Life span of prime ministers of Australia*

Data analysis of life span of prime ministers of Australia until 2024-2-6

Shipeng Zhang

February 6, 2024

1 Table and Graph

Table 1: Prime Ministers Death (until 2024-2-6)

Table 1: Life span of Prime ministers of Australia

Name	Born Date	Died Date	Age of 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
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

*Code and data are available at: https://github.com/zero616/Mini-essay-5a.

Name	Born Date	Died Date	Age of Death
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

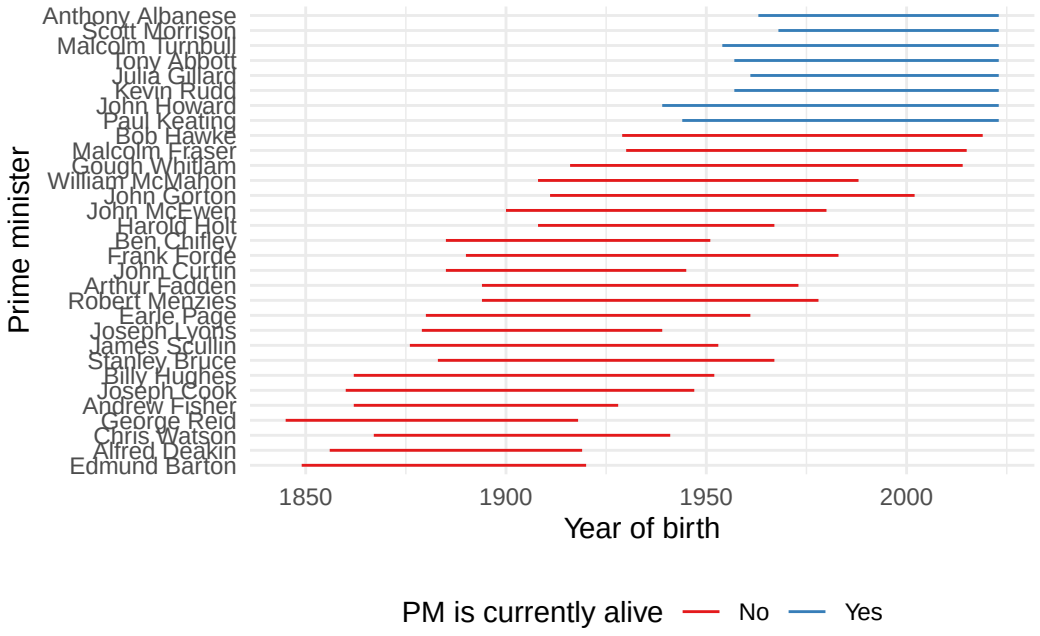

Figure 1: Death analysis by year

2 Findings

2.1 From table

Table 1 above provides a complete list of the life spans of Australian Prime Ministers, including their names, dates of birth and deaths. By analyzing this data, we can look at the life span of Australian politicians over the years.

2.2 Data sources and collection process

The data comes from the Wikipedia page titled "List of prime ministers of Australia". The web scraping process involves extracting information from the HTML structure of the page, specifically targeting the relevant tables containing the Prime Minister's details. The source data table contains very detailed information such as name, term, party affiliation, constituency, etc. However, in order to study the life span of Australian prime ministers, I deleted all data except name-year of birth and death. After some sorting, I summarized four types of data including the Prime Minister's name, year of birth and death, and age at death.

2.3 Challenge and Enjoyment

One challenge encountered along the way was using regular expressions to get the Prime Minister's name and birthday. I had to start working on the basics of regular expressions. I also had difficulties trying to distinguish living prime ministers and had to spend a lot of time solving them.

Despite the challenges, the process became fun as I dug deeper into exploring and visualizing the data. After looking at the images of each prime minister's life span, I discovered some deeper research themes from the images that could be analyzed. For example, the distribution pattern of prime life span, factors that may affect life span, etc. This discovery fueled my curiosity about the stories behind the data.

2.4 Reflection and future improvements

Some improvements I might make in future similar projects include more flexible data cleaning methods and more complex regular expressions to handle different formats of information. In addition, I will also consider introducing some automated tools to help identify and deal with potential data quality issues and improve the efficiency of the entire process.

References

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

Allaire, JJ, Yihui Xie, Christophe Dervieux, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, et al. 2023. *Rmarkdown: Dynamic Documents for r*. https://github.com/rstudio/rmarkdown.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/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. 2022. *Rvest: Easily Harvest (Scrape) Web Pages*. https://rvest.tidyverse.org/.

———. 2023a. *Stringr: Simple, Consistent Wrappers for Common String Operations*. https://stringr.tidyverse.org.

———. 2023b. *Tidyverse: Easily Install and Load the Tidyverse*. https://tidyverse.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.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown.

Xie, Yihui, Christophe Dervieux, and Emily Riederer. 2020. *R Markdown Cookbook*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown-cookbook.