Australian Life Expectancy Process Book

COS30045 - Data Visualisation
Mercury Host Link -
https://mercury.swin.edu.au/cos30045/s103798793/Project%20Final/Life%20Expectancy%20of%20Australia.html

By Jed Pagtakhan (103068823) and Hridoy Ahmed (103798793)
Tutorial Dates: **Mondays 10:30 AM-12:30 PM EN305**
Due Date: **7th of June 2024**
Year & Semester: **2024 Semester**
Word count: **6287 Words**

Table of Contents

Introduction

This data visualisation project aims to develop an interactive web-based platform that represents life expectancy trends in Australia. The visualisation is segmented by gender and examines various factors contributing to life expectancy, such as health spending and cancer mortality rates. By leveraging the data visualisation capabilities of D3.js, users can explore differences and trends across Australia's states and territories. The primary objective is to provide insightful and accessible visualisations to understand how life expectancy has evolved over the years and the multifaceted influences affecting it. This interactive tool is intended for public health officials, policymakers, researchers, and the general public, offering a comprehensive view of life expectancy trends and their determinants in Australia.

Background and Motivation

Understanding life expectancy trends and the factors that influence them is essential for public health officials, policymakers, researchers, and the public. Life expectancy serves as a critical indicator of a population's overall health and well-being. By analysing these trends, we can derive insights into the effectiveness of health policies, identify areas requiring intervention, and comprehend the impact of various socioeconomic factors on public health outcomes.

This project focuses on developing an interactive web-based visualisation that showcases life expectancy trends in Australia, segmented by gender and various contributing factors such as health spending and cancer mortality rates. Our primary objective is to leverage D3.js's powerful data visualisation capabilities to allow users to explore and understand differences and trends in life expectancy across Australia's states and territories. This interactive visualisation aims to provide a comprehensive view of how life expectancy has evolved over the years and the multifaceted influences affecting it.

Visualisation Purpose

- **Educate** a diverse audience about the history and trends in life expectancy across different times in history in Australia, giving the audiences a better understanding of public health outcomes over time.
- **Support** for other countries by providing a clear visual tool that highlights the factors that contribute to a greater life expectancy for the general population and helps guide healthcare decisions in the future.
- **Promote** Public Engagement by making complex data accessible and interactive, encouraging the public to explore how health trends affect their communities and understand them more easily.
- **Facilitate** Research by offering a resource that shows and analyses historical life expectancy data and different factors that contribute to it to help aid health and demographic studies

Data

Data Source (Proposal)

When conducting our research regarding which data sets we were going to be using, we made sure that we were going to be using at least one OECD health dataset to make sure that we stayed within the scope of the project. For this, we decided to filter through all the available datasets on the OECD website regarding factors that could relate to a higher/lower life expectancy in Australia. Below are some of the datasets that we found and were deciding to use for the project.

1. **Health Spending Dataset:** This data set that we found on the OECD website was related to the health spending throughout the history of each country per capita. We found this dataset useful and attractive as when we evaluated the datasets in tables, we can see that there is a clear trend of life health spending per person in each and advancing year. As we filtered out all the countries and only put on Australia this was made even clearer. This dataset essentially showed us that each person was spending more and more money on themselves either through appointments, doctor checkups or medications throughout history. Pairing this with another data set that we had in mind, we figured that this was a good data set to use and thus kept it to advance to the final dataset application.

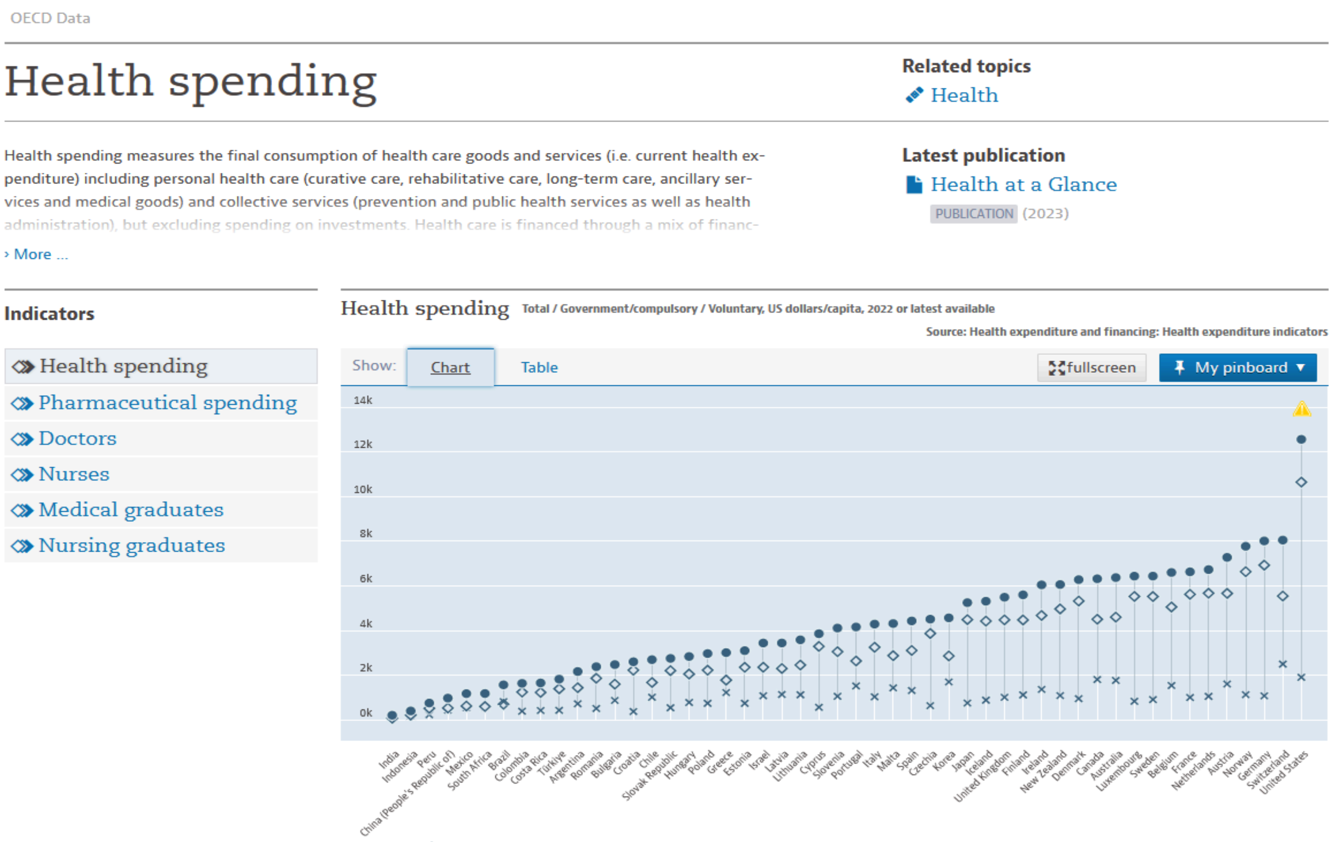


Figure 1 - OECD Health Spending Data

2. **OECD Life Expectancy table/dataset:** The dataset in question that was mentioned above was regarding the OECD's data set of the average life expectancy of each person in Australia dating back from 1980. This dataset not only showed the life expectancy of Australians but went into greater detail by also showing the two genders' life expectancy as the years went on. This data shows the average male, female and both average life expectancy from 1980 to 2022. Our whole project completely revolves around the idea of life expectancy and how we deem to explore it and research it. This data is undoubtedly necessary regarding our final visualisation implementation

Applied filters 4 Reference area: × Australia × | Age: × 0 years × | Frequency of observation: Annual | Time period: × Start: 1980 ×

Clear all filters ×

209 data points

Overview Table Chart Labels Layout Share Download Developer API Full screen

Life expectancy

Reference area: Australia ● ***Frequency of observation:*** Annual ● ***Age:*** 0 years

Unit of measure: Years

Time period	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Measure																				
Sex: Total																				
Life expectancy	..	74.8	74.6	75.4	75.7	75.5	76	76.2	76.2	76.4	76.9	77.3	77.4	77.9	77.9	77.8	78.1	78.4	78.6	78.9
Life expectancy difference (female-male)	7.1	7	7	6.7	6.5	6.4	6.3	6.4	6.4	6.3	6.2	6	5.9	5.9	5.9	5.8	5.9	5.7	5.6	5.6
Life expectancy difference (male-female)	-7.1	-7	-7	-6.7	-6.5	-6.4	-6.3	-6.4	-6.4	-6.3	-6.2	-6	-5.9	-5.9	-5.9	-5.8	-5.9	-5.7	-5.6	-5.6
Sex: Female																				
Life expectancy	78.1	78.4	78.2	78.8	79	78.8	79.2	79.5	79.5	79.6	80.1	80.4	80.4	80.9	80.9	80.8	81.1	81.3	81.5	81.8
Sex: Male																				
Life expectancy	71	71.4	71.2	72.1	72.5	72.4	72.9	73.1	73.1	73.3	73.9	74.4	74.5	75	75	75	75.2	75.6	75.9	76.2

Time period	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Measure																				
Sex: Total																				
Life expectancy	79.2	79.6	79.9	80.2	80.5	80.8	81	81.3	81.4	81.5	81.7	81.9	82	82.1	82.3	82.4	82.4	82.5	82.7	82.9
Life expectancy difference (female-male)	5.4	5.4	5.2	5	4.9	4.8	4.8	4.7	4.5	4.6	4.5	4.5	4.4	4.2	4.1	4.1	4.2	4.1	4.2	4.1
Life expectancy difference (male-female)	-5.4	-5.4	-5.2	-5	-4.9	-4.8	-4.8	-4.7	-4.5	-4.6	-4.5	-4.5	-4.4	-4.2	-4.1	-4.1	-4.2	-4.1	-4.2	-4.1
Sex: Female																				
Life expectancy	82	82.4	82.6	82.8	83	83.3	83.5	83.7	83.7	83.9	84	84.2	84.3	84.3	84.4	84.5	84.6	84.6	84.9	85
Sex: Male																				
Life expectancy	76.6	77	77.4	77.8	78.1	78.5	78.7	79	79.2	79.3	79.5	79.7	79.9	80.1	80.3	80.4	80.4	80.5	80.7	80.9

© Life expectancy OECD

Figure 2 - OECD Life Expectancy Data

3. **Vaccine rates for children:** Another potential dataset that we explored was the dataset of immunisations for all children including Aboriginals and Torres strait islander children coming from different ages. This was supposed to be a supplementary dataset as it did not come from the OECD but its main purpose was to complement the main datasets we had. This was a potential choice as having immunisations for children is another contributing factor to why life expectancy could go up. However, we thought this may not be relatable enough as the data only goes back to 1999 and we deemed that there were more important factors that could relate to why life expectancy is rising in Australia.

