Lifespans of Prime Ministers of Canada*

Irene Huynh

February 6, 2024

1 Planning, Gathering, and Cleaning Sources

The statistical programming language R was used to retrieve, clean, and process the data (R Core Team 2022). In particular, the following R packages were used: tidyverse (**tidyverse?**), rvest [], and xml2 [] for data acquisition, knitr (**knitr?**), janitor [], dplyr [], and here [] for data cleaning and processing, and ggplot2 (**ggplot2?**) for creating figures.

2 Planning

A simulation of the expected data set is found below, having the Prime Minister's name, their birth year, death year, and years lived as columns. The simulation was created using the following R packages: tidyverse [], babynames [], and pdftools [].

```
# A tibble: 10 x 4
   prime_minister birth_year death_year years_lived
   <chr>               <int>      <int>       <int>
 1 Kevin                1813       1908          95
 2 Karen                1832       1896          64
 3 Robert               1839       1899          60
 4 Bertha               1846       1915          69
 5 Jennifer             1867       1943          76
 6 Arthur               1892       1984          92
 7 Donna                1907       2006          99
 8 Emma                 1957       2031          74
 9 Ryan                 1959       2053          94
10 Tyler                1990       2062          72
```

*Code is available at: https://github.com/IreneeeH/STA302_Mini-Essay_5a.git

A sketch of the expected graph can be found below:

3 Gathering

The data utilized in this paper was retrieved from a Wikipedia page of the List of Prime Ministers of Canada [] using the rvest R package [].

There is another element in pms.html with a wikitable class that precedes the desired table of Prime Ministers of Canada. This caused html_element(".wikitable") to use that table instead of the desired one. To resolve this issue, I went into pms.html and removed the wikitable class from that table. Now, html_element(".wikitable") should be able to identify the correct wikitable.

4 Cleaning

While using the original code provided in Telling Stories With Data - 7 Gather data [], cleaning the data was difficult as the birth dates of Prime Ministers who are still alive did not appear. After some trial and error, I realized that the "born" attribute must be manipulated. Additionally, the finalized cleaned_data included a final row with extra information from the html file. To remove this, I had to filter it out.

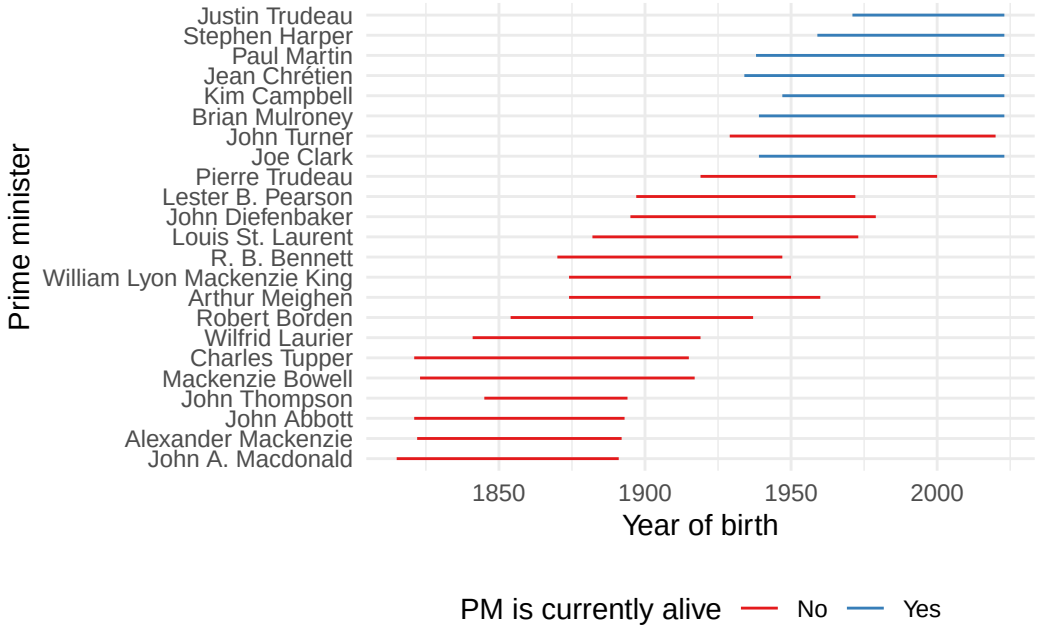

Figure 1: The lifespan of all Prime Ministers of Canada

5 Results

Figure 1 reveals that John Thompson had the shortest lifespan, Additionally, 7 Prime Ministers are still alive.

6 Results

References

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.