

An Analysis of the lifespan of Indian Prime Ministers*

Aryaman Sharma

February 5, 2024

Table of contents

1	Introduction	1
2	Data	2
3	Results	3
4	Reflection	3
	References	4

1 Introduction

India got independence in 1947 and since then there have been 14 prime ministers that the country has seen (*List of Prime Ministers of India* 2024). India has seen a majority of civil unrest over the years which have caused Prime Ministers from getting assassinated, namely Indira Gandhi and Rajiv Gandhi.

This report gives us the opportunity to delve deeper into how these assassinations affect the overall trend of Indian Prime Ministers' lifespans. In this paper, we made use of web scraping to obtain data from the Wikipedia page of a list of Indian Prime Ministers (*List of Prime Ministers of India* 2024). It is essential to note that essential principles of web scraping were followed in this report.

*Code and data are available at: <https://github.com/Ary4m3n/lifespan-indian-pms.git>

This paper is structured using the following sections: Data, Results, Reflection. In the Data (Section 2) section, the process of obtaining the data by web scraping and cleaning it is outlined. In the Results (Section 3) section, the data findings and the plot has been discussed. The paper ends with the Reflection (Section 4) section, where I have reflected on the process of web scraping.

2 Data

As mentioned above, the data analysed in this paper was obtained from web scrapping the Wikipedia page cited above. The data was cleaned and analysed using the open source R programming language (R Core Team 2022). R libraries and packages such as **tidyverse** (Wickham et al. 2019), **janitor** (Firke 2023), **ggplot2** (Wickham 2016), **knitr** (Xie 2023), **readr** (Wickham, Hester, and Bryan 2023), **dplyr** (Wickham et al. 2023), **xml2** (Wickham, Hester, and Ooms 2023) and **rvest** (Wickham 2022).

The first step in web scraping was to download the raw html code of the webpage. We only scraped the webpage material once and saved it as **raw_data.csv**. However, the raw data contained the code for the whole page which was unnecessary. We observed that all the data we required was in the html element called **wikitable**. Hence, we focused on the table and cleaned the names and removed unnecessary columns such as portrait, political part, head of state etc. We were specifically interested in the column that contained the name of the Prime Minister, along with the year born, year died and constituency in it. We then separated the data in each row of the column selected in order to have just the name, the important years and year born for Prime Minsters who are still alive.

The data was cleaned further to finally obtain a table showing the name of the Prime Minister, the Birth Year, the Death Year and the Age at Death as shown below in Table 1 which shows the first 6 rows of the data obtained.

Table 1: Cleaned Data showing birth year, death year and lifespan for each Indian Prime Minister

Prime Minister	Birth year	Death year	Age at death
Jawaharlal Nehru	1889	1964	75
Gulzarilal Nanda	1898	1998	100
Lal Bahadur Shastri	1904	1966	62
Indira Gandhi	1917	1984	67
Morarji Desai	1896	1995	99
Charan Singh	1902	1987	85

3 Results

A segment plot (Figure 1) was made to visualize the findings reported in Table 1.

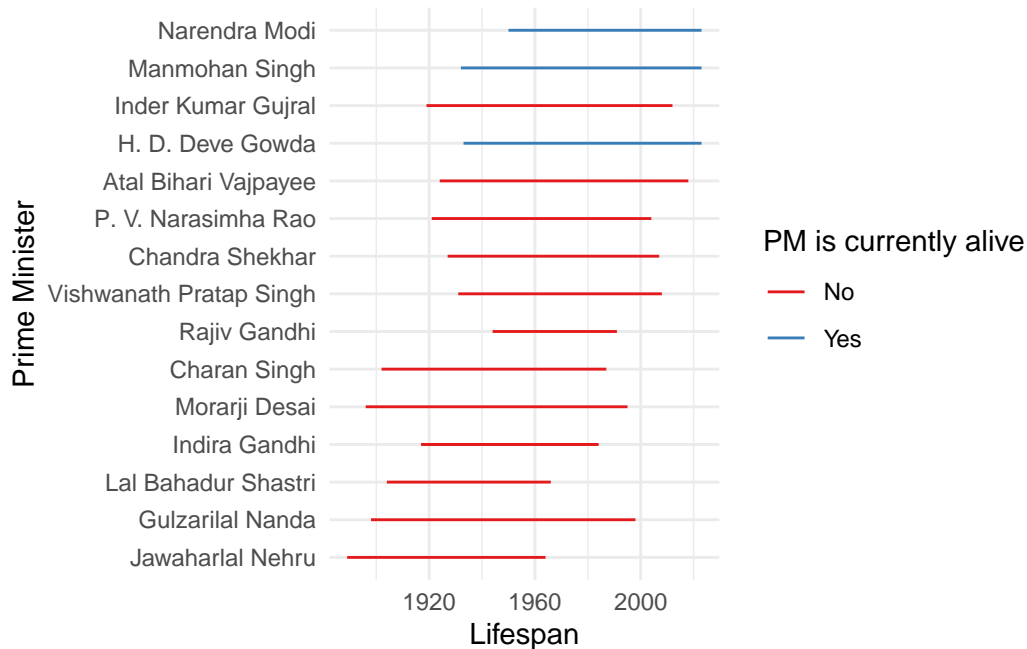


Figure 1: Lifespan of all Indian Prime Ministers

As we can see in Figure 1, the plot shows us the lifespans of all Indian Prime Ministers. Visualizing the data obtained helps us to delve deeper into assessing the different outliers in this situation. We can see that Prime Ministers Rajiv Gandhi, Indira Gandhi and Lal Bahadur Shastri are seen to have shorter lifespans when compared to others. This is because of Rajiv Gandhi and Indira Gandhi were assassinated in 1991 and 1984 respectively (*List of Prime Ministers of India* 2024). Similarly, Lal Bahadur Shastri was reported dead in Tashkent, Uzbekistan by heart attack after signing the peace treat to end the 1965 Indo-Pakistan War, but it was claimed by his wife that there was foul play involved wherein he was poisoned (*Lal Bahadur Shastri* 2024).

Visualizing the plot (Figure 1) allows us to delve deeper into analyzing such outlier cases and researching about them.

4 Reflection

Overall, this paper provided me with an opportunity to try out web scraping hands-on. In this report, cleaning the data took longer than expected as I tried to try different things and play

around with it. In general the process of web scraping seemed intuitive after the first try. The main challenge was to only extract the data needed as for that certain html elements had to be chosen. It became a lot of fun when the tables started coming together and felt extremely rewarding after making the plot.

The next time I do this, I will certainly make it a point to map out what the data on the webpage looks like to make the data cleaning process easier for myself. In general, it was a fun experience making this report.

References

- Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. <https://CRAN.R-project.org/package=janitor>.
- Lal Bahadur Shastri. 2024. *Wikipedia*. https://en.wikipedia.org/wiki/Lal_Bahadur_Shastri.
- List of Prime Ministers of India. 2024. *Wikipedia*. https://en.wikipedia.org/wiki/List_of_prime_ministers_of_India.
- 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. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2022. *Rvest: Easily Harvest (Scrape) Web Pages*. <https://CRAN.R-project.org/package=rvest>.
- 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://CRAN.R-project.org/package=dplyr>.
- Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. <https://CRAN.R-project.org/package=readr>.
- Wickham, Hadley, Jim Hester, and Jeroen Ooms. 2023. *Xml2: Parse XML*. <https://CRAN.R-project.org/package=xml2>.
- Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.