Figure | Data table

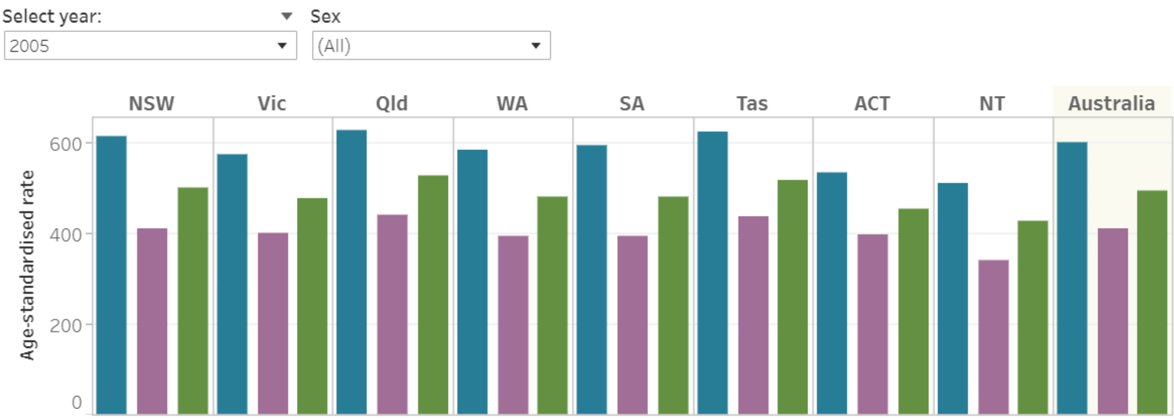
Figure 1: Immunisation coverage rates for all children and Aboriginal and Torres Strait Islander children aged 1, 2 and 5, 1999 to 2021

	Immunisation and vaccination					
	All children			Indigenous children		
Year	1 year	2 years	5 years	1 year	2 years	5 years
1999	86.2	73.6				
2000	89.2	81.8				
2001	91.1	86.6				
2002	90.9	88.5				
2003	91.3	89.8		84.7	86.5	
2004	91.1	91.5		84.4	90.4	
2005	90.9	91.9	74.4	84.4	90.7	
2006	90.7	92.3	79.8	84.0	91.3	
2007	91.2	92.5	83.7	84.3	90.6	78.0
2008	91.2	92.7	79.4	84.1	91.1	76.8
2009	91.7	92.5	82.0	84.9	90.3	77.8
2010	91.5	92.5	88.0	84.8	91.9	83.6
2011	91.5	92.7	89.5	85.0	92.5	85.8
2012	91.7	92.6	90.8	85.4	92.2	89.8
2013	90.7	92.2	91.9	85.6	91.7	93.1
2014	90.7	91.2	92.1	87.0	90.6	93.2
2015	92.3	89.3	92.6	88.7	86.2	93.9
2016	93.4	91.4	93.2	91.2	89.1	95.2
2017	94.0	90.5	94.0	92.2	88.2	96.2
2018	94.0	90.8	94.7	92.6	88.2	96.7
2019	94.3	91.6	94.8	92.6	90.0	97.0
2020	94.9	92.6	95.1	93.8	91.4	97.3
2021	94.6	92.6	95.0	92.4	91.4	96.8

Source: Department of Health 2022a.
http://www.aihw.gov.au/

Figure 3 - Vaccination Rates History for Children Data

4. Cancer Deaths in Australia:



Age-standardised rates (ASR) are expressed as per 100,000 population.

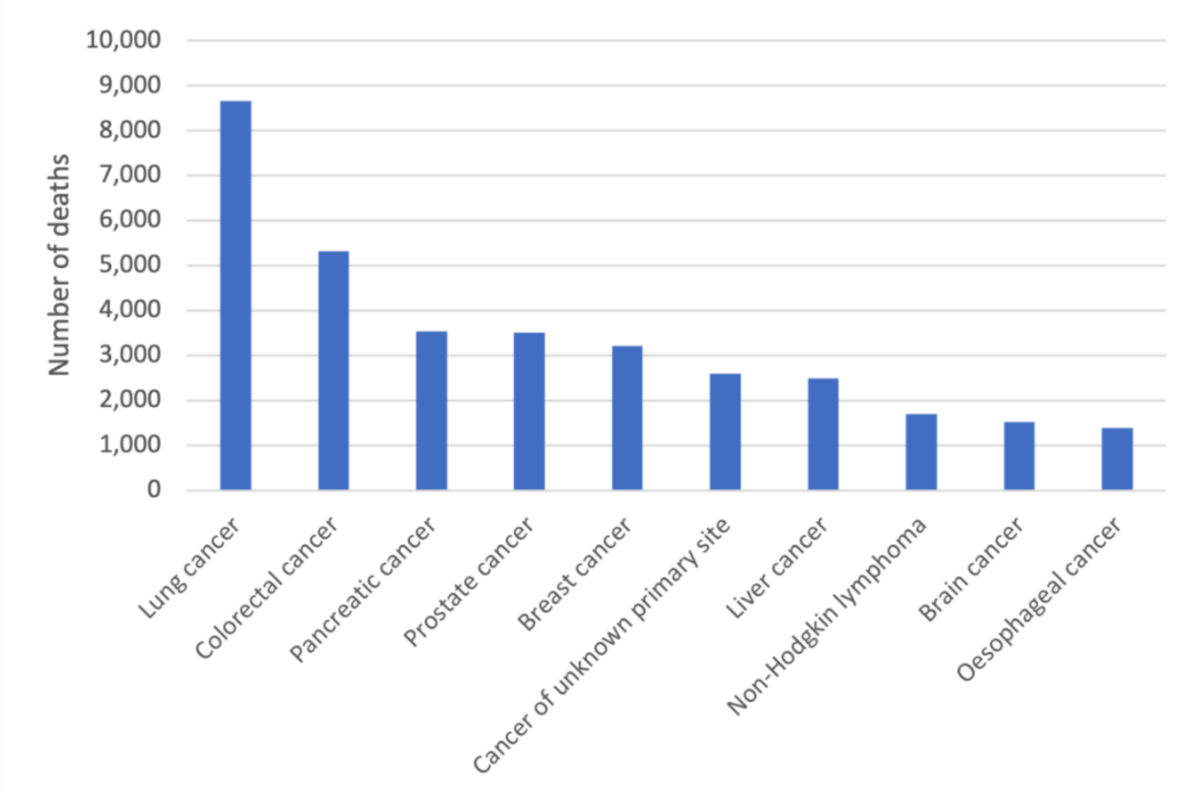


Figure 4- Cancer Deaths in Australia

The visual representation shows the age-standardised incidence rates and the deaths for all cancers combined, segmented by sex, state, and territory for the year 2005. The data, expressed as age-standardised rates (ASR) per 100,000 population, allows for accurate comparisons by eliminating age distribution differences across regions. This comprehensive overview includes both yearly breakdowns and a focus on 2005.

For our visualisation website, we used this data to represent cancer incidence rates, segmented by sex, state, and territory. Instead of yearly data, we averaged the data in five-year intervals (e.g., 2000-2005, 2005-2010, up to 2020) to smooth out yearly variations and highlight longer-term trends. This approach provides a clearer and more concise overview, making it easier for users to understand cancer incidence trends and regional differences in Australia(Cancer Data in Australia, Cancer by State and Territory Data Visualisation, 2023).

Data Source (Final)

We chose to use the OECD's datasets on average life expectancy and health spending per capita for Australia because they are directly relevant to our project's focus on life expectancy. The life expectancy dataset includes detailed information for both males and females from 1980 to 2022, giving us a comprehensive view of how life expectancy has changed over time. This allows us to see long-term trends and patterns in the health of Australians. The health spending dataset is equally important as it shows how much money has been spent on healthcare yearly in each territory. By examining this data, we can see how increasing health expenditures may be related to changes in life expectancy. Using these two datasets together, we can explore whether higher health spending has contributed to longer lives and better health outcomes. This combination provides a strong foundation for our final visualisations and helps us understand the broader picture of health and longevity in Australia. In addition to this, we have also chosen to continue with the dataset regarding cancer deaths in Australia as it is one of the leading causes of death in Australia and deemed it to be useful as a supplementary visualisation to show cancer death trends over time alongside life expectancy.

Applied filters 4 Reference area: × Australia × | Age: × 0 years × | Frequency of observation: Annual | Time period: × Start: 1980 ×
Clear all filters ×
209 data points

Overview Table Chart Labels Layout Share Download Developer API Full screen

Life expectancy
Reference area: Australia • ***Frequency of observation:*** Annual • ***Age:*** 0 years
Unit of measure: Years

Time period	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Measure																				
Sex: Total																				
Life expectancy	..	74.8	74.6	75.4	75.7	75.5	76	76.2	76.2	76.4	76.9	77.3	77.4	77.9	77.9	77.8	78.1	78.4	78.6	78.9
Life expectancy difference (female-male)	7.1	7	7	6.7	6.5	6.4	6.3	6.4	6.4	6.3	6.2	6	5.9	5.9	5.9	5.8	5.9	5.7	5.6	5.6
Life expectancy difference (male-female)	-7.1	-7	-7	-6.7	-6.5	-6.4	-6.3	-6.4	-6.4	-6.3	-6.2	-6	-5.9	-5.9	-5.9	-5.8	-5.9	-5.7	-5.6	-5.6
Sex: Female																				
Life expectancy	78.1	78.4	78.2	78.8	79	78.8	79.2	79.5	79.5	79.6	80.1	80.4	80.4	80.9	80.9	80.8	81.1	81.3	81.5	81.8
Sex: Male																				
Life expectancy	71	71.4	71.2	72.1	72.5	72.4	72.9	73.1	73.1	73.3	73.9	74.4	74.5	75	75	75	75.2	75.6	75.9	76.2

Time period	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Measure																				
Sex: Total																				
Life expectancy	79.2	79.6	79.9	80.2	80.5	80.8	81	81.3	81.4	81.5	81.7	81.9	82	82.1	82.3	82.4	82.4	82.5	82.7	82.9
Life expectancy difference (female-male)	5.4	5.4	5.2	5	4.9	4.8	4.8	4.7	4.5	4.6	4.5	4.5	4.4	4.2	4.1	4.1	4.2	4.1	4.2	4.1
Life expectancy difference (male-female)	-5.4	-5.4	-5.2	-5	-4.9	-4.8	-4.8	-4.7	-4.5	-4.6	-4.5	-4.5	-4.4	-4.2	-4.1	-4.1	-4.2	-4.1	-4.2	-4.1
Sex: Female																				
Life expectancy	82	82.4	82.6	82.8	83	83.3	83.5	83.7	83.7	83.9	84	84.2	84.3	84.3	84.4	84.5	84.6	84.6	84.9	85
Sex: Male																				
Life expectancy	76.6	77	77.4	77.8	78.1	78.5	78.7	79	79.2	79.3	79.5	79.7	79.9	80.1	80.3	80.4	80.4	80.5	80.7	80.9

