FINAL REPORT

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHTN
ĐHQG-HCM

SEMESTER II
2022 - 2023

Data Visualization

Life Expectancy

PRESENTED TO

Dr. Bui Tien Len

PRESENTED BY

GROUP 09

Table of Contents

About Us

Group 09

Vo Van Hoang

20127028

Nguyen Duc Minh

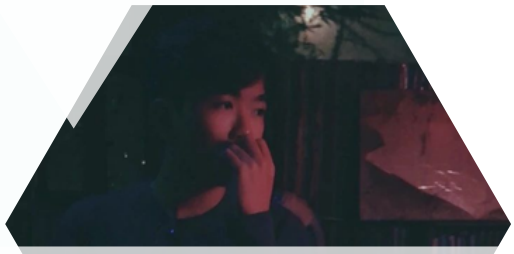
20127049

Ngo Van Trung Nguyen

20127054

Nguyen Minh Tuan

20127092



Nguyen Truong Minh Khoi

20127214

We are all K20 students of Faculty of Information Technology with Data Science major.

Data visualizaion is a subject which all of us are interested in.

Task Allocation

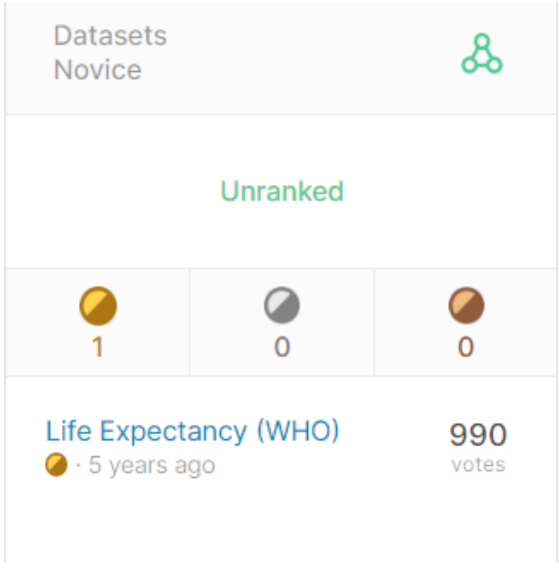
Group 09

Task Description	Person in charge
Writing Report Data Visualization (Continent part)	Vo Van Hoang 100% 20127028
Writing Report Web Visualization Machine Learning Model	Nguyen Duc Minh 100% 20127049
Writing Report Data Visualization (World part)	Ngo Van Trung Nguyen 100% 20127054
Statistical Description Machine Learning Model	Nguyen Minh Tuan 100% 20127092
Choosing Dataset Data Visualization (Country part)	Nguyen Truong Minh Khoi 100% 20127214

Information of Dataset

The happier you are,
the longer your life expectancy is.

Dataset Source



Reasons choosing this dataset

01 Life expectancy is a critical measure of population health and well-being.

02 By visualizing life expectancy data, we can understand many factors, and challenges that influence health outcomes.

03 Comparative analysis helps understand variations in life expectancy across different populations and informs global health strategies.

04 Interdisciplinary connections can be made by exploring the relationships between life expectancy and various variables.

05 Visualizations of life expectancy data serve as educational tools to promote health literacy and raise awareness.

This dataset is scraped from WHO (World Health Organization). Therefore, it is a beliable dataset.

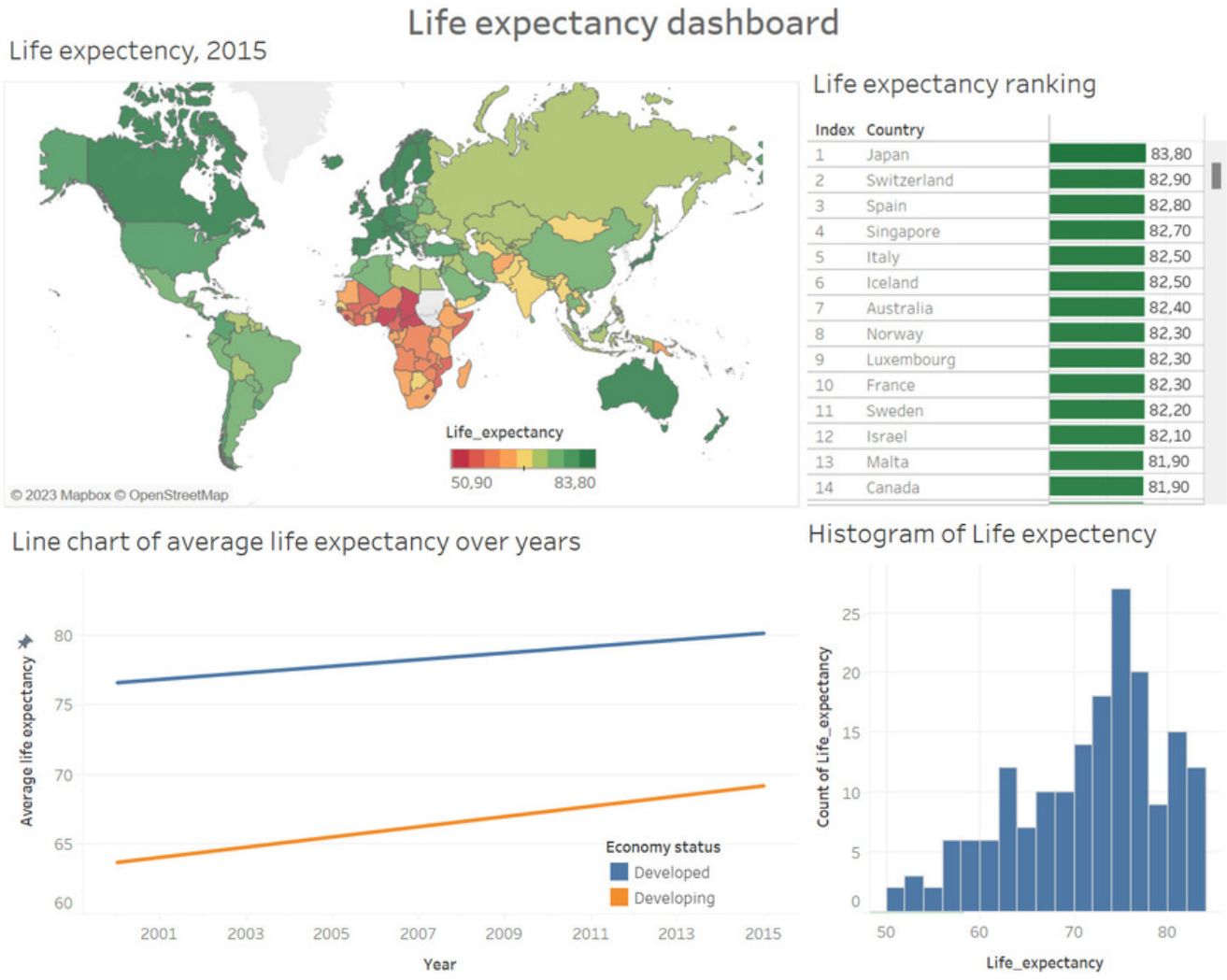
Attributes definition

Attribute	Definition
Country	The name of the country being measured for life expectancy.
Region	The region where the country is located.
Year	The year in which the life expectancy data was collected.
Infant_deaths	The number of deaths of infants under 1 year old per 1000 live births.
Under_five_deaths	The number of deaths of children under 5 years old per 1000 live births.
Adult_mortality	The probability of dying between the ages of 15 and 60 years per 1000 population.
Alcohol_consumption	The average annual alcohol consumption per capita in liters of pure alcohol.
Hepatitis_B	The percentage of 1-year-olds who have been vaccinated against Hepatitis B.
Measles	The number of reported measles cases per 1000 population.
BMI	The average Body Mass Index (BMI) of the population.
Polio	The percentage of 1-year-olds who have been vaccinated against polio.

Attribute	Definition
Diphtheria	The percentage of 1-year-olds who have been vaccinated against diphtheria.
Incidents_HIV	The number of new HIV infections per 1000 population.
GDP_per_capita	The Gross Domestic Product (GDP) per capita in US dollars.
Population_mln	The total population of the country in millions.
Thinness_ten_nineteen_years	The percentage of the population aged 10-19 years who are thin.
Thinness_five_nine_years	The percentage of the population aged 5-9 years who are thin.
Schooling	The average number of years of schooling for the population aged 15 and over.
Economy status	The average annual alcohol consumption per capita in liters of pure alcohol.
Life_expectancy	The average life expectancy at birth in years.

DASDBOARD

This is the big picture of our data visualization

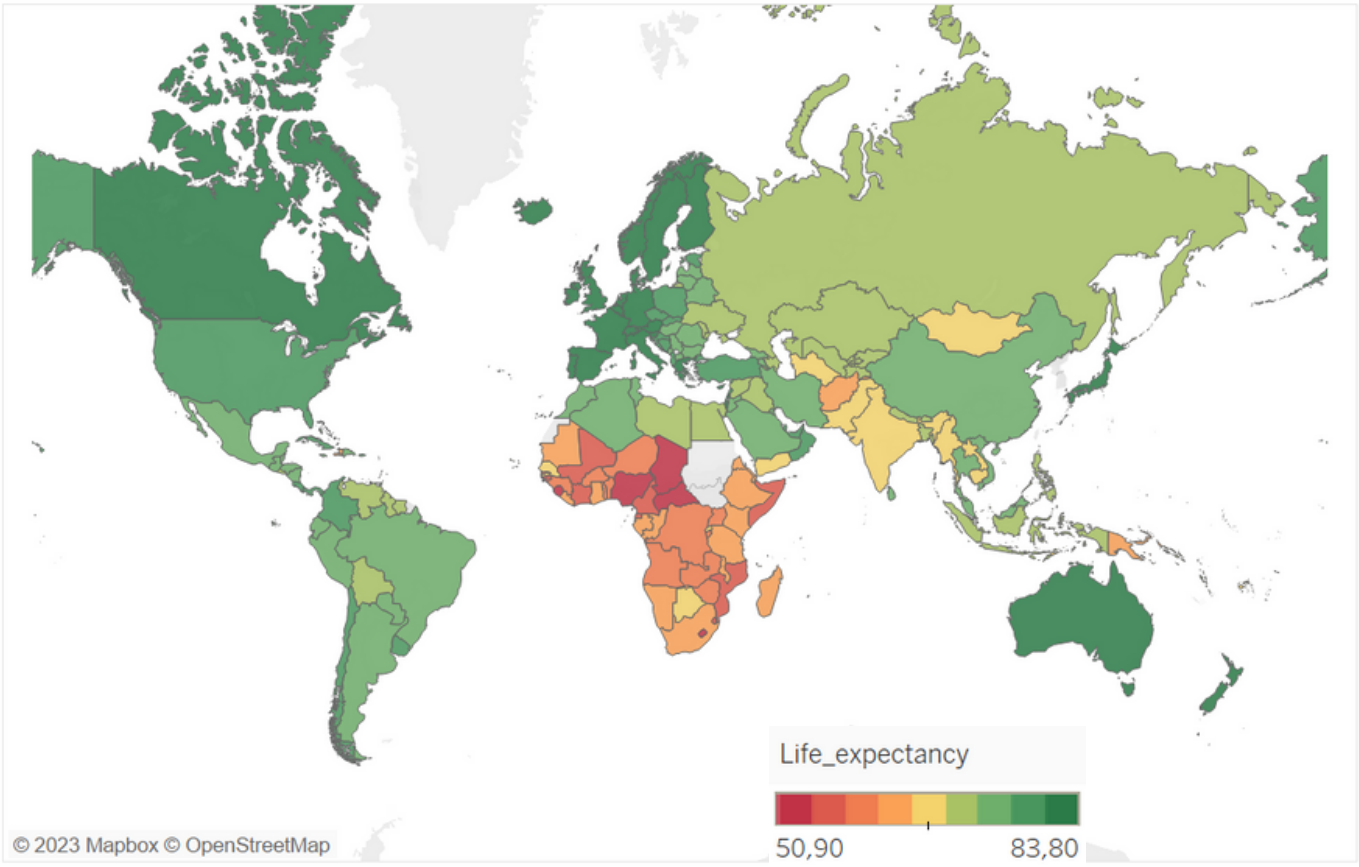


The dashboard shows life expectancy world map, rankings and it's histogram (2015). And it compares the average life expectancy between developed and developing countries over time.

World Visualization

In this section, we want to visualize something related to the world in general.

Life expectency, 2015



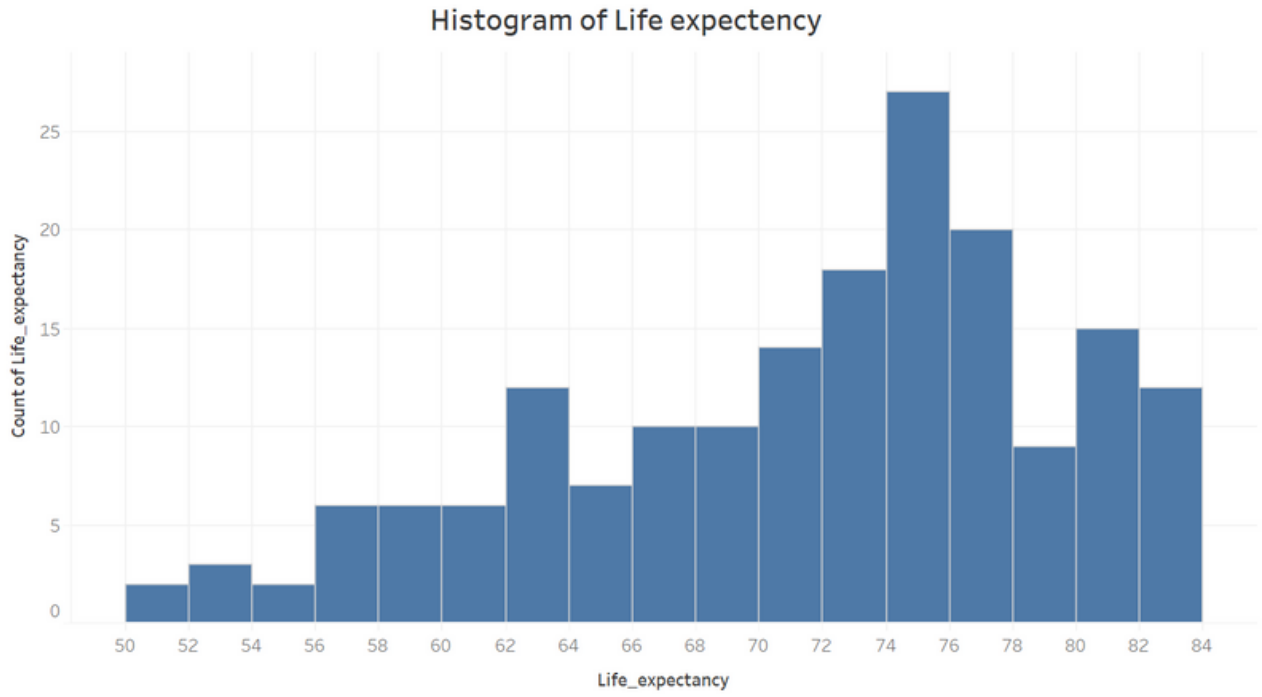
Top 10 highest life expectency countries

Index	Country	
1	Japan	83,800
2	Switzerland	82,900
3	Spain	82,800
4	Italy	82,500
5	Iceland	82,500
6	Australia	82,400
7	France	82,300
8	Sweden	82,200
9	Israel	82,100
10	Canada	81,900

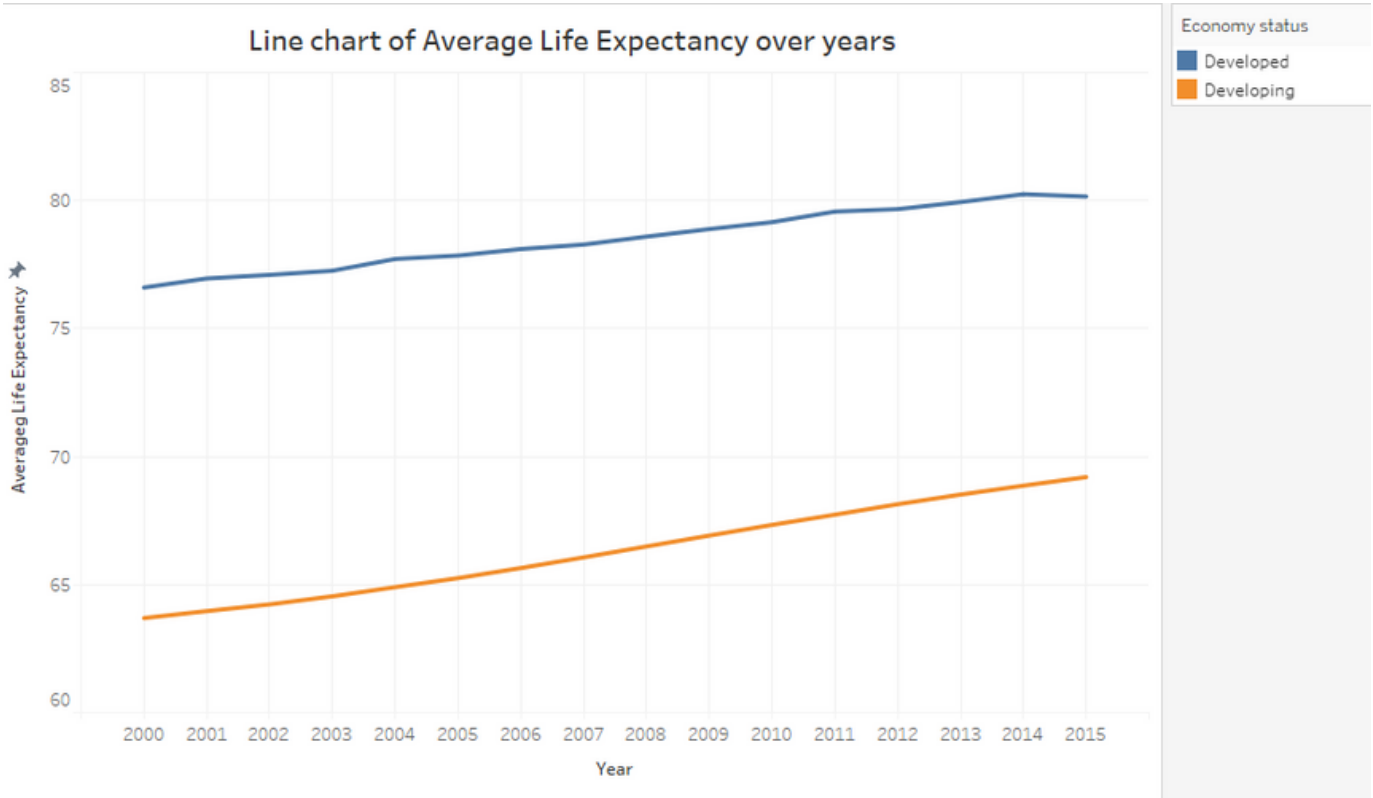
Top 10 lowest life expectency countries

Index	Country	
1	Central African Republic	50,90
2	Lesotho	51,00
3	Sierra Leone	52,90
4	Chad	53,10
5	Nigeria	53,10
6	Eswatini	55,40
7	Cote d'Ivoire	56,10
8	Mozambique	57,20
9	Zimbabwe	59,50
10	Malawi	62,00

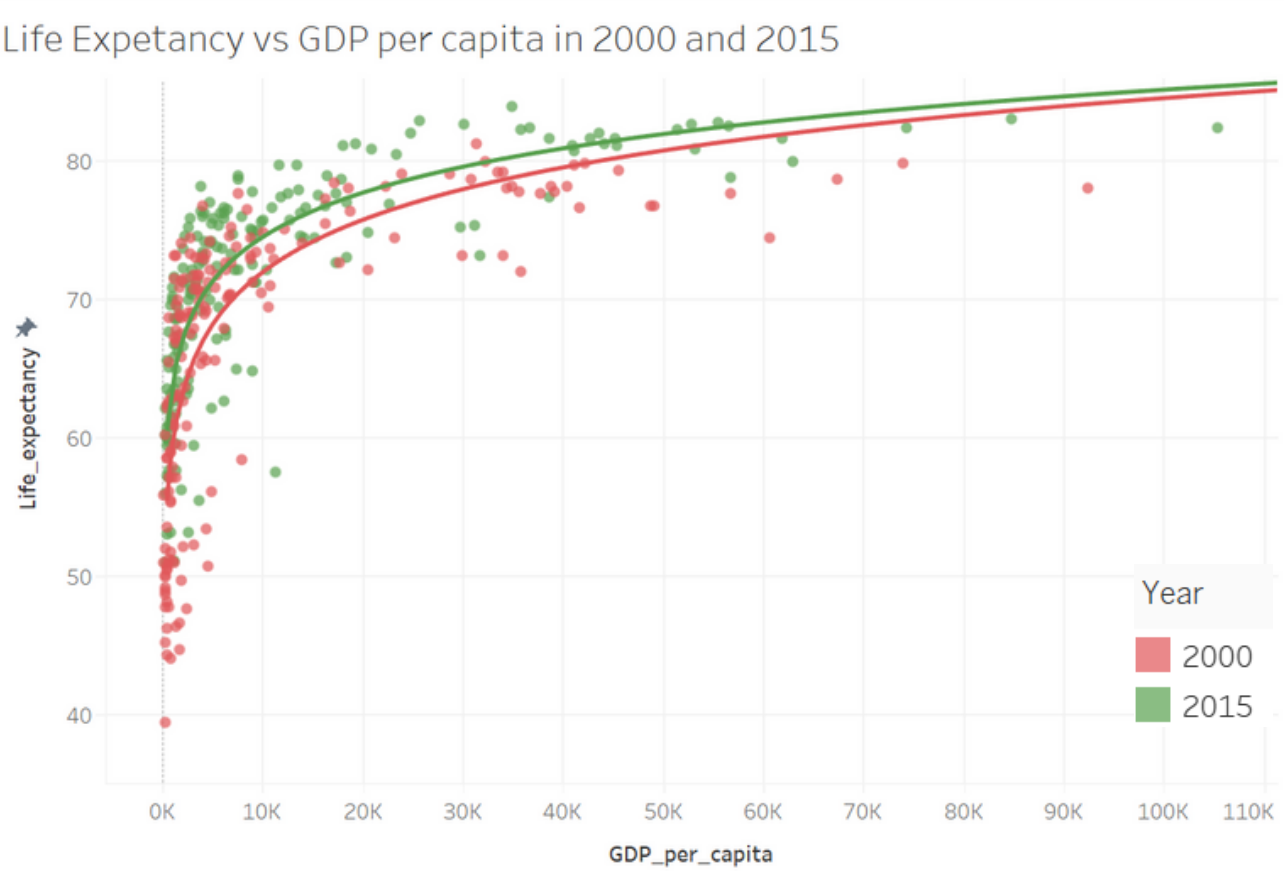
- The population of many of the richest countries in the world have life expectancies of over 80 years. In 2015, the life expectancy in Switzerland, Spain, Italy, Iceland, Australia, France, Sweden and Israel was over 82 years. In Japan it was the highest with close to 84 years.
- In the countries with the worst health (mostly in Africa), life expectancy is between 50 and 60 years. The population of the Central African Republic has the lowest life expectancy in 2015 with 50 years.



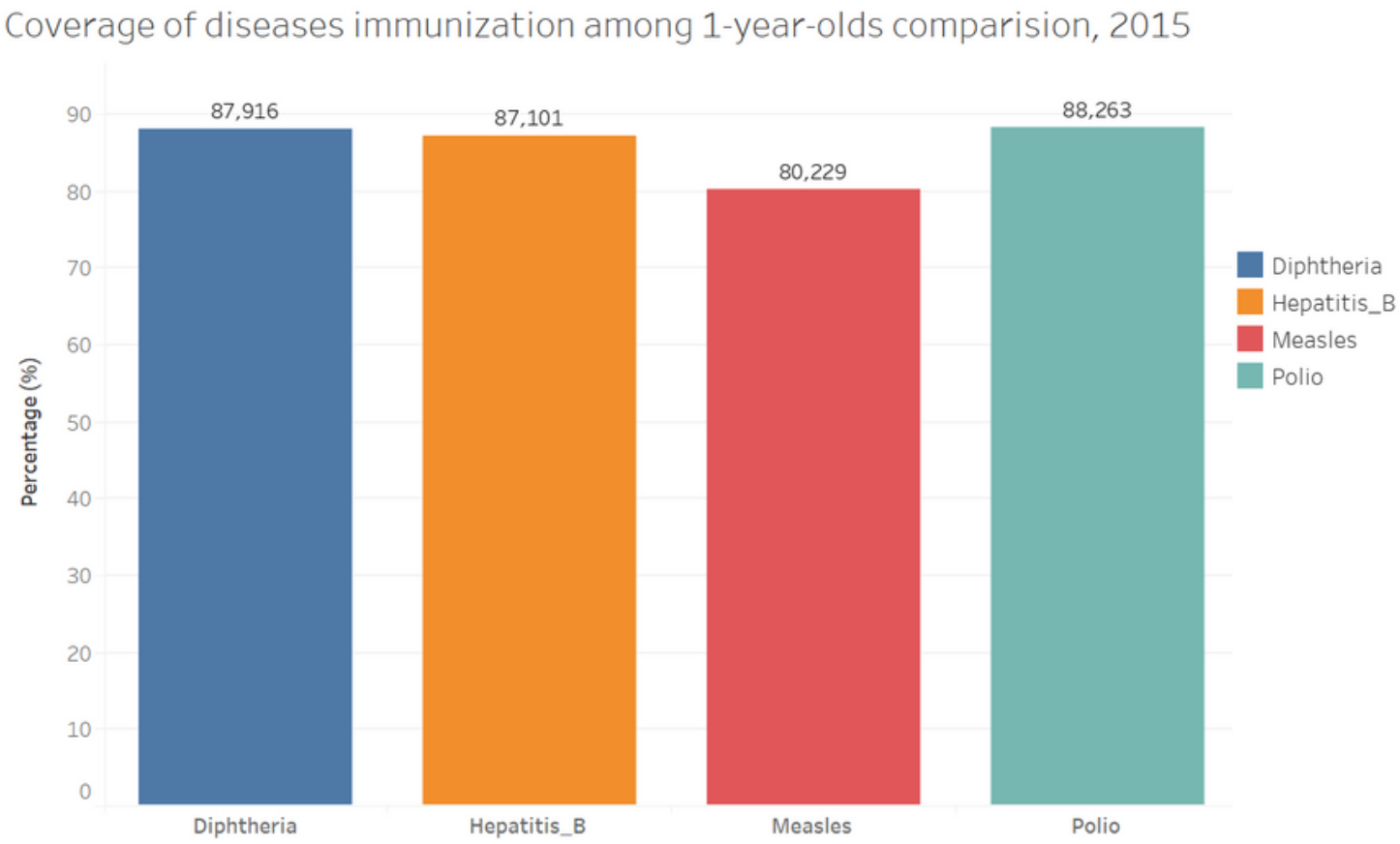
The life expectancy lies between 50 to 84, with an average life expectancy of 71.46 years. Most countries are in 74 to 76 years old range. In conclusion, it can be observed that human beings nowadays have a significantly high life expectancy.



It is easier to notice that developed countries always have higher life expectancy than developing countries. This can be argued to be achieved through higher standards of living, more effective health systems, and more resources invested in determinants of health.



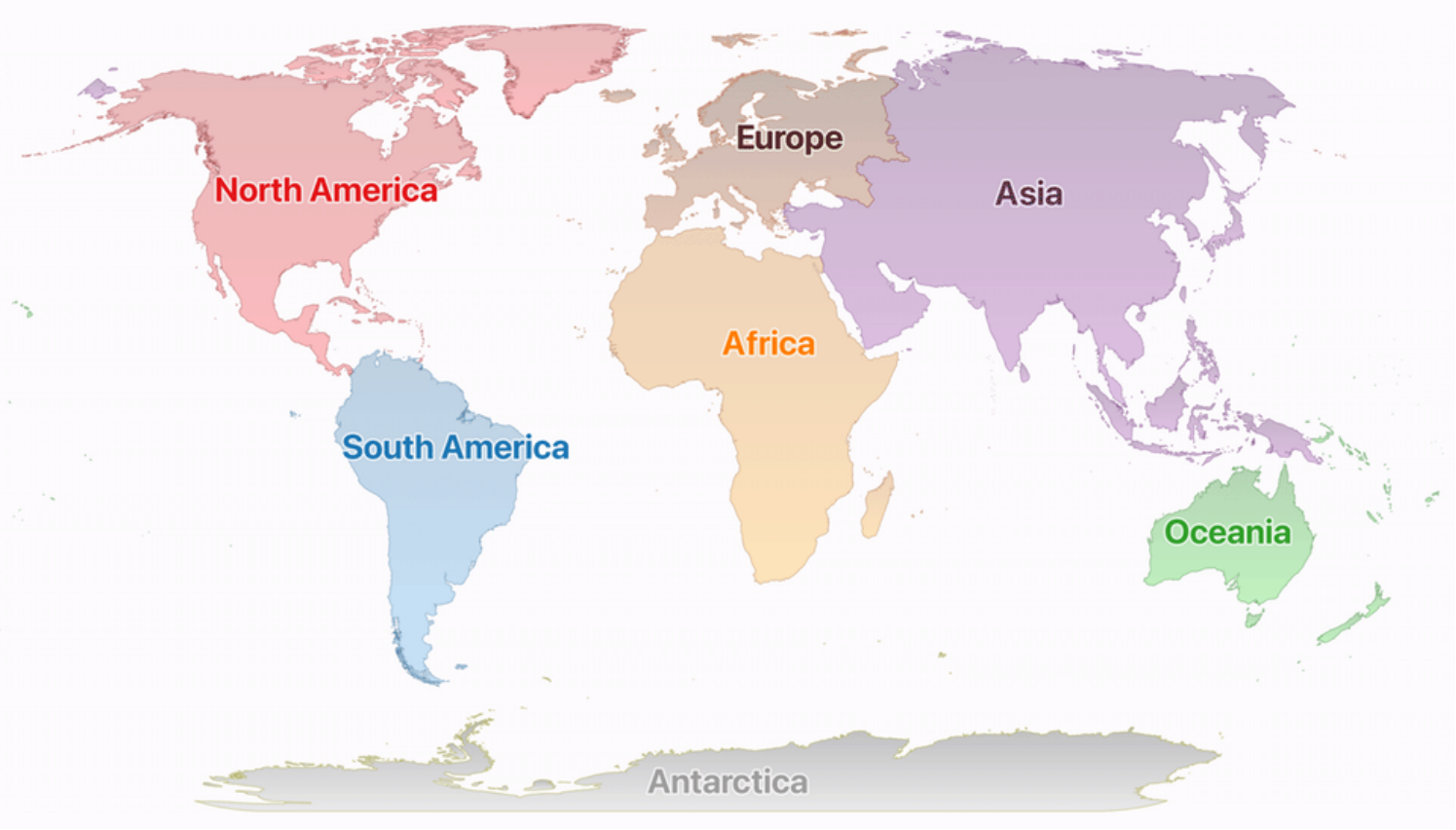
- The chart above shows that countries with higher levels of GDP per capita tend to have higher life expectancy. Wealthier nations typically have better healthcare systems, access to education, better nutritional standards, improved sanitation, and safer working conditions. All these factors contribute to improved overall health outcomes, which translates into longer life expectancy.
- From the chart above, we can find that the life expectancy associated with GDP per capita is rising over time. If economic development was the only determinant of health countries then we would see a steady relationship between the two metrics and the curve would not shift over time. Since this is not the case, we can infer that economic growth cannot be the only factor affecting health. A possible explanation for this changing relationship is that the advancement of science and technology has led to some very effective public health interventions and also make them more affordable for people.

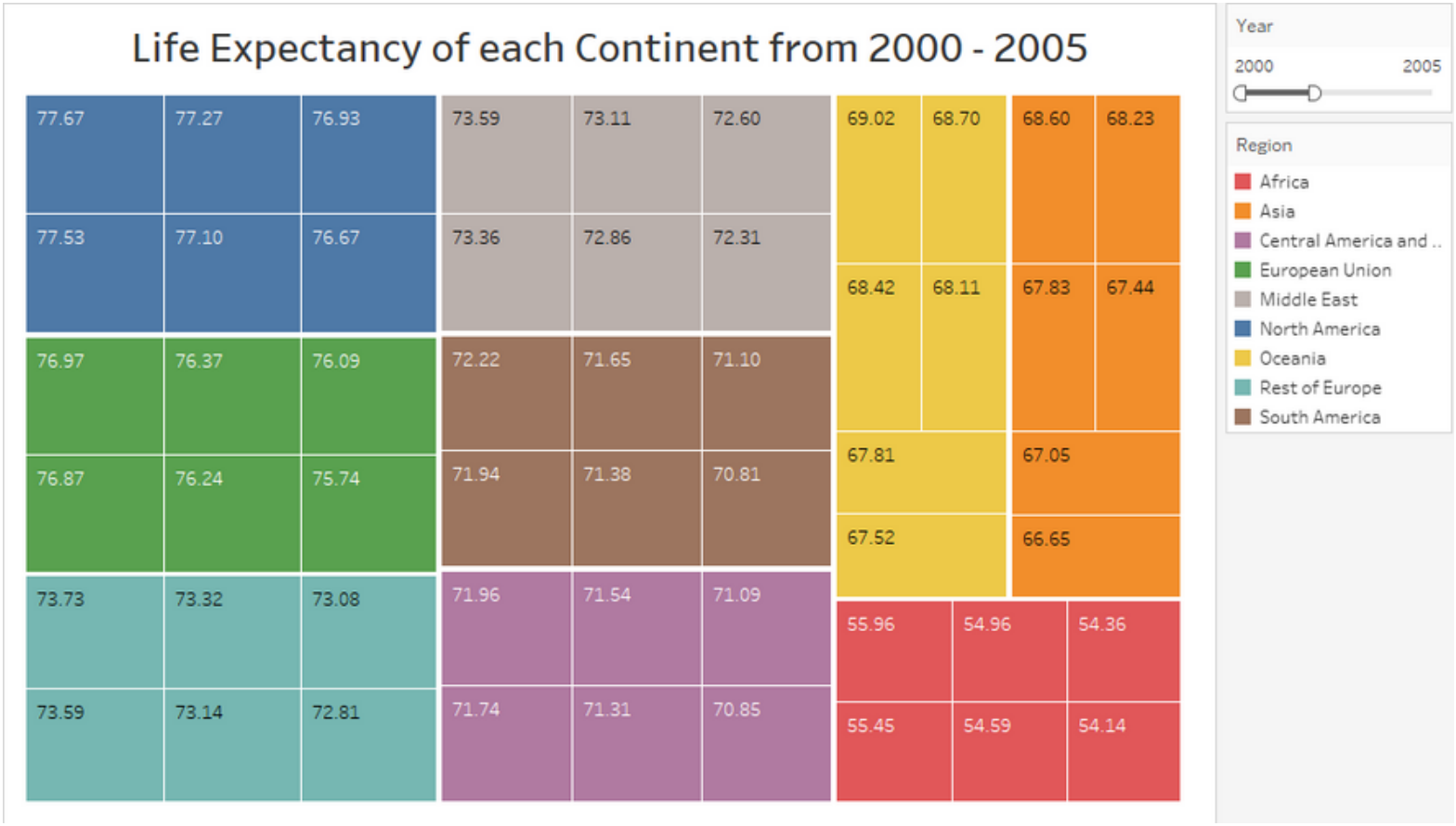


- Immunization coverage is essential to protecting young children from preventable diseases that can cause serious health complications and even death. By ensuring high vaccination rates among 1-year-old, we can help keep them healthy and safe.
- Although immunization coverage rates vary widely between regions and countries, but from the chart above (average percentage), we can be assured that most babies are receiving proper health care in the present era.

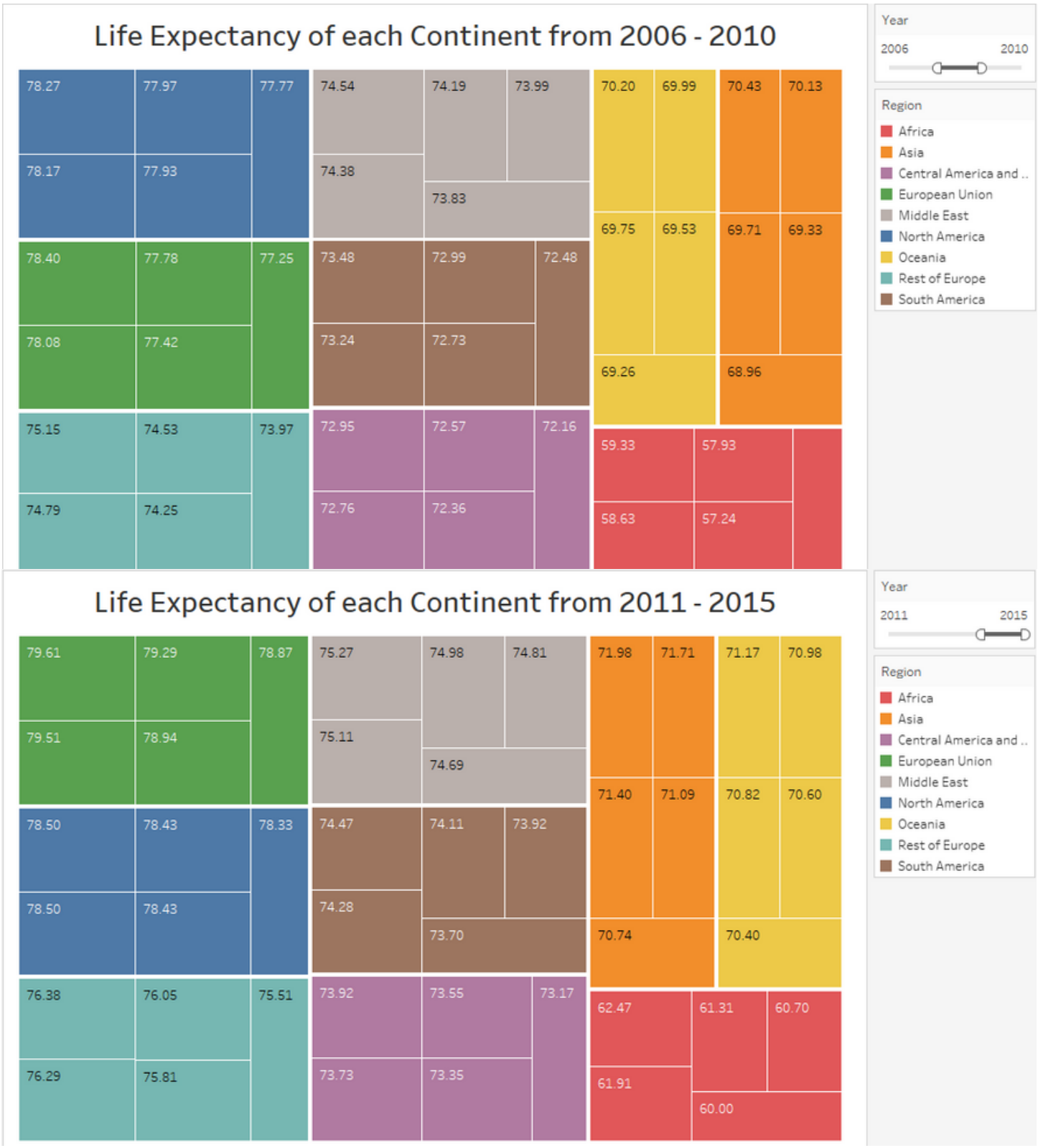
Continent Visualization

In this section, we want to visualize something related to the continent in general.



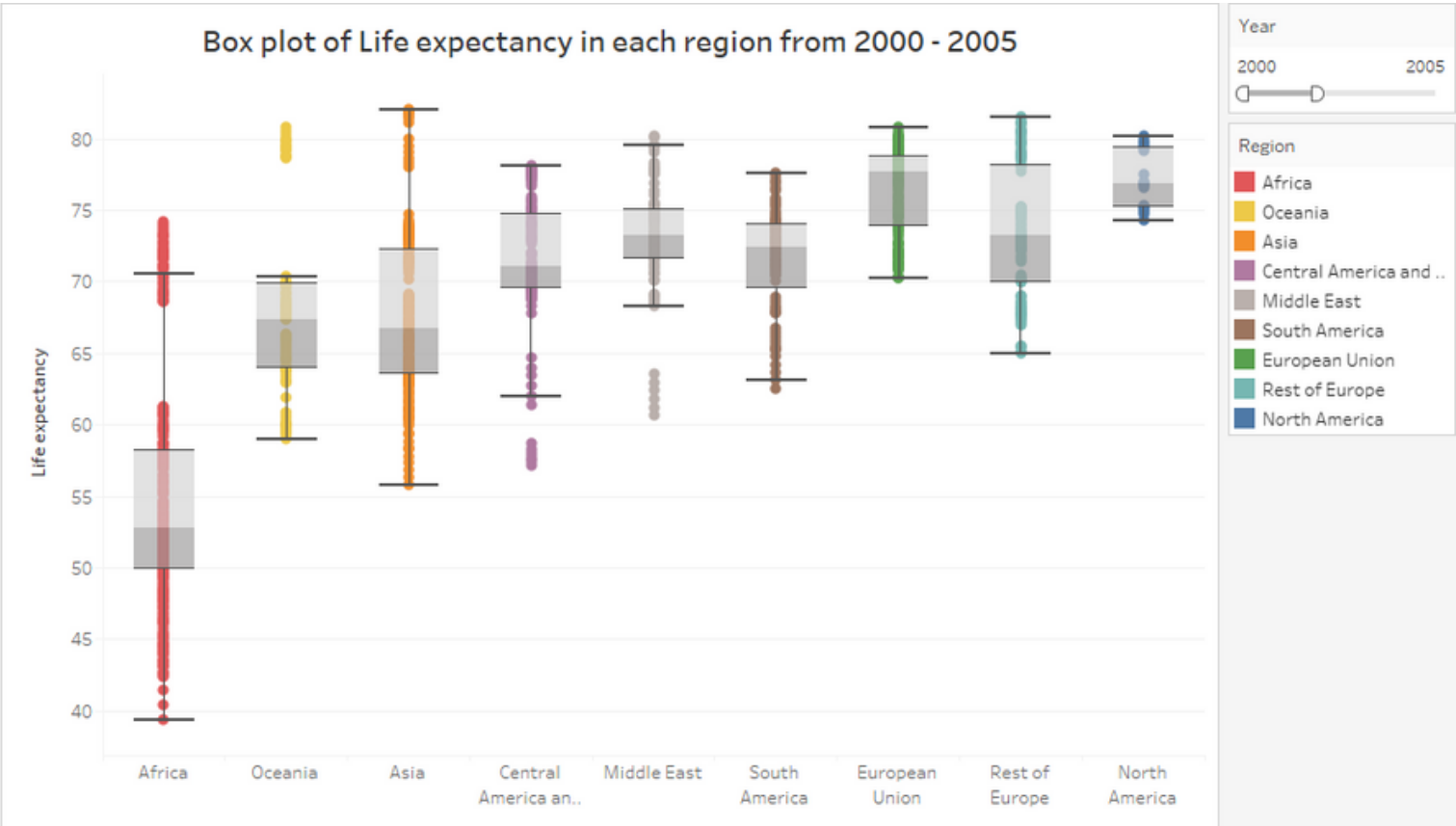


- We split 15 years of data into 3 period: first period, middle period and final period.
- In this first period we can see clearly that the range of life expectancy is from at least 54 year-old (Africa) and up to nearly 78 (North - America).
- Top 3 continents have the highest life expectancy is North - America, European Union and Rest of Europe, all is over 72 year - old.
- In contrast, Africa, Asia and Oceania has the lowest longevity, all is under 69 year - old (over 57).
- In general, we can confirm that the longevity of every continent increased year by year although that value is small.
- To be more specific, there is no doubt to say that there is a really big distance in life expectancy of each continent, which is about 23 in 2000 and approximately 22 in 2005. The distance was narrowed but still big.

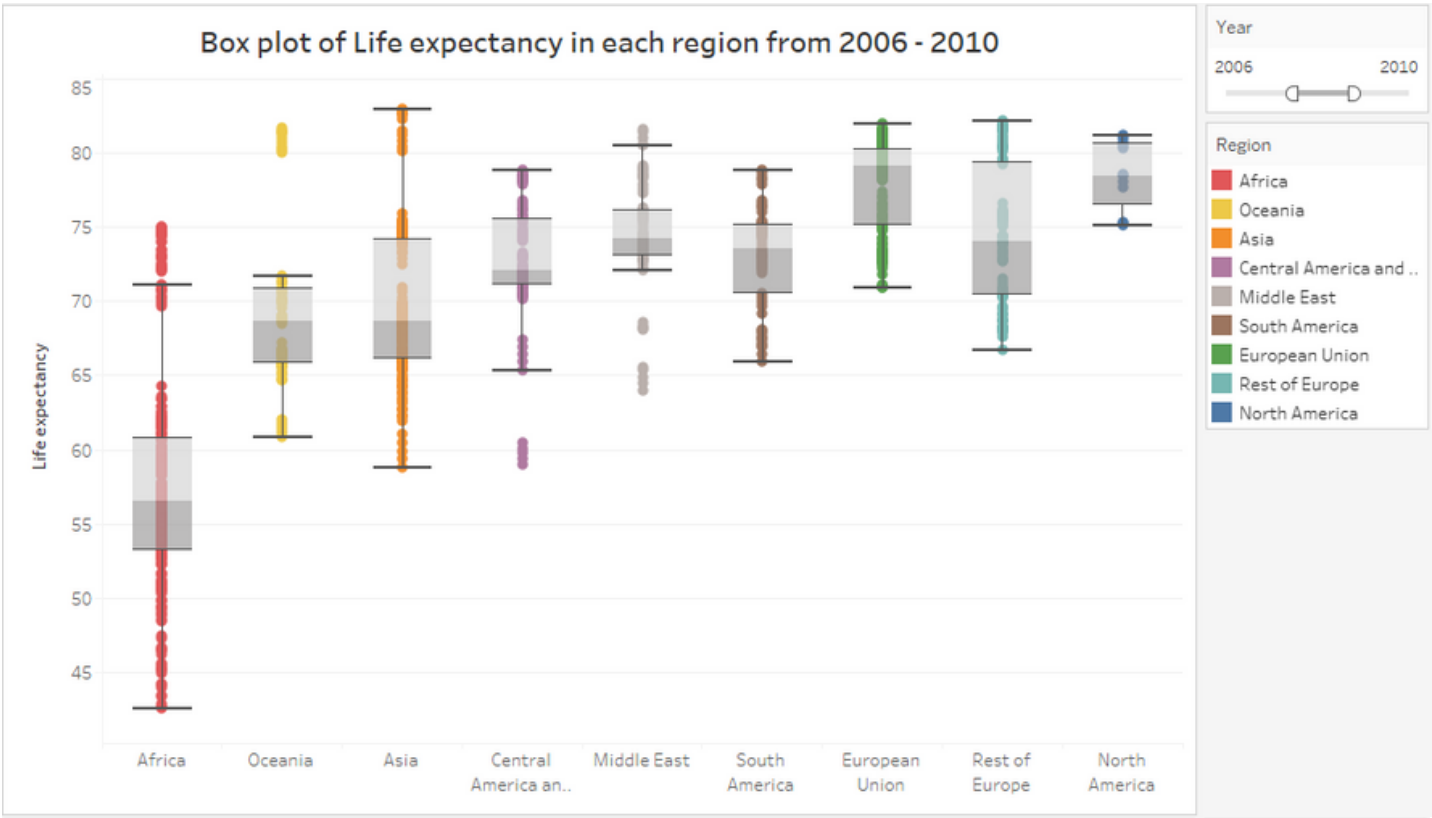


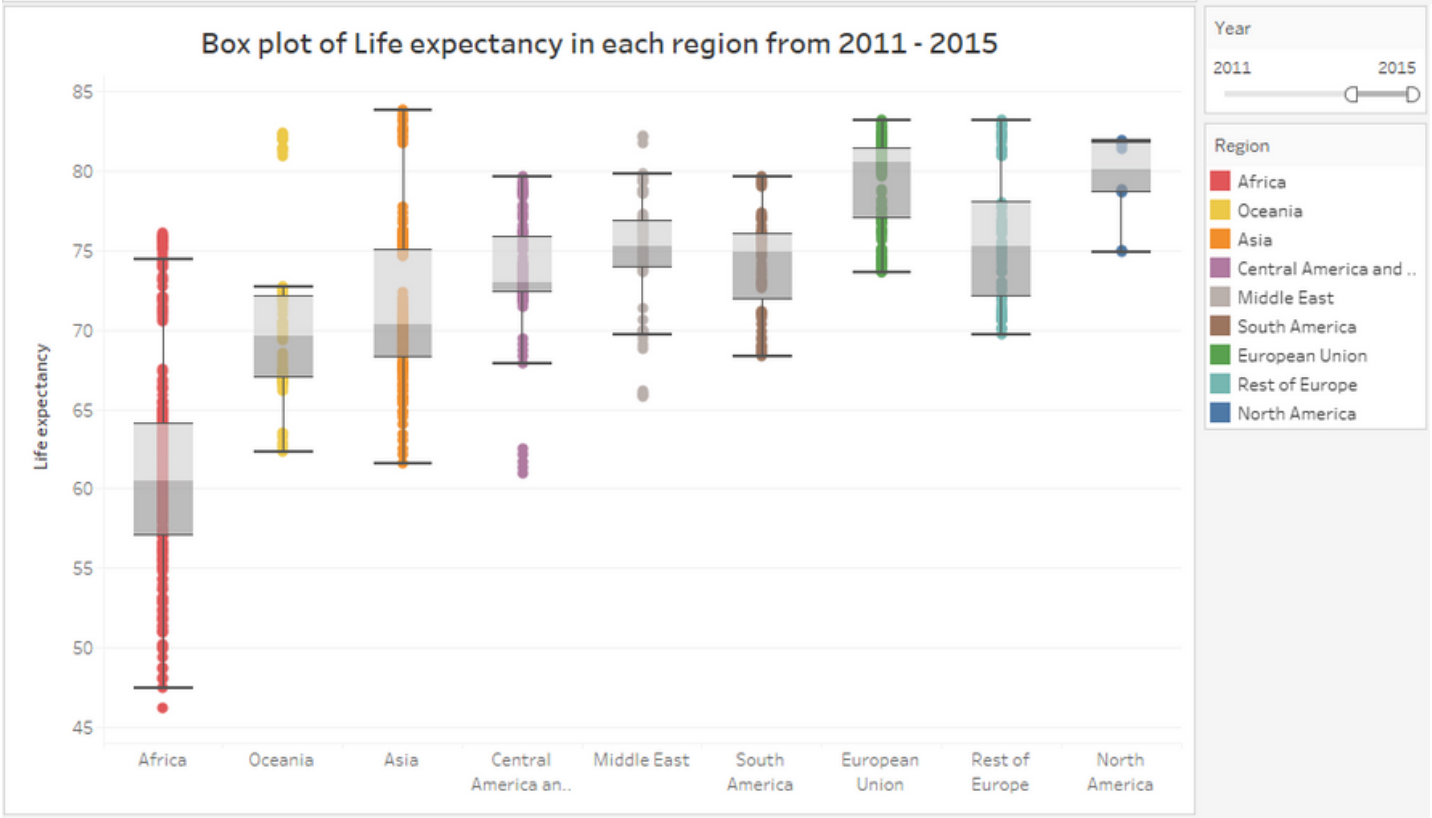
- Follow the middle and final in researching period, there is some changes in the order of longevity. In the final period, European Union has the highest longevity, and also in this period, the longevity of Asia is higher than Ocenia.
- But there is positive sign, every continent has the increase in life expectancy, specially, Africa up to 8 years (from 54 in 2000 to 62 in 2015). Although the distance range between continent to continent is still so far - 17 years.

Let's look at a few statistics



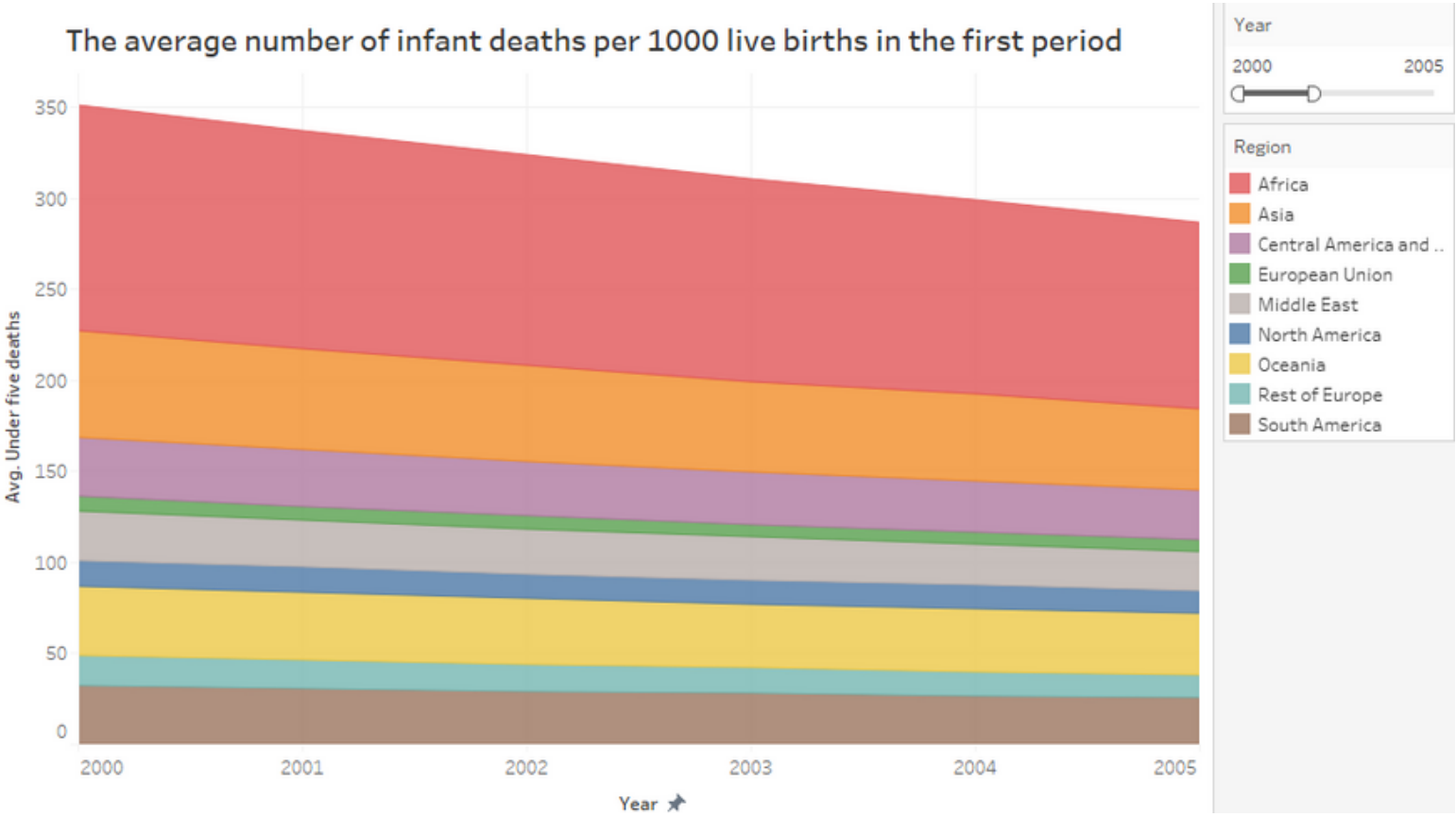
- During the time between 2000 and 2005, this box plot can prove that the range of longevity in each continent is so different. Almost countries in the continent they belong to have the same trend, there are still some exceptions.
- In a big view, top 3 continents which have the highest average life expectancy also have the smallest range in distribution, there is no outliers in these 3 continents.
- This can prove that life expectancy in every countries of these 3 continents are high together.
- Otherwise, Africa has the largest range from 40 up to about 70 in this period. In this continent there are some countries which are over the range such as: Algeria, Tunisia,...
- In Oceania the range is narrow but low. The highest index in the range of this region is about 70 - which is still lower than the minimum index of life expectancy in North America. Although the range of this continent is short and low, there are still some countries which have the longevity up to 80 like: Australia or New Zealand.
- Central America & Caribbean and Middle East is normal, not to high and also not to low, but there are some outliers: Haiti of Central America & Caribbean or Yemen of Middle East are low, this index is about 60.





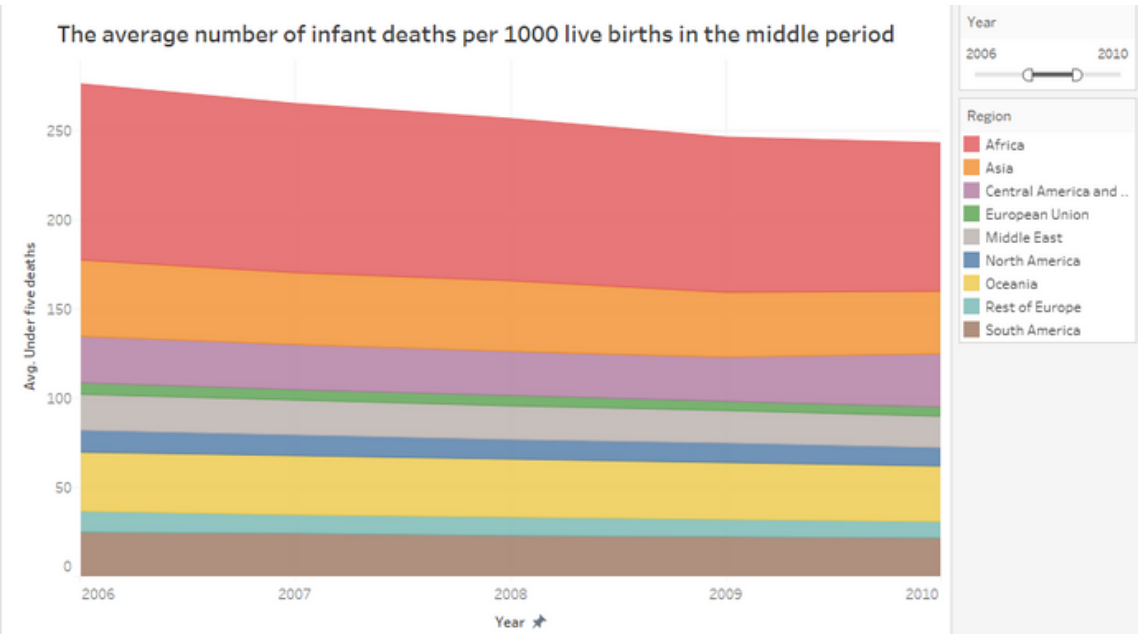
- Over the 2-period the box plot distribution shows that there is no remarkable changes, all of continents have the positive increase steadily.

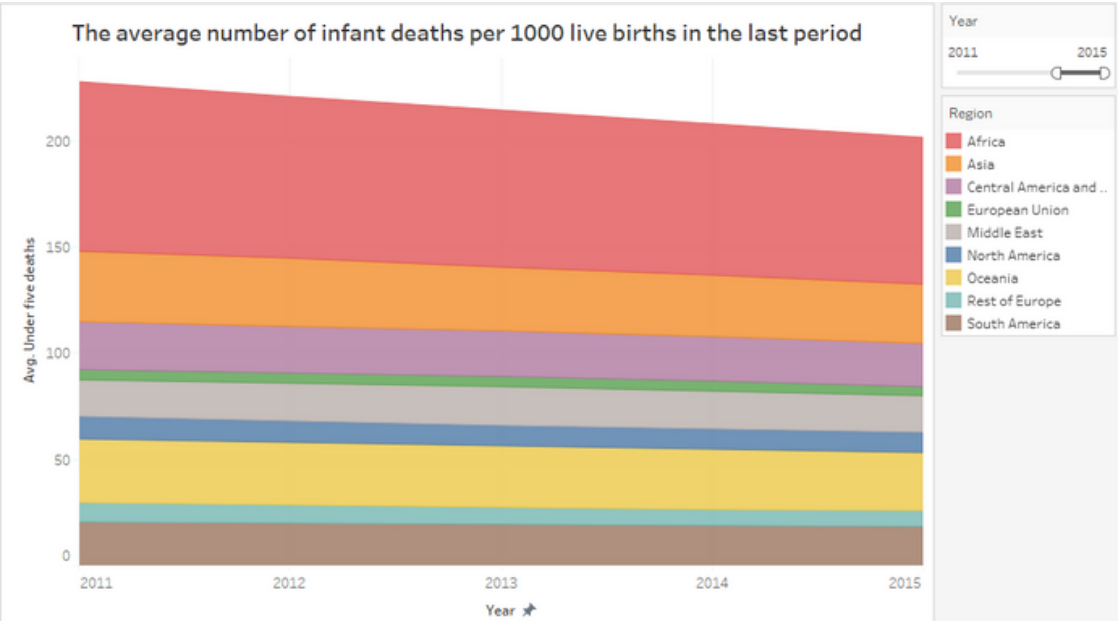
Concentrate more about the economy, education and society in each continent



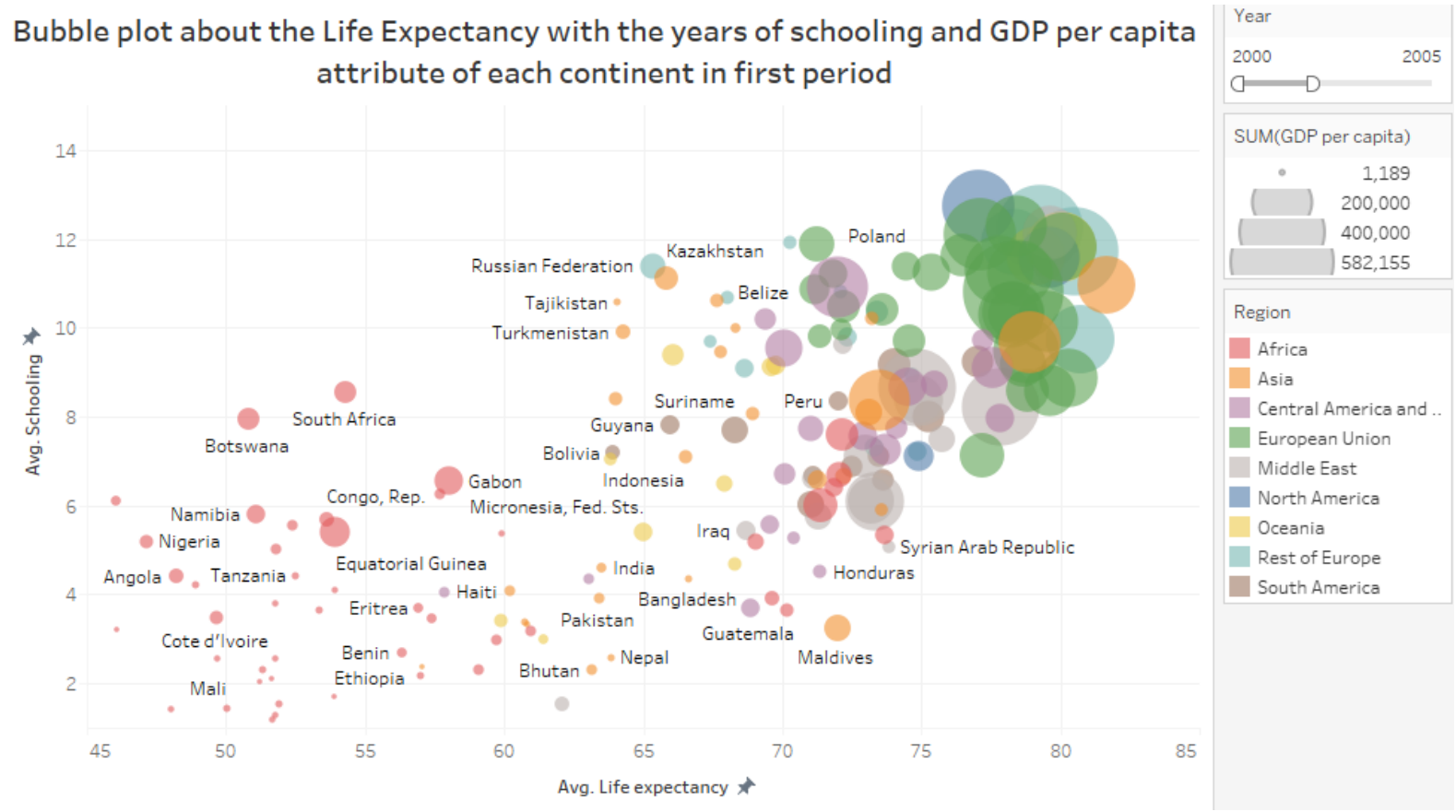
- From 2000 to 2005, we can clearly find out that, Africa had the highest number of infant deaths compared to the orther continents (up to 350 death cases per 1000 live cases. After that is Asia, which is about 230 cases in 2000.
- In contrast, South Ameria and Rest of Europe had the lowest number of death infants in 2000 (under 50 cases). This order of these regions keep the same till 2005 with the trend of decreasing gradually.
- Now, we will compare the first period to the middle and last period of researching.

Concentrate more about the economy, education and society in each continent

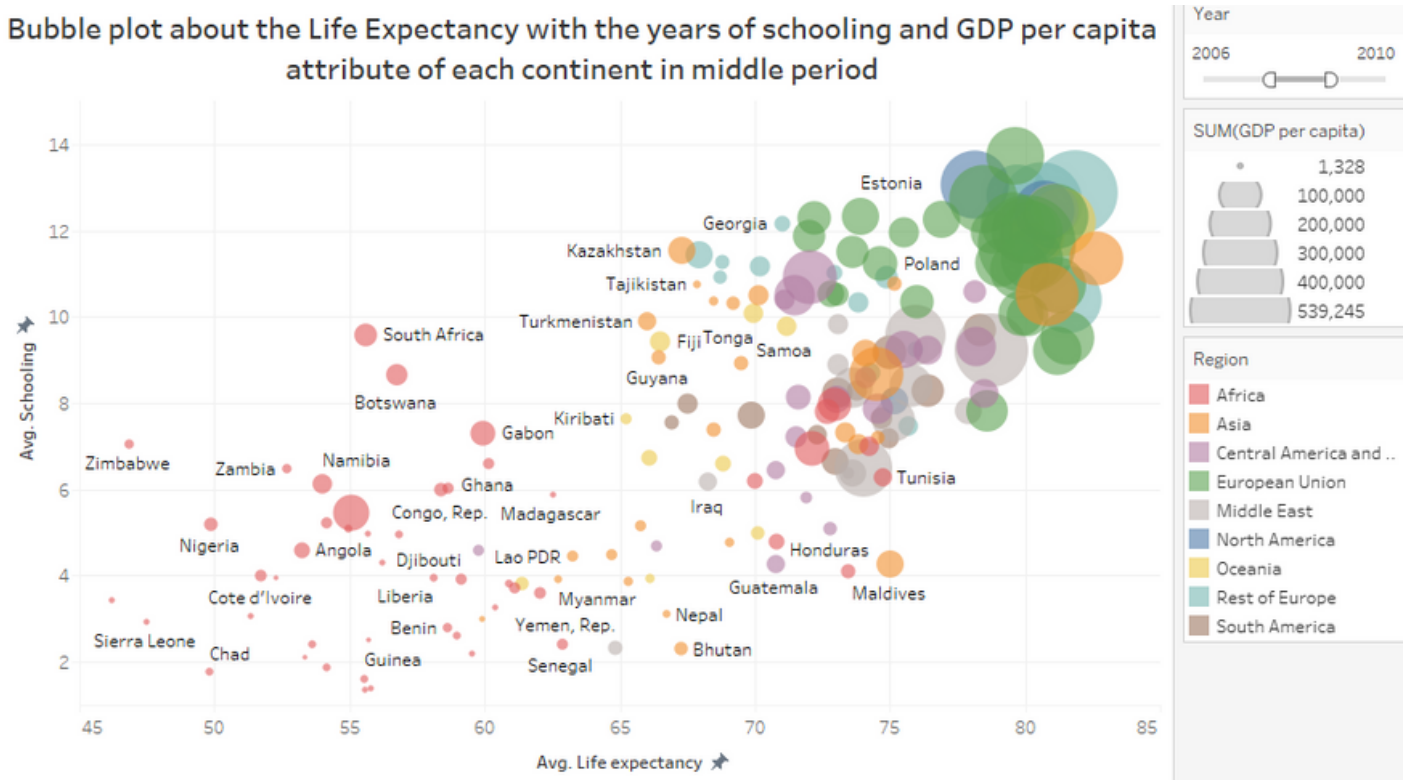


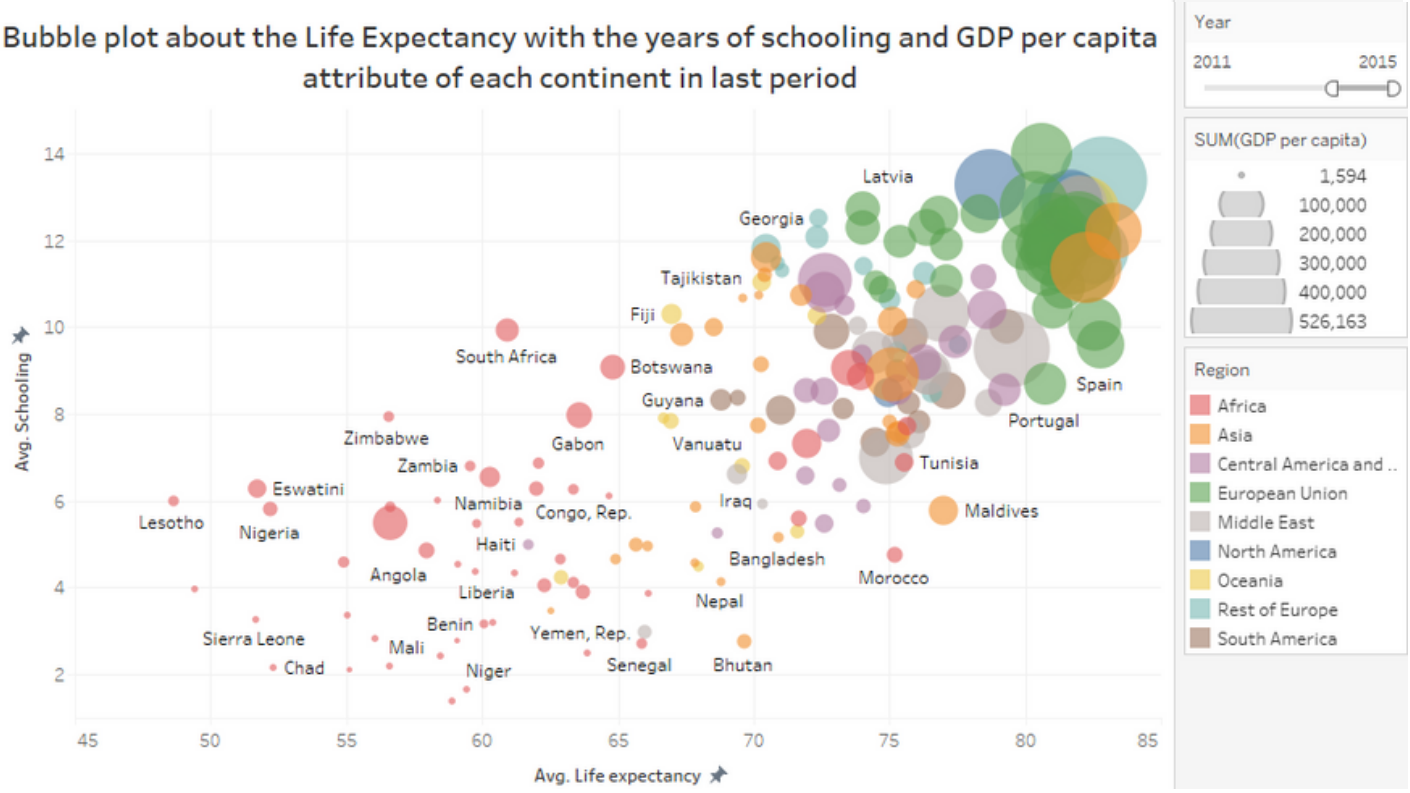


- Look at these period, there is no changes about the order of death infants per 1000 live births, so that we are quitely confident to comment that this trend is still true nowadays.
- Furthermore, we can remember that Africa and Asia are in top 3 of lowest life expectancy continents. Therefore, it can be a clue to prove that infant deaths can affect at a level to the life expectancy.
- Through these 3 area charts we can also have an idea that the medical quality of Africa should be improved, this may be a good way to enhance the life expectancy of this continent.



- This is a bubble plot showing the trend of life expectancy for different continents.
- The size of bubble plot defines the GDP per capital of each country.
- We can see that the red colour appears alot in the 45 - 60 year-old. There is no doubt to say that it is the continent which has the lowest longevity. Besides, almost countries in Africa have the years of schooling under 8.
- The majority of life expectancy is in the range from 65 to 75. Central America & Caribbean has most countries in this range.
- We can also see that the size of bubbles in this range is small which can confirm that the economy there is still poor.
- European Union and North America are the two continents which has most countries have the longevity higher than 75.
- Interestingly, not only have the highest life expectancy, the bubbles in those locations are the biggest which can show that in these countries, economy developped strongly.
- Before make sure some ideas, we will also visualize the middle period and last period.

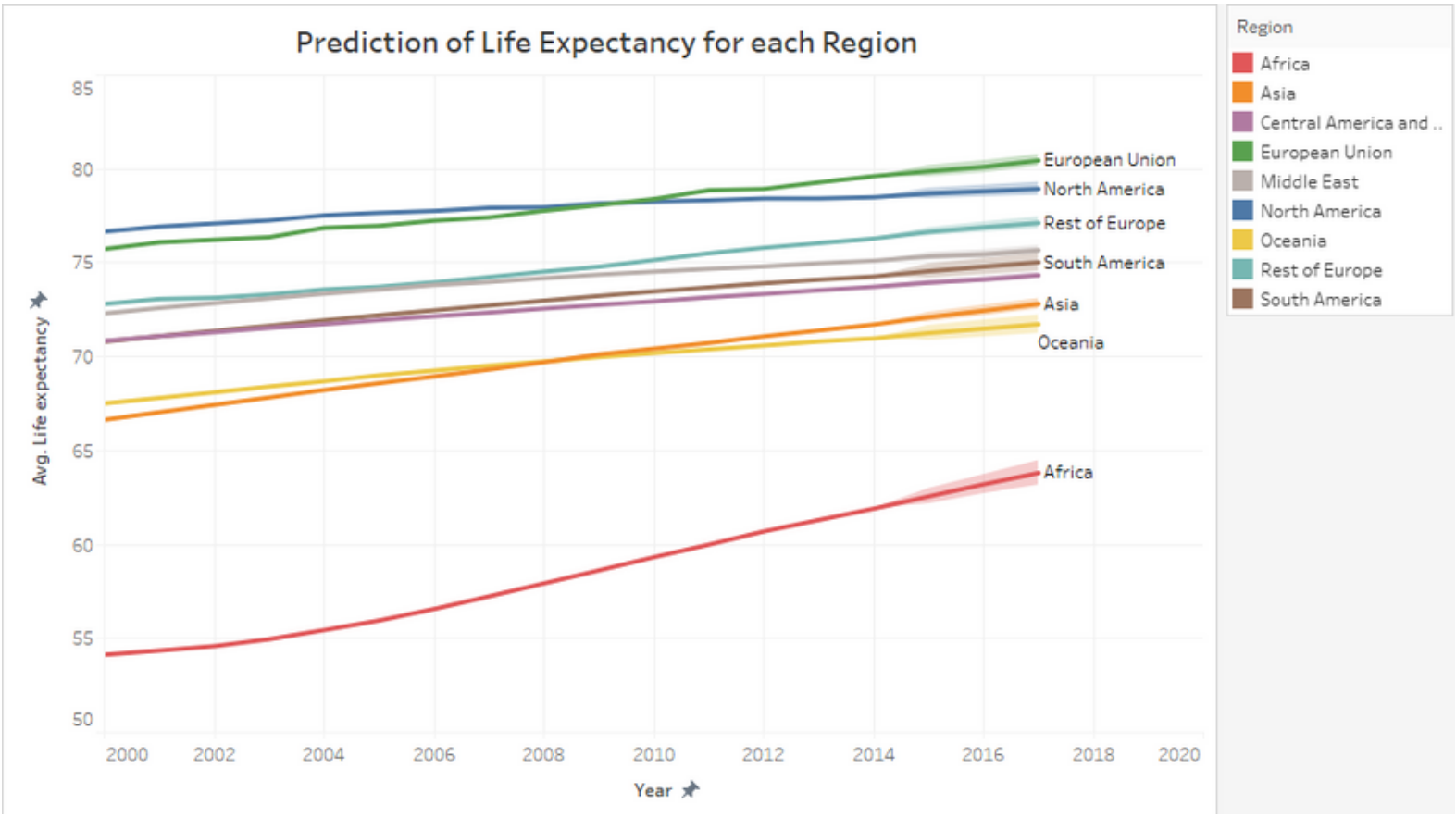




- There is no remarkable changes through all 3 period of researching time. Therefore, we can find out some interesting ideas: First, the more years we study, the more life expeactancy we can get. This can be explained if we study more, we can have more scientific knowledge about food, sport, lifestyle,... which can help us aware of what is dangerous to liminate and life expectancy can be longer as a result.

- Second, the richer a country can be, the longer of life expectancy their citizens can get. In is not difficult to explain this, if a country has a good economy, their citizens can improve their cost of living, they care more about healthy meals, medical care,...which can lead to higher longevity. To sum up, not only medical quality affects on longevity, but also economy and education can affect.

Predict life expectancy for each region

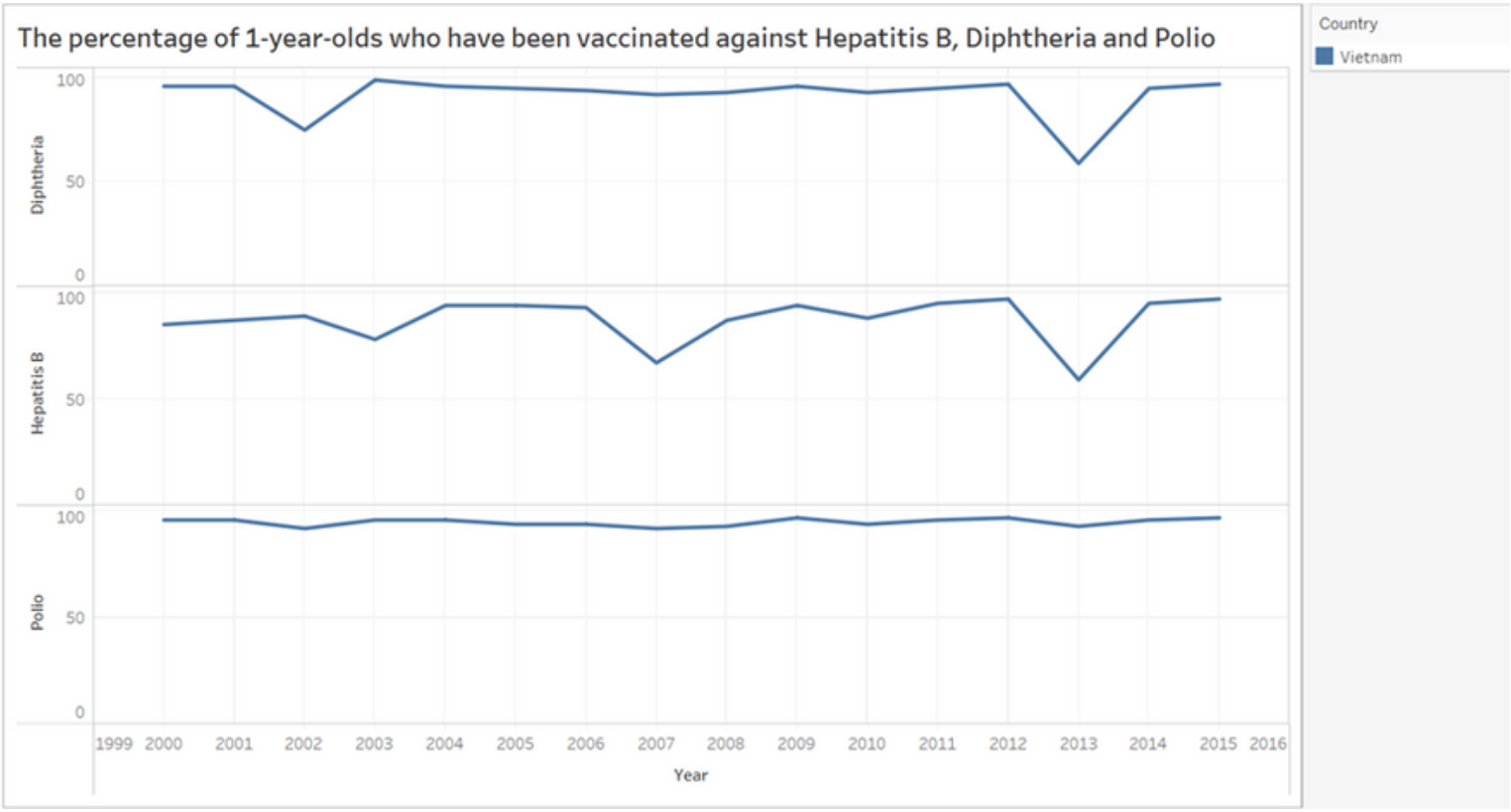


- This is a tool in Tableau, as we can see: all of continents have a rise in life - expectancy, Africa has a significant increase although after the growth, thê longevity of this continent still much lower than the others.
- Every continent has a positive trend.

