Investigating How Long Each Prime Minister of Australia Lived*

Shuhan Yang

2024-02-06

To investigate the question of how long each prime minister of Australia lived, this research utilized data scrapped from Wikipedia and created a graph based on the scrapped information. Results revealed that among the 31 different ministers, only 8 of them are still alive, and the increasing lifespan also reflects the advancement of the nation's healthcare system.

Table of contents

*Code and data are available at: https://github.com/Jasmineee35/web-scraping-prime-minister-of-Australia.git

1. Introduction

Inspired by the example of 'Prime Minister of the United Kingdom', this research aims to analyze on the question:"how long each prime minister of Australia lived". The results of the study not only reveals the generational iteration of the ministers but also reflecting broader society changes as well. Spanning from the 1850s to the present day, the increasing lengths of the ministers' lifespan suggest Australia's healthcare advancements overtime.

2. Data Source

Using web-scrapping, I retrieved the dataset from Wikipedia 'List of prime ministers of Australia' (2023b). This data from the webpage contains 31 observations and various information regarding each minister's lifespan, political party, the time in the office...etc. This research will only focus on their birth and death year. By using R (R Core Team 2023), and R packages "tidyverse" (Wickham et al. 2019), "janitor" (Firke 2023), "lubridate" (Grolemund and Wickham 2011), knitr" (Xie 2023). Loading "rvest" (Wickham 2022) to scrape the data from Wikipedia, I the used "read_html()" to download the page. Using the tool "SelectorGadget" allows high efficiency when selecting elements we want. I first cleaned the dataset by filtering out the unused information, then gathered data into the format of "name, date, born" to make the data organized. Then finishing the data preparation process by cleaning each column so the data can be directly used to generate graphics using (Wickham 2016).

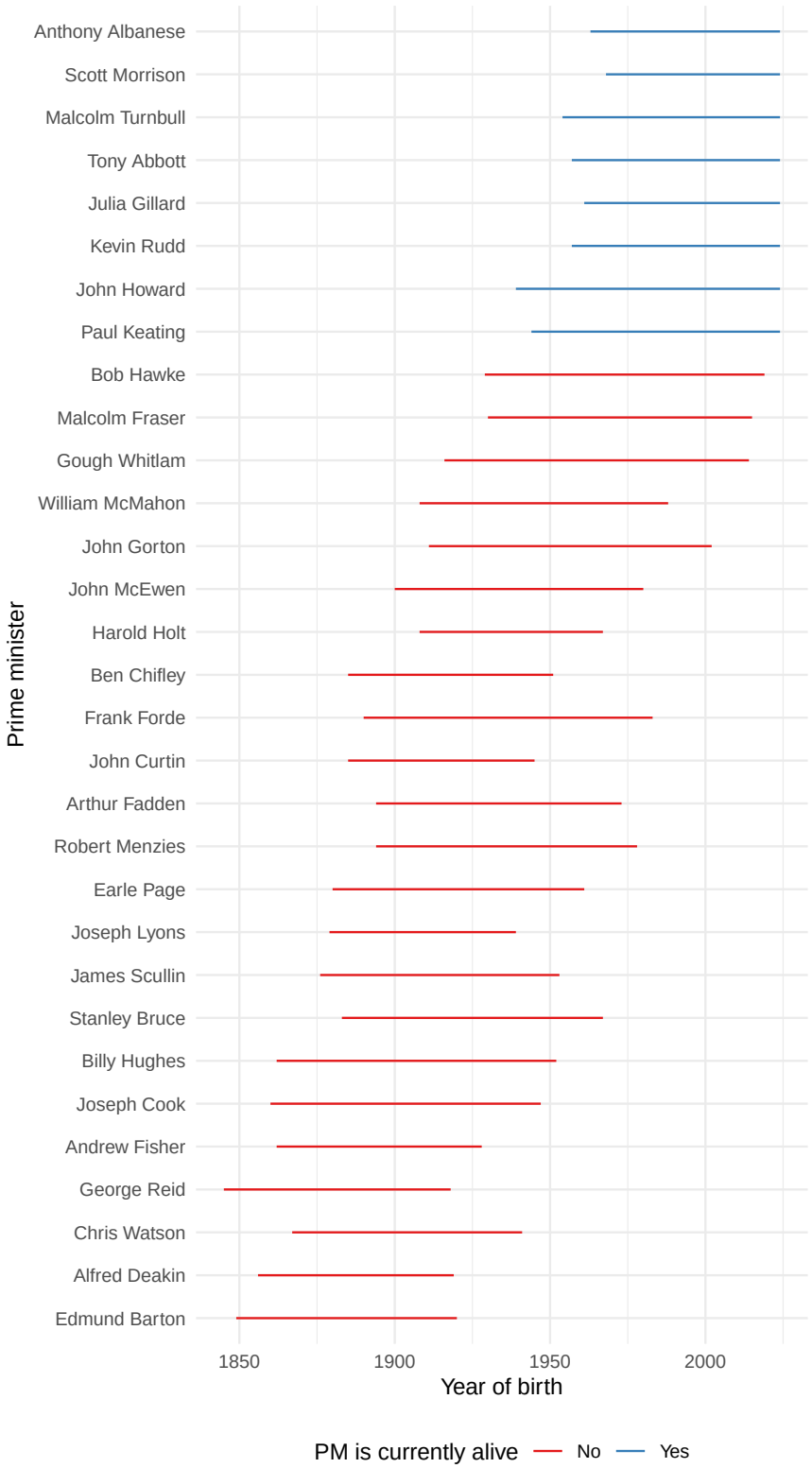
The following table will illustrate 6 rows for the prime minister data, to gain a general idea of what the dataset looks like.

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

This table illustrates the first six prime ministers of Australia, showing their name, corresponding birth year, death time, and lifespan. The first prime ministers of Australia is Edmund Barton, who was born in 1849 and then passed away during 1920. Edmund took office from 1901, when he was 52 years old, and left during the year of 1903.

3. Findings

At this point, to better visualize the whole data, the below graph was created in order to illustrates how long each prime minister lived. To make the graph visually appealing, colour code 'red' represents already passed prime ministers and 'blue' for those who are still alive. The vertical axis of the graph represents each prime minister, and the horizontal axis corresponding to their year of birth. In this case, the lines follow a chronological ascension from historical to contemporary, with the top line representing the current president and the bottom line representing Edmund, the first president of Australia.


Anthony Albanese
Scott Morrison
Malcolm Turnbull
Tony Abbott
Julia Gillard
Kevin Rudd
John Howard
Paul Keating
Bob Hawke
Malcolm Fraser
Gough Whitlam
William McMahon
John Gorton
John McEwen
Harold Holt
Ben Chifley
Frank Forde
John Curtin
Arthur Fadden
Robert Menzies
Earle Page
Joseph Lyons
James Scullin
Stanley Bruce
Billy Hughes
Joseph Cook
Andrew Fisher
George Reid
Chris Watson
Alfred Deakin
Edmund Barton
Prime minister
1850
1900
1950
2000
Year of birth
PM is currently alive
No
Yes

Analyzing the trends and results observed from this graph, we found that among the 31 different prime ministers in Australia, only 8 of them are still currently alive with the rest all passed away. Taking a closer look into it the the 23rd prime minister, Bob Hawke was the most recent minister that passed away during 2019, 5 years from now. Among the 31 prime ministers, a majority of them was born during 1900s which corresponds to formative years of Australia's federation and also reflects the quick development the country was experiencing during that period. While all the 8 prime ministers who are still alive were born around or after 1950s, indicating a shift towards modern-generation leadership.

Noticing that the lines for the more recent prime ministers are typically longer than the ones in 1850s. This would reflect Australia's healthcare system developed with the world-wide economic blooms and invention of new technologies (2023a). This is also one of the reasons why this study can be important to society as the health of the ministers can indicate a broader epitome of the population's health trends.

4. Reflection on this Exercise

Doing this Exercise, I found the most time consuming part is actually preparing the data scrapped from Wikipedia into the format we want. Since the example provided used the United Kingdom's data, which is slightly different with Australia's so it took me awhile to extract the columns we need, and then separate them into the right format. During this data cleaning process, there were many trial and errors but successfully I was able to transform the raw data into the correct format and conduct graphical analysis.

It became fun every time when I figured out why my code indicates an error and the process of fixing it. Although it can be frustrating at first but the excitement every time I debugged an error made me enjoy this process. If I would redo this exercise again, I think I would carefully examine the prime minister's data at the first hand. In this way, having a clearer understanding of the data can better prepare me extract and separate it more efficiently. As during this exercise I frequently need to go back to the website and check whether I got the variable information correct. However, this exercise enables me to practice the skill of web-scraping practicality and overall I think it was a fun experience.

5. Reference

2023a. *James Cook University*. https://online.jcu.edu.au/blog/impact-of-technology-on-healthcare-in-australia#:~:text=Technology%20in%20Australian%20healthcare%20is,skills%20for%20managing%20patient%20care.

———. 2023b. *Wikipedia*. Wikimedia Foundation. https://en.wikipedia.org/wiki/List_of_prime_ministers_of_Australia.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Grolemund, Garrett, and Hadley Wickham. 2011. "Dates and Times Made Easy with lubridate." *Journal of Statistical Software* 40 (3): 1–25. https://www.jstatsoft.org/v40/i03/.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2022. *Rvest: Easily Harvest (Scrape) Web Pages*. https://rvest.tidyverse.org/.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.