INF552 - INDIVIDUAL PROJECT REPORT

Visualizing WHO Life Expectancy Dataset with D3js

16 décembre 2022

Chinonso Stanislaus Ngwu & Changjie Wang

ÉCOLE POLYTECHNIQUE

IP PARIS

1
WHO LIFE EXPECTANCY DATASET VISUALIZATION

1.1 Dataset Description

Taking inspiration from this **Kaggle** competition which examines the overall trend in The Global Health Observatory (GHO) data repository, maintained by the World Health Organization (WHO), which tracks the health status and other related factors for all countries. We wish to show the relationships between the variables that affect life expectancies, such as the relationships between GDP and life expectancy or unemployment and life expectancy (e.g., do nations with higher lower GDP often have shorter life expectancies ?). The dataset, which includes 21 properties, was meticulously displayed using D3js and other auxiliary packages, as discussed below and will be covered in detail in the next section.

This project made use of two datasets : the first is a collection of WHO life expectancy data from 2000 to 2019 for 193 countries that were cleaned to remove null or invalid inputs ; the resulting dataset has 23 columns and 8049 rows. This dataset comprises the following attributes : 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 and, Suicide Rate.

The second dataset contains life expectancy statistics from 1950 to 2019 for seven geographic regions, which are organised like planets and include Africa, the Americas, Asia, Europe, Latin America and the Caribbean, Northern America, and Oceania. After cleaning the dataset, we have 4 columns and 491 rows. Life expectancy and year are the only interesting characteristics. Using an animated horizontal Bar chart, we were able to illustrate the relative change in life expectancy for the seven regions over a longer period (70 years).

1.2 Visualizations

According to our data, the value of Life expectancy is a sequence. Furthermore, there are 21 attributes that can be utilised to explore the relationship between life expectancy and other factors. These 21 variables are differentiated and can simply be divided into two categories : the first is discrete, such as Country, Year, and Gender, which implies that we can consider directly visualising the relationship between each optional value of these attributes and life expectancy because the choice of these attributes is limited. The second category is the sequence attribute, which includes things like clean fuels and cooking technology, GDP, road traffic injury

mortality, and so forth. This type of attribute requires us to visualise the relationship between two sequential variables. These two groups must be considered separately and will be discussed in the following sections.

1.3 Visualization of the Discrete Data Category

We begin by considering the display of discrete data. The statistics of discrete attributes differ ; the country provides nearly 200 possible options, whereas gender has only two. In our data, Year has 20 possible values ranging from 2000 to 2019. The amount of different alternative values distinguish the visualisation concept.

We began by illustrating the relationship between country and life expectancy. We learnt how to build visualisations of drinking water problems in each country in session 6. We take great inspiration from the concept and directly linked the Life expectancy to each country's section on the world map. Since Life Expectancy is a continuous value, 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 shown in Figure 1

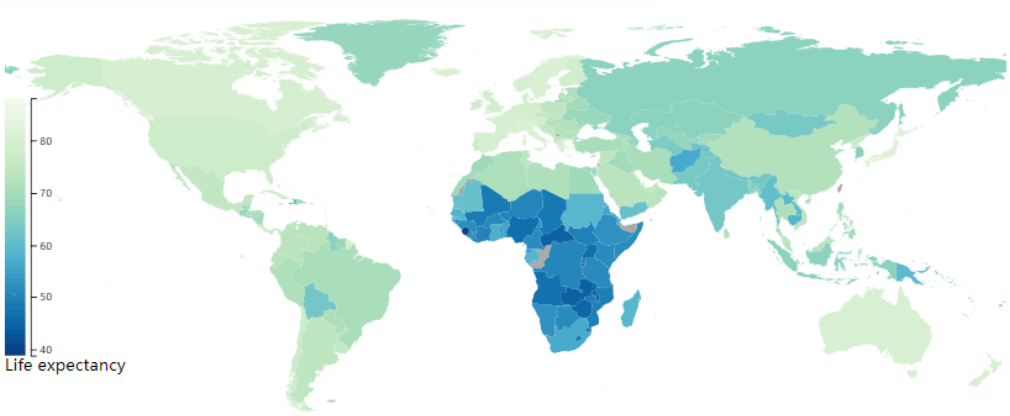

Figure 1 – World map showing the heatmap of Life Expectancy

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 from geojson, 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.

