The Prime Ministers of Canada*

Alaina Hu

February 4, 2024

Data

Prime Minister	Birth Year	Death Year	Age at Death
John A. Macdonald	1815	1891	76
Alexander Mackenzie	1822	1892	70
John Abbott	1821	1893	72
John Thompson	1845	1894	49
Mackenzie Bowell	1823	1917	94
Charles Tupper	1821	1915	94
Wilfrid Laurier	1841	1919	78
Robert Borden	1854	1937	83
Arthur Meighen	1874	1960	86
William Lyon Mackenzie King	1874	1950	76
R. B. Bennett	1870	1947	77
Louis St. Laurent	1882	1973	91
John Diefenbaker	1895	1979	84
Lester B. Pearson	1897	1972	75
Pierre Trudeau	1919	2000	81
Joe Clark	1939	NA	NA
John Turner	1929	2020	91
Brian Mulroney	1939	NA	NA
Kim Campbell	1947	NA	NA
Jean Chrétien	1934	NA	NA
Paul Martin	1938	NA	NA
Stephen Harper	1959	NA	NA
Justin Trudeau	1971	NA	NA

*Code is available at: https://github.com/alainahu/mini_essay5

Reflection

From the data shown above, I learned much about the lifespans and history of the Canadian prime ministers. Firstly, I discovered that the first Canadian prime minister was born in 1815, meaning that the history of Canadian prime ministers began after the start of the U.S presidency. Through web scraping and cleaning the data, I learned that there have been 23 different prime ministers with many prime ministers serving multiple terms or returning to office after leaving office. Prime Minister John Thompson lived for 49 years, and Prime Ministers Mackenzie Bowell and Charles Tupper both lived for 94 years. Additionally, I learned that many former prime ministers are still alive; 7 prime ministers are still alive today.

Data for the birth and death years for the Canadian prime ministers was web scraped from the Wikipedia page titled `List of prime ministers of Canada` (Wikipedia contributors 2024). Data gathering and data cleaning was conducted in `R` (R Core Team 2022) with help from libraries such as `tidyverse` (Wickham et al. 2019), `dplyr` (Wickham et al. 2023), `knitr` (Xie 2023), `rvest` (Wickham 2022), and `janitor` (Firke 2023). The web scraping process was guided by code provided in Chapter 7 of `Telling Stories with Data` (Alexander 2023). For this analysis, I gathered data from a table on the Wikipedia page. Since all in the information I needed about the birth and death years of the Canadian prime ministers was located on one table, I only had to gather information from that table on the page, easily done with the `Selector Gadget`. However, I ran into some trouble when trying to gather information from this table because there were two tables with the same element tag. The `Selector Gadget` was retrieving the table right above the table I wanted, so I had to specify that I wanted the second table element on the page and modify my original code. After successfully retrieving the table I wanted, I began my data cleaning process. For this paper, I only needed information on the names, birth years, and death years of prime ministers, so I filtered out unnecessary columns. To further enhance my data, I cleaned up the values of each column, separated birth years and death years, and calculated the ages at death for prime ministers that had passed away.

The web scraping and data cleaning process started becoming fun when I had successfully separated the birth and death years into multiple columns. Originally, the prime ministers' name, birth year, and death were joined together in one single column. First, I separated the names from the years, then I separated the birth and death years for the prime ministers. Achieving this step was fun because my table started looking like the table in the web scraping example with U.K prime minister data. Data cleaning took longer than I expected. Initially, I expected the lengthiest process to be scraping the data from the website, but surprisingly, cleaning the data took much longer than retrieving the data. In the future, I would dedicate more time to cleaning the data and have more realistic expectations for the time needed to thoroughly gather the data I need.

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman & Hall.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://github.com/sfirke/janitor.

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 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.

Wikipedia contributors. 2024. "List of Prime Ministers of Canada." https://en.wikipedia.org/wiki/List_of_prime_ministers_of_Canada.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.