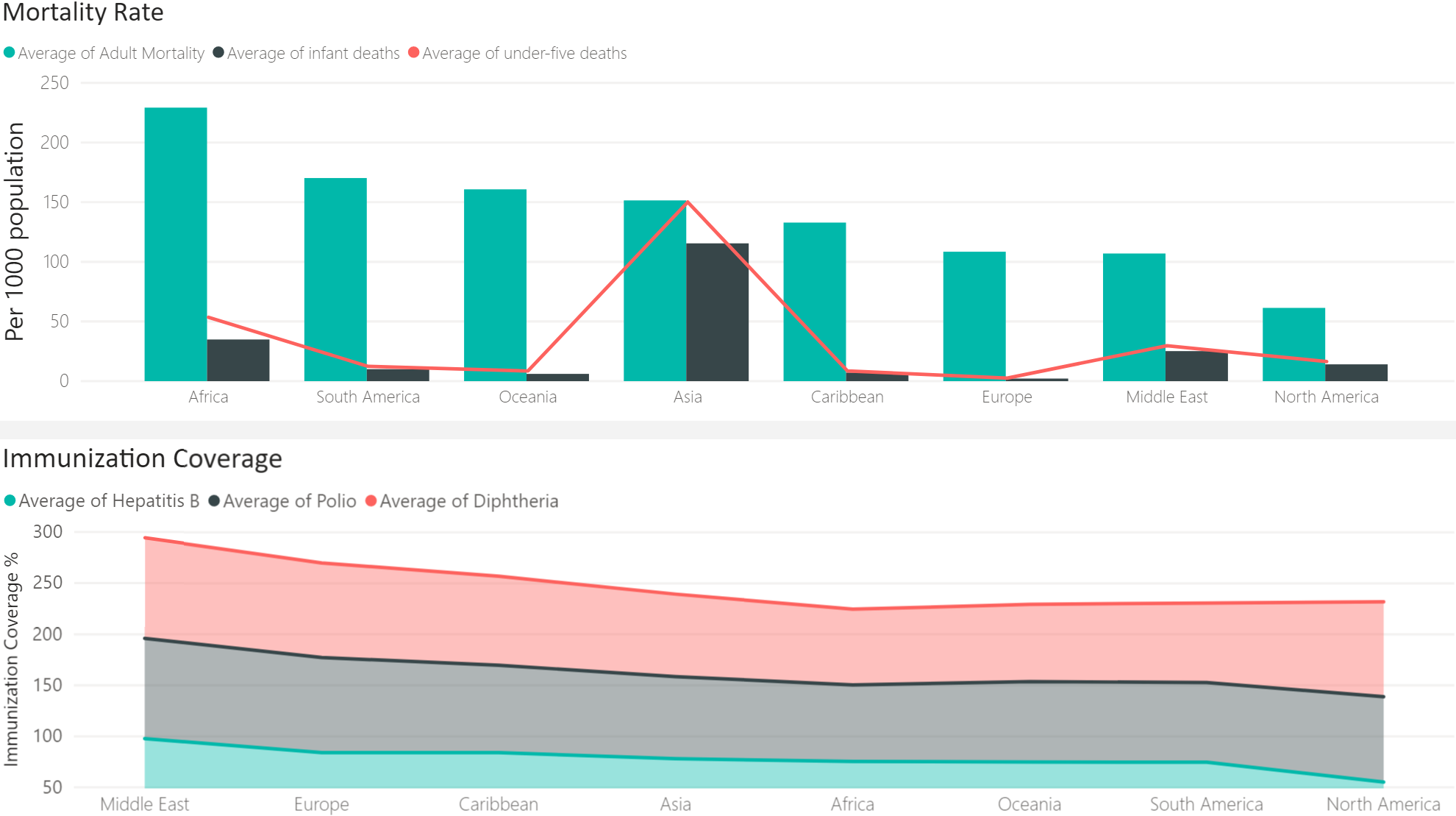
The line graph shows that Africa is the region with the highest rate of HIV/AIDS. The rate is significantly higher compare to other regions as North America, Europe, Mid East, and Asia. The bar chart indicates Asia has significant number of measles cases compare to other regions. The reason is that huge number of people, especially children in Asia who more likely to have Measles comparing to other regions and low number of vaccines in this area which cause the high rate of Measles in Asia. Average BMI of all countries is 38.42 and North America is the region with the highest BMI as 57 and follow up by Europe, Oceania, and South America, respectively. The reason is that those Western countries usually consume more junk food than Asia countries. The other reason is genetic differences between Western people and Asian. Additionally, as can be seen in the bar chart that North America and Europe also contain more alcohol than other regions which result in unhealthy lifestyle and higher BMI. The consumption of alcohol in Europe and America are much higher than other countries because they are home of producing wine and beer. Additionally, the difference of culture and regulations between Asian and Western also cause the lower rate of alcohol. One notice is that data from Asia countries is more challenging to collect compare to other countries which may cause the actual rate of alcohol of Asia region is higher than the below chart. However, the consumption of alcohol of Asia is noticeably depreciated comparing to other regions including Europe and North America.



**Immunization & Mortality**

Due to the uncommon of Hepatis B, Polio, and Diphtheria diseases in North America and South America compare to Middle East and Europe, the immunization coverage of those diseases among 1 year old is minimal in America.

