Data Visualization
Global Health Indicators Dataset Analysis

Participant	E-Mail	Exam-Number
Ann-Kathrin Mösel	anmos23@student.sdu.dk	4123808
Michael Gruber	migru23@student.sdu.dk	4123882
Daniel Sakarli	dasak23@student.sdu.dk	4123822
Jakob Ladegaard Møller	jamoe20@student.sdu.dk	486387

Date of submission: 28.12.2023

Abstract

The dataset contains health and development indicators over the years of 2000 until 2015 and for different countries all over the world. The indicators contain for example the life expectancy, infant deaths, alcohol consumption, Hepatitis B immunization coverage, among others. These parameters were visualized to give an in-depth understanding of the developmental aspect of diseases in our world. Different forms of visualization, such as bar chart, world maps, bubble plots as well as scatter plots helped in this aim. It was possible to achieve a good understanding of the development of diseases with the animated plots over the years.

Background and Motivation

The COVID-19 pandemic has shown us how important health is globally. It has affected the whole world and highlighted the need for strong healthcare systems and surveillance of possible upcoming threats. This made us interested in researching health and development indicators globally.

We are looking at various aspects, not just the immediate health crisis. We want to understand the connection between parameters like money, education, and overall well-being for people everywhere. To do this, we are using different ways to display the information, like charts and graphs.

In our group, we have students with different backgrounds. Some are studying medical technology, bringing in medical knowledge and technology skills, some have specializations in data science. This mix of perspectives helps us understand the dataset better.

The connection between medical technology and global health makes our project even more interesting. It lets us not only study healthcare and diseases but also think about how technology can help make health better and promote sustainable development.

Basically, we are curious to understand global health, and our group's different backgrounds help us do that. The mix of medical technology, public health, data science and data visualization give us tools to study the dataset deeply.

Project Objectives

In our project, we used various charts to delve into significant aspects and gain a better understanding of global health. The central question is: "How do factors like a country's status, alcohol consumption, and the spread of diseases impact the world, and what insights can be drawn from the generated charts?"

We initiated our exploration with an animated plot showcasing the life expectancy in different countries and years. This plot is shown in ***Figure 1: Bubble plot – Life Expectancy***. What factors might contribute to different life expectancies in various countries and years? We aim to discover if there are specific years or places where people live longer and if this is connected to factors such as healthcare access, income, or education. ***Figure 2: Boxplot - Life Expectency*** shows this. Furthermore, how does a country's status influence people's life expectancy? Therefore, we created another animated boxplot depicting life expectancy in developed and undeveloped countries.

We expanded our investigation with a world map, in ***Figure 3: World Map - Disease***, illustrating the spread of different diseases in various years. Are there trends in the spread of diseases over the years? This aids in understanding the global movement of diseases and identifying changes over time. Delving deeper into specific diseases, we crafted a chart for Hepatitis and alcohol consumption, exploring trends and connections between the two in different countries and years. This is shown in ***Figure 4: Bubble plot - Hepatitis / Alcohol***. Given that Hepatitis B affects the liver, and alcohol can impair liver

function, we are exploring potential connections. If someone already has Hepatitis B, alcohol consumption can worsen the disease and elevate the risk of severe liver damage.
Utilizing an animated bubble plot, in ***Figure 5: Bubble plot – Alcohol / Adult Mortality***, we scrutinized the effects of alcohol consumption on adult mortality in different countries each year. We are keen to understand if there is a correlation between alcohol consumption and life expectancy. Additionally, the chart helps us discern changes in drinking habits over the years. We are also intrigued by the connection between a country's status and alcohol consumption, for which we crafted another boxplot. This plot is shown in ***Figure 6: Boxplot - Alcohol.*** What role does a country's status play in relation to alcohol consumption?
Through Scatter Plots, we also investigate the relationships between health indicators in different countries. The first plot explores HIV/AIDS and Mortality Probability (ages 15 to 60), identifying trends and variations. This is shown in ***Figure 7: Scatterplot – HIV/AIDS / Adult Mortality***. The second plot focuses on child health, studying Death under the age of 5 and Hepatitis B immunization coverage. And this is shown in ***Figure 8: Scatterplot – Hepatitis B / Death under 5***. By using continent-specific colors, we aim to highlight patterns across regions. These visualizations offer insights into global health trends, making complex information accessible.
Our analysis includes a bar chart for BMI and adult mortality shown in ***Figure 9: Bar chart - BMI / Adult Mortality***. What conclusions can be drawn from variations in BMI across different countries and years? To what extent is BMI linked to adult mortality? We seek to understand how BMI varies across countries and years and if there are notable differences. Additionally, we are curious if BMI influences adult mortality.
In summary, our project aims to answer questions such as where infant deaths are more common, how alcohol and diseases spread, and the connections between alcohol, life expectancy, BMI, and adult mortality. When we examine these charts collectively, patterns emerge, enhancing our understanding of how various factors influence global health. These charts are crucial tools for sharing information and discussing ways to create a healthier world for everyone.

Data

Our project relies on the comprehensive "Life Expectancy Data" obtained from Kaggle, accessible through the following link: Health & Development Indicators: Global Insights (kaggle.com).
The dataset covers a rich array of variables providing insights into life expectancy, spanning various countries and years. The dataset includes datasets from multiple countries, with each dataset providing a snapshot for a specific year. The essential variables included are:

- Country: Identifies the country for each data entry.
- Year: Represents the year of observation, allowing for temporal analysis.
- Status: Categorizes countries into "Developed" or "Developing," offering a socio-economic perspective.
- Life Expectancy: The primary outcome variable, reflecting the average expected lifespan.
- Adult Mortality: Signifying the probability of dying between the ages of 15 and 60 per 1000 population
- Alcohol: Indicates the average alcohol consumption per capita.
- Hepatitis B: Presents the Hepatitis B immunization coverage among 1-year-olds.
- BMI (Body Mass Index): Represents the average Body Mass Index across the population.

