

```
library(tidyverse)
library(rvest) #scrpe fform internet
```

Warning message in system("timedatectl", intern = TRUE):

"running command 'timedatectl' had status 1"

Warning message:

"Failed to locate timezone database"

— Attaching packages — tidyverse 1.3.1

```
✓ ggplot2 3.3.5    ✓ purrr  0.3.4
✓ tibble  3.1.5    ✓ dplyr  1.0.7
✓ tidyr   1.1.4    ✓ stringr 1.4.0
✓ readr   2.0.2    ✓ forcats 0.5.1
```

— Conflicts — tidyverse\_conflicts()

```
✗ dplyr::filter() masks stats::filter()
✗ purrr::flatten() masks jsonlite::flatten()
✗ dplyr::lag()     masks stats::lag()
```

Attaching package: 'rvest'

```
url <- "https://www.imdb.com/search/title/?groups=top_100&sort=user_rating,desc"
```

```
#read html
imdb <- read_html(url)
```

```
#movie title
title <- imdb %>%
  html_nodes("h3.lister-item-header") %>%
  html_text2() #จะลบอักขระอื่นๆทิ้ง
```

```
imdb %>%
```

```
html_node("div.ratings-bar") %>%  
html_text2() #จะลบอักขระอื่นๆทิ้ง
```

```
'9.3\nRate this\n1 2 3 4 5 6 7 8 9 10 9.3/10 X \n81 Metascore'
```

```
ratings<- imdb %>%  
  html_nodes("div.ratings-imdb-rating") %>%  
  html_text2() #จะลบอักขระอื่นๆทิ้ง
```

```
#number of vote  
num_vote<-imdb %>%  
  html_nodes("p.sort-num_votes-visible") %>%  
  html_text2() #จะลบอักขระอื่นๆทิ้ง
```

```
#build a dataframe  
df <- data.frame(  
  title = title,  
  rating = ratings,  
  num_vote = num_vote  
)
```

```
df
```

A data.frame: 50 × 3

title	rating	num_vote
<chr>	<chr>	<chr>
1. The Shawshank Redemption (1994)	9.3	Votes: 2,664,096   Gross: \$28.34M   Top 250: #1
2. The Godfather (1972)	9.2	Votes: 1,846,148   Gross: \$134.97M   Top 250: #2
3. The Dark Knight (2008)	9.0	Votes: 2,637,049   Gross: \$534.86M   Top 250: #3
4. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,836,721   Gross: \$377.85M   Top 250: #7
5. Schindler's List (1993)	9.0	Votes: 1,349,035   Gross: \$96.90M   Top 250: #6
6. The Godfather Part II (1974)	9.0	Votes: 1,264,515   Gross: \$57.30M   Top 250: #4
7. 12 Angry Men (1957)	9.0	Votes: 786,749   Gross: \$4.36M   Top 250: #5
8. Pulp Fiction (1994)	8.9	Votes: 2,039,147   Gross: \$107.93M   Top 250: #8
9. Inception (2010)	8.8	Votes: 2,336,794   Gross: \$292.58M   Top 250: #14
10. The Lord of the Rings: The Two Towers (2002)	8.8	Votes: 1,658,487   Gross: \$342.55M   Top 250: #13
11. Fight Club (1999)	8.8	Votes: 2,109,503   Gross: \$37.03M   Top 250: #12
12. The Lord of the Rings: The Fellowship of the Ring (2001)	8.8	Votes: 1,865,793   Gross: \$315.54M   Top 250: #9
13. Forrest Gump (1994)	8.8	Votes: 2,064,835   Gross: \$330.25M   Top 250: #11
14. Il buono, il brutto, il cattivo (1966)	8.8	Votes: 759,278   Gross: \$6.10M   Top 250: #10
15. The Matrix (1999)	8.7	Votes: 1,903,504   Gross: \$171.48M   Top 250: #16
16. Goodfellas (1990)	8.7	Votes: 1,154,595   Gross: \$46.84M   Top 250: #17
17. The Empire Strikes Back (1980)	8.7	Votes: 1,286,528   Gross: \$290.48M   Top 250: #15
18. One Flew Over the Cuckoo's Nest (1975)	8.7	Votes: 1,004,660   Gross: \$112.00M   Top 250: #18
19. Interstellar (2014)	8.6	Votes: 1,810,678   Gross: \$188.02M   Top 250: #26
20. Cidade de Deus (2002)	8.6	Votes: 754,987   Gross: \$7.56M   Top 250: #23
21. Sen to Chihiro no kamikakushi (2001)	8.6	Votes: 759,103   Gross: \$10.06M   Top 250: #31
22. Saving Private Ryan (1998)	8.6	Votes: 1,384,638   Gross: \$216.54M   Top 250: #24
23. The Green Mile (1999)	8.6	Votes: 1,294,551   Gross: \$136.80M   Top 250: #27
24. La vita è bella (1997)	8.6	Votes: 692,670   Gross: \$57.60M   Top 250: #25
25. Se7en (1995)	8.6	Votes: 1,642,118   Gross: \$100.13M   Top 250: #19
26. Terminator 2: Judgment Day (1991)	8.6	Votes: 1,094,643   Gross: \$204.84M   Top 250: #29
27. The Silence of the Lambs (1991)	8.6	Votes: 1,425,364   Gross: \$130.74M   Top 250: #22
28. Star Wars (1977)	8.6	Votes: 1,359,100   Gross: \$322.74M   Top 250: #28

29. Seppuku (1962)	8.6	Votes: 57,302   Top 250: #44
30. Shichinin no samurai (1954)	8.6	Votes: 345,722   Gross: \$0.27M   Top 250: #20
31. It's a Wonderful Life (1946)	8.6	Votes: 454,505   Top 250: #21
32. Gisaengchung (2019)	8.5	Votes: 792,717   Gross: \$53.37M   Top 250: #34
33. Whiplash (2014)	8.5	Votes: 852,711   Gross: \$13.09M   Top 250: #42
34. The Intouchables (2011)	8.5	Votes: 854,357   Gross: \$13.18M   Top 250: #45
35. The Prestige (2006)	8.5	Votes: 1,326,988   Gross: \$53.09M   Top 250: #41
36. The Departed (2006)	8.5	Votes: 1,319,164   Gross: \$132.38M   Top 250: #39
37. The Pianist (2002)	8.5	Votes: 828,410   Gross: \$32.57M   Top 250: #33
38. Gladiator (2000)	8.5	Votes: 1,492,934   Gross: \$187.71M   Top 250: #37
39. American History X (1998)	8.5	Votes: 1,118,985   Gross: \$6.72M   Top 250: #38
40. The Usual Suspects (1995)	8.5	Votes: 1,081,911   Gross: \$23.34M   Top 250: #40
41. Léon (1994)	8.5	Votes: 1,155,985   Gross: \$19.50M   Top 250: #35
42. The Lion King (1994)	8.5	Votes: 1,053,385   Gross: \$422.78M   Top 250: #36
43. Nuovo Cinema Paradiso (1988)	8.5	Votes: 261,077   Gross: \$11.99M   Top 250: #52
44. Hotaru no haka (1988)	8.5	Votes: 276,847   Top 250: #46
45. Back to the Future (1985)	8.5	Votes: 1,198,356   Gross: \$210.61M   Top 250: #30
46. Apocalypse Now (1979)	8.5	Votes: 665,930   Gross: \$83.47M   Top 250: #53
47. Alien (1979)	8.5	Votes: 879,416   Gross: \$78.90M   Top 250: #50
48. Once Upon a Time in the West (1968)	8.5	Votes: 329,283   Gross: \$5.32M   Top 250: #48
49. Psycho (1960)	8.5	Votes: 670,433   Gross: \$32.00M   Top 250: #32
50. Rear Window (1954)	8.5	Votes: 491,177   Gross: \$36.76M   Top 250: #49