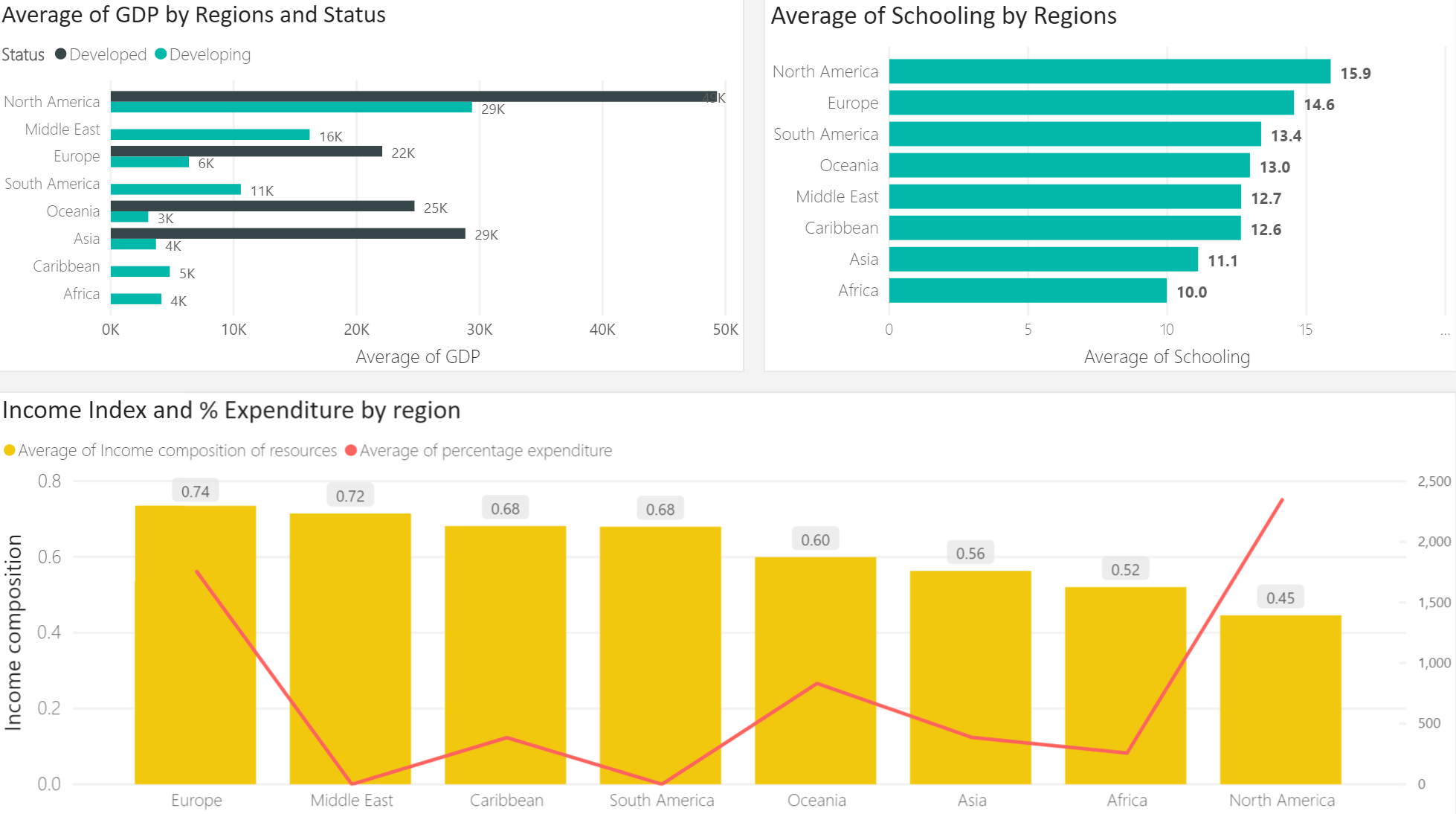
Africa has the highest rate of adult mortality which can be explained by poor health care system and environment compare to other regions. Asian countries have remarkable rate of infant and under-five death because some disease like Diphtheria, Measles, and Polio are easily spread out in tropical weather which affected children’s health in Asia region.



**Economic Factors**

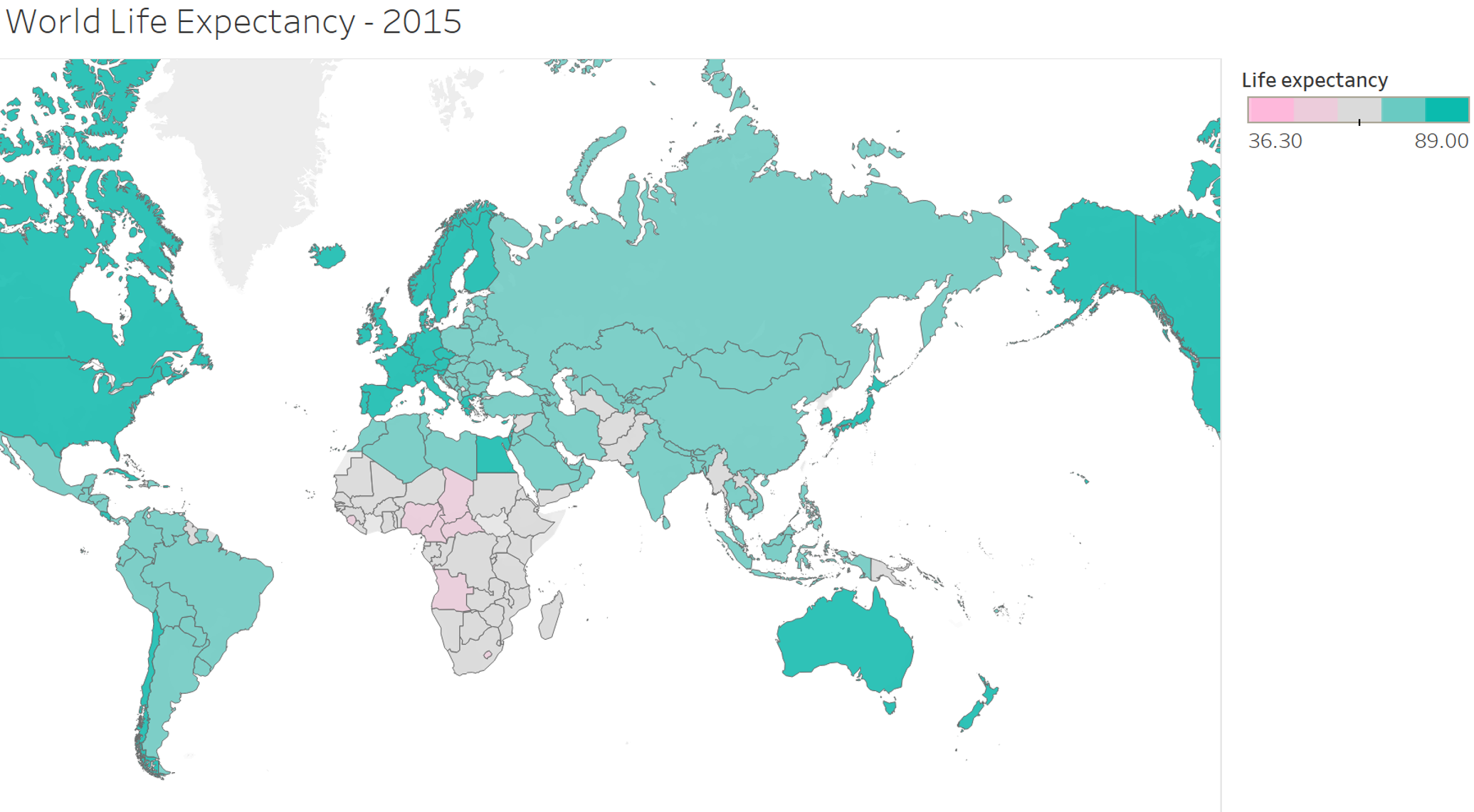
The bar chart of indicates that the average GDP per capita of developed countries are much higher than developing country in most of regions which including North America, Europe, Oceania, and Asia. North America region has the highest GDP per capita in both developing and developed countries. In 2015, Africa and South America have no developed countries. The gap between developing and developed countries in Asia and Europe is much higher than North America because huge difference of GDP per capital between developed countries as China (approximately $8,000) compare to the developing as Cambodia ($1000) which is 8 times higher.

