

Class Activity 17

Your name here

2024-05-03

Group Activity 1

1. Go to the [the numbers webpage](https://www.the-numbers.com/movie/budgets/all) and extract the table on the front page.

```
session1 <- bow(url = "https://www.the-numbers.com/movie/budgets/all") %>% scrape() %>%
```

Error: <text>:4:0: unexpected end of input

2:

3:

^

2. Find out the number of pages that contain the movie table, while looking for the changes in the url in the address bar. How does the url changes when you go to the next page?

Answer:

3. Write a for loop to store all the data in multiple pages to a single data frame. Also, do the same using `purrr::map_df`.

```
library(tidyverse)
library(rvest)

new_urls <- "https://www.the-numbers.com/movie/budgets/all/"

# Create an empty list
df1 <- list()

# Generate a vector of indices
index <- seq(1, 6301, 100)
```

Group Activity 2

1. Go to [scrapethissite](https://www.scrapethissite.com/) and extract the table on the front page.

```
session1 <- read_html("https://www.scrapethissite.com/pages/forms/")
```

2. Find out the number of pages that contain the movie table, while looking for the changes in the url in the address bar. How does the url changes when you go to the next page?
3. Write a for loop to store all the data in multiple pages to a single data frame. Also, do the same using `purrr::map_df`.