



Data Visualization – COS30045

Health Topics
in Australia and the world

Tutorial 10 - Tuesday
10:30 a.m. – 12:30 p.m

Semester 1 – 2024
Word Count: 2402

https://mercury.swin.edu.au/cos30045/s104309099/data_visualisation/

Team 6
Minh Nguyen - 104309099
Trong Nhan Vo – 103812143

Due date: June 7th, 2024

Table Of Content

SUMMARY.....	2
1. INTRODUCTION.....	3
1.1 BACKGROUND AND MOTIVATION	3
1.2 VISUALIZATION PURPOSE.....	3
1.3 PROJECT SCHEDULE	4
2. DATA	5
2.1 DATA SOURCE	5
2.2 DATA PROCESSING	7
3. REQUIREMENTS	9
4. VISUALIZATION DESIGN	10
4.1 INITIAL DESIGN	10
4.2 FINAL DESIGN	12
5. VALIDATION	15
6. CONCLUSION	15
REFERENCES.....	16

Summary

This report highlights the visualization background, data findings and cleaning, visualization requirements, design, and validation of a comprehensive data visualization project. This project illustrated health situations in Australia and worldwide through many aspects (physicians, tobacco, and alcohol consumption). The primary purpose of this project is to create a user-centered website that demonstrates data visualizations and interactive features.

The project's primary features include interactive effects to read the information of each data point, a concise and straightforward design, accurate data representation, and a range slider or a dropdown to show data fluctuation and various data aspects.

This project mainly concerns healthcare, such as the number of physicians in each state in Australia, alcohol consumption, and tobacco consumption in both OECD and non-OECD countries. The initial design used screenshots from charts made in Excel or simple drafts using D3. The final design utilized some graphs, such as donut and bubble graphs, to clearly and accurately demonstrate the data.

Validating data involves checking for sources, methods to process data correctly, and data representation in the visualization. Standard design guidelines were followed, and interactive features were added to understand the data better and illustrate effects.

In conclusion, this project has contributed to a better understanding of many aspects of the health situation in Australia and other countries, providing valuable insights for readers and policymakers. The project emphasizes the importance of informed discussions and decisions in health-related law.

1. INTRODUCTION

Health and its contributing factors have been essential for countries, reflecting their economic status and civilian living standards. Hence, there is an increased demand for health data visualization to show health features in many countries worldwide.

1.1 BACKGROUND AND MOTIVATION

Health is an essential field of a country, dramatically impacting the economy, demographics, and even policies. However, understanding this trend by reading articles or searching for data requires massive effort and time to process all the information.

That is why we decided to do this project: to make the data easily understandable regardless of your educational background or occupation. Statisticalisation presents health data on tobacco consumption, alcohol consumption, and the number of physicians in Australia.

Beyond being a tool for educational purposes, our visualization aims to raise people's awareness of the health situation in Australia and worldwide and broaden their perspectives by presenting different health aspects.

1.2 VISUALIZATION PURPOSE

The project will help the authorities have a great point of view on health to identify potential challenges to well-being so that they can modify the policies based on real-time situations in many countries. Furthermore, some tools will be added so that users can interact with the data for better illustration. For instance, users can use visualized charts to identify the consumption of alcohol and tobacco. Donut charts are helpful for people who want to analyze the percentage of physicians in Australia over a long period. Moreover, the project explains why the health system dramatically impacts people's well-being and immunization. This point is vital for the government to make important decisions about allocating resources, developing healthcare infrastructure, and modifying health policies.

1.3 PROJECT SCHEDULE

- Week 7: Prepare for the project by reading the requirements of the project and create a GitHub repository.
- Week 8: Work on the project.
 - Prepare the outline of the process book.
 - Add some information in the introduction.
 - Set up a basic website for demonstrating the charts.
 - Plan which types of chart will be used for the project.
- Week 9 - 10: Search for the data and analyze.
 - Search for suitable data.
 - Process data and convert them into suitable datasets for the charts using D3.
 - Writing about data processing and cleaning in the process book.
- Week 11: Finalize the website and the charts.
 - Add functions to present data properly.
 - Add CSS effects to decorate the website interface.
 - Add JS to convert datasets into charts using D3.
 - Ask for feedback from the tutor.
- Week 12: Finish the process book.
 - Finalize the book.
 - Add references.
 - Check for any grammar and vocabulary mistakes and fix them.
- Week 13 - 14: Finalize the project.
 - Get feedback from the tutor.
 - Submit the work on Canvas before deadline.

2. DATA

2.1 DATA SOURCE

The project mainly uses the OECD database as the data source. The OECD database contains data in table form on many aspects of the health system, including the number of physicians in Australia and tobacco and alcohol consumption. We retrieve these data tables by downloading the files from the OECD database in XLSX files and then converting them into CSV files for later analysis. These datasets give us a broad view of Australia's and other countries' health systems. Our goal is to help the readers understand this multifaceted situation through a thorough collection of essential datasets and data analysis to give different insights and informed discussions on health.

Here is where we retrieve the data tables from: <https://data-explorer.oecd.org/>

We analyze many data files, focusing on different aspects of health:

- Physicians in Australia (physicians.csv): The dataset contains these fields:

Field Name	Description	Types of Data
Regions	The states of Australia (New South Wales, Victoria, etc.)	Nominal
Period	The years of the data recorded	Ratio
Period (%)	The years of the data recorded in percent	Ratio

The 'Regions' field presents data for each region in Australia, like New South Wales or Victoria, from 2010 to 2021. The data statistics for each year represent the number and proportion of physicians in each region during this period. We will use all the data in this dataset except for the title of the dataset and the unit of measurement. These details will be added later using HTML.

- Alcohol consumption (alcohol.csv): The dataset contains these fields:

Field Name	Description	Types of Data
Countries	The countries with alcohol consumption	Nominal
Alcohol	The amount of alcohol consumed in liters	Ratio

The 'Countries' field represents the countries in the dataset. The 'Alcohol' field represents the amount of tobacco consumption by people over 15 years old, using liters as the unit of measurement. We will use the data recorded in 2017 in this dataset and remove the data recorded in other years and unnecessary features like the title of the table, the unit of measurement, and the subject of the table. These details will be added later using HTML.

- Tobacco consumption (tobacco.csv): The dataset contains these fields:

Field Name	Description	Types of Data
Countries	The countries with tobacco consumption	Nominal
Tobacco	The amount of tobacco consumed in grams	Ratio

The 'Countries' field represents the countries in the dataset. The 'Tobacco' field represents the amount of tobacco consumed by people over 15 years old, using grams as the unit of measurement. We will use the data recorded in 2018 in this dataset and remove the data recorded in other years and unnecessary features like the title of the table, the unit of measurement, and the subject of the table. These details will be added later using HTML.

2.2 DATA PROCESSING

```

1 Health care - Regions
2 Frequency of observation: Annual
3 Measure: Active physicians
4 Unit of measure: Persons
5 Time period,2010,2011,2012,2013,2014,2015,2016,2017,2018,2019,2020,2021
6 Australia,"76,198","78,922","79,640","82,408","85,491","87,999","91,341","95,194","98,395","101,841","105,293","108,501"
7 New South Wales,"24,348","25,468","25,566","26,676","27,374","28,095","28,995","30,037","30,809","31,817","32,681","33,622"
8 Victoria,"18,704","19,437","19,621","20,230","21,186","21,893","22,712","23,762","24,641","25,684","26,822","27,590"
9 Queensland,"15,200","15,656","15,897","16,224","16,950","17,548","18,563","19,553","20,334","21,041","21,785","22,485"
10 South Australia,"6,137","6,337","6,296","6,457","6,618","6,712","6,836","7,009","7,161","7,365","7,500","7,640"
11 Western Australia,"7,600","7,671","7,906","8,232","8,667","8,937","9,251","9,631","9,974","10,276","10,692","11,119"
12 Tasmania,"1,778","1,814","1,777","1,845","1,900","1,902","1,964","2,059","2,211","2,259","2,386","2,483"
13 Northern Territory,934,974,"1,008","1,002","1,053","1,101","1,164","1,221","1,256","1,282","1,286","1,339"
14 Australian Capital Territory,"1,497","1,565","1,569","1,654","1,654","1,715","1,768","1,820","1,893","1,965","2,020","2,102"
15 |

```

Figure 1: Physicians in Australia

```

1 Region,2010,2011,2012,2013,2014,2015,2016,2017,2018,2019,2020,2021
2 Australia,76198,78922,79640,82408,85491,87999,91341,95194,98395,101841,105293,108501
3 New South Wales,24348,25468,25566,26676,27374,28095,28995,30037,30809,31817,32681,33622
4 Victoria,18704,19437,19621,20230,21186,21893,22712,23762,24641,25684,26822,27590
5 Queensland,15200,15656,15897,16224,16950,17548,18563,19553,20334,21041,21785,22485
6 South Australia,6137,6337,6296,6457,6618,6712,6836,7009,7161,7365,7500,7640
7 Western Australia,7600,7671,7906,8232,8667,8937,9251,9631,9974,10276,10692,11119
8 Tasmania,1778,1814,1777,1845,1900,1902,1964,2059,2211,2259,2386,2483
9 Northern Territory,934,974,1008,1002,1053,1101,1164,1221,1256,1282,1286,1339
10 Australian Capital Territory,1497,1565,1569,1654,1654,1715,1768,1820,1893,1965,2020,2102
11 |
12

```

Figure 2: Cleaned Physicians in Australia

Figure 1 and **Figure 2** present the process of cleaning the dataset, which is essential for data analysis. This process focuses on many steps, such as removing unnecessary elements, such as the table title or the unit of measurement, which often contribute little to the need for visualization. By removing these elements, the usability and integrity of the datasets will be increased, which help create a more concise and suitable representation for the subsequent process.


```

1 country,tobacco
2 Australia,647
3 Czechia,2367
4 Denmark,1113
5 France,905
6 Germany,1507
7 Israel,769
8 Japan,1171
9 Korea,1606
10 Switzerland,1271
11 USA,1061
12 UK,446
13 Finland,734
14 Hungary,1521
15 Norway,793
16 ✨
17

```

Figure 3: Cleaned Tobacco Consumption

```

country,alcohol
Australia,9.5
Czechia,11.6
Denmark,9.7
France,11.7
Germany,10.9
Israel,2.9
Japan,7.2
Korea,8.7
Switzerland,8.6
USA,8.9
UK,9.7
Finland,8.4
Hungary,10.7
Norway,6

```

Figure 4: Cleaned Alcohol Consumption

Figure 3 and **Figure 4** presented the cleaned data and the critical characteristics of Physicians in Australia, Tobacco Consumption, and Alcohol Consumption. Each figure demonstrates some aspects of the datasets and their correlations, which offers an overview of the aspects of health. The visualizations help enlighten readers on the correlations and the patterns within these different characteristics. Hence, visualizations play an essential role in the analysis and visualization of these aspects of health.

3. REQUIREMENTS

The requirements are the critical features needed for a practical project.

Must-have Features:

- To encourage users' engagement, we create interactive visualizations with the mouse hovering to show the statistics.
- We plan to design three distinct charts with three datasets of three factors
- to provide information for various health factors.
- The visualization aims to adopt a concise and user-friendly design.
- It is essential to demonstrate a concise and up-to-date data representation to guarantee valid points of view.
- Interactive tools are utilized to encourage the interaction between the users and the visualizations.

Optional Features:

- The integration of smooth animations for the website is proposed to encourage data exploration and comprehension with intrigue and fascination.
- Legends provide information about colors and chart descriptions, helping users understand the chart's purpose. A legend and interactive effects to show data statistics are also made for clarification.
- Sliders for years are made to provide context for the data and compare the data in each year.

4. VISUALIZATION DESIGN

4.1 Initial Design

Question 1: How has the number of physicians in Australia changed over time?

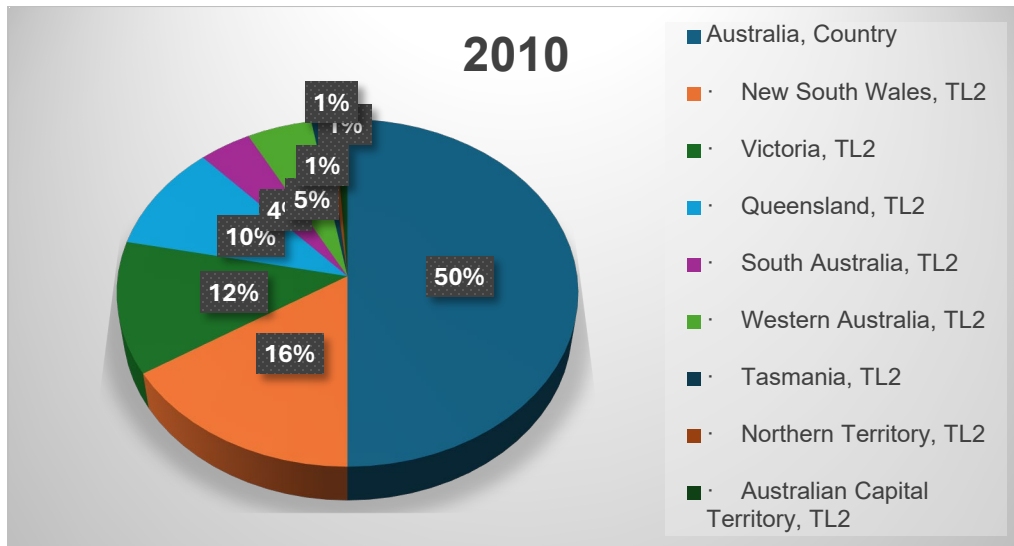


Figure 5: Initial Number of Physicians in Australia

Figure 5 shows a screenshot from an Excel chart. This preliminary chart outlines the percentage of physicians in each region in 2010. The dataset's design and chosen year provide a brief overview of the number of physicians in Australia. The legends are added for visualization layers and are marked using different colors. The percentages of each region are shown on the chart.

Question 2: What is the consumption of tobacco by children over 15 years old?

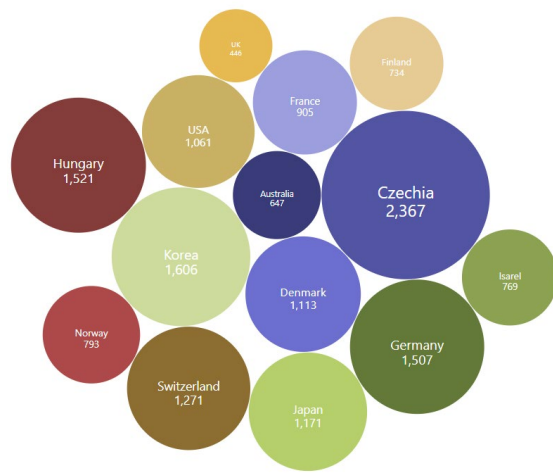


Figure 6: Initial Tobacco Consumption

A simple visualization for Figure 6 is made with D3 to present the primary characteristics of tobacco consumption. This chart presents a draft of the amount of tobacco consumed by children in 2018. Inside the bubbles are the names of the countries and the amount of tobacco consumed below. The higher the amount of tobacco consumed, the bigger the bubble is.

Question 3: What is the consumption of alcohol by children over 15 years old?

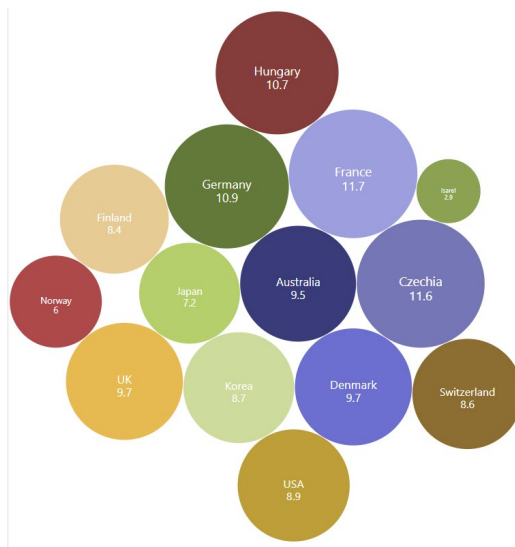


Figure 7: Initial Alcohol Consumption

A simple visualization for Figure 7 is made with D3 to present the primary characteristics of alcohol consumption. This chart presents a draft of the amount of alcohol consumed by children in 2017. Inside the bubbles are the names of the countries and the amount of alcohol consumed below. The higher the amount of alcohol consumed, the bigger the bubble is.

4.2 Final Visualization Design

Question 1: How has the number of physicians in Australia changed over time?

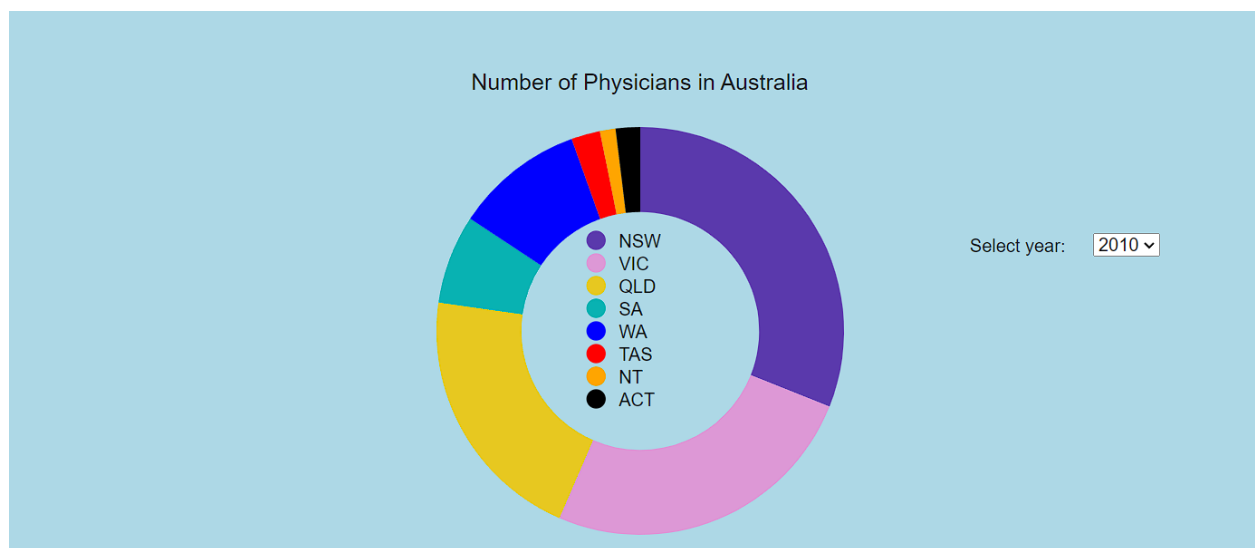


Figure 8: Final Number of Physicians in Australia

Figure 8 illustrates a donut chart that visualizes the number and percentage of physicians across various Australian states from 2010 to 2021. Users can access the precise count and percentage of physicians for each state by hovering over the chart segments. A central legend bar employs distinct colors to denote each state. Additionally, users can utilize a dropdown menu to select specific years and explore physician distributions for those selected years.

Question 2: What is the consumption of tobacco by children over 15 years old?

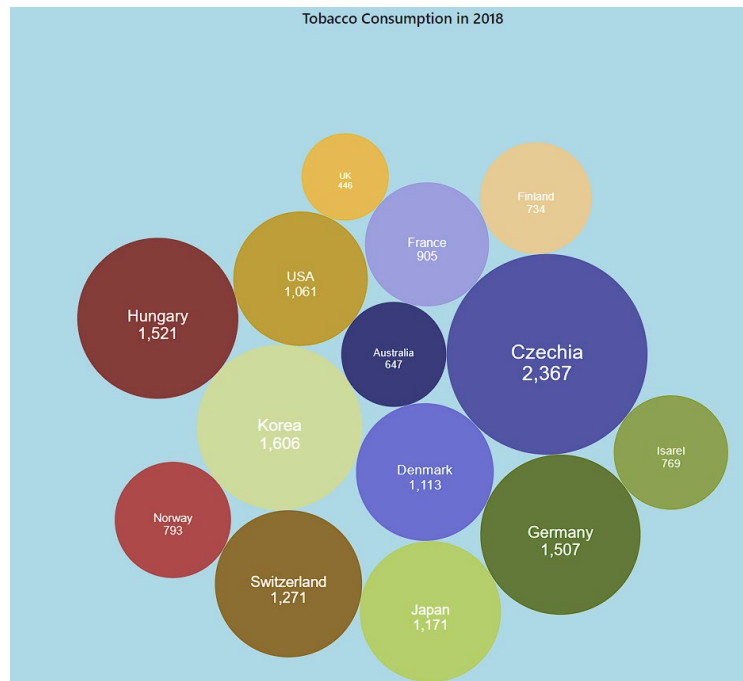


Figure 9: Final Alcohol Consumption (liters)

Figure 9 uses a bubble chart to visualize tobacco consumption per person aged 15 and above in countries like Switzerland and Germany. The chart includes bubbles representing each country, with bubble size indicating consumption levels. Below the country names, the chart displays the amount of alcohol consumed per person, providing additional context to the data visualization.

Question 3: What is the consumption of alcohol by children over 15 years old?



Figure 10: Final Tobacco Consumption (grams)

Figure 10 employs a bubble chart to illustrate tobacco consumption per person aged 15 and above across various countries. Each country is represented by a bubble, with different colors enhancing the visual differentiation. Below the country names, specific data for tobacco consumption in each country is provided, offering detailed insights into the consumption levels.

5. VALIDATION

Data validation involves verifying reliable sources, applying proper methods to process data, and adequately representing data through visualizations.

The datasets' sources are verified since all were taken from the OECD database, a trusted intergovernmental organization worldwide. The data in the datasets is chosen randomly and is used only for data analysis and visualization. No racism or discrimination is intended.

The data were cleaned to ensure they were appropriate for analysis and visualization. The process comprised removing unnecessary features such as unused columns and rows, titles, and units of titles and units. Furthermore, data normalization was used to maintain the data format and comparison.

We complied with design guidelines so that the visualizations concisely demonstrated the datasets. These comprised suitable data visualization types to demonstrate the dataset and proper use of titles, legends, and references. We also added interactive effects to encourage the users' interaction with data and make the charts more intriguing.

6. CONCLUSION

Healthcare has an essential impact on the world's demographics and economic status. The project questions highlight the various aspects of health situations in Australia and many other countries, both in good and bad terms.

This report offers a broad view of health patterns in Australia and other countries. It simplifies the large volume of data for a clearer understanding of health systems and some health factors that impact well-being. The tools used in this project help demonstrate the consumption of alcohol and tobacco and the number of physicians in Australia.

In conclusion, the visualization emphasizes the importance of health systems worldwide and contributes to a better understanding of this unfamiliar situation. Hence, understanding health status is vital for everyone, especially policymakers, who must modify their policies to suit the current situations in their countries better.

REFERENCES

Murray, S. (2017). *Interactive Data Visualization for the Web*. "O'Reilly Media, Inc."

http://books.google.ie/books?id=NGwvDwAAQBAJ&printsec=frontcover&dq=Interactive+Data+Visualization+for+the+Web+:+An+Introduction+to+Designing+with+D3&hl=&cd=1&source=gbs_api