Swinburne University of Technology

Semester 1, 2023

COS30045- Data Visualization

Impact of COVID-19 on Migration in Australia **Process Book**

**Mercury link:** https://mercury.swin.edu.au/cos30045/s103533868/FinalProject/index.html

**Tutorial session: Monday, 16:30-18:30**

**Authors: Group 8**

Nghia Hieu Pham 103533868

Kim Duong Pham 103792850

Thanh Phat Vo 103792850

**Word Count: 3748 words**

Table of Contents

[1. Introduction 3](#_Toc136814594)

[1.1 Background and Motivation 3](#_Toc136814595)

[1.2 Visualization Purpose 3](#_Toc136814596)

[1.3 Project Schedule 3](#_Toc136814597)

[2. Data 5](#_Toc136814598)

[2.1 Data Source 5](#_Toc136814599)

[Donut Chart: 5](#_Toc136814600)

[Bar Chart (integrated with Donut Chart): 6](#_Toc136814601)

[State Map: 6](#_Toc136814602)

[Lollipop Chart (integrated with State Map) 7](#_Toc136814603)

[2.2 Data Processing 7](#_Toc136814604)

[Donut Chart: 7](#_Toc136814605)

[Bar Chart (integrated with Donut Chart): 8](#_Toc136814606)

[State Map: 9](#_Toc136814607)

[3. Requirements 10](#_Toc136814608)

[3.1 Must-Have Features 10](#_Toc136814609)

[3.2 Optional Features 10](#_Toc136814610)

[4. Visualization Design 11](#_Toc136814611)

[4.1. First Visualization – Choropleth with Bar Chart 11](#_Toc136814612)

[4.2. Second Visualization – Donut Chart with integrated Bar Chart 13](#_Toc136814613)

[5. Validation 18](#_Toc136814614)

[6. Conclusion 24](#_Toc136814615)

[References 25](#_Toc136814616)

[For scholarly ideas: 25](#_Toc136814617)

[For our project: 25](#_Toc136814618)

# 1. Introduction

## 1.1 Background and Motivation

The COVID-19 pandemic had created a significant disruption globally in every aspect of lives. As a result, millions of people lost their jobs, schools are closed, and cities are locked down, altering the socioeconomic flow of the world, and world migration is no better. Between 10 March 2020 and 28 February 2022, 122,823 movement restrictions were implemented around the world.  Assuming zero growth in the number of migrants between 1 March and 1 July 2020, it has been estimated that we have experienced a decrease of 2 million international migrants compared to the expected increase between mid-2019 and mid-2020 (Migration Data Portal 2022).

This visualization should be used by people interested in knowing how COVID-19 impacted Australia’s migration trend, including both experts who wants to gain specific insights on both overall and state specific migration data, or normal users who wish to explore the impact of the pandemic on migration.

## 1.2 Visualization Purpose

Having experienced the pandemic ourselves, our team worked on this visualization project to further investigate the impact of this virus on migration trends, particularly in Australia, overall, and its regions separately, before, during, and after COVID-19 (2017-2022).

The users will be able to identify the overall migration trends by states and through time periods through the first visualization, which makes use of the net migration data in Australia, while the second visualization will help users to identify the specific migration trend that can be referred to socioeconomic conditions in Australia in different periods, with the use of the temporary visas data, therefore referring to existed policies and real-life events to explain the changes in visa categories.

## 1.3 Project Schedule

* Week 2: Initialization
  + Arrange weekly meeting schedule.
  + Create a GitHub repository.
  + Create communication channels (Discord, Facebook)
  + Create a Google Drive folder for sharing documentations.
* Week 3-4: Start working on the project.
  + Complete Section 1 of the Project Process Book
  + Discuss some approaches for visualization.
  + Discuss the progress with tutor in the Project Stand Up 1
* Week 5-6: Data research and analysis
  + Research from multiple sources
  + Process data to be ready for use in the visualizations.
  + Complete Section 2 of the Project Process Book
* Week 7:
  + Draw design sketches for the visualizations.
* Week 8: Start of the implementation phase.
  + Start building the website.
  + Integrate the visualizations into the website.
  + Complete Section 3 of the Project Process Book
* Week 9-11: Improvements & Adjustments
  + Add CSS to style the website.
  + Discuss some improvements for the visualization design in Project Stand Up 2, 3.
  + Implement extra features in the visualization.
  + Complete Sections 4, 5, 6 of the Project Process Book
* Week 12: Project finalization
  + Final feedback from the tutor in Project Stand Up 4
  + Review the source codes to ensure it is well-formatted and commented.
  + Review the Project Process Book to ensure it is properly styled and has no typos or grammatical errors.

