

Australian Prime Minister Lifespan Web Scraping*

Mary Cheng

February 6, 2024

Introduction

This report uses web scraping techniques in R (R Core Team 2020) to gather information on the lifespan of Australian prime ministers. The lifespan of these prime ministers can provide insights into political trends, societal changes, and the impact of various factors on the longevity of people as the technology develops.

This report is going to first scrape the data on Wikipedia page (Foundation 2023), then clean the data to only include things we are interested in, and then plot a graph to illustrate the dataset.

Data Source

The data was gathered from the Wikipedia page about Australian prime ministers. I first used `rvest` and then downloaded the page using `read_html()`.

Then using selector gadget, I identified the items I want, which was the `wikitable`. It downloaded the entire table which contained a lot of information that I did not need. So I performed some data cleaning.

I filtered the column that I was interested in, which was the called “Name(Birth–Death)Constituency”.

But this column contained both the name, birth year, death year of the prime minister. Therefore, I used `separate()` to separate the information into different columns. Then I performed some data cleaning to make sure column names make sense and the entries are correct.

*Code and data are available at:https://github.com/marycx/Australian_PM_lifespan.git

Here, I took a quick look at the final cleaned data to see if the format matches with what I think it should look like. Since Table 1 looks like my sketched dataset, I proceeded to the next step.

Table 1: Lifespan of Australian prime minister

Prime Minister	Birth year	Death year	Age at death
Edmund Barton	1849	1920	71
Alfred Deakin	1856	1919	63
Chris Watson	1867	1941	74
George Reid	1845	1918	73
Andrew Fisher	1862	1928	66
Joseph Cook	1860	1947	87

After making sure the cleaned data is correct, I plotted Figure 1 to illustrate how long each prime minister lived. I used two different colors in the graph to differ prime ministers who are alive and who passed away.

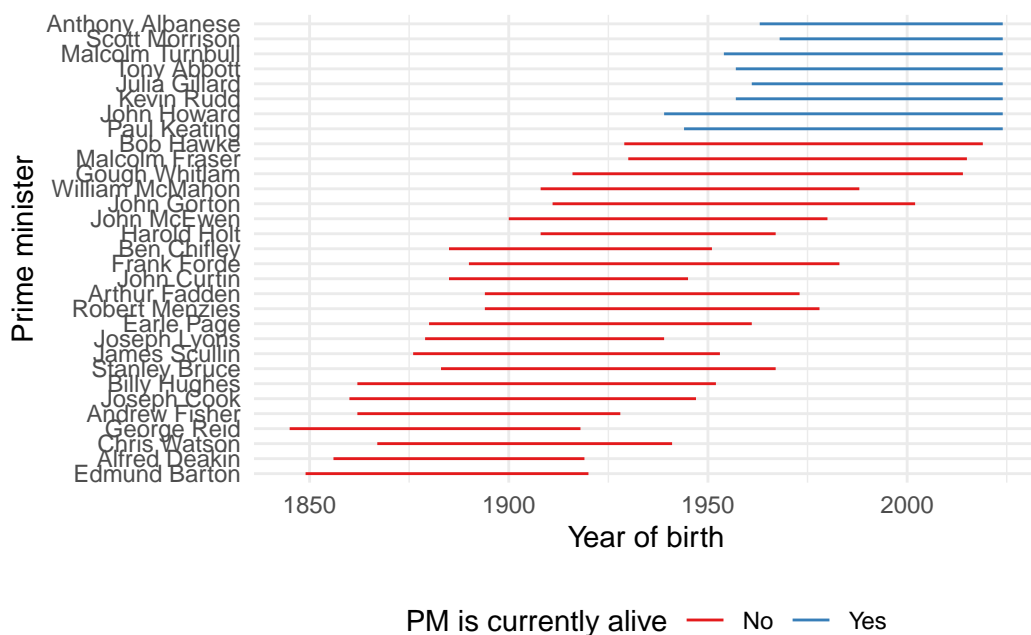


Figure 1: Lifespan of Australian prime ministers

Findings

Figure 1 shows that 8 out of 31 prime ministers in Australia are still alive. The oldest alive prime minister is John Howard, who was born in 1939. The youngest alive prime minister is Scott Morrison, who was born in 1968.

Not considering the Prime Ministers who are still alive and well, the longest lifespan is John Gorton who was born in 1911 and passed away in 2002. His age at death was 91 years old. The shortest lifespan is Harold Holt, who was born in 1908 and passed away in 1967. His age at death was 59 years old. It is observed that there is a trend of gradual rise in lifespans over the years, although some minor decreases in lifespan are also evident. This is normal since no one can control their lifespan or death. It can be seen that the most recent three prime ministers who passed away (Gough Whitlam, Malcolm Fraser, and Bob Hawke), their age at death was 98, 85, 90 respectively. It is possible that due to the modern technology and medical development, people's lifespan is increasing gradually.

Reflection

The separating name and dates took longer than I expected, because I was unfamiliar with the `str_extract()` function and its argument format so I searched online for a long time to see how I should write it. It became fun when I finally figured out how to separate names and dates. When I saw the cleaned data, I felt very motivated and proud. Also, it was fun to finally make that graph and see the lifespan of prime ministers visually. Next time, I think I will try to change the ordering of the prime ministers. Right now, the ordering is what appears on Wikipedia page. I think I will try to arrange them based on their birth year, so that it would be clearer to see the general trend of lifespan when I plot the graph.

Reference

- Foundation, Wikimedia. 2023. *List of Prime Ministers of Australia*. https://en.wikipedia.org/wiki/List_of_prime_ministers_of_Australia.
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.