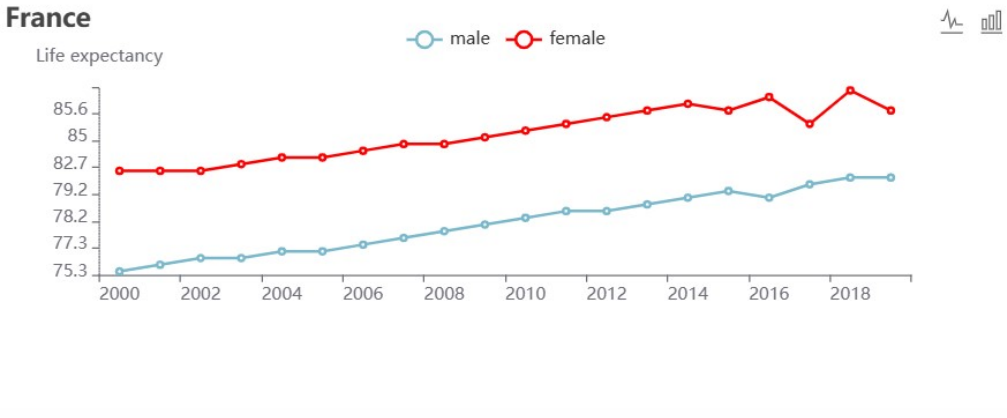

Figure 2 – A line chart showing the trend of Life expectancy and Year (2000 - 2019)

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

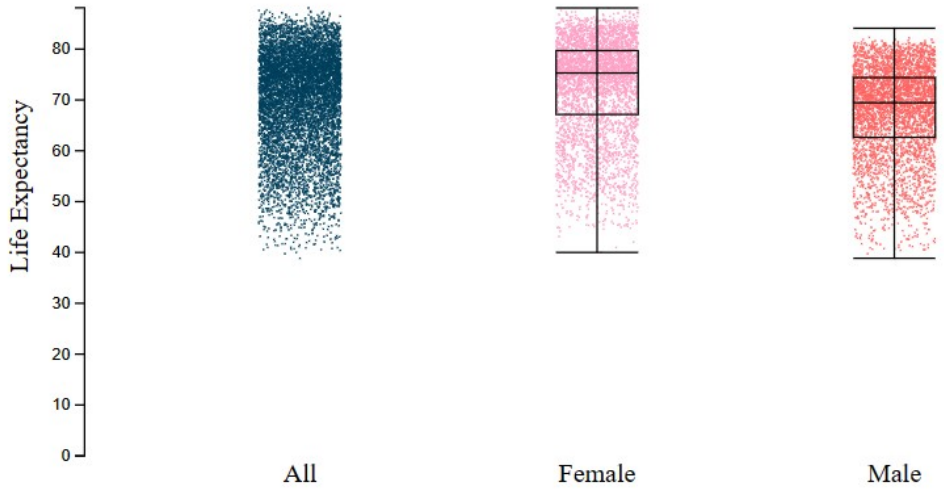

Figure 3 – Distribution of Life expectancy, broken down by gender

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 which we included as an additional feature in our visualization.

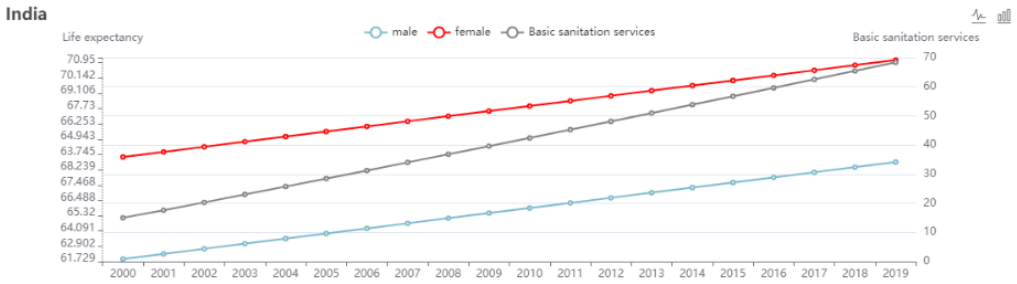

Figure 4 – A line chart showing the trend of Life expectancy and Year (2000 - 2019) and also the Basic sanitation services trend

