

RWorksheet_Lomibao#5(lomibao and rabago)

lomibao and rabago

2024-11-11

1.

```
library("rvest")
library("polite")
library("httr")
library("dplyr")

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union
url <- "https://www.imdb.com/chart/toptv/"

session <- bow(url,
               user_agent = "Student education purpose")
session

## <polite session> https://www.imdb.com/chart/toptv/
##   User-agent: Student education purpose
##   robots.txt: 35 rules are defined for 3 bots
##   Crawl delay: 5 sec
##   The path is scrapable for this user-agent
# session_page <- scrape(session)
# page <- read_html(url)

rank_title <- character(0)
links <- character(0)

titles <- scrape(session)%>%html_nodes("h3.ipc-title__text")%>%html_text()
titles

## [1] "IMDb Charts"
## [2] "1. Breaking Bad"
## [3] "2. Planet Earth II"
## [4] "3. Planet Earth"
## [5] "4. Band of Brothers"
## [6] "5. Chernobyl"
## [7] "6. The Wire"
```

```
## [8] "7. Avatar: The Last Airbender"
## [9] "8. Blue Planet II"
## [10] "9. The Sopranos"
## [11] "10. Cosmos: A Spacetime Odyssey"
## [12] "11. Cosmos"
## [13] "12. Our Planet"
## [14] "13. Game of Thrones"
## [15] "14. Bluey"
## [16] "15. The World at War"
## [17] "16. Fullmetal Alchemist: Brotherhood"
## [18] "17. Rick and Morty"
## [19] "18. Life"
## [20] "19. The Last Dance"
## [21] "20. The Twilight Zone"
## [22] "21. The Vietnam War"
## [23] "22. Sherlock"
## [24] "23. Attack on Titan"
## [25] "24. Batman: The Animated Series"
## [26] "25. Arcane"
## [27] "Recently viewed"
```

```
title_list <-as.data.frame(titles[2:26])

split_df<-strsplit(as.character(title_list$rank), ".",fixed = TRUE)

split_df <- data.frame(do.call(rbind,split_df))

split_df <-split_df[- c(3,4)]

#colnames(split_df) <-c("ranks","title")

str(split_df)
```

```
## 'data.frame': 0 obs. of 0 variables
```

```
split_df<-strsplit(as.character(title_list$rank), ".",fixed = TRUE)
split_df <- data.frame(do.call(rbind,split_df))
split_df <-split_df[- c(3,4)]
#colnames(split_df) <-c("ranks","title")
str(split_df)
```

```
## 'data.frame': 0 obs. of 0 variables
```