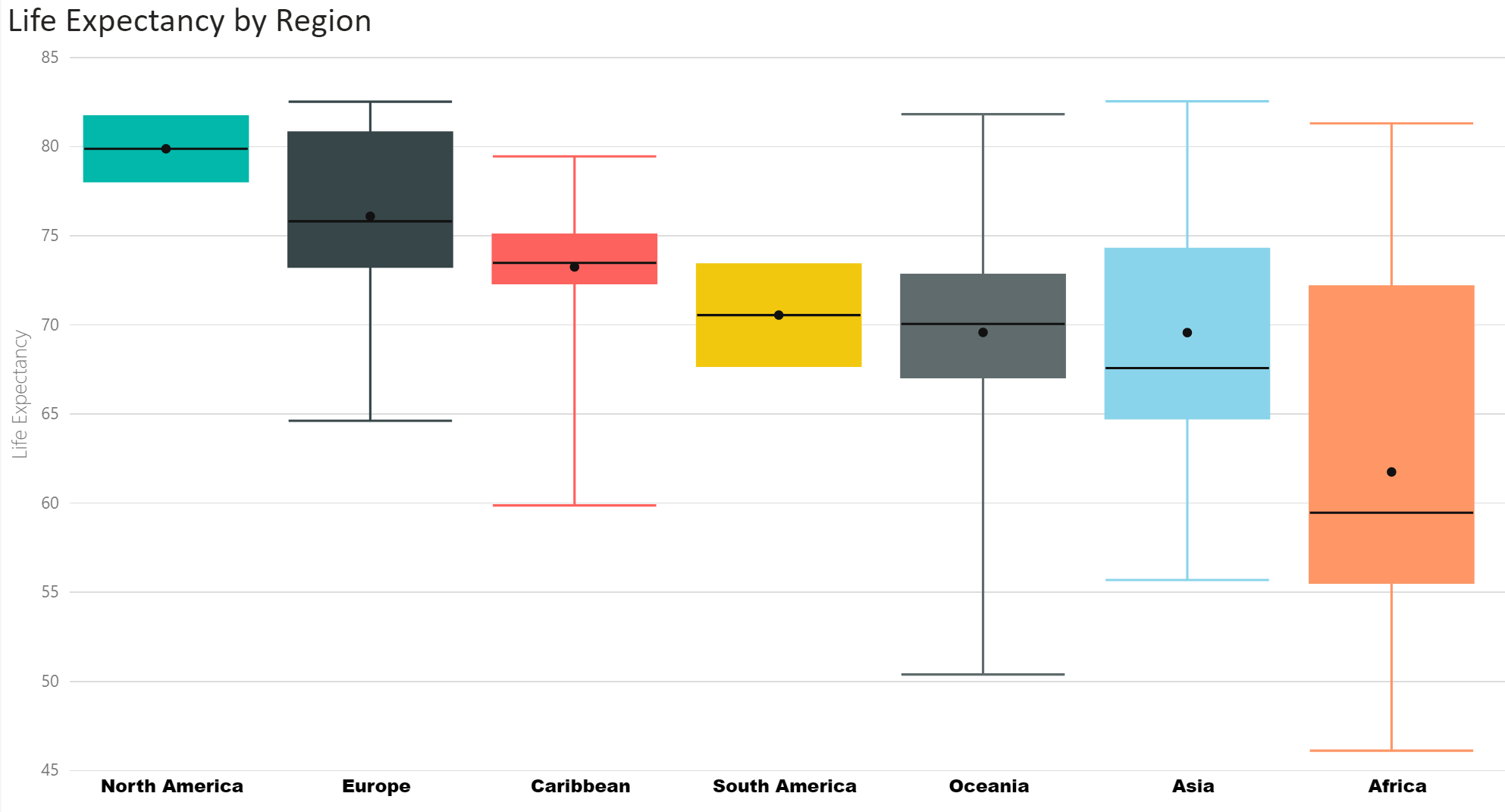
From the chart of schooling by regions, North America and Europe have the high number of schooling years than other regions, 15.9 and 14.6 years, respectively. The reason is that those regions are home of many leading institutions and government pay attention on developing education systems to encourage people go to schools. The bar and line graph states Europe and North America have the highest average income index and percentage expenditure by regions, respectively. Europe’s government utilizes its productively resources to the citizens. North America spend huge fund on health care system as a proportion of GDP per capital which means health care is crucial element in their countries.



**Life Expectancy**

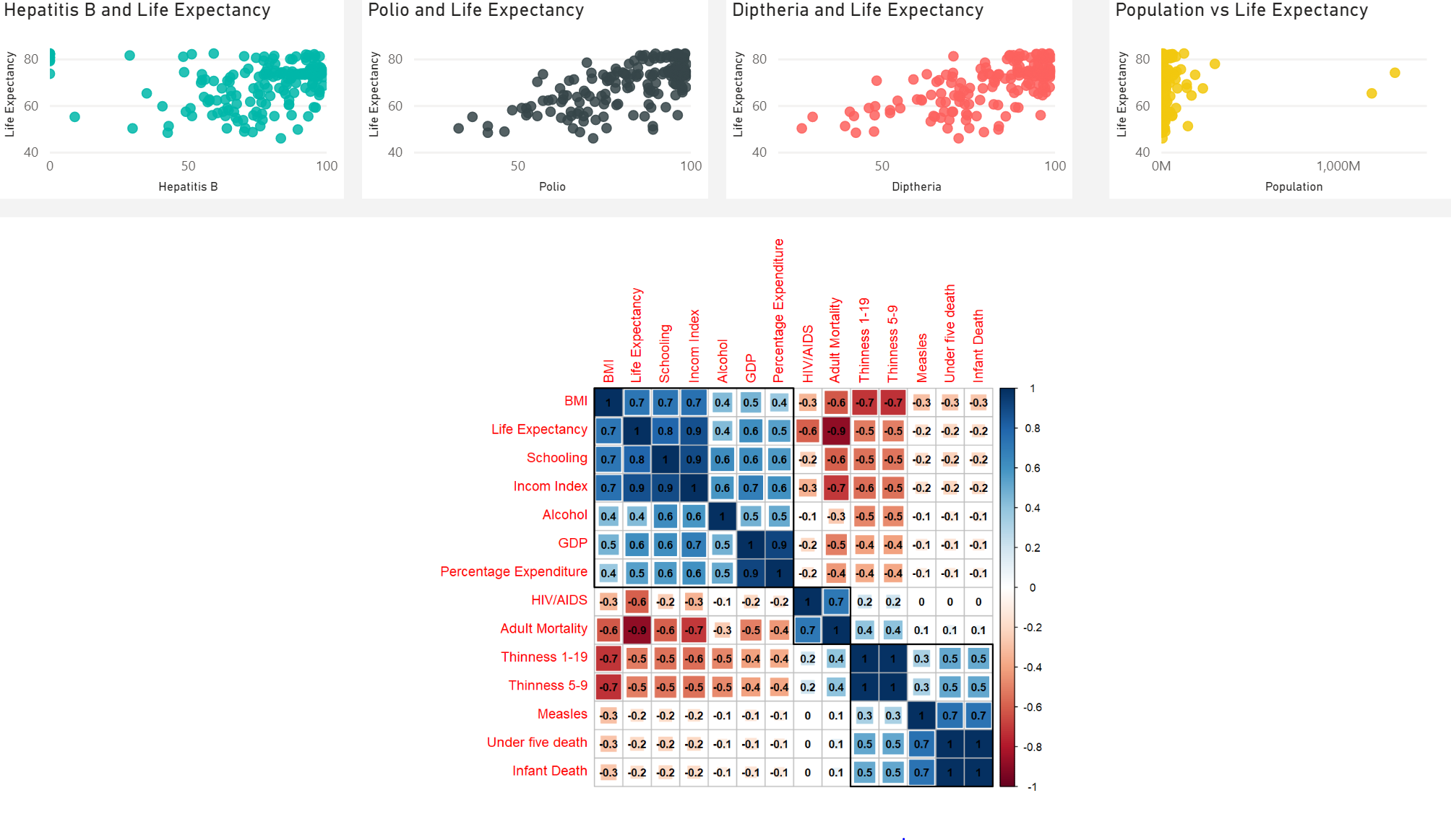
During the period from 2000 to 2015, the trend of average life expectancy was gradually increased from 66.9 to 71.3. Around the world, most people live longer 64.32 years and only 4.2% people live shorter than 50 years. From the box plot, North America and Europe are the region with high life expectancy, however the range between lowest and highest point of Europe is larger compare to America. It indicates that most of countries in North America have high life expectancy as above 77 years. On the other hand, some of countries in Europe with the life expectancy lower than 75 years. North America has the good rate of most of mentioned factors (health, GDP per capita, schooling, and income index) which can be explained why this is the region with highest life expectancy. In contrast, Africa has lowest life expectancy with some countries have life’s duration less than 60 years. The reason is that Africa has high number of diseases which cause high rate adult mortality, low GDP per capital, and less schooling years compare to other regions.

