

```
library(tidyverse)
library(rvest) #scrap data
```

```
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'
```

## Mini Project 01 = IMDB web scraping

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

```
print(url)
```

```
[1] "https://www.imdb.com/search/title/?groups=top_100&sort=user_rating,desc"
```

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

```
imdb
```

```
{html_document}
<html xmlns:og="http://ogp.me/ns#" xmlns:fb="http://www.facebook.com/2008/fbml"
[1] <head>\n<meta http-equiv="Content-Type" content="text/html; charset=UTF-8 .
[2] <body id="styleguide-v2" class="fixed">\n          <img height="1" width .
```

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

```
titles[1:10]
```

```
'1. The Shawshank Redemption (1994)' · '2. The Godfather (1972)' · '3. The Dark Knight (2008)' ·
'4. The Lord of the Rings: The Return of the King (2003)' · '5. Schindler\'s List (1993)' ·
'6. The Godfather Part II (1974)' · '7. 12 Angry Men (1957)' · '8. Pulp Fiction (1994)' · '9. Inception (2010)' ·
'10. The Lord of the Rings: The Two Towers (2002)'
```

```
#rating
```

```
ratings <- imdb %>%
  html_nodes("div.ratings-imdb-rating")%>%
  html_text2()%>%
  as.numeric()
```

```
ratings[1:10]
```

```
9.3 · 9.2 · 9 · 9 · 9 · 9 · 9 · 8.9 · 8.8 · 8.8
```

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

```
#build a dataset
df <- data.frame(
  title = titles,
  rating = ratings,
  num_votes = num_votes
)

head(df)
```

A data.frame: 6 × 3

	title	rating	num_votes
	<chr>	<dbl>	<chr>
1	1. The Shawshank Redemption (1994)	9.3	Votes: 2,674,338   Gross: \$28.34M   Top 250: #1
2	2. The Godfather (1972)	9.2	Votes: 1,853,567   Gross: \$134.97M   Top 250: #2
3	3. The Dark Knight (2008)	9.0	Votes: 2,647,331   Gross: \$534.86M   Top 250: #3
4	4. The Lord of the Rings: The Return of the King (2003)	9.0	Votes: 1,842,959   Gross: \$377.85M   Top 250: #7
5	5. Schindler's List (1993)	9.0	Votes: 1,353,779   Gross: \$96.90M   Top 250: #6
6	6. The Godfather Part II (1974)	9.0	Votes: 1,269,040   Gross: \$57.30M   Top 250: #4

## Mini Project 02 - Specphone Phone Database

```
url = read_html("https://specphone.com/Samsung-Galaxy-A04.html")
```

```
library(tidyverse)
library(rvest) #scrap data
```

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'

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

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

```
df <- data.frame(  
  topic = att,  
  detail = detail  
)
```

```
df
```

A data.frame: 31 × 2

topic	detail
<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
samsung <- read_html("https://specphone.com/brand/Samsung")
```