- Diseases: Includes information on the prevalence of various diseases.
- Infant Deaths: Specifies the number of infant deaths per 1,000 live births.

There are a lot of records in the dataset, which helps us look at many different situations and make more reliable conclusions.

Visualization/Dashboard

In our data visualization project, we carefully considered how to present our data for a comprehensive understanding. We used the following plots: world maps, animated bubble plots, animated boxplots and scatter plots as well as bar charts. The dashboard can be viewed under following URL: https://data-viz.ladegaardmoeller.dk/

Bubble plot – Life Expectancy

We have chosen an animated bubble plot as the primary visualization tool. This dynamic plot allows us to investigate the changing dynamics between the gross domestic product (GDP) per capita and life expectancy across different countries (see Figure 1). Each bubble on the plot corresponds to a specific country, with the x-axis representing GDP per capita in US-Dollars (USD) and the y-axis depicting life expectancy in years.

To make it easier, we have given each group of countries from the same part of the world a different color. Also, the size of the bubbles shows how many people live in each country. This way, we can understand not just how rich a country is or how long people live but also how many people are there. Our special Animated Bubble Plot lets us see how things change over the years, and we hope to learn more about how money and health are connected all around the world.

In summary, our choice of an animated bubble plot with specific visual encodings like color-coded continents and bubble size for population aims to offer a dynamic and comprehensive exploration of the intricate relationship between GDP per capita and Life Expectancy. This design allows for an engaging and informative analysis of global health trends and economic indicators.

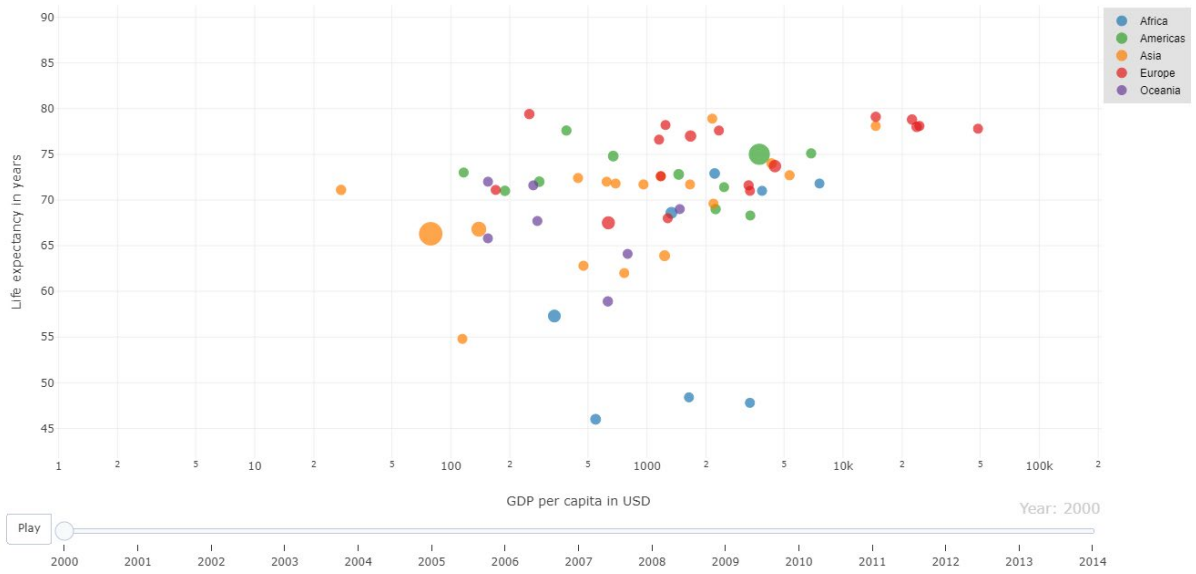

Figure 1: Bubble plot – Life Expectancy

Boxplot – Life Expectancy

To represent the life expectancy we chose an animated boxplot. The y-axis displays life expectancy, while the x-axis categorizes countries into developed and developing countries (see Figure 2). The animation progresses through each year, showcasing changes in life expectancy over time. Additionally, the plot integrates an interactive element.

The boxplot is ideal for analysis as it summarizes the parameter of alcohol consumption and lets us compare the two categories. The categories are the countries which are divided into developed and developing countries to allow for a simplified but meaningful classification. This binary division aids in quickly discerning broad patterns and disparities. The animation feature enables the audience to witness trends unfolding over time. The interactive feature makes it easier for users by showing exact numbers when you hover over the plot. It helps you get a closer look at the details of how life expectancy is spread out.

The boxplot allows us to compare the life expectancies of developed and developing countries across different years. This visual representation aids in identifying trends and variations. The animation feature adds a dynamic element, enabling us to observe how life expectancy evolves yearly. This dynamic aspect is important for capturing the temporal dimension. Including all countries in the plot ensures a comprehensive overview. This provides us with a holistic understanding of global life expectancy patterns.

