Tutorial 7: Indian PMs Scraping Example*

INF312: Worlds Become Data - Prof. Rohan Alexander

Luca Carnegie

February 27, 2024

Please redo the web scraping example, but for one of: Australia, Canada, India, or New Zealand. Use Quarto, and include an appropriate title, author, date, link to a GitHub repo, and citations. Submit a PDF. {.unnumbered}

My Choice - India.

0.1 Simulate Data

Our goal is a table that looks somewhat like this:

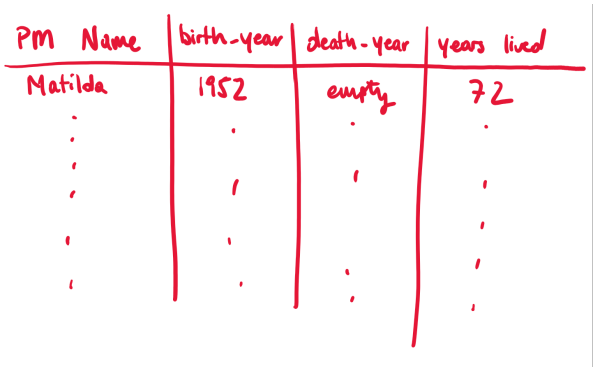

Figure 1: goal

Moving into the R environment, we are aiming for a table containing the Prime Minister's name, birth year, death year (if they are dead), as well as lifespan (in years). If they are dead, the death year is empty. To simulate, we do exactly as what is done in the textbook.

*Code available at: https://github.com/lcarnegie/INF312/tree/main/W7%20-%20Gather%20Data

```r
set.seed(853)

simulated_dataset <-
  tibble(
    prime_minister = babynames |>
      filter(prop > 0.01) |>
      distinct(name) |>
      unlist() |>
      sample(size = 10, replace = FALSE),
    birth_year = sample(1947:1990, size = 10, replace = TRUE),
    years_lived = sample(50:100, size = 10, replace = TRUE),
    death_year = birth_year + years_lived
  ) |>
  select(prime_minister, birth_year, death_year, years_lived) |>
  arrange(birth_year)

simulated_dataset
```

```
# A tibble: 10 x 4
   prime_minister birth_year death_year years_lived
   <chr>               <int>      <int>       <int>
 1 Ryan                 1949       2000          51
 2 Donna                1950       2022          72
 3 Emma                 1958       2052          94
 4 Jennifer             1964       2033          69
 5 Bertha               1965       2030          65
 6 Kevin                1969       2023          54
 7 Tyler                1981       2032          51
 8 Robert               1983       2033          50
 9 Karen                1983       2078          95
10 Arthur               1986       2046          60
```

This gives us a good goalpost to aim for as we are scraping data from Wikipedia on Indian Prime Minister data.

0.2 Gathering and Cleaning Data

Data is scraped from Wikipedia's list of Indian Prime Ministers

0.3 Creating the Table

0.4 Discussion/Reflection

0.4.1 Data Source

0.4.2 My Findings

0.4.3 Reflections