Lifespans of the Prime Ministers of Australia*

Mini-Essay 5a

Sima Shmuylovich

February 6, 2024

In this paper we are interested in how long prime ministers of the Australia lived, based on the year they were born. We will scrape data from Wikipedia using rvest, clean it, and then make a graph.

Table of contents

1 Results

1.0.1 Australian Prime Ministers and their Respective Birth and Death Years.

Table 1: Australian Prime Ministers and their Respective Birth and Death Years.

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

*Code and data are available at: https://github.com/sshmuylovich/australia-pm.git.

Prime Minister	Birth year	Death year	Age at death
Andrew Fisher	1862	1928	66
Joseph Cook	1860	1947	87
Billy Hughes	1862	1952	90
Stanley Bruce	1883	1967	84
James Scullin	1876	1953	77
Joseph Lyons	1879	1939	60
Earle Page	1880	1961	81
Robert Menzies	1894	1978	84
Arthur Fadden	1894	1973	79
John Curtin	1885	1945	60
Frank Forde	1890	1983	93
Ben Chifley	1885	1951	66
Harold Holt	1908	1967	59
John McEwen	1900	1980	80
John Gorton	1911	2002	91
William McMahon	1908	1988	80
Gough Whitlam	1916	2014	98
Malcolm Fraser	1930	2015	85
Bob Hawke	1929	2019	90
Paul Keating	1944	NA	NA
John Howard	1939	NA	NA
Kevin Rudd	1957	NA	NA
Julia Gillard	1961	NA	NA
Tony Abbott	1957	NA	NA
Malcolm Turnbull	1954	NA	NA
Scott Morrison	1968	NA	NA
Anthony Albanese	1963	NA	NA

1.0.2 The Lifespan of Australian Prime Ministers.

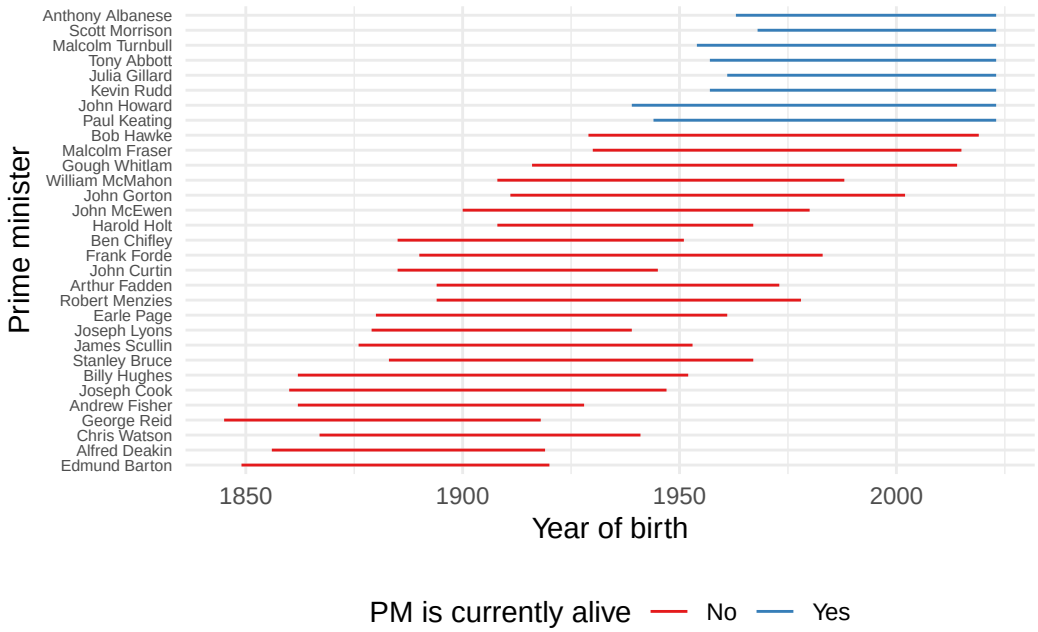

Figure 1: The Lifespan of Australian Prime Ministers.

2 Discussion

This study highlights several key findings on the lifespans of Australian Prime Ministers. Firstly, it outlines a clear timeline of birth and death years, providing a historical context for each leader's tenure. The analysis suggests a potential increase in lifespan over generations, possibly reflecting improvements in healthcare and living conditions. Moreover, the research touches upon outliers and their impact on the overall lifespan trends. This in-depth examination not only sheds light on the personal health histories of these political figures but also mirrors broader societal changes affecting longevity in Australia. This study offers valuable insights into the historical and health-related aspects of Australia's leaders.

My experience with web scraping for this study involved collecting data from Wikipedia on the birth and death years of Australian Prime Ministers. This task required me to navigate challenges such as understanding the structure of Wikipedia's data, modifying scripts taught introduced in the textbook, and ensuring the accuracy and completeness of the gathered information. Through this process, I gained practical skills in data collection, programming, and data analysis, providing a hands-on learning experience in handling real-world data. The project also offered insights into the historical trends of political leadership in Australia, showcasing the practical applications of web scraping in research.

The data source for this project is the Wikipedia page titled "List of Prime Ministers of Australia." The objective of this data extraction and processing project was to gather information about the Prime Ministers of Australia, specifically their names, birth and death dates, and

from there calculate the age at which they passed away. The Wikipedia page used in this study contains a table that lists the Prime Ministers' names alongside their birth and death years, as well as other information prevalent to their time in office.

To extract and process the relevant data from the Wikipedia page, the project utilized the R programming language and several libraries such as `rvest`. The extraction process involved several key steps. Initially, the project used the `read_html()` function to fetch the HTML content of the Wikipedia page and saved it locally as "pms.html" for further processing. SelectorGadget, a web scraping tool, was used to identify the specific HTML table element that contained the desired data. In this case, the table with the class "wikitable" was selected for extraction. The extracted data contained a header row with column names and some unwanted information.

To clean the data, column names were standardized using the `clean_names()` function, the "raw_text" column was selected, and the header row was removed. Duplicate rows were also eliminated to ensure data consistency. The "date" column was split into "birth" and "died" columns, taking into account variations in the format. The "born" column was also processed to accommodate different representations, such as "b." preceding birth years. Age at death was calculated by subtracting birth years from death years. Finally, any remaining duplicate rows were removed to obtain a clean dataset.

I initially wanted to conduct a Study on Canada's Prime Ministers but because of the inconsistent cell merging for the table containing the information I wished to scrape, I found it difficult to identify specific HTML table elements. I understood that I would have had to combine several scrapes from smaller HTML elements to get all the information. My next choice was to conduct a Study on New Zealand's Prime Ministers but on their Wikipedia page the information was split among more than one table. Again, I found it difficult to identify the specific HTML table elements I would have had to combine to get all the information. In the end, I chose Australia because the information I wished to scrape from Wikipedia was contained in a single table that was easy to identify using Selector Gadget.

I have past experience scraping websites in academic settings- as a Computer Science major- and in real-life- whether it be for a personal project or during a summer internship. This was my first time scraping a website in R however, and While the functions used were different, I enjoyed how similar the process was at its core or perhaps better said "in theory". Next time, I will challenge myself more and go for a more difficult web scrape so that I may learn more through the process of trying.

3 References

https://tellingstorieswithdata.com/07-gather.html#exercises

https://en.wikipedia.org/wiki/List_of_prime_ministers_of_Australia

https://rvest.tidyverse.org/articles/selectorgadget.html

https://rvest.tidyverse.org/

https://www.tidyverse.org/

https://www.r-project.org/