Scraping HTML At Home Exercises Key

Carolina Alves de Lima Salge

2/1/2021

## Exercise 1

See below the code to complete the first exercise:

library(rvest)

## Loading required package: xml2

library(tidyverse)

## ── Attaching packages ─────────────────────────────────────── tidyverse 1.3.0 ──

## ✓ ggplot2 3.3.3 ✓ purrr 0.3.4  
## ✓ tibble 3.0.4 ✓ dplyr 1.0.2  
## ✓ tidyr 1.1.2 ✓ stringr 1.4.0  
## ✓ readr 1.4.0 ✓ forcats 0.5.0

## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## x dplyr::filter() masks stats::filter()  
## x readr::guess\_encoding() masks rvest::guess\_encoding()  
## x dplyr::lag() masks stats::lag()  
## x purrr::pluck() masks rvest::pluck()

ebay <- read\_html("https://www.ebay.com/sch/i.html?\_from=R40&\_nkw=mac+laptop&\_sacat=0&rt=nc&\_pgn=1")  
  
title <- ebay %>%   
 html\_nodes(".s-item\_\_title") %>%  
 html\_text()   
title

## [1] "MACBOOK PRO 15 LAPTOP | INTEL CORE 2.5GHz | 500GB | WARRANTY MAC"   
## [2] "Apple MacBook 13\" Laptop UPGRADED 8GB+1TB HD \*\* MAC OS X High Sierra \*\* "   
## [3] "APPLE MACBOOK PRO 15 INCH MAC LAPTOP / QUAD CORE I7 / 1TB SSD / 16GB / OS2020"   
## [4] "Apple Macbook Pro 13 Laptop / 16GB RAM 1TB SSD / Mac OS / 3 YEAR WARRANTY"   
## [5] "APPLE MACBOOK PRO 15\" MAC LAPTOP / QUAD CORE I7 / 1TB SSD / 16GB / OS2018"   
## [6] "Apple Macbook Pro 15 inch Laptop | 250GB SSD | UPGRADED | MAC OS2017 | WARRANTY "   
## [7] "Apple MacBook 13\" Laptop - UPDATED 2017 MAC OS +plus ANY UPGRADE CHOICE !"   
## [8] "New ListingApple MacBook Pro 13\" Laptop | Used | 500 GB | Mac OS | WARRANTY!"   
## [9] "Apple Macbook 13\" Laptop | UPGRADED 8GB RAM 1TB HD | MAC OS 2019 | WARRANTY"   
## [10] "Apple Macbook 13\" Laptop / UPGRADED 8GB RAM 1TB HDD / MAC OS 2017 + WARRANTY"   
## [11] "Apple MacBook 13\" Laptop | UPGRADED 8GB RAM+1TB HD | MAC OS 2017. WARRANTY"   
## [12] "New Listing13\" Apple MacBook Air Laptop | TURBO BOOST i5 | SSD | Mac OS | 3 YEAR WARRANTY"   
## [13] "Apple Macbook Pro 15 inch Laptop | i5 8GB 250GB SSD | MAC OS2017 | 2 YR Warranty"   
## [14] "Apple Macbook Pro 13\" MD101LL/A i5-3210m 2.5GHz 4GB RAM 500GB HD Mac OS 10.11"   
## [15] "MACBOOK PRO 15 LAPTOP / INTEL CORE 2.5GHz / 500GB / OS2015 / WARRANTY MAC"   
## [16] "Apple Macbook 13\" Laptop /UPGRADED 8GB RAM 1TB HD /MAC OS CATALINA/WARRANTY #777"   
## [17] "APPLE MACBOOK PRO 15\" RETINA MAC LAPTOP \* QUAD CORE I7 \* 1TB SSD \* 16GB \* OSx"   
## [18] "Apple Macbook 13\" Pre-Retina / UPGRADED 8GB RAM 1TB HDD / MAC OS 2017 / WARRANTY"   
## [19] "Apple MacBook White 13\" MC516LL/A 250GB HDD Intel 2.40GHz 4GB LATEST MAC OS 2017"   
## [20] "Macbook Pro 15\" Core 2 Duo 2.2 GHz (2007) 4 GB Ram 120 GB HDD Mac OS X"   
## [21] "APPLE MACBOOK PRO 15\" MAC LAPTOP / QUAD CORE I7 / 1TB SSD / 16GB / OS2020"   
## [22] "APPLE MACBOOK AIR 13 INCH LAPTOP - 1 YEAR WARRANTY - SSD - OS2015"   
## [23] "New Listing1993 Apple Macintosh PowerBook 165c WORKING Incl. Power Cable Bundle"   
## [24] "Apple Macbook 13\" Laptop / UPGRADED 8GB RAM 1TB HD / MAC-OS 2017 / WARRANTY"   
## [25] "MACBOOK PRO 15\" MAC LAPTOP | 500GB | OS2019 | APPLE LAPTOP | CORE i5 | SILVER"   
## [26] "Apple MacBook Pro13\"New 256GB SSD.Intel i5 /New 16GB RAM Mac OS High Sierra 2017"   
## [27] "Excellent Condition 16inch Mac Book Pro"   
## [28] "New ListingApple MacBook Pro 12,1 Intel i5-5257U @2.70GHz 8GB RAM 256GB Mac OS El Capitan"   
## [29] "Apple MacBook Pro13\" 500GB HDD/ Intel i5 /New 16 GB RAM. Mac OS Mojave"   
## [30] "Apple MacBook Air A1369 13.3 Laptop MC503LL/A 2010 Mac OS X Lion READ"   
## [31] "Apple MacBook 13.3\" Laptop - MA700LL/A (November, 2006) Mac OS Lion  160 GB HD "   
## [32] "APPLE MACINTOSH POWERBOOK G4 15\" \*\*READ\*\* "   
## [33] "Apple MacBook 13\" Laptop | UPGRADED 8GB RAM+1TB HD | MAC OS HIGH SIERRA 2017"   
## [34] "New ListingWorking Apple iBook G4 Laptop A1133 1 GB Ram, Mac OS 10.5"   
## [35] "Apple Mac Mini 2014 MGEM2LL/A i5-4260U 1.4GHz, 4GB RAM, 500GB HDD PC655362"   
## [36] "Apple MacBook Air 13.3\" Laptop i5 1.8GHz 8GB 128GB Mac OS (MQD32LL/A)"   
## [37] "Macbook 1TB RAM 4GB Mac Mojave + Linux + Windows10 Pro"   
## [38] "2020 Mac OS Catalina - READ ENTIRE DESCRIPTION - Upgraded to current model year "   
## [39] "Apple MacBook 13\" Laptop - UPDATED 2017 MAC OS + UPGRADE to SSD or MORE RAM ?"   
## [40] "Apple MacBook 13\" Laptop UPGRADED 8GB+500GB HD MAC OS 2017"   
## [41] "Apple Mac book 13 in. 2008 model"   
## [42] "Apple MacBook Pro 15 inch Touch Bar Core i7 16GB 512GB SSD Mac OSx2020 Big Sur"   
## [43] "APPLE MACBOOK PRO 15\" RETINA MAC LAPTOP | QUAD CORE I7 | 1TB SSD | 16GB | OS2020"   
## [44] "mac mini late 2014 big sur 8gig 1600 ddr3 ram 1tb sata hd"   
## [45] "APPLE MACBOOK PRO 15\" MAC LAPTOP PRE-RETINA / CORE I5 / 500GB / OS2020"   
## [46] "Apple 12\" PowerBook G4 Aluminum (Mac OS 10.4 and iLife '06 install discs)"   
## [47] "Apple MacBook Air 13 Inch Laptop 256GB SSD 8GB RAM Core i7 2.20GHz Mac OS 2020"   
## [48] "A1398 MacBook Mac Pro 512GB SSD15 Inch Retina Laptop 2.0 i7 8GB Ram Refurbished"   
## [49] "Apple Mac book Pro 13\" Intel Core 2 Duo 2.4GHz 4GB RAM 250GB Mid 2010 "   
## [50] "Mac Mini MC270LL/A Mid-2010 8GB Core 2 Duo 2.4GHz 320GB+WARRANTY!"   
## [51] "Apple Mac Mini Mid 2011 Intel Core i7 2.7GHz 8GB 750GB HDD"   
## [52] "Apple Macbook Air 13\" 256GB SSD Great Condition Mac OS Big Sur"   
## [53] "Apple Macbook Pro 15\" Laptop i7 / UPGRADED 16GB RAM + 2TB SSD / MAC OS 2018"   
## [54] "Apple 12\" MacBook Intel Core M3 256GB SSD 8GB RAM MAC OS Gray Webcam Student "   
## [55] "Apple Mac Mini 500GB A1347"   
## [56] "New Listing0885909223138"   
## [57] "New ListingApple MacBook Air A1466 Intel Core i7-4650U 1.7GHz CPU 8GB RAM 128GB SSD Mac OSX"  
## [58] "Apple MacBook Air 13\" early 2015 i7 8gb 500gb Mac laptop notebook OSX 10.14"   
## [59] "MacBook Pro, 13-inch, Mid-2010, Silver"   
## [60] "▓ APPLE MACBOOK PRO 15 IN RETINA MAC LAPTOP | QUAD CORE I7 | 512GB SSD OS-x ▓"   
## [61] "15\" APPLE MACBOOK PRO / INTEL / 8GB / 1TB / 3YR WARRANTY / Mac OSx-2015"   
## [62] "Apple MacBook Pro 13\" Mac Laptop | Used | 500 GB | OSX-2019 | WARRANTY!"   
## [63] "Apple Macbook Pro 13 Mac Laptop / i5 2.3GHZ 250GB SSD / Warranty + MAC OS 2017"

price <- ebay %>%  
 html\_nodes(".s-item\_\_detail--primary:nth-child(1) .s-item\_\_price") %>%  
 html\_text()  
price

## [1] "$458.50" "$194.95 to $294.95" "$689.00"   
## [4] "$635.00" "$649.00" "$435.00"   
## [7] "$137.00 to $335.00" "$399.00" "$249.00 to $359.00"  
## [10] "$269.10 to $359.10" "$244.00 to $344.00" "$479.00"   
## [13] "$419.99" "$239.00" "$479.00"   
## [16] "$349.00" "$979.00" "$269.99 to $359.99"  
## [19] "$299.97" "$56.00" "$649.00"   
## [22] "$329.00" "$49.95" "$269.10 to $359.10"  
## [25] "$479.00" "$499.97" "$1,699.00"   
## [28] "$399.98" "$399.95" "$189.99"   
## [31] "$70.00" "$99.99" "$275.00 to $375.00"  
## [34] "$79.00" "$212.00" "$555.39"   
## [37] "$348.00" "$1,063.00" "$137.00 to $335.00"  
## [40] "$339.00" "$110.00" "$1,339.00"   
## [43] "$999.00" "$320.00" "$579.00"   
## [46] "$76.00" "$549.00" "$550.00"   
## [49] "$170.00" "$199.99" "$275.00"   
## [52] "$395.00" "$845.00" "$609.99"   
## [55] "$274.99" "$175.00" "$339.99"   
## [58] "$499.99" "$195.00" "$742.00"   
## [61] "$615.62" "$499.00" "$399.99"

condition <- ebay %>%  
 html\_nodes(".SECONDARY\_INFO") %>%  
 html\_text()   
condition

## [1] "Pre-Owned" "Refurbished" "Refurbished" "Pre-Owned" "Refurbished"  
## [6] "Pre-Owned" "Refurbished" "Pre-Owned" "Refurbished" "Pre-Owned"   
## [11] "Refurbished" "Pre-Owned" "Pre-Owned" "Refurbished" "Pre-Owned"   
## [16] "Refurbished" "Pre-Owned" "Pre-Owned" "Refurbished" "Pre-Owned"   
## [21] "Refurbished" "Pre-Owned" "Pre-Owned" "Refurbished" "Pre-Owned"   
## [26] "Refurbished" "Pre-Owned" "Pre-Owned" "Refurbished" "Pre-Owned"   
## [31] "Refurbished" "Pre-Owned" "Refurbished" "Refurbished" "Pre-Owned"   
## [36] "Pre-Owned" "Refurbished" "Refurbished" "Refurbished" "Pre-Owned"   
## [41] "Pre-Owned" "Refurbished" "Refurbished" "Pre-Owned" "Refurbished"  
## [46] "Pre-Owned" "Refurbished" "Refurbished" "Pre-Owned" "Pre-Owned"   
## [51] "Pre-Owned" "Pre-Owned" "Pre-Owned" "Open Box" "Pre-Owned"   
## [56] "Pre-Owned" "Pre-Owned" "Pre-Owned" "Pre-Owned" "Pre-Owned"   
## [61] "Pre-Owned" "Pre-Owned" "Pre-Owned"

photo <- ebay %>%  
 html\_nodes(".s-item\_\_image-img") %>%  
 html\_attr("src")  
photo

## [1] "https://i.ebayimg.com/thumbs/images/g/oNkAAOSw2sBgEwZr/s-l225.jpg"  
## [2] "https://i.ebayimg.com/thumbs/images/g/Q7MAAOSwILpf~1kC/s-l225.jpg"  
## [3] "https://i.ebayimg.com/thumbs/images/g/jkYAAOSwpk9f9OAa/s-l225.jpg"  
## [4] "https://i.ebayimg.com/thumbs/images/g/r~IAAOSwloxf80dw/s-l225.jpg"  
## [5] "https://i.ebayimg.com/thumbs/images/g/EuQAAOSwtllf9N~-/s-l225.jpg"  
## [6] "https://i.ebayimg.com/thumbs/images/g/kPcAAOSwgVxf-Ny7/s-l225.jpg"  
## [7] "https://i.ebayimg.com/thumbs/images/g/K88AAOSw0FRfrrpp/s-l225.jpg"  
## [8] "https://i.ebayimg.com/thumbs/images/g/JWkAAOSwRodf6QYm/s-l225.jpg"  
## [9] "https://i.ebayimg.com/thumbs/images/g/GCsAAOSwiBhgAIfd/s-l225.jpg"  
## [10] "https://i.ebayimg.com/thumbs/images/g/sFoAAOSwZmJcnpjo/s-l225.jpg"  
## [11] "https://i.ebayimg.com/thumbs/images/g/ZpYAAOSwo1Ze7qSF/s-l225.jpg"  
## [12] "https://i.ebayimg.com/thumbs/images/g/pJIAAOSwVFNgAc~9/s-l225.jpg"  
## [13] "https://i.ebayimg.com/thumbs/images/g/YSYAAOSwv8NgAOZZ/s-l225.jpg"  
## [14] "https://i.ebayimg.com/thumbs/images/g/T4UAAOSwKptgEt7S/s-l225.jpg"  
## [15] "https://i.ebayimg.com/thumbs/images/g/xdsAAOSwFE1gFEaY/s-l225.jpg"  
## [16] "https://i.ebayimg.com/thumbs/images/g/hMMAAOSwGFVefTUZ/s-l225.jpg"  
## [17] "https://i.ebayimg.com/thumbs/images/g/QHAAAOSwD4xgBYlF/s-l225.jpg"  
## [18] "https://i.ebayimg.com/thumbs/images/g/WjