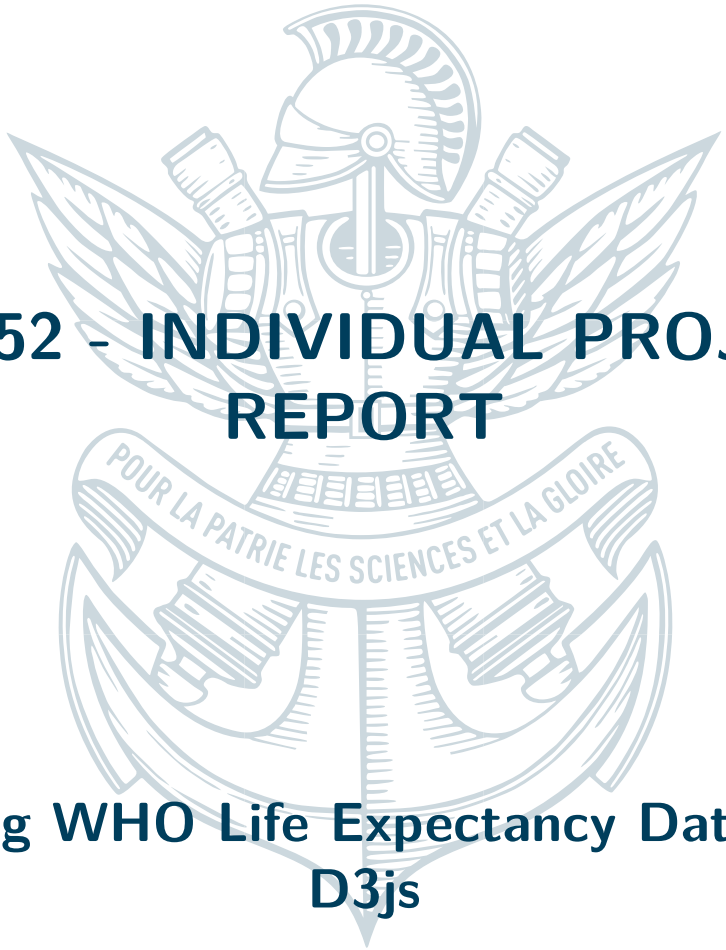

INF552 - INDIVIDUAL PROJECT REPORT

Visualizing WHO Life Expectancy Dataset with D3js

15 décembre 2022

Chinonso Stanislaus Ngwu & Changjie Wang

1
WHO LIFE EXPECTANCY DATASET VISUALIZATION

1.1 Dataset Description

Taking inspiration from this **Kaggle** competition which aims to analyze the general trend in The Global Health Observatory (GHO) data repository, maintained by the World Health Organization (WHO), that tracks the health status and other related factors for all countries. We want to visualize the correlation between the factors affecting life expectancy for example, to see the different correlations between the different attributes in the dataset such as the correlation between GDP and life expectancy or unemployment and life expectancy (e.g., do densely populated countries tend to have lower life expectancy ?). As described below, the dataset has 21 attributes which were carefully visualized using D3js and other supporting libraries as will be discussed in the following section.

Two datasets were used in the course of this project ; the first is a collection WHO life expectancy data from the year 2000-2019 for 193 countries which was cleaned to remove null or invalid inputs, the resulting dataset consists of 23 columns and 8049 rows. The attributes represented in this dataset includes the following ; Country, Code, Year, Gender, Life expectancy, Unemployment, Infant Mortality, GDP, GNI, Clean fuels and cooking technologies, Per Capita, Mortality caused by road traffic injury, Tuberculosis Incidence, DPT Immunization, HepB3 Immunization, Measles Immunization, Hospital beds, Basic sanitation services, Tuberculosis treatment, Urban population, Rural population, Non-communicable Mortality, Suicide Rate.

The second dataset is a Life expectancy data from the year 1950 - 2019 for seven geographic regions which is grouped like the planets and includes ; Africa, Americas, Asia, Europe, Latin America and the Caribbean, Northern America and Oceania. After cleaning the dataset, we have 4 columns and 491 rows. The interesting attributes are just the life expectancy and year. This helped us to visualize the relative change in life expectancy for the 7 regions over a longer period of time (70 years) using an animated horizontal Bar chart.

