

**Health and Economics Exploration**

*Visualization & Presentation*

Manuel Andrés Hernández Alonso

Alexandru Ioan Oarga Hategan

Madison Elizabeth Chester

[**1. Exploratory Data Analysis 2**](#_5e00obi1na4g)

[1.1. Gapminder dataset 2](#_1midc2xp30qw)

[1.2. World Health Organization dataset 3](#_jtq7zbrox7ko)

[1.3. Exploratory Charts 3](#_9ld7vnlswh4w)

[**2. Audience 9**](#_q4uw4usidxso)

[2.1. Persona 9](#_5aif4a85gyce)

[2.2. Napkin design 10](#_agqizgi1rpxh)

[**3. Selection of charts and encoding 12**](#_sn9p7ek7muo7)

[3.1. General Dashboard: Worldwide Health Inequities 12](#_af7ldkhcnpvm)

[3.2. Specific Dashboard: Government Health Expenditure 15](#_u8sncpx3brl5)

[**4. Implementation of the Dashboards 19**](#_ngwvxwn0hi6x)

# Exploratory Data Analysis

This section provides a general description of the data used for our project,

and the general process followed for data selection. Firstly, we looked for data related to our main subject for the project: Health and Economics. The first data we used came from Gapminder which included lots of indexes and statistics we could work with, but after inspecting it further and working with it, we noticed that it had a lot of missing data. Thus, we had to look for some additional datasets to supplement the Gapminder one.

## 1.1. Gapminder dataset

We looked into the different indexes of the dataset of Gapminder and chose a specific list of indexes that we were interested in:

* **GDP:** This data is given in a continuous value of $USD per country. We find that this data is in the trillions for some countries.
* **GDP per Capita:** This data is given in a continuous value of $USD per country. This value is nearly 100.000 in some countries.
* **Life Expectancy:** This data is related to the life expectancy per country, it is a continuous value in years. The value ranges from 45 to 82.1 years.
* **Adult Mortality/Child Mortality:** These two indexes are closely related as both are ratios per 1000 population per country, these are integer-valued. These values range from 23 to 547.
* **Government Health Spending per Person:** This data is given as a continuous value of $USD per person, per country. This variable ranges from 0.87$ to 5690$.
* **Countries:** We also had to take the information on the names of the countries as well as geographical data related to them.

As we can see, all of the data is country-wise, so we’ll use this to our advantage to plot geographical-natured charts. Since there was a lot of data missing from this dataset we also decided to take the time range from 2000 to 2010, included. So this data is not really that valid for the current situation, but it is useful as an overview and guideline.

We found that most of these indexes follow a distribution where the less developed countries have less life expectancy, GDP and Health spending. We also saw that adult mortality and child mortality rose for those less developed countries.

Some of the indexes have outliers such as the USA having a lot of health spending but not that large of a life expectancy, as well as a high child and adult mortality rate for the GDP they possess.

## 1.2. World Health Organization dataset

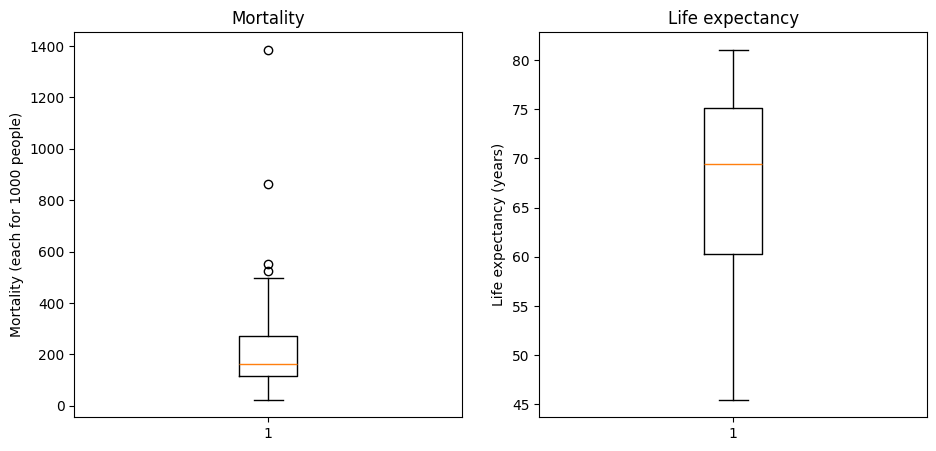
As we commented on the previous dataset, it was missing some data, so we resorted to another dataset to supplement it. We chose to look into some public datasets from the World Health Organization, here we found different columns of variables that are interesting. It is also separated by countries and years. Finally, we only took these columns from this additional dataset:

* **Life Expectancy:** Same as the other dataset, it is given in years and per country.
* **Income Composition of Resources:** Measures how good a country is at managing its resources, ranging from 0 to 1.
* **Infant/Adult Mortality Rate:** Similar to the previous dataset, it is a per 1000 population rate.
* **Status:** This gives the status of each country as a categorical value of “Developed” or “Developing”.

We found the same distribution patterns as with the other dataset, so they are consistent with each other. We also had to choose a time range, so we used all of the dataset ranges from 2000 to 2014.

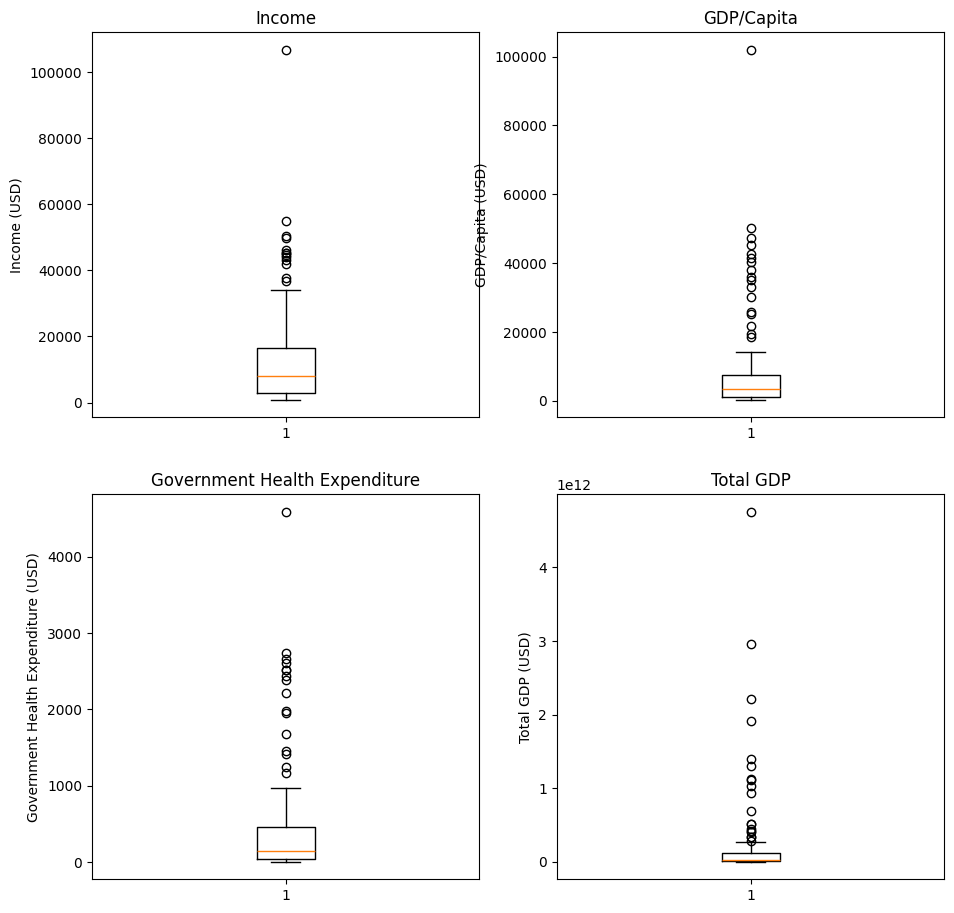
## 1.3. Exploratory Charts

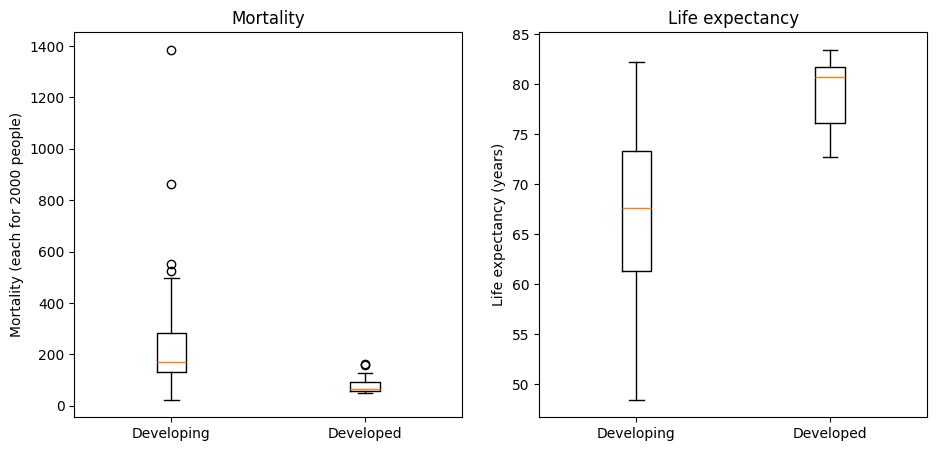
In this section, first we study the distribution of all the variables individually, and then we study the relationships that appear between economic and health variables.

First, we will study the distribution of each variable individually. To do so, here we plot the box plot of each of the variables we are interested in. We start by plotting the distribution of the health-related variables, this is, life expectancy and mortality. For simplicity, child and adult mortality have been aggregated into one single variable (thus, making it mortality per 2000 people rate). The box plots can be found in the figures below.

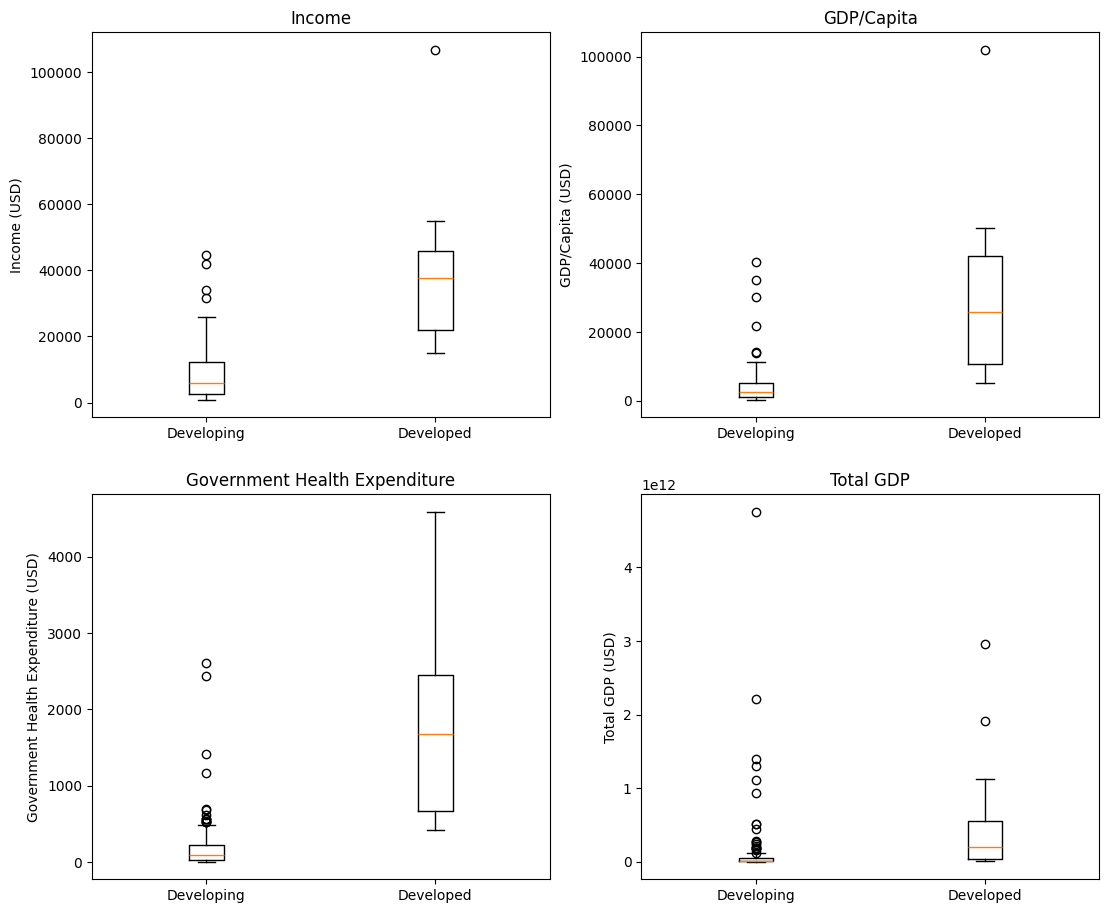
From the above box plots, we can see that the mortality distribution has an average of nearly 150 per 2000 people and that the life expectancy has an average of 70. We see that the mortality distribution has a few outliers that differ significantly from the rest of the distribution, reaching until a 1400 per 2000 people mortality rate. Finally, it can be seen that the mortality distribution is skewed towards a lower mortality while the life expectancy distribution is skewed towards a higher life expectancy.

We will now explore the distribution of the economic variables, this is income, GDP per capita, government health expenditure and total GDP. In the four figures below we show a box plot of each of the distributions respectively. Although at different scales, we can see that each of them shows a similar pattern, with each distribution being skewed toward the lower range of values.

We want to study now whether, for each of the previous variables, there is a significant difference between developed and developing countries. To do so, we will repeat the box plots shown before, but this time, we will differentiate between the two categories. In the figures below, we show the resulting box plots for the mortality and life expectancy data.

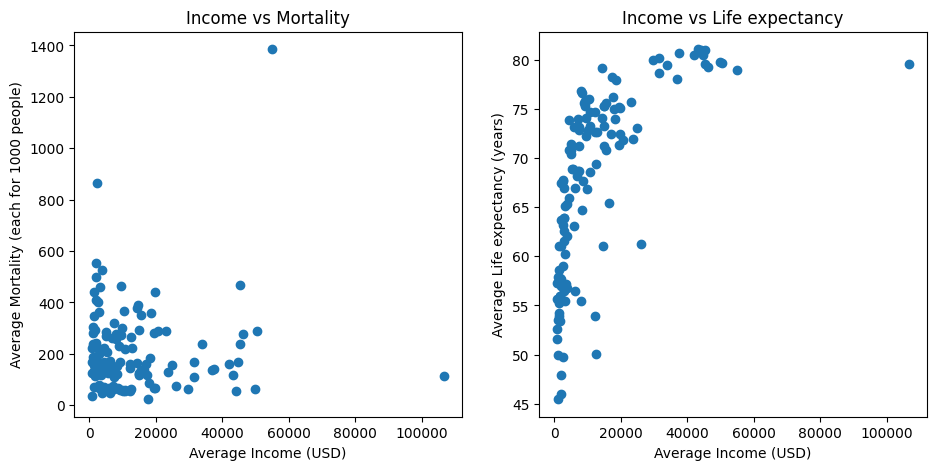


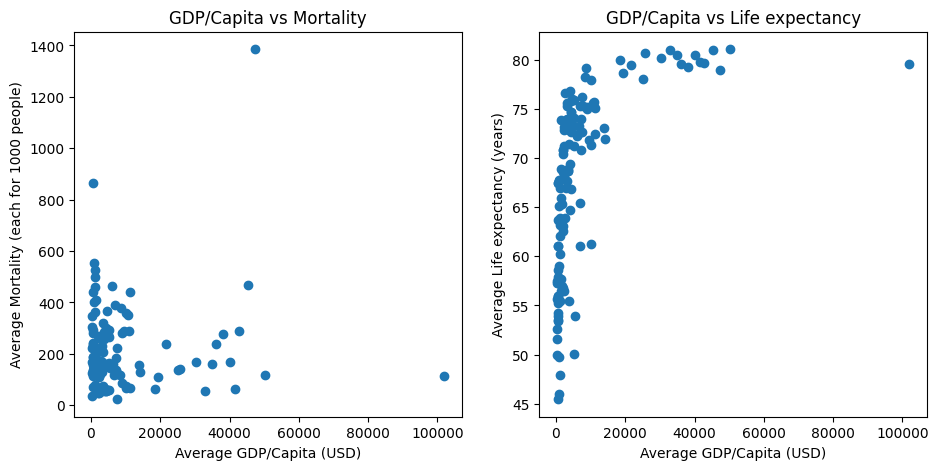
From the above box plots, we can see that there is a clear difference between developed and developing countries, both in terms of life expectancy and mortality. We can see that both the mortality distribution is more spread towards higher mortality rates in developing countries and that the life expectancy is more spread towards lower values in these countries. In our dashboard, we will be hence interested in visualising these clear divergences between the two types of countries shown.

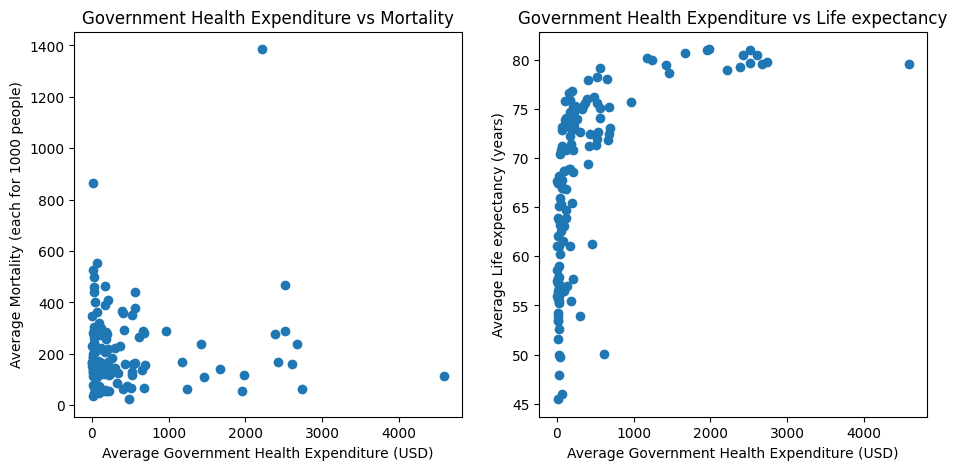


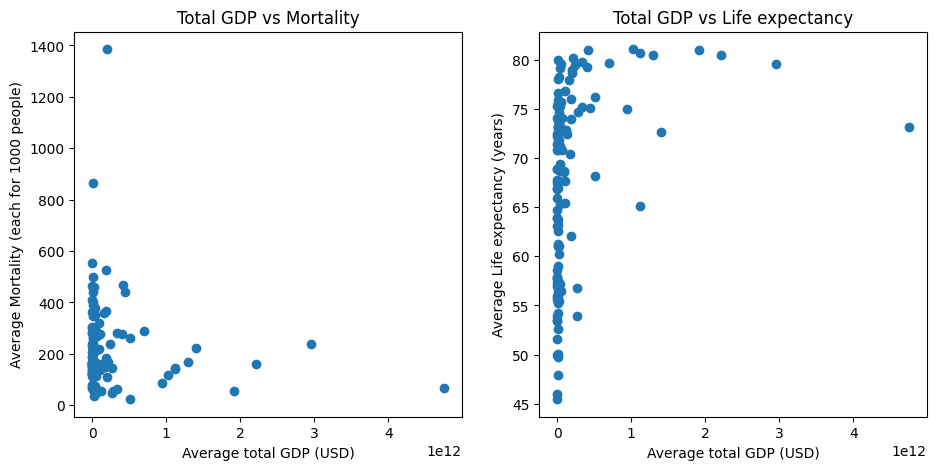
Finally, we will plot the box plot for each of the economic variable distributions. These box plots are shown in the four figures above. As it happened with the health-related variables, the economic indicators also show a clear difference between developing and developed countries, with the developed countries having a distribution shifted towards larger values of GDP, income or government health expenditure. Consequently, we will be also particularly interested in reflecting these inequalities in our visualization.

Once we have reviewed our variables distributions, we proceed to explore the relationship between the different variables available in our databases. We will be interested in studying the relationship between economic indicators (GDP, income, government health expenditure) and health indicators (mortality and life expectancy). We are particularly interested in finding relationships between both types of variables. To make the comparison of variables easier, we will be computing for each dimension the average over all the selected years.





First, we plot the relationship between average income and average GDP per capita with mortality and life expectancy (shown in the four figures above). From these graphs, we can see that both variables are surprisingly similar (if not identical in some cases). We can see therefore that one variable may not provide any more significant information than the other and thus, we can focus on only one of the two for our visualization purposes.



The previous two figures show a scatter plot where the government health expenditure and total GDP of a country are respectively shown in the horizontal axis and the mortality and life expectancy in the vertical axis. If we look at these diagrams, and the previous one (GDP per capita and income), we can notice that, even though some of them show significantly different scales, they generally show the same pattern for mortality and life expectancy. Regarding mortality, it can be seen that countries are usually grouped together and that there are always a few outliers that can be clearly identified from the diagrams. Generally, no clear relationship can be drawn from these. If we look now at life expectancy, a more clear relationship appears in all of the figures. Even though there is no linear correlation, we can see that there is a curve in all the diagrams and that countries with higher GDP or health expenditure tend to have a higher life expectancy and vice-versa.

From the previous diagrams, we can conclude that all economic indicators show similar relationships with our selected health variables. We will thus, be focusing on particular variables that might be more specific than other ones, such as government health expenditure.

# Audience

Our target audience for this project is managers of charitable NGO’s so they can distribute the resources they have and better know the current situation of health and economics around the world. They could also use it to check if they’ve had any impact in subsequent years.

The audience is not necessarily experts in charts and statistics, so we have to simplify the charts. They should consult the visualization around once every quarter so they can define and plan out strategies of aid and support.

## 2.1. Persona

We defined a persona to better help us create a dashboard for the specific use of planning and creating strategies of aid and supply:

***Sarah Rodriguez***, the Manager of Operations at the Global Aid Initiative (GAI), brings over a decade of experience in the nonprofit sector with a background in international relations. Passionate about positively impacting vulnerable communities, she has dedicated the past three years to overseeing GAI's operations. Sarah's primary objective is to efficiently plan and distribute available resources to provide aid to countries in need.

These are the following questions that Sarah wants answered:

* **Resource Allocation**: How can we distribute our resources effectively to maximize impact in regions facing the most urgent needs?
* **Impact Assessment**: What is the impact of our ongoing aid programs? Are we meeting our goals, and where are we falling short?
* **Geographical Focus**: Which countries or regions are currently experiencing the most critical situations that require immediate intervention?
* **Logistical Failures**: Are there countries where the resources aren’t being used effectively?

***Scenario of Use:***

Sarah begins her day by accessing the annual world situation dashboard, which provides a comprehensive overview of the global landscape. Given that the dashboard only contains information about the situation of the world yearly, Sarah relies on quarterly revisions to inform her decisions. Today, she notices a disparity in aid requirements across various countries. Utilizing the quarterly data, Sarah identifies regions facing escalating crises and determines the countries that should receive aid promptly.

Upon closer examination, the dashboard highlights areas where resources are underutilized or not effectively addressing the identified needs. Sarah focuses on devising strategies to support these countries better. She considers targeted interventions, reallocation of resources, or collaboration with local partners to enhance the impact of aid programs. Throughout the day, Sarah utilizes the dashboard insights to make informed decisions, ensuring GAI's resources are directed where they are most urgently needed and used efficiently.

***Device specifications:***

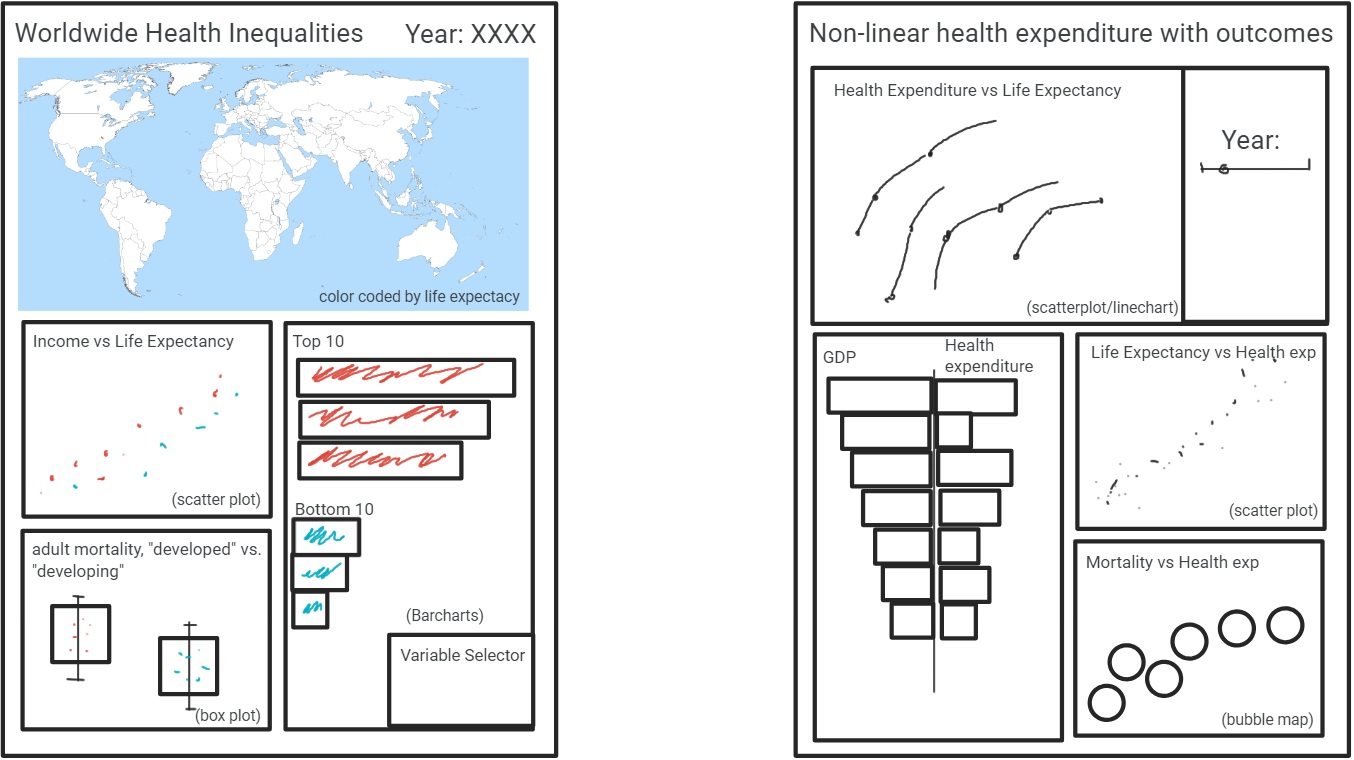
Sarah would use her personal computer with a standard 1920x1080px monitor in a well-lit space in her office, so the visualization can be of standard use.

With this persona and the data we have we decided that the main message we want to transmit is “Health Inequalities and Health Expenditure related to Outcomes”. This will help our target audience find the answers to the questions previously commented.

Since we noticed that spending more on health doesn’t always correlate to better outcomes such as life expectancy, mortality rates, etc.. We wanted to answer this question to get to know what countries are misusing their resources.

## 2.2. Napkin design

Lastly, we created a “napkin” design that is low-fi:



Here we created two dashboards, one for the general overview of health inequalities, and one to compare how well is the health expenditure working in each country. This design will help us transmit our message with the use of easily digestible and understandable charts, where we highlight the disparity between countries.

For example, we use the top 10 countries and the bottom 10 countries with different variables to see how distant they are. We also used a map to showcase the general state of the world visually. Additionally, we color-coded the bottom 3 charts by “developed” and “developing” to see the different tendencies between the two statuses. Lastly, since we also want to emphasize the evolution over time, we added a year filter on the top right of the dashboard where you can choose the year or let it play out.

On the other hand, for the specific dashboard, we used different variables to compare it to the Government Health Expenditure. In this way, we can showcase that there are countries that aren’t using their resources efficiently. On the main chart, we will use just a few countries with very distinct patterns to demonstrate the different behaviours between health expenditure and the outcomes. Since we also want to show the evolution over time we also added a year filter, but just for the main chart.

# Selection of charts and encoding

After some iterations and considerations we decided to change the layout of the dashboards to landscape instead of vertical so that the user doesn’t have to scroll and that it fits on a single screen.

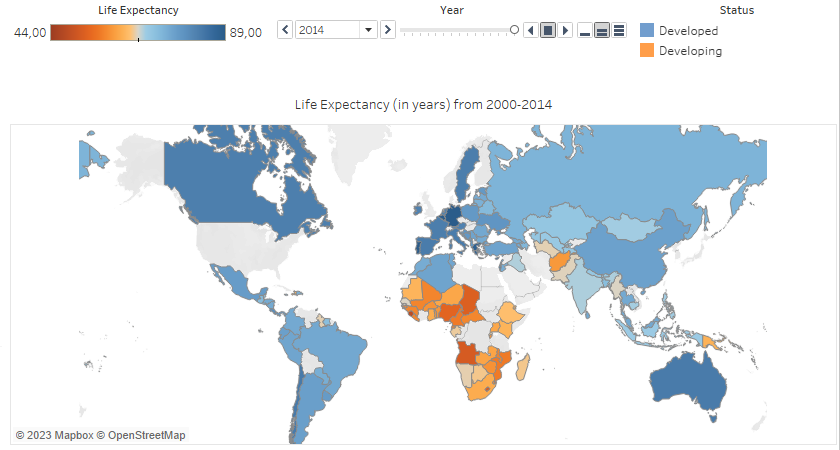
## 3.1. General Dashboard: Worldwide Health Inequities

For this dashboard we decided to use the next charts:

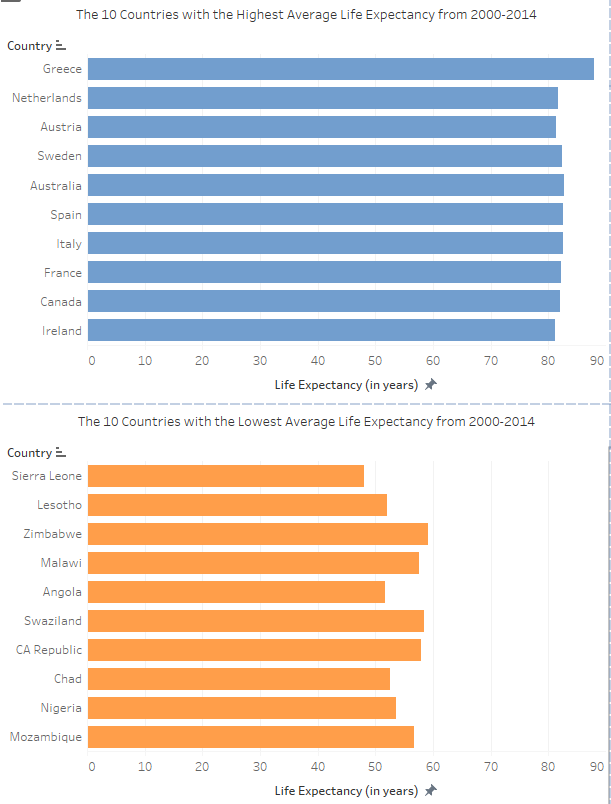
* **Life Expectancy:** We wanted to show the overall life expectancy tendency over the world, so we chose to show it with a world map box with two distinct color palettes, orange for developing countries and blue for developed countries.

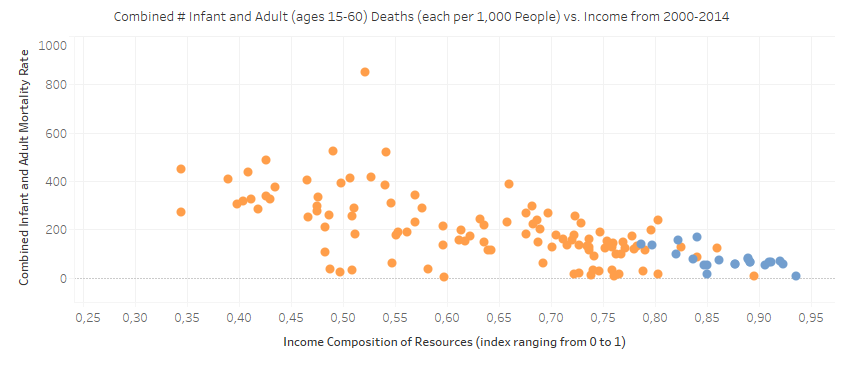
We also added gradients in those color palettes so we can see the life expectancy per country visually. This main chart is used as a guidance on where to locate the countries that the other charts refer to by means of the same color encoding.

Notice that here we used color and position as preattentive dimensions to encode information. The Gestalt laws of perception that can be highlighted in this diagram are proximity (because of the position of each country) and similarity (in the intensity of the color of each country).



* **Top and Bottom Countries Life Expectancy:** We chose two barcharts to showcase the different rankings between the countries, here we can see the top 10 countries by life expectancy. We chose to show the top countries in blue and the bottom countries in orange to show that these countries are correlated with the developed and developing countries shown earlier. Since both are bar charts we start the axis at 0. We also used the dimension of length as a means of separating the data. Recall that length is not an easily separable dimension. However, this is not an issue here, since we do not want to provide an exact comparison of lengths, but only provide general information that there are countries with larger values than others. If the user wants to get the exact value in each case, it can resort to the interaction to get the exact value for each country.

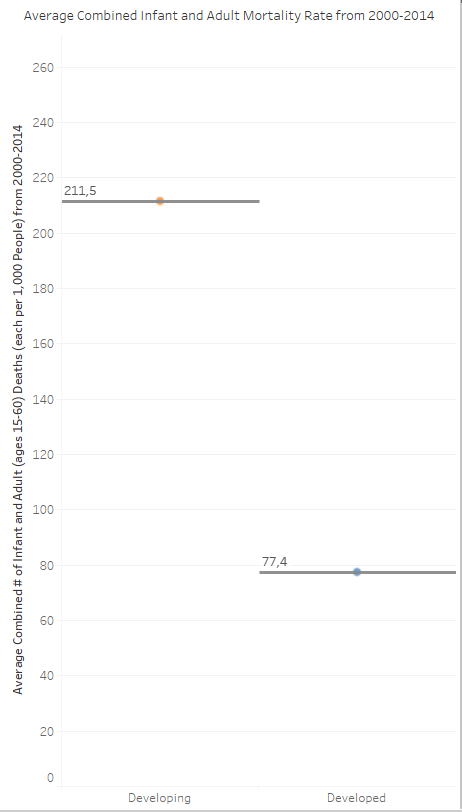


* **Combined Infant and Adult Mortality vs Income:** We wanted to showcase how income correlates to the mortality rate of most of the population. To this end, we chose to use a scatter plot with the same color encodings of “Developing” and “Developed” countries. We used the combined number of Infant and Adults deaths per 1000 people. 

Here we are using color and position encodings which are highly separable dimensions. The Gestalt laws of perception that apply here are similarity (we clumped together the orange and blues in the scatter plot), and proximity (in the 2d distribution).

* **Average Combined Infant and Adult Mortality:** We also wanted to show the disparity between the mortality of developing countries and the mortality of the developed countries. For that, we used a boxplot with no variance so it’s only the average as a line.

We didn’t use a full boxplot in the end because it is harder to understand and our audience doesn't necessarily have a statistical background.



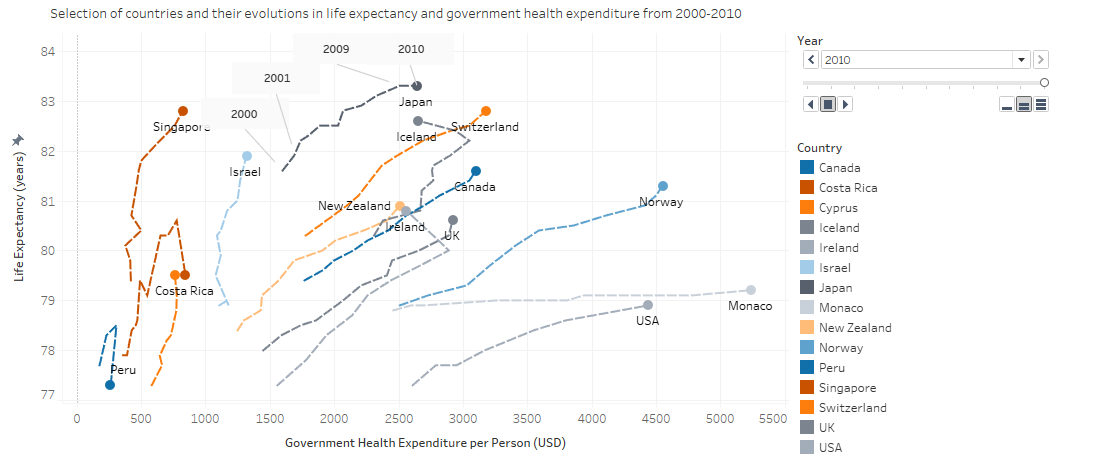
Generally speaking, we mostly separated all the charts with colors and space, we didn’t use more dimensions to separate the different data as we think these follow the necessary best practices in that regard.

Layout wise we distributed the bar charts together since they are similar. We also used the same colors all around the layout to be consistent and keep the relationship between blue and orange to developed and developing.

## 3.2. Specific Dashboard: Government Health Expenditure

For this dashboard we decided to use the next charts:

* **Life Expectancy vs. Government Health Expenditure**: We wanted to show how the health expenditure of governments per person varied over time, so we correlated this to the life expectancy. Thus, we created a connected-scatter animated-chart to show the different patterns in the data because different countries show different patterns, even inside the same status. This has the intention of sending a message on efficacy with health expenditure.

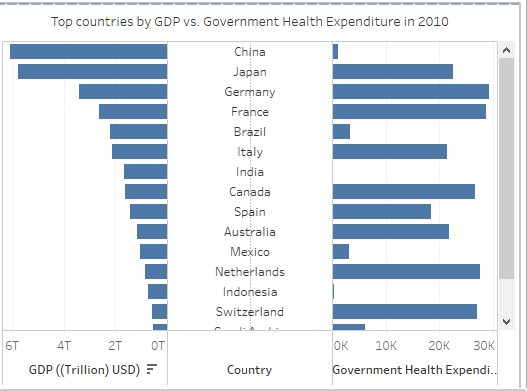
The following Gestalt of perception can be highlighted in this chart: similarity, as countries share similar patterns; focal point, since the last year is emphasized; connectedness, since the points from the same country are connected together; and common faith, sin some countries show similar tendencies. We also set that by default the graph shows the complete trajectory without any animation, since the animation can be more difficult cognitively, the user can choose to play an animation or go year by year.

Notice that this chart encodes information by color and position, which are highly separable and preattentive dimensions. If the animation of the chart is played, the monition preattentive dimension is also involved.

Generally, we expect the progress of the years to go from left to right and down to up, and this is the case for this chart. However, to dismiss any doubt, we also included some manual annotations on one of the countries with the label of each year, just to make it clearer for the end user.

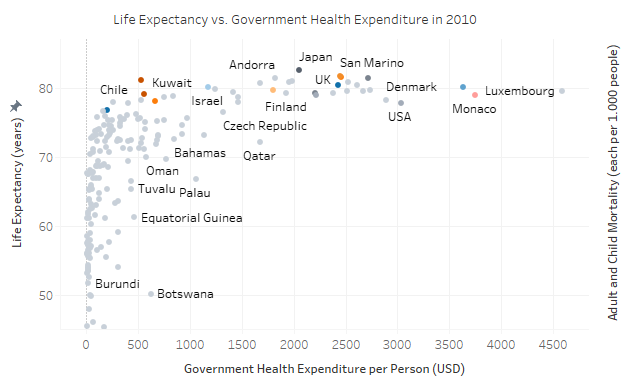
* **Top countries by GDP vs. Government Health Expenditure:** Here we wanted to show the discrepancy between the large GDPs that some countries have and their health expenditure per person, this sends the message that larger GDPs don’t always correlate to large health expenditures by the government. Thus we used a butterfly chart to showcase the disparity between them since we want to compare two distributions. We ordered the left-hand side of the chart by GDP, so it is also a ranking of the largest GDPs worldwide.

We only used the dimension length to separate the data. This is fine since the idea of this chart is not to give exact data but to give a general idea of the situation between the indexes. Some gradient could be used for redundancy in this graph but we decided to omit this redundancy information. The Gestalt laws that apply here are symmetry, given the symmetric nature of the graph; other Gestalt laws that can be found are proximity and continuity, as the bars are grouped together by GDP and Health Expenditure and each row has a feel that it belongs to a single country.

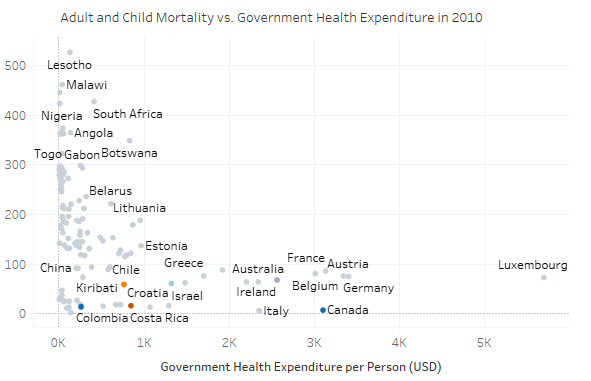


* **Life Expectancy vs. Government Health Expenditure:** The objective of this graph is to show that there’s a clear correlation between two continuous variables. We thus resorted to a scatter plot to show this correlation and highlighted those countries that appeared on the first and main charts of the dashboard.

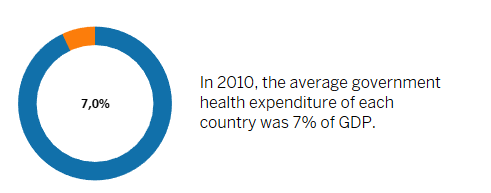
We used the location and color as dimensions since they are both highly separable. The main Gestalt law that applies here is proximity since they are countries that are close together in the 2D space.



* **Adult and Child Mortality vs. Government Health Expenditure:** The objective of this graph is the same as the last one, showing the correlation between the indexes. We also used the scatter plot to demonstrate it, everything we applied to the last one is also used in this one.



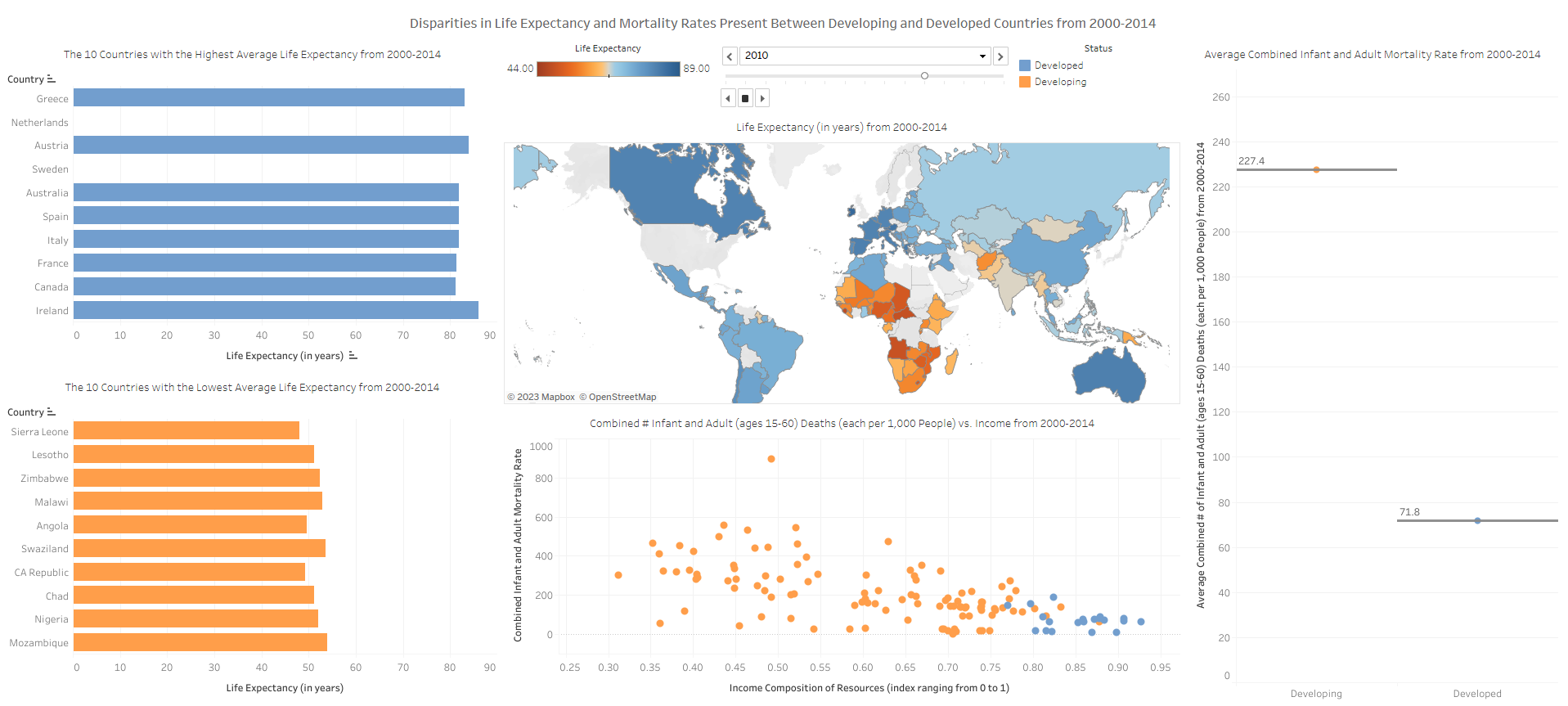
* **Donut chart:** We added a small informative doughnut chart under the tile to act as an opener to thought. This chart shows the average percentage of the GDP used in health expenditure. We chose orange and blue since they are opposite colors.

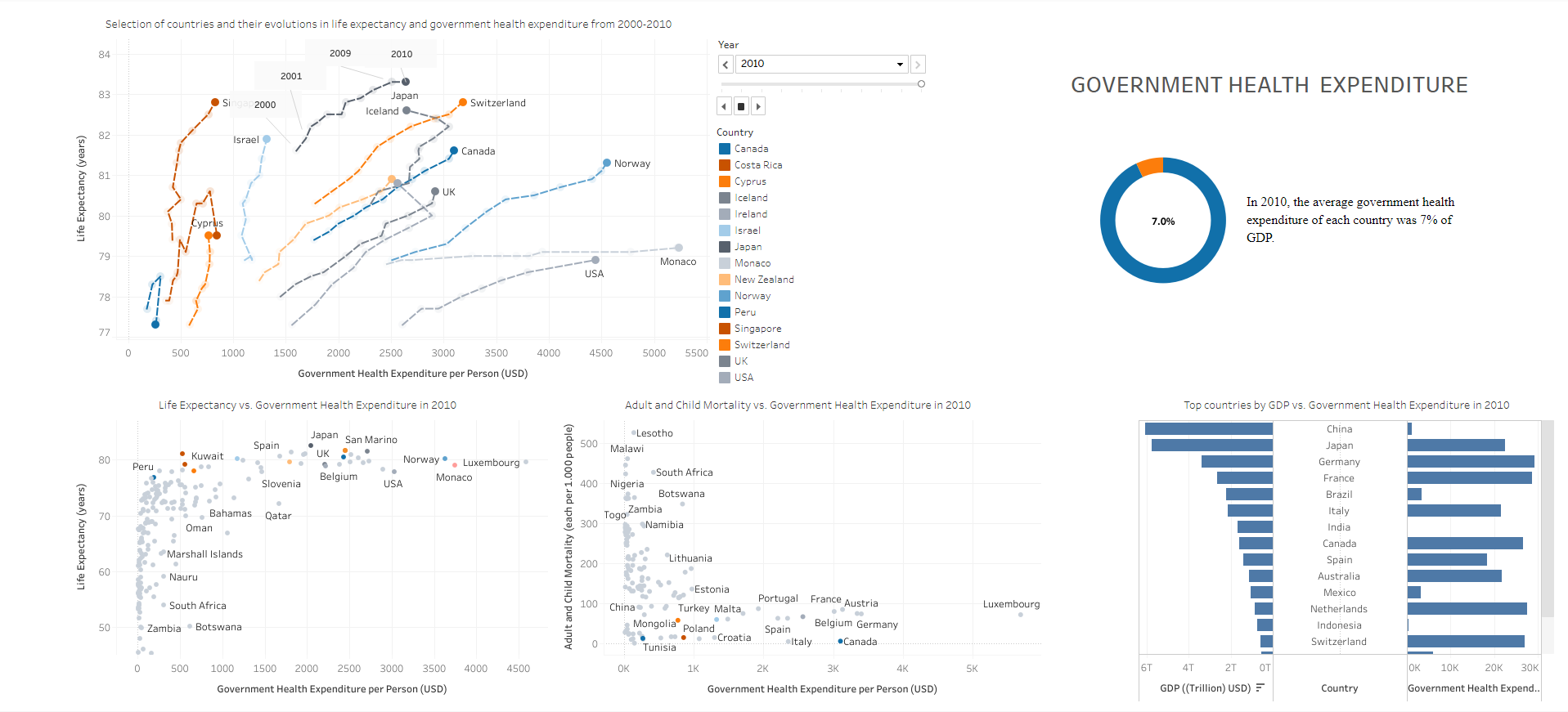


Layout wise, this dashboard was grouped into the scatter plots and the animated chart since they share colors (following the Gestalt law of proximity, which says that closer items are perceived to belong to the same group). Additionally, due to sharing colors we grouped the charts into the same legends to showcase the same countries in all three charts. On the other hand, the butterfly plot doesn’t share any colors or anything with the other ones, so we placed it further away. Additionally, since we want to highlight the connected-scatter diagram over the others, we placed it at the top-left of the dashboard.

# Implementation of the Dashboards

To implement these two dashboards we used Tableau. We added some interactions in the dashboards:

* **General Dashboard: Worldwide Health Inequities**
  + Inside the Mapbox we can zoom and pan the map to see countries of interest for the user. Here we find a zoom interaction and a navigation interaction.
  + We can also change the visualization over time with a slider or selector of year on the top of the dashboard, this would be considered as a navigation interaction.
  + There’s also the possibility of ordering the bar charts with an order interaction button. This lets us display the information in ascending or descending order to better compare it.
* **Specific Dashboard: Government Health Expenditure**
  + We find a year selector to change the visualization over time, here the year affects the scatter plot and the animated chart. This is a navigation interaction.
  + We can also filter countries by including or excluding them in the legend, as a filter interaction. This lets us focus on a single country on all three charts.
  + We can also order the butterfly chart in descending or ascending order, letting us better compare the variables in this plot.



* **Public access**  
    
  We can find the general dashboard in this link:

<https://public.tableau.com/app/profile/madison.chester/viz/DisparitiesinLifeExpectancyandMortalityRatesPresentBetweenDevelopingandDevelopedCountriesfrom2000-2014/PVDashboard?publish=yes>

And we can find the specific dashboard in this one:

<https://public.tableau.com/app/profile/alex.oarga/viz/Dashboard_17004993879660/Dashboard?publish=yes>

**Note:** even though the dashboard was developed using the generic desktop setting, some users with smaller screen ratios have encountered problems visualizing it properly. If you encounter any issues visualizing the above dashboard, please try using less zoom on your default browser and reload the webpage.

These dashboards can also be found in their respective .twbx in the same zip file this document is in.