

COS-30045 Data visualization process book

Semester 1 2024



Tsung Lee 103497841

[Date]

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 methods to reduce peoples' tension and relief their stress. As well as one of the great factors that guide as non-medical determinant death of the population. As numerous of quantity growth on the consumption, people may not still aware the impact of it. Hence, in this project, the motivation that guide the group to do is to present a clear visualization of the graph to further attracts several groups. The visitor may utilize these graphs as collection of reference to gain and spread alert on the potential health hazard of continue consumes alcohol and tobacco.

# Visualization Purpose

The visualization of the OECD Health Statistics dataset serves various purposes for users. It allows them to compare and analyze healthcare resources across countries, identifying trends and disparities. Users can explore questions related to workforce demographics and infrastructure, enabling benchmarking and best practice identification. Ultimately, the visualization empowers informed decision-making, policy development, and healthcare system improvement efforts.

# 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 of the website

Week 5 (Week 10)

-Finish website

Week 6 (Week 11)

-Finish Process book

Week 7 (Week 12)

-Start and finish the reflection

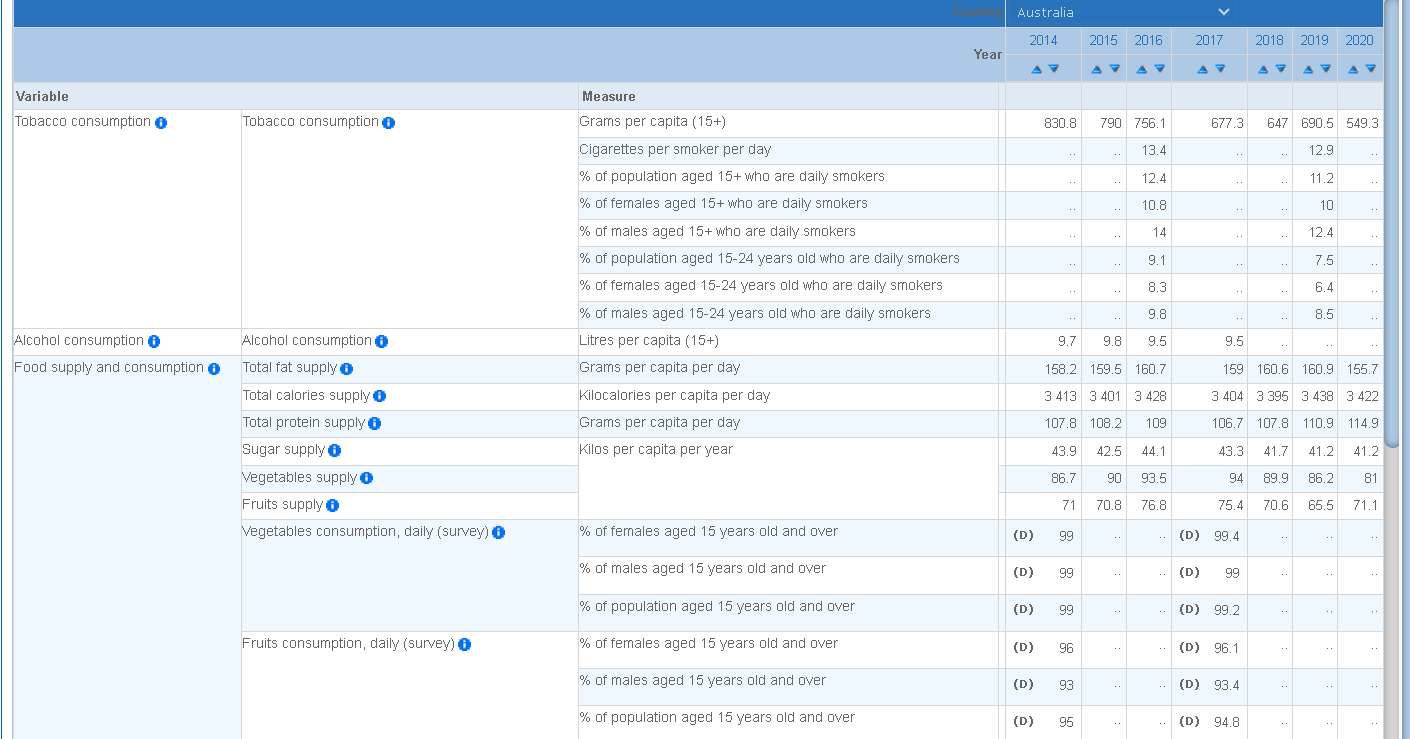
# Data

## Data Source

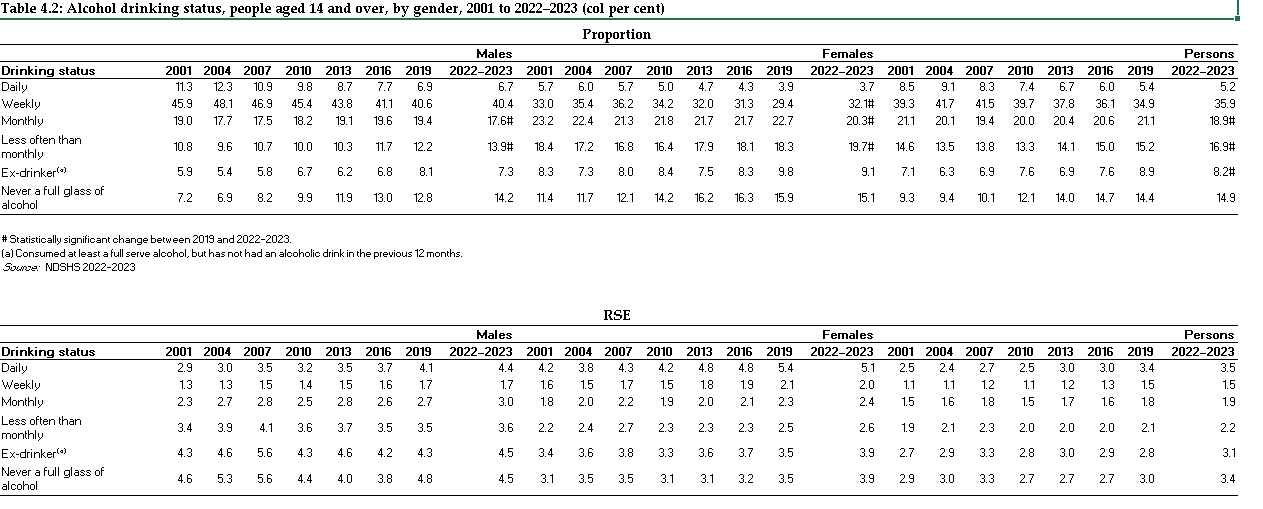
The main data in the project is referred to:

<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>

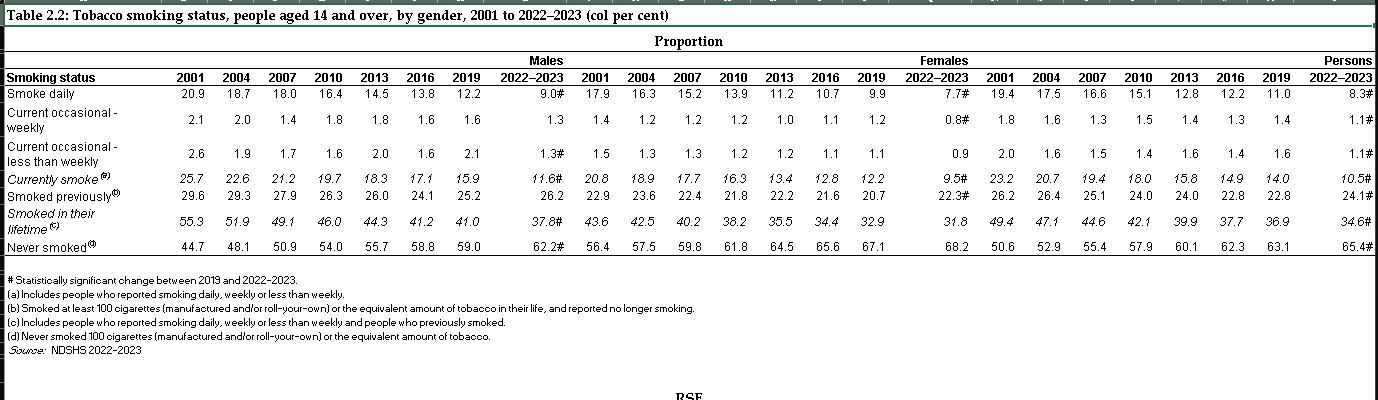
From the original source dataset (Fig. 1 OECD non-medical determinant of health),



The teams’ main objectives are to present the visualize of the tobacco (cigarette) consumption and alcohol consumption. However, according the dataset, huge numbers of data are missing (present as Not Available), therefore, although by find the NA via mean, mode, or median. The overall presentation of the data will not be precise and accurate. Hence during the development of the project, the group was managed to find external dataset source with richer content.



(fig2. External dataset source of alcohol drinking status)



(fig.3 External dataset source of tobacco/ cigarette smoking status )

## Data Processing

### Data Clean Up

### Filling NA (Not Available) Value Gap

The team was decided to fill the NA column with the mean value

A screenshot of a graph

Description automatically generated

(fig. Empty value refill)

# Website Requirement

## Website required features

Australia map visualization

* The choropleth map of Australia with either hover effect or direct visual color gradient effect to display the viewer the data.

Statistic graph (Bar chart)

* To discrete the contrast of two different group in one category.

For example: number of cigarettes consume by person between male and female group.

graph (Line chart)

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

graph (pie chart)

* To display the portion of each group in the category

## Optional features

Animation of data visual line transition

# Visualization design

## Initial idea designs

Australia choropleth Map:

• The idea was to use a choropleth map to display the number of population consume cigarette which is tobacco and alcohol of the most recent years’ record.

• When viewer hovers the mouse on each state, it will display the numbers

## 

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

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

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

## Justification

# Conclusion

# References