movieReviews.rmd

2024-02-06

```r
library(rvest)
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
```

```r
library(polite)


#1ST MOVIE(Titanic)
session <- bow(url = 'https://www.imdb.com/title/tt0120338/reviews?ref_=tt_urv',
               user_agent = "Educational")
session
```

```
## <polite session> https://www.imdb.com/title/tt0120338/reviews?ref_=tt_urv
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
  scrapeReviews <- function(page_url) {
  movieReviews <- read_html(page_url)

    #title
    movieReviews %>%
      html_nodes('.title') %>%
      html_text() -> title

    #username
    movieReviews %>%
      html_nodes('.display-name-link') %>%
      html_text() -> username

    #content
    movieReviews %>%
      html_nodes('.content') %>%
      html_text() -> content
    #date
    movieReviews %>%
      html_nodes('.review-date') %>%
      html_text() -> date

    #stars
      movieReviews %>%
        html_nodes('.rating-other-user-rating') %>%
        html_text() -> stars

    #dataframe
      titanic_df = data.frame(Title = title[1:25],
                              Username = username[1:25],
                              Content = content[1:25],
                              Date = date[1:25],
                              Stars = stars[1:25]
    )}

    #URLs
      titanic_urls<- c('https://www.imdb.com/title/tt0120338/reviews?ref_=tt_urv',
                       'https://www.imdb.com/title/tt0120338/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt0120338/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt0120338/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt0120338/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt0120338/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt0120338/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt0120338/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt0120338/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt0120338/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt0120338/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt0120338/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                      )
      allReviews <- lapply(titanic_urls, scrapeReviews)

      movieReviews1 <- do.call(rbind, allReviews)



#2ND MOVIE(Wonka)
session2 <- bow(url = 'https://www.imdb.com/title/tt6166392/reviews?ref_=tt_urv',
                     user_agent = "Educational")
session2
```

```
## <polite session> https://www.imdb.com/title/tt6166392/reviews?ref_=tt_urv
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
    scrapeReviews2 <- function(page_url) {
    movie2Reviews <- read_html(page_url)

        #title
        movie2Reviews %>%
          html_nodes('.title') %>%
          html_text() -> title

        #username
        movie2Reviews %>%
          html_nodes('.display-name-link') %>%
          html_text() -> username

        #content
        movie2Reviews %>%
          html_nodes('.content') %>%
          html_text() -> content

        #date
        movie2Reviews %>%
          html_nodes('.review-date') %>%
          html_text() -> date

        #stars
        movie2Reviews %>%
          html_nodes('.rating-other-user-rating') %>%
          html_text() -> stars

        #dataframe
        wonka_df = data.frame(Title = title[1:25],
                                Username = username[1:25],
                                Content = content[1:25],
                                Date = date[1:25],
                                Stars = stars[1:25]
        )}

      #URLs
      wonka_urls<- c('https://www.imdb.com/title/tt6166392/reviews?ref_=tt_urv',
                       'https://www.imdb.com/title/tt6166392/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt6166392/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt6166392/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt6166392/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt6166392/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt6166392/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt6166392/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt6166392/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt6166392/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt6166392/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo
                       'https://www.imdb.com/title/tt6166392/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo

      )
      allReviews2 <- lapply(wonka_urls, scrapeReviews2)

      movieReviews2 <- do.call(rbind, allReviews2)



#3RD MOVIE(The Shawshank Redemption)
```

```
session <- bow(url = 'https://www.imdb.com/title/tt0111161/reviews?ref_=tt_urv',
               user_agent = "Educational")
session
```

```
## <polite session> https://www.imdb.com/title/tt0111161/reviews?ref_=tt_urv
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
scrapeReviews3 <- function(page_url) {
  movie3Reviews <- read_html(page_url)

  #title
  movie3Reviews %>%
    html_nodes('.title') %>%
    html_text() -> title

  #username
  movie3Reviews %>%
    html_nodes('.display-name-link') %>%
    html_text() -> username

  #content
  movie3Reviews %>%
    html_nodes('.content') %>%
    html_text() -> content

  #date
  movie3Reviews %>%
    html_nodes('.review-date') %>%
    html_text() -> date

  #stars
  movie3Reviews %>%
    html_nodes('.rating-other-user-rating') %>%
    html_text() -> stars

  #dataframe
  redemption_df = data.frame(Title = title[1:25],
                             Username = username[1:25],
                             Content = content[1:25],
                             Date = date[1:25],
                             Stars = stars[1:25]
  )}

#URLs
redemption_urls<- c('https://www.imdb.com/title/tt0111161/reviews?ref_=tt_urv',
                    'https://www.imdb.com/title/tt0111161/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                    'https://www.imdb.com/title/tt0111161/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                    'https://www.imdb.com/title/tt0111161/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                    'https://www.imdb.com/title/tt0111161/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                    'https://www.imdb.com/title/tt0111161/reviews/_ajax?&paginationKey=g4w6ddbmqyz
                    'https://www.imdb.com/title/tt0111161/reviews/_ajax?&paginationKey=g4w6ddbmqyz
```

```
                          'https://www.imdb.com/title/tt0111161/reviews/_ajax?&paginationKey=g4w6ddbmqy:
                          'https://www.imdb.com/title/tt0111161/reviews/_ajax?&paginationKey=g4w6ddbmqy:
                          'https://www.imdb.com/title/tt0111161/reviews/_ajax?&paginationKey=g4w6ddbmqy:
                          'https://www.imdb.com/title/tt0111161/reviews/_ajax?&paginationKey=g4w6ddbmqy:
                          'https://www.imdb.com/title/tt0111161/reviews/_ajax?&paginationKey=g4w6ddbmqy:
      )

      allReviews3 <- lapply(redemption_urls, scrapeReviews3)

      movieReviews3 <- do.call(rbind, allReviews3)



#4TH MOVIE(Argylle)
      session <- bow(url = 'https://www.imdb.com/title/tt15009428/reviews?ref_=tt_urv',
                     user_agent = "Educational")
      session
```

```
## <polite session> https://www.imdb.com/title/tt15009428/reviews?ref_=tt_urv
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
      scrapeReviews4 <- function(page_url) {
        movie4Reviews <- read_html(page_url)

        #title
        movie4Reviews %>%
          html_nodes('.title') %>%
          html_text() -> title

        #username
        movie4Reviews %>%
          html_nodes('.display-name-link') %>%
          html_text() -> username

        #content
        movie4Reviews %>%
          html_nodes('.content') %>%
          html_text() -> content

        #date
        movie4Reviews %>%
          html_nodes('.review-date') %>%
          html_text() -> date

        #stars
        movie4Reviews %>%
          html_nodes('.rating-other-user-rating') %>%
          html_text() -> stars

        #dataframe
        argylle_df = data.frame(Title = title[1:25],
                                   Username = username[1:25],
                                   Content = content[1:25],
                                   Date = date[1:25],
                                   Stars = stars[1:25]
        )}

      #URLs
      argylle_urls<- c('https://www.imdb.com/title/tt15009428/reviews?ref_=tt_urv',
                          'https://www.imdb.com/title/tt15009428/reviews/_ajax?&paginationKey=g4w6ddbmq
                          'https://www.imdb.com/title/tt15009428/reviews/_ajax?&paginationKey=g4w6ddbmq
                           'https://www.imdb.com/title/tt15009428/reviews/_ajax?&paginationKey=g4w6ddbm
                           'https://www.imdb.com/title/tt15009428/reviews/_ajax?&paginationKey=g4w6ddbm
                           'https://www.imdb.com/title/tt15009428/reviews/_ajax?&paginationKey=g4w6ddbm
                           'https://www.imdb.com/title/tt15009428/reviews/_ajax?&paginationKey=g4w6ddbm
                           'https://www.imdb.com/title/tt15009428/reviews/_ajax?&paginationKey=g4w6ddbm
                           'https://www.imdb.com/title/tt15009428/reviews/_ajax?&paginationKey=g4w6ddbm
                           'https://www.imdb.com/title/tt15009428/reviews/_ajax?&paginationKey=g4w6ddbm
                           'https://www.imdb.com/title/tt15009428/reviews/_ajax?&paginationKey=g4w6ddbm
                           'https://www.imdb.com/title/tt15009428/reviews/_ajax?&paginationKey=g4w6ddbm
          )

      allReviews4 <- lapply(argylle_urls, scrapeReviews4)

      movieReviews4 <- do.call(rbind, allReviews4)



#5TH MOVIE(The Marvels)
      session <- bow(url = 'https://www.imdb.com/title/tt10676048/reviews?ref_=tt_urv',
                     user_agent = "Educational")
      session
```

```
## <polite session> https://www.imdb.com/title/tt10676048/reviews?ref_=tt_urv
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
      scrapeReviews5 <- function(page_url) {
        movie5Reviews <- read_html(page_url)