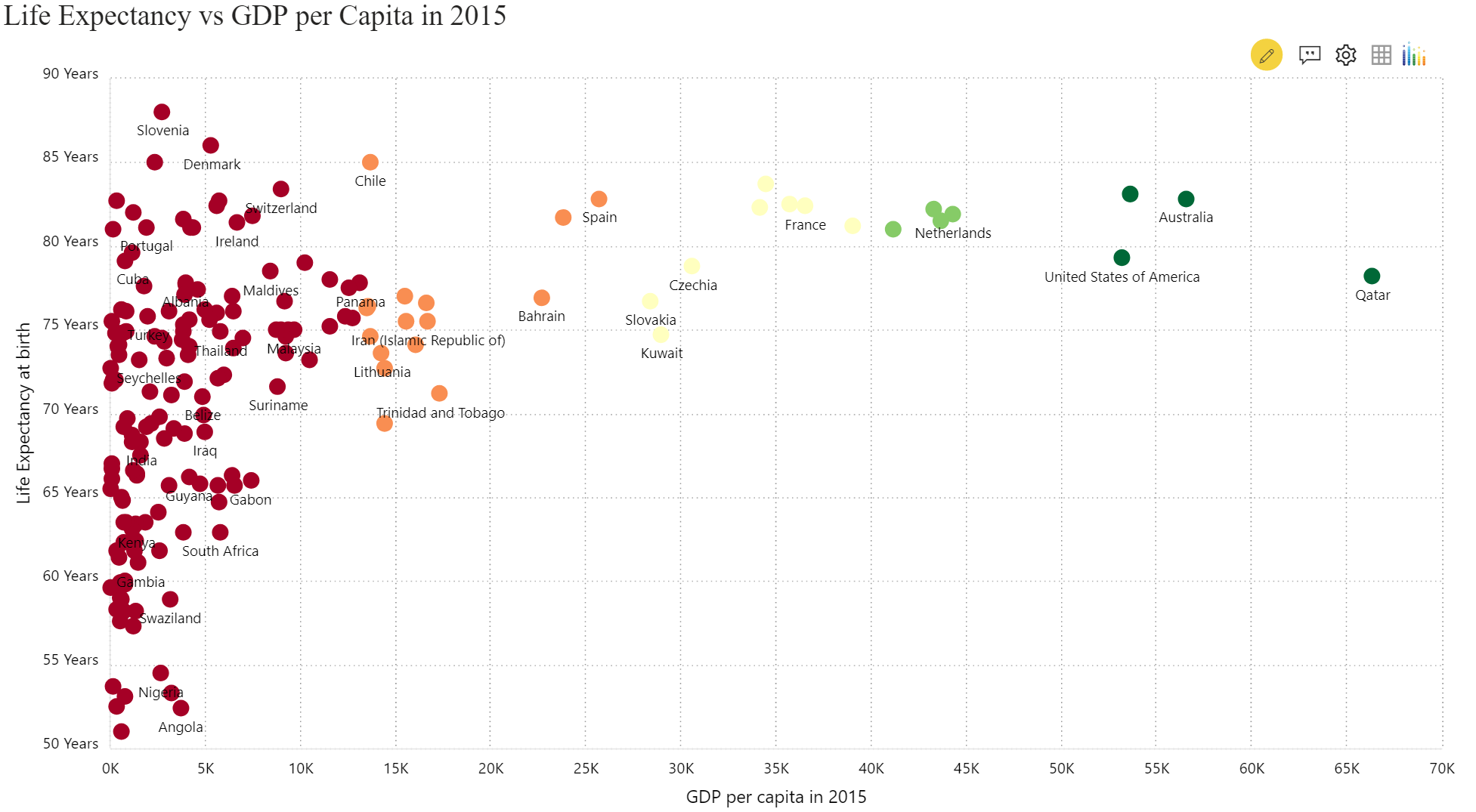




# Relation between Life Expectancy and factors

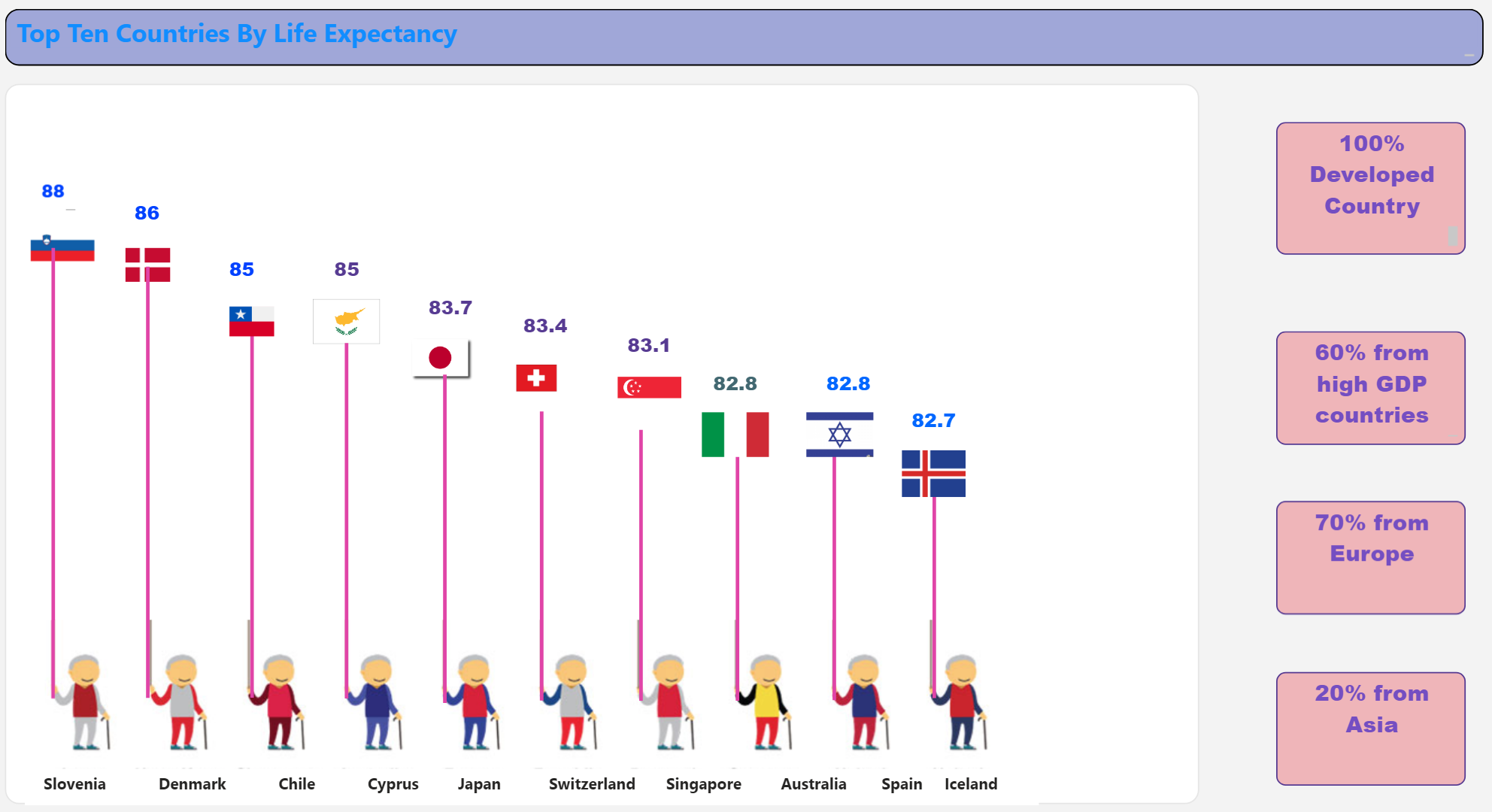
From the correlation matrix chart, income composition of resource and schooling are strong correlated with life expectancy. Income index has the highest correlation coefficient of 0.91 which means country utilizes its productively resources, it is more likely the citizen’s life expectancy is longer than expected. The correlation coefficient of schooling is 0.8 which indicates people tend to live longer than expected with higher education. The reason is that people who have higher education more likely to earn more income and better chance to support for their life and taking care of their health which help them to live longer. The scatter plot of GDP and life expectancy indicates that there is minimal relation between expectancy of life and GDP per capita. When country’s economy upswing, people can suffer more from medical problems and die faster regarding to the relation to industrial activity and air pollution. For instance, in 2015 Qatar and Czechia have higher GDP per capita than Demark but they are shorter in life’s duration.





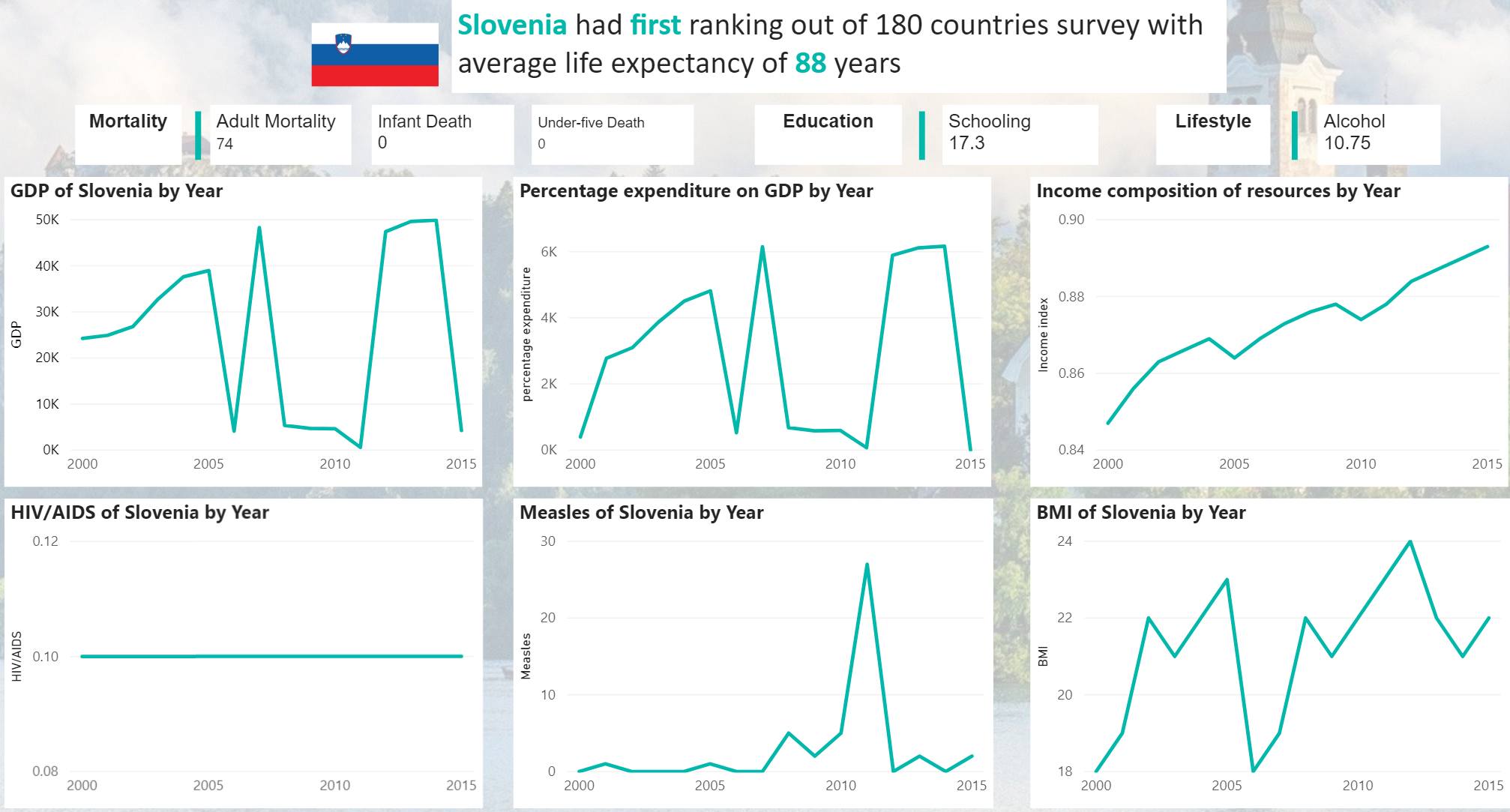
**Top Ten Countries by Life Expectancy**

Slovenia is the country with the highest life expectancy (88 years). Demark and Chile are placed at second and third place with 86 and 85 years, respectively. All of countries in the top ten are developed countries and six out of ten countries have high GDP per capita. 70% of them are from Europe region and only 20% of them are from Asia.



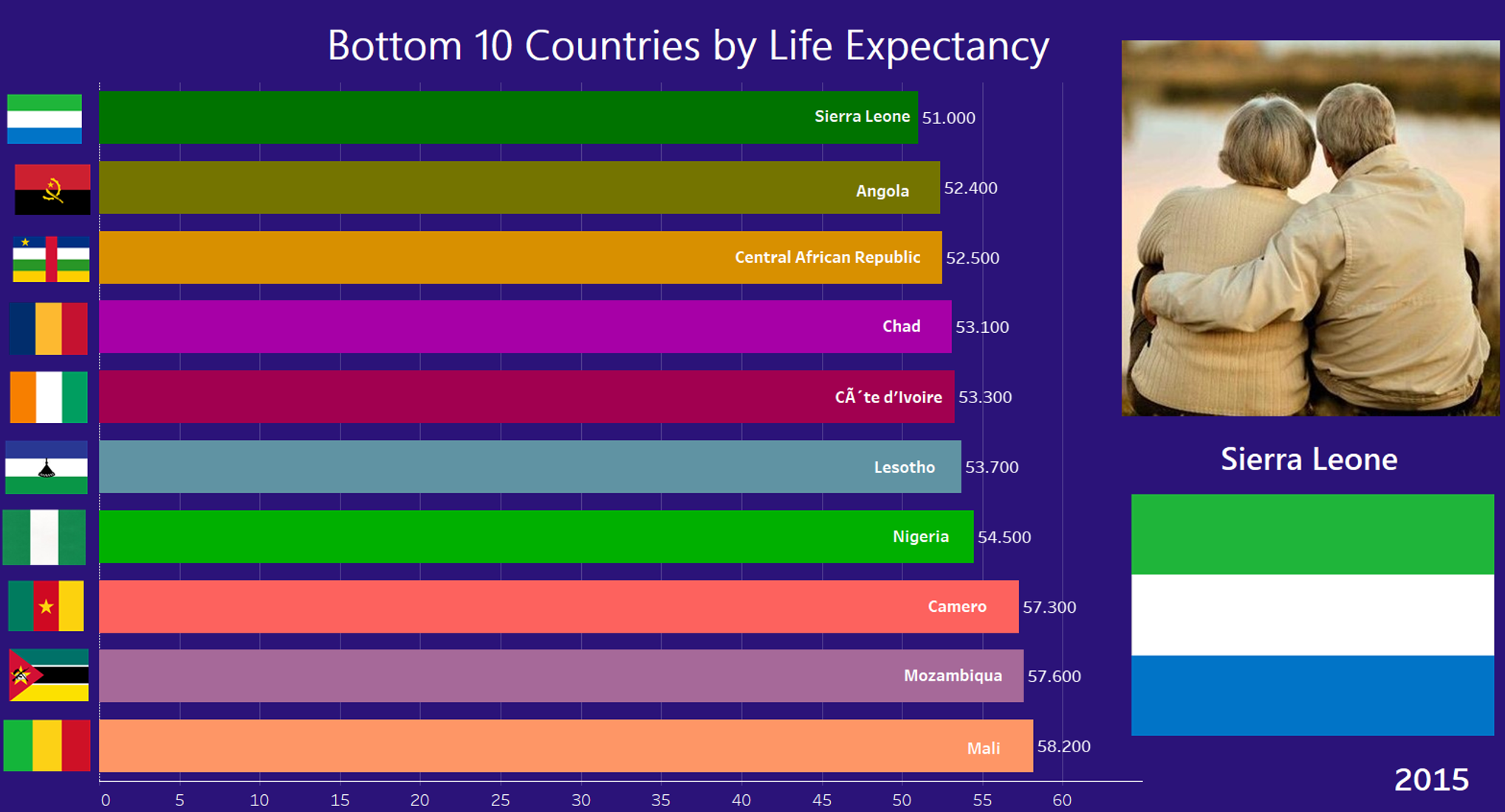
**Slovenia**

Out of 180 surveyed countries, Slovenia had first ranking of life expectancy with 88 years which is located in Europe. The noticeable points of Slovenia are schooling and income index are significantly high, and they are two factors that influence life’s duration the most. The schooling is 17.3 years which is even higher than the leading North America region of education years. The trend of income index was steadily increased which means country more and more pay attention on productively aggregate income to the area. Those factors play essential role to make citizen in Sloven live longer.



**Bottom ten countries by Life Expectancy**

Sierra Leone has the shortest life expectancy (51 years). Angola and Central African Republic are placed as second and third place with 52.4 and 52.5 years, respectively which are lower than by approximately 26% compare to the average life expectancy of the world. Most of countries in bottom ten list are from Africa region.



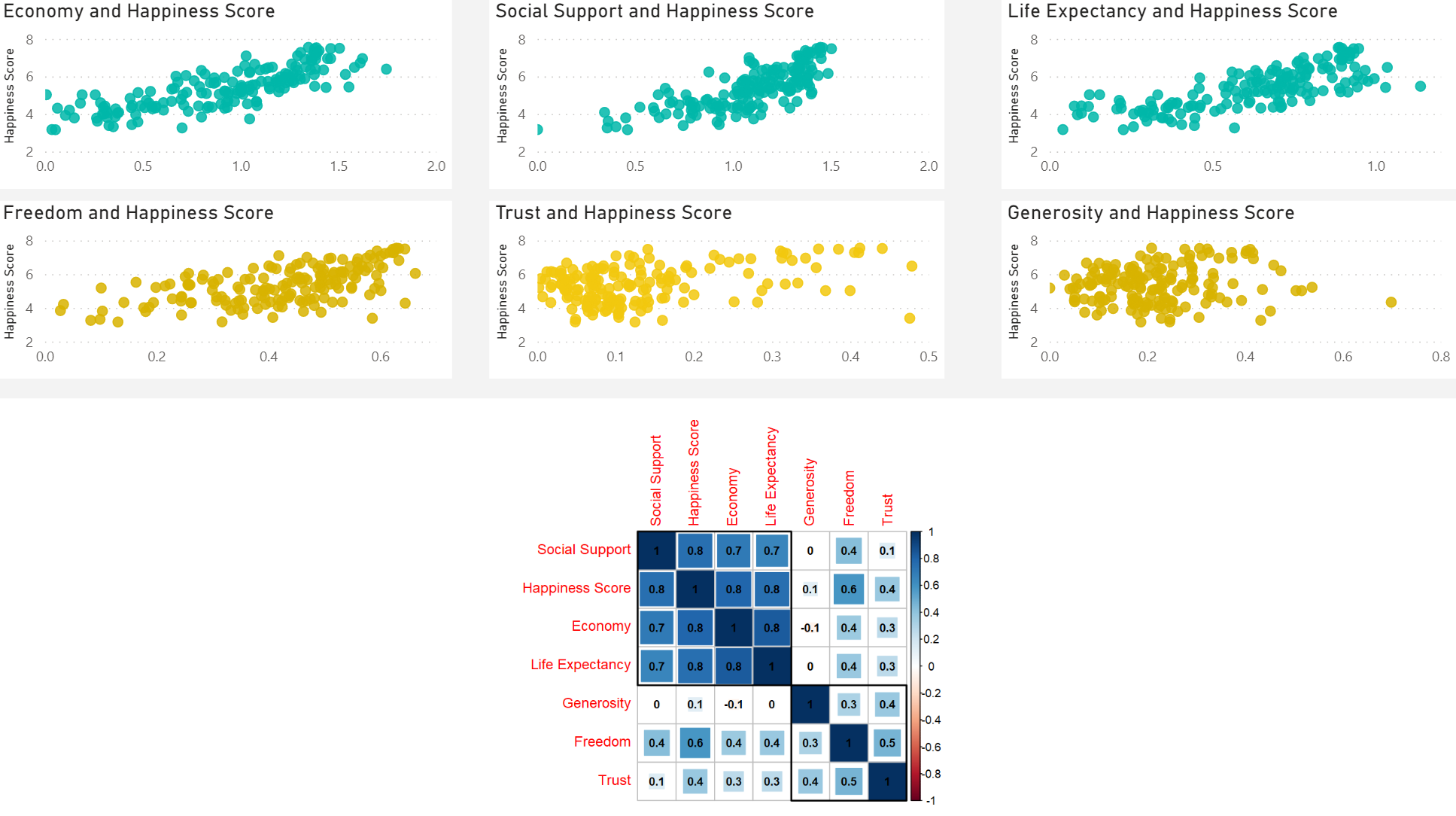
**Sierra Leone**

Sierra Leone has the lowest expectancy of life out of surveyed 180 countries which is located in West Africa with the population is nearly 7 million. From the line chart, it can be seen that income index which is the factor affected life’s duration the most are remarkable low in this country. Particularly, income composition resources is 0.4 which is lower by 50% compare to Slovenia and 35% compare to average of the world. Average schooling years is 9.5 years in this country which indicates that most people do not have chance to go to high school. This leads to minimal income to support from their life and health. Due to the low rate of education, labor force is mainly blue-collar which indicates higher chance of involving in industrial accidents. Those mentioned along with disadvantageous weather, and disease which are explained why Sierra Leona has shorter life expectancy.



