Table of Canadian Prime Minister Lifespans Scraped from Wikipedia*

Week 7 tutorial

Francesca Ye

February 28, 2024

Table of contents

1 Lifespan Table for Each Australian Prime Minister

Table 1: Full Table of Australian Prime Ministers and their Age at Death

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

*Data and code can be accessed at https://github.com/Francesca-Ye/aus_pm_lifespan

Table 1: Full Table of Australian Prime Ministers and their Age at Death

Prime Minister	Birth Year	Death Year	Age at Death
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

2 Timeline Graph for Each Australian Prime Minister's Lifespan

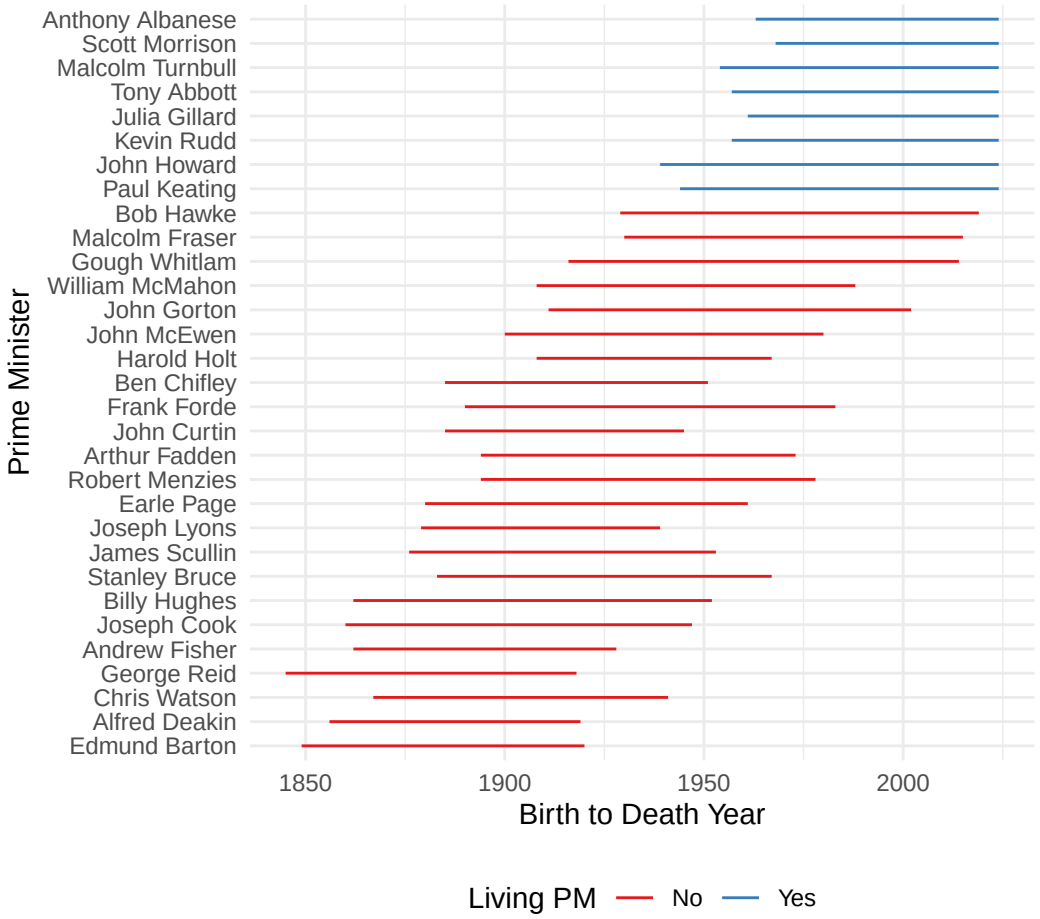

Figure 1: Australian Prime Ministers from their Birth to Death Years

3 Findings

As seen in Table 1, there have been 31 different Australian Prime Ministers since the office was created in 1901. 8 of the 31 prime ministers are currently alive as of February 2024. They are Paul Keating, John Howard, Kevin Rudd, Julia Gillard, Tony Abbott, Malcolm Turnbull, Scott Morrison and Anthony Albanese, the latter being the incumbent prime minister. The average lifespan of Australian prime ministers based on current data is approximately 77.9 years based on the lifespan of the 23 deceased prime ministers. As well, the median lifespan is 80 years based on the data. The modes of the lifespans of deceased Australian prime ministers are 66, 90, 84, 60, and 80. Two prime ministers for each aforementioned number have passed away at those ages.