        #title
        movie5Reviews %>%
          html_nodes('.title') %>%
          html_text() -> title

        #username
        movie5Reviews %>%
          html_nodes('.display-name-link') %>%
          html_text() -> username

        #content
        movie5Reviews %>%
          html_nodes('.content') %>%
          html_text() -> content

        #date
        movie5Reviews %>%
          html_nodes('.review-date') %>%
          html_text() -> date

        #stars
        movie5Reviews %>%
          html_nodes('.rating-other-user-rating') %>%
          html_text() -> stars

        #dataframe
        marvels_df = data.frame(Title = title[1:25],
                                Username = username[1:25],
                                Content = content[1:25],
                                Date = date[1:25],
                                Stars = stars[1:25]
        )}

      #URLs
      marvels_urls<- c('https://www.imdb.com/title/tt10676048/reviews?ref_=tt_urv',
                             'https://www.imdb.com/title/tt10676048/reviews/_ajax?&paginationKey=g4w6ddl
                             'https://www.imdb.com/title/tt10676048/reviews/_ajax?&paginationKey=g4w6ddl
                             'https://www.imdb.com/title/tt10676048/reviews/_ajax?&paginationKey=g4w6ddl
                             'https://www.imdb.com/title/tt10676048/reviews/_ajax?&paginationKey=g4w6ddl
                             'https://www.imdb.com/title/tt10676048/reviews/_ajax?&paginationKey=g4w6ddl
                             'https://www.imdb.com/title/tt10676048/reviews/_ajax?&paginationKey=g4w6ddl
                             'https://www.imdb.com/title/tt10676048/reviews/_ajax?&paginationKey=g4w6ddl
                             'https://www.imdb.com/title/tt10676048/reviews/_ajax?&paginationKey=g4w6ddl
                             'https://www.imdb.com/title/tt10676048/reviews/_ajax?&paginationKey=g4w6ddl
                             'https://www.imdb.com/title/tt10676048/reviews/_ajax?&paginationKey=g4w6ddl
                             'https://www.imdb.com/title/tt10676048/reviews/_ajax?&paginationKey=g4w6ddl

      )

      allReviews5 <- lapply(marvels_urls, scrapeReviews5)

      movieReviews5 <- do.call(rbind, allReviews5)



#6TH MOVIE(Aquaman)
      session <- bow(url = 'https://www.imdb.com/title/tt9663764/reviews?ref_=tt_urv',
                     user_agent = "Educational")
      session
```

```
## <polite session> https://www.imdb.com/title/tt9663764/reviews?ref_=tt_urv
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
 scrapeReviews6 <- function(page_url) {
   movie6Reviews <- read_html(page_url)

   #title
   movie6Reviews %>%
     html_nodes('.title') %>%
     html_text() -> title

   #username
   movie6Reviews %>%
     html_nodes('.display-name-link') %>%
     html_text() -> username

   #content
   movie6Reviews %>%
     html_nodes('.content') %>%
     html_text() -> content

   #date
   movie6Reviews %>%
     html_nodes('.review-date') %>%
     html_text() -> date

   #stars
   movie6Reviews %>%
     html_nodes('.rating-other-user-rating') %>%
     html_text() -> stars

   #dataframe
   aquaman_df = data.frame(Title = title[1:25],
                           Username = username[1:25],
                           Content = content[1:25],
                           Date = date[1:25],
                           Stars = stars[1:25]
   )}

 #URLs
aquaman_urls<- c('https://www.imdb.com/title/tt9663764/reviews?ref_=tt_urv',
                   'https://www.imdb.com/title/tt9663764/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                   'https://www.imdb.com/title/tt9663764/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                   'https://www.imdb.com/title/tt9663764/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                   'https://www.imdb.com/title/tt9663764/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                   'https://www.imdb.com/title/tt9663764/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                   'https://www.imdb.com/title/tt9663764/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                   'https://www.imdb.com/title/tt9663764/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                   'https://www.imdb.com/title/tt9663764/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                   'https://www.imdb.com/title/tt9663764/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                   'https://www.imdb.com/title/tt9663764/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                   'https://www.imdb.com/title/tt9663764/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
 )

 allReviews6 <- lapply(aquaman_urls, scrapeReviews6)
```

```
      movieReviews6 <- do.call(rbind, allReviews6)



#7TH MOVIE(Barbie)
      session <- bow(url = 'https://www.imdb.com/title/tt1517268/reviews?ref_=tt_urv',
                     user_agent = "Educational")
      session
```

```
## <polite session> https://www.imdb.com/title/tt1517268/reviews?ref_=tt_urv
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
      scrapeReviews7 <- function(page_url) {
        movie7Reviews <- read_html(page_url)

        #title
        movie7Reviews %>%
          html_nodes('.title') %>%
          html_text() -> title

        #username
        movie7Reviews %>%
          html_nodes('.display-name-link') %>%
          html_text() -> username

        #content
        movie7Reviews %>%
          html_nodes('.content') %>%
          html_text() -> content

        #date
        movie7Reviews %>%
          html_nodes('.review-date') %>%
          html_text() -> date

        #stars
        movie7Reviews %>%
          html_nodes('.rating-other-user-rating') %>%
          html_text() -> stars

        #dataframe
        barbie_df = data.frame(Title = title[1:25],
                                Username = username[1:25],
                                Content = content[1:25],
                                Date = date[1:25],
                                Stars = stars[1:25]
        )}

      #URLs
      barbie_urls<- c('https://www.imdb.com/title/tt1517268/reviews?ref_=tt_urv',
                        'https://www.imdb.com/title/tt1517268/reviews/_ajax?&paginationKey=g4xojermtizc
                        'https://www.imdb.com/title/tt1517268/reviews/_ajax?&paginationKey=g4u6dermtizc
                        'https://www.imdb.com/title/tt1517268/reviews/_ajax?&paginationKey=g4v6jermtizc
                        'https://www.imdb.com/title/tt1517268/reviews/_ajax?&paginationKey=g4w6ddbsqyxd
                        'https://www.imdb.com/title/tt1517268/reviews/_ajax?&paginationKey=g4w6hcjsqyxd
                        'https://www.imdb.com/title/tt1517268/reviews/_ajax?&paginationKey=g4w6jdbsqyxd
                        'https://www.imdb.com/title/tt1517268/reviews/_ajax?&paginationKey=g4w6ncbsqyxd
                        'https://www.imdb.com/title/tt1517268/reviews/_ajax?&paginationKey=g4w6rbjsqyxd
                        'https://www.imdb.com/title/tt1517268/reviews/_ajax?&paginationKey=g4xohcbsqyxd
                        'https://www.imdb.com/title/tt1517268/reviews/_ajax?&paginationKey=g4xolbjsqyxd
                        'https://www.imdb.com/title/tt1517268/reviews/_ajax?&paginationKey=g4xorbjsqyxd
      )

      allReviews7 <- lapply(barbie_urls, scrapeReviews7)

      movieReviews7 <- do.call(rbind, allReviews7)



#8TH MOVIE(Five Feet Apart)
      session <- bow(url = 'https://www.imdb.com/title/tt1517268/reviews?ref_=tt_urv',
                     user_agent = "Educational")
      session
```

```
## <polite session> https://www.imdb.com/title/tt1517268/reviews?ref_=tt_urv
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
      scrapeReviews8 <- function(page_url) {
        movie8Reviews <- read_html(page_url)

        #title
        movie8Reviews %>%
          html_nodes('.title') %>%
          html_text() -> title

        #username
        movie8Reviews %>%
          html_nodes('.display-name-link') %>%
          html_text() -> username

        #content
        movie8Reviews %>%
          html_nodes('.content') %>%
          html_text() -> content

        #date
        movie8Reviews %>%
          html_nodes('.review-date') %>%
          html_text() -> date

        #stars
        movie8Reviews %>%
          html_nodes('.rating-other-user-rating') %>%
          html_text() -> stars

        #dataframe
        fivefeet_df = data.frame(Title = title[1:25],
                               Username = username[1:25],
                               Content = content[1:25],
                               Date = date[1:25],
                               Stars = stars[1:25]
        )}

      #URLs
      fivefeet_urls<- c('https://www.imdb.com/title/tt6472976/reviews?ref_=tt_urv',
                      'https://www.imdb.com/title/tt6472976/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6
                      'https://www.imdb.com/title/tt6472976/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6
                      'https://www.imdb.com/title/tt6472976/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6
                      'https://www.imdb.com/title/tt6472976/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6
                      'https://www.imdb.com/title/tt6472976/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6
                      'https://www.imdb.com/title/tt6472976/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6
                      'https://www.imdb.com/title/tt6472976/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6
                      'https://www.imdb.com/title/tt6472976/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6
                      'https://www.imdb.com/title/tt6472976/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6
                      'https://www.imdb.com/title/tt6472976/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6
                      'https://www.imdb.com/title/tt6472976/reviews/_ajax?&paginationKey=g4w6ddbmqyzdo6

      )

      allReviews8  <- lapply(fivefeet_urls, scrapeReviews8)

      movieReviews8 <- do.call(rbind, allReviews8)



#9TH MOVIE(Ready or Not)
      session <- bow(url = 'https://www.imdb.com/title/tt7798634/reviews?ref_=tt_urv',
                     user_agent = "Educational")
      session
```

```
## <polite session> https://www.imdb.com/title/tt7798634/reviews?ref_=tt_urv
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
      scrapeReviews9 <- function(page_url) {
        movie9Reviews <- read_html(page_url)

        #title
        movie9Reviews %>%
          html_nodes('.title') %>%
          html_text() -> title

        #username
        movie9Reviews %>%
          html_nodes('.display-name-link') %>%
          html_text() -> username

        #content
        movie9Reviews %>%
          html_nodes('.content') %>%
          html_text() -> content

        #date
        movie9Reviews %>%
          html_nodes('.review-date') %>%
          html_text() -> date

        #stars
        movie9Reviews %>%
          html_nodes('.rating-other-user-rating') %>%
          html_text() -> stars

        #dataframe
        readyOrNot_df = data.frame(Title = title[1:25],
                                 Username = username[1:25],
                                 Content = content[1:25],
                                 Date = date[1:25],
                                 Stars = stars[1:25]
        )}

      #URLs
      readyOrNot_urls<- c('https://www.imdb.com/title/tt7798634/reviews?ref_=tt_urv',
                        'https://www.imdb.com/title/tt7798634/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt7798634/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt7798634/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt7798634/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt7798634/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt7798634/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt7798634/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt7798634/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt7798634/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt7798634/reviews/_ajax?&paginationKey=g4w6ddbmqyzd
                        'https://www.imdb.com/title/tt7798634/reviews/_ajax?&paginationKey=g4w6ddbmqyzd

      )

      allReviews9  <- lapply( readyOrNot_urls, scrapeReviews9)

      movieReviews9 <- do.call(rbind, allReviews9)



#10TH MOVIE(To All the Boys I've Loved Before)
      session <- bow(url = 'https://www.imdb.com/title/tt3846674/reviews?ref_=tt_urv',
                     user_agent = "Educational")
      session
```

```
## <polite session> https://www.imdb.com/title/tt3846674/reviews?ref_=tt_urv
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```
scrapeReviews10 <- function(page_url) {
  movie10Reviews <- read_html(page_url)

  #title
  movie10Reviews %>%
    html_nodes('.title') %>%
    html_text() -> title

  #username
  movie10Reviews %>%
    html_nodes('.display-name-link') %>%
    html_text() -> username

  #content
  movie10Reviews %>%
    html_nodes('.content') %>%
    html_text() -> content

  #date
  movie10Reviews %>%
    html_nodes('.review-date') %>%
    html_text() -> date

  #stars
  movie10Reviews %>%
    html_nodes('.rating-other-user-rating') %>%
    html_text() -> stars

  #dataframe
  toAllBoys_df = data.frame(Title = title[1:25],
                             Username = username[1:25],
                             Content = content[1:25],
                             Date = date[1:25],
                             Stars = stars[1:25]
  )}

#URLs
toAllBoys_urls<- c('https://www.imdb.com/title/tt3846674/reviews?ref_=tt_urv',
                     'https://www.imdb.com/title/tt3846674/reviews/_ajax?&paginationKey=g4w6ddbmq
                     'https://www.imdb.com/title/tt3846674/reviews/_ajax?&paginationKey=g4w6ddbmq
                     'https://www.imdb.com/title/tt3846674/reviews/_ajax?&paginationKey=g4w6ddbmq
                     'https://www.imdb.com/title/tt3846674/reviews/_ajax?&paginationKey=g4w6ddbmq
                     'https://www.imdb.com/title/tt3846674/reviews/_ajax?&paginationKey=g4w6ddbmq
                     'https://www.imdb.com/title/tt3846674/reviews/_ajax?&paginationKey=g4w6ddbmq
                     'https://www.imdb.com/title/tt3846674/reviews/_ajax?&paginationKey=g4w6ddbmq
                     'https://www.imdb.com/title/tt3846674/reviews/_ajax?&paginationKey=g4w6ddbmq
                     'https://www.imdb.com/title/tt3846674/reviews/_ajax?&paginationKey=g4w6ddbmq
                     'https://www.imdb.com/title/tt3846674/reviews/_ajax?&paginationKey=g4w6ddbmq
                     'https://www.imdb.com/title/tt3846674/reviews/_ajax?&paginationKey=g4w6ddbmq
```

```
      )

      allReviews10  <- lapply(  toAllBoys_urls, scrapeReviews10)

      movieReviews10 <- do.call(rbind, allReviews10)


#RBind(I combined all movie reviews I scraped)
combined_df <- rbind(movieReviews1, movieReviews2,movieReviews3,movieReviews4,movieReviews5,movieReviews
```