© Life expectancy OECD

Figure 5 - OECD Life Expectancy Data

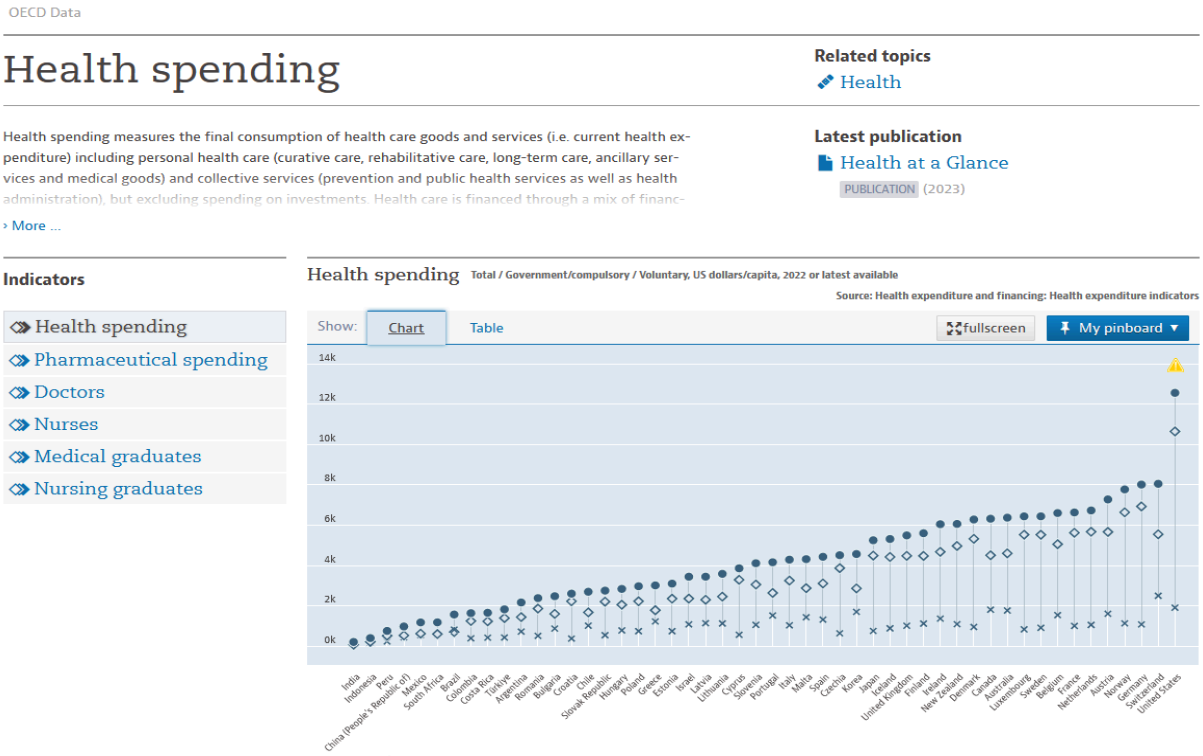


Figure 6 - OECD Health Spending Data

Life expectancy at birth, males

Graph Table Download

State or territory	2010-2012	2019-2021	2020-2022
NSW	79.9	81.4	81.3
Vic.	80.5	81.7	81.6
Qld	79.5	80.9	80.7
SA	79.8	81	81
WA	80.1	81.7	81.7
Tas.	78.7	80.3	80.3
NT	74.7	76.3	76.2
ACT	81.2	82.7	82.2
Australia(a)	79.9	81.3	81.2

a. Includes Other Territories

Figure 7- Life expectancy at birth, males and females in Each Territory of Australia

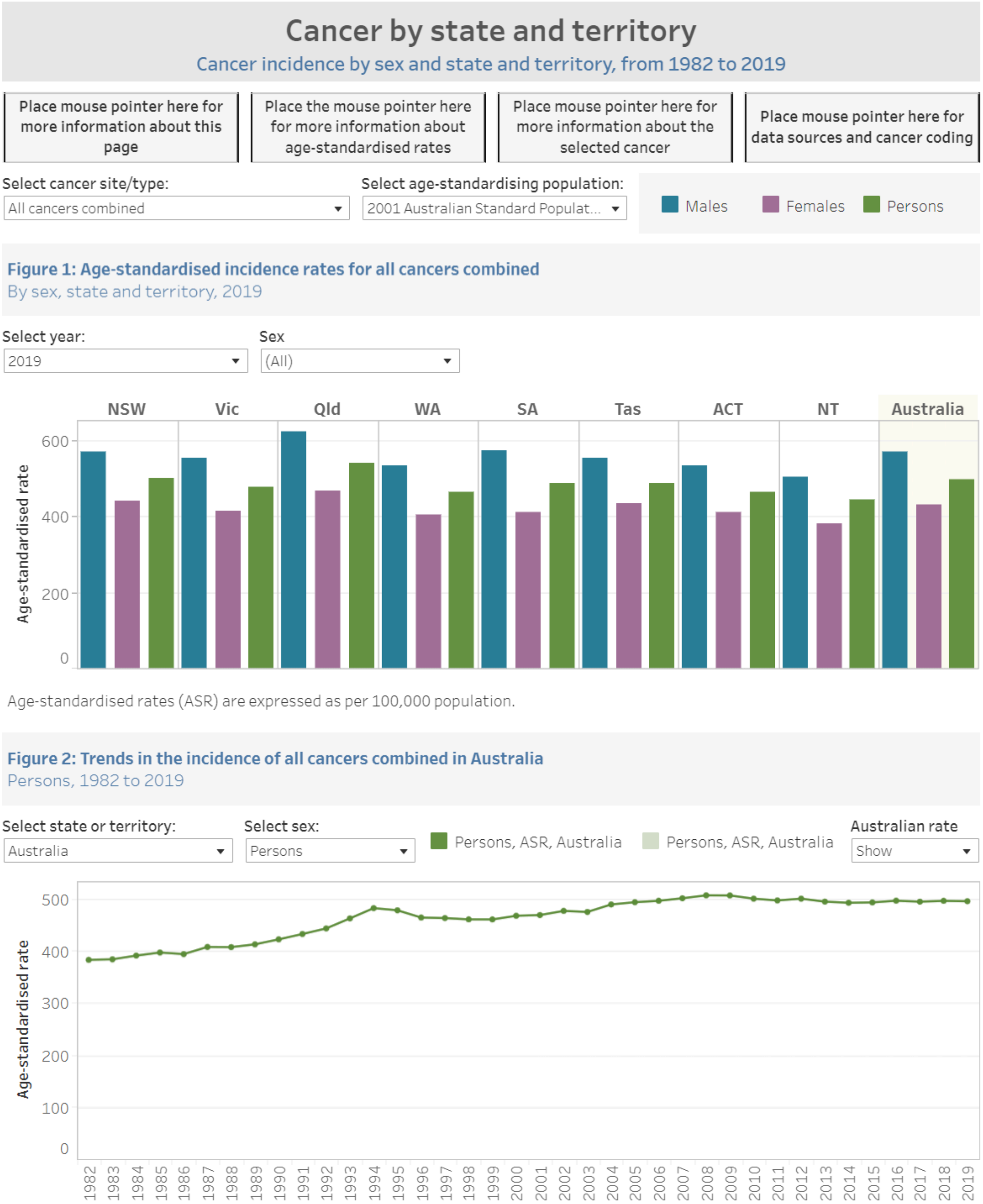


Figure 8 -Cancer Deaths in Each Territory of Australia

Data Processing

For our visualisation, we are aiming to have an interactive website that can separate the male and female life expectancy indicative of the data gathered from the OECD website regarding Australia's life expectancy. We will have a button that says “male” that will

transform the data to show the life expectancy of males in Australia in history, another button saying “female” will transform the data to show the life expectancy of females and a button saying “both” that shows and compares both male and female life expectancy.
To do this, we have cleaned the data gathered from OECD with Excel where we manually deleted extra sections that were originally in the OECD table so that unnecessary data was removed. We can interpret the data better regarding how we plan to use it for our visualisations.

Data Processing (Proposal)

Although the dataset that we had gathered is readable for us to understand its intentions, we discovered that to be able to use it for d3.visualisations and such, we must first transform it into the right format, and layout and simplify it so that we can implement it the right way. For this, we aim to follow these steps.

1. **Data Cleaning in Excel:**
    a. **Consolidate Data:** First, we will transform the data from multiple sheets or sources gathered from OECD and other websites into a single Excel sheet for each dataset (life expectancy, health spending and cancer deaths.).
    b. **Handle Missing Values:** We will identify and handle missing or incomplete data points, either by filling them in where possible or by removing those rows to maintain data integrity. We will be filling in missing values with the “null” value.
    c. **Standardised Formats:** Ensure all date formats are consistent and convert any text-based numbers into numerical format for ease of processing.
    d. **Labelling and Structuring:** Properly label columns with descriptive headers such as "Year," "Male Life Expectancy," "Female Life Expectancy," "Total Life Expectancy," and "Health Spending per Capita."
2. **Transforming Data into CSV Files:**
    a. Once the data is clean and organised, we will save each dataset as a CSV file. CSV format is ideal for data interchange because it is simple, lightweight, and easily readable by most data processing tools, including D3.js and we have been taught to use the CSV format this semester.
    b. **Export Process:** In Excel, we will use the "Save As" function to export the cleaned and structured data into CSV format.
3. **Declaring Data Types:**
    a. Once we have figured out which datasets we are going to be using, we must then declare what data types they are going to be declared as.
    b. Make a table declaring what data types each category is so members are on the same page and correct data types are being used.

Life expectancy

Reference area: Australia ● ***Frequency of observation:*** Annual ● ***Age:*** 0 years

Unit of measure: Years

Time period		1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991
Measure													
Sex: Total													
Life expectancy	i	..	74.8	74.6	75.4	75.7	75.5	76	76.2	76.2	76.4	76.9	77.3
Life expectancy difference (female-male)		7.1	7	7	6.7	6.5	6.4	6.3	6.4	6.4	6.3	6.2	6
Life expectancy difference (male-female)		-7.1	-7	-7	-6.7	-6.5	-6.4	-6.3	-6.4	-6.4	-6.3	-6.2	-6
Sex: Female													
Life expectancy	i	78.1	78.4	78.2	78.8	79	78.8	79.2	79.5	79.5	79.6	80.1	80.4
Sex: Male													
Life expectancy	i	71	71.4	71.2	72.1	72.5	72.4	72.9	73.1	73.1	73.3	73.9	74.4

Figure 9 - Life Expectancy data to be transformed according to the steps mentioned above

Health spending Total, US dollars/capita, 1981 – 2022

Location	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Argentina	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Australia	668.2	705.5	757.4	803.8	855.5	904.5	952.6	1 027.9	1 097.4	1 166.3	1 234.4	1 305.5	1 373.2	1 454.6	1 537.5	1 643.0	1 727.8	1 839.2	1 956.9	2 153.3
Austria	659.7	704.1	736.4	770.3	828.9	889.3	982.7	1 051.1	1 162.8	1 482.3	1 585.8	1 706.9	1 849.9	2 017.3	2 086.1	2 160.5	2 265.2	2 428.2	2 569.3	2 803.9
Belgium	668.9	741.7	786.2	813.5	865.9	944.8	1 009.7	1 111.8	1 203.7	1 301.9	1 421.0	1 525.7	1 583.7	1 621.7	1 670.7	1 768.1	1 814.9	1 881.7	2 011.7	2 298.2
Brazil	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Bulgaria	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Canada	898.8	996.1	1 066.7	1 135.0	1 212.8	1 278.8	1 357.5	1 461.3	1 579.5	1 699.8	1 805.2	1 897.5	1 930.9	1 962.2	1 984.9	1 999.8	2 092.0	2 200.5	2 278.3	2 450.6
Chile	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	589.9
China (People's Republic of)	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Colombia	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Costa Rica	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Croatia	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Cyprus	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Czechia	..	..	..	..	..	..	..	..	..	490.9	438.4	462.6	658.3	725.1	808.2	858.8	869.4	888.9	951.9	1 026.0
Denmark	871.6	956.1	1 003.4	1 025.3	1 106.7	1 167.6	1 251.1	1 313.4	1 359.3	1 441.5	1 484.0	1 554.4	1 646.9	1 708.5	1 715.0	1 814.0	1 891.7	1 914.2	2 181.7	2 345.7
Estonia	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	510.2	543.0
Finland	554.4	608.7	659.1	715.3	805.5	863.6	946.6	1 028.5	1 134.3	1 261.6	1 350.3	1 334.3	1 226.3	1 218.2	1 346.2	1 418.2	1 475.3	1 524.0	1 609.6	1 876.8
France	..	..	..	..	1 001.1	..	..	..	..	1 459.1	1 558.0	1 651.1	1 753.5	1 817.0	2 100.9	2 169.5	2 236.6	2 321.9	2 431.3	2 687.5
Germany	1 014.7	1 044.5	1 104.6	1 196.6	1 298.6	1 386.5	1 480.1	1 616.3	1 596.2	1 724.3	..	2 019.3	2 040.3	2 188.7	2 349.1	2 480.2	2 496.2	2 566.0	2 697.9	2 895.6
Greece	..	..	..	..	..	..	..	617.2	736.2	773.7	796.1	887.7	1 001.5	1 140.6	1 191.4	1 234.0	1 289.1	1 328.6	1 390.3	1 418.4
Hungary	..	..	..	..	..	..	..	..	..	..	589.8	627.2	647.8	727.7	685.4	692.8	698.3	768.8	823.2	906.9
Iceland	822.5	925.3	984.3	962.9	1 046.3	1 248.2	1 472.0	1 597.4	1 557.3	1 575.2	1 662.6	1 665.7	1 697.8	1 751.6	1 810.7	1 886.5	2 019.1	2 334.4	2 575.6	2 658.9
India	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..	..
Indonesia																				

Figure 10 - Health Spending data to be transformed according to the steps mentioned above

Data Processing (Progress)

During our data transformation, we unfortunately encountered some obstacles regarding how we conducted our data transformations. These shortcomings included the likes of:

1. **Miscommunication:**
    - **Team Coordination:** Initially, our team faced miscommunication issues regarding task assignments and data processing protocols. Different team members had different understandings of how to clean and structure the data, leading to inconsistencies in the initial datasets.
    - **Data Interpretation:** There were misunderstandings about the meaning of certain data points and how they should be represented. For instance, some team members interpreted life expectancy data differently, resulting in variations in how the data was processed.
2. **Data Cleaning Errors:**
    - **Incomplete Data Handling:** In our initial cleaning efforts, we missed some incomplete data points, which led to errors during the analysis phase. This required us to revisit the datasets to ensure all missing values were appropriately handled.
    - **Standardisation Issues:** We encountered problems with inconsistent date formats and numerical representations, which caused issues when trying to merge datasets. Some dates were in the 'YYYY/MM/DD' format while others were in 'MM-DD-YYYY' format, leading to misalignment in our data.
3. **Saving and Exporting Files Incorrectly:**
    - **Incorrect File Saving:** On several occasions, we did not save the cleaned and manipulated data correctly. This resulted in loss of progress and the need to redo significant portions of our work. For example, some team members saved files in proprietary Excel formats (.xlsx) instead of the required CSV format, causing compatibility issues.
    - **Export Errors:** When exporting data to CSV format, we sometimes overlooked critical settings such as proper delimiter choices (comma vs. semicolon) and text encoding (UTF-8), leading to corrupted or improperly formatted files. This created problems when importing the data into D3.js for visualisation.
4. **Data Transformation Challenges:**
    - **Incorrect Parsing:** During the transformation of data within our JavaScript environment, we faced issues with incorrect parsing of dates and numerical values. For instance, some numeric fields were inadvertently treated as strings, which caused errors in our visualisations.
    - **Data Merging Issues:** Combining the life expectancy and health spending datasets based on the year was more challenging than anticipated due to discrepancies in data formats and missing years in one dataset but not the other. This required additional steps to ensure alignment.

Time Period	Age
1980	null
1981	74.8
1982	74.6
1983	75.4
1984	75.7
1985	75.5
1986	76
1987	76.2
1988	76.2
1989	76.4
1990	76.9
1991	77.3
1992	77.4
1993	77.9
1994	77.9
1995	77.8
1996	78.1
1997	78.4
1998	78.6
1999	78.9
2000	79.2
2001	79.6
2002	79.9
2003	80.2
2004	80.5
2005	80.8
2006	81
2007	81.3
2008	81.4
2009	81.5
2010	81.7
2011	81.9
2012	82
2013	82.1
2014	82.3
2015	82.4
2016	82.4
2017	82.5
2018	82.7
2019	82.9
2020	83.2
2021	83.3

Time Period	US Dollars in Health Spending per person
1980	592
1981	668
1982	706
1983	757
1984	804
1985	855
1986	905
1987	953
1988	1028
1989	1097
1990	1166
1991	1234
1992	1306
1993	1373
1994	1455
1995	1537
1996	1643
1997	1728
1998	1839
1999	1957
2000	2153
2001	2291
2002	2481
2003	2579
2004	2818
2005	2872
2006	2996
2007	3200
2008	3350
2009	3418
2010	3593
2011	3809
2012	3854
2013	4088
2014	4556
2015	4776
2016	5037
2017	5075
2018	5191
2019	5127
2020	5692
2021	6226

Figure 11 - Before Data Transformation

Data Processing (Final)

During our project, we faced several challenges while transforming the OECD datasets on average life expectancy and health spending per capita for Australia. These issues included miscommunication, errors in data handling, incorrect file saving, and difficulties in data transformation. However, by addressing these shortcomings systematically, we were able to fully transform the datasets as originally planned.

1. Improved Communication and Coordination

- **Regular Meetings:** We scheduled regular team meetings and check-ins to ensure all members were aligned with the project goals and data processing protocols. This helped us quickly address any misunderstandings and ensure consistent approaches across the team.
- **Clear Documentation:** We created detailed documentation outlining the steps for data cleaning, transformation, and exporting. This included specific instructions on handling missing values, standardising formats, and using the correct delimiters for CSV files.

2. Enhanced Data Cleaning and Handling

- **Thorough Data Checks:** We implemented rigorous checks to identify and address incomplete data points. Automated scripts were developed to highlight anomalies and missing values, which were then corrected or filled in using appropriate methods.
- **Standardised Formats:** We standardised all date formats and numerical representations across the datasets. Dates were converted to a consistent numerical field and were ensured to be in a proper numerical format rather than text.

3. Proper Saving and Exporting Practices

- **File Naming Conventions:** We established clear file naming conventions and version control practices to track changes and avoid overwriting important data. This included using meaningful names and timestamps for different versions.
- **Correct Export Settings:** We created detailed protocols for exporting files to CSV format. These protocols ensured that all necessary settings, such as proper delimiters (commas) and text encoding (UTF-8), were correctly applied to prevent format issues.

4. Data Source

The dataset used for this project includes various types of data sourced from public health records and statistical databases. The primary datasets are structured as tables, which include attributes related to life expectancy, health spending, and cancer deaths across different states of Australia. Additionally, a GeoJSON file is used to represent the geographical boundaries of the Australian states for mapping purposes.

Life Expectancy Data: This dataset includes attributes such as state, male life expectancy, female life expectancy, and overall life expectancy. The values in this dataset are quantitative, representing the number of years.

Health Spending Data: This dataset contains attributes for state and spending, with spending being a quantitative attribute measured in millions of dollars.

Cancer Deaths Data: This dataset includes the year, state, count of cancer deaths, and sex (both). The attributes are a mix of categorical (state, sex) and quantitative (year, count).

GeoJSON File: The GeoJSON file includes the geographical coordinates that define the boundaries of each state in Australia. This file is used to create the map visualisation.

5. Data Processing

To enhance the clarity and conciseness of the visualisations, the data was averaged into five-year intervals. This method involved averaging the data for periods such as 2000-2005, 2005-2010, and so on, up to 2020. By averaging the data over these intervals, we were able to smooth out yearly variations and highlight longer-term trends, making it easier for users to interpret and gain insights into the visualisations.

The data processing steps included: Cleaning and preparing the data by handling missing values and ensuring consistency across different datasets. Averaging the data in five-year intervals to provide a clearer overview of the trends.Integrating the GeoJSON file with the life expectancy, health spending, and cancer death data to create an interactive map visualisation.

.

year	state	count	sex
2000	New Soutl	8523	both
2000	Victoria	7891	both
2000	Queenslar	7321	both
2000	South Aus	6457	both
2000	Western A	5983	both
2000	Tasmania	4832	both
2000	Northern '	3142	both
2000	Australian	4098	both
2005	New Soutl	6784	both
2005	Victoria	7456	both
2005	Queenslar	6897	both
2005	South Aus	6231	both
2005	Western A	5782	both
2005	Tasmania	4795	both
2005	Northern '	2876	both
2005	Australian	3584	both
2010	New Soutl	7893	both
2010	Victoria	7285	both
2010	Queenslar	7142	both
2010	South Aus	6483	both
2010	Western A	5961	both
2010	Tasmania	4831	both
2010	Northern '	2935	both
2010	Australian	3741	both
2015	New Soutl	5123	both
2015	Victoria	8684	both
2015	Queenslar	7891	both
2015	South Aus	7457	both
2015	Western A	5998	both
2015	Tasmania	3728	both
2015	Northern '	3981	both
2015	Australian	3429	both
2020	New Soutl	7392	both
2020	Victoria	3897	both
2020	Queenslar	5291	both
2020	South Aus	8415	both
2020	Western A	6783	both
2020	Tasmania	3598	both
2020	Northern '	4894	both
2020	Australian	6367	both

state	spending
New South Wales	5191
Victoria	4776
Queensland	5075
South Australia	5037
Western Australia	5692
Tasmania	6226
Northern Territor	3854
Australian Capital	3809

A1 state

state	male	female	both
New Soutl	80	85	82.5
Victoria	79	84	81.5
Queenslar	78	83	80.5
South Aus	77	82	79.5
Western A	76	81	78.5
Tasmania	75	80	77.5
Northern '	74	79	76.5
Australian	81	86	83.5

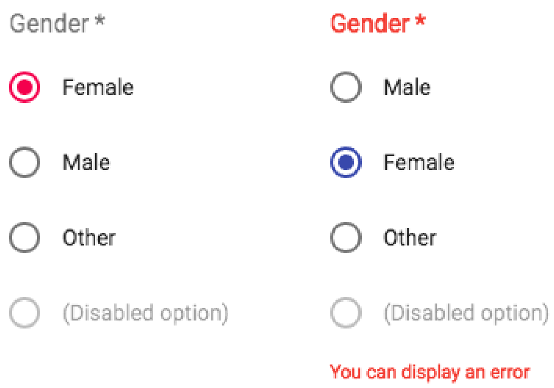
Figure 12 - After Data Transformation

Requirements

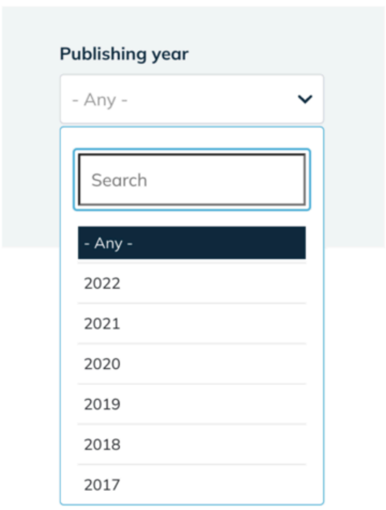
Must have features (Proposal)

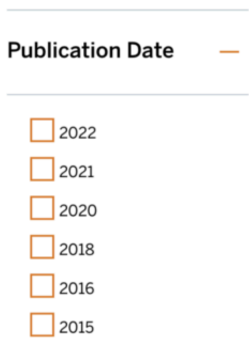
Having a project that revolves around making datasets more appealing and interactive through the use of javascript and D3, it is essential that our project has the right amount of features that are implemented to make sure it is interactive, usable, understandable and as appealing as possible. For this, we have decided to add the following things

1. **Buttons for Male, Female and both:** We plan to implement a usable button that changes our data visualisation depending on what the user has clicked. There would be a drop-down button that has the button names “Male”, “Female” and “Both”. Where “Male” will transform the chart of our choice into the male life expectancy averages, and clicking the “Female” will transform the chart into the female life expectancy averages and finally the “Both” button transforming the chart showing the average of both the male and female’s life expectancy.

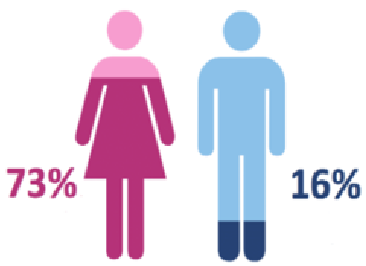


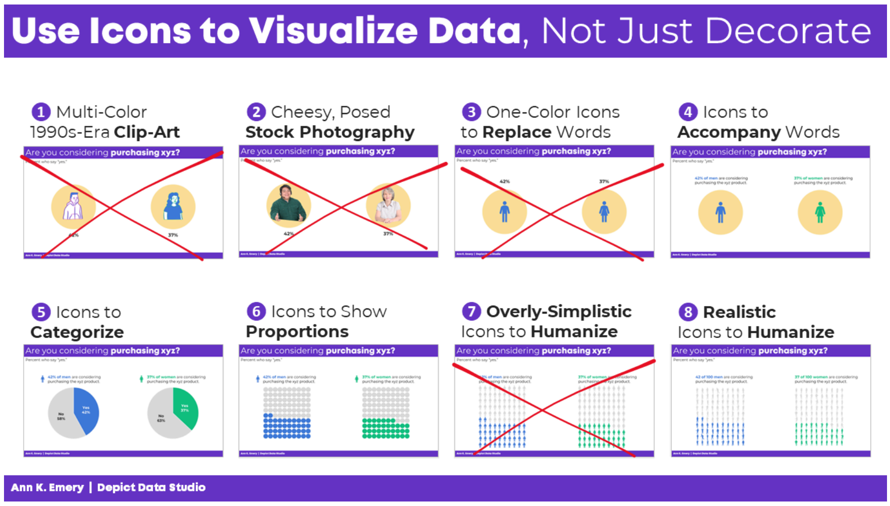
2. **Year segments/filter**: We also aim to add a function that segments the year for the health spending data set so that it is more readable and usable for the user. This feature will allow the user to segment years from a drop-down menu which they can then manipulate and choose the years that they want to segment on their own which would then transform the dataset and cut out the years that weren’t chosen and only focus on the years “segmented” to allow for easier viewing.



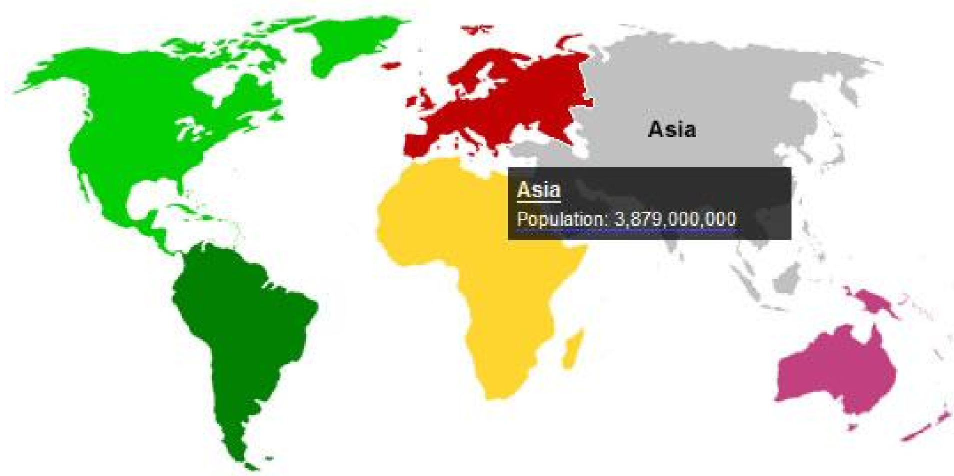


3. **Added visualisations**: We also aim to add little visualisation icons and such to keep the visualisation more user-friendly and easier to interpret. We aim to do this by adding things such as people icons, year icons, gender icons, money icons and more UI-based visualisations through the use of D3 to make the visualisations more engaging and appealing.





4. **Interactive Visualisations:** The last thing we aim to add to make our visualisations more appealing and easier to interpret for our audiences would be to add in a functional and interactive feature on our visualisations in regards to the types of designs that we are aiming to add. This could be something such as a hover feature that shows the value of where the user is hovering above or a feature that removes and adds datasets (for example - states in a country) depending on what the user desires.



Must have feature (Final)

In our project proposal, we outlined several must-have features aimed at enhancing the interactivity, usability, and appeal of our data visualisation project. These features include:

1. **Buttons for Male, Female, and Both:**
    - Implementation of buttons to toggle between male, female, and both genders for life expectancy data visualisation. This feature provides users with the flexibility to explore different demographic segments of the data and gain insights specific to each gender.
2. **Year Segments/Filter:**
    - Addition of a function to segment the years for the health spending dataset, making it more readable and usable for users. This feature allows users to select specific years from a dropdown menu, effectively filtering out unnecessary data and focusing only on the selected time periods.
3. **Added Visualisations:**
    - Incorporation of visual elements such as icons (e.g., people icons, year icons, gender icons, money icons) to enhance the user experience and make the visualisation more engaging and appealing.
4. **Interactive Visualisations:**
    - Making the visualisations more functional and interactive will make the user more likely to interact with it, encouraging engagement and usability that elevates the user experience.

While we aim to implement these features to improve the overall quality of our project, we acknowledge that we are still in the process of completing the project. At this stage, we are evaluating whether these features align with our project goals and whether they will effectively enhance the user experience.

Given that we are still determining the feasibility and relevance of these features, they have not all been applied yet. However, we remain committed to ensuring that our final project meets the highest standards of interactivity, usability, understandability, and appeal. As we continue to develop and refine our project, we will carefully consider the implementation of these features to create the best visualisations regarding life expectancy as best as we can.

Visualisation Designs

Drafted Designs

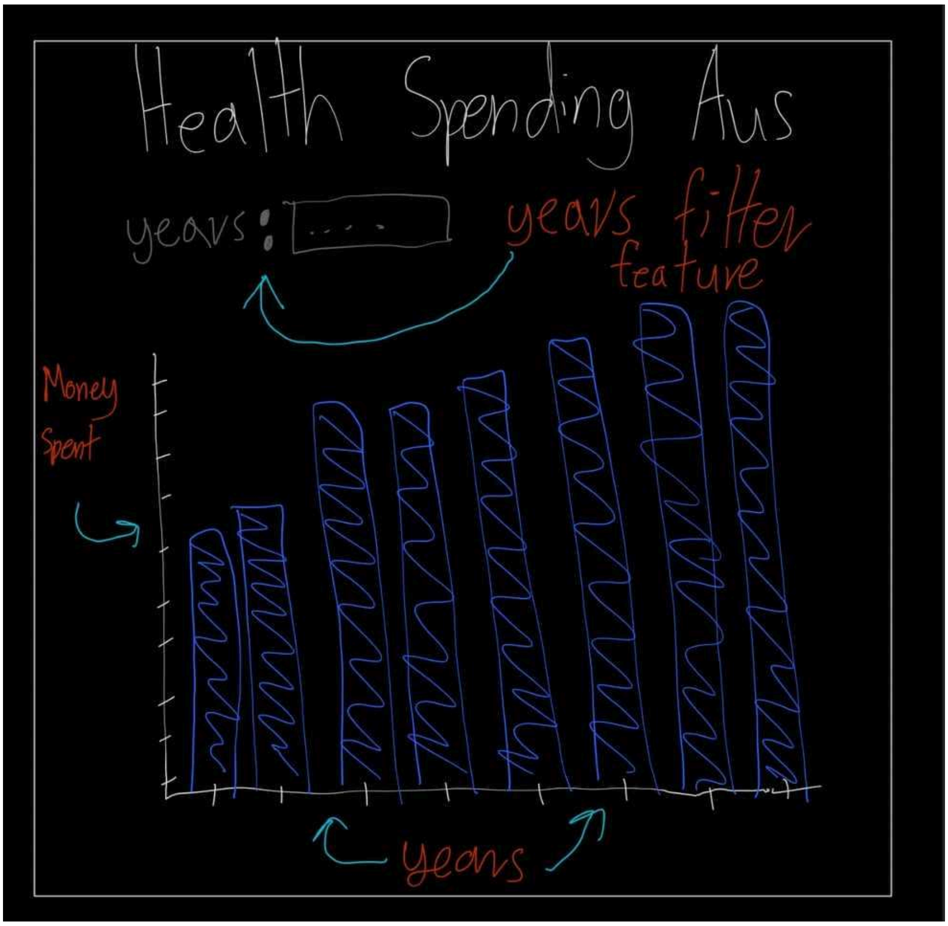


Figure 13 - Health Spending Visualisation Draft Design

The first data visualisation design is the initial design that we have come up with for the health spending dataset in Australia. In this design, we aim to add a central main heading and title showing the main intention of the data visualisation being displayed. It follows the descriptions of the main features that were supposed to be added as mentioned above.

This design includes a simple bar chart where each bar depicts the years in history and the height of the bars depicting the amount of money spent per person regarding health spending. This design also includes a tab at the top called “years” where the user will be able to filter through the years of health spending in Australia where a user can accurately filter the years that they want to target allowing them to more easily see and visualise the health spending values in the years selected.

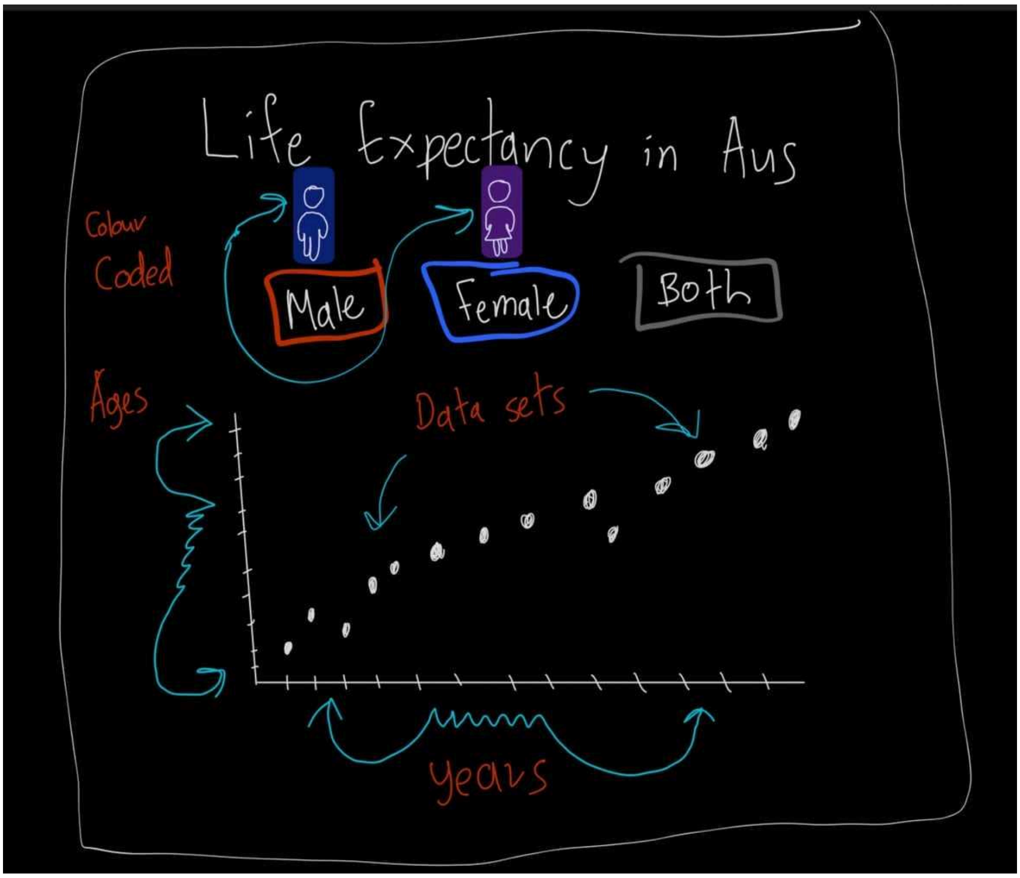


Figure 14 - Life Expectancy Visualisation Draft Design

The second visualisation design that we created was related to the Life Expectancy in Australia dataset from OECD. This design follows the same design aspects of the Health Spending visualisation where it has a central main heading letting the audience know what data it is showing followed by the actual data visualisation underneath.

This time around, we have chosen to do a scatterplot design where the x-axis dictates the year values and the y-axis dictates the average age values. This design also has added functionality with the added male, female and both buttons where the scatterplot values transform depending on which button the user chooses to click. This design also aims to include more visualisation and aesthetic elements by adding a “male” icon above the male button and a “female” icon above the female button.

With these two initial designs finished, we decided to finally start and try to implement them on our website to see if they would work as we had expected them to.

Implemented Designs V1

Figure 15 - Health Spending Data Initial Implementation Design

In our first initial implementation, we used the scatterplot design (which was supposed to be used for the life expectancy) instead of our original plan of using the bar chart to display our health spending in Australia data. We thought that the scatterplot design made more sense as it was there were a lot of data sets that were to be displayed as our visualisation including datasets coming from 1980 to 2020. This meant that if we had used a bar chart as we had planned, the chart would've looked pretty messy as all the bars would be skinny and be "squished" together in order for all datasets to be able to fit inside the bar chart window.

As this was our first implementation, we weren't exactly satisfied with it yet as right now it is very bare bones and doesn't do much for the audience in regards to aesthetics and appeal. To solve this, we went back to brainstorming and drafting new designs in hopes of finding the one we deemed the best fit.

Implemented Designs V2

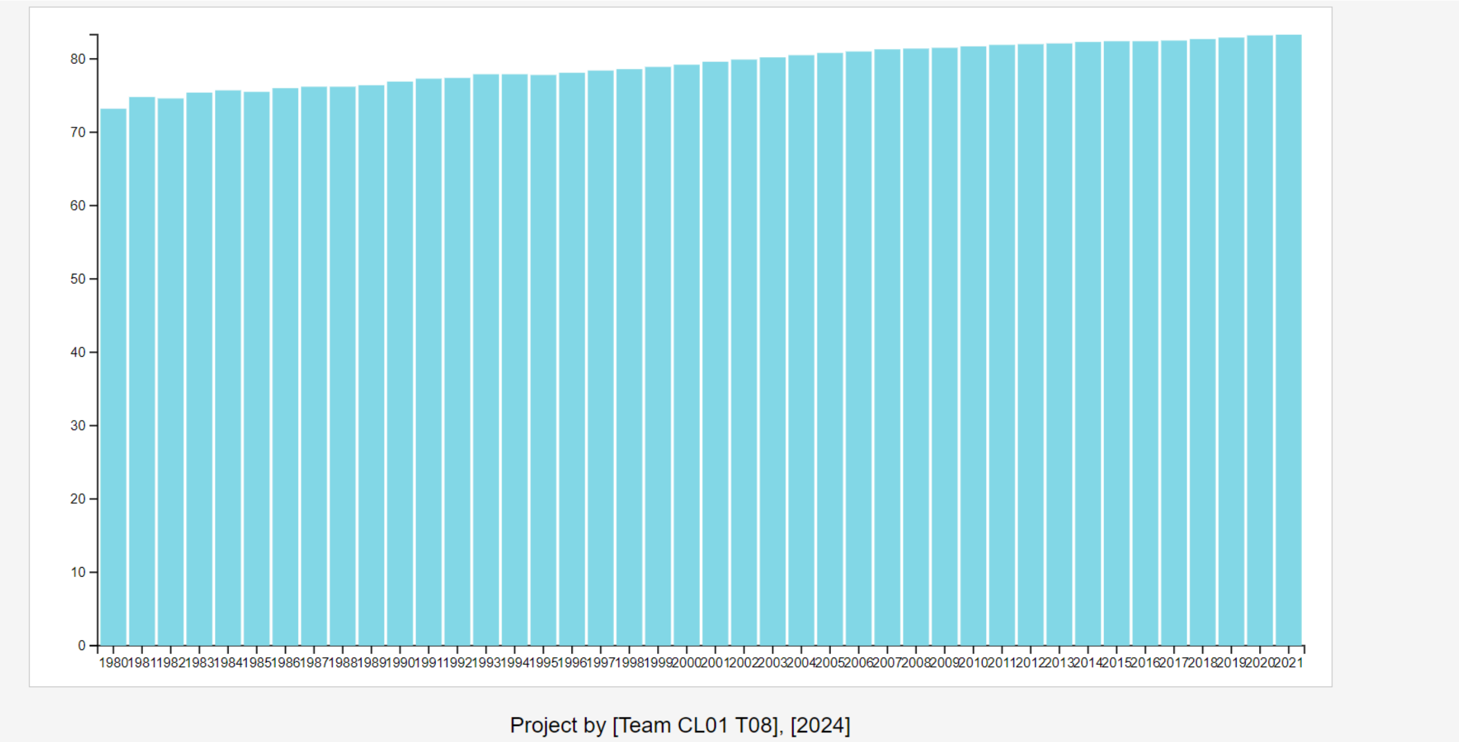


Figure 16 - Life Expectancy Data Initial Implementation Design

Moving forward to the next version implementation of our visualisations, we started of trying to use one of our other original draft ideas of using a bar chart to implement our life expectancy visualisation. This turned out messier than we had anticipated in our designs because there were way too many datasets in the chart for it to be readable and appealing to the audience. The X-axis depicting the years is unreadable as data values are extremely close to each other and the differences in value based on the heights of the bar are barely visible as there are only slight increments that life expectancy increases per year.

This draft design proved to be pretty unusable as it fails the most important requirement we have which is having a readable and appealing design for our audiences. This draft design will need more improvement.

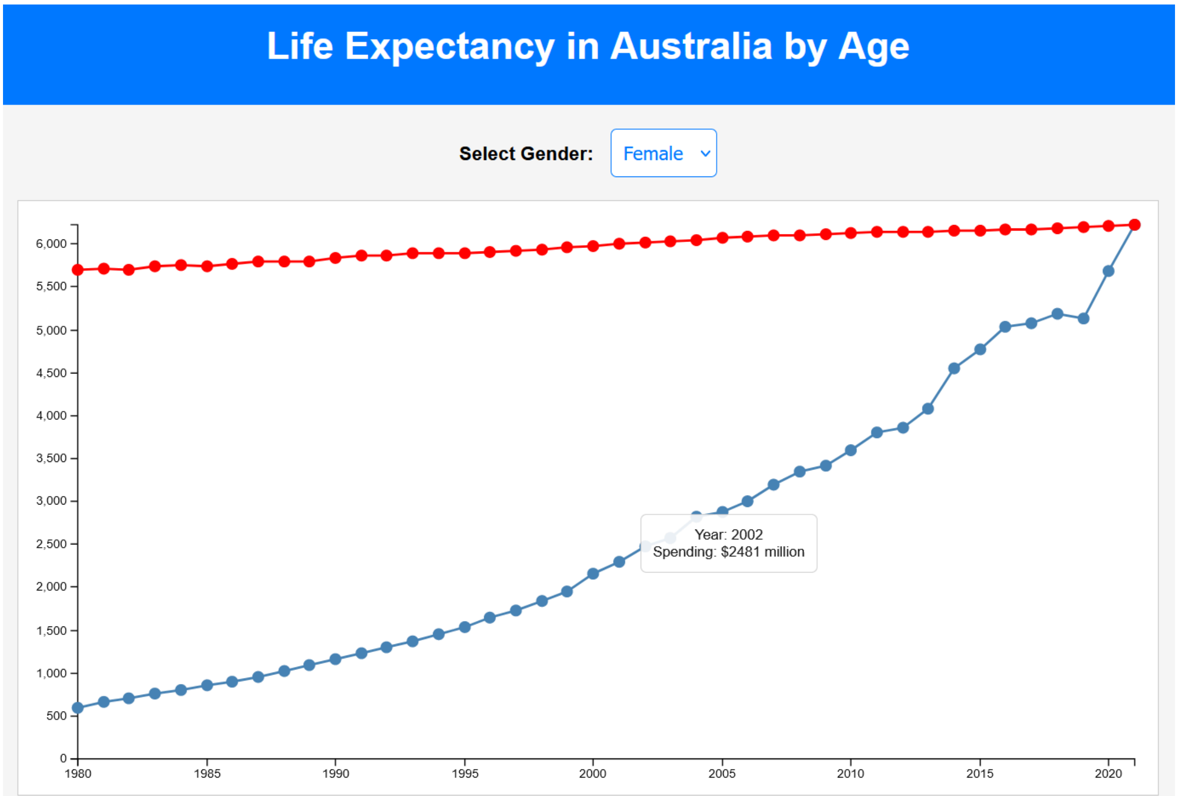
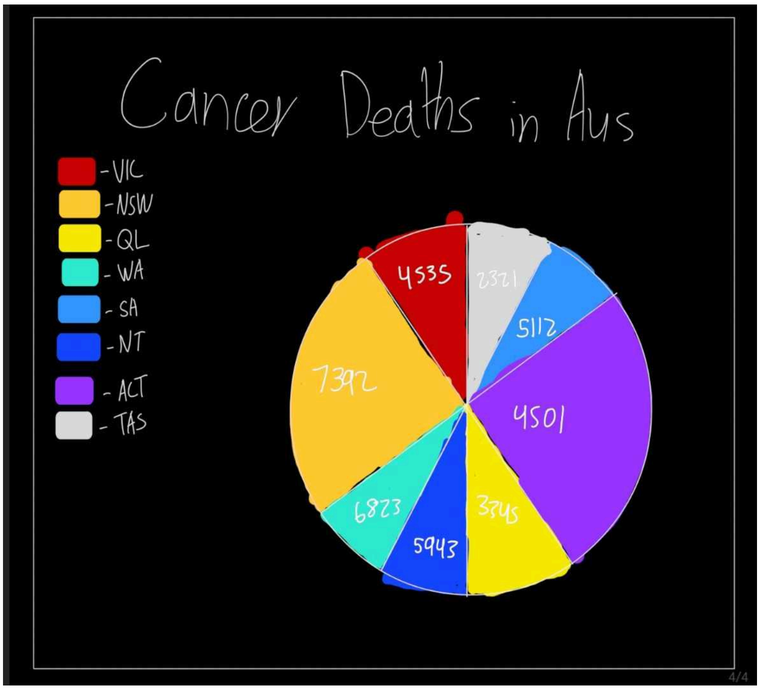


Figure 17 - Life Expectancy and Health Spending Visualisation

The next thing we did in version 2 of our implementation to supplement the bar chart that we had created was to make another visualisation that spliced both datasets from Health Spending and Life expectancy together. This scatterplot had two different data sets on it which were differentiated by having different colours where the red dots signified the life expectancy of each year and the blue represents the health spending per capita each year. A hover function is added so that when a user hovers on a specific dot plot, it shows which year it belongs to and the life expectancy and/or spending per capita that year depending on which coloured dot the user hovers over.

Although on paper this sounded like a good idea as not only did it provide more information to the users even with just using one visualisation, it was much easier to interpret compared to the bar chart for life expectancy that we had created earlier. However, it doesn't take long to see its flaws either. The biggest flaw in this design implementation is that the Y-axis could only store one type of dataset. This means that it was impossible to see the age values and the spending values per capita simultaneously. This would make it confusing and uninspiring for our audience as they would have a harder time trying to interpret the data that they are being presented with. The other big flaw is that the life expectancy scatter plot increments are very difficult to visualise. With life expectancy only going up by a few decimal values each year, it is hard to see how much it has changed over the years. For this to be fixed we would have to scale down the life expectancy values from 0 - 85 years old to around 75 - 85 years old to see any meaningful changes. This is, however, impossible as life expectancy is also spliced with the health spending data. With yet more obstacles being faced and new lessons being learned we realised we needed to go back even further and rethink how we should be doing our designs.

Revised Draft Designs



Figures 18 & 19 - Cancer Deaths in Australia Design and Revised Life Expectancy In Australia Design

Coming back to the drawing board after being unhappy and unsuccessful with the implementation of our previous designs, we had to think outside the box to find a better solution to display our datasets in a more meaningful and appealing way. Through trial and error and conducting extra research, we decided to draft two new visualisation ideas.

Pie Chart Design - Being stuck on the other two datasets with the Life expectancy and health spending data, we completely neglected the other dataset that we have finalised which was the cancer deaths in Australia dataset. For this dataset, we came up with an idea that we have never used before which was to use a pie chart. This pie chart aims to show all the cancer deaths in each state of Australia where each state is represented by a colour according to the legend and each pie having the values of deaths in that state of that year. We also changed our plans and planned to implement a new feature with the "central" year selector that changes every single visualisation on the whole page depending on what year the user has selected. This means that all of the visualisations will depend on this new year selector in regard to what data is being displayed at the moment.

Choropleth Design - Another design that we had come up with and thought would work perfectly in conjunction with our new pie chart design was to implement a choropleth visualisation. This new design would take over the current life expectancy designs which we had a lot of trouble with finding a good visualisation that was appealing and wasn't hard to read. This new choropleth design will show a map of Australia where all the states are visible and segmented. The map and states of Australia will then be coloured in according to the legend which shows the colour red getting darker and darker which corresponds to the average age in that state of that year.

Implemented Designs V3

Figures 20 & 21 - Revised Visualisation Designs for Life Expectancy and Cancer Deaths Datasets

Having gone back to the drawing board and eventually coming up with the designs that we were happy and optimistic about, we finally went back to implementing our designs on our website. Using the designs we have drafted, we made the basic implementations of the cancer deaths in Australia visualisation with the use of a pie chart. This pie chart at the moment was merely hardcoded to show if it would work and didn't have any functioning features such as showing the values of the states and the cancer deaths along with it. This basic skeleton however, was able to show us that using a pie chart as a visualisation was a viable choice and enabled us to continue with this design instead of trying to draft a different one.

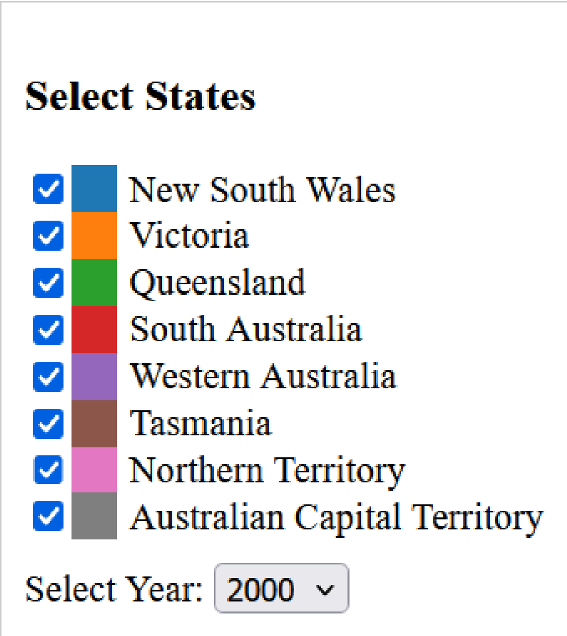
Moving on to our other implementation with the choropleth design, we had to make minor adjustments and tweaks as we encountered problems and challenges in regard to making the "heatmap" feature of the choropleth visualisation work. For this, instead of making a traditional choropleth, we decided to show the map of Australia with the segmented states which are colour-coded to its matching legends with the aim of having a hover-over feature that enabled users to see information about that state's life expectancy for both males and females depending on the year that the user has selected. We also added a function that allowed states to be filtered out signified by a checkbox next to each of the states in the clickable legend which "removes" and greys out that state if unticked by the user allowing them to see other ones that were left instead more easily.

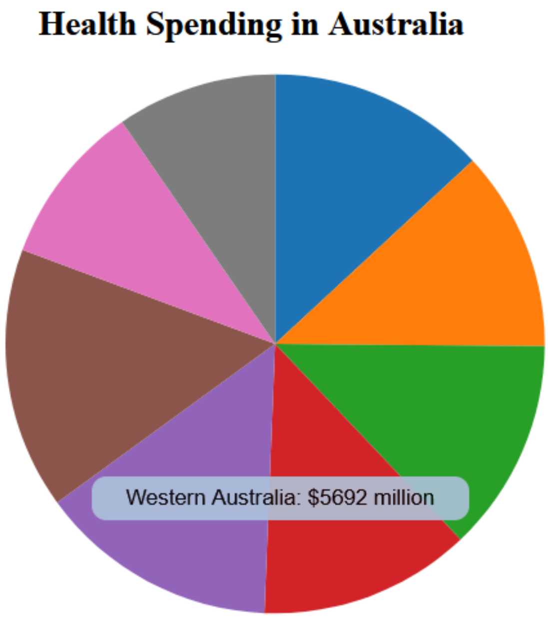
After much trial and error, we were finally happy with the designs that we had come up with. Moving forward these two designs were the two main designs that we were going to be continuing with for our final implementations.

Final Designs/Implementation

Figure 22 - Final Life Expectancy Visualisation Implementation

For our final implementation, we continued with the revised visualisation drafts that we came up with and tweaked them to be more functional and appealing as we had planned. For our choropleth-based design, we were able to implement the feature of being able to hover over the states which prompted a window that showed the life expectancy for males, females and both for that state depending on the year. We tested thoroughly and ensured that each state was working and that each state's data was changing according to the year the user had selected.





Figures 23 & 24 - Revised Visualisation Designs for Life Expectancy and Cancer Deaths Datasets

To control, understand and change the values of the map design we have implemented, we made sure to have a matching legend along with it that has the same colours as the states on our map. This legend has two main functionalities in that it has a “tick box” that selects or unselects the states of the user's choosing. It also houses the most important function of our whole website in that it contains the “year” selector. This year-selector shows a drop-down box when clicked and shows the years from 2000 to 2020 in 5-year increments. A change of using the original plan of going from 1980 to 2020. We have chosen to narrow down the years as we were having issues finding all relevant datasets that related to each other from 1980 to 2020 as some datasets (such as cancer deaths) only went as far back as the year 2000. We have also decided on doing 5-year increments instead of 1-year increments for the “Select Year” function as it proved to be too challenging and with time running out until the submission date, we finalised on going forward with 5-year increments instead.

The next main implementation is with the pie chart visualisation. This pie chart visualisation now fully functions as we had planned where each pie shows a message depending on which pie the user hovers over. The pie's colours are also directly related to the legend just like the choropleth. A change of which data set that was going to be displayed was changed from Cancer deaths in Australia to Health Spending in Australia instead, as we found a better way to visualise the cancer deaths dataset with an alternate visualisation.

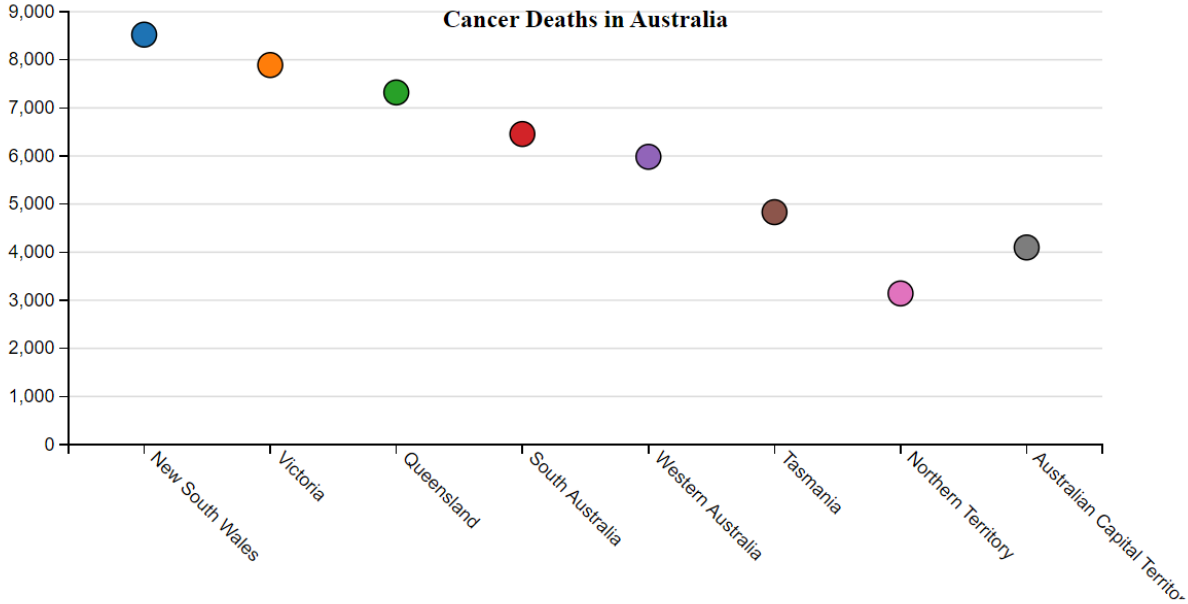


Figure 25 - Final Life Expectancy Visualisation Implementation

The alternate visualisation in question that we used for the cancer deaths in Australia dataset was none other than the scatter plot visualisation that we had originally used for the life expectancy and health spending dataset during our first implementations. This scatterplot visualisation is also directly related to the legend where each dot is colour-coded to match just like the other two visualisations discussed. This scatterplot shows which state the dot is with the X-axis and the range of death values with the Y-axis. This visualisation also has the added hover-over function like the others in that when the user hovers over a specific dot on the scatterplot, a window will pop up showing information related to cancer deaths such as how many deaths there were, in what year and for which sexes in relation to its position on the Y-axis and state.

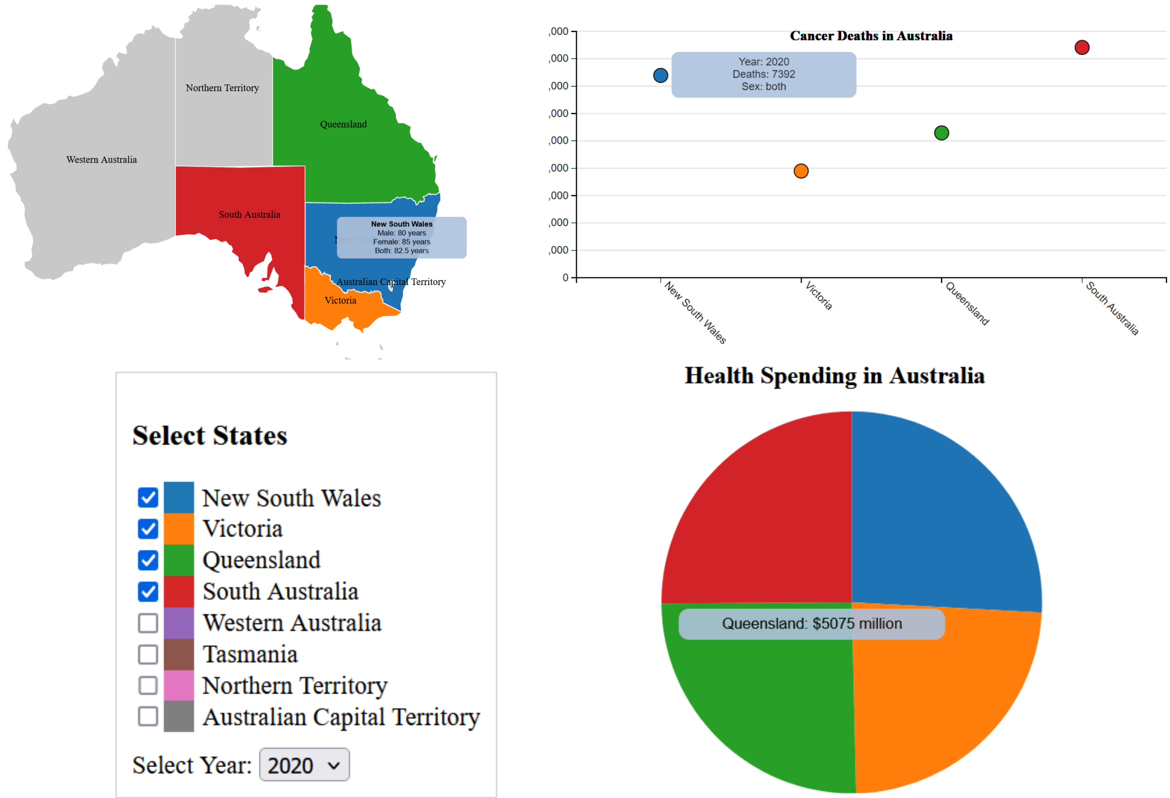


Figure 26 - Legend Filtering Function In Action

The figures above show the main functionality of the website and its interactive nature with the ability to section out and filter states according to the user's desires. In this example, we can see that the user has decided to only show the states of NSW, VIC, QL and SA and has selected for the year 2020. This is reflected throughout all the visualisations on the page and changes can be seen for each one. These changes include:

- **Choropleth Visualisation:** Has NSW, VIC, QL and SA coloured in and the rest of the states greyed out
- **Scatterplot Visualisation:** Removes other scatter plot dots and only shows the 4 states that were selected
- **Pie Chart:** Removes the other pie chart data values and only shows the colours of the 4 states that were chosen

Conclusion

The Australian Life Expectancy data visualisation project successfully developed an interactive web-based platform to depict life expectancy trends in Australia, segmented by gender and influenced by factors such as health spending and cancer mortality rates. Utilising D3.js, we created dynamic visualisations that allow users to explore differences and trends across Australia's states and territories. Our primary objective was to provide insightful and accessible visualisations, offering a comprehensive view of life expectancy trends and their determinants over time. We aimed this tool at public health officials, policymakers, researchers, and the general public equipping them with a resource to understand and analyse life expectancy trends effectively.

Despite initial challenges in data transformation and visualisation, we systematically addressed these issues, ensuring accurate and effective data representation. Key improvements in communication, data cleaning, and handling, along with proper saving and exporting practices, enabled us to overcome obstacles. By averaging data in five-year intervals, we provided a clearer and more concise overview of long-term trends. The final implementations, including the choropleth map for life expectancy, the pie chart for health spending, and the scatterplot for cancer deaths, offer a robust and interactive platform for exploring life expectancy trends in Australia. Each visualisation is designed to be engaging, informative, and user-friendly, facilitating in-depth analysis and exploration of public health data.

This project demonstrates the power of data visualisation in enhancing the understanding of public health outcomes, fostering greater public engagement, and supporting research. By making complex data accessible and interactive, we contribute to better-informed healthcare decisions and policies, underscoring the potential of data visualisation to drive positive societal change.

References

1. *Cancer data in Australia, Cancer by state and territory data visualisation*. (2023, August 31). Australian Institute of Health and Welfare. https://www.aihw.gov.au/reports/cancer/cancer-data-in-australia/contents/state-and-territory
2. *Life expectancy by state and territory, 2020 - 2022*. (2023, November 8). Australian Bureau of Statistics; Australian Bureau of Statistics. https://www.abs.gov.au/statistics/people/population/life-expectancy/latest-release
3. Australian Institute of Health and Welfare. (2020, July 23). *Immunisation and vaccination*. Australian Institute of Health and Welfare. https://www.aihw.gov.au/reports/australias-health/immunisation-and-vaccination
4. OECD. (2022). *Health resources - Health spending - OECD Data*. Organisation for Economic Co-Operation and Development; OECD. https://data.oecd.org/healthres/health-spending.htm
5. OECD. (2024). *Health resources - Life Expectancy - OECD Data*. Organisation for Economic Co-Operation and Development; OECD. https://data-explorer.oecd.org/vis?fs[0]=Topic%2C0%7CHealth%23HEA%23&pg=20&fc=Topic&bp=true&snb=47&vw=tb&df[ds]=dsDisseminateFinalDMZ&df[id]=DSD_HEALTH_STAT%40DF_LE&df[ag]=OECD.ELS.HD&df[vs]=1.0&pd=1980%2C&dq=AUS.A...Y0........&ly[cl]=TIME_PERIOD&ly[rs]=MEASURE%2CSEX&to[TIME_PERIOD]=false