VIETNAM WITH COMPARISON

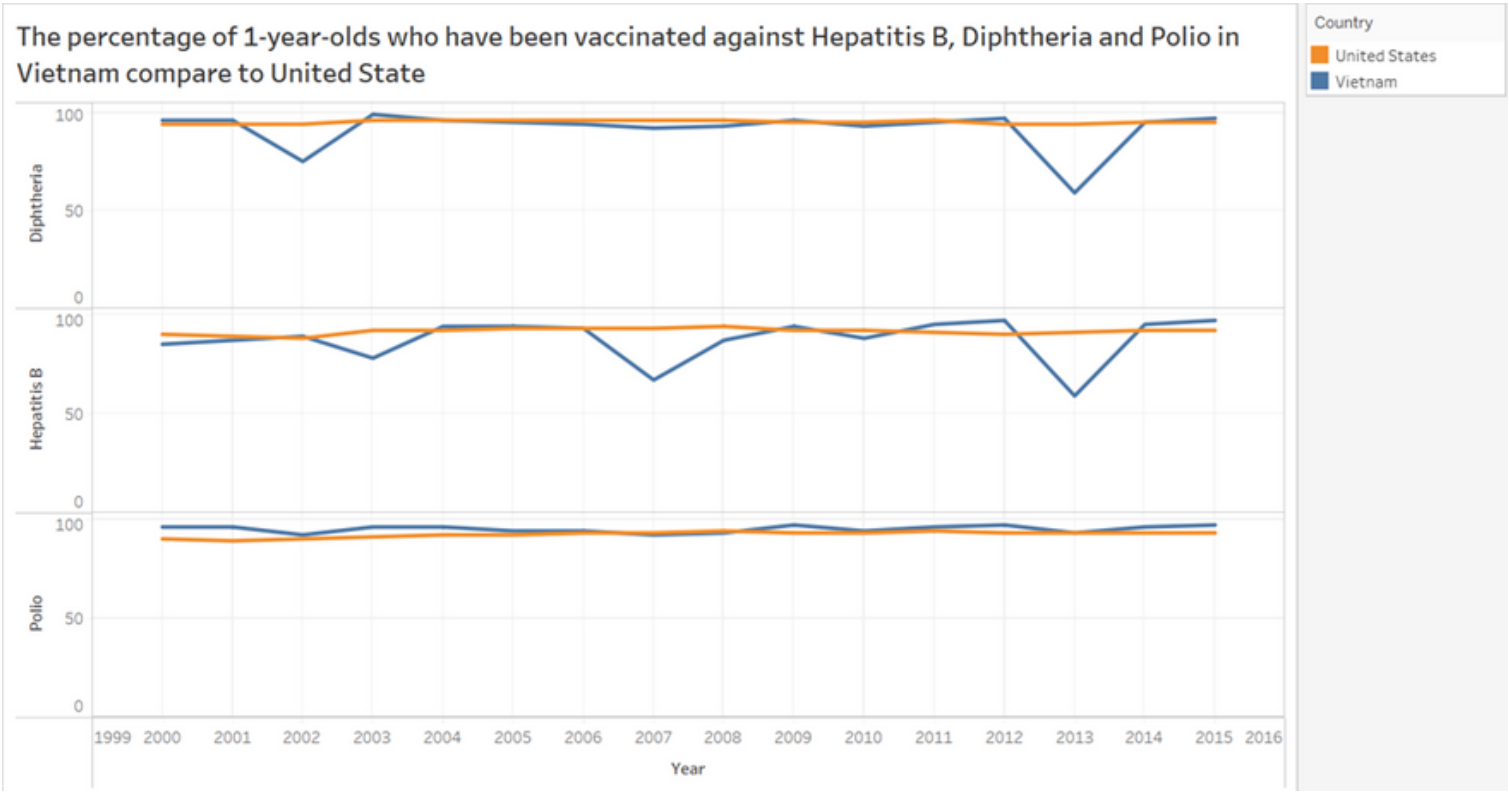
In this section, we want to visualize something related to Viet Nam.

Vaccination against Hepatitis B, Diphtheria and Polio for 1-year-olds in Vietnam



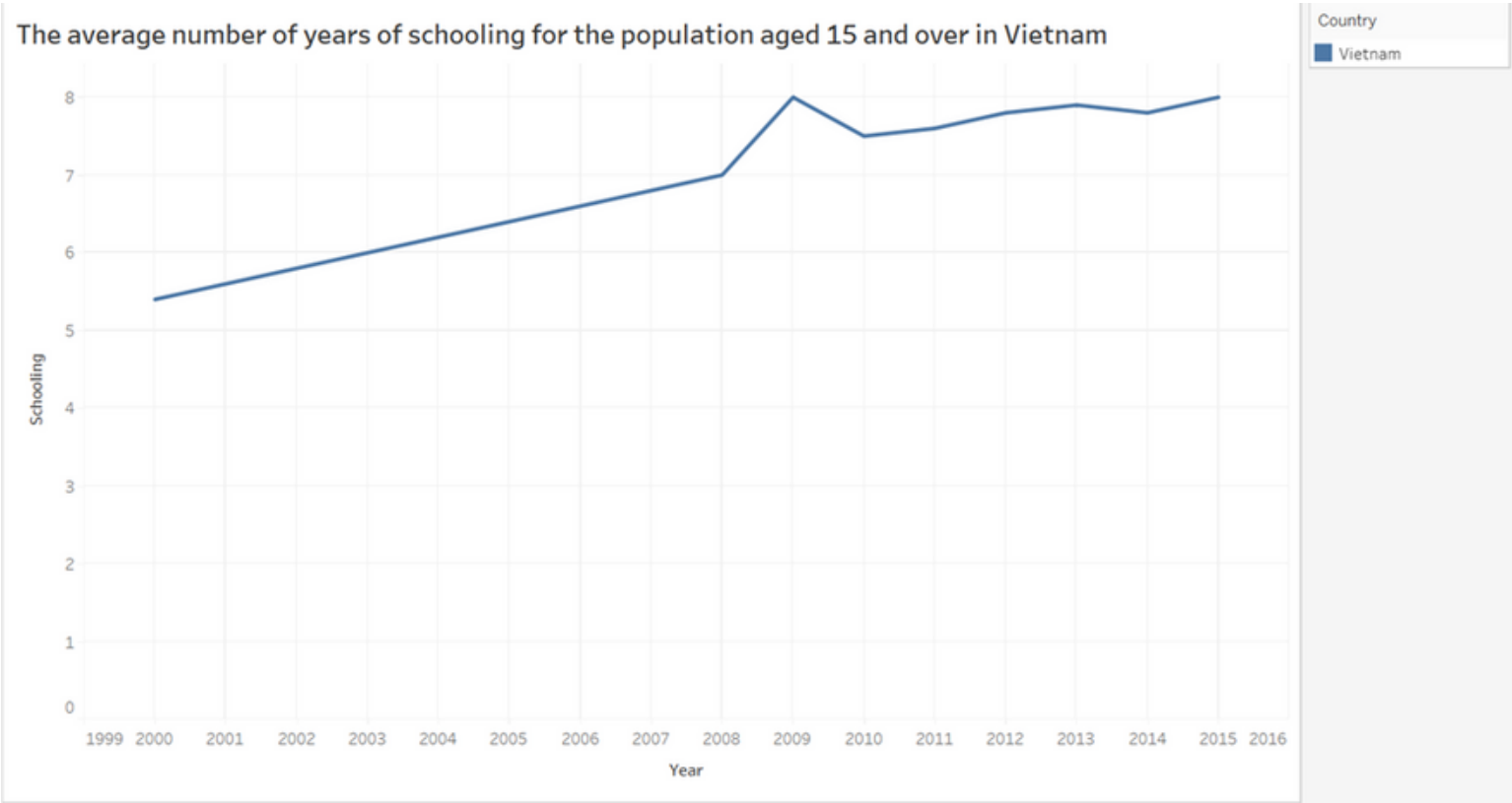
- Chart used: lines chart
- The reason why we want to use lines chart to visualize the vaccination status in Vietnam is:
  - The percentage of 1-year-olds who have been vaccinated changed over time.
  - Show the trend.
- Comments:
  - Generally, Vietnam is a country that is relatively focused on people's health through the percentage of 1-year-olds who have been vaccinated against some diseases is quite high.
  - To some dangerous diseases such as Diphtheria and Polio, the percentage of vaccination is consistent and very high (approximate 100%) especially to Polio.
  - The percentage of vaccination against the Hepatitis B is not consistent maybe because Vietnamese don't find this a dangerous disease, so they don't get vaccinated. The lowest percentage is 59% in 2013 but this percentage is still acceptable.

Vaccination against Hepatitis B, Diphtheria and Polio for 1-year-olds in Vietnam compared to United States



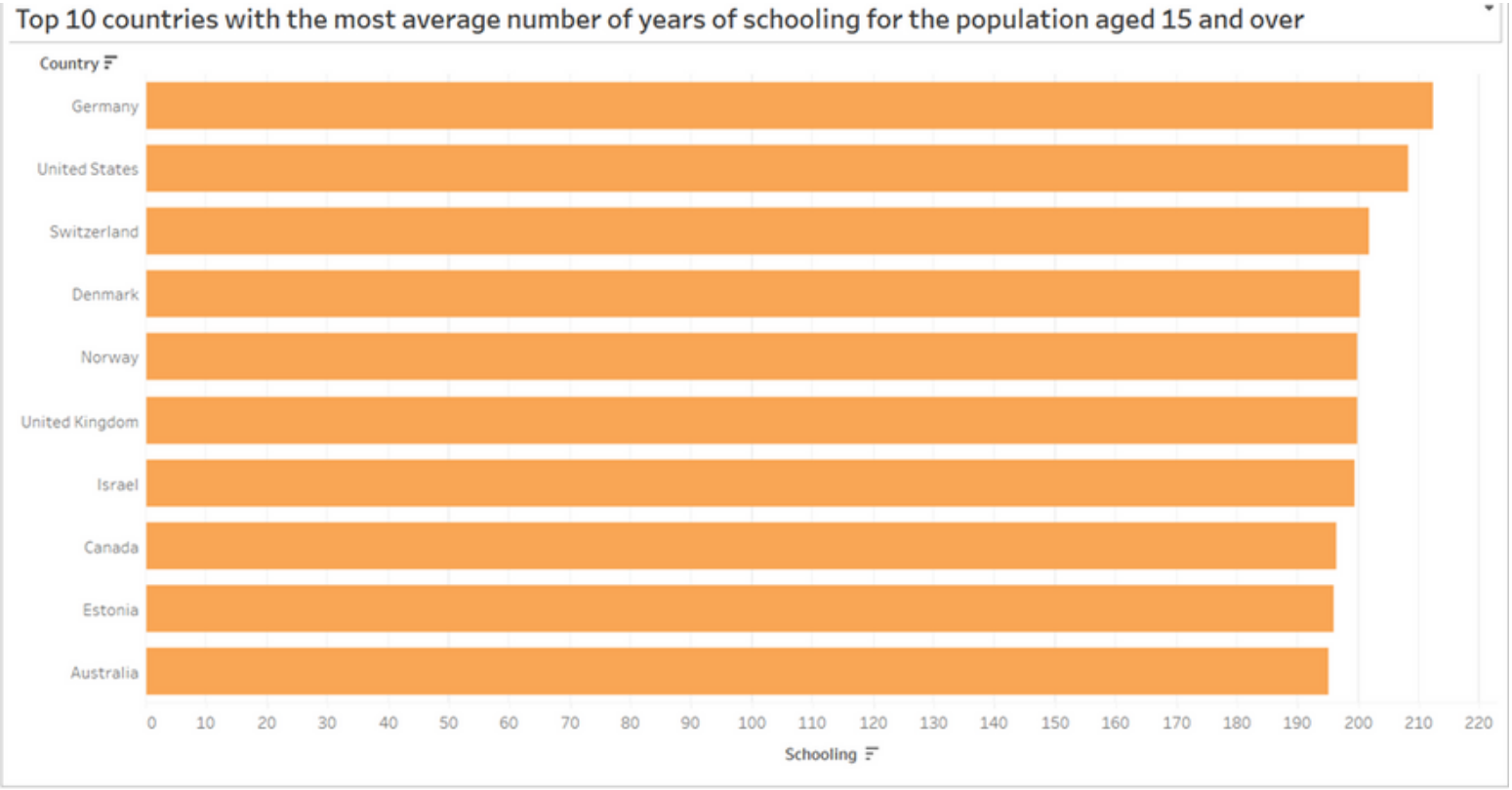
- Chart used: lines chart
- The reason why we want to use lines chart to visualize the vaccination status in Vietnam is:
  - The percentage of 1-year-olds who have been vaccinated changed over time.
  - Show the trend.
  - Compare indexes between Vietnam and United States.
- Comments:
  - As we can see, the percentages of vaccination against Hepatitis B, Diphtheria and Polio are all consistent and very high which means USA has a very developed medical background. No matter what kind of disease, they always have well prepared through their vaccination.
  - Although Vietnam is a developing country, but we can see that the percentage of vaccination in Vietnam is just a bit lower compared to the percentage of vaccination in USA which means the medical background of Vietnam has made a great effort to protect Vietnamese's health and has made a significant improvement.
  - Especially to Polio, which is a dangerous disease, the percentage of vaccination in Vietnam is a bit higher than the percentage of vaccination in USA.

The schooling situation in Vietnam



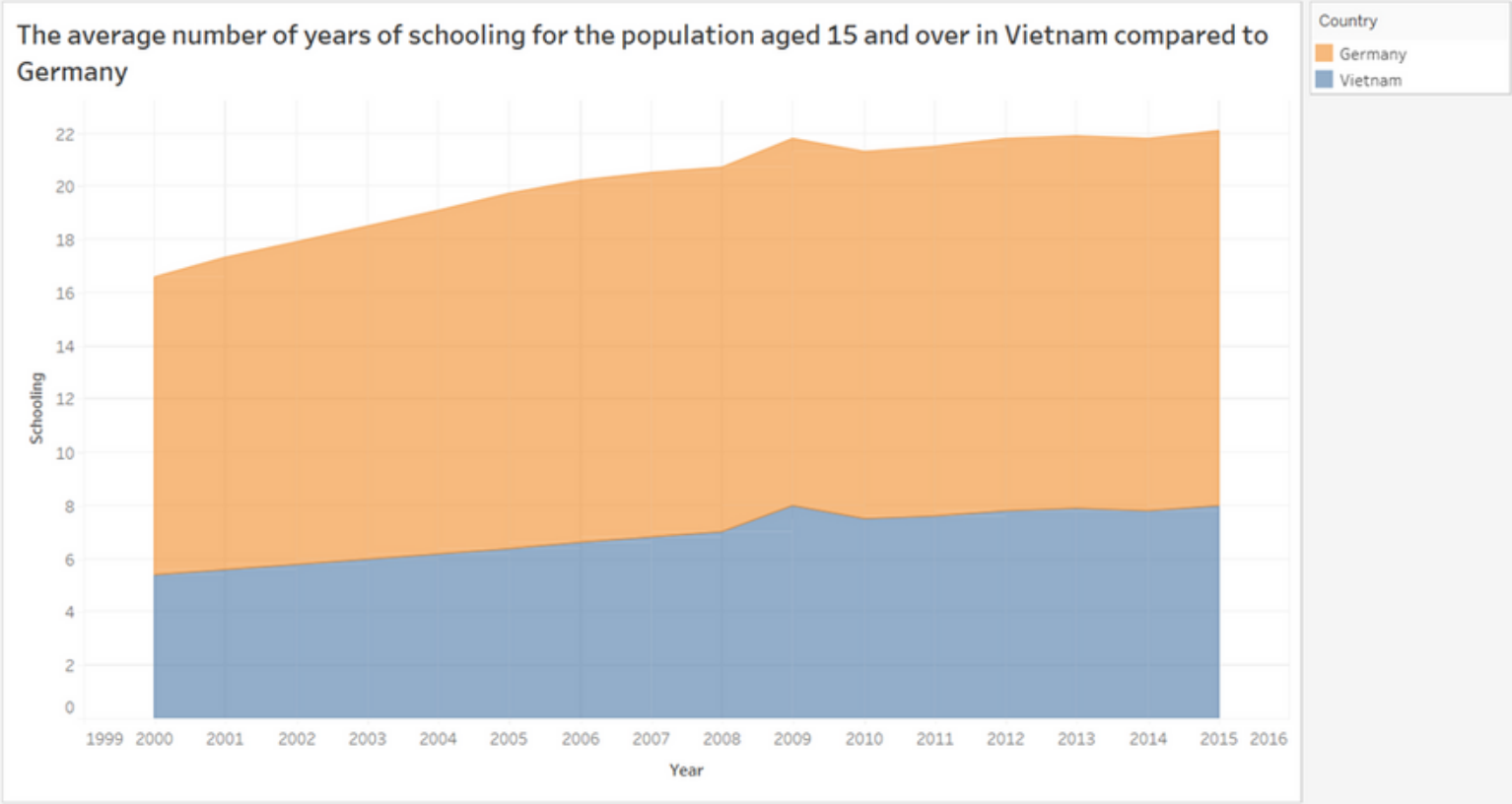
- Chart used: line chart
- The reason why we want to use line chart to display the schooling situation in Vietnam is:
  - The average number of years of schooling for the population aged 15 and over changed over time.
  - Show the trend.
- Comments:
  - As we can see, of Vietnam the average number of years of schooling for the population aged 15 and over is increasing over time which means Vietnam has made a great effort to gain more knowledge, becoming developed country through the intelligence of netizen.
  - The lowest value of schooling is 5,4 in 2000 and the highest is 8 in 2009. We can see that although we have significantly improved the value of schooling, it is still quite low based on the age of people.
  - This problem maybe come from the reason that: the economic of Vietnam is still developing so it takes time to improve the schooling, there are ethnic minorities that are unconditionally to gain knowledge,...

Top 10 countries with the most average number of years of schooling for the population aged 15 and over



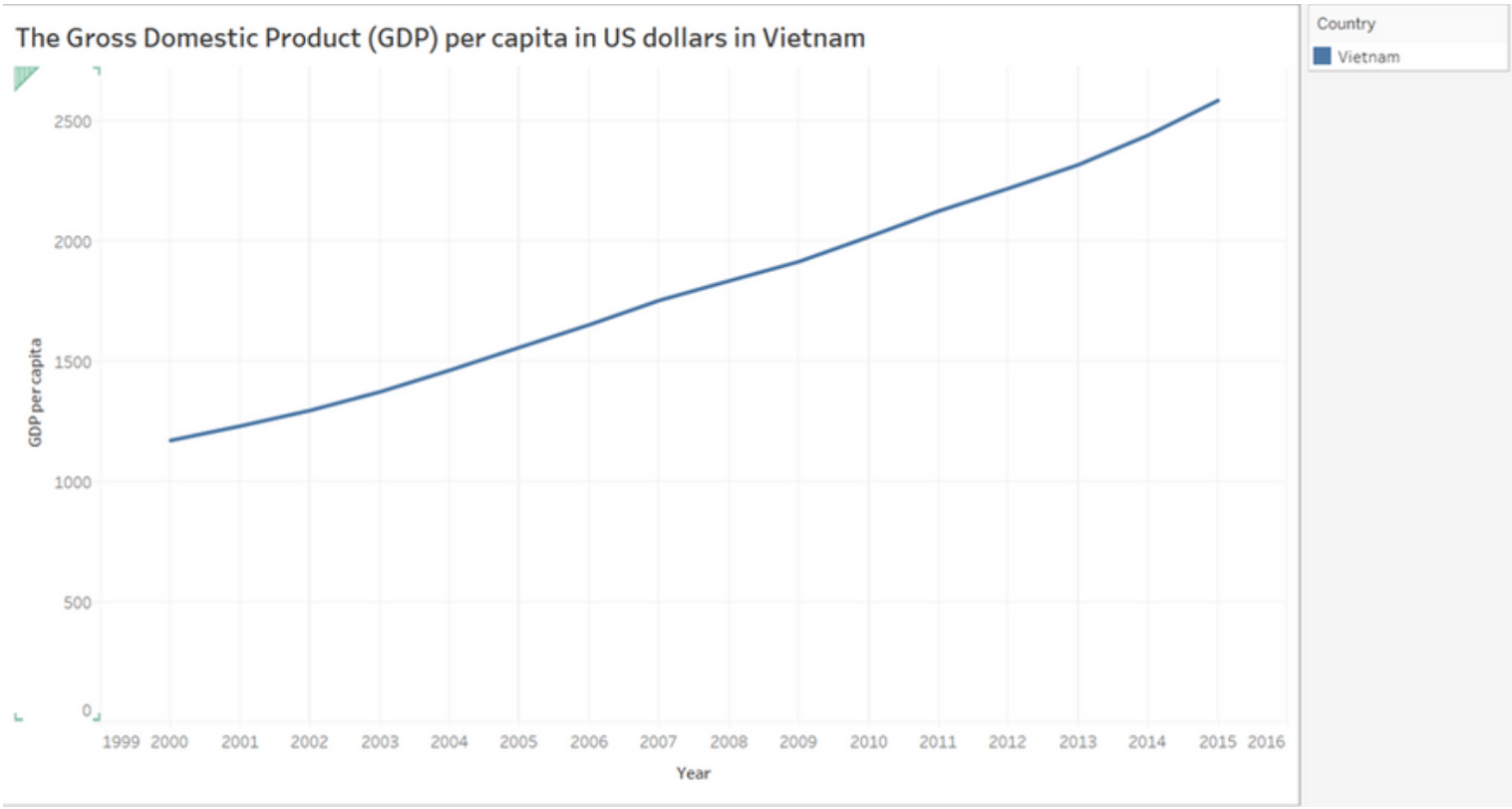
- Chart used: horizontal bars chart
- The reason why we want to use horizontal bars chart to display Top 10 countries with the most average number of years of schooling for the population aged 15 and over:
  - Show the difference between each country in the top 10.
  - Quantitative values overlap qualitative values.
- Comments:
  - Overall, we can see that there are no differences in education between developed countries because the schooling value between them is quite similar.
  - The intellectual level of people in developed countries is very high.
  - The most educated country is Germany.

The average number of years of schooling for the population aged 15 and over in Vietnam compared to Germany



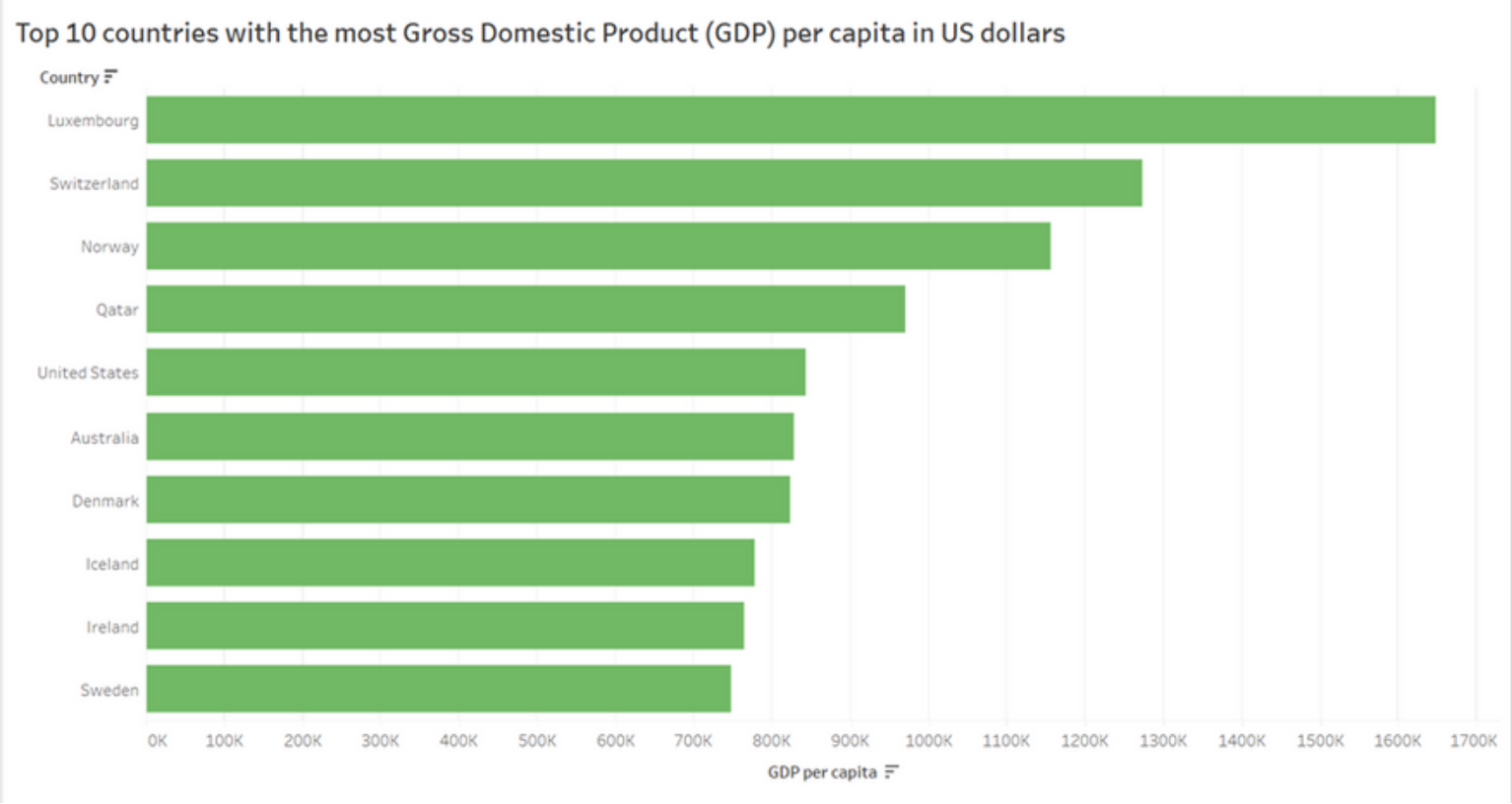
- Chart used: stacked area chart
- The reason why we want to use stacked area charts to display the difference between Vietnam and Germany is:
  - Show the relation between Vietnam and Germany.
  - Show the relation in a period.
  - Data type is numeric.
- Comments:
  - As we can see, Germany is a developed country, so their schooling value is very high and increasing over years.
  - The German is well educated and has wide knowledge since the average number of years of schooling for the population aged 15 and over of them is approximate to their actual age.
  - Vietnam has made a great effort to improve knowledge and education since number of years of schooling for the population aged 15 and over of them is increasing over years. Although the increase in schooling, this value is still very low compared to correspond value of Germany.
  - Vietnam needs more effort by learning from developed countries.

The economy of Vietnam



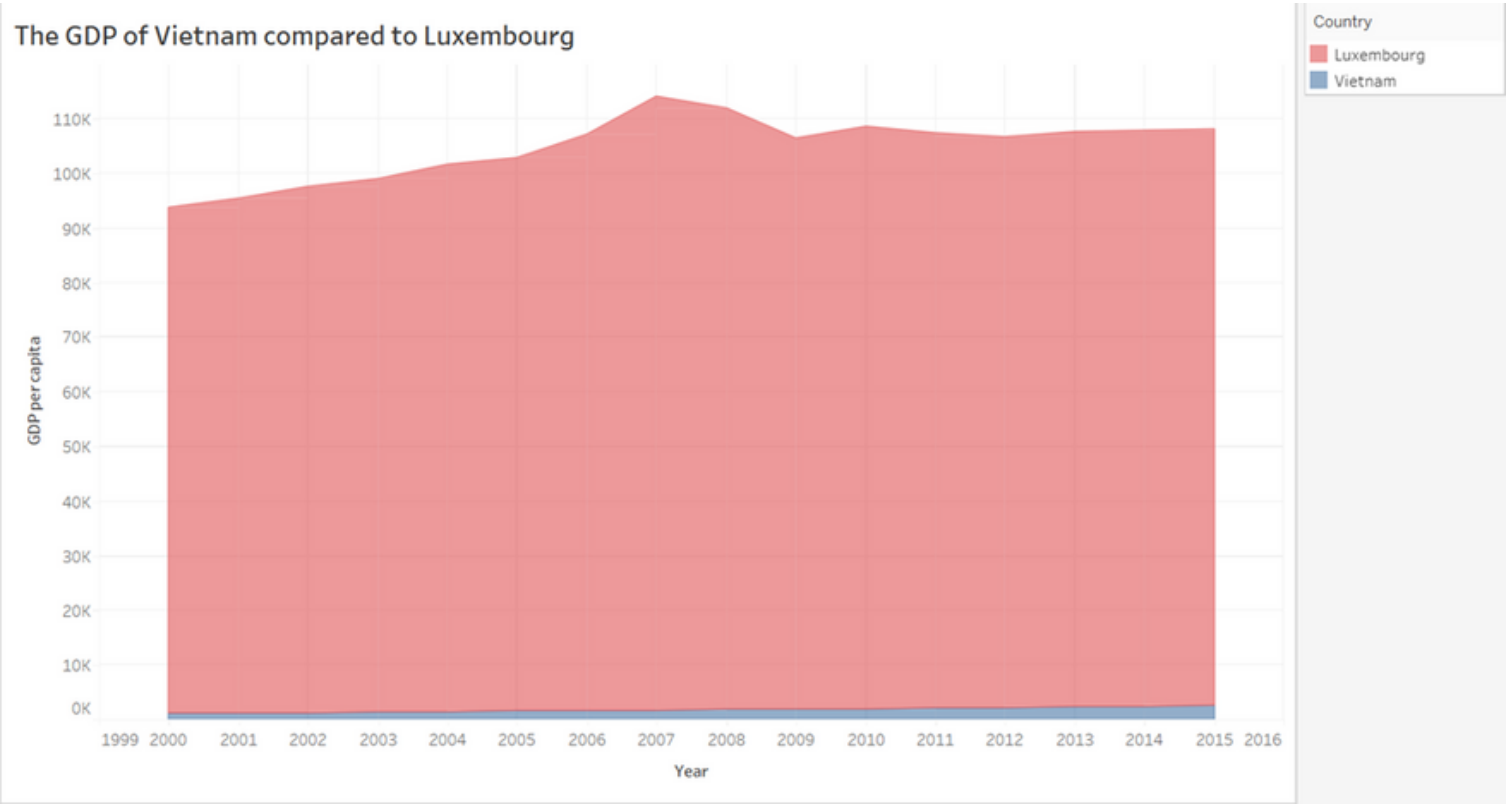
- Chart used: line chart
- The reason why we want to use line chart to display the schooling situation in Vietnam is:
  - The Gross Domestic Product (GDP) per capita in US dollars changed over time.
  - Show the trend.
- Comments:
  - Generally, the economy of Vietnam is increasing over years and no deterioration through the GDP which means Vietnam has made a great effort to improve the economy, becoming developing country.
  - Although the economic crisis in 2008, GDP of Vietnam is still increasing which means we controlled the situation quite well.
  - Those are good signals that predict GDP in Vietnam will grow even more in the future.

Top 10 countries with the most Gross Domestic Product (GDP) per capita in US dollars



- Chart used: horizontal bars chart
- The reason why we want to use horizontal bars chart to display Top 10 countries with the most Gross Domestic Product (GDP) per capita in US dollars is:
  - Show the difference between each country in the top 10.
  - Quantitative values overlap qualitative values.
- Comments:
  - As we can see, Luxembourg is the most GDP per capita country with 1,647,563 US dollars which is a huge amount of money. This value is quite larger than the GDP of remaining countries in the top 10.
  - Qatar is a remarkable country, the reason why they have high GDP is rich source of oil.
  - Sweden is the country with the lowest GDP in top 10 countries with the most GDP.

The GDP of Vietnam compared to Luxembourg

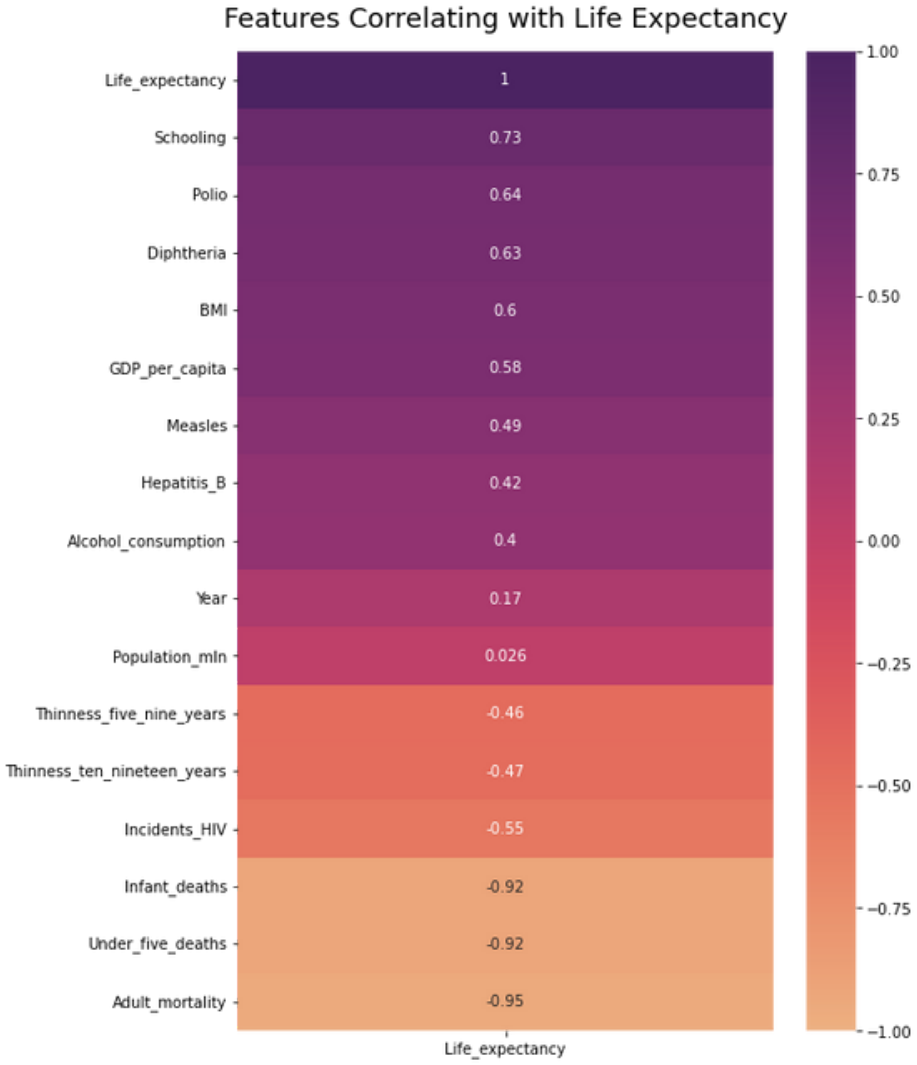


- Chart used: stacked area chart
- The reason why we want to use stacked area charts to display the GDP of Vietnam compared to Luxembourg:
  - Show the relation between Vietnam and Luxembourg.
  - Show the relation in a period.
  - Data type is numeric.
- Comments:
  - Although Vietnam has made a great effort to improve the economy since the increase in GDP over years, the GDP of Vietnam is still much smaller than the GDP of Luxembourg.
  - Although the economic crisis in 2008 impacted to the economy of Luxembourg and the GDP of Luxembourg is not consistent over years, their GDP is still leading.
  - Vietnam needs to make more effort.

Applying Machine Learning

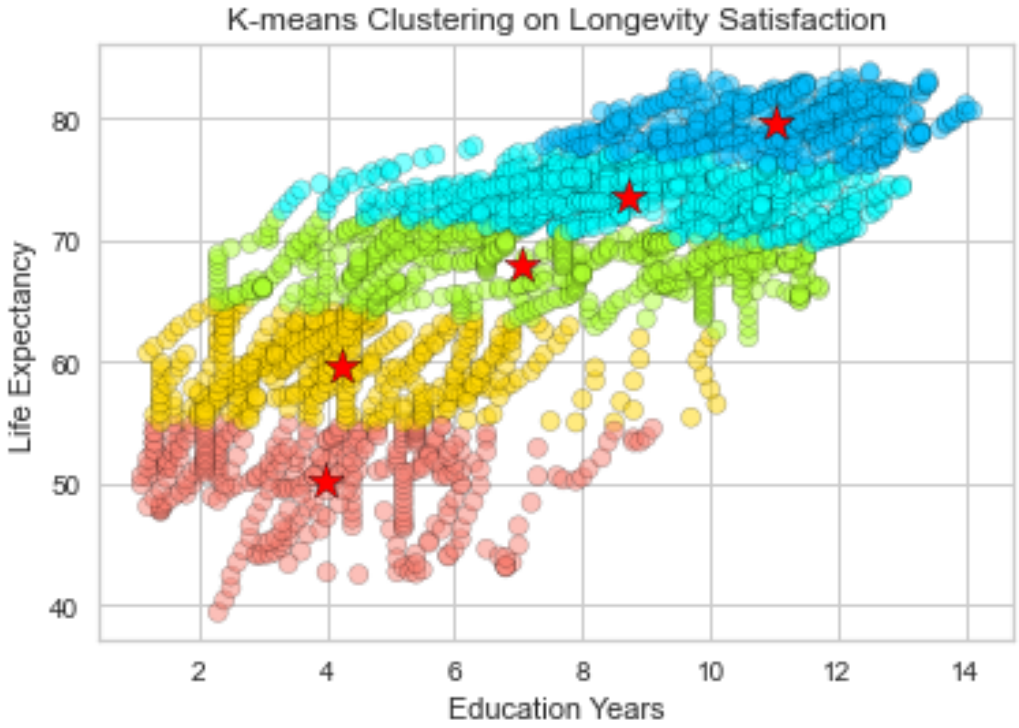
We would love to know how other aspects affect of Life Expectancy, so that we choose some of simple Machine Learning model in order to explore some interesting details

Plotting Correlation Table over Life Expectancy



- The main Purpose of our team to build Machine Learning Model is to explore which aspects could affect of Life Expectancy.
- Therefore, at first, we create a table correlation heatmap of every attributes over 'Life Expectancy'. As we could observe, almost all categories have correlation scores that impact on Longevity.
- Population shows that it is not so relevant to the target due to its score reaching 0.
- However, we have researched that in reality, population does extremely play a big role on a person life so it is not a big deal to keep the 'population_mln' in the input data of Model.

Levels of Longevity Satisfaction



- During the Research, we inquire that is better schooling experience could lead you to have a happy life. In addition, the scatter plot illustrates the positive correlation of schooling experience over longevity. Here is a reference, and definitely, with deeper knowledge base and experience, a person could avoid dangerosity, accident and prepare for their future better than the rest.
- However, in the scatter plot, there are many countries have higher longevity than expected by their low indicates of number of education years, so we would like to know life satisfaction level of a country.
- We decided to apply K-means that presents for 5 levels of satisfaction of Life Expectancy (very dissatisfied, dissatisfied, neutral, satisfied, very satisfied) and finally plot out this chart. The level of life satisfaction of a Country in our model is eventually based on numbers of schooling years and average age.
- We sequentially set color for each group with (red - very dissatisfied), (yellow - dissatisfied), (green - neutral), (light blue - satisfied), (dark blue - very satisfied). With group of very satisfied, we easliy could comprehend that their longevity is sufficiently highwhich means they have a healthy life and also long terms educated results in happiness of their life.
- About Neutral group, the range of education years is wide but they only live about 65 to 72 which is quite long enough for a human. With dissatisfied groups, they have not had well educated also have faced diversified difficulties in life, it is easy to infer to the lack of positive attitude toward life.
- This model have not been optimised yet, we built this to seek which country is worth to live and the other governments must improve their strategy and management to enhance people life quality.

Predict Longevity by simple Linear Regression Model

Our group has decided to split the dataset for training phase and test phase to validate the model, the two thirds for training model and the remaining data is for validating the model With categorical columns, we one-hot encoded them to let them participate in the Model. After training, we compare the prediction of test dataset to the actual values and earn up to 0.98 confidence score. The model is used for predicting the life expectancy of a country to manage the life quality of residents to identify whether that country is good enough to live or have too many challenges to encounter in upgrading life quality. Predicting longevity helps you to:

1. Compare among nations to know which aspects lengthen their age
2. Prepare foods and goods for your residents
3. Design an appropriate for social development
4. Also a vital attribute devoting in Economics, Sociology, Geography Prediction
5. A remind to live in a healthier life

Our Model has a formula:

$$y = \sum a * x$$

With y is longevity and a is coefficeints and x is the factor effects of age

```
Linear Regression coefficients:
--------------------------------------------------------------------------------
With Adult_mortality : -0.04778110968910544
With Alcohol_consumption : -0.0037891963131536322
With BMI : -0.14419341046732137
With Diphtheria : -0.010016970010398892
With GDP_per_capita : 1.8848930326352905e-05
With Hepatitis_B : -0.008303553370222741
With Incidents_HIV : 0.0998567169973774
With Infant_deaths : -0.053489481307448176
With Measles : 0.00233122181758527
With Polio : 0.01025360682945442
With Population_mln : -0.00025834944434633707
With Schooling : 0.11382154983825984
With Thinness_five_nine_years : 0.03146345493403617
With Thinness_ten_nineteen_years : -0.03844003901469193
With Under_five_deaths : -0.05051876554187024
With Economy status_Developed : 1.2493013848202978
With Economy status_Developing : -1.2493013848202987
With Region_Africa : -0.2941120869474811
With Region_Asia : -0.11243650014191173
With Region_Central America and Caribbean : 1.6157051477325655
With Region_European Union : -1.2161598660620743
With Region_Middle East : -0.18047666772737236
With Region_North America : 0.32142005814690455
With Region_Oceania : -1.2861710462131601
With Region_Rest of Europe : -0.16018491201329915
With Region_South America : 1.312415873225829
--------------------------------------------------------------------------------
```

Validating Model Chart

To guarantee the model, we draw many types of charts to validate the residuals of predicting values.

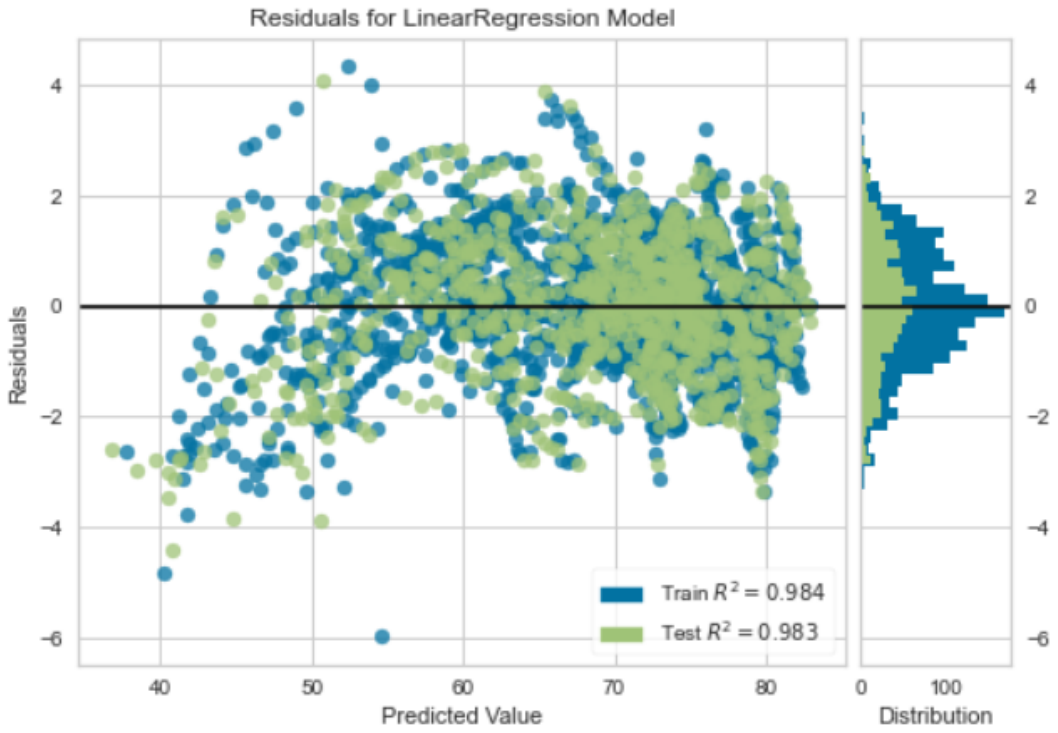


Figure Validate.1

- Fortunately, the model is quite fit and not many residuals go far too much. At the figure Validate.1, the residuals almost distributed at 0 and the other placed very closed to 0. The model predict right to almost objects

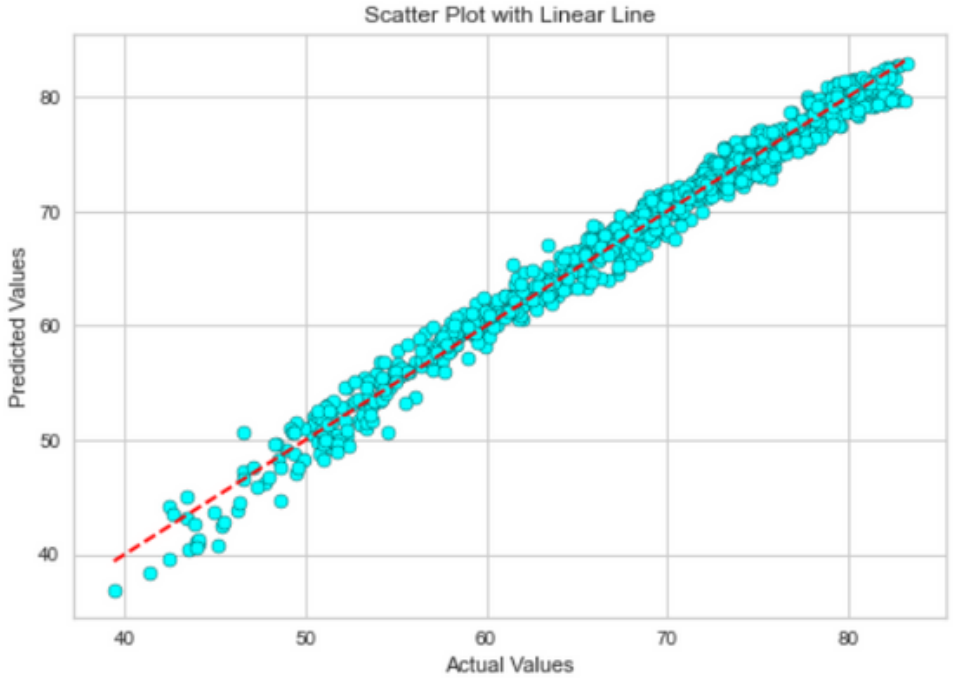


Figure Validate.2

In Figure Validate.2
We easily observe that all the points placed nearly on the identity line which means the predicted values similar to the actual values.

THANK-YOU
NOTE

Thank you Teacher for your Review

From Group 09 with Love ♡♡♡