# 2. Data

## 2.1 Data Source

### Donut Chart:

**A picture containing text, screenshot, number, font

Description automatically generated**

*Figure 2.1.1: A section of the initial dataset of Pie Chart.*

* **Source:** https://www.abs.gov.au/statistics/people/population/overseas-migration/latest-release#data-downloads
* **Method of data collection:** data is extracted from the Excel spreadsheet from the Australian Bureau of Statistics, containing overseas migrant arrivals and departures data by visa groupings by state or territory and financial year, during the period from 2004 to 2022.
* **Type of dataset:** table in Excel file (.xlsx format)
* **Attributes in the dataset:** categorical (visa types, period) and numerical (migrant arrivals and departures amount)
* **Data excluded:** For this chart, data is extracted from the “Table 2.1” sheet, which contains the data related to overseas migrant arrivals and departures by visa grouping throughout Australia.

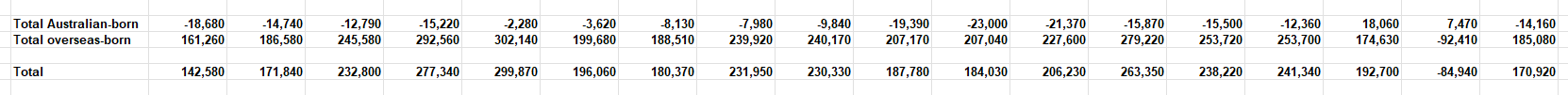
Also, Since the project only evaluates data from 2017 to 2022, any irrelevant period and data related will be excluded from the final dataset. In addition, the chart will visualize the temporary visa types of values, so values belonging to the permanent visa types will be excluded. Also, since the numerical data in the initial dataset represented the number of migrants by category, it needs to go through the transformation process to achieve the desired value (percentage of each category compared to the total amount) to be represented on the Donut Chart.

### A screenshot of a computer Description automatically generated with medium confidenceBar Chart (integrated with Donut Chart):

*Figure 2.1.2: Initial version of Bar Chart Dataset in New South Wales*

* **Source:** https://www.abs.gov.au/statistics/people/population/overseas-migration/latest-release#data-downloads
* **Method of data collection:** Like the Donut Chart, data is extracted from the same excel spreadsheet from the Australian Bureau of Statistics. The difference is that the donut chart uses data for the whole Australia country, and the bar chart uses data for each state in Australia.
* **Type of dataset:** Table in excel file (.xlsx format)
* **Values’ datatypes:** categorical (visa types, period, state name) and numerical (migrant arrivals and departures amount)
* **Data excluded:** Like the Donut Chart, any data which was not from 2017 to 2022, or belonged to the permanent visa types were excluded.

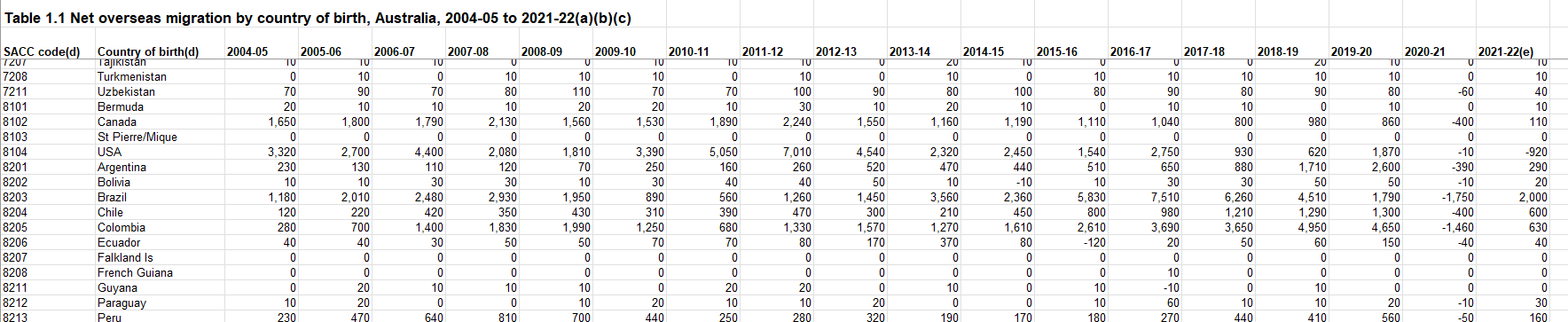
### State Map:



*Figure 2.1.3: A section of initial dataset for State Map*

* **Source:** https://www.abs.gov.au/statistics/people/population/overseas-migration/latest-release#data-downloads
* **Method of data collection:** The data is extracted from the same excel spreadsheet from the Australian Bureau of Statistics. The state map chart using the total net migration of each year data for each state in each year in Australia.
* **Type of dataset:** Table in excel file (.xlsx format)
* **Values’ datatypes:** categorical (countries) and numerical (net overseas migration)
* **Data excluded:** Like the all the other Chart, any data which was not from 2017 to 2022.

### Lollipop Chart (integrated with State Map)



*Figure 2.1.4: A section of the Bar Chart’s initial dataset*

* **Source:** https://www.abs.gov.au/statistics/people/population/overseas-migration/latest-release#data-downloads
* **Method of data collection:** The data is extracted from the same excel spreadsheet from the Australian Bureau of Statistics. Same as state map but instead look at the total net of each state in each year, we look at top 5 countries have the biggest positive values in each year for each state.
* **Type of dataset:** Table in excel file (.xlsx format)
* **Values’ datatypes:** categorical (countries) and numerical (net overseas migration)
* **Data excluded:** Like the all the other Chart, any data which was not from 2017 to 2022, or not in Top 5

## 2.2 Data Processing

Initially, we did expect to deal with a considerable amount of data transformation work, since the dataset is not ready to be used by d3 and is filled with irrelevant text, sheets, and data to be processed as a visualization. From the original dataset, for every visualization we plan to extract the data (categories and values) that lies in the predefined period of before, during, and after COVID19 (2017-2022).

We processed the data by extracting them from the original file, organizing records and exporting those newly extracted data files to the suitable format to be processed. Details of the process will vary by different visualizations, which will be clarified in the upcoming section:

### Donut Chart:

**A screenshot of a spreadsheet

Description automatically generated with medium confidence**

*Figure 2.2.1: Transformed dataset.*

After extracting the desired values from the original dataset to a separate fille, the data used for this chart was acquired by converting the number of temporary visas by type to the percentage of that compared to the total amount of temporary visas, using this formula:

*Percentage of temporary visa type 100%*

After implementing the formula, the result will then be rounded off to 1 decimal place.This process was implemented in Excel by using Excel functions ROUND, and SUM to calculate the percentages values from the original values, without having to calculate each value by ourselves. When all values are processed, the file will be saved in CSV format, and ready to be processed.

### Bar Chart (integrated with Donut Chart):

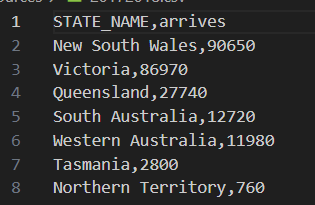
**A screenshot of a computer program

Description automatically generated with medium confidence**The original spreadsheet presents data for each state in a separate table, which is not suitable for the visualization. Specifically, the visualization requires the value of each temporary visa type for all Australia states in a specific year within 2017 to 2021. To achieve this, each table’s data of temporary visa types was extracted and stored in a JSON file.

*Figure 2.2.2: Bar Chart data after processing*

This can also be achieved by creating 5 CSV files (each represent temporary visa data for a specific year within 2017-2021), but it is not as efficient as storing all of them in 1 JSON file.

### State Map:



*Figure 2.2.3: State Heat Map data after processing*

After extracting the desired values from the original dataset to a separate fille, the data used for this chart was acquired by taking the total net migration of each state from 2017-2022. All the data has been calculated in the dataset has been provided.

This can also be achieved by creating 7 CSV files as 7 states.

**Bar Chart (integrated with State Map) ￼**



*Figure 2.2.4: Bar Charts (State Heat Map) data after processing*

Same with above but with this one, we need to select the top 5 countries that have the most net migration of each year, from 2017 to 2022.

This can also be achieved by creating 7 CSV files as 7 states.

# 3. Requirements

## 3.1 Must-Have Features

The visualizations in the project must include these features:

* **Title:** The name of the visualization (what it is, and what it is about)
* **Axes & Labels:** Help clarify data values used in the visualizations.
* **Legends:** List of labels to clarify each color used in the visualization
* **Color:** Simple, contrast color scheme to distinguish different data values
* **Interactivity:**
  + **Buttons**: Allow users to choose period of the data (within 2017 – 2021)
  + **Hovering effects**: When hovered, the element will be uniquely styled to distinguish with other elements.
  + **Tooltips:** Display the exact value of the hovered element
* **Transition:** Smoothly displays the changes happening to the visualization
* **Brief explanations:** The visualization may be complex for some people. Therefore, it is necessary to include a brief explanation about how to use and the context around the visualizations.
* **Charts:** 
  + **Pie chart:** Able to display data in correct portions.
  + **Bar chart:** Able to update the data according to the selected visa type and periods.
  + **Heatmap:** Able to display data in the correct state position's
  + **Bar chart (Heat Map): A**ble to update the data according to the selected state and year.

## Optional Features

The following features are nice to have but not mandatory:

* **Feedback Form:** A form allowing users to submit their feedback on visualizations.
* **Zoom function:** A function allowing users to zoom in/out the visualization which will help the user to see sections clearer at will.
* **Animation:** Include animations to smoothen the data-changing process in the visualizations.

# 4. Visualization Design

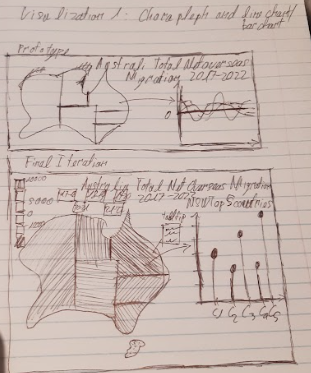
### 4.1. First Visualization – Choropleth with Bar Chart

#### 4.1.1. General ideas

By using Choropleth, we can show the total net overseas migration from 2017 - 2022 for each state, using a combination of buttons and mouse over, we can illustrate our data in each year and each state with just a single click and look at the color, for the specific data, they can just mouse hover the state to see it. Allow the user to interact more with the chart. If the user wants to know more about each state in that specific year, they can simply click on that state. It will show the top 5 countries that contribute most to the net migration of that state in that year as a bar chart.

#### 4.1.2. Design evolution

First, we produce a visualization that uses a Background map of states just to interact with then show the actual data in the line chart one the side. How Think about how much data we need to include I single chart, let the other do nothing, and make confusion. Instead, we produce an ideal that change the background map to choropleth map and make it show the total net migration and change the other line chart to the BarCharts and show the top 5 countries to support the heat map, make both charts meaningful. Plus, we, add button change the choropleth map, so we



*Figure 4.1.2.1: Final Iteration of Choropleth map and bar chart*

#### A picture containing background pattern Description automatically generated

#### 4.1.3. Alternative ideas

- Before producing the final iteration, we thought about using 2 different charts to provide information and interaction by using a button and a mouse. Still, then we consider how to make the visualization provide meaningful data for the user in the easy was too much and take more time to understand the visualization. So, a better way is using one to interact with the other is a better.

*Figure 4.1.2.2: Another ideal of Choropleth map and bar chart*

#### 4.1.4. Visual encoding and idioms

* The Idiom of this visualization was combining combination of a choropleth and bar chart. In these ways, a lot of information can be provided to the reader without reading, they can simply understand the visualization just by looking at it. By using choropleth and button, we can show data of each state at each year and using the bar chart to show the top 5 countries that contribute to the net. Both charts support each other to provide the full meaning of the visualization. And the difference between year support by button, it shows the impact of the covid-19 to the net of migration to Australia.
* In addition, in this visualization, color hues and saturations on each state to clarify the difference between total net overseas migration, while improving the overall aesthetic, visibility on the website (as it does not conflict with the website colors).
* We also provide a highlight went hover the state by using the thickness of lines with colors saturation around the state border and blur other state by decreasing the capacity.

### 4.2. Second Visualization – Donut Chart with integrated Bar Chart

#### 4.2.1. General ideas

The pie chart will show the proportions of temporary visa types in Australia in different periods, which will allow the user to analyze the change of values by time, during the period of 2017-2022. The specific percentage values will be shown to the users by hovering through the desired slices. Moreover, the user can zoom in the pie chart and drag at will to see the difference between pie slices or indicate a minor section in the bar chart.

To further analyze the data trend by states, a bar chart is also integrated, which show the specific number of temporary visas granted for specific states in Australia. This chart will change according to the user’s hover action, so that when the user hovers the mouse over a specific visa type in a period, the bar chart will display the number of visas granted that belongs to that type in every state, and in the same period as the pie chart.

While the pie chart was implemented with a color scheme with different hues and saturation to clarify the different visa types, the bar chart uses assorted color saturations to clarify the difference in quantities.

#### 4.2.2. Design evolution

A piece of paper with writing on it

Description automatically generated with low confidence

*Figure 4.2.2.1: Second Visualization’s Iteration 1*

A close-up of a graph

Description automatically generated with low confidence

*Figure 4.2.2.2: Second Visualization’s Final Iteration*

* The first design simply includes a donut chart with legends to help identify the color of each visa type in the donut. There are also buttons that allow users to select year, and the donut is updated when a button is clicked. It is nice but quite simple.
* The final design includes a side bar graph which interacts with the pie chart. Specifically, whenever users hover over a slice of the donut chart (bound to a visa type) to, the bar graph will update its data to show the number of that visa type grouped by Australia states. In addition, the titles of both graphs are updated if users select different data.

#### 4.2.3. Alternative ideas

While producing the final design, we have also considered alternatives to visualize our data, such as a stacked bar chart, side bar chart, and the pie charts, integrated with parallel chart.

A piece of paper with writing on it

Description automatically generated with medium confidence

*Figure 4.2.3.1: Stacked bar chart design*

The stacked bar chart will allow us to show the weight of various categories, in comparison to the overall value, by multiple categories. However, as we wanted more complexity and effectiveness for the visualization, this option was not choses since it is simple and did not have the ability to extend the data range further, such as specific States's number of immigrants by visa types.

A notebook with writing on it

Description automatically generated with low confidence

*Figure 4.1.3.2: Side bar chart design*

The number of migrants for different granted temporary visa types can also be represented with a side bar chart, however, it did not offer the capability to represent more context and depth of data, same as the previous option. As a result, we also did not choose this to be our final implementation and move on with another attempt to find the suitable visualization for our needs.

-A picture containing text, handwriting, number, notebook

Description automatically generated

*Figure 4.1.3.3: Pie chart with integrated parallel chart design*

This alternative offers the integrated chart type that we aim for, with multiple potential interactivity and data representation, however, we thought that the parallel chart might not be a suitable representation for our selected data, as it may provide unneeded complexity for the users to understand and extract information from the visualization. In addition, we thought that compared to the simple bar chart representation, this chart is not as effective and “information-dense” as its peer.

#### 4.1.4. Visual encoding and idioms

The idioms used in this visualization are donut (pie) charts and bar charts. These two idioms are integrated with each other to fully maximize the properties of each type. The goal for this visualization is to not only displays the temporary visas data in Australia, but also to visualize the data of that visa type in different states of Australia, allowing the user to have an overall view of the problem domain, as well as assessing and evaluating the data in a wider picture.

In addition, in this visualization, color hues and saturations for different visa types were used on the donut chart to clarify the difference between categories, while improving the overall aesthetic, visibility on the website (as it does not conflict with the website colors).

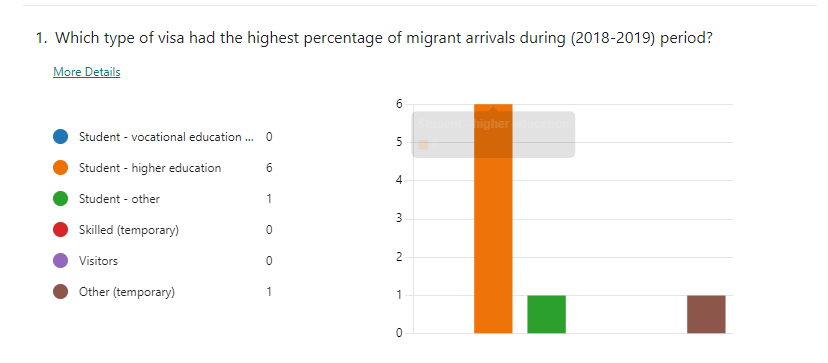
For the bar chart, color saturations were used to display the numerical data, which helps the user to identify the difference of the number of visas between states more clearly and effectively.