```
samsung %>%  
  html_nodes("div.name")%>%  
  html_text()
```

```
'Galaxy M13' · 'Galaxy A23' · 'Galaxy A13' · 'Galaxy M32 5G' · 'Galaxy A12 Nacho' · 'Galaxy Pocket Neo' ·  
'Galaxy Young' · 'Galaxy J1 Mini' · 'Galaxy A01 Core (1+16GB)' · 'Galaxy V PLUS' · 'Galaxy Young 2' · 'Galaxy M02' ·  
'Galaxy A11' · 'Galaxy J2 Pro (2018)' · 'Galaxy A12 (2021)' · 'Galaxy A21s (3+32GB)' · 'Galaxy J5' · 'Galaxy J4' ·  
'Galaxy Core 2 Duos' · 'Galaxy Ace Plus' · 'Galaxy A20' · 'Galaxy Chat' · 'Galaxy Gio' · 'Galaxy Tab A7 Lite LTE' ·  
'Galaxy Tab A (10.5)WIFI' · 'Galaxy Alpha' · 'Galaxy S3 Slim' · 'Galaxy S4 zoom' · 'Galaxy Xcover 2' ·  
'Galaxy Tab 8.9 3G 16GB' · 'Galaxy Tab A8 LTE 2021' · 'Galaxy A8 2018' · 'Galaxy Tab4 8.0 wifi' · 'Galaxy M33 5G' ·  
'Galaxy A50' · 'Galaxy E7' · 'Galaxy S6' · 'Galaxy S20 FE' · 'Galaxy Tab S4 WIFI' · 'Galaxy S7' · 'Galaxy Note 5 (Exynos)' ·  
'Galaxy TabPRO 12.2 LTE' · 'Galaxy S4 Active' · 'Galaxy Tab Active 3' · 'Galaxy Tab S3 9.7' · 'Galaxy S6 edge' ·  
'Galaxy Note 4 (Exynos)' · 'Galaxy Round' · 'Galaxy Note 20 Ultra 5G' · 'ATIV Q' · 'ATIV Smart PC PRO' ·  
'Galaxy S22 Ultra(12+128GB)' · 'Galaxy Z Flip 5G' · 'Galaxy Z Flip' · 'Galaxy Tab S8 Ultra 5G' ·  
'Galaxy S21 Ultra (16+512GB)' · 'Galaxy S10 Plus Ram 12GB' · 'Galaxy Z Fold 3' · 'Galaxy Z Fold4' ·  
'Galaxy Z Fold 2 5G'
```

```
links <- samsung %>%  
  html_nodes("li.mobile-brand-item a")%>%  
  html_attr("href")
```

```
links
```

```

'/Samsung-Galaxy-M13.html' · '/Samsung-Galaxy-A23.html' · '/Samsung-Galaxy-A13.html' ·
'/Samsung-Galaxy-M32-5G.html' · '/Samsung-Galaxy-A12-Nacho.html' · '/Samsung-Galaxy-Pocket-Neo.html' ·
'/Samsung-Galaxy-Young.html' · '/Samsung-Galaxy-J1-Mini.html' · '/Samsung-Galaxy-A01-Core-1-16GB.html' ·
'/Samsung-Galaxy-V-PLUS.html' · '/Samsung-Galaxy-Young-2.html' · '/Samsung-Galaxy-M02.html' ·
'/Samsung-Galaxy-A11.html' · '/Samsung-Galaxy-J2-Pro-2018.html' · '/Samsung-Galaxy-A12-2021.html' ·
'/Samsung-Galaxy-A21s-3-32GB.html' · '/Samsung-Galaxy-J5.html' · '/Samsung-Galaxy-J4.html' ·
'/Samsung-Galaxy-Core-2-Duos.html' · '/Samsung-Galaxy-Ace-Plus.html' · '/Samsung-Galaxy-A20.html' ·
'/Samsung-Galaxy-Chat.html' · '/Samsung-Galaxy-Gio.html' · '/Samsung-Galaxy-Tab-A7-Lite-LTE.html' ·
'/Samsung-Galaxy-Tab-A-10.5WIFI.html' · '/Samsung-Galaxy-Alpha.html' · '/Samsung-Galaxy-S3-Slim.html' ·
'/Samsung-Galaxy-S4-zoom.html' · '/Samsung-Galaxy-Xcover-2.html' · '/Samsung-Galaxy-Tab-8.9-3G-16GB.html' ·
'/Samsung-Galaxy-Tab-A8-LTE-2021.html' · '/Samsung-Galaxy-A8-2018.html' ·
'/Samsung-Galaxy-Tab4-8.0-wifi.html' · '/Samsung-Galaxy-M33-5G.html' · '/Samsung-Galaxy-A50.html' ·
'/Samsung-Galaxy-E7.html' · '/Samsung-Galaxy-S6.html' · '/Samsung-Galaxy-S20-FE.html' ·
'/Samsung-Galaxy-Tab-S4-WIFI.html' · '/Samsung-Galaxy-S7.html' · '/Samsung-Galaxy-Note-5-Exynos.html' ·
'/Samsung-Galaxy-TabPRO-12.2-LTE.html' · '/Samsung-Galaxy-S4-Active.html' ·
'/Samsung-Galaxy-Tab-Active-3.html' · '/Samsung-Galaxy-Tab-S3-9.7.html' · '/Samsung-Galaxy-S6-edge.html' ·
'/Samsung-Galaxy-Note-4-Exynos.html' · '/Samsung-Galaxy-Round.html' ·
'/Samsung-Galaxy-Note-20-Ultra-5G.html' · '/Samsung-ATIV-Q.html' · '/Samsung-ATIV-Smart-PC-PRO.html' ·
'/Samsung-Galaxy-S22-Ultra12-128GB.html' · '/Samsung-Galaxy-Z-Flip-5G.html' · '/Samsung-Galaxy-Z-Flip.html' ·
'/Samsung-Galaxy-Tab-S8-Ultra-5G.html' · '/Samsung-Galaxy-S21-Ultra-16-512GB.html' ·
'/Samsung-Galaxy-S10-Plus-Ram-12GB.html' · '/Samsung-Galaxy-Z-Fold-3.html' · '/Samsung-Galaxy-Z-Fold4.html' ·
'/Samsung-Galaxy-Z-Fold-2-5G.html'

```

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

```

result <- data.frame()

for (link in full_links[1:10]) {
  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)
}

```



```
result
```



A data.frame: 319 × 2

attributes	value
<chr>	<chr>
วันเปิดตัว	มิถุนายน 2565
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	165.40 x 76.90 x 8.40 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
5G	-
ความเร็ว	HSPA 42.2/5.76 Mbps, LTE-A
ประเภท	PLS LCD
ขนาดหน้าจอ	6.60 นิ้ว
ความละเอียด	1080 x 2408 pixels
ระบบปฏิบัติการ	Android 12
ชิปประมวลผล	Samsung Exynos 850 S5E3830 2 GHz
ชิปกราฟิก	Mali-G52 MP1
หน่วยความจำ	4 GB
ความจุ	64 GB
Memory Card	microSD (1)
กล้องหลัก	ตัวที่ 1: 50 MP, f/1.8, (wide), PDAF ตัวที่ 2: 5 MP, f/2.2, 123° (ultrawide) ตัวที่ 3: 2 MP, f/2.4, (depth)
ความละเอียดวิดีโอ	1080p@30fps
กล้องหน้า	ตัวที่ 1: 8 MP, f/2.2, (wide)
Bluetooth	5.0, A2DP, LE
Wi-Fi	802.11 a/b/g/n/ac, dual-b
USB	Type-C
GPS	A-GPS, GLONASS, GALILEO,
NFC	รองรับ
ความจุ	5,000 mAh
:	:
ขนาด	121.40 x 62.90 x 10.70 มม.
น้ำหนัก	123 กรัม
วัสดุ	Plastic
SIM	รองรับ 2 ซิมการ์ด (micro sim, micro sim)
Technology	EDGE, HSPA
2G	850/900/1800/1900
3G	900/2100
4G	-

5G	-
ความเร็ว	EDGE, HSPA

```
print(head(result))
```

attributes	สามารถให้ PC ใช้อินเทอร์เน็ตของสมาร์ทโฟนได้ด้วย Internet Teaming สามารถใช้งาน Mobile Hotspot		
ฟีเจอร์เพิ่มเติม	เปิดโดยแชร์อินเทอร์เน็ตบนสมาร์ทโฟนให้กับเครื่องอื่นๆเช่น ทีวี สามารถใช้งานเป็น USB Mass Storage ได้		
2 วันวางจำหน่าย	เพื่อโอนหรือย้ายข้อมูลบน PC ใช้โทรศัพท์ TouchWiz UX มาพร้อมกับ Android 4.4 รองรับ 2 ซิม		
ระบบปฏิบัติการ	Android 4.4	ขนาด	165.40 x 76.90 x 8.40 มม.
4 การ์	น้ำหนัก		192 กรัม
5 ปีประมวลผล	Intel Quad Core M3		
6 อุปกรณ์กราฟิก	VideoCore IV รองรับ 2 ซิมการ์ด (nano sim, nano sim)		
หน่วยความจำ	512 MB		
ความจุ	4 GB		
Memory			

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

กล้องหน้า	ตัวที่ 1: MP
Bluetooth	4
Wi-Fi	802.11b/g/n
USB	micro USB
GPS	Glomass, GPS
NFC	ไม่รองรับ
ความจุ	1,500 mAh
ประเภท	Li-ion