Class Activity 17

Your name here

May 03 2024

Group Activity 1

1. Go to the the numbers webpage and extract the table on the front page.

```
session1 <-
```

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.

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

# creating two empty data-frames

table_new <-data.frame()
df2 <- data.frame()


idx <-  seq(__, ___, ___)

for (i in 1:length(idx)) {
  new_webpage <-
  table_new <- html_____(new_webpage)[[1]] %>%
  tibble::as_tibble(.name_repair = "unique")
  df2 <- rbind(df2, table_new)
}
```

```
# alternate using map
df3 <- data.frame()


# alternate using lapply
```

4. Display the data table using `knitr::kable(data, format = "html")`

5. Store the data in a `.csv` file with `write_csv(data_object, "/path/name.csv")`

Group Activity 2

```
mca <- data.frame(percent = c(23.5, 34.7, 19.8, 22.0, 27.6, 32.1, 20.2, 20.2, 26.0, 30.7, 22.1, 21.1),
                  portfolio = factor(rep(c("Commodities", "Stocks", "Real State", "Other Assets"), 3),
                                     levels=c("Commodities", "Stocks", "Real State", "Other Assets")),
                  year = rep(c("2020","2021","2022"), each=4))
glimpse(mca)
head(mca,6)
```

a. Make an interactive bar plot showiung the distribution of the various portfolio allocation for each year in a side-by-side format. Hint: use `fill = portfolio` as one of your aesthetics.

```
library(plotly)
mybar <- ggplot(mca, aes(x=    , y=     , fill=     )) +
  labs(title="Investment Portfolio") +
  geom_bar(stat = "identity", position = "dodge") +
  theme(legend.position = "bottom") +
  scale_fill_wsj()
ggplotly(mybar)
```