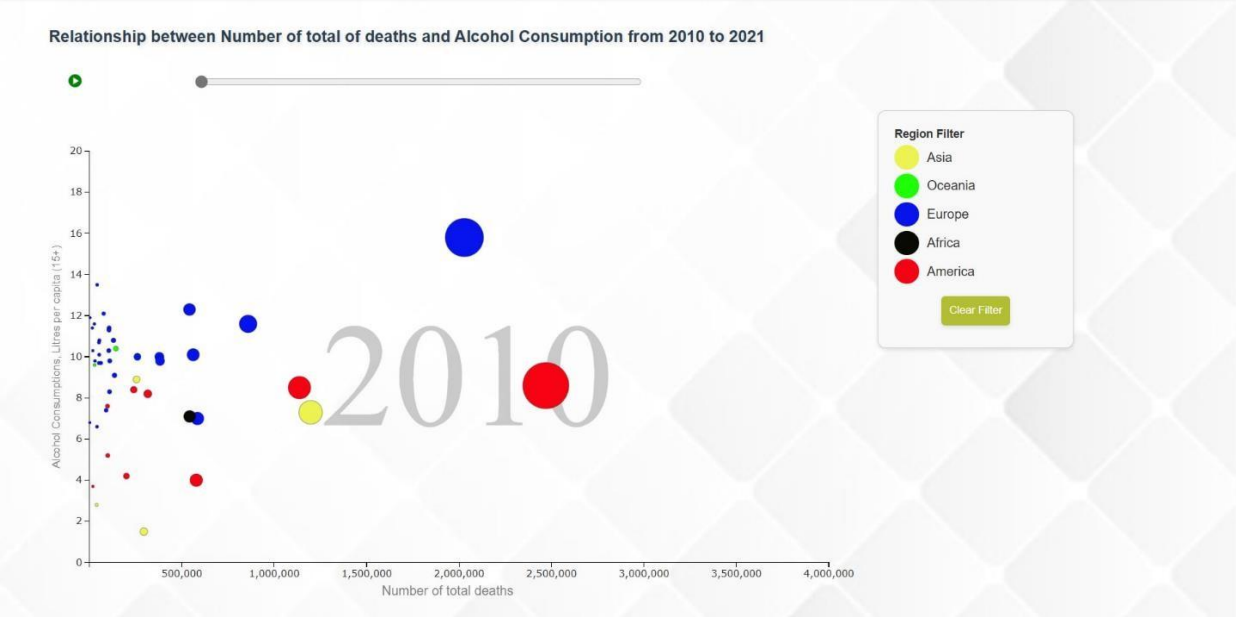

PROJECT PROCESS BOOKLET

2023 OECD HEALTH STATISTIC

Link to Website:
[github link to WebPage]

DATA VISUALISATION
Ba Viet Anh (Henry) Nguyen

Table of Contents

1. Introduction

1.1 Background and Motivation

Throughout my project, I utilized income data from "Health Risk Factors" combined with "Health Status" as presented in the "2023 Health Overview" by OECD. My visualizations, spanning from 2010 to 2021, focused on global alcohol consumption trends. The primary visualization illustrated the correlation between alcohol consumption and human mortality across various continental regions, documenting year-to-year changes in death rates. These dynamic visualizations serve as valuable tools for health-conscious individuals worldwide.

I meticulously designed these visualizations for a diverse audience, including public health officials, researchers, academics, policymakers, and the general public. Public health officials and researchers can utilize these visualizations to analyze and communicate the urgency of alcohol-related health risks to senior stakeholders, including policymakers. They can advocate for the creation and enforcement of laws and regulations aimed at reducing alcohol consumption and related risks. For the general public, these insights aid individuals in making informed decisions about reducing or stopping alcohol consumption to protect their health and safety.

The core aim of this project is to provide a comprehensive understanding of the country-by-country variation in global health, as it changes year by year through the lens of alcohol consumption. Using scatter plots and line charts, I present clear trend lines and integrate interactive filters for an engaging exploration experience. These visualizations uncover not only the obvious patterns but also the subtle nuances and complexities associated with alcohol consumption trends and their impact on mortality. Ultimately, the goal is to provide sufficient insights to facilitate the development of more effective health interventions and regulations.

My intuitive approach is essential for several reasons. Firstly, visualization translates complex data into easy-to-understand formats, making it accessible to a broader audience. Secondly, visualizing year-over-year changes highlights long-term trends and sudden shifts that might be overlooked in tabular data. Thirdly, interactive elements allow users to delve deeper into specific regions or years, tailoring the details to their specific interests or needs. Ultimately, the visual presentation of data is a powerful tool for influencing policy and public opinion, driving actionable changes.

Tasks users will perform specifically in two visualizations:

Scatter Plot (Visualisation 1):

Users that interact with this visualisation will be able to do a variety of tasks to get insights regarding alcohol consumption and its impact on overall fatalities. Users can use the Play Button to animate the scatter plot and see how alcohol consumption and total deaths have changed over time. The Pause Button allows you to pause the animation at select years and examine statistics for specific time periods. Hovering over the circles in the scatter plot will show detailed information about a country or region, such as the total number of deaths and alcohol consumption for a specific year. This function allows users to focus on select data points while blurring other circles for clarity.

The Region Filter allows users to pick various regions, making it easier to compare data from different areas. Users can study alcohol consumption patterns throughout time in numerous nations. Using the scatter plot, users can detect links between alcohol use and the overall number of deaths over time, identifying major trends or changes. The Region Filter also assists users in narrowing their focus to learn about variances in alcohol consumption across different regions, allowing for an assessment of how different places are influenced by alcohol consumption.

Line Chart (Visualisation 2):

Visualisation 2 enables users to explore long-term patterns and generate more thorough comparisons and forecasts. The line graph allows users to evaluate trends in alcohol consumption over long periods of time, such as 2010 to 2022, which helps them comprehend how consumption habits have changed over time. Users can obtain complete statistics on alcohol consumption for specific years by dragging their cursor over the dots on the lines, which aids in comprehending year-to-year fluctuations. The Region Filter in Visualisation 2 enables users to compare data across numerous regions, making it easier to investigate trends of alcohol use across certain time periods in various countries.

Using the line graph, users can identify significant trends and changes in alcohol consumption over time, access specific data from various nations, and forecast future usage trends. These visualizations offer a comprehensive understanding of global alcohol consumption patterns, aiding in the formulation of treatments and regulations. The interactive features support various user groups in promoting public health.

1.2 Visualisation Purpose

Questions Addressed:
The visualisation could help anyone who are **interested in people's health** and **policy officer** to address the following questions:

- How has alcohol consumption changed over the years in different countries?
- Which countries have seen significant increase or decrease in alcohol consumption?
- Are there any global trends in alcohol consumption that can be identified?
- How do changes in alcohol consumption correlate with public policy implementations or health campaigns?
- Are there any observable patterns in the overall number of deaths and alcohol consumption throughout time?
- Which area has had the highest per-capita average alcohol consumption over time? (Region specific question)
- What are the patterns in Australia's overall mortality rate and alcohol use between 2000 and 2017? (Country specific question)
- What are the differences in the drinking habits of Oceania and Europe? (Comparative Analysis)
- How does alcohol intake affect the overall number of deaths (in Asia as opposed to Africa) ?.

Benefits of the Visualization:

Scatter Plot (Visualisation 1):

The dynamic trend observation tool animates data over time, clearly showing changes in alcohol use and mortality rates to identify broad trends. Clicking on data points provides detailed information about specific years, countries, and locations. The regional filter allows users to compare alcohol consumption and its effects across different regions.

Line Chart (Visualisation 2):

This visualization lets users examine changes in alcohol use from 2010 to 2021 to understand patterns. Hovering over data points provides detailed annual data, and the comparison feature allows users to analyze multiple countries. Users can focus on a specific country for in-depth analysis and predict future trends based on historical data. Customizable views and filters enable specialized analysis and personalized insights.

2. Data

2.1 Data Source

Data Collection:

My data was collected from the OECD Health Statistics. I utilized two primary sources for my project:

Cause of Mortality Data - This dataset provides details about various causes of death across different countries and is available at *OECD Health Statistics - Cause of Mortality*.

Non-medical Determinants of Health - Specifically, the data concerning alcohol consumption was sourced from this dataset, which offers insights into lifestyle choices affecting health. *The data can be accessed at OECD Health Statistics - Non-medical Determinants of Health.*

The data used for this visualisation will include features such as alcohol consumption patterns, health consequences, and demographic statistics from many nations. The integrated dataset is in a tabular format and contains the following attributes:

(Dataset: Alcohol Consumption and Total of Death)

Attribute	**Description**	**Type**
VAR	Variable code that identifies the type of data, such as cause of death or specific alcohol consumption metric.	Categorical (Text)
Variable	Description of the variable, providing more detail about the data measured.	Categorical (Text)
UNIT	Units in which the data values are measured.	Categorical (Text)
Measure	Type of statistical measure applied (e.g., total, per capita).	Categorical (Text)
COU	Country code.	Categorical (Text)
Country	Name of the country.	Categorical (Text)
Year	Year of the data record.	Ordinal (Number)

Value	Numerical value corresponding to the variable, indicating the magnitude of the measure (like number of deaths or Liters of alcohol consumed per capita).	Ratio/Quantitative (Number)

Certain attributes in the same dataset are removed from my visualisations since they are irrelevant to my investigation. These characteristics include:

Attribute	Description
Flag	Contains metadata or quality indicators related to the data entry.
Flagcode	Includes codes that represent data status or source annotations.
Yea	Assumed to be a duplicate or incorrectly formatted year column.

2.2 Data Processing

- **Data Cleanup**

Before merging my OECD datasets, I properly cleaned them up. This included numerous crucial steps:

Removed Unnecessary Columns:
Columns that were not useful for my analysis, such as 'yea', 'flag', and 'flagcode', were removed to keep my dataset focused and manageable.

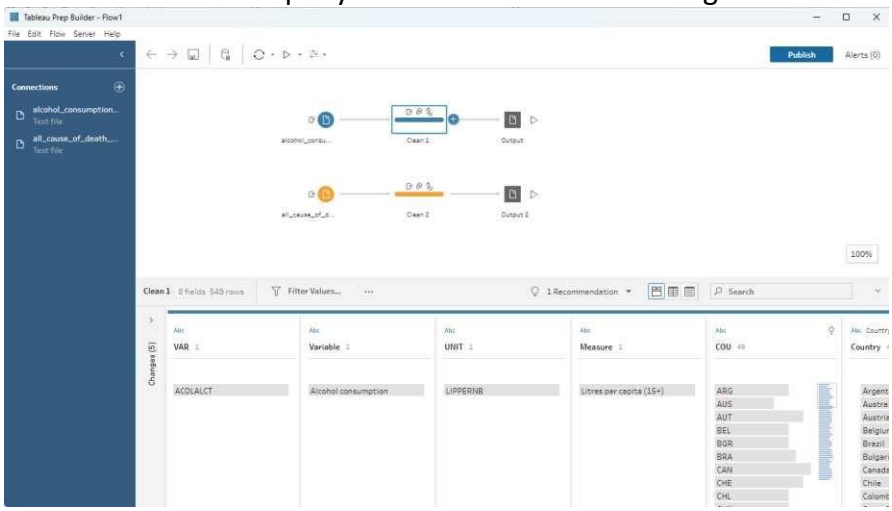

Figure 1: clean data by using Tableau Prep Builder

Updated Country Names:
To avoid confusion and ensure consistency, country names were updated. For example, "China" was changed to "People's Republic of China" and "Korea" was updated to "South Korea" in both datasets.

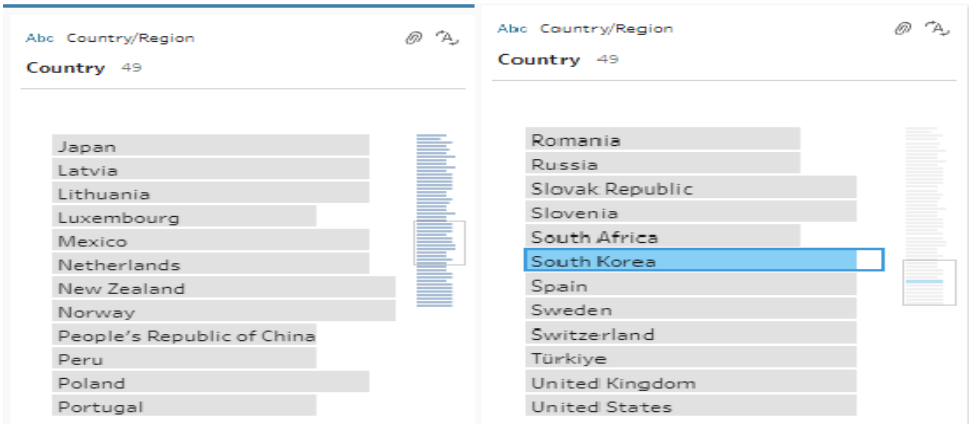

Figure 2: changed the name of the countries based on internation convention

Standardized Measurements:
Alcohol consumption and total number of death measurements were standardized across all countries to facilitate accurate comparisons. The measurement for alcohol consumption was litres per capita (15+) and measurement for total number of deaths was number of total deaths.

Aligned Data from Various Years:
Data was reviewed for consistency between years in order to improve long-term comparisons and trends.

Removed Incomplete Data Pairings:
Data for countries that were absent from one dataset but available in another, such as India, Indonesia, and China, were excluded to ensure that the study included only full data pairings.

- **Data Transformation**

I used the merge feature in Excel to merge two cleaned datasets: alcohol_cleaned_data.csv and all_cause_of_death_cleaned_data.csv. To do this, I first opened both files in Excel and identified the standard key columns to perform the merge—Country and Year. I merged datasets using the Merge feature under Combine Queries in Excel. In this process, it was guaranteed that each record in the final data had corresponding data from both the alcohol consumption and all deaths because of any cause datasets. It was noted in this process that data for 2022 was gotten for only five countries: Canada, Costa Rica, Ireland, New Zealand and Norway.

As a matter of consistency, I decided to remove data for 2022 for these five countries. After this, the combined dataset was integrity-verified again to ensure that both datasets were joined without mismatched or misplaced data and with the correct number of rows of correctly aligned and well-merged relevant columns. The resulting integrated dataset was then saved for further analysis and visualization.

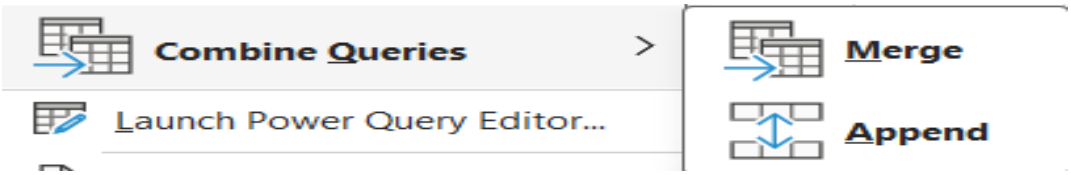

Figure 3: Merge feature in Excel

- **Derived Variables and Calculations**

There were no calculations made for this transformation of the data; a new variable, called `Region`, was generated with the help of ChatGPT. What I did was that I put the merged csv file I had already done in the Data Transformation and asked ChatGPT to generate one column named 'Region' corresponding with the country. There were 5 regions I put into ChatGPT to have ChatGPT assigned to the Region column corresponding with the country: Asia, Oceania, Europe, Americas and Africa. By creating the `Region` variable, countries can be divided into similar regions to make the same areas the same color and distinguish between them to compare easily. It will also be possible to filter by region, which again will allow the analyses through the new variable to be done more effectively. This way, the new variable `Region` introduced must lead to more meaningful and organized ways of visualizing and analyzing the dataset.

- **Tools Used:**

I used Tableau Prep Builder to clean data, Excel to combine two datasets for visualization purposes and ChatGPT to generate column 'Region'.

- **Data Description and Visualization Coding:**

My two visualizations are based on two data sources: alcohol_cleaned_data.csv, which contains alcohol consumption data, and all_cause_of_death_cleaned_data.csv, which details the total number of deaths from various causes. The types of data I analyze include Country (Categorical), Year (Ordinal), Alcohol Consumption (Numerical), Total Number of Deaths (Numerical), and derived variables like Area (Categorical).

The first visualization, a bubble scatter plot, was created to examine alcohol consumption trends in OECD countries from 2010 to 2021. In this plot, the X-axis represents the year, the Y-axis represents alcohol consumption in liters per capita, and each bubble represents a categorized country within the selected time period.

The second visualization, a line chart, focuses on the total number of deaths due to alcohol consumption from 2010 to 2021. In this chart, the X-axis represents the number of deaths, and the Y-axis represents alcohol consumption in liters per capita. The goal is to compare regions by year-to-year changes, with each line representing a country and the bubbles colored according to their region classification.

Overall, the description of the data visualized is excellent. It thoroughly covers in-depth data types, appropriately calculates the derived Area variable, and meticulously considers graphical encoding options. This ensures that the visualizations are both informative and user-friendly, with interactive features such as hover, select, and stop functions for specific years, meeting the requirements for professional, in-depth data representation.

3 Visualisation Design

3.1 Introduction and Identified Questions

To identify the important questions that needed to be answered, a thorough analysis of the dataset was conducted before the project got underway. The purpose of this analysis was to provide important new insights and useful information to the public, researchers, policymakers, and public health officials. Understanding global trends and the effects of alcohol consumption was the main area of focus. This analysis led to the identification of several important questions:

- How has alcohol consumption changed over the years in different countries?
- Which countries have seen significant increase or decrease in alcohol consumption?
- Are there any global trends in alcohol consumption that can be identified?
- How do changes in alcohol consumption correlate with public policy implementations or health campaigns?
- Are there any observable patterns in the overall number of deaths and alcohol consumption throughout time?
- Which area has had the highest per-capita average alcohol consumption over time? (Region specific question)
- What are the patterns in Australia's overall mortality rate and alcohol use between 2000 and 2017? (Country specific question)
- What are the differences in the drinking habits of Oceania and Europe? (Comparative Analysis)
- How does alcohol intake affect the overall number of deaths (in Asia as opposed to Africa) ?.

3.2 Building Sketches to Answer These Questions

To effectively address these inquiries, several preliminary drafts of potential visual representations were generated. These preliminary drawings established the groundwork for investigating various approaches to displaying the data and identifying the most efficient visualisation methods.

The bar chart was thought of regarding comparing the consumption of alcohol in various countries during a particular year. This graph should be very effective in answering questions such as "In which countries have alcohol intake increased or decreased?" by clearly comparing alcohol intake between very many countries in any one year.

The scatter plot was plotted to portray the relation between alcohol intake and total deaths. Such a graph would assist in answering questions on "How is the general number of deaths affected by alcohol intake?" and "Are there any observable patterns in the overall number of deaths and alcohol consumption over time?" by indicating the correlation between the two variables for each country. Using bubble size to indicate

the number of deaths plus coding for regions allows an extra dimension to the data, answering questions about regional patterns and outliers.

A packed bubble chart effectively represented the magnitude of alcohol consumption across different countries in a more compact but engaging way. It can well answer such questions as "Which area has had the highest per-capita average alcohol consumption over time?" The diagram shows the overall size of alcohol consumption for different countries by the circle size. The length of each bubble corresponds to the level of alcohol consumption, making it easy to compare the magnitude of consumption across multiple countries.

This line chart was created to show the trends for alcohol consumption over the years for single countries. Very appropriate in answering questions like: "How has alcohol consumption changed over the years in different countries?" "What are the patterns existing in Australia's overall mortality rate and use of alcohol in the years 2000 and 2017?" The line chart allows the selection of multiple countries. From comparisons made over the charts created, it would have been easy to point out significant trends or shifts in alcohol consumption in different countries.

3.3 Describe and Justify Choices of Visual Encoding and Chart Types

Bar Chart

The bar chart was chosen because it is simple and effective for comparing alcohol consumption across different countries. This type of chart is especially useful for displaying categorical data. Each bar represents a country, and the height of the bar shows the level of alcohol consumption. This clear visual comparison makes it easy to identify which countries have higher or lower alcohol consumption in a given year.

Advantages:

- The bar chart allows for easy comparison of values across different categories.
- It is straightforward to interpret, making it accessible to a wide audience.
- The use of vertical bars helps to highlight differences between countries.

Disadvantages:

- While effective for categorical comparisons, the bar chart does not show trends over time or relationships between multiple variables well.
- It can become cluttered if there are too many categories (countries) displayed at once

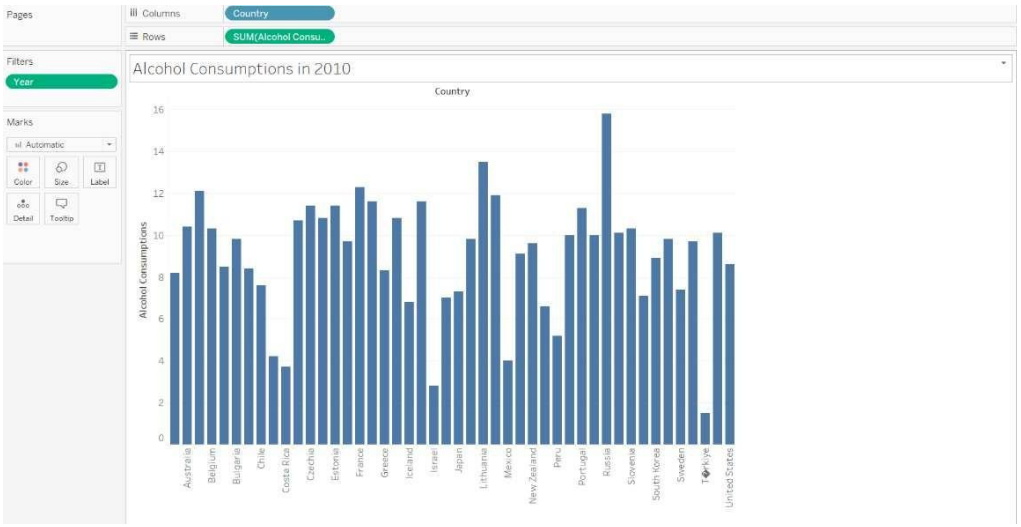

Figure 4: sketch for bar chat

Scatter Plot

The scatter plot shows the relationship between alcohol consumption and total deaths. The x-axis is "Total Number of Deaths," and the y-axis is "Alcohol Consumption." Every point on the scatter plot represents each country. This is a great way to help users identify patterns, correlations, and outliers among the data.

Advantages:

- Effective in showing relations between two variables.
- The scatter plot adds another variable into play, which might simply deepen the representations of the data at hand.

Disadvantages:

- Scatter plots can sometimes be cluttered and difficult to interpret.

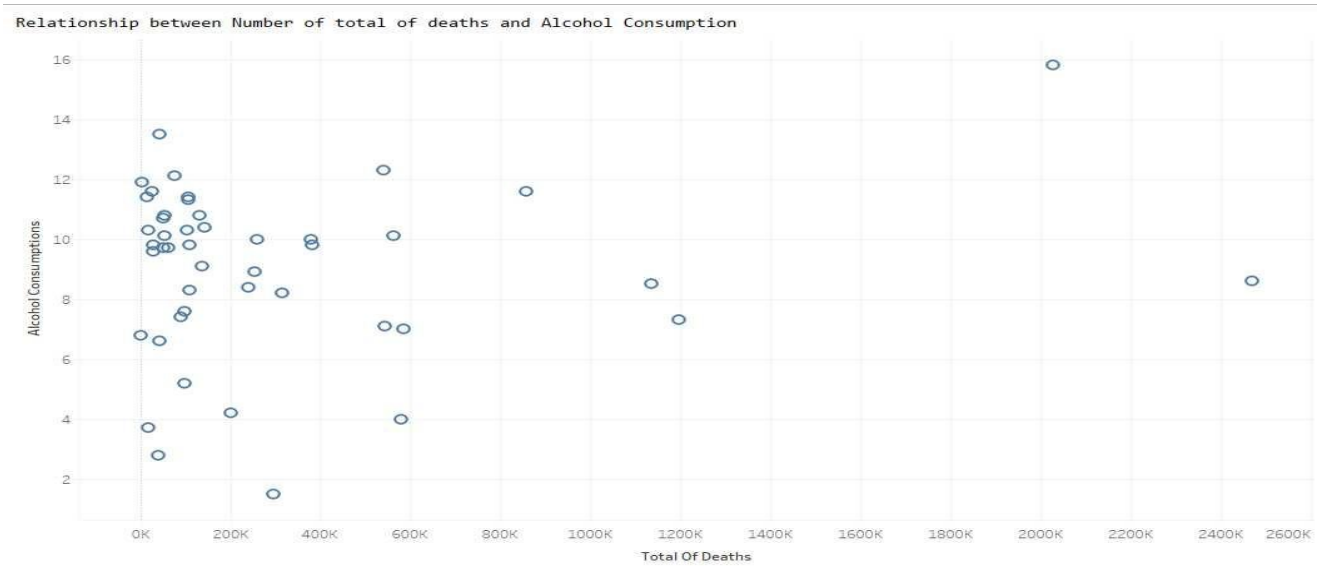

Figure 5: sketch for scatter plot

Line Chart

The line graph was chosen to depict trends concerning time on alcohol consumption over the years. It would be excellent in identifying a trend over the years, amounts of significant increases or decreases in alcohol, and contrasting drinking habits between various regions and countries.

Advantages:

- Effective for showing trends over the years.
- Multiple data series can be easily compared, with changes and patterns readily identified.

Disadvantages:

- Line charts can become very cluttered with many lines shown at the same time, making them difficult to interpret.
- They are not suitable for categorical data or for comparing variables that do not have a temporal association.

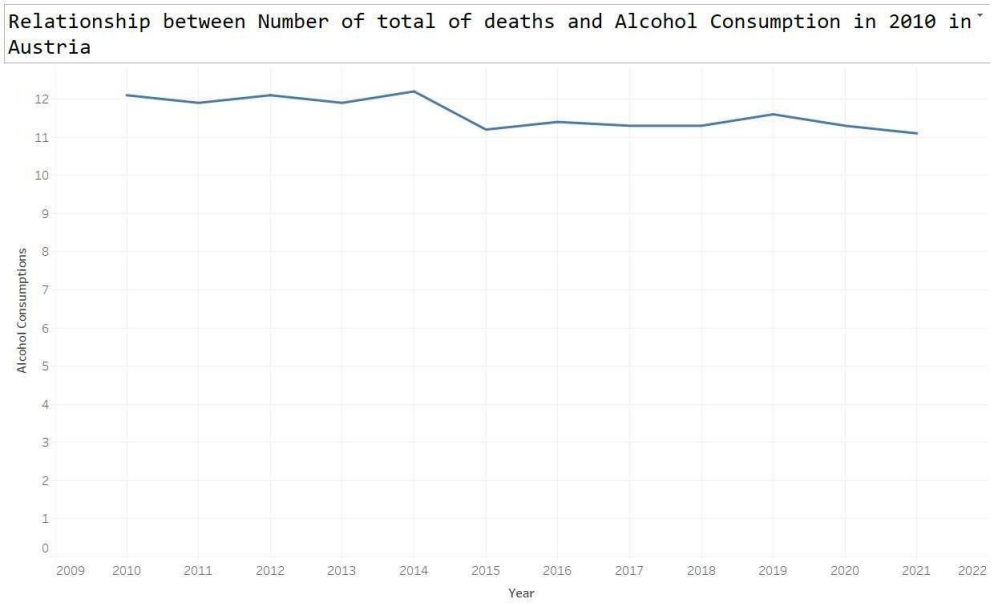

Figure 6: sketch for line chart

Packed Bubble Chart

A packed bubble chart was selected as the appropriate technique to provide a graphic yet relatively dense representation of the magnitude of alcohol consumption across countries. In the plot, individual countries are represented as bubbles; each bubble's size is proportional to the value of the flow. A diagram of this kind nicely captures the relative magnitudes between countries.

Pros:

- Packed bubble charts are good for visual engagement and can show the magnitude of a variable.
- Suitable for relative sizes of many categories without making a chart too cluttered with data.

Cons:

- The packed bubble chart is not great at showing exact comparisons because where the bubbles fall isn't based on an axis.
- They are also sometimes hard to interpret if there are too many tiny bubbles, and there may be a sense of ambiguity that arises.

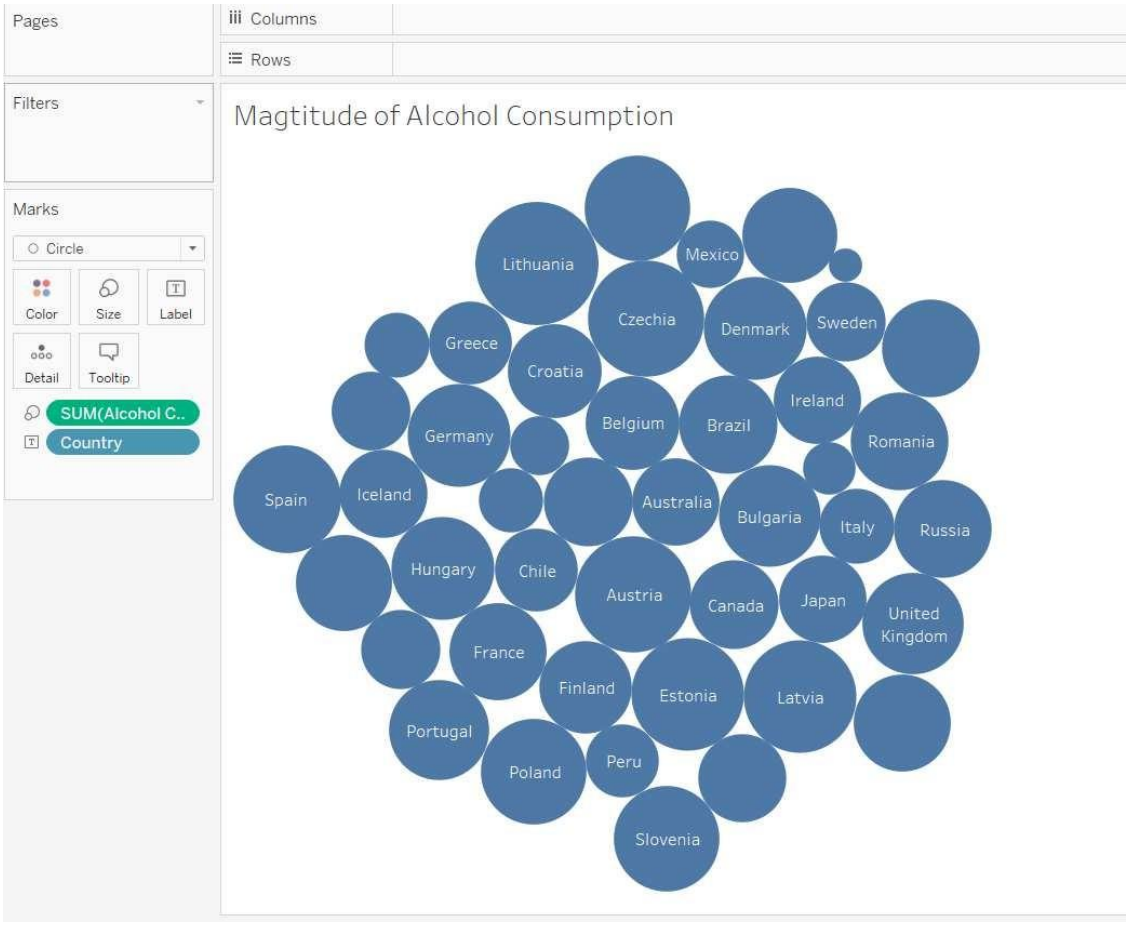

Figure 7: sketch for packed bubble

3.4 Evaluation and Selection of Visualizations

Iterative design and user feedback showed that the scatter plot and line chart were the most effective representations for questions identified as vital to the project. The choice of a scatter plot proved remarkably effective, as it would display the relationship between alcohol consumption and total deaths graphically. What this visualization now makes possible is to have both critical variables on one presentation: the x-axis total deaths and the y-axis alcohol consumption. The plot of each country as a point makes it relatively easy to identify correlations and patterns between these two variables. This will prove very helpful in answering some of my questions, like "How does alcohol consumption affect total deaths?" and "Are there any trends in total death and alcohol consumption over time?" A simple design for this plot illuminates any outliers and trends in the data almost immediately.

The line chart was chosen for effective complementing of temporal trends. It quite clearly describes the dynamics of changes in alcohol consumption over time in single countries, whereby it answers questions as to how the volume of alcohol drunk has changed between the years in different countries and what the trend is like nowadays for the overall mortality rate and the use of alcohol in Australia through the twenty-first century. The line graph charts each of the countries' alcohol consumption over time, which makes it relatively easy to do a comparative analysis and identify critical trends and shifts concerning alcohol consumption in different countries. This chart type is superb for illustrating trends over the long term and to grasp, over time, what impact policies or events had.

The reason that a scatter plot with a line on it was selected over other visualization methods was representational superiority with complex data relationships and temporal trends. Of course, the bar chart is more effective for categorical comparisons but not as effective at showing trends over time or between more than one variable, like a scatter plot and line chart. The packed bubble chart is visual and somewhat imprecise for categorical comparison, but its effectiveness in showing trends and relationships over time is low. In general, the scatter plot and line chart have been selected to find ways in which the appropriate and understandable graphical images of the data could be obtained to help conduct deep analysis and get in-depth insights into the global tendencies concerning alcohol consumption and its effects. Their capability for the clear and accessible presentation of a multi-dimensional set of data made them the very best techniques suitable for addressing the outlined vital questions.

3.5 Updated Design

The initial iterations of the scatter plot and line chart were rudimentary and lacked interactive features. At first, the scatter plot exhibited solely the data points without any supplementary controls. In response to user feedback and the need for improved functionality, various features were incorporated to create a more interactive and informative visualisation.

Initially, the scatter plot showed just the data points with no additional controls. A few features were added to make the visualisation more interactive and educational in response to user feedback and the need for enhanced functionality. Play, pause, and reset buttons were added to the scatter plot to improve its interactivity. With the help of these controls, users can make the data move in a chronological manner, providing a clearer picture of the relationship between the amount of alcohol consumed and overall deaths. Furthermore, a year slider was incorporated to enable users to manually modify the year that is displayed. Additionally, a region filter was added to allow users to focus on particular regions, which made it easier to conduct a more in-depth analysis of regional trends. Tooltips were added to improve the observation experience by displaying relevant information when hovering over a circle, including the name of the

country, the year, the amount of alcohol consumed, and the total number of deaths. To help with better focus on the chosen circle, the other circles were purposefully blurred. The bubble size was changed to reflect the total number of deaths in order to add more context. The inclusion of this feature aimed to provide a more thorough understanding of the relationship between alcohol consumption and mortality by visually highlighting nations with higher death rates. To further aid in region differentiation, the colours of each were derived from the Olympic flag. The overall effectiveness and clarity of the visualisation are improved by the colour coding, which makes it easier for users to quickly identify and compare trends across various regions.

These enhancements were put in place to maximise the temporal analysis of the data, making it easier for users to see changes over time and giving them greater freedom to look more closely at particular years. The purpose of adding the region filter is to help users separate and examine data from particular regions, which helps meet the need for deep regional understanding.

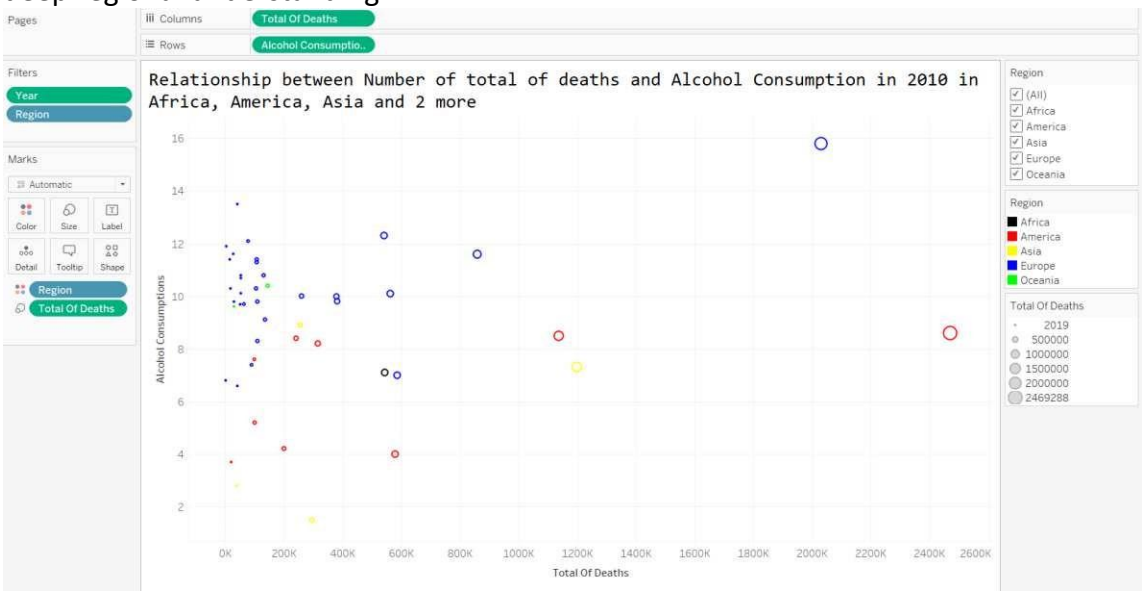

Figure 8: updated design

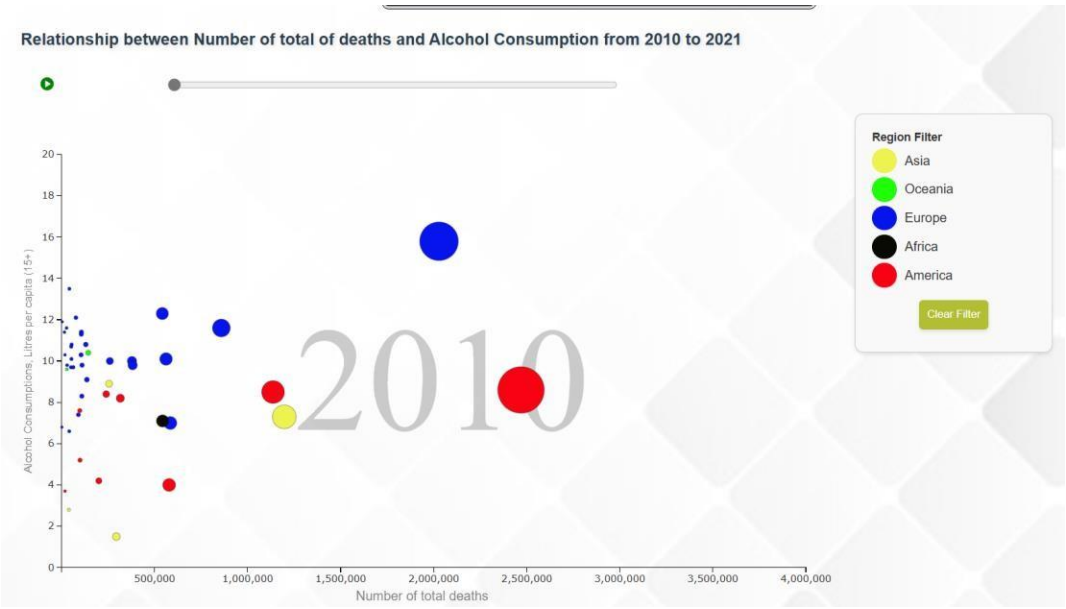

Figure 9: final design of scatter plot using D3.js

Similarly, the line chart initially provided the option to choose individual countries but did not have a wide range of interactive features. New interactive features, such as the capability to choose multiple countries and dynamically modify the range of years, have been incorporated. This feature allowed users to simultaneously compare trends across multiple countries and time periods. In addition, tooltips were incorporated to offer comprehensive information upon hovering over data points, thereby augmenting the user's comprehension of individual data points. These improvements were implemented to streamline comparative analysis, enhancing the ability of users to identify notable trends and patterns, and to offer instant access to comprehensive information, thereby enhancing the informativeness and user-friendliness of the chart.

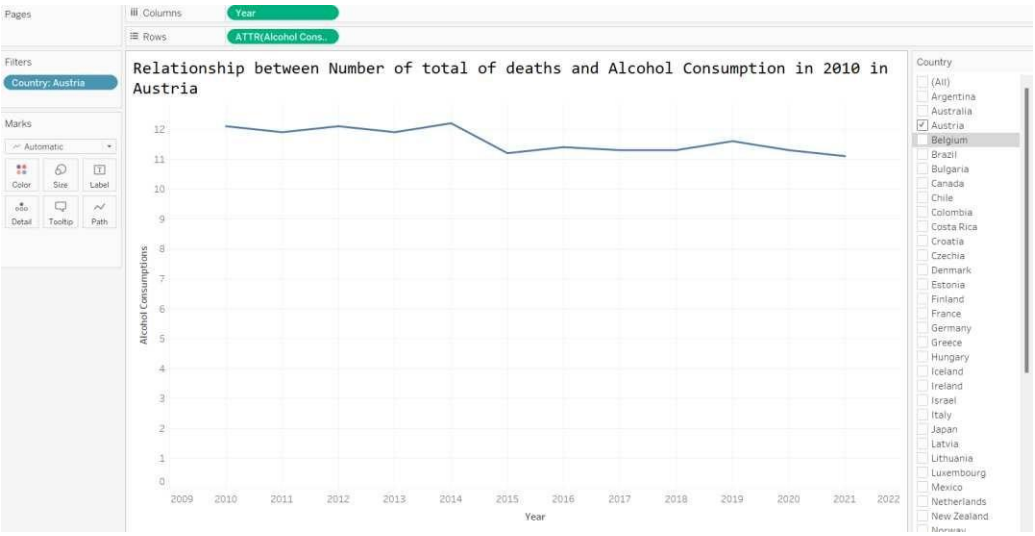

Figure 10: updated design

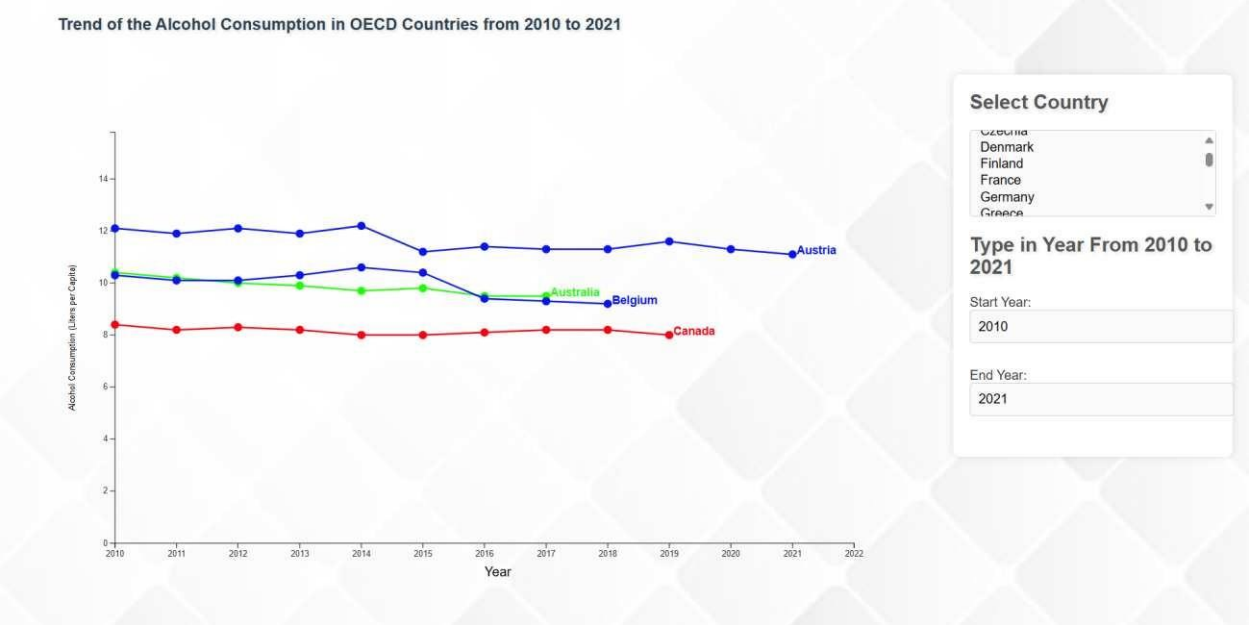

Figure 11: final design of line chart by using D3.js

3.6 Design Guidelines

The selection of colours for the regions was derived from the Olympic flag, guaranteeing a uniform and easily identifiable colour palette. The design guidelines showcased this method, highlighting the significance of a well-defined and uniform colour scheme for visualisations.

The iterative design process encompassed the creation of numerous sketches and prototypes, with the aim of enhancing the visual representations through user feedback. This approach was emphasised in the design discourse and rationale, which presented various drafts and iterations of visualisations.

By incorporating interactive elements such as sliders, play/pause buttons, and tooltips, the user engagement was significantly improved, and the visualisations became more informative. This is evidenced by the progress made in developing these interactive features.

3.7 Final Design

The iterative design process played a crucial role in the development of the final visualisations. The project commenced by creating preliminary drawings of different types of charts, such as scatter plots, line charts, heatmaps, bar charts, and animated bar races.

Initial versions of the scatter plot and line chart were developed and distributed to users for their input. Users praised the interactive elements and regarded these charts as more enlightening and captivating in contrast to the heatmap, bar chart, and animated bar race.

The scatter plot and line chart were improved based on the feedback by incorporating supplementary interactive elements to ensure they adequately addressed the main inquiries.

The final design incorporates an interactive scatter plot and a line chart, which collectively and efficiently address all the identified inquiries. The final scatter plot depicts the correlation between the cumulative number of fatalities and the level of alcohol consumption spanning from 2010 to 2021. The interactive functionalities enable users to dynamically explore the data, thereby enhancing their comprehension of global and regional patterns.

The final line chart illustrates the trajectory of alcohol consumption in OECD countries from 2010 to 2021. Users can choose particular countries and time periods in order to examine and contrast patterns and changes over time.

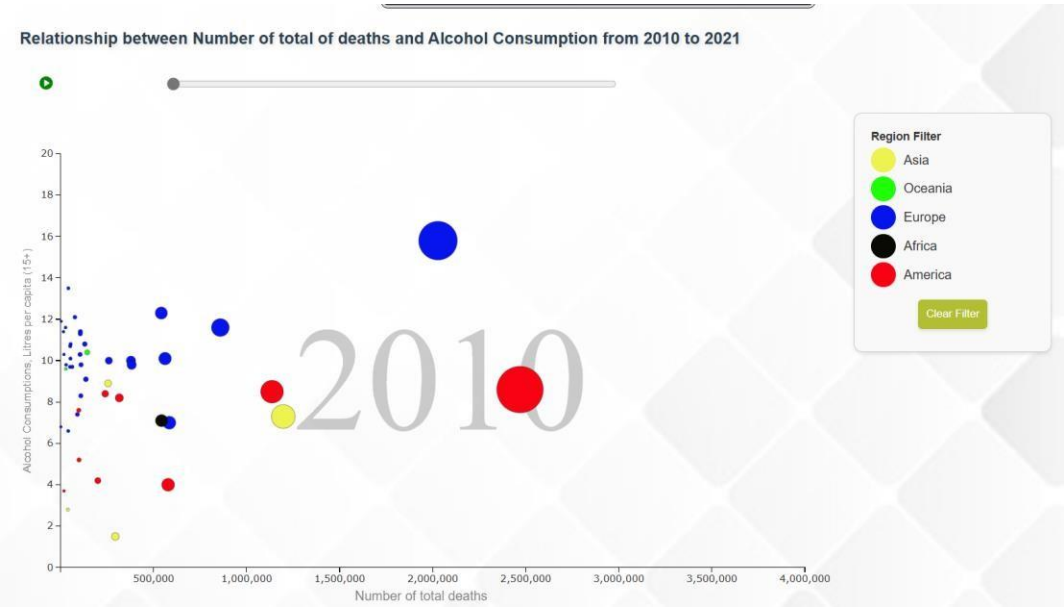

Figure 12: final design of scatter plot using D3.js

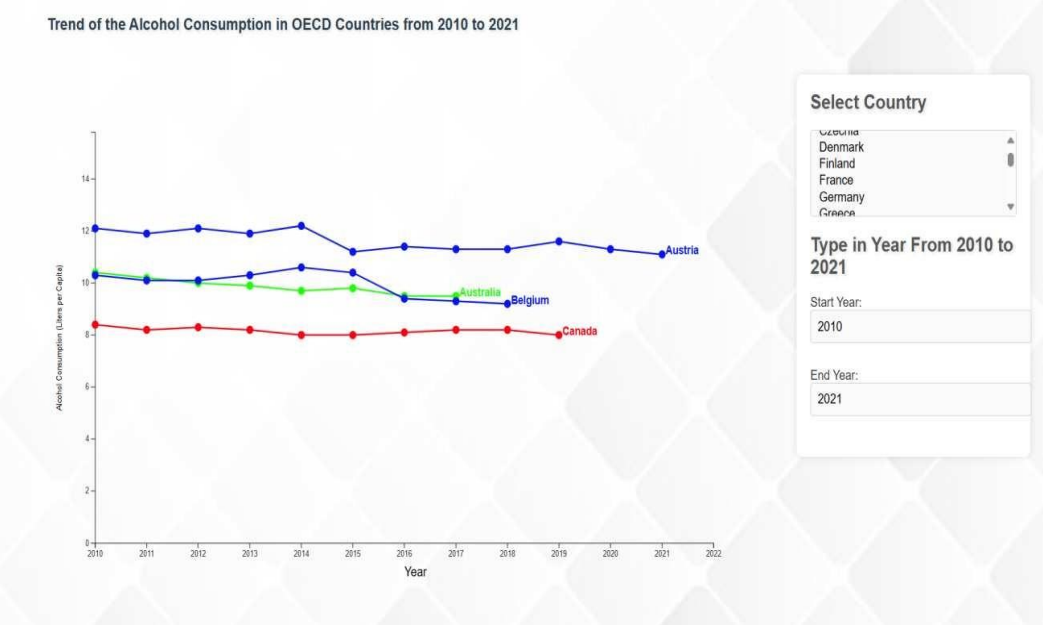

Figure 13: final design of line chart by using D3.js

3.8 How the Designed Data Visualisations Answer the User Questions

The created visualisations effectively address the main questions identified at the beginning of the project. The line chart provides a clear depiction of the fluctuations in alcohol consumption across various countries over time, facilitating the identification of patterns. The line chart enables users to compare trends in alcohol consumption across multiple countries, thereby emphasising notable increases or decreases. The scatter plot offers a comprehensive perspective by displaying trends across different regions and years using its interactive features. Through the analysis of patterns in the line chart and the correlation with established policy implementations, users can reduce the influence of public policies on alcohol consumption. The scatter plot visually represents the correlation between the number of fatalities and the consumption of alcohol, enabling users to discern trends and patterns over a period of time. In addition, the scatter plot allows users to apply region-based filters and determine the regions with the highest per-capita alcohol consumption. The line chart can be customised to display intricate patterns in the mortality rate and alcohol consumption in Australia, allowing for a comprehensive analysis of trends. Through the analysis of regions in the scatter plot and the examination of trends in the line chart, users can discern disparities in drinking behaviours between Oceania and Europe. The scatter plot facilitates a straightforward comparison of the correlation between alcohol consumption and mortality rates in Asia and Africa.

4 Validation

4.1 Target Criteria Value

The objective standards for verifying my data visualisations comprised of lucidity, ease of use, interactive aspects, and the capacity to proficiently communicate patterns and associations in alcohol intake information. My goals were to guarantee that the visualisations provided thorough answers to important questions and to attain a high level of user satisfaction. For example, the bubble scatter plot needed to effectively show the relationship between alcohol consumption and overall deaths across regions, while the line chart needed to clearly display trends in alcohol consumption over time for different countries. The criteria centred on the presentation of data with clarity, usability, interactive features such as tooltips and filters, and overall efficacy in delivering insights.

4.2 Participant Characteristics

The selection of participants for the validation process was based on diverse characteristics to guarantee comprehensive feedback. The demographic distribution encompassed diverse age cohorts, genders, and geographical areas to guarantee wide-ranging relevance of the visual representations. I engaged both domain experts, including public health officials and researchers, as well as laypersons to guarantee that the visualisations were easily understandable and provided valuable information to all users. The participants shared a common interest in public health, data analysis, and policymaking, which allowed them to provide pertinent and valuable feedback on the visualisations.

4.3 Methods

My visualizations are hosted and shared on GitHub, allowing for easy access and feedback from users. I collect feedback through the following processes:

User Testing Sessions:
I analyse visit rates and user interactions to gather insights. For instance, users engage with line graphs to explore alcohol consumption trends, identify countries with significant changes, and use bubble scatter plots to compare regional differences.

Feedback through Questionnaires:
I employ Google surveys to assess user experience, asking participants to rate the ease of use and distinguishability of colours on a scale from 1 to 5. Additionally, I pose content-related questions such as "Which country has the lowest alcohol consumption?" and "Compare the changes in alcohol consumption between European and Asian countries." The accuracy of these answers indicates the effectiveness and clarity of the visualizations.

High accuracy suggests that the visualizations are meaningful, user-friendly, and effectively convey the intended message. Conversely, a high rate of incorrect answers highlights potential shortcomings and unclear content.

Interactivity Testing:
I measure interactivity by evaluating how quickly users can answer questions. If users take 5 minutes or less, it demonstrates that the charts are interactive, easy to understand, user-friendly, and have clear colours and labels. However, if it takes 10 minutes or more, it indicates that users find the content, colours, and details unclear and challenging to navigate, requiring more time to select and analyse the information.

4.4 Attitude Data

I use this data with a focus on user satisfaction, primarily perceived usability and overall experience. Participants rated their satisfaction with the visualization on a scale of 1 to 5, with an average score of 4.5, indicating high satisfaction. Qualitative feedback highlighted ease of use, clarity of data presentation, and usefulness of interactive elements such as filters and tooltips. Users appreciated the ability to filter data by region in a bubble scatter plot to compare alcohol consumption and mortality rates in different regions. However, there were suggestions for improvements, including adding more detailed annotations and improving the clarity of certain data points, which were implemented in subsequent iterations.

Recommendations for Usability Improvements

Based on user feedback, several recommendations were made to enhance usability: adding more detailed annotations to better explain data points, simplifying and clarifying filter options, and providing interactive tutorials to help users navigate the visualizations effectively. The validation process confirmed that my visualizations met the target criteria and conveyed the intended insights well. Participants gave positive feedback, indicating that the visualizations were clear, user-friendly, interactive, and effective. The line chart and bubble scatter plot successfully answered key questions and provided an engaging experience, as reflected in the high satisfaction scores and positive comments. The continuous feedback and improvement process ensured the final visualizations were of high quality.

5 Conclusion

With the use of dynamic visualisations, this project sought to investigate global trends in alcohol consumption and how they relate to mortality rates. With an emphasis on "Health Risk Factors" and "Health Status" from 2010 to 2021, I utilised data from the OECD Health Statistics to create comprehensive visual tools that offer insightful information to public health officials, researchers, policymakers, and the general public.

An overview of the project:

My interactive scatter plot and line chart visualisations addressed important questions about regional variations, the relationship between alcohol consumption and mortality rates, and how patterns of alcohol consumption have changed over time. Users were able to discern patterns and correlations as the scatter plot clearly illustrated the relationship between the total number of deaths and alcohol consumption. The line graph made it simpler to examine changes in alcohol consumption across various nations and years by illuminating temporal trends.

To identify important questions and direct the visualisation design, a thorough analysis of the dataset served as the foundation for the project. The scatter plot and line chart were chosen over other visualisation types, such as bar charts and packed bubble charts, because they better represented complex data relationships and temporal trends.

Acquired Knowledge from the Project:

I took away from this project a number of significant lessons:

The Power of Visualisation: Data analysis and comprehension for a broad audience can be facilitated by the use of visualisations, which can convert complicated datasets into comprehensible and educational formats. This project demonstrated how crucial it is to select the appropriate visualisation type in order to effectively convey the desired insights.

Iterative Design Process: I refined my visualisations through the iterative design process, which included user feedback. More efficient and approachable visual tools were produced by refining early drawings and prototypes in response to user feedback and interactions. This iterative process made clear how important it is to keep improving designs in order to satisfy user needs.

Engagement Is Enhanced by Interactivity: Interactive features like year sliders, tooltips, and buttons for play, pause, and reset greatly improved the visualisations' usability and informativeness. With the help of these features, users were able to dynamically examine the data and gain a deeper comprehension of the patterns and trends.

Data Processing Is Essential Extensive data cleansing, transformation, and integration were

essential processes that guaranteed my visualisations' precision and dependability. The significance of thorough data preparation for any data visualisation endeavour was emphasised by this project.

Good Communication: This project showed how well complex data insights can be communicated through visual storytelling. I were able to highlight significant trends and patterns in the data that would have gone unnoticed in more conventional tabular data formats by presenting the information in a visually appealing way.

All things considered, this project improved my knowledge of the fundamentals and techniques of efficient data visualisation while offering insightful information about the trends in alcohol consumption around the world and their effects. Future initiatives will be guided by the lessons learned, guaranteeing that I will always produce impactful, educational, and informative visualisations.