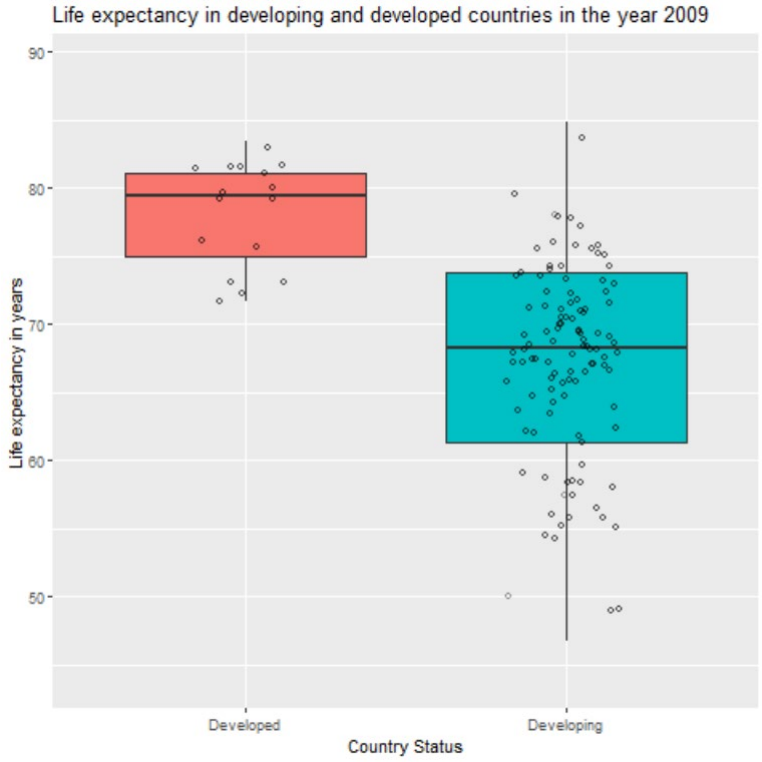

Figure 2: Boxplot - Life Expectency

World Map -Disease

We expanded our exploration to incorporate a world map showcasing various diseases over different years. The featured illnesses include Measles, Hepatitis B, Polio, HIV/AIDS, and Diphtheria. Each map is color-coded to depict the percentage of the population affected by a specific disease (see Figure 3).

An exception is the HIV/AIDS parameter which is shown as prevalence. That is how many people of a specific population currently have HIV/AIDS and is typically as the number of affected people per 10,000 or 100,000 people. Unfortunately, the documentation of this dataset does not specify which prevalence is shown, either per 10,000 or 100,000 people. Nevertheless, a comparison between different countries is still possible. The world maps are generated for each year individually, and there is also the possibility of illustrating the averaged parameters over the years.
The coloring scheme on the map offers a swift comprehension of the proportion of the population affected by each disease in different regions. This representation was chosen to facilitate the identification of trends in the spread of diseases.
The selection of diseases aims to provide insights into the dynamics of global health for diverse ailments. This allows for a comparative analysis of how differently these diseases are spread worldwide.
Similar to the animated boxplot, our world map ensures inclusivity by incorporating data for all countries. This approach grants a comprehensive understanding of how diseases have globally evolved over the specified years, contributing to a more thorough insight into global health patterns.

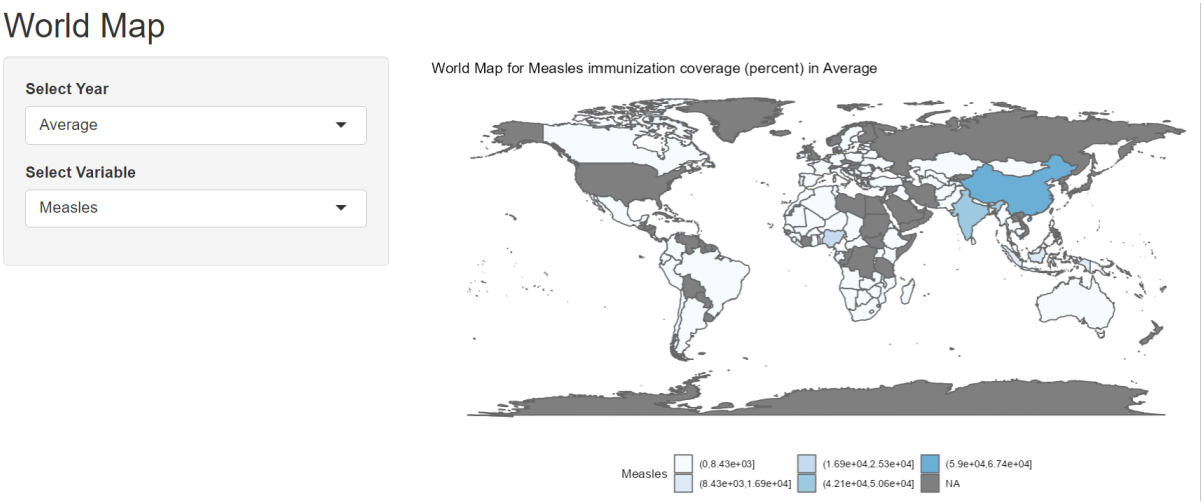

Figure 3: World Map - Disease

Bubble plot – Hepatitis B / Alcohol

A moving bubble chart was chosen to show how alcohol consumption and Hepatitis B are connected in different countries. This special chart changes over the years. The x-axis of the plot represents alcohol consumption, while the y-axis showcases the percentage of the population affected by Hepatitis B as can be seen in Figure 4. Each country has its own bubble, and the color of the bubble shows which continent it belongs to. Also, when moving the cursor over a country, it shows more details such as the country's name, how long people live on average, and the exact numbers for alcohol consumption and Hepatitis B immunization coverage.
The choice of a dynamic bubble chart aligns with the intention to present an informative representation of the relationship between alcohol and Hepatitis B. The dynamic aspect of the chart allows for a time-centric exploration, showcasing changes and trends over the years. The use of bubbles, with each country having its own, enables a clear differentiation of data points, facilitating easy identification and comparison.

The decision to use color to represent continents aids in immediately associating data points with geographical regions. This visual encoding simplifies the understanding of global patterns and allows users to quickly discern the distribution of countries across different parts of the world.
The inclusion of an interactive feature, where detailed information appears upon hovering over a data point, enhances user engagement. This interactivity enables users to delve into specific details about each country, fostering a more in-depth exploration of the dataset.
The size of each bubble corresponds to the population size of the country. This way you can visually represent the size and influence of each country.
The dynamic of the bubble chart aligns with our goal of showcasing changes over time. This visual representation helps us look at different times and find out how alcohol use and Hepatitis B sickness change.

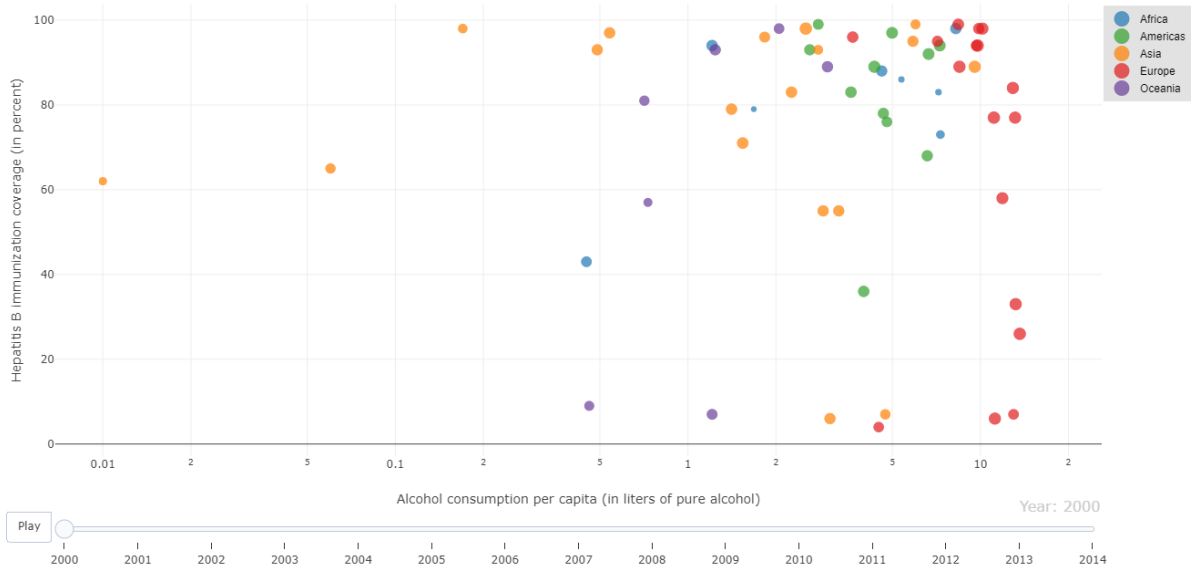

Figure 4: Bubble plot - Hepatitis / Alcohol

Bubble plot – Alcohol / Adult Mortality

We created a graph to understand how adult mortality and alcohol consumption relate across countries. In this visual, each country is shown as a bubble in the plot. The x-axis displays adult mortality per capita, the y-axis shows alcohol consumption, and the size of each bubble reflects the country's population. To make it easy to see, different colors for each continent were used as can be seen in Figure 5.
A bubble plot was chosen, because it helps to see how the two parameters – adult mortality and alcohol consumption – are connected. In the bubble plot patterns and differences in global health are quickly observable.
The size of each bubble represents the country's population, adding another layer of information. Bigger bubbles stand out, showing which countries have a bigger impact on the overall picture.
Colors were used to group countries by continent. This helps the user to see trends in different parts of the world more easily. Each continent has its own color for every generated plot, making it simple to compare.

Furthermore, an interactive feature was added. When moving the cursor over a bubble, it tells the user more about that country, such as its name and the average life expectancy. This makes it easier for anyone to explore and learn from the graph.

In conclusion, the design choices, such as the bubble plot, bubble size, continent colors, and interactivity, work together to create a visual that is easy to understand. This way, users of the graphs can get a good sense of global health trends and explore specific countries for more details.

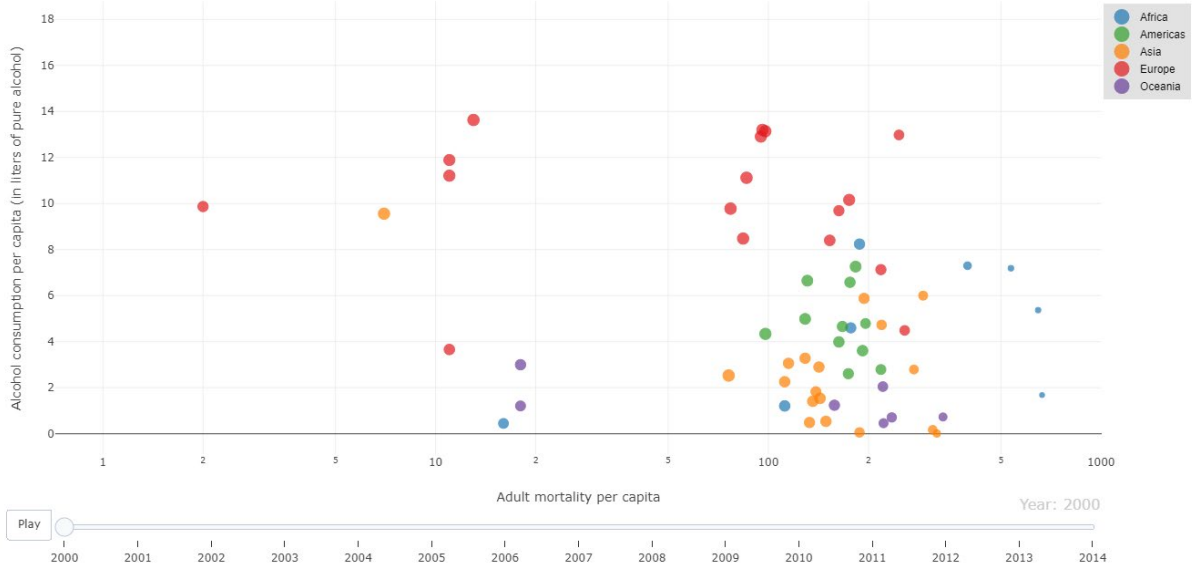

Figure 5: Bubble plot – Alcohol / Adult Mortality

Boxplot – Alcohol

We delved into the intricate patterns of alcohol consumption across various country statuses, i.e. either a developed or a developing country, spanning multiple years. Our tool of choice was a boxplot graph, where the x-axis unravels the country status (e.g. developed or developing), while the y-axis lays bare the nuances of alcohol consumption (see Figure 6). Each plotted point corresponds to a specific country, and the timeline encapsulates several years.

The boxplot unveils the rich tapestry of alcohol consumption variations within distinct country status groups. With country status neatly aligned on the x-axis, we aimed for a straightforward comparison of alcohol consumption trends, making it easy to discern patterns between developed and developing countries. This arrangement was pivotal in swiftly spotting trends and potential distinctions.

By extending our plot over the expanse of multiple years, the stage for a keen observation of the ebb and flow of alcohol consumption over time was set. This strategic approach empowers viewers to unravel evolving patterns and subtle shifts in drinking habits across the spectrum of country status groups.

In conclusion, the boxplot design, where country status graces the x-axis and alcohol consumption takes center stage on the y-axis over an extended timeline, provides a crystal-clear framework for dissecting trends. This method allows for a swift yet comprehensive comparison between different country status groups, offering valuable insights into the potential transformations in alcohol consumption patterns over the course of time.

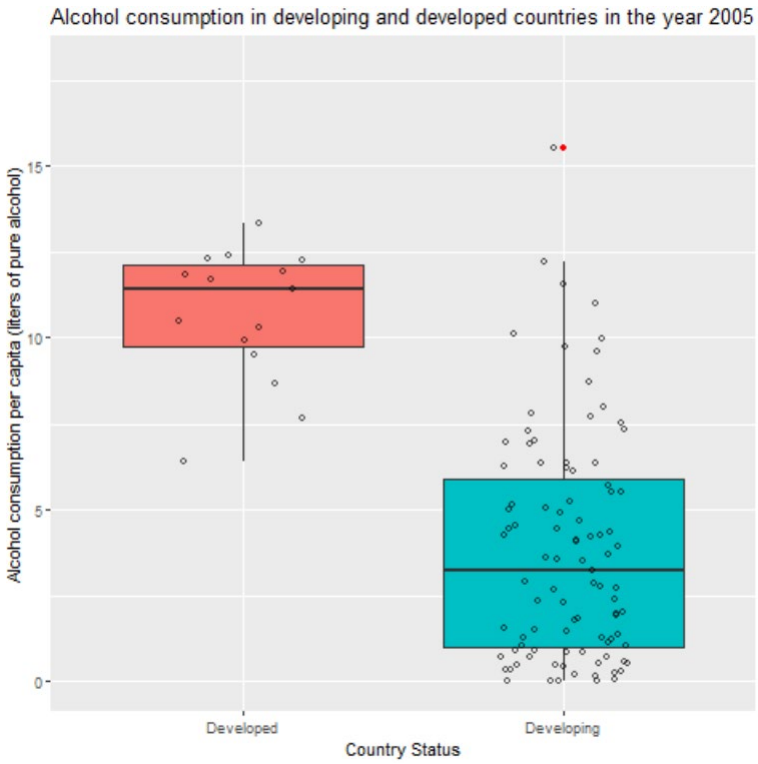

Figure 6: Boxplot - Alcohol

Scatterplot – HIV/AIDS / Adult Mortality

We have opted for a scatter plot to illustrate the relationship between HIV/AIDS prevalence and mortality probability between 15 and 60 years. Each point on the plot represents a different country, and to enhance clarity, we have employed distinct colors for each continent. The x-axis displays HIV/AIDS prevalence, while the y-axis showcases mortality probability, with the x-axis being in percentage and the ongoing problem, that the documentation of the dataset does not offer clear explanation on which prevalence was chosen for the HIV/AIDS parameter. The inclusion of a linear regression line helps in identifying the overall trend in the data, offering a visual guide for understanding how these health factors relate (see Figure 7). Importantly, an interactive feature allows users to hover over each point, revealing specific details such as the country's name and the exact values for HIV/AIDS prevalence and mortality probability.

This design choice aims to make the visualization accessible and informative. The scatter plot enables a quick understanding of the connection between health factors for each country. The use of colors facilitates a straightforward comparison between continents, and the linear regression line provides a visual guide to discern trends. The interactive feature adds a layer of engagement, allowing users to explore detailed information about each country. Overall, the design intends to offer a clear and comprehensive exploration of the relationship between HIV/AIDS and mortality probability in different countries across the globe.

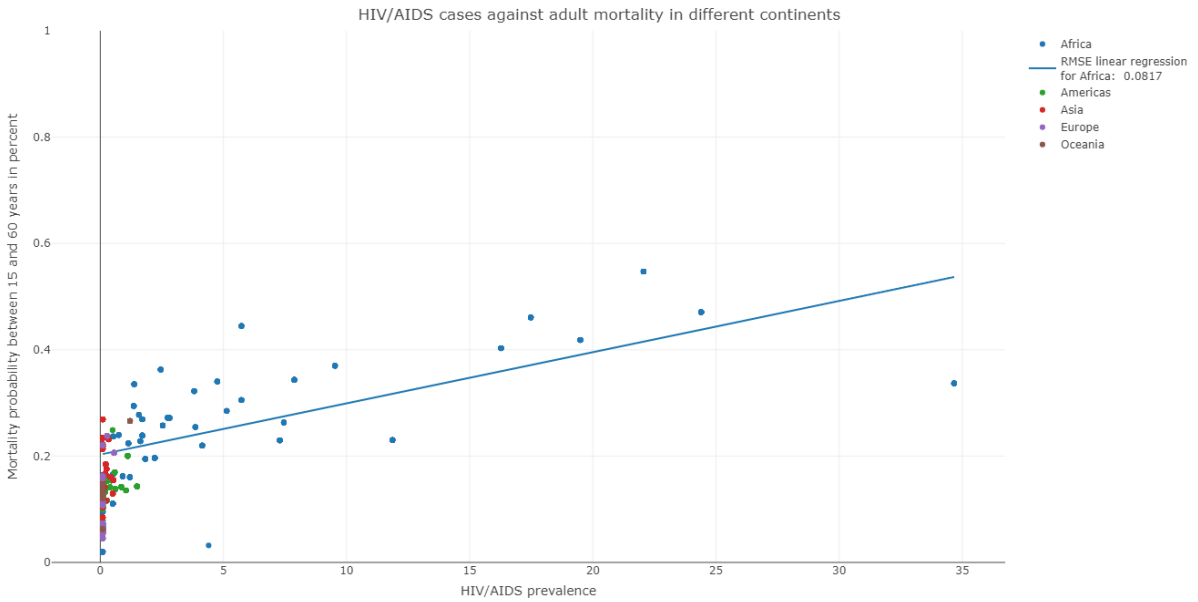

Figure 7: Scatterplot – HIV/AIDS / Adult Mortality

Scatterplot – Hepatitis B / Death under 5

We also have opted for a scatter plot to explore the relationship between death under the age of 5 and Hepatitis B immunization coverage. In this visual representation, each point corresponds to a different country. To make it more straightforward, each continent is represented by a unique color. Additionally, an interactive feature allows users to hover over the points, revealing specific details about each country, including the exact values for death under the age of 5 and Hepatitis B immunization coverage.

The x-axis represents death under the age of 5, and the y-axis illustrates Hepatitis B immunization coverage, both in percentages as can be seen in Figure 8. The use of distinct colors for each continent aids in easily grouping countries, providing insights into geographical patterns. The interactive feature adds an extra layer of engagement, allowing users to gather detailed information by simply hovering over each data point.

The scatter plot design is chosen for its effectiveness in illustrating the relationship between these two health indicators for each country. The color-coded continents help users quickly identify geographic trends, while the interactive feature enhances the overall exploration experience. This design aims to offer a clear and engaging understanding of child health, examining the interplay between death under the age of 5 and Hepatitis B immunization coverage across different countries and continents.

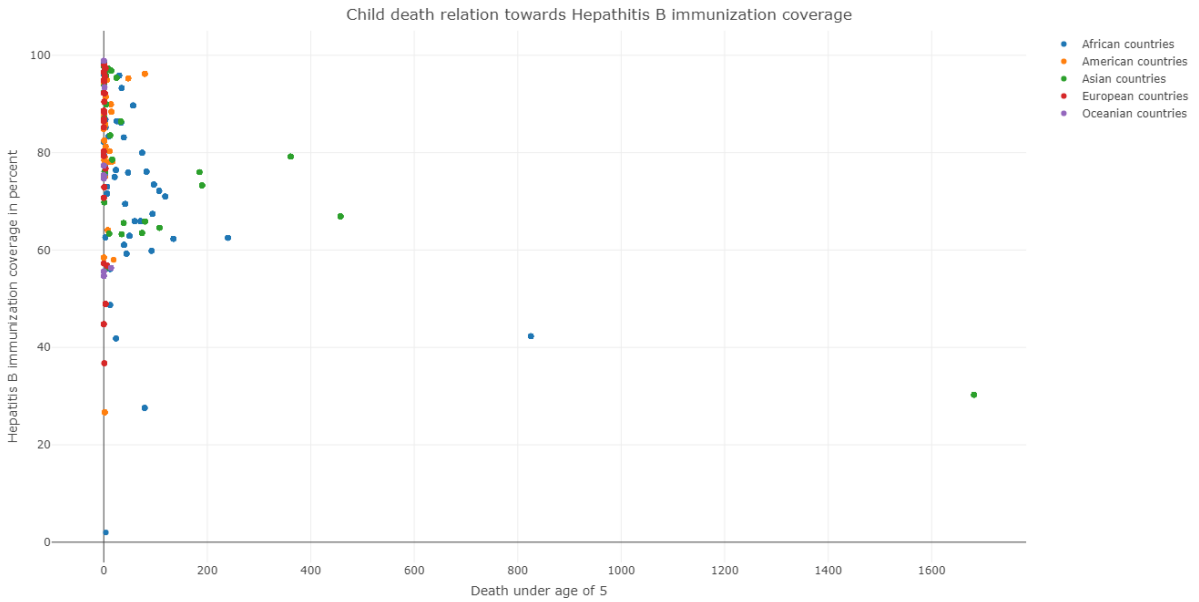

Figure 8: Scatterplot – Hepatitis B / Death under 5

Bar chart – BMI / Adult Mortality

We chose a bar chart to study the connection between BMI (Body Mass Index) and adult mortality in different countries as can be seen in Figure 9. The chart has countries on the x-axis grouped by continent, and the y-axis shows the values for BMI and adult mortality. Specific continents and years can be picked to look at, and there is also a graph showing the average values over different years for each continent. When you move the cursor over the bars, you can see the exact values for BMI and adult mortality.