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. It is therefore 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 at the same time. 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 and observe the particular year. 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 that we wanted to compare. We then encoded the life-expectancy as 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 control 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

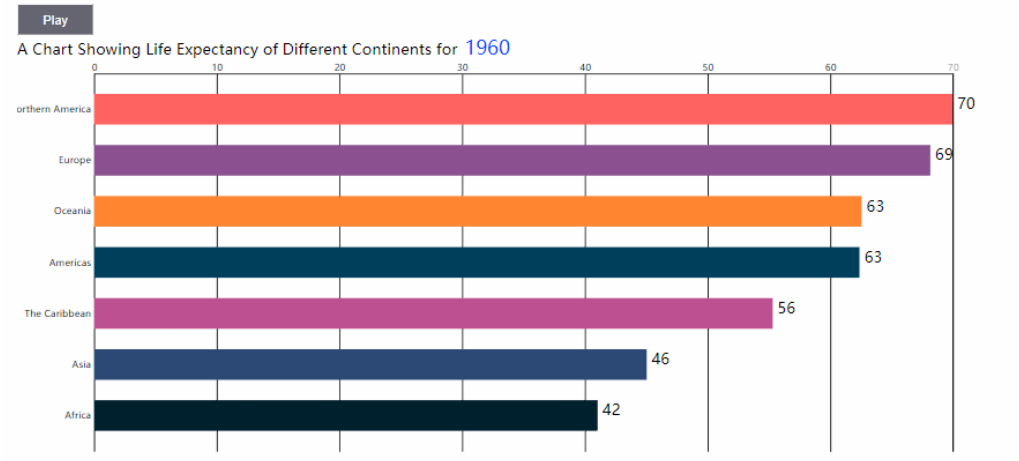

Figure 6 – A chart Showing Life Expectancy of Different Continents between 1950 & 2019

2 ILLUSTRATIONS

The complete video illustration can be accessed in our Github repository by clicking here.

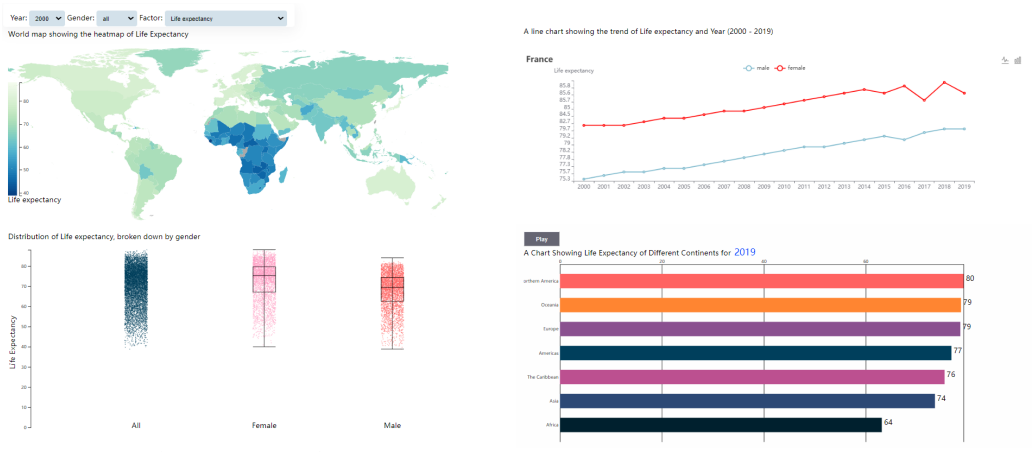

Figure 7 – The complete visualization showing the default plots.