## Mini project 02 - Specphone Phone Database

```
library(tidyverse)
library(rvest) #scrpe fform internet
```

```
url <- "https://specphone.com/Samsung-Galaxy-A04.html"
```

```
att<- url %>%
  read_html() %>%
  html_nodes("div.topic") %>%
  html_text2()
```

```
value <- url %>%
  read_html() %>%
  html_nodes("div.detail") %>%
  html_text2()
```

```
df <- data.frame(
  attribute = att,
  value = value
)
```

```
df
```

A data.frame: 31 × 2

attribute	value
<chr>	<chr>
วันเปิดตัว	ตุลาคม 2565
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	164.40 x 76.30 x 9.10 มม.
น้ำหนัก	192 กรัม
วัสดุ	Glass front, plastic back, plastic frame
SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)
Technology	HSPA 42.2/5.76 Mbps, LTE-A
2G	850/900/1800/1900
3G	850/900/1900/2100
4G	850/900/1900/2100/2600
5G	-
ความเร็ว	HSPA 42.2/5.76 Mbps, LTE-A
ประเภท	PLS LCD
ขนาดหน้าจอ	6.50 นิ้ว
ความละเอียด	720 x 1600 pixels
ระบบปฏิบัติการ	Android 12
ชิปประมวลผล	Spreadtrum Unisoc SC9863A 1.6 GHz
ชิปกราฟิก	PowerVR GE8322
หน่วยความจำ	3 GB
ความจุ	32 GB
Memory Card	microSD (1)
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.8, (wide), AF ตัวที่ 2: 2 MP, f/2.4, (depth)
ความละเอียดวิดีโอ	1080p@30fps
กล้องหน้า	ตัวที่ 1: 5 MP, f/2.2
Bluetooth	5.0, A2DP, LE
Wi-Fi	802.11 a/b/g/n/ac, dual-b
USB	Type-C
GPS	GLONASS, GALILEO, BDS
NFC	ไม่รองรับ
ความจุ	5,000 mAh
ประเภท	Non-removable Li-Po Batt

```
#all samsung smartphone
samsung_url <- read_html("https://specphone.com/brand/Samsung")
```

```
links <- samsung_url %>%
  html_nodes("li.mobile-brand-item a") %>% #งหาตัว a ที่อยู่ภายใน li
  html_attr("href") # เจอ a แล้วดึง attribute href
```

```
full_links <- paste0("https://specphone.com", links)
```

```
result <- data.frame()

for(link in full_links[1:5]){
  ss_topic <-link %>%
    read_html() %>%
    html_nodes("div.topic") %>%
    html_text2()
  ss_detail <-link %>%
    read_html() %>%
    html_nodes("div.detail") %>%
    html_text2()
  tmp <- data.frame(attributes = ss_topic,
                    value = ss_detail)
  result <- bind_rows(result, tmp)
  print("Progress...")
}

print(result)
```

```
[1] "Progress..."
[1] "Progress..."
[1] "Progress..."
[1] "Progress..."
[1] "Progress..."
```

```
attributes
1      วันเปิดตัว
2      วันวางจำหน่าย
3      ขนาด
4      น้ำหนัก
5      วัสดุ
6      SIM
7      Technology
8      2G
9      3G
10     4G
11     5G
```

12           ความเร็ว  
13           ประเภท

---

```
print(head(result),3)
```

	attributes	value
1	วันเปิดตัว	มิถุนายน 2565
2	วันวางจำหน่าย	ยังไม่วางจำหน่าย
3	ขนาด	165.40 x 76.90 x 8.40 มม.
4	น้ำหนัก	192 กรัม
5	วัสดุ	Glass front, plastic back, plastic frame
6	SIM	รองรับ 2 ซิมการ์ด (nano sim, nano sim)

```
#write csv  
write_csv(result, "result_ss_phone.csv")
```