1.2 Visualizations

According to our data, it can be found that the value of Life expectancy is a sequence. In addition to that, we can find that there are 21 attributes that can be used to study the relationship between life expectancy and other factors. These 21 factors are differentiated and can simply be separated into two categories : the first is discrete, such as Country, Year and Gender, given that the choice of these attributes are limited, it implies that we can consider

directly visualizing the relationship between each optional value of these attributes and life expectancy. The second category is the sequence attribute, such as clean fuels and cooking technologies, GDP, Mortality caused by road traffic injury, etc. For this type of attribute, it means that we should visualize the relationship between two sequential variables. These two categories need to be discussed separately.

1.3 Visualization of the Discrete Data Category

We first Consider the visualization of the discrete data. The statistics of discrete attributes is different ; Country contains about 200 optional values ; and gender undoubtedly has only two optional values. There are 20 optional values for Year in our data : from 2000 to 2019. The number of different optional values makes the idea of visualization different.

We first visualized the relationship between Country and Life expectancy. As we learned in session 6, about how to draw visualizations of drinking water conditions in each country. Here we are deeply inspired and directly linked the Life expectancy to each country's section on the world map. Since Life Expectancy is a continuous value here, we thought of using a **Colorscaler()** to reflect Life Expectancy as the color of the layout.

The specific technical realization was done using D3, which we learned in the course. We downloaded the geojson data of 248 countries around the world from the website https ://geojson-maps.ash.ms/. The reason we did not use the geojson data provided in session 6 is because its resolution is not very satisfactory. We want to show the borders of each country more clearly. Then, we load the dataset and bind the data for each country (all Years and all genders) with the corresponding geojson. For the choice of color, we choose the color schema of **d3.interpolateGnBu()**. The world map image highlighting the different attributes of our dataset is shouwn in Figure 1

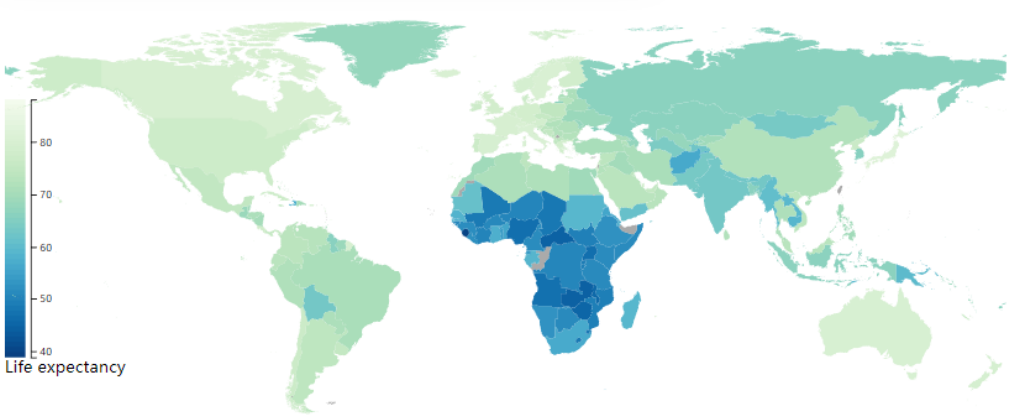

Figure 1 – World map showing the heatmap of Life Expectancy

Similarly, although our main visualization goal is to explore the relationship between Life expectancy and other factors, we found that on the world map, we can also compare different countries with attribute values of other parameters in Figure 1, such as the heat map of GDP and Suicide rate in different countries. These are also features worth exploring. We will discuss this specifically when visualizing continuous attribute values in later sections.

Secondly, we visualize the discrete data **Year** and **Gender**. For the year, what we want to see is the relationship between the value of each parameter as the year changes. With the increase of the years, (of course, the advancement of technology), we expect that there will be a more or less increasing trend in life expectancy, and we want to show this trend more clearly. Our data can be classified by country and the value differences between different countries may be more volatile. For example, in the past two decades, the Life expectancy range of Chinese men is (70, 74), while that of Namibia men is (50, 60). Putting the pictures of the two together in the past two decades will make people feel confused, since this is more biased towards highlighting the gap between the two countries' data than the relationship between Life Expectancy and age, which we have shown in the map to some degree.

Therefore, we choose to separate each country first, and first visualize the development trend of Life expectancy in a single country over time. Line charts are very effective for showing trends. Of course, we can't just show only the trend in one country. We need to show trends in all countries likewise, but it's not realistic to draw these lines on the same canvas, we have 248 countries, and therefore will need 248 lines plots. Therefore, we bind the canvas of the line chart to the world map and used an event binding to the country, whenever an image is clicked, the line chart of the corresponding country will be displayed. A life expectancy trend line plot of France is shown in Figure 2.

