

# Tutorial 5\*

## Prime Ministers of Australia Information

Sinan Ma

February 6, 2024

### Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data</b>	<b>2</b>
<b>3</b>	<b>Method</b>	<b>3</b>
<b>4</b>	<b>Discussion</b>	<b>4</b>
4.1	Data Source and Gathering . . . . .	4
4.2	Data Cleaning . . . . .	4
4.3	Challenges . . . . .	4
	<b>References</b>	<b>5</b>

## 1 Introduction

This project analyzes the lifespans of Australian Prime Ministers using Wikipedia data. Through the process of gathering, cleaning, and analyzing this data, it is aiming to show insights into the country's political leadership patterns and historical governance dynamics.

The data is gathered from *List of Prime Ministers of Australia* (Wikipedia 2023). We used the following packages in R(R Core Team (2022)) to analyze the data: Wickham et al. (2019), Xie (2023), Müller (2020), Firke (2023), Wickham et al. (2023), and the knowledge from *Telling Stories with Data* (Alexander 2023).

---

\*Code and data are available at: <https://github.com/Sinanma/TUT-5>

## 2 Data

The table (Table 1) lists the birth and death years of Australian Prime Ministers, along with their ages at death. Some of recent Prime Ministers are still alive, as indicated by the ‘NA’ in their death year and age at death columns, reflecting the recency of their terms in office.

Table 1: How long each prime minister of the Australia lived

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
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

The graph(Figure 1) shows the lifespans of Australian Prime Ministers, with red lines representing those who have passed away and blue for those who still alive. It have a time period from the 19th to the 21st century, and illustrates a historical overview of their lives in relation to their years of birth.

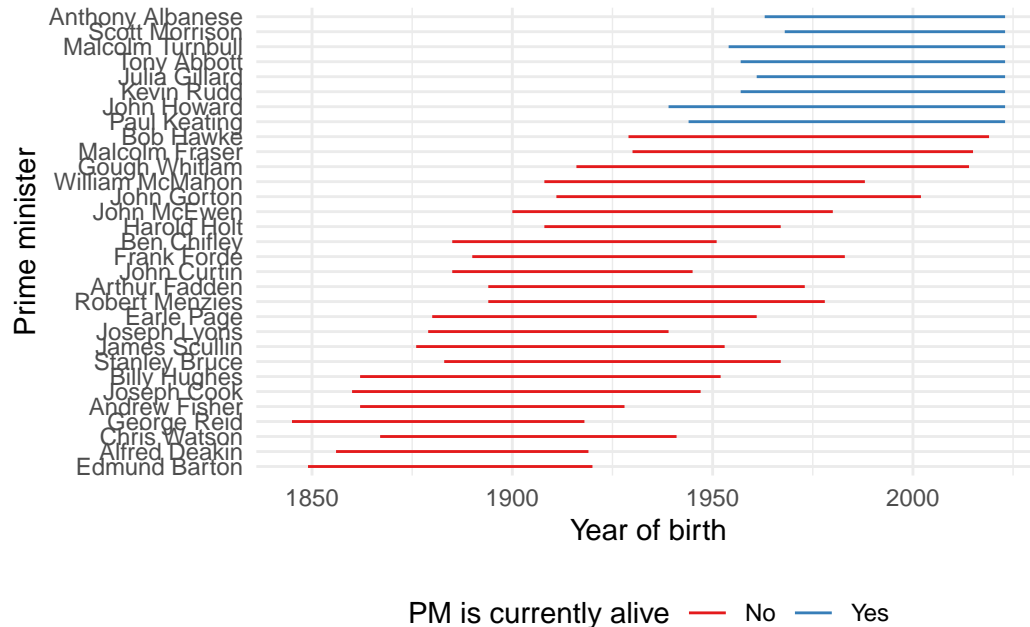


Figure 1: How long each prime minister of the Australia lived

#Result From the data we can determine that there was no clear correlation between the lifespan of Prime Ministers and their time role, indicating that factors determining political longevity are varied. The tenure patterns provided insights in significant political events, such as world wars and economic crises, that may led to shifts in leadership.

### 3 Method

Data was gathered using web scraping techniques. Tools such as rvest(Wickham 2022) in R is used to extract the required information from the website page.

## **4 Discussion**

### **4.1 Data Source and Gathering**

Wikipedia is a rich source of all information, but extracting structured data from HTML needed to pay more attention detail, especially in correctly identifying and analyzing the relevant table elements.

### **4.2 Data Cleaning**

After gathering the data, the information has to be cleaned and structured to be useful for analysis. This process involved: extracting individual data points (names, birth and death years), handling missing values, and etc..

### **4.3 Challenges**

The most time-consuming part was cleaning the data, especially handling the various formats of dates and dealing with missing information. However, resolving these issues is quite engaging for me.

## References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Boca Raton: CRC Press. <https://tellingstorieswithdata.com/>.
- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://github.com/sfirke/janitor>.
- Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. <https://here.r-lib.org/>.
- 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://rvest.tidyverse.org/>.
- Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Golemund, 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>.
- Wikipedia. 2023. “List of Prime Ministers of Australia.” [https://en.wikipedia.org/wiki/List\\_of\\_prime\\_ministers\\_of\\_Australia](https://en.wikipedia.org/wiki/List_of_prime_ministers_of_Australia).
- Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.