Prime Ministers of Australia*

Tutorial 7

Samantha Barfoot

February 27, 2024

1 Findings

Figure 1 graphs the life span of all former and current Australian prime ministers with the prime ministers that are still alive marked in blue. Whereas, the deceased presents are the ones marked in red. The prime ministers are also graphed in order from the first prime minister to the current prime minister starting at the bottom and making our way to the top.

Looking at Figure 1 we can make a few observations. We can see that the majority of the prime ministers live for a long time. This is interesting given that the timeline is fairly large. The first prime minister was in 1901 and the most recent one was in 2022. From the list of prime ministers, we can also see that there has only been one female prime minister in this 120-year range, Julia Gillard.

One interesting aspect that, just by looking at the date of birth, we can see in Figure 1 is how the ages of election greatly range. For example, the 26th prime minister, Kevin Rudd, was born in 1957 and the 29th prime minister, Malcolm Turnball was elected 10 years and three prime ministers later despite being even older than Kevin Rudd. I think that it is interesting how the age of a given prime minister does not affect them being elected. We can also see that there are a fair amount of former prime ministers still alive, eight of them in total including the current prime minister.

Some aspects that we can not see on Figure 1 are the prime minister's affiliated political party, their age at death, and age at appointment. All of these would be interesting aspects that could be added in the future.

The analysis and graphs in this assignment were created with R in RStudio (R Core Team 2023). To create Figure 1 I used the packages `tiddyverse` (Wickham et al. 2019), `rvest` (Wickham 2024), `xml2`(Wickham, Hester, and Ooms 2023), `janitor` (Firke 2023), `knitr` (Xie

*Code and data are available at:

2021), and `ggplot2` (Wickham 2016). I scraped data from the Wikipedia page *List of Prime Ministers of Australia* using code from Alexander (2023).

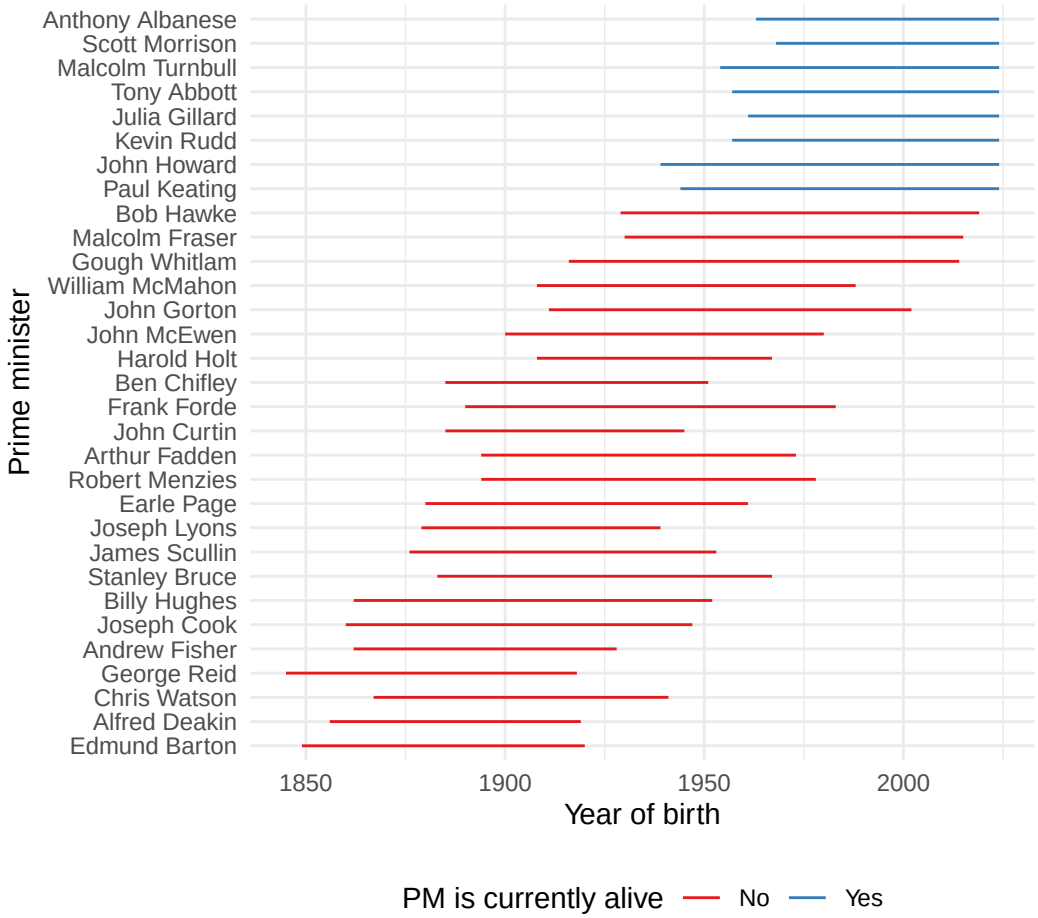

Figure 1: Graph of all Australian Prime Ministers and their life span

2 Data

The data used to create this graph was web scraped from the Wikipedia page *List of prime ministers of Australia* using `rvest` using `read_html()`. SelectorGadget was then used to select the table of prime ministers from the page to use as the data. This table contained 31 prime ministers and included their portraits, names, birth and death dates, office terms, and other data that were not relevant to the assignment. For Figure 1 we only need to know the prime minister's name, birth date, and death date.

After web scraping the data needed to be cleaned extensively. Initially, a tibble was created with the uncleaned data. Then the data was cleaned by removing all information other than

the prime minister's name and birth and death dates. Next, the birth and death dates were separated into two different columns, and the age at death of the prime minister was calculated. For the prime ministers that were still alive, the death year and age at death columns were filled as "NA". From this tibble, the graph could be created as seen in Figure 1.

3 What Took Longer Than Expected

One aspect of this assignment that took longer than expected was plotting the prime ministers who were still alive. Since they had no death date this created difficulties in trying to display their lifespan on the graph. Initially, they would not appear on the graph at all. Eventually, I realized that this was because the string being extracted from the scrapped data was not being cleaned and extracted properly. This is because in the script I was assuming that they would have a four-character death date included in the original table. After fixing this issue they were properly plotted on Figure 1.

4 When It Became Fun

One aspect of the assignment that I liked was practicing web scraping. I found it rally interesting that with a few lines of code, I could collect all the data I needed for Figure 1. This is something that I have never done before so I found it really interesting.

5 What I Would Do Differently

Something that I would do differently would be to also include the political party each prime minister was affiliated with. I think that it would be interesting to see how the political party changed over the century. It would also be interesting to see the frequency of how many prime ministers were a part of each party. This could be done by including this information in the same graph or creating a separate one just for political parties.

References

Alexander, Rohan. 2023. *Telling Stories With Data*. https://tellingstorieswithdata.com.

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

———. 2024. *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.

Wickham, Hadley, Jim Hester, and Jeroen Ooms. 2023. *Xml2: Parse XML*. https://xml2.r-lib.org/.

Xie, Yihui. 2021. "Knitr: A General-Purpose Package for Dynamic Report Generation in R." https://yihui.org/knitr/.