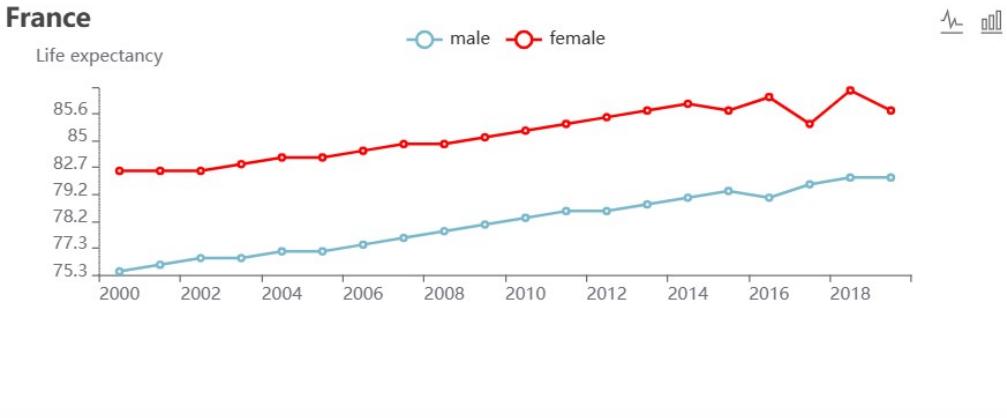

Figure 2 – A line chart showing the trend of Life expectancy and Year (2000 - 2019)

From figure 2, we have added a separate line plot for both genders, this is because, for men and women in a single country, the development trend of their life expectancy differs, with the female gender generally having greater life expectancy than the male. This difference between both genders is not only in value, but also in trend. The display of the gender attributes is configurable and can be switched ON/OFF from the line plot which in turn is controlled from

the gender drop-down selection box in the control panel.

For the last discrete attribute, we also show the relationship between Gender and Life expectancy. Notice that there are only two options for Gender here, male and female. To be specific, we need to show the distribution difference and summary statistics between the two genders. Therefore, we thought of the Box Plot used in session 4. We draw all data points, box plots of male data and female data on the same canvas, and we can see the difference between the two as shown in Figure 3

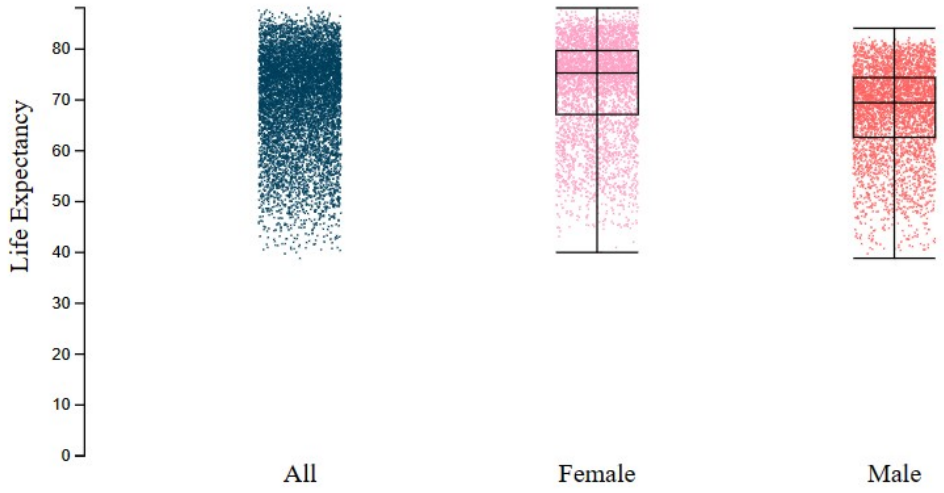

Figure 3 – Distribution of Life expectancy, broken down by gender

1.4 Visualization of Continuous Values

To handle the visualization of continuous values, we use the attribute ;Basic sanitation services as an example whose values are in the range (0, 100). We want to establish a connection between two sequential values. Going back to the line chart we created earlier (Figure 2, which shows the trend of life expectancy with year. We take Inspiration from it and add the development trend of Basic sanitation services to the graph which makes comparison between the two data trends possible. Therefore, we consider adding another y-coordinate axis on the right to reflect this as shown in figure 4.

It is worth noting that for the line charts, we did not make use of **D3**.This is Because it requires a lot of effort to draw multiple line charts and multiple coordinate axes on the same canvas with the **D3**. Instead, we apply an integrated API **Echarts** to make the visualization. We found the figures drawn by **Echarts** to be smoother and more beautiful than that of **D3**. **Echarts** also makes it easy to switch between line plot and bar charts we included as an additional feature in our visualization.

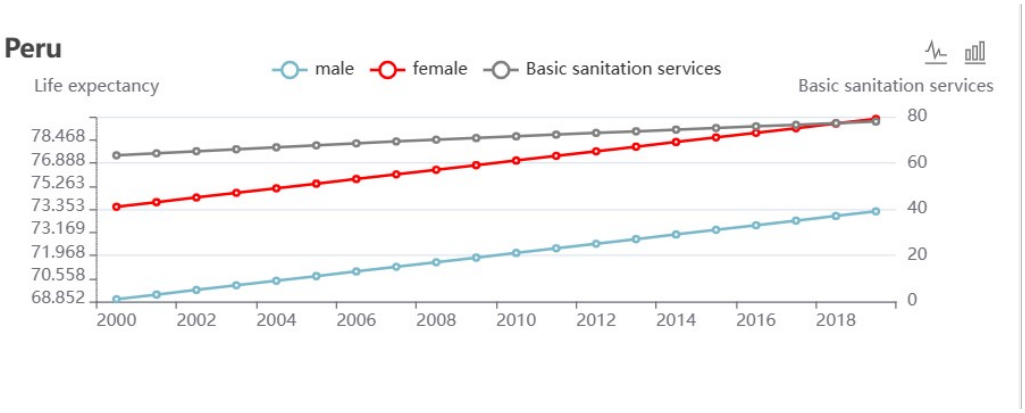

Figure 4 – A line chart showing the trend of Life expectancy and Year (2000 - 2019) and also the Basic sanitation services trend

Thinking back to the world map and line charts we have drawn, we observed that we have been comparing data within a country and, also comparing the development trends of various factors in a single country. Our data is much more than that, we have data for every country in the entire world. We therefore should not only show a single country, but also visualize the development trend of the whole world, since the data of a country may be more or less accidental. Therefore, it is essential to show all data for all countries.

Back to the question of how to present data for all countries. The data to be displayed are : year, gender and life expectancy and other attributes for example, GDP. There are four channels of data that need to be displayed. Among them, life expectancy and Basic sanitation services are continuous parameters, so we thought of scatter plot and used these two as the main coordinate axes. For gender, we can show it by two different colors. As for the year, we thought of another implementation method, using animation to show the trend. We can first show the life expectancy with Basic sanitation services for all countries in year 2000, grouped by gender and then animate it over the years to show the trend. With this animation, we can see that the collection of points for the entire world has a somewhat slow-moving trend over the years. This is exactly what we want. Of course, if we want to see the details of a specific year or a certain gender, we can also stop the animation observation. The animated scatter plot is shown in figure 5.

Finally we wanted to compare the life expectancy of different regions of the world instead of just countries. We decided to go with a horizontal bar chart since we have discrete number of regions/continents we wanted to compare. We then encoded the life-expectancy and the lengths of each corresponding bar. Since we have data up to years each for the continents, we used animation to show the general trend of life expectancy across the different select regions. Figure 6 below shows the animated horizontal barchart Showing Life Expectancy of Different Continents between 1950 & 2019

It is worthy to note that all the charts aside the continent chart is controlled by the three select drop-downs which we refer to as the panel. This helps to select the Year, gender or attribute you want to visualize.

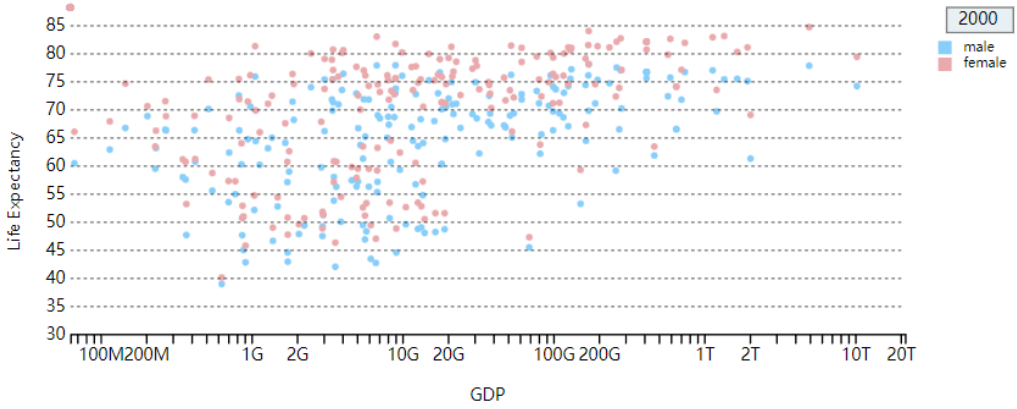

Figure 5 – An animated scatter plot of Life Expectancy versus GDP

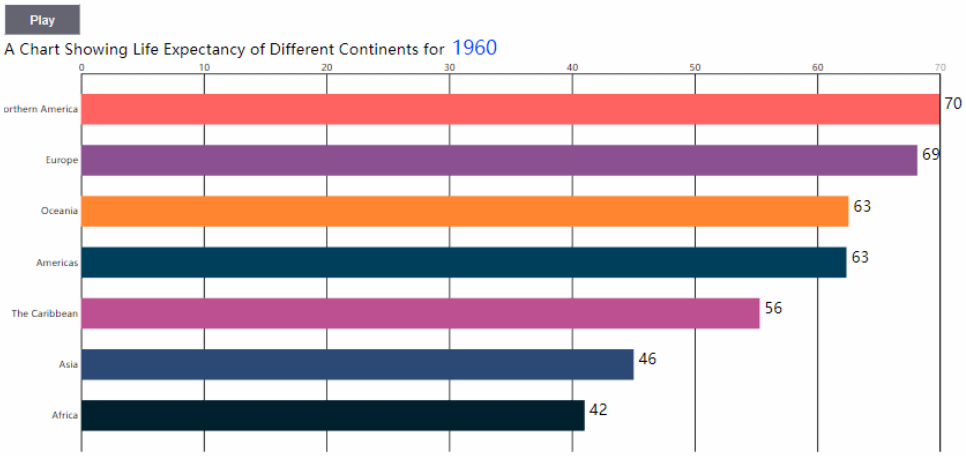

Figure 6 – A chart Showing Life Expectancy of Different Continents between 1950 & 2019

2 ILLUSTRATIONS

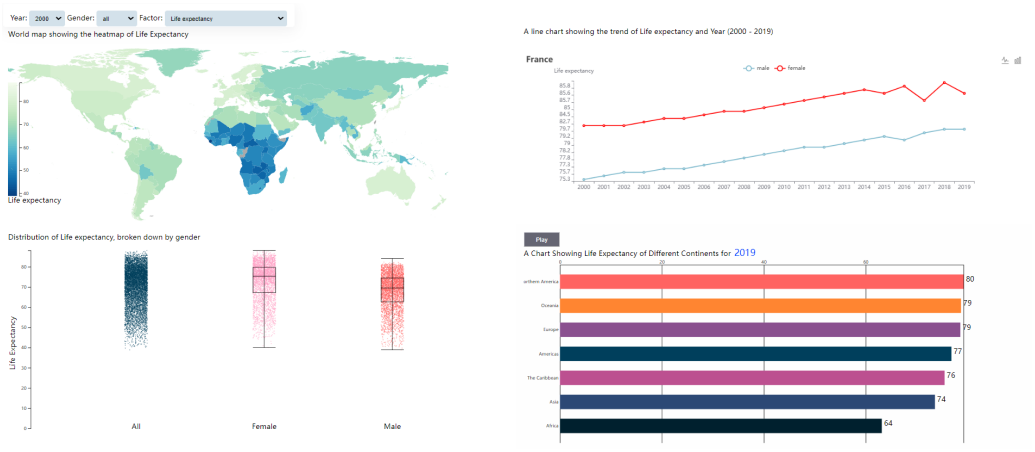

Figure 7 – The complete visualization showing the default plots.

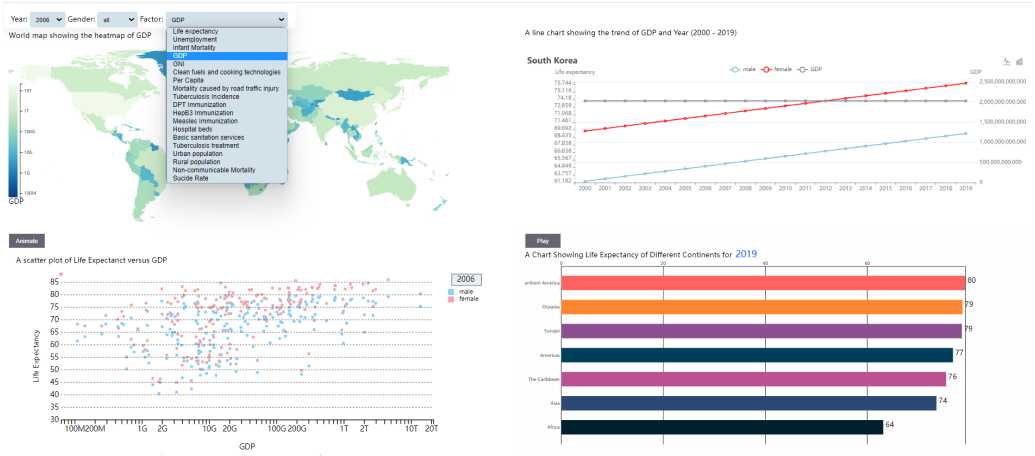

Figure 8 – The complete visualization showing the control panel for selection drop-down

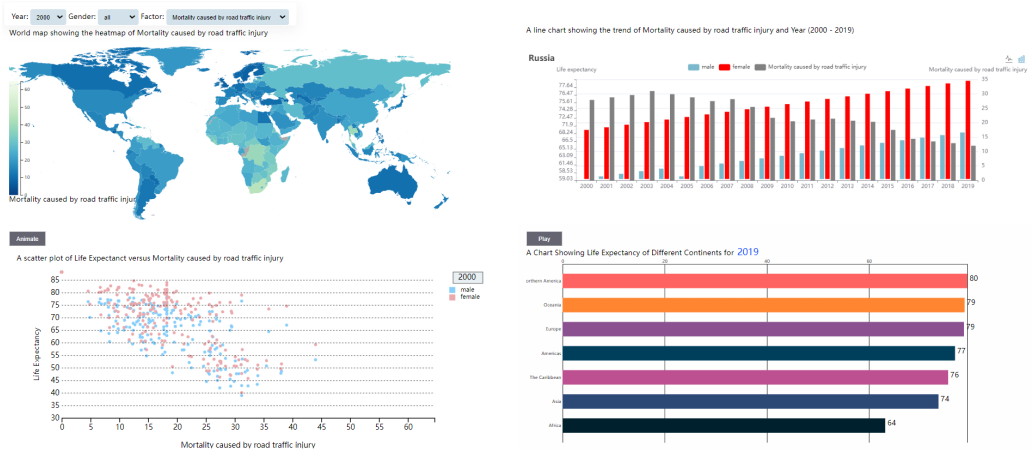

Figure 9 – The complete visualization showing the bar charts instead of line plots.

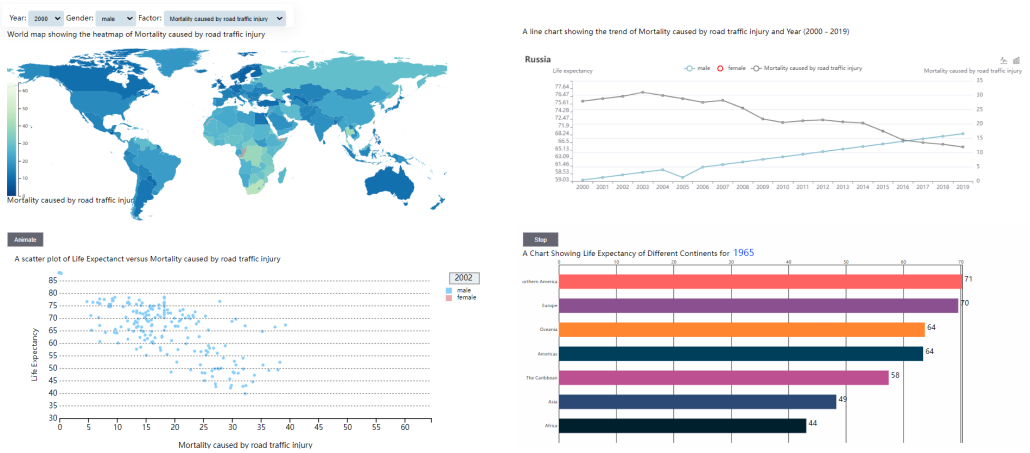

Figure 10 – The complete visualization showing only male gender data.