COS30045 Data Visualisation
Process Book
Trends and Comparisons in Health Statistics: An OECD Perspective

CL01_T07 (Tutorial Monday 10:30)

Student Name	Student ID	Year
Vihanga	103833681	Year 3 Semester 1
Cheryl Larissa	104158958	Year 3 Semester 1

Word Count: 3021 Words

Mercury link: Trends and Comparisons in Health Statistics: An OEDC Perspective

Google Docs: COS30045-Data Visualisation Project - Google Drive

Table of Contents

Introduction

Background and Motivation

The relationship between healthcare expenditure, healthcare workforce availability, and the overall health of the population is quite an important one. This stands out as a discussion on whether to invest more resources into healthcare and whether it has an impact on the total health status of a country. In this document we aim to outline the statistics of Australia in these fields, the results of notable workforce expansion and investments put into the healthcare sector as well as its impact. Through the observed trends and associations we aim to highlight the value of investment in healthcare systems to policymakers, government officials, healthcare professionals, and even the public.

We also want to examine the way that Australia invests in healthcare, including the costs and the effects on the country's citizens. Through the use of data visuals and in-depth statistical analysis, we can comprehend the relationship between healthcare investments and a nation's overall health.

Visualization Purpose

The visualization aims to depict the correlations between healthcare expenditure, healthcare workforce practitioners, and key health indicators in Australia. Through this visualization, one can answer the questions of:

- Does more money in the health sector result in more health practitioners?
- Does investing more money into the health sector affect overall health?
- What is the relationship between more practicing doctors and overall health?

As taxation is a primary way that the Australian government funds the health sector (*Health expenditure Australia 2021-22, Government spending on health relative to government expenses* 2023), the results of this case study may help to alleviate the stigma behind revenue tax and provide a positive outlook for the general public.

This study also comprises the additional benefits of promoting public health awareness, supporting research, and driving evidence-based decision-making for policies.

Data

Data Source

All data is gathered from OECD Health Statistics (link). All of the data we chose falls under the Health Status section of the health statistics.

Datasets Chosen

Mortality of people around the world in the year 2023.
We took the first 5 weeks of 2023 data and combined it. The data that we use inside this dataset are:

1. Country code as a primary key: Categorical (nominal), string
2. Country name: Categorical (nominal), string

3. Week/Week number: Ordinal, integer
4. Year: Ordinal, integer
5. Value: Ordinal, integer

We sum all of the values from week 1 to week 5 to get a new value to show.

Yearly Disease Incidence from 2010 to 2020

We chose some of the diseases that are common and used them as our dataset. The data inside are:

1. VAR as a primary key: Categorical (nominal), string
2. Variable, descriptive name of the disease or condition: Categorical (nominal), string
3. Country: Categorical (nominal), string
4. Year: Ordinal, integer
5. Value: Ordinal, integer

Health Expenditure of OECD in 2022

1. Country code as primary key: Categorical(Nominal), string
2. Country name: Categorical(Nominal), string
3. Year: Ordinal, integer
4. Health Expenditure(% of Capita): Continuous, float

Practicing Nurses of OECD from 2019 to 2021

1. Country code as primary key: Categorical(Nominal), string
2. Country name: Categorical(Nominal), string
3. Year: Ordinal, integer
4. Number of Nurses: Continuous, Integer

Practicing Physicians of OECD from 2019 to 2021

1. Country code as primary key: Categorical(Nominal), string
2. Country name: Categorical(Nominal), string
3. Year: Ordinal, integer
4. Number of Physicians: Continuous, integer

Preventable Deaths of OECD from 2019 to 2021

1. Country code as primary key: Categorical(Nominal), string
2. Country name: Categorical(Nominal), string
3. Year: Ordinal, integer
4. Number of Preventable Deaths: Continuous, integer

Treatable Deaths of OECD from 2019 to 2021

1. Country code as primary key: Categorical(Nominal), string
2. Country name: Categorical(Nominal), string
3. Year: Ordinal, integer
4. Number of Treatable Deaths: Continuous, integer

Values from 2019 to 2021 are averaged and data is processed to impute estimated values for 2023.

Data processing

A Python script is used to process the data of Practicing Nurses, Practicing Physicians as well as Preventable and Treatable Deaths. Integer values from 2019 to 2021 are averaged and an estimated value is imputed to the data to be used.

The processed datasets have the following structure:

1. Country name: Categorical (Nominal), string
2. Year: Ordinal, integer (2023)
3. Value: Continuous, integer

Countries with missing values were discarded when coding visualizations.
The **Mortality of people around the world in the year 2023** dataset also required slight data processing. Some of the country names did not align between the JSON and mortality datasets. These were changed to be similar to the JSON so that values on the map display accurately when the webpage is loaded.

Visualization Design

Visual 1

When we examined the data, we saw that only one year—2023, with numerous countries and values ranging from 0 to 69134—was displayed. We decided to select a world map visualization over other visualizations as a result.

For example, if we choose a bar chart: the list of the countries is too much for the chart to show one by one since there are a lot of countries listed in the dataset and that will disadvantage us in making the visualization because it will only make people who see it confuse. We conclude that making a world map will be the best option we can have to reach our goal because it will give a clean user-friendly appearance to show all the data.

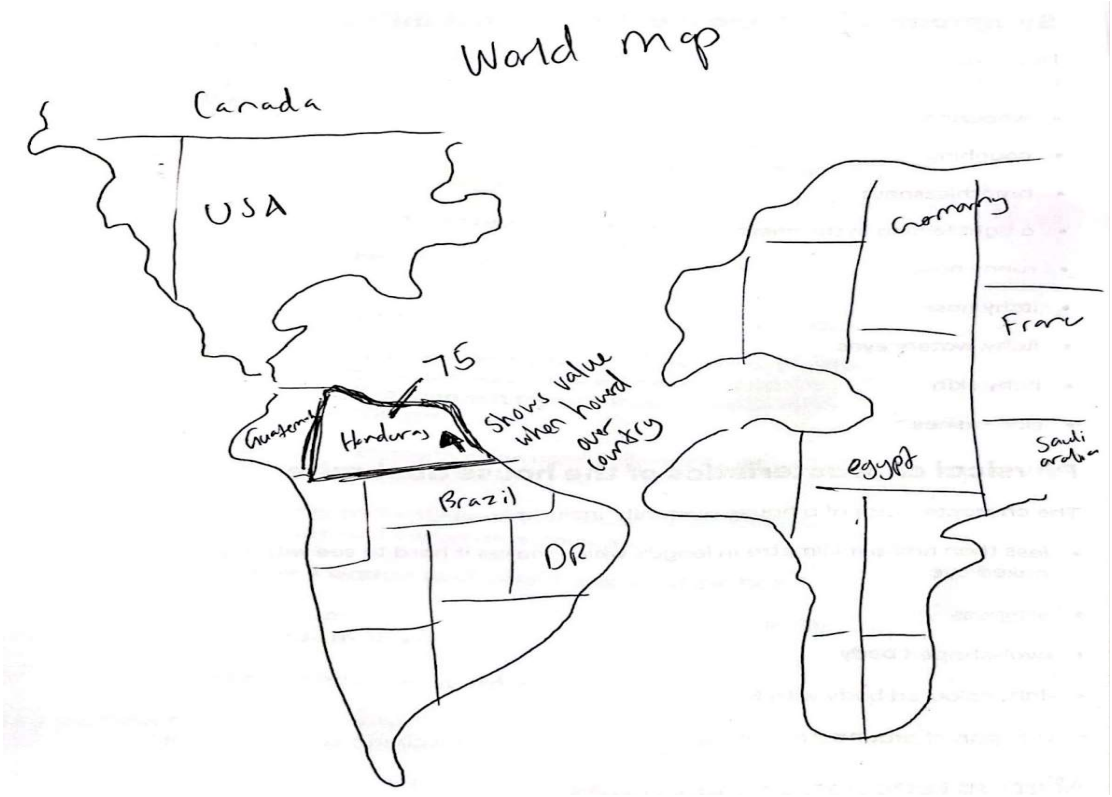

Figure 1. Sketch of world map

We sketched our idea and it come up to a world map that can be hovered and when a country is hovered it will show the information about the country.

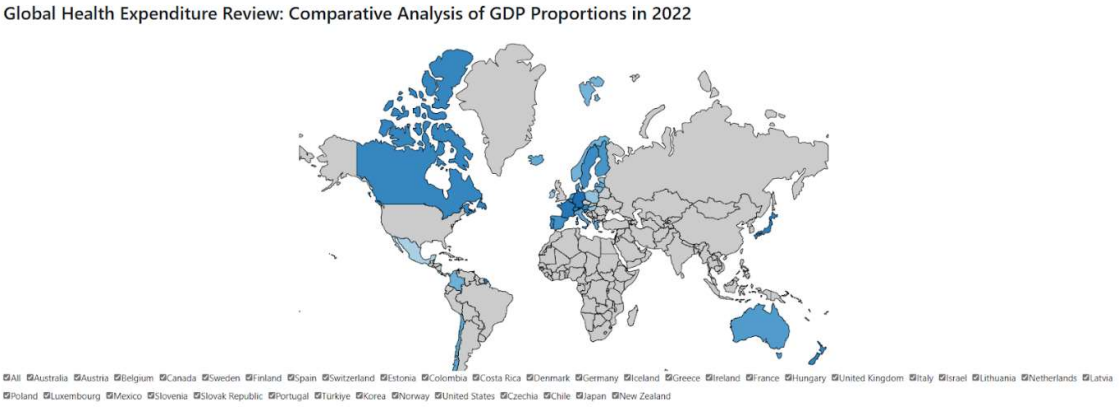

Figure 2. World map data visualization

Next, to assist viewers in sorting the data they want to examine, we tried to find a sort event after the creation of the world map. Initially, we considered creating a checkbox system in which the country whose checkbox is checked would display the selected color. We came into two main issues. Initially, the large amount of non-user-friendly checkboxes made it difficult to navigate the country's selection process. Second, due to their small size and the lack of

significant differences in the colour results, some of the smaller countries we looked at were unable to provide us with significant outcomes.

Figure 3. World map data visualization result

In the end, we replaced the checkboxes with a dropdown button to fix the sorting process. Once a country is selected, the visualisation will direct the viewer to the location where they can select the data. Dropdown lists contain all countries with values in the data that has more than zero. Additionally, the viewer can move the map, zoom in and out, and hover over the desired area to see information about it, including the name of the country and its value.

Visual 2

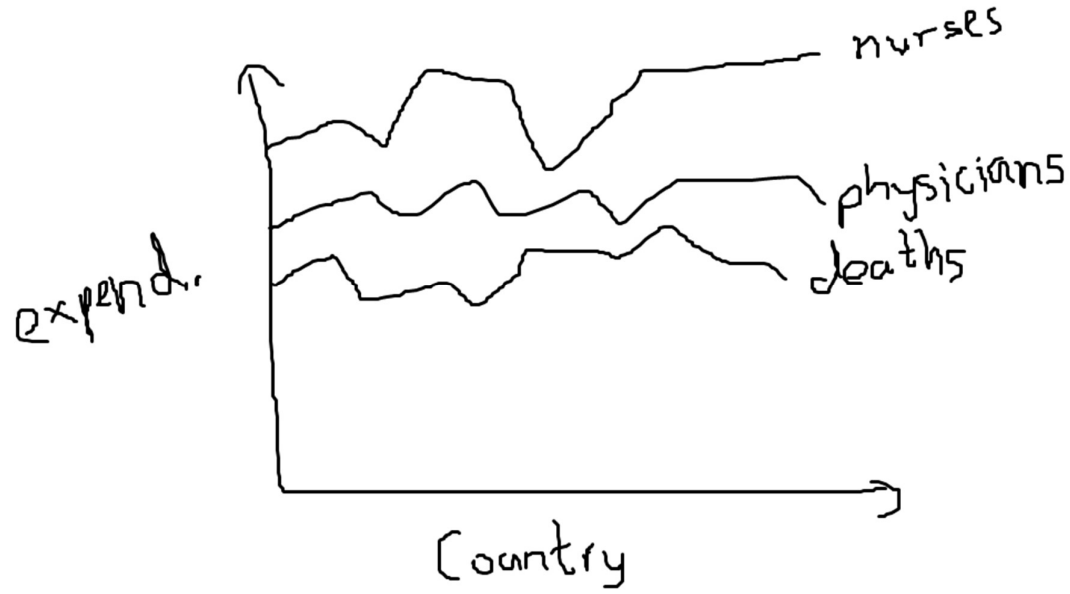

Figure 4. Sketch of the line chart

Initially, it was thought that a line graph with all the data would be an appropriate way to dissect the correlations between all of the information and construct the narrative that we are aiming for, however, the large variations and diversity of information in a singular graph makes the visualization hard to read and looks extremely cluttered. The graphs were then split into multiple graphs.

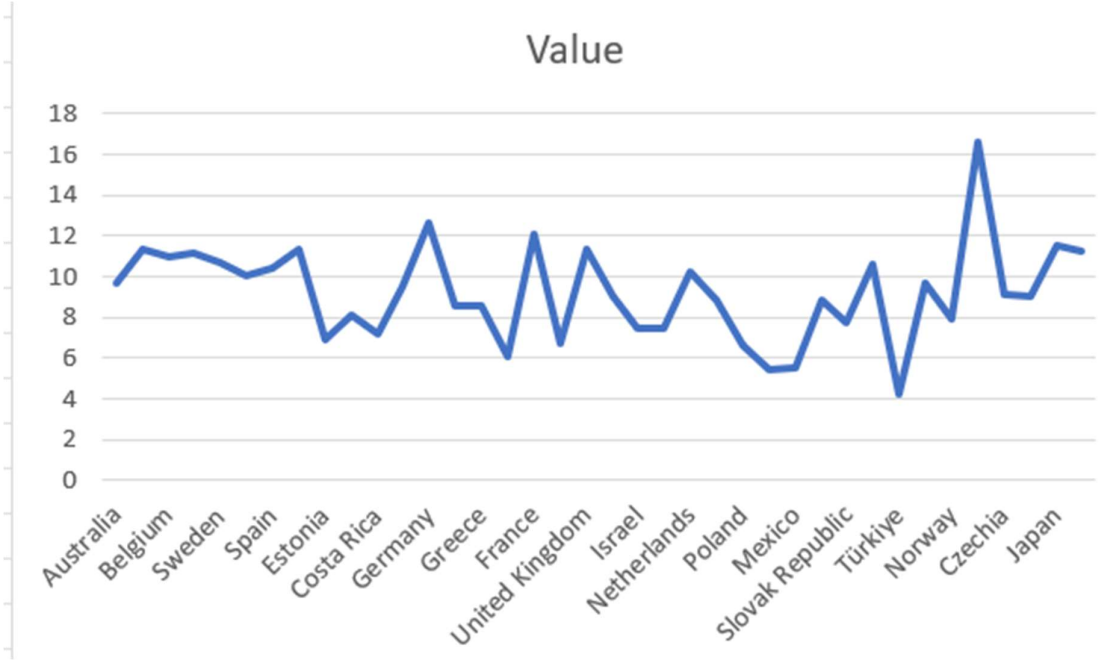

Figure 5. Line graph testing

A line graph was tested with fewer types of data. First comparing the Health Expenditure of OECD countries against each other. This prototype proved to be inadequate in displaying the information in the way we needed it to be conveyed. Since there is no progressive data such as time, the line graph serves less purpose. For this reason, a bar graph was ultimately decided upon.

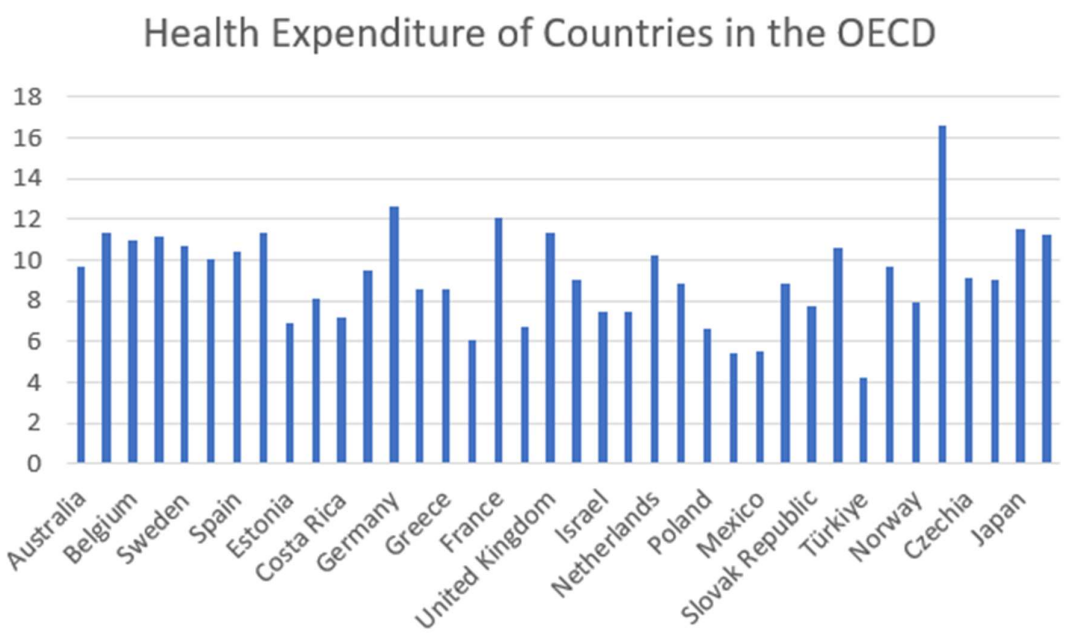

Figure 6. Change line graph to bar chart

The bar graph was far easier to read in terms of how much money each country was spending in the health sector. It is also far easier to discern which countries spend far more and which countries spend far less.

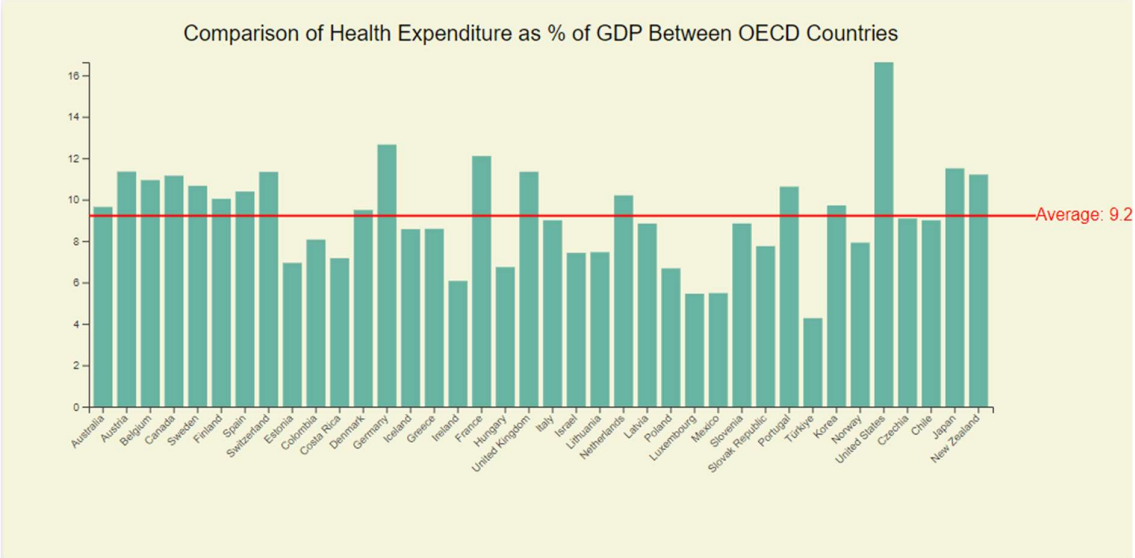

Figure 7. Bar chart result

We updated the style to make it more presentable and added average information.

Visual 3

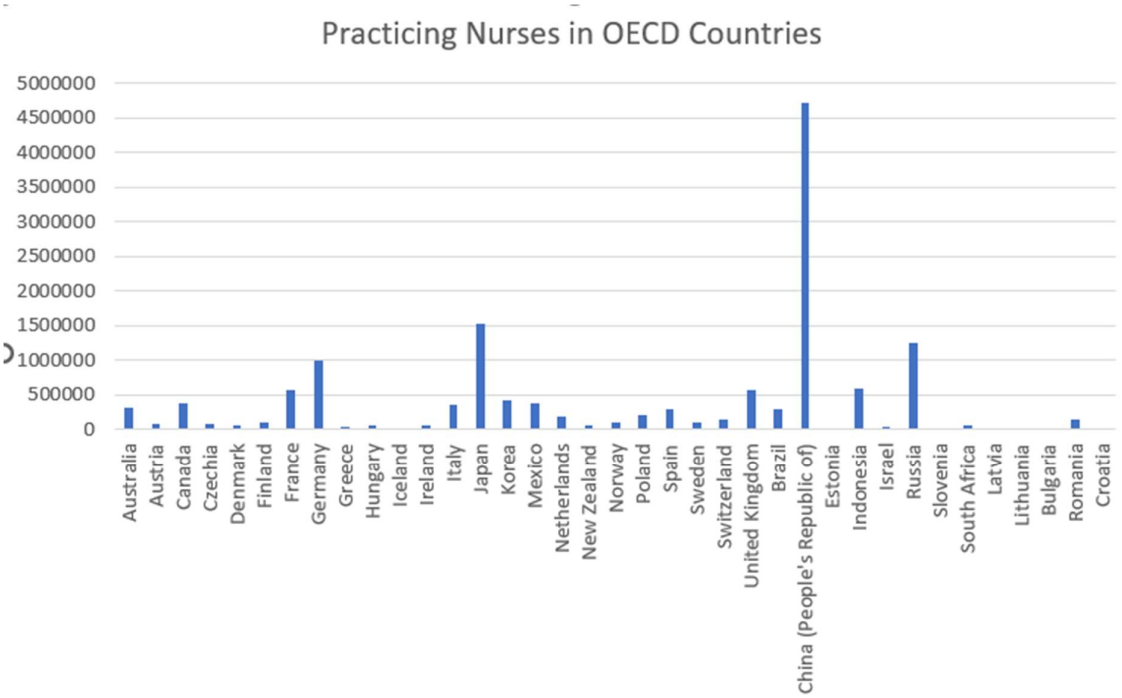

Figure 8. Making Bar Chart for Practicing Nurses

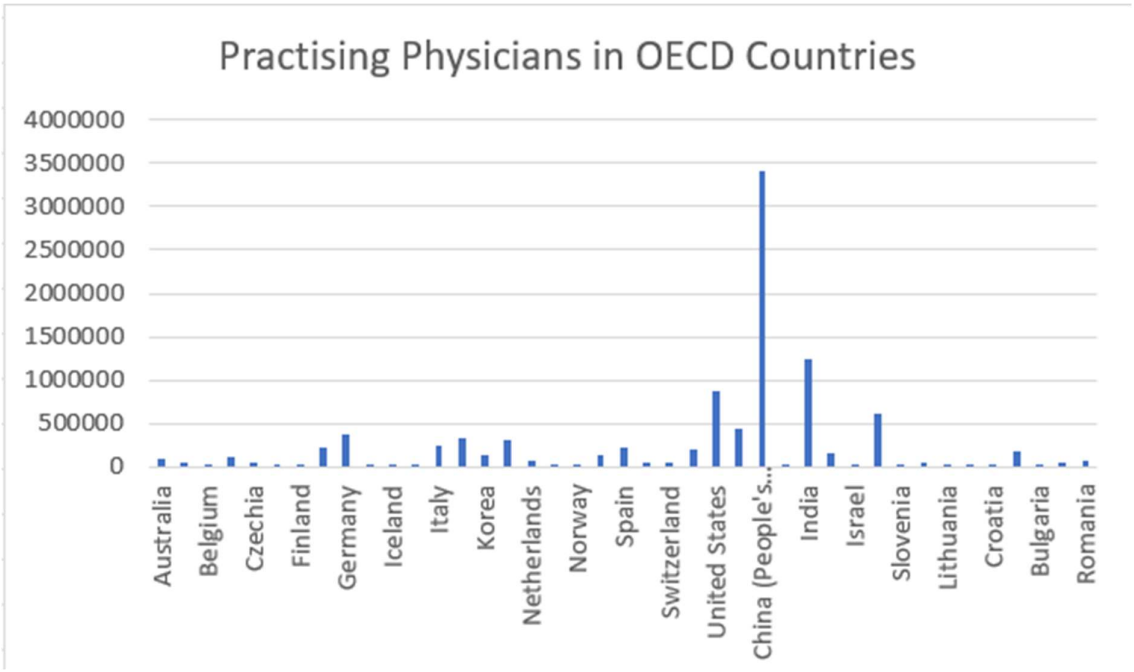

Figure 9. Making Bar Chart for Practising Physicians

After making the bar chart for Health expenditure, the same was done for practicing nurses and physicians. Although the discrepancy caused by the large population in China makes the graph harder to read, it would be manipulative to cut off the graph prematurely simply to

make Australia's number look larger than it is comparatively. It is clear regardless that the size of the population itself plays a role in the number of employees in the health sector.

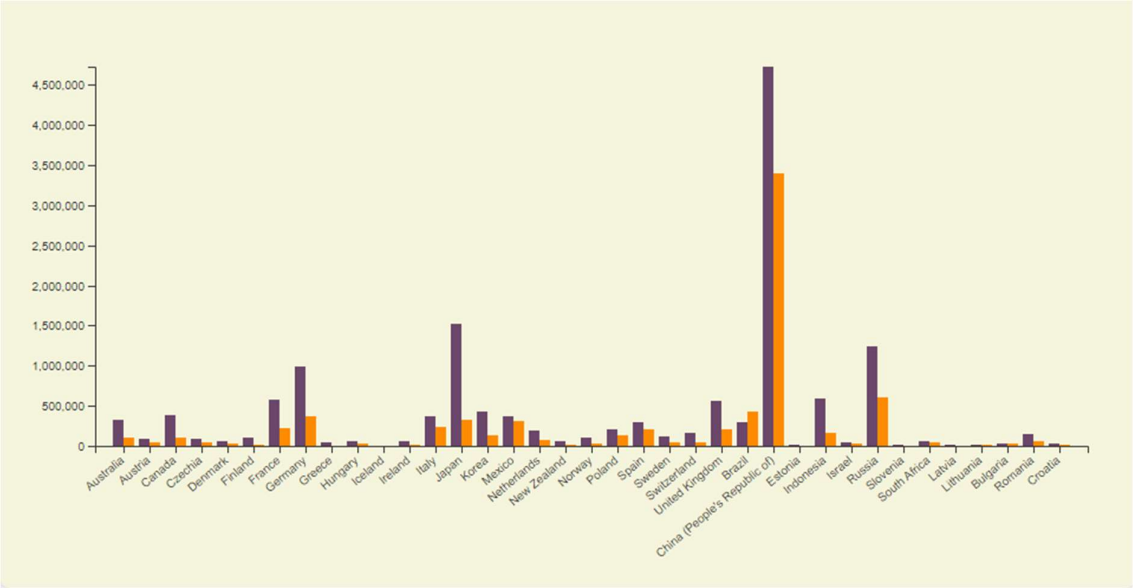

Figure 10. Bar chart result

Showing practicing nurses and physicians in two separate graphs proved to be redundant and simply taking up more space on the webpage. It was decided that both data would be presented in a side-by-side bar chart visualization. This also allows for a comparison on the professions themselves and how much more or less nurses are employed in a country compared to physicians and vice versa.

Visual 4

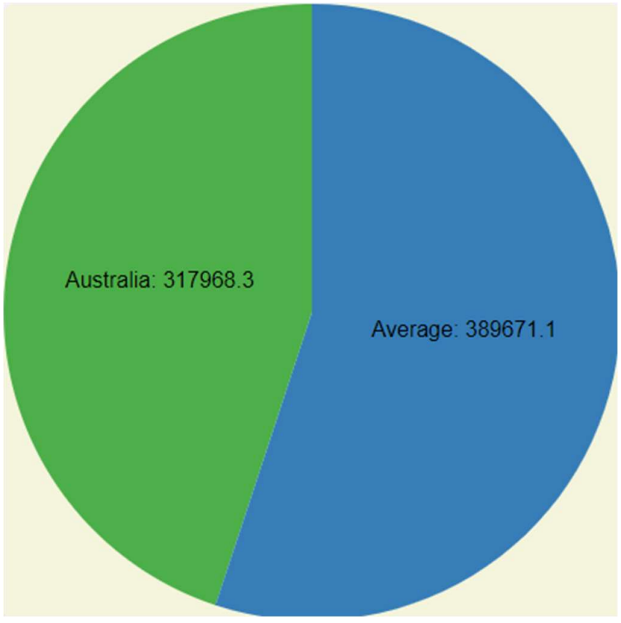

Figure 11. Adding pie chart 1

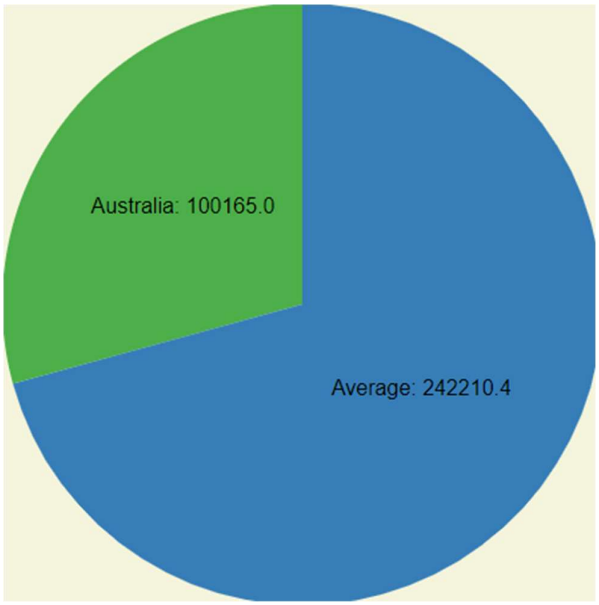

Figure 12. Adding pie chart 2

After implementing the side-by-side bar chart, due to Australia's level of healthcare employment being hard to discern, a separate pie chart was implemented to compare the overall number of healthcare professionals with the OECD average. This is to show that even though the average may be influenced by the outlier that is China, with its extremely immense population, Australia's number of healthcare professionals is still high in comparison.

Visual 5

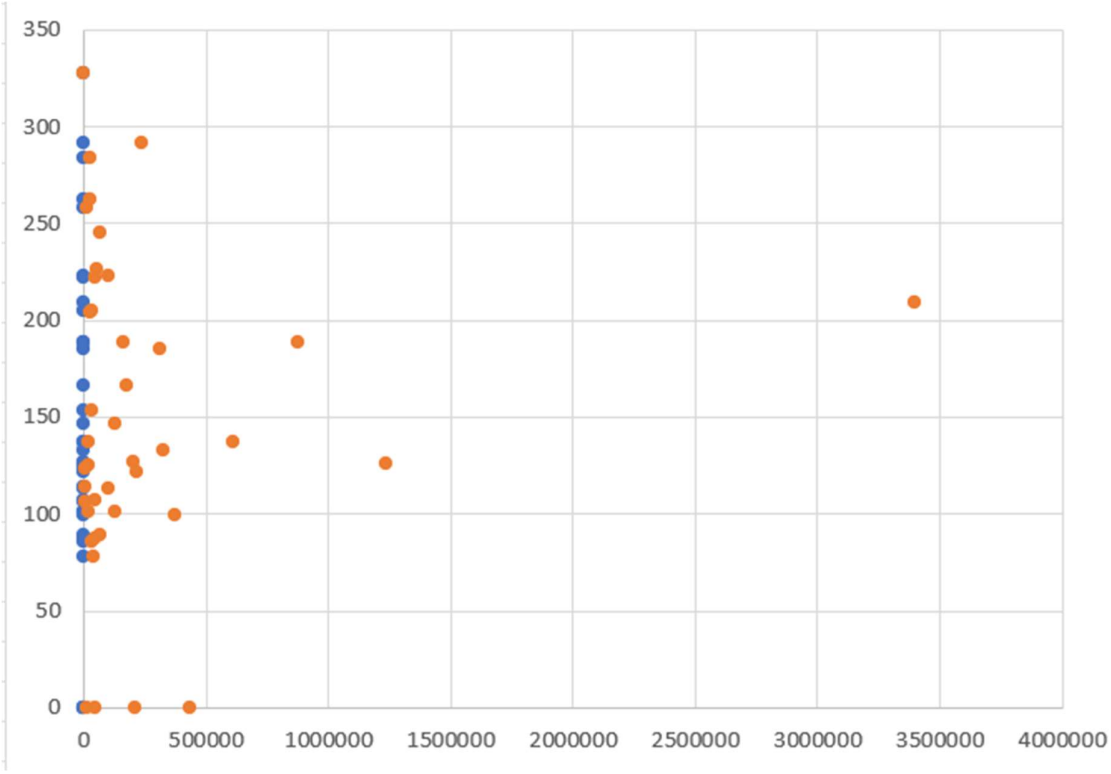

Figure 13. Scatter plot testing

We decided to use a scatter plot to specifically show the correlations between health expenditure, employed professionals in the health sector, and population mortality in terms of preventable and treatable deaths collectively. However, the outliers and certain error values were proving to make the visualization difficult to read and once again the large variety in information was affecting the purpose of the visualization.

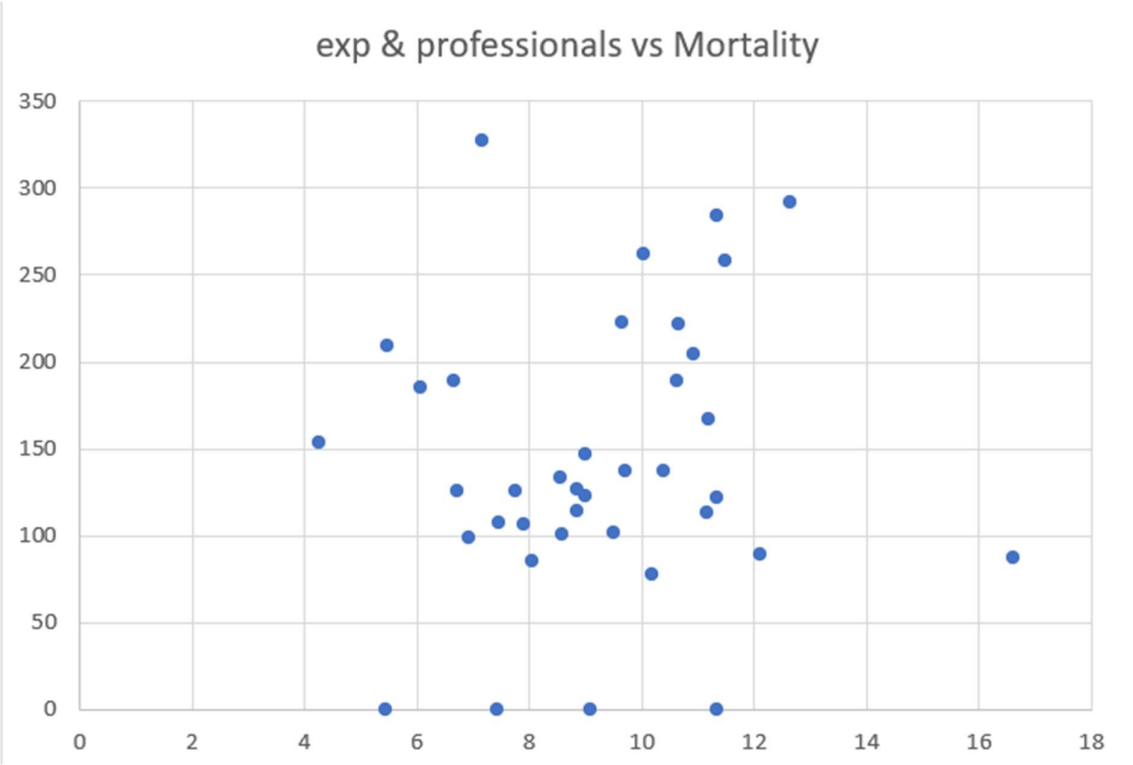

Figure 14. Scatter plot 1

Figure 15. Scatter plot 2

It was far easier to display the data in two separate scatter plots. The correlation between them is still difficult to tell without the aid of a linear regression line. For this purpose, calculations were done and during the execution and implementation of the graph, a line was drawn to show that there is a negative correlation between both health expenditure and mortality as well as number of healthcare professionals and mortality.

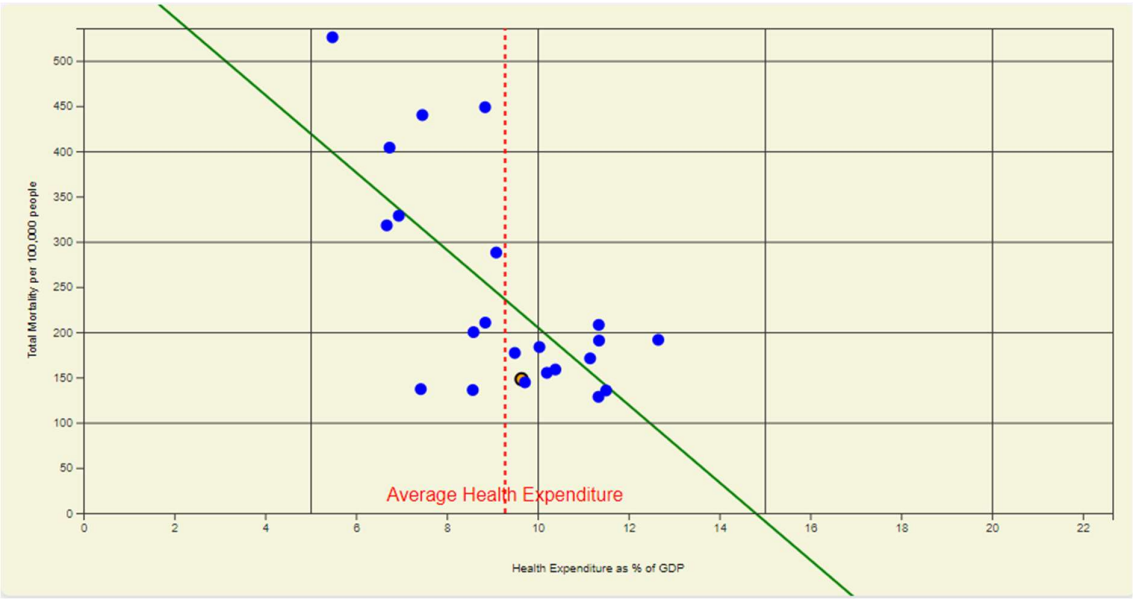

Figure 16. Scatter plot of health expenditure result

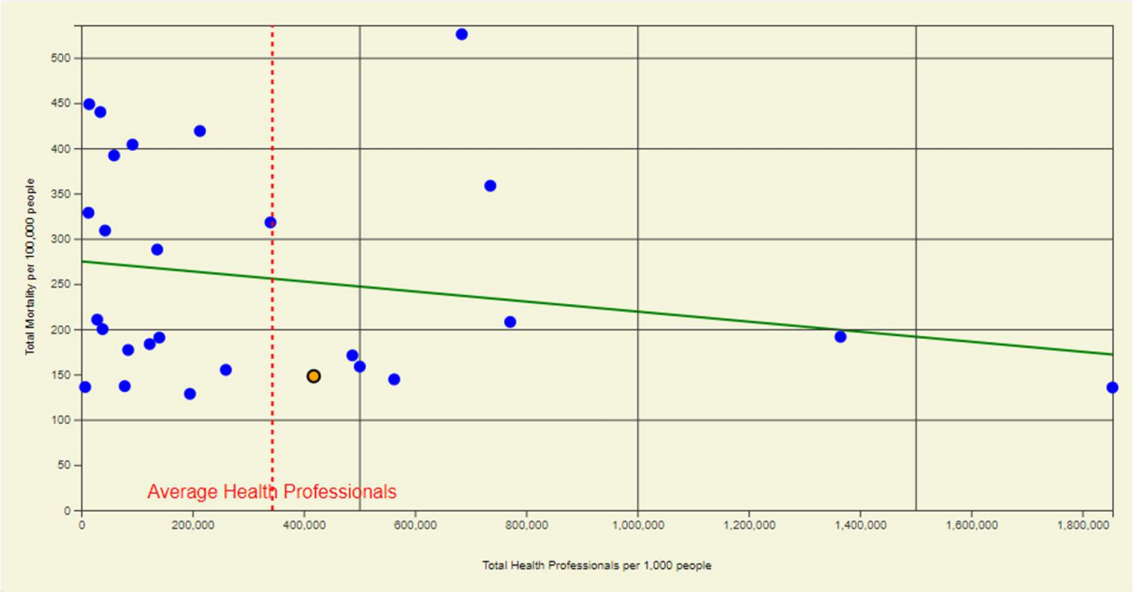

Figure 17. Scatter plot of health professionals result

Once again, a line for the average was drawn to show that Australia is above average in health expenditure as well as healthcare professionals even with a lower population compared to some other powerhouse countries.

Visual 6

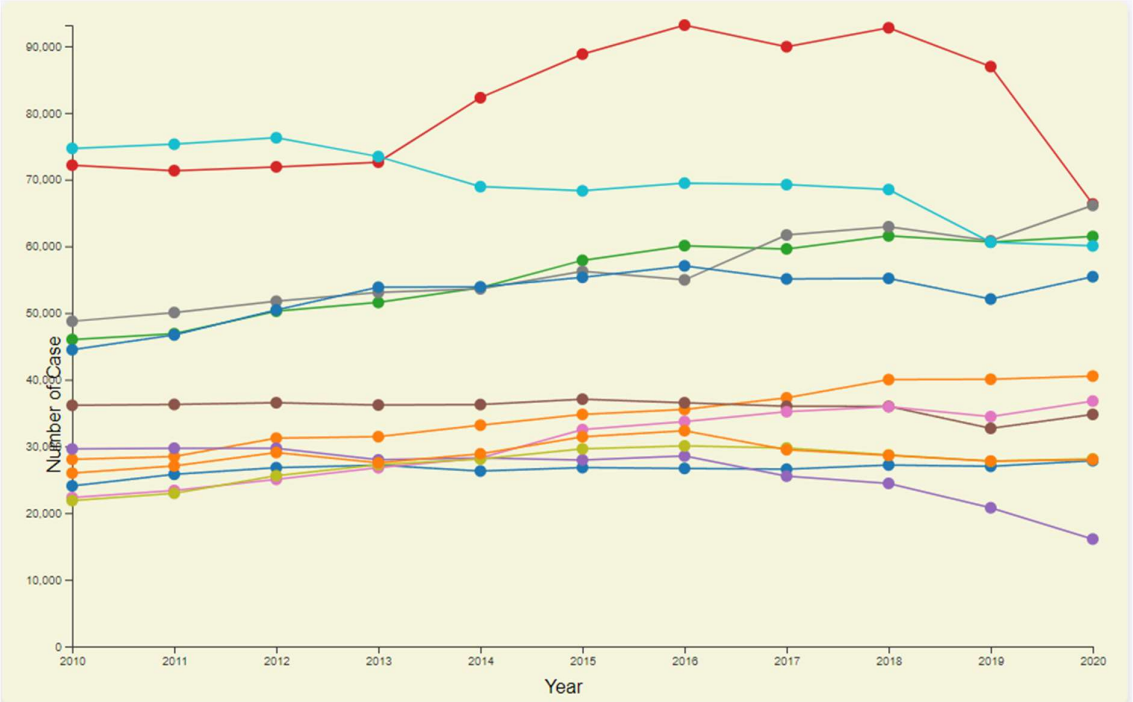

Figure 18. Line graph result

We still wanted to implement a line graph and we did have a dataset that had progressive time data from 2010 to 2020. There was not much of an iterative process to this graph as the visualization worked out well from the start. It is easy to discern which ailments affect the population more as well as how those numbers change through time. Different colours were used between the illnesses to make them easier to differentiate.

Validation

For the validation of our visualizations, we decided to test how effective the visualizations were in terms of how well the data can be interpreted as well as how easily a user can identify certain information or use the data that is presented.

Task scenarios were developed to test both metrics:

- Identify the countries with the highest and lowest global all-cause mortality in 2023.
- Identify the cause of mortality that had the biggest impact on the population.
- Compare the all-cause mortality rates between Canada and New Zealand.
- Identify the country with the highest % of health expenditure.
- Identify how many more nurses are in Australia and Japan compared to physicians.
- Identify how many more healthcare professionals Australia has compared to the OECD average.

The task scenarios were evaluated on:

- How successful the participants were on each task.

- How long do the participants take for each task.
- How subjectively difficult participants found the task. (Ranging from Easy, Moderately Difficult, and Difficult)

Participants for testing:
Due to time constraints, we were unable to have any third-party users evaluate and take part in the task scenarios, so we attempted to evaluate the metrics ourselves.

Results:
1. Identify the countries with the highest and lowest global all-cause mortality in 2023.
Success: 100%
Time taken: 10s
Difficulty: Easy

This was relatively easy to discern as the intensity of the colors is a large aid in deciding which has a higher or lower mortality rate. However, the lack of a legend, showing whether red or orange indicates a low mortality rate created delays in the task.

2. Identify the cause of mortality that had the biggest impact on the population.
Success: 100%
Time taken: 2s
Difficulty: Easy

Although the task was easy and the required data could be discerned fast, if asked to pick out data from a median region in the line graph or compare values, it may prove to be far more time-consuming as the graph can be more difficult to read in the average areas.

3. Compare the all-cause mortality rates between Canada and New Zealand.
Success: 100%
Time taken: 15s
Difficulty: Easy

Visualization provides the necessary information when hovered with a mouse. If the participant already knows the general location of the countries, it is even easier. However, not having the name of the country displayed by default may mean that if it is a country that is between other countries or not as commonly known, it may be difficult to locate.

4. Identify the country with the highest % of health expenditure.
Success: 100%
Time taken: 15s
Difficulty: Easy

The bar graph makes it easy to spot the lowest and highest values. Comparing two values from two different countries may take slightly longer, although it would still be relatively easy as all the information is easy to read and discern. The average is provided as well, which means that any user task related to using the average will be far easier to carry out.

5. Identify how many more nurses are in Australia and Japan compared to physicians.
Success: 50%
Time taken: 90s

Difficulty: Difficult

The bar graphs for most of the countries can be difficult to get the value of as there is no tooltip and must be manually calculated. The graph is also smaller compared to the other visualizations due to China having such a large number of healthcare professionals.

6. Identify how many more healthcare professionals Australia has compared to the OECD average.
Success: 100%
Time taken: 60s
Difficulty: Moderately Difficult

The average is made clear through the graph and Australia itself is also highlighted. This makes the task far easier than initially anticipated, however, the exact value for the average is not displayed, which means that the manual calculation adds time to the task. If it was a country that was not highlighted, it would prove to be far more time-consuming as well, considering that a user would have to go and hover over each scatter point to figure out which country is where.

The validation process proved that the visualizations were extremely effective and efficient in the tasks that were carried out. However, there is an inherent bias due to the participants not being a third party. Having the creator of the visualization, who has already worked with it, carrying out task scenarios will always have more success and they would take less time. For this reason, we must consider a margin of error and assume that the tasks will be slightly more difficult for an external user to carry out.

Conclusion

The project consisted of gathering datasets, analyzing them and identifying the potential story that could be told through the data and subsequent comparisons. A compelling narrative was created to inform users through visualizations created using the data. It was important to have a clear motivation for the website, as the visualizations would have been extremely difficult to flesh out and iteratively improve upon. The purpose for which the visualizations were being constructed proved to be the largest factor in effectively developing them to portray the message that we are trying to communicate.

Through the project and its development, we have learnt important data visualization techniques, data processing and analysis methods, web development skills, web design skills as well as more information about the world through the data that we worked with. Iteratively improving the plans and designs for the visualization provided insight into how data can be portrayed more efficiently to deliver the necessary narrative. The importance of appropriate visualization was also made clear and more apparent through the work done for the website.

Validation of visualizations showed the flaws in the design even in the final stages. It highlighted the importance of task scenarios and how diverse they need to be to appropriately understand how effective visualizations and the website structure itself is.

Overall, this project demonstrated how important appropriate data visualisation is to the understandability and interest of complex information. It also shows how crucial it is to have

a goal and motivation behind the creation of powerful visualisations. Our goal was to develop a tool that would support and guide evidence-based decision-making in the healthcare industry by iteratively improving our method.

References

"Health expenditure Australia 2021-22, Government spending on health relative to government expenses" 2023, *Australian Institute of Health and Welfare*, viewed <https://www.aihw.gov.au/reports/health-welfare-expenditure/health-expenditure-australia-2021-22/contents/overview/government-spending-on-health>.