

# RWorksheet#5\_group.R,

Noblezada, Camasa, Cabilia

2024-11-06

```
library(polite)
library(kableExtra)
library(rmarkdown)

url <- 'https://www.amazon.com/ref=nav_logo'

session <- bow(url,
               user_agent = "Educational")
session

## <polite session> https://www.amazon.com/ref=nav_logo
##   User-agent: Educational
##   robots.txt: 137 rules are defined for 4 bots
##   Crawl delay: 5 sec
##   The path is scrapable for this user-agent

page <- scrape(session)

## No encoding supplied: defaulting to UTF-8.

library(polite)
library(kableExtra)
library(rmarkdown)
library(rvest)

url <- 'https://www.imdb.com/search/title/?title_type=tv_series&sort=num_votes,desc'

session <- bow(url,
               user_agent = "Educational")
session

## <polite session> https://www.imdb.com/search/title/?title_type=tv_series&sort=num_votes,desc
##   User-agent: Educational
##   robots.txt: 35 rules are defined for 3 bots
##   Crawl delay: 5 sec
##   The path is scrapable for this user-agent

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

title_list <- scrape(session) %>%
  html_nodes('h3.ipc-title__text') %>%
  html_text

title_list_sub <- as.data.frame(title_list[1:50])
```

```
head(title_list_sub)
```

```
##      title_list[1:50]
## 1  1. Game of Thrones
## 2    2. Breaking Bad
## 3  3. Stranger Things
## 4        4. Friends
## 5 5. The Walking Dead
## 6        6. Sherlock
```

```
tail(title_list_sub)
```

```
##      title_list[1:50]
## 45                <NA>
## 46                <NA>
## 47                <NA>
## 48                <NA>
## 49                <NA>
## 50                <NA>
```

```
colnames(title_list_sub) <- "ranks"
```

```
split_df <- strsplit(as.character(title_list_sub$ranks), ".", fixed = TRUE)
split_df <- data.frame(do.call(rbind, split_df))
```