The focus is on the bar chart, where each bar represents a country grouped by continent. This setup makes it easy to compare health indicators in a user-friendly way. Grouping countries by continent helps understand regional patterns. The interactive feature adds to the experience by providing detailed information when you hover over the bars.

In conclusion, using a bar chart with features like continent grouping and interactivity aims to make it simple and engaging to explore BMI and adult mortality. This design allows for easy comparisons across countries, continents, and years, providing useful insights into global health trends.

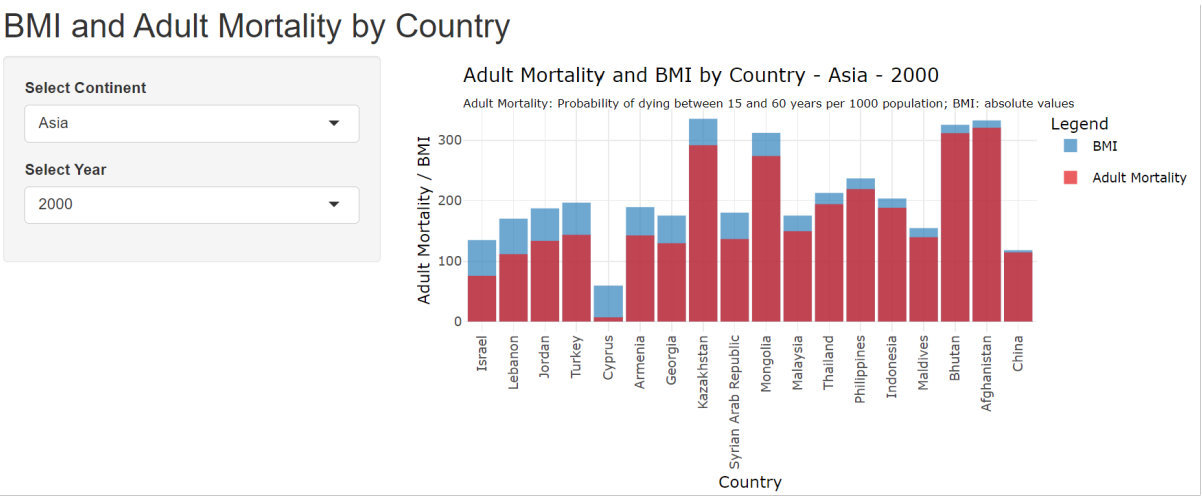

Figure 9: Bar chart - BMI / Adult Mortality

Story/Results

Bubble plot – Life Expectancy

We started by looking at a moving graph that shows how long people live in different countries over the years. It is a bubble plot, and in the dashboard, you can see it in action. We wanted to find out what makes life expectancy different in various places.

As we looked at the moving graph, we noticed interesting things. Some places saw more people living longer as their economy grew. But in other places, things were more complicated, and it wasn't just about money. The graph helped us see these patterns and differences.

Our graph did a good job of answering our main question about how money influences how long people live. It also showed us a more detailed story than we expected. It made us realize that other things, besides money, play a role in health. The graph changes over time, helping us see how things are getting better or worse.

Our graph is good at showing trends over time, but we could make it even better. We could add more information, like how much a country spends on healthcare or how educated the people are.

In summary, our moving bubble graph was helpful in telling us about life expectancy worldwide and how it relates to money. It met our basic expectations and taught us more than we thought.

Boxplot – Life Expectancy

We wanted to find out if there are specific years or places where people live longer, and if this has to do with things like healthcare, money, or education. The graph we made, Figure 1, helps us see this. We also wondered how a country being rich or not affects how long people live, so we made another moving graph for developed and undeveloped countries.

What we found in the graph is that developed countries generally have more people living longer. But as we looked over the years, we saw some changes and differences. This made us realize that the connection between life expectancy and a country's development is not so simple.

Our graph did a good job of answering our questions about life expectancy and development. It showed us that the relationship is not straightforward, and things change over time. The interactive feature helps a lot, but we could make the graph even better by adding more information, like education and money.

To sum up, our moving boxplot helped us understand how life expectancy changes over time and with a country's development. It met our basic expectations. To make it even better, we could add more details and ways for people to explore the information.

World Map -Disease

We used a world map, shown in , to see how different diseases spread across the globe in different years. We wanted to find out if there are patterns in how diseases spread over time. The map includes diseases like Measles, Hepatitis B, Polio, HIV/AIDS, and Diphtheria.

Our goal was to learn more about global health for different diseases and compare how they spread worldwide.

We expected to see changes in disease patterns over the years and to identify places with more or fewer cases. The data showed us that disease prevalence changes a lot over time, with different diseases affecting people in various regions. The maps helped us understand the differences in health between regions.

The world map worked well in showing us how diseases spread globally. It effectively communicated the story of how different diseases impact people over time and in different parts of the world. The colors on the map made it easy to understand. To make it even better, we could add more information about things like vaccination rates or healthcare and make it interactive so users can explore specific diseases or regions.
In summary, our world map successfully showed us how diseases spread globally over time. It met our expectations by revealing trends and differences in disease prevalence, providing valuable insights into global health patterns. Adding more details and an interactive feature could make it even better for users.

Bubble plot – Hepatitis / Alcohol

We used a bubble plot, to show how alcohol and Hepatitis are connected in different countries over the years. We chose this dynamic chart to show how alcohol and Hepatitis are linked.
Our plot met our expectations by showing how alcohol and Hepatitis immunization coverage are connected. To make it even better, we could change the Hepatitis immunization coverage to Hepatitis diseases. That would help us to see the connection between Hepatitis and Alcohol better. For now, we just see the Hepatitis immunization coverage and that makes it hard to understand the connection between alcohol and Hepatitis in general.
In summary, our dynamic chart successfully showed how alcohol and Hepatitis immunization coverage are linked across countries and years. Changing the Hepatitis immunization coverage to Hepatitis B diseases could make the plot even better.

Bubble plot – Alcohol / Adult Mortality

We used a special animated chart, shown in , to see how alcohol affects adult mortality in different countries each year. We wanted to know if there's a connection between alcohol and how long people live. The chart also helps us see changes in drinking habits over time.
As for what we learned, the chart doesn´t shows us patterns in how alcohol consumption might be linked to adult mortality in different countries. It doesn´t answered our questions about this connection.
But the chart still works well in telling the story of alcohol and adult mortality. There is just no connection. To make it even better, we could add more information about things like healthcare or money.
In summary, the animated bubble plot does a good job of telling the story of alcohol and adult mortality but there is no connection between these two variables.

Boxplot – Alcohol

We made another chart, a boxplot, to look at how a country's status is connected to alcohol consumption. You can see it in . We wanted to find out how much a country's status influences how much alcohol is consumed.
We looked closely at how much alcohol is consumed in different country statuses over many years.
In terms of what we learned, the chart showed us that there are differences in how much alcohol is consumed between Developed and Developing countries. It effectively answered our questions about the role of a country's status in alcohol consumption, providing insights into global drinking habits.
The chart works well in telling the story of alcohol consumption based on country status. To make it even better, we could add more information about things like money or healthcare.

In conclusion, the boxplot successfully revealed patterns in alcohol consumption based on country status. It met our expectations and is a useful tool for understanding the details of alcohol consumption trends in different types of countries.

Scatterplot – HIV/AIDS / Adult Mortality

We used Scatter Plots to look at how health indicators are connected in different countries. Thise plot focuses on HIV/AIDS and Mortality Probability (ages 15 to 60), showing trends. You can see this in Figure 7 .

In terms of what we expected, we wanted to learn about the connection between HIV/AIDS and mortality probability in different countries. The data shows a clear linear pattern, revealing the links between these health factors.

In summary, the scatter plot successfully shows how HIV/AIDS and mortality probability are connected in different countries. It met our expectations and is a useful tool for understanding global health trends.

Scatterplot – Hepatitis B / Death under 5

The plot in Figure 8 is about child health, looking at Death under the age of 5 and Hepatitis B immunization coverage.

Our goal was to understand how child deaths and Hepatitis B immunization coverage are connected worldwide. The data doesn´t show a clear patterns, indicating links between these health factors. The scatter plot answered our questions, that there is no connection.

The scatter plot works well to show the connection between child deaths and Hepatitis B immunization coverage. To make it better, we could think about adding more information, like about money or healthcare, for a deeper understanding. Also, getting feedback from people who use the chart could help us make it even easier to understand.

Bar chart – BMI / Adult Mortality

Our analysis also includes a bar chart for BMI and adult mortality shown in Figure 9. We seek to understand how BMI varies across countries and years and if there are notable differences. Additionally, we are curious if BMI influences adult mortality.

Our plot did not met our expectations. One reason for that is the dataset. The dataset, while complete, is lacking consistency and by times shows incorrect values of parameters. The inconsistency is e.g. observable for the population sizes which jump over the years as an error in the placement of the decimal point. The incorrect values are e.g. clearly visible in the bar chart where averaged BMIs of a country of sometimes under 2 or over 50 are observed. That is of course neither logical or possible.

Also the type of plot is not the best for visualizing this data. To make it better, another plot type could be used to show the connection between BMI and adult mortality. In Conclusion one can say, that this graph is not an example for a good data visualization.

Conclusion/Discussion

We managed to i) show trends, ii) see developments over the years and iii) differentiate between the health status of different regions worldwide. Firstly, the trends that we were able to show were e.g. a constant increase in immunization coverage worldwide from the year 2000 until the year 2014. These trends were observable over different parameters as the colors were unified over each plot. So, each color had its specific meaning, no matter which graph was being observed. Secondly, an achievement was being able to see these beforementioned trends as a temporal sequence, so over the years. This was possible by animating the plots over the years. Thirdly, health development differences between regions and continents were able to be seen. Whether a certain disease is prominent in one continent and might spread in the future to other regions of the world if necessary counteractions are not taken into account. A counteraction might e.g. be a higher vaccination rate in adjacent regions where a disease is currently spreading. These insights can be crucial and are aided by one of the features of the graphs. That is that each continent has its own specific color in every graph. As the colors do not change, it is very easy for the user to compare different aspects of health indicators across the globe and gain in-depth insights such as possible counteractions for specific disease spreads.

While having acquired valuable information on health and development indicators there were challenging aspects of this work. There is i) a lack of consistency and ii) incorrect values in the dataset. Fixing the inconsistency in e.g. the population sizes as an error in the placement of the decimal point proved to be a time-consuming task. However, incorrect values such as the BMI were not able to be corrected, as there was no data found on the correct BMI values of each country of the world over the observed period of 2000 until 2015.