

Non-medical determinant of health factor: Alcohol and tobacco consumption

Semester 1 2024

Tsung Lee 103497841

Yong Yuan Chong 101224021

Completed by 7/June/2024

Total word count:

Contents

[Background and Motivation 3](#_Toc166868226)

[Visualization Purpose 3](#_Toc166868227)

[Schedule 4](#_Toc166868228)

[Data 5](#_Toc166868229)

[Data source 5](#_Toc166868230)

[Data processing 5](#_Toc166868231)

[Requirement 6](#_Toc166868232)

[Website required features 6](#_Toc166868233)

[Optional features 6](#_Toc166868234)

[Visualization design 7](#_Toc166868235)

[Initial idea designs 7](#_Toc166868236)

[Justification 7](#_Toc166868237)

[Conclusion 8](#_Toc166868238)

[References 9](#_Toc166868239)

# Background and Motivation

# In the modern era, tobacco and alcohol is one of the common methods to reduce peoples' tension and relieve their stress. As well as one of the great factors that guide as non-medical determinant death of the population. As large quantity growth on consumption, people may not still be aware of the impact of it. Hence, in this project, the motivation that guided the group was to present a clear visualization of the graph to further attract several groups such as policymakers, and health and medical carers. The visitor may utilize these graphs as a collection of references to gain and spread the potential health hazards of continuing to consume alcohol and tobacco. As well as engage the visitor to promote this reference to enhance social awareness of this potential hazard.

# Visualization Purpose

After the visitor views the data visualization graph of the trendline(line chart), the population of consumption between males and females (bar chart), and the proportion of both smoking and alcohol status (pie chart).

The visitors can understand the following questions:

-Comparison of the trend between time intervals.

-specific time point of where change starts.

-Comparative consumption levels of males versus females.

-Relative proportion of the status relating to smoking versus alcohol consumption.

Furthermore, the possible benefits of completed visualization include:

-Informing the stakeholders relative to a specific field about the impact on the health of the average person.

-Enhance public awareness about the consumption trend and health risks.

-Further study with the research of trend analysis and potential health risk.

# Schedule

Week 1 (Week 6)

-Team up, figure out the objectives of the project.

Week 2 (Week 7)

-Discussion of the topic to pick, start the process of background motivation, visualization purpose, team schedule, and Data source.

Week 3 (week 8)

-Group stand-up, start the coding with the implementation of D3 chart.

-Start the data processing.

Week 4 (Week 9)

-Continue with the implementation of the website which includes the d3 chart, animation, interactivity of the website.

Week 5 (Week 10)

-Complete the data processing.

-review the data and investigate how to find or method to fill the NA (Not available) value.

-Complete the structure of the website.

Week 6 (Week 11)

-Continue with constructing bar chart, pie chart, and line chart.

Week 7 (Week 12)

-Finish the website.

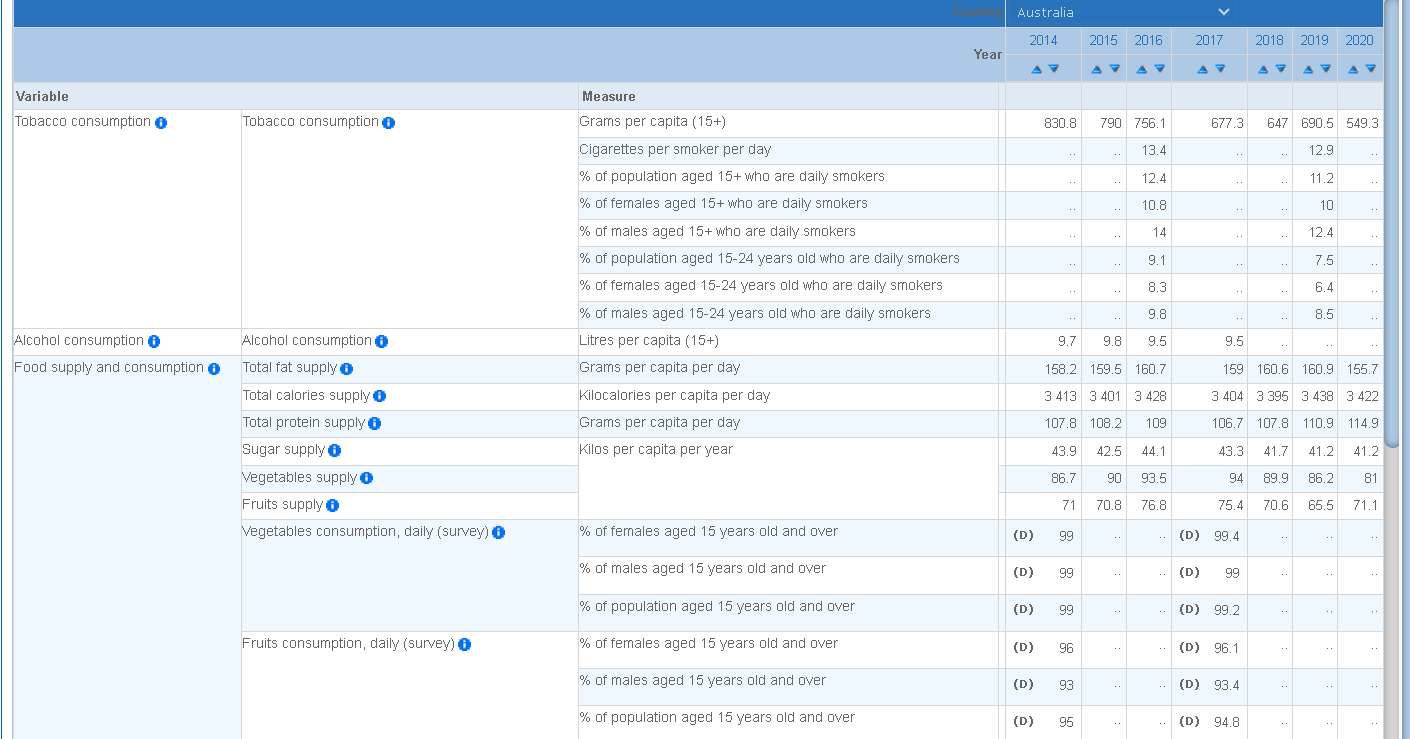
-Finish Process book.

# Data

## Data Source

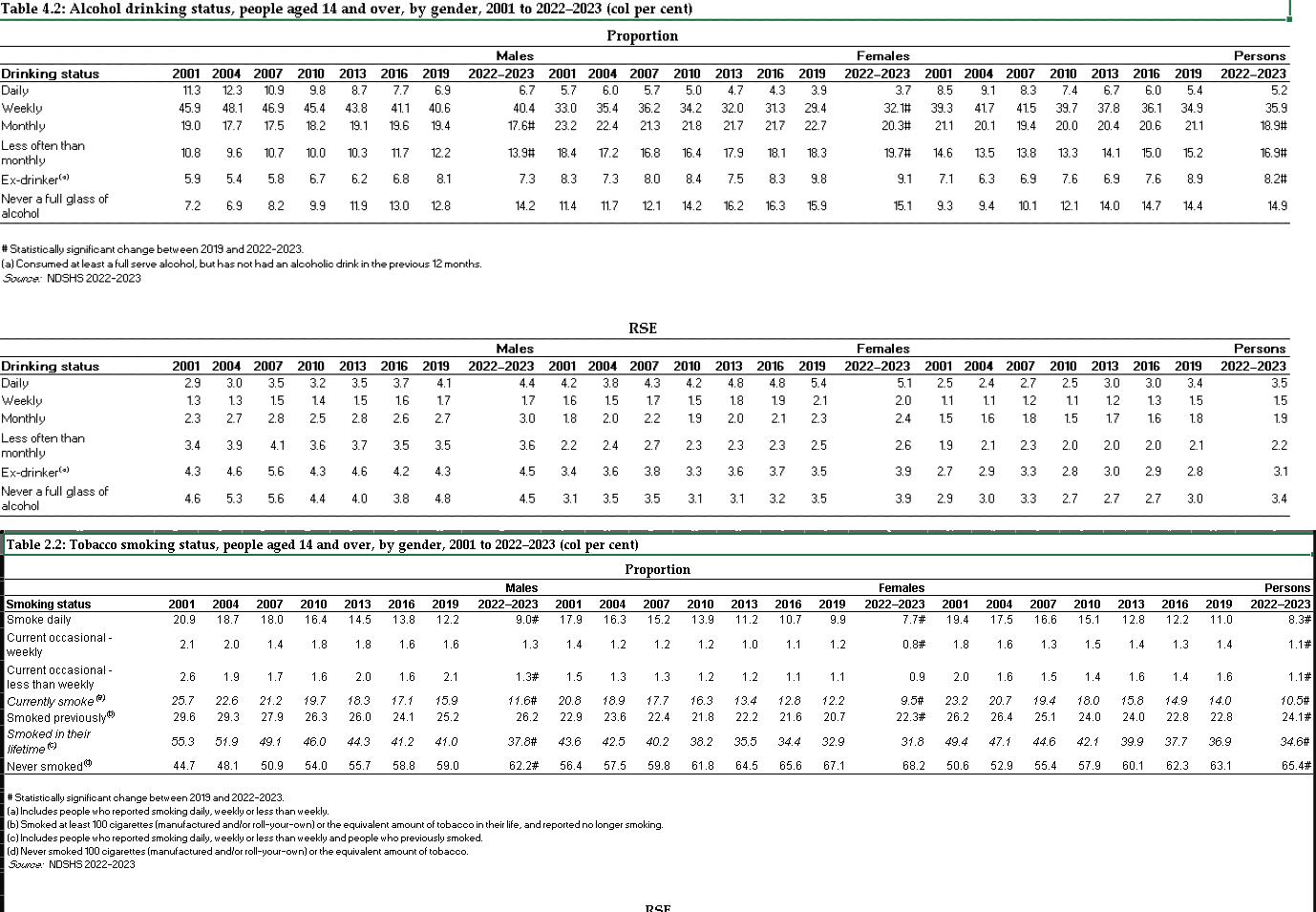
The project uses OECD (Organization for economic cooperation and development) statistics as a primary data source. The main data in the project is referred to fig.1:

<https://www.oecd-ilibrary.org/social-issues-migration-health/data/oecd-health-statistics/oecd-health-data-non-medical-determinants-of-health_data-00546-en?parentId=http%3A%2F%2Finstance.metastore.ingenta.com%2Fcontent%2Fcollection%2Fhealth-data-en>

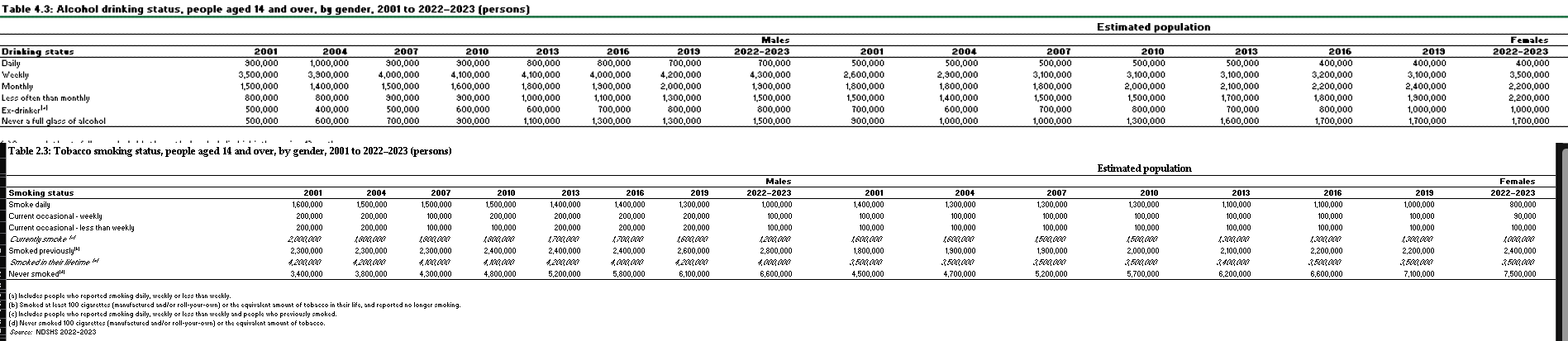


(Fig. 1 OECD non-medical determinant of health dataset)

From the source dataset, the teams’ main objectives are to present the visualization of tobacco, in other words, cigarette and alcohol consumption. However, when extracting data value to the dataset, huge numbers of data are missing (present as Not Available). Therefore, although by find the NA via mean, mode, or median value of the dataset could solve the issue. Yet the overall presentation of the data will not be precise nor accurate. Hence during the development of the project, the group used the OECD dataset as a reference and explored further resources relevant to the project objective. Ultimately, the team managed to find external dataset sources with richer content from the Australian Institute of Health and Welfare, known as AIHW an acronym shown in fig.2 and fig.3.



(fig2. AIHW dataset source of alcohol drinking and tobacco smoking status)



(fig.3 AIHW dataset source of population of Australia smoking and drinking status between male and female)

## Data Clean Up

During the data cleanup process, what is crucial is to remove unneeded data to keep the data table optimize for efficient coding. Thus, in the tobacco smoking status dataset from Fig. 2. The team decided to remove the ‘currently smoke’ and ‘smoked in their lifetime’ because it doesn't contribute to a clear visual presence. The team figures these kinds of 'combination' type data could drag down the preciseness of the data visualization as it could cause the individual data point to be uncertain as well as complicating the data integration.

The second chart focuses on the population proportion of the different statuses of smoking and alcohol consumption between males and females. The original dataset also includes the RSE (relative standard error) type of population. The team decided to remove the data between 2001 to 2013 to simplify the data visualization graph length and analysis. As well as remove the RSE type because it required to understanding of standard error and statistical concepts, which are not user-friendly to visitors after export to visualize the graph. In contrast with the proportion type data, it is more straightforward to understand as they present a percentage of the whole.

## Data Processing

### As the data visualization purpose mentioned, the team has a clear objective with constructing three following charts to provide clear visualization of the topic from each different characteristic.

### Line chart

* Method of collection:

The data is taken from OECD dataset. Data from the table is put collectively of numerical data used for the visualization, which is then presented through visualization in numerical (ordinal)and numerical (ratio) data.

### Pie chart

* Method of collection

The data is taken from AIHW spreadsheet table 2.2 and 4.2. Data from the table is composed with categorical/numerical data used for the visualization, which is then presented through visualization in through categorical (nominal)and numerical (ratio) data.

### Bar chart

* Method of collection

The data is taken from AIHW spreadsheet table 2.3 and 4.3. data. Data from the table is put collectively of categorical used for the visualization.

## Filling NA (Not Available) Value Gap

According from fig.4 the data set from the OECD graph, the missing value starts occurring in the later year. Therefore, the team decided to replace NA values with the mean value of the dataset as it is a straightforward approach that maintains the dataset's “overall central tendency” (Duxburry Press, 1999). Using the mean helps ensure that the data remains balanced, allowing for more accurate analysis and predictions. In contrast, replacing NA values with the median does not utilize all data points, potentially leading to less precise estimation. Using the mode can be useful for categorical data but may not be suitable for continuous data, as it can lead to oversimplification and may not accurately represent the dataset's divergence.

A screenshot of a graph

Description automatically generated

(fig.4 Empty value refill with the mean value of the dataset)

# Website Requirement

## Website required features

Statistic graph (Bar chart)

* To discrete the contrast of two different groups in one category.
* Highlighting Differences: Display the difference between the two groups, such as higher or lower consumption rates.

Graph (Line chart)

* To display the trend of the flow, visualize and display the increase or decrease of the trend
* By using line connecting datapoint. It could best present the pattern and figure out changes in timeline.

Graph (pie chart)

* To display the clear portion of each group in the category
* Visually demonstrates the relative sizes of smoking and alcohol consumption groups, making it easy to see which group is larger

## Optional features

* Animation of data visual line transition: Enhance the interactive of the visual line graph.
* Hover the mouse on any bar in the bar chart and display the number.

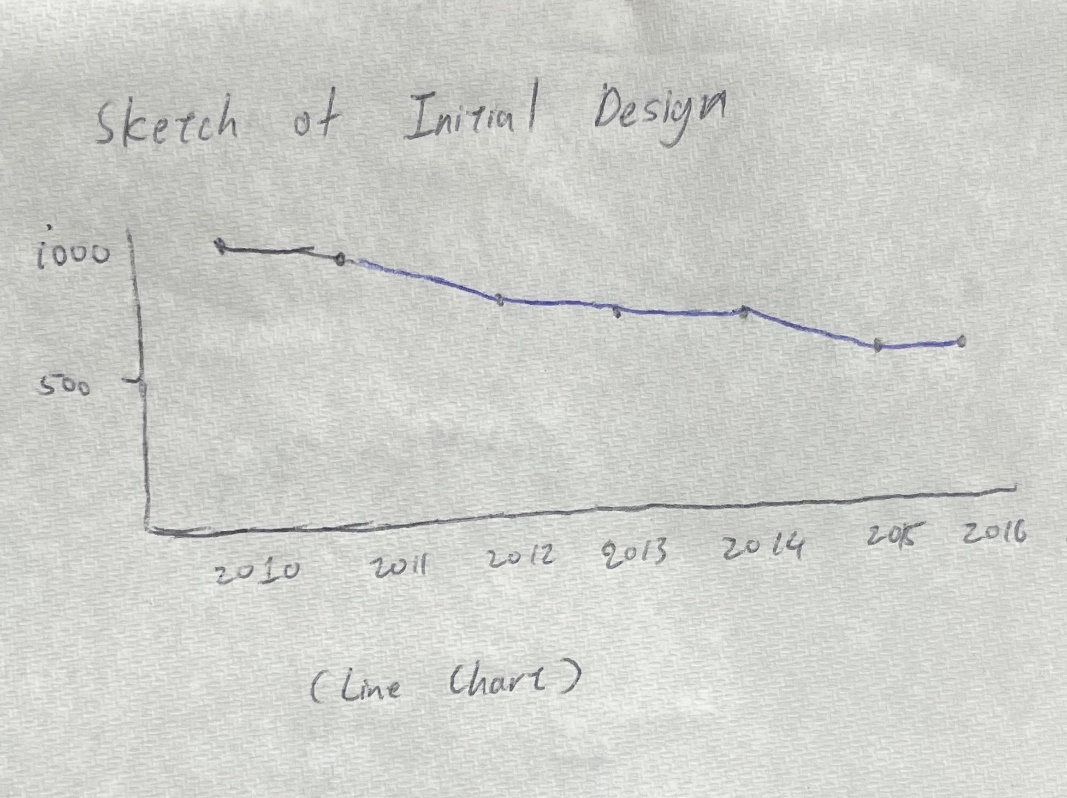
# Visualization design

## Initial idea designs

Whenever designing the visualize of graph on the website, it is crucial to understand that each graph usually correspond to a specific question or purpose. Hence the team came up with a few graphs type with clear subject to demonstrate the goal of each graph and the design purpose.

### The trend of average consumption alcohol and tobacco of person over year (Line chart)

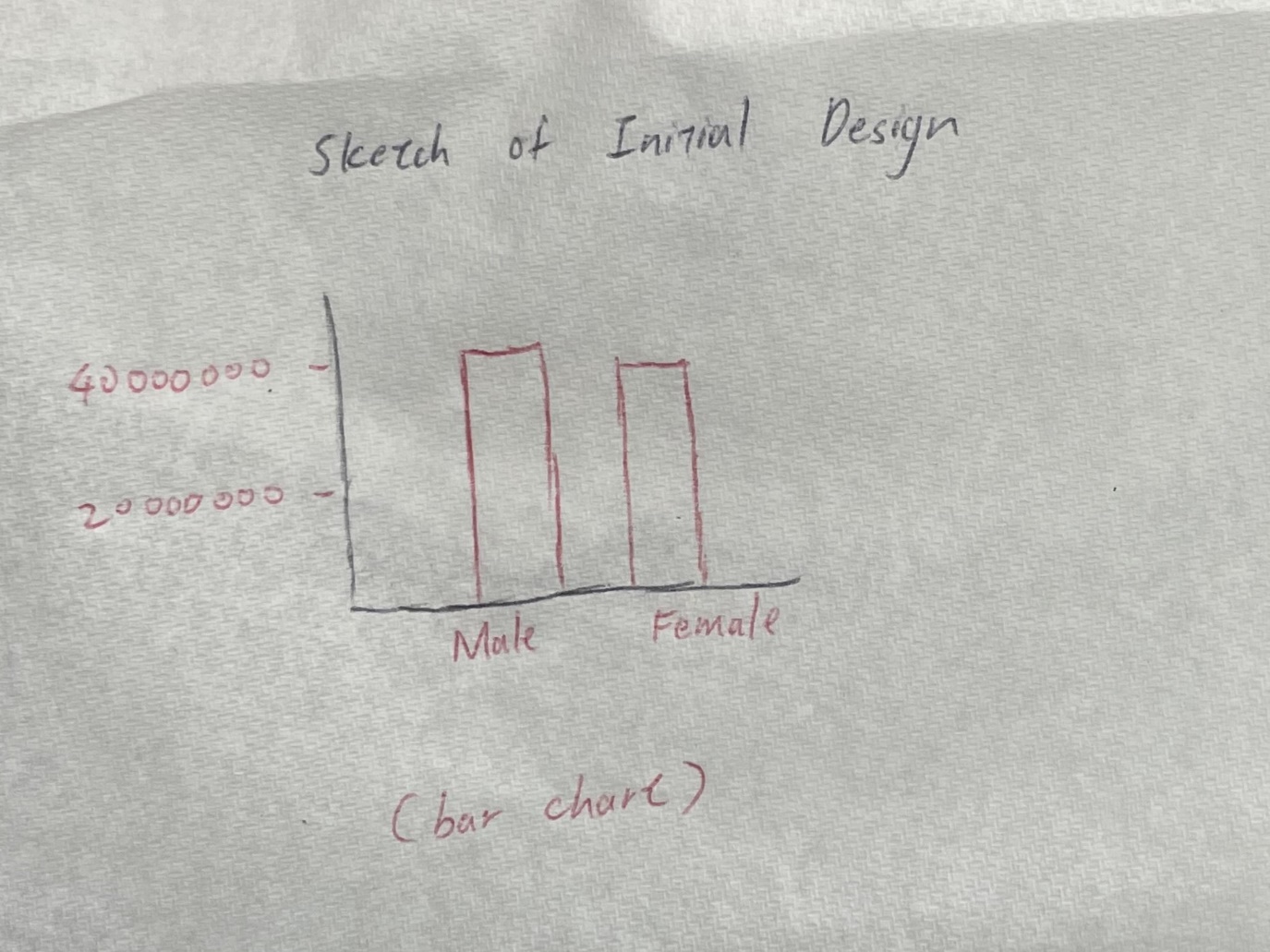
Question 1: What is the average consumption per person in a year?



* The primary design of the line graph is to indicate the average tobacco and alcohol consumption per person over the year.
* There will be two individual line chart for each corresponding item.
* The x-axis represents the year, and the y-axis represents the quantity consumed over the year.

### The population of consuming alcohol and tobacco between male and female (Bar chart)

Question 2: How do the rates of tobacco and alcohol use compare between male and female groups?



* The bar illustrates the number of populations consuming alcohol or tobacco between male and female group.
* The x-axis stands for each discrete group which is male and female, and the y-axis stands for the number of populations who smoke or drink matter currently or previously.
* The design will specify the year of this data visualize statistic occurs.

For example:

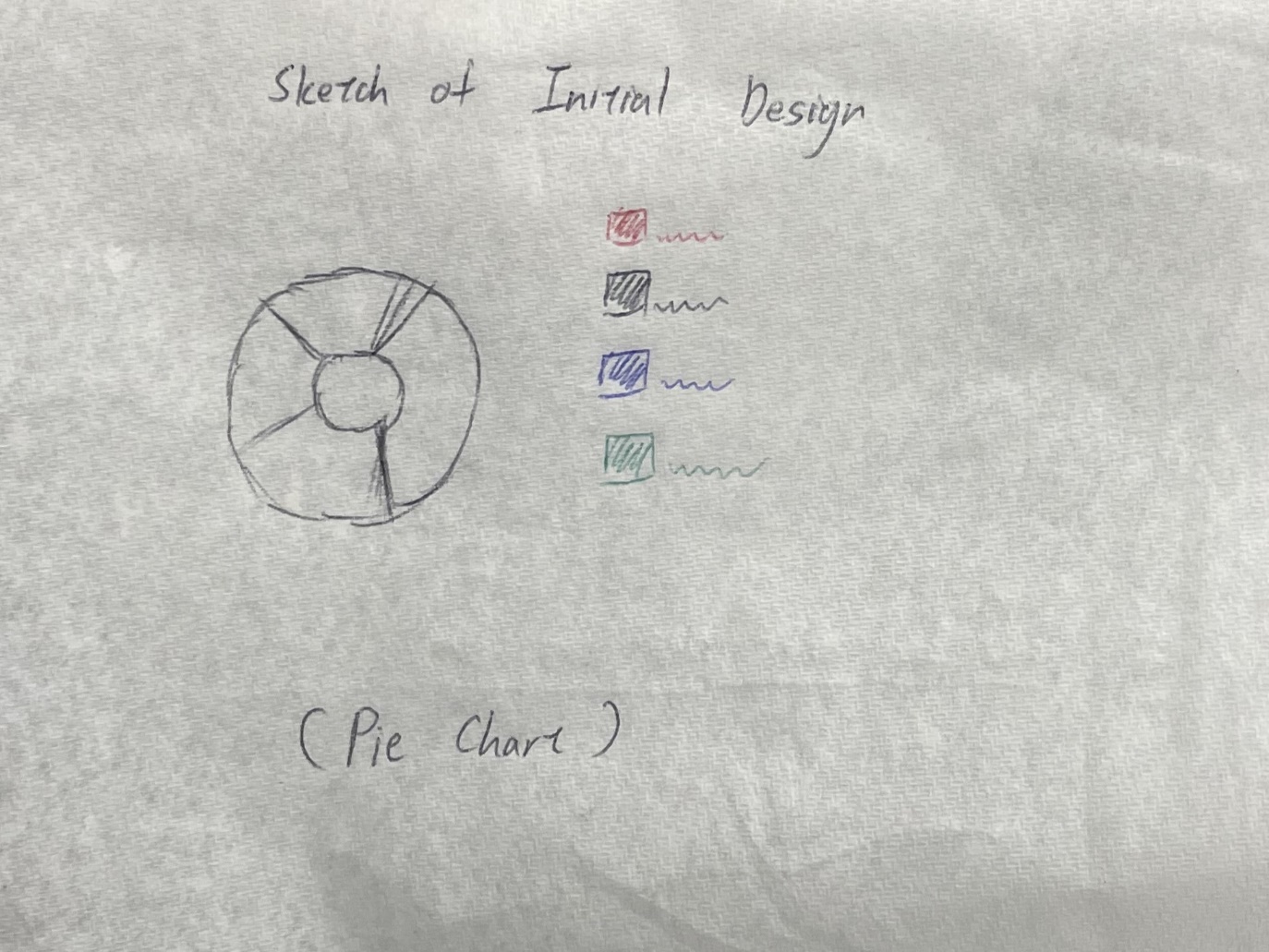
-2016

-2019

-2022

### The portion percentage of group consuming alcohol or tobacco (pie chart)

Question: What is the percentage of the status that consumes tobacco or alcohol?



* The pie chart illustrates the percentage of the status of consuming alcohol or tobacco between male and female group
* The status include:

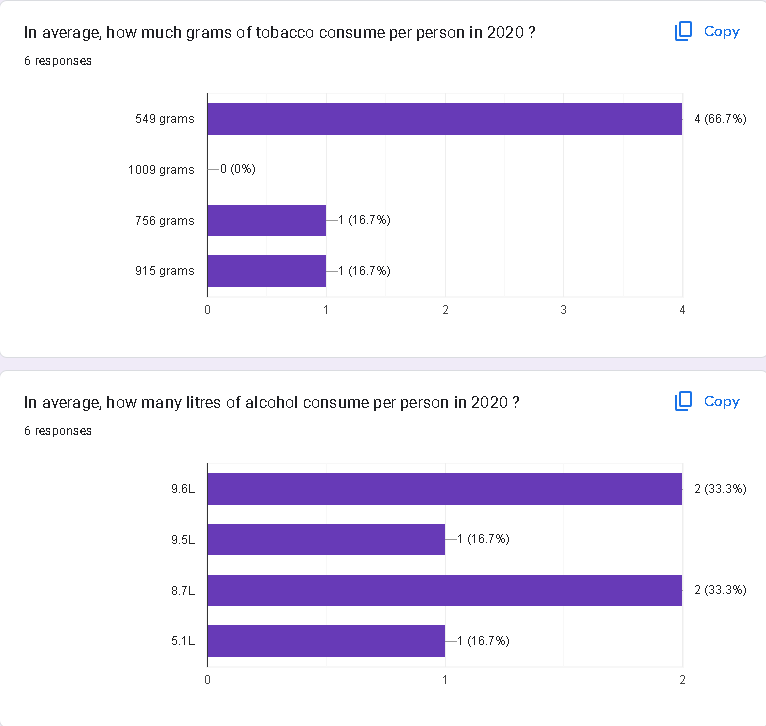
1. Daily
2. Weekly
3. Monthly
4. Smoked previous
5. Never smoked

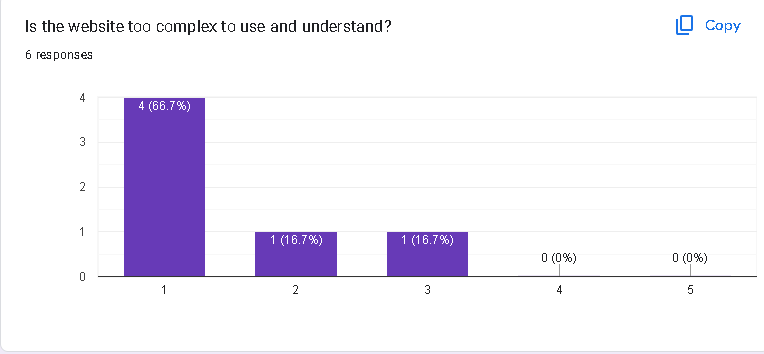
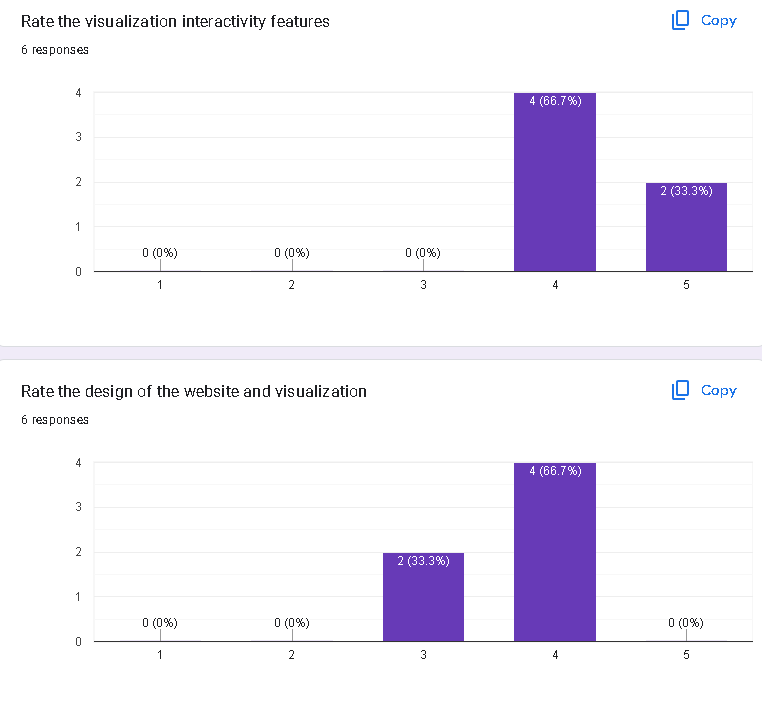
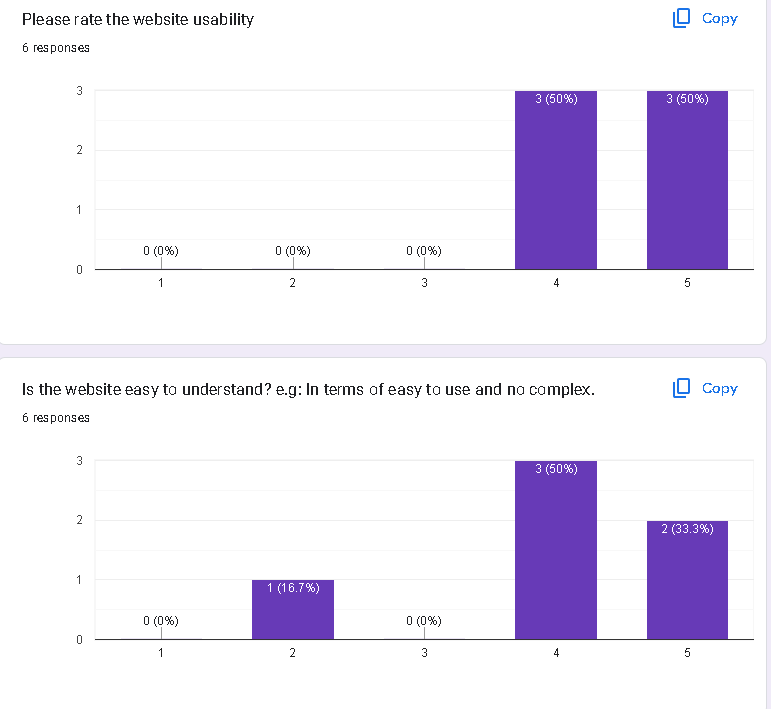
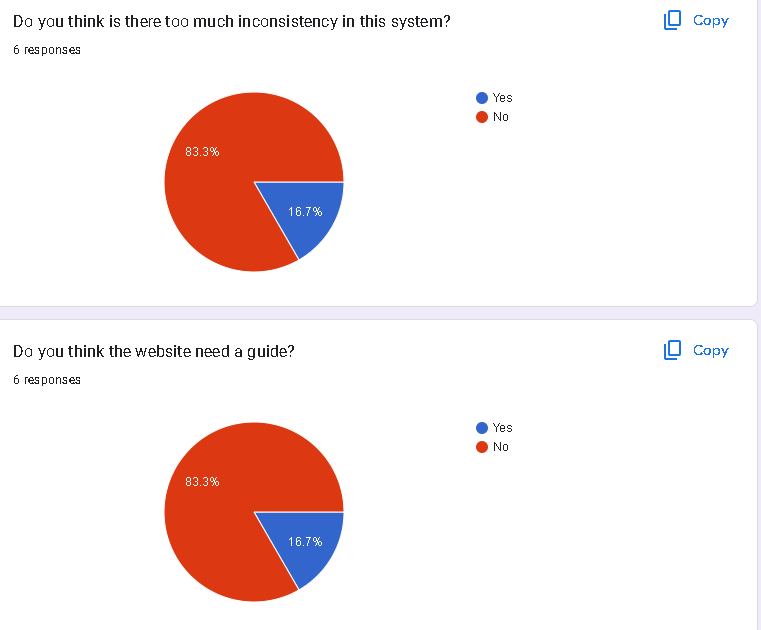
* Like the bar chart, the design will specify the year of data visualization.

## Justification of visual coding

# Validation

The user rating assessment came up with 6 different responses to the following questions with figures as evidence.





Ultimately, with the yielded practical data, the response proposes the website exists with the inconsistency of the page format. Besides the graph page format may slightly differ from the index page.

The on-page elements such as images, heading, colour theme, and background images. Another potential factor such as the level of interactivity could potentially as mainly cause of why the rating drop-down. On the other hand, the response also suggests that the website rating toward positive and user-friendly. The data could be used as a reference in the future iteration of the website to offer better productive method of visualization.

# Conclusion

Ultimately, the project aimed to visualize the consumption patterns of tobacco and alcohol to raise awareness about their health impacts. The team developed clear and instructive visualizations by leveraging datasets from the OECD and AIHW to aid various stakeholders, such as policymakers and healthcare professionals. These visualizations included line charts to track consumption trends over time, bar charts to compare consumption between males and females, and pie charts to show the relative proportions of different consumption statuses. The process involved proper data cleaning and processing, including filling in missing values with the mean to maintain overall data balance. This approach ensured a consistent dataset, smoothing accurate analysis, and predictions. The need to provide particular answers to inquiries concerning consumption patterns, like trend identification, gender group comparisons of the level of consumption, and understanding of consumption proportions. Despite some challenges, such as handling missing data and ensuring the accuracy of visual representations, the project was achieved in its primary goal. The visualizations provided a clear, easy way to understand complex data, enhancing public awareness and informing stakeholders about the potential health risks associated with tobacco and alcohol consumption. Additionally, user feedback highlighted areas for improvement, such as the need for more consistent page formatting and enhanced interactivity, which will be valuable for future website iterations. Overall, this project demonstrated clear data visualization via the concept of applying the programming technique of d3 javascript to export the message of important health information and engage diverse audiences.

# References

OECD library, non-medical determinant of health, Available at: <https://www.oecd-ilibrary.org/social-issues-migration-health/data/oecd-health-statistics/oecd-health-data-non-medical-determinants-of-health_data-00546-en?parentId=http%3A%2F%2Finstance.metastore.ingenta.com%2Fcontent%2Fcollection%2Fhealth-data-en> (Accessed at: 24/ 04/2024)

Australian Institute of Health and Welfare, National Drug Strategy Household Survey 2022–2023, Available at:<https://www.aihw.gov.au/reports/illicit-use-of-drugs/national-drug-strategy-household-survey/data> (Accessed at: 24/ 04/2024)

Faculty Washington(1999),Available at: <http://faculty.washington.edu/swithers/seestats/SeeingStatisticsFiles/seeing/center/meanvmedian3.html> (Accessed at: 12/05/2024)

Laerd, Measures of Central Tendency(2018), Available at: <https://statistics.laerd.com/statistical-guides/measures-central-tendency-mean-mode-median.php> (Accessed at: 17/05/2024)

Stackflow, D3.js donut chart... arc.centroid(d) is not influenced by d.innerRadius and d.outerRadius (2013), Available at: <https://stackoverflow.com/questions/16062257/d3-js-donut-chart-arc-centroidd-is-not-influenced-by-d-innerradius-and-d-ou> (Accessed at: 25/05/2024)

D3-graph-gallery, Interactive pie chart with input data selector , Available at:<https://d3-graph-gallery.com/graph/pie_changeData.html> (Accessed at: 25/05/2024)