

# Loading form JSON, HTML, XML

By Tony Fraser

2023-09-30

## Assignment

Load data from within HTML, JSON and XML into data frames and compare.

## Load from HTML

```
library(rvest)
html_content <- readLines("./mountains.html", warn = FALSE)
html_content <- paste(html_content, collapse = "\n")
html_parsed <- read_html(html_content)

# Extract data using CSS selectors
titles <- html_parsed %>% html_nodes(".book h2") %>% html_text()
authors <- html_parsed %>% html_nodes(xpath = '//*[@contains(text(),"Authors:")]') %>% html_text()
descriptions <- html_parsed %>% html_nodes(xpath = '//*[@not(descendant::*[contains(text(),"Year Published:")]')]') %>% html_text()
years <- html_parsed %>% html_nodes(xpath = '//*[@contains(text(),"Year Published:")]') %>% html_text()

# Create a dataframe
df <- data.frame(Title = titles, Authors = authors, Description = descriptions, Year_Published = years)

print(df)
```

```
1                                     Title
2                                     The Wildest Dream: The Biography of George Mallory
3 Last Hours on Everest: The Gripping Story of Mallory and Irvine's Fatal Ascent
4                                     Deborah: A Wilderness Narrative
5                                     The Wildest Dream: The Biography of George Mallory
6                                     K2: Life and Death on the World's Most Dangerous Mountain
7 Last Hours on Everest: The Gripping Story of Mallory and Irvine's Fatal Ascent
8                                     Deborah: A Wilderness Narrative
```

```
Authors
1 Authors:
2 Authors:
3 Authors:
4 Authors:
5 Authors:
6 Authors:
7 Authors:
8 Authors:
```

```
1
2                                     This biography delves into the life of George Mallory, the
3
4 This book provides a detailed account of the history and allure of K2, the world's second-highest mountain.
```

5

6

Building on the story of Mallory and Irvine's ill-fated 1921

7

8

This account details the gripping

Year\_Published

1 Year Published:

2 Year Published:

3 Year Published:

4 Year Published:

5 Year Published:

6 Year Published:

7 Year Published:

8 Year Published: