

```
library(tidyverse)
library(rvest) # scrape data from 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
```

```
✓ 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,des
```

```
imdb <- read_html(url)
```

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

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

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

```
df <- data.frame(
  title = titles,
  rating = ratings,
  num_vote = num_votes
)
head(df)
```

```
ERROR: Error in data.frame(title = titles, rating = ratings, num_vote = num_
```

## Mini Project 02 - Specphone

```
library(tidyverse)
library(rvest)
```

```
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
```

```
✓ 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 <- read_html("https://specphone.com/Apple-iPhone-12-Mini.html")
```

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

```
data.frame(attribute = att, value = value)
```

A data.frame: 34 × 2

attribute	value
<chr>	<chr>
วันเปิดตัว	กันยายน 2563
วันวางจำหน่าย	ยังไม่วางจำหน่าย
ขนาด	0.00 x 0.00 x 0.00 มม.
น้ำหนัก	ไม่รองรับ
วัสดุ	Front/back glass, stainless steel frame
SIM	รองรับ 1 ซิมการ์ด (nano sim)
Technology	5G ,LTE Advanced,HSPA+ ,EDGE
2G	850/900/1800/1900
3G	850/900/1900/2100
4G	700/800/850/900/1800/1900/2100
5G	-
ความเร็ว	5G ,LTE Advanced,HSPA+ ,EDGE
ประเภท	Super Retina HD
ขนาดหน้าจอ	5.40 นิ้ว
ความละเอียด	1792 x 828 pixels
ฟีเจอร์เพิ่มเติม	หน้าจอ HDR รองรับ Dolby Vision , HDR10 กล้องหน้าสามารถถ่ายเซลฟีแบบสโลว์โมชั่น กล้องหลัง 2 ตัวเป็นแบบ Wide , Ultra Wide กล้องรองรับ night mode,Portrait mode,Smart HDR กันน้ำกันฝุ่น IP68
ระบบปฏิบัติการ	iOS 14
ชิปประมวลผล	Apple A14 0 MHz
ชิปกราฟิก	Yes
หน่วยความจำ	4 GB
ความจุ	128/256 GB
Memory Card	ไม่รองรับ
กล้องหลัก	ตัวที่ 1: 12 MP ตัวที่ 2: 12 MP
ความละเอียดวิดีโอ	4K at 60fps
กล้องหน้า	ตัวที่ 1: 12 MP
Bluetooth	5.0, A2DP, LE
Wi-Fi	802.11 a/b/g/n/ac/ax
USB	Lightning
GPS	A-GPS, GLONASS, GALILEO,
NFC	รองรับ
ความจุ	3,110 mAh
ประเภท	Non-removable Li-Ion
Wireless Charging	รองรับ
Fast Charging	รองรับ (18W)

```
# all apple smartphone
apple_url <- read_html("https://specphone.com/brand/Apple")
```

```
# link to all apple smartphone
links <- apple_url %>%
  html_nodes("li.mobile-brand-item a") %>%
  html_attr("href")
```

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

```
result <- data.frame()

for (link in full_links[1:10]) {
  apple_topic <- link %>%
    read_html() %>%
    html_nodes("div.topic") %>%
    html_text2()

  apple_detail <- link %>%
    read_html() %>%
    html_nodes("div.detail") %>%
    html_text2()

  tmp <- data.frame(attribute = apple_topic, value = apple_detail)

  result <- bind_rows(result, tmp)
  print("Profress ...")
}

print(result)
```

```
[1] "Profress ..."
[1] "Profress ..."
[1] "Profress ..."
[1] "Profress ..."
[1] "Profress ..."
[1] "Profress ..."
[1] "Profress ..."
[1] "Profress ..."
[1] "Profress ..."
[1] "Profress ..."
      attribute
1      วันเปิดตัว
2      วันวางจำหน่าย
3      ขนาด
```

4            น้ำหนัก  
5            วัสดุ  
6            SIM  
7    Technology  
8            จอ

---

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

	attribute	value
1	วันเปิดตัว	พฤศจิกายน 2555
2	วันวางจำหน่าย	พฤษภาคม 2556, สินค้าจำหน่ายหมดแล้ว
3	ขนาด	200.00 x 134.70 x 7.20 มม.
4	น้ำหนัก	308 กรัม
5	วัสดุ	Aluminium, Plastic
6	SIM	รองรับ 1 ซิมการ์ด

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