Mini Essay 5a*

Hritik Shukla

February 6, 2024

1 Scraping the Data

Using R (R Core Team 2022) and the rvest library (Wickham 2022), the list of Australian Prime Ministers and their birth and death years were scraped from Wikipedia (*List of Prime Ministers of Australia*, n.d.). First the raw HTML data was scraped and saved locally to avoid sending repeated calls to Wikipedia's servers. From within this data, the ".wikitable" element was extracted, which was an HTML table which contained all of the data regarding Prime Ministers.

NThe table obtained from the raw HTML file contained our required information in a single column, in the either format "Name(YOB-YOD)Constituency" for deceased Prime Ministers or "Name(b. YOB)Constituency" for PMs who are currently alive. To extract this information into a usable data frame, we used the following libraries - tidyverse (Wickham et al. 2019), dplyr (Wickham et al. 2023), janitor (Firke 2023).First, the column was extracted from the raw data, and the information was separated into individual columns. Finally, the age of the prime ministers was calculated from the given year of birth and year of death. Note that these ages are not completely accurate as we have only been provided with their year of birth and death, and not the particular dates.

2 Exploring the Data

After we cleaned up our data, we ended up with the dataset shown in Table 1. Some of the challenges encountered while acquiring this data was separately managing the ages of still alive and deceased Prime Ministers, as both cases had different ways to representing their birth relevant (and death) years. Figuring out which element in the HTML file contained the relevant information was much easier when it was done on the locally saved file, saved as

*Code and data are available at: https://github.com/hritikshuklas/miniessay5a

Table 1: Lifespan Data

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

"pms.html" in the data folder within inputs. Once these finer details were figured out, it was smooth sailing from that point on.

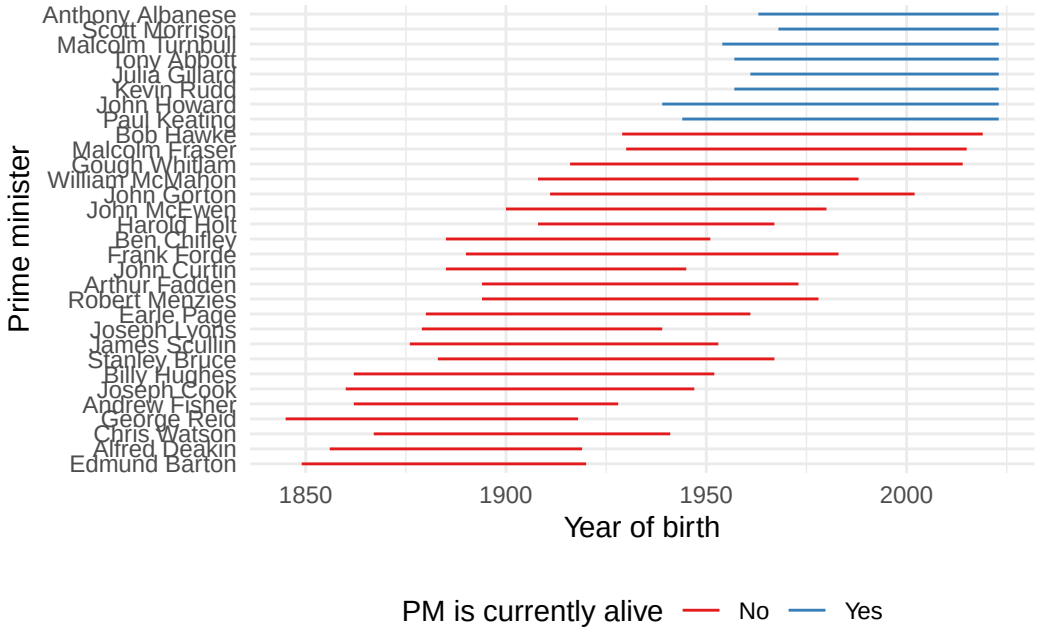

Figure 1: Lifespans of Australian Prime Ministers

Looking at Figure 1, we can see the deceased Prime Ministers' lifespans are depicted in read, and currently alive Prime Ministers are shown in blue. The Prime Ministers are listed from most recent to oldest, in terms of when they held the office. While there isn't a huge age gap between consecutive prime ministers generally, there are some outliers that stand out, such as from Chris Watson to George Reid, as well as from Stanley Bruce to James Scullin. There is no trend in preference for a younger or older consecutive prime minister. The age preference is seemingly random.

References

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

List of Prime Ministers of Australia. n.d. https://en.wikipedia.org/wiki/List_of_prime_ministers_of_Australia.

R Core Team. 2022. *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://CRAN.R-project.org/package=rvest.

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