Stock

Rebecca Weiss

5/3/2022

```
#load all necessary libraries
library(tidyverse)
library(rvest)
```

Clear the workspace:

```
rm(list = ls())
```

Scrape data from yahoo finance for stocks (1 year)

Function to scrape 1 year of historical data from Yahoo Finance

```
scrapestock <- function(url){
  df <- url %>%
  # read the HTML from the webpage
  read_html() %>%
  # get data table from copied xpath for historical data
  html_nodes(xpath = ('//*[@id="Col1-1-HistoricalDataTable-Proxy"]/section')) %>%
  html_table()
}
```

Use function to get data from Apple, Tesla, Amazon, and Netflix

```
tsla <- scrapestock('https://finance.yahoo.com/quote/TSLA/history?p=TSLA')
aapl <- scrapestock('https://finance.yahoo.com/quote/AAPL/history?p=AAPL')
nflx <- scrapestock('https://finance.yahoo.com/quote/NFLX/history?p=NFLX')
amzn <- scrapestock('https://finance.yahoo.com/quote/AMZN/history?p=AMZN')
```

Make sure data looks correct

```
glimpse(amzn)
```

```
## List of 1
##  $ : tibble [101 x 7] (S3: tbl_df/tbl/data.frame)
##   ..$ Date       : chr [1:101] "May 03, 2022" "May 02, 2022" "Apr 29, 2022" "Apr 28, 2022" ...
##   ..$ Open       : chr [1:101] "2,481.07" "2,448.02" "2,596.98" "2,843.56" ...
##   ..$ High       : chr [1:101] "2,524.41" "2,493.36" "2,615.22" "2,918.75" ...
```

```
##   ..$ Low        : chr [1:101] "2,456.50" "2,367.50" "2,432.50" "2,806.00" ...
##   ..$ Close*     : chr [1:101] "2,485.07" "2,490.00" "2,485.63" "2,891.93" ...
##   ..$ Adj Close**: chr [1:101] "2,485.07" "2,490.00" "2,485.63" "2,891.93" ...
##   ..$ Volume     : chr [1:101] "3,952,000" "7,439,400" "13,616,500" "5,865,800" ...
```