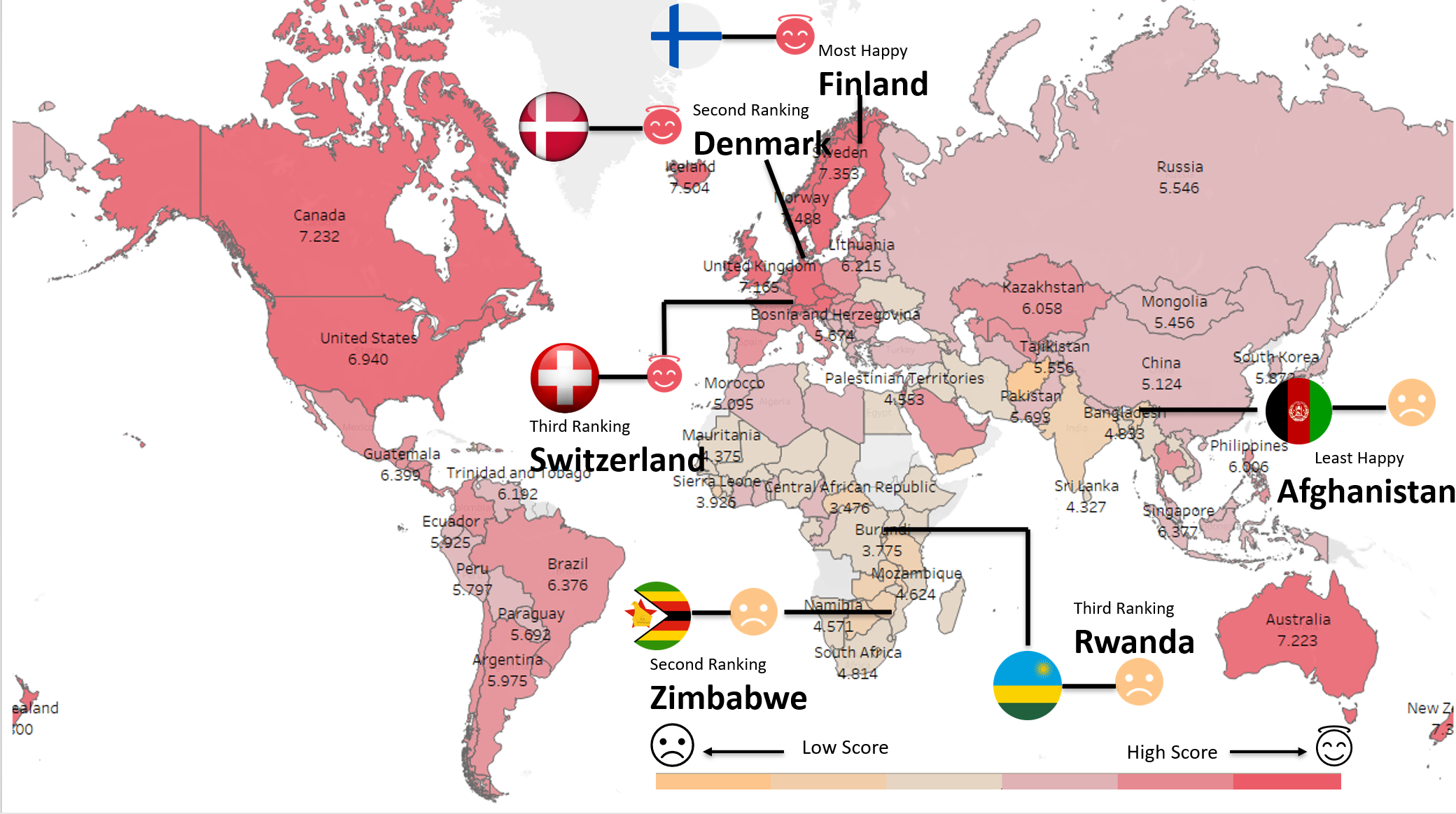
# Relation between Happiness and factors

From the correlation matrix between happiness and factors, the social support has the strongest the relation to happiness. Economy and life expectancy are the second and third important dimensions to happiness, respectively. In the scatter plot, there is positive relation between social support and expectancy of life. Social support is emotional and physical comfort that received from friends, family, organizations, and other people. In recent time, people more and more suffer from stress thus the support which is not limited emotional, tangible, and informational support which can help people reduce and balance unfavorable impacts of mental pressure. Especially, with elder people the role of social support even more important to increase their life’s satisfactions. Those reasons will result in people live happier when they received social support. The coefficient of economy is 0.8 indicates if economy is doing well, it is more likely that citizen live happier than expected. When economy grow, people have more income to support their life and less stressful on finance which is one reason to feel happy. The scatter plot of life expectancy of happiness states that there is positive relation between longevity and happiness. Happiness should be take into consideration to support long-living life of citizens.



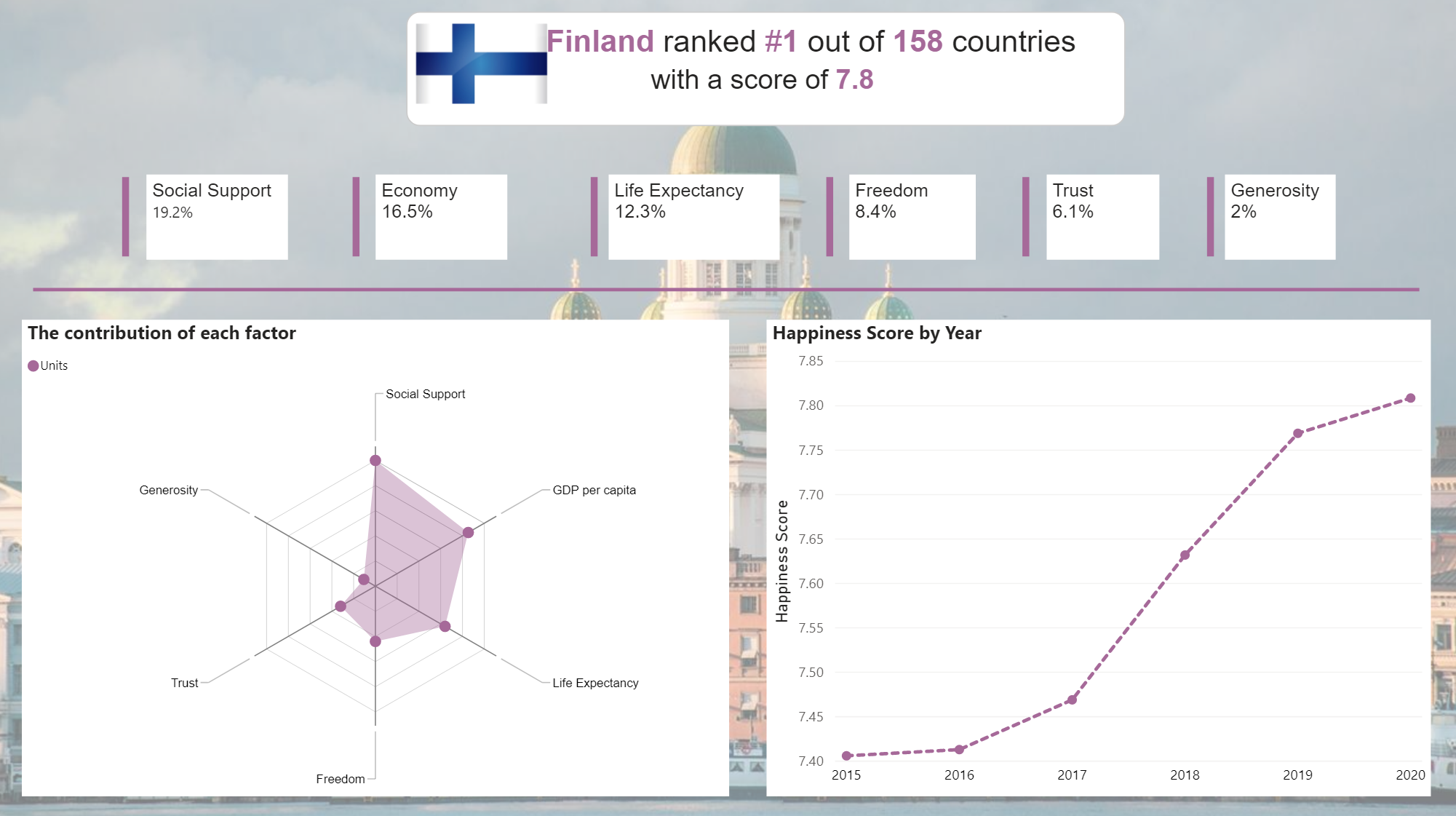
**Top Three and Bottom Three Countries by Life Expectancy**

In 2020, Finland, Denmark, and Switzerland are placed at the first, second, and third places of happiness out of 158 countries around the world, respectively. All of three countries are developed countries from Europe. Denmark and Switzerland are in the top ten countries by life expectancy. In contrast, Afghanistan has the lowest score of happiness with 2.56 compare to other country. Zimbabwe and Rwanda are ranked as second and third least happy country in the world which are located in Africa where have most countries in the bottom list of life expectancy.



**Finland**

In 2020, Finland is ranked number one out of 158 countries with a score of 7.8. The line chart states there is an increasing trend of happiness score in Finland from 2015 to 2020. The radar chart indicates social support, economy as GDP per capita and life expectancy are highly contributed to the happiness of Finland which align with the result of correlation matrix between happiness and dimensions.



**Afghanistan**

Afghanistan placed at the least happy country around the world in 2020 which is located Central Asia. The happiness ranking of Afghanistan sharply dropped by 14 degrees from 2017 to 2020. In this country, the contribution of social support, economy, and life expectancy dimensions to happiness are lower than other countries. However, the main reasons which make this country become least happy are related to freedom and trust to government. There are many conflicts between different political parties, fights, and terrorism issues in Afghanistan which result in unhappy country.



# Conclusion

An analysis of life expectancy and happiness not only help with people identify which place is best place to live but also proving insight about life expectancy and happiness. There are numerous advantages of happiness which are related to individual and community. Being happy is long-term solution for health’s problems, especially mental issues. Additionally, majority of happy countries have low rate of crime rate which will be positively affect on tourism. In workplace, people are happy which lead to lower rate of turn over and higher productivity. It is undeniable that country with high longevity will bring a lot of advantages, it also contains some obstacles. First, an aging population which means there are fewer working-age labor force in the country. This result in shortage of skilled and qualified employees, especially it is more challenging for business to recruit demanding roles. The shortage of supply will cause higher employment cost and decreasing productivity. Second, when facing with this problem, company will try to hire foreign employees which will increase cost to manage immigrant issues. In addition, rapidly aging population which cause high spending of social support and health care.

# Appendix

## **Use of charts**

* Bar chart: using height or length of bar to make categorical comparison of proportion and concepts among dimensions of data. In this case, the bar chart is used to compare average schooling by regions.
* Column chart: is similar to bar chart. In this case bar chart is use to compare average income index by regions.
* Line chart: encodes data using position and often shows trend over time. In this case line chart is used to show the trend of some factors such as GDP or happiness score.
* Area chart: is extension of line chart which shows the trend over time. In this case, it is used to make comparison of immunization coverage of three different diseases.
* Histogram: encodes data using height to show a distribution of data. It is used to show distribution of life expectancy by years.
* World map: encodes data using color and position to show data geographically. In this case, using color and position to show life expectancy around the world.
* Box plot: encodes data using position and height/length to shows distribution and variability of data. It is used to show the range of life expectancy in different regions.
* Scatter plot: encodes data using position to show relationship between two variables. The size can also be used to show a secondary comparison. The scatter plots are used to show the relationship of life expectancy to other factors.
* Correlation Matrix: is a table showing correlation coefficients between variables. It shows the correlation coefficients of happiness to other factors (economy, freedom, social support,...)
* Radar chart: is used to compare two or more items or groups on various features or characteristics. In this case, it is used to compare the contribution of each factors as social support, economy, freedom, trust, and generosity to happiness.

## **Data Preparation:**

### **Data Source:**

There are two main data files from Kaggle and WHO including life expectancy and happiness which contains 3000 and 1000 observations, respectively of more than 150 countries.

### **Data Cleaning:**

We checked the missing, inconsistent, invalid values of both files: world life expectancy and world happiness. The file world happiness is cleaned without missing, inconsistent and invalid values.

However, there are some missing values in file world life expectancy. For some countries which lack information of 1-2 years, we calculate the average of remaining values to fill in the missing values. However, there is one values which have no records from 2000 – 2015, so we decided to remove this record out of data set. We created one more variable called “region” which including North America, Europe, Caribbean, Oceania, Africa, South America and Asia for further analyze later.

### **Data Dictionary:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Column Name** | **Position** | **Data Type** | **Description** |
| Country | 1 | Character | Name of Country |
| Region | 2 | Character |  |
| Happiness Rank | 3 | Numeric | Score of Happiness |
| Happiness Score | 4 | Numeric | Ranking of Happiess |
| Economy | 5 | Numeric | The contribution of GDP per capita on Happiness Score |
| Social Support | 6 | Numeric | The contribution of Social Support on Happiness Score |
| Health | 7 | Numeric | The contribution of Life Expectancy on Happiness Score |
| Freedom | 8 | Numeric | The contribution of Freedom on Happiness Score |
| Trust | 9 | Numeric | The contribution of Government Corruption on Happiness Score |
| Generosity | 10 | Numeric | The contribution of Generosity on Happiness Score |
| Year | 11 | Date | Year: From 2015 -2020 |