Gough Whitlamis the prime minister with the longest recorded lifespan at 93 years. He was born in 1916 and passed away in 2014. Harold Holt is the prime minister with the shortest recorded lifespan at 59 years. He was born in 1908 and passed away in 1967. George Reid is the chronologically oldest of all prime ministers as he was born in 1845. He was also the first ever Australian prime minister to pass away in 1918. The chronologically youngest of all prime ministers is Scott Morrision who was born in 1968. The most recently deceased prime minister was Bob Hawke in 2019.

In Figure 1, the timeline graph that shows the length of time between a prime minister's birth and death year. From graphical interpretation, the graph highlights how Australian prime ministers are getting born later than their predecessors which has been a consistent trend throughout the graph. As well, the graph highlights the relatively long life prime ministers generally live as they are alive during the same time as multiple successors.

4 Data Description and Learning

The data was scraped from a Wikipedia page titled "List of prime ministers of Australia" on February 27, 2024 (Wikipedia 2024). Wikipedia itself uses resources to write their content, thus, the table on Wikipedia references the Australian National University's *Australian Dictionary of Biography* and the Museum of Australian Democracy as the sources used to inform the information placed in each prime minister's table row. A majority of the references were retrieved on September 1, 2022 (Wikipedia 2024).

The data gathered from the Wikipedia page to create the graph and table were the names of each Australian prime minister, their year of birth, and year of death if applicable. Data was scraped and cleaned using the statistical programming language, R (R Core Team 2023). Code used to download and clean the data was heavily referenced from Chapter 7 of *Telling Stories with Data* (Alexander 2023). Using this data, the lifespan of deceased prime ministers was calculated by subtracting their year of birth from their year of death. The SelectorGadget tool (Wickham 2022) was used to find the html element name of the table that needed to be

scraped. The full html table was read and saved using the `rvest` (Wickham 2022) and `xml2` (Wickham, Hester, and Ooms 2023) packages.

To clean the data, the `janitor` (Firke 2023), `dplyr` (Wickham et al. 2023), `tidyr` (Wickham, Vaughan, and Girlich 2023), `stringr` (Wickham 2023), `retroharmonize` (Antal 2021) and `knitr` (Xie 2014) packages were used. The table had all irrelevant columns rows with duplicated data removed. The content from the column rows were split into prime minister name, date, and born as a new data frame. Finally, the date column was split into birth and death to be then used to populate the lifespan column values. For living prime ministers, only their birth was recorded. After all this cleaning, a cleaned data set was created with columns featuring the prime minister's names, birth, death, and age at death values. The graph was created using `ggplot` (Wickham 2016) and the paper was formatted using `tinytex` (Xie 2023)

Reflecting on this exercise, cleaning the data took longer than expected because the Wikipedia page for the Australian prime ministers list differed slightly from the UK prime ministers list that was used in *Telling Stories with Data*. As a result, the code did not work without changes to the logic and `if statements`. This meant taking the time to understand the code in the example and applying similar concepts to the Australian data set took a lot longer than I expected. There were some points of the data scraping that was fun. Using the SelectorGadget and looking at the different html classes was entertaining. Creating the table and graph were also fun as it did not require learning new R concepts. One thing that I would do differently next time is to refine the logic used to split the values as I personally got slightly confused when debugging. Another thing that I would do differently next time is maybe add the dates of each prime minister's time in the position to see how much of their life was spent in office.

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC.

Antal, Daniel. 2021. *Retroharmonize: Ex Post Survey Data Harmonization*. https://CRAN.R-project.org/package=retroharmonize.

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

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://CRAN.R-project.org/package=rvest.

———. 2023. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://CRAN.R-project.org/package=stringr.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Jim Hester, and Jeroen Ooms. 2023. *Xml2: Parse XML*. https://CRAN.R-project.org/package=xml2.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2023. *Tidyr: Tidy Messy Data*. https://CRAN.R-project.org/package=tidyr.

Wikipedia. 2024. "List of Prime Ministers of Australia." https://en.wikipedia.org/wiki/List_of_prime_ministers_of_Australia.

Xie, Yihui. 2014. "Knitr: A Comprehensive Tool for Reproducible Research in R." In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC.

———. 2023. *Tinytex: Helper Functions to Install and Maintain TeX Live, and Compile LaTeX Documents*. https://github.com/rstudio/tinytex.