

Question2B

Nabila Doctor

March 14, 2020

```
library(rvest)
library(tidyverse)
library(stringr)
```

Modified Function

```
boxoffice <- function() {
  movie_html = read_html("https://www.imdb.com/chart/boxoffice")
  movie_table = html_table(html_nodes(movie_html, "table")[[1]])

  movie_table$Weekend = as.numeric(gsub("[M]", "", movie_table$Weekend))
  movie_table$Gross = as.numeric(gsub("[M]", "", movie_table$Gross))

  boxoffice_data = tibble(Name = movie_table$Title,
                           'BoxOffice ($M)' = movie_table$Weekend,
                           Gross = movie_table$Gross,
                           Weeks = as.numeric(movie_table$Weeks),
                           RT = NA)

  boxoffice_data = boxoffice_data %>%
    mutate('PerWeek ($M)' = Gross / Weeks) %>%
    select(Name, 'BoxOffice ($M)', 'PerWeek ($M)', RT)

  # will add to this url
  rt_url = 'https://www.rottentomatoes.com/m/'

  # for each movie name in boxoffice_data, this loop will run
  for (i in 1:nrow(boxoffice_data)) {
    moviename = tolower(boxoffice_data$Name[i])
    moviename = gsub("[:.-]", "", moviename)

    # create a character list of the movie name
    if (grepl(" ", moviename)) {
      moviename = strsplit(moviename, '\\s+')[[1]] # split on spaces
    }

    if (length(moviename) > 1) {

      url_end = paste(moviename[1], moviename[2], sep = "_")
      j = 2
```

```

# Goal of while: adding each word to the url, one by one
# and stopping once we found a 'working' url
while(j<=length(moviename)){
  rt_movieurl = paste(rt_url,url_end,sep="")

  rt_html = try(read_html(rt_movieurl), silent = T) # testing the url

  if ("try-error" %in% class(rt_html)) { # if error

    if (j == length(moviename)) {
      url_end = paste(url_end,"2020",sep = "_")
      rt_movieurl = paste(rt_url,url_end,sep="")

      j = length(moviename)+1
    } else {
      url_end = paste(url_end,moviename[j+1],sep = "_")

      j = j+1
    }
  } else { # otherwise, no error
    j = length(moviename)+1

    # check date of movie is greater than or equal to year 2019
    tmp = readLines(rt_movieurl)

    tmp = grep("<div class=\"meta-value\">",t)[5]
    tp = trimws(t[tmp:tmp+1])
    tp = gsub("<.*?>", "", tp)
    tp = strsplit(gsub("<.*?>", "", tp), '\\s+')[[1]]

    if (!(as.numeric(tp[3]) >= 2019)){
      rt_movieurl = paste(rt_movieurl,"2020",sep="_")
    }
  }
}

} else { # For length of one
  rt_movieurl = paste(rt_url,moviename,sep="")

  rt_html = try(read_html(rt_movieurl), silent = T)

  if ("try-error" %in% class(rt_html)) {
    rt_movieurl = paste(rt_movieurl,"2020",sep = "_")
  }
}

rt_html = try(read_html(rt_movieurl), silent = T) # testing the url

```

```

    if ("try-error" %in% class(rt_html)) { # ratings for movies I couldn't ma
tch
      rt_rating = "NA"

    } else {
      rt_html = readLines(rt_movieurl)

      a = grep('<span class="mop-ratings-wrap__percentage">', rt_html)[1]

      if (is.na(a)) {
        rt_movieurl = paste(rt_movieurl, "2020", sep="_")

        rt_html = readLines(rt_movieurl)
        a = grep('<span class="mop-ratings-wrap__percentage">', rt_html)[1]
        rt_rating = trimws(rt_html[a:a+1])

      } else {
        rt_rating = trimws(rt_html[a:a+1])
      }
    }

    boxoffice_data$RT[i] = rt_rating
  }

  return(boxoffice_data)
}

```

Calling Function and Providing the Database

```
dat = boxoffice()
```

```

> dat
# A tibble: 10 x 4
  Name                                     `BoxOffice ($M)` `PerWeek ($M)` RT
  <chr>                                     <dbl>         <dbl> <chr>
1 Onward                                   39.1           39.1 87%
2 The Invisible Man                       15.1           26.4 91%
3 The Way Back                             8.2            8.2 83%
4 Sonic the Hedgehog                      7.7           35.1 64%
5 The Call of the wild                    6.8           19.1 61%
6 Emma.                                    4.8            2.23 85%
7 Bad Boys for Life                       3             25.2 77%
8 Birds of Prey: And the Fantabulous Emancipation of One Harley Quinn 2.1           16.5 78%
9 Impractical Jokers: The Movie            1.8            3.23 26%
10 My Hero Academia - Boku no hīrō akademia THE MOVIE - Heroes: Rising~ 1.7            6.45 NA
> |

```