Also, we have grouped the change of pie slices with the bar chart, so that whenever the user hovers through a pie slice of a particular time period, the bar chart (along with its axes) will change accordingly to match the period and the temporary visa category in a short amount of time, creating a smooth data transition for the user to interact and explore with the visualization while improving the user’s visual experience and satisfaction.

# 5. Validation

To gain more insights into the visualizations we created, we have included a survey for users to test the website’s usability and the user’s experience. The survey included several questions related to the information that can be extracted from the website, as well as the evaluation of the users’ experience with the website and their recommendations to improve the website further in the future.

We have conducted the survey and collected 8 responses:



*Figure 5.1: First question of the survey*

A picture containing text, screenshot, software, plot

Description automatically generated

*Figure 5.2: Second question of the survey*

A screenshot of a graph

Description automatically generated with low confidence

*Figure 5.3: Third question of the survey*

A picture containing text, software, colorfulness, plot

Description automatically generated

*Figure 5.4: Fourth question of the survey*

A screenshot of a test

Description automatically generated with low confidence

*Figure 5.5: Fifth question of the survey*

A screenshot of a graph

Description automatically generated with low confidence

*Figure 5.6: Sixth question of the survey*

A screenshot of a graph

Description automatically generated with low confidence

*Figure 5.7: Seventh question of the survey*

A picture containing text, screenshot, diagram, line

Description automatically generated

*Figure 5.8: Eighth question of the survey*

A screenshot of a survey

Description automatically generated with medium confidence

*Figure 5.9: Ninth question of the survey*

A screenshot of a survey

Description automatically generated with medium confidence

*Figure 5.10: Tenth question of the survey*

From the collected data, our website has received positive responses from users about their experience. Most users managed to get the correct answer for the data-related questions. In addition, the results suggest that the website is user friendly to most users, and the visualizations interactivity, as well as the website’s aesthetics are appealing and improved the users’ experience.

The text feedback suggests that the chart with integration of another chart is appealing to users, as well as the color scheme selected and the hover interaction. However, some suggest that the website's overall aesthetic and layout can be improved with more pictures or organization of contents.

# 6. Conclusion

In conclusion, through the implementation of the website using D3.js library, and the development of the visualization project about Migration to Australia, with respect to the recent COVID-19 pandemic, our group has gained critical knowledge on utilizing the D3.js library, and on understanding and applying visualization concepts to build graphs using real data. Moreover, as the project was designed to be like a significant business project, we have gained invaluable experience on working as a team collaborating on GitHub, documenting progress, and reporting it every project stand up, and testing, debugging the final product. We believe those experiences are beneficial to our career path.

# References

## For scholarly ideas:

Australian Bureau of Statistics, Overseas Migration, Australian Bureau of Statistics, viewed 3 June 2023, <<https://www.abs.gov.au/statistics/people/population/overseas-migration/latest-release?fbclid=IwAR2OYGNIUeyPvTDCCul5qXju-TH2_9ywwTkwqSvXiETLkYtm_4A9hsS48rg#data-downloads>>.

Migration Data Portal 2022, Migration data relevant for the COVID-19 pandemic, Migration Data Portal, viewed 1 June 2023, < <https://www.migrationdataportal.org/themes/migration-data-relevant-covid-19-pandemic>>.

## For our project:

<https://gist.github.com/dbuezas/9306799>: used to learn how to build the donut chart with a data transition animation.

<https://www.youtube.com/watch?v=6HFpw5fcaD8>: used to learn how to build the navigation bar on the webpage.

<https://youtu.be/oYRda7UtuhA>: used to learn how to build the website with proper first page and sections.

<https://medium.com/@kj_schmidt/making-an-animated-donut-chart-with-d3-js-17751fde4679>: used to learn how to build the donut chart and tooltips.

<https://stackoverflow.com/questions/38971140/how-to-use-zooming-with-pie-in-d3-charts>: used to learn how to implement the zooming function.

<https://d3-graph-gallery.com/graph/line_change_data.html>: used to learn how to change data in graphs.

<https://raw.githubusercontent.com/tonywr71/GeoJson-Data/master/australian-states.json>: for Australia GeoMap