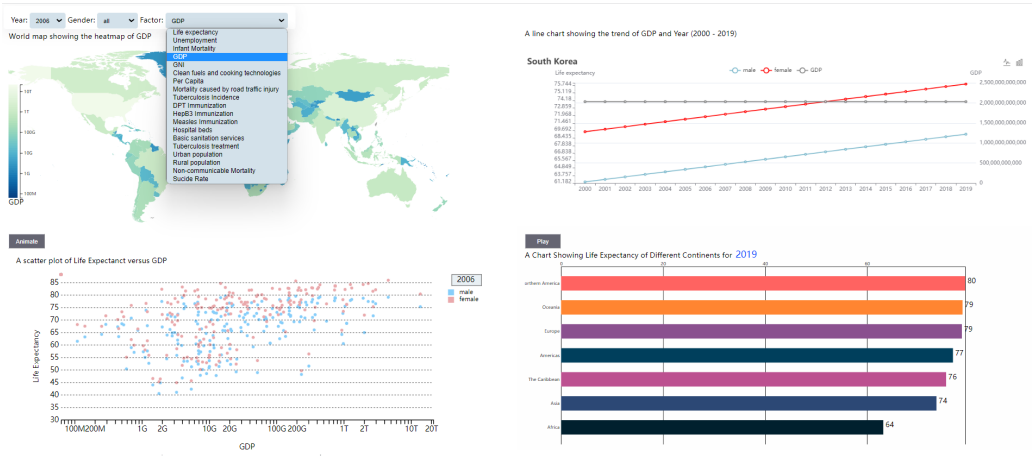

Figure 8 – The complete visualization showing the control panel for selection drop-down

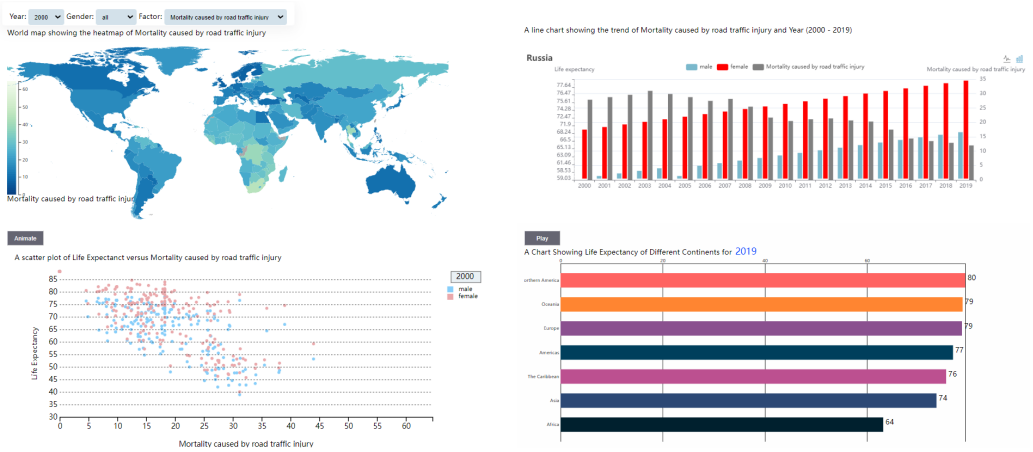

Figure 9 – The complete visualization showing the bar charts instead of line plots.

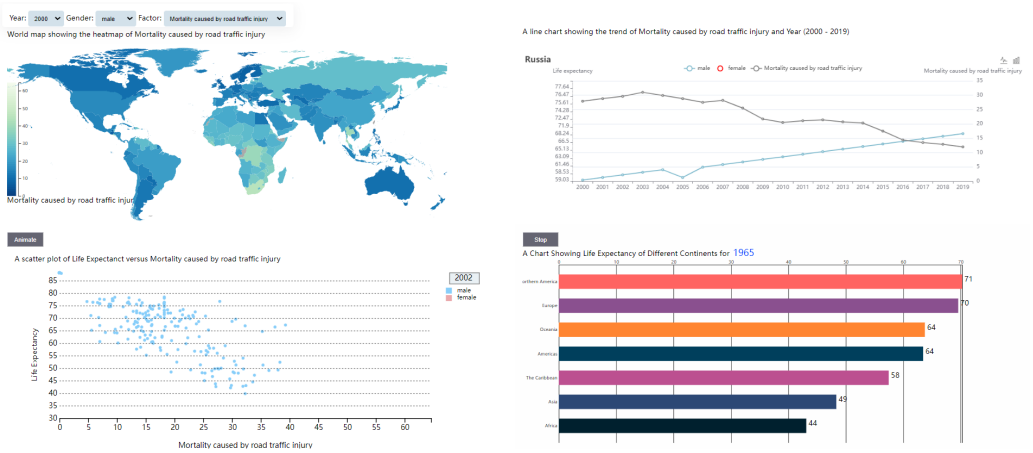

Figure 10 – The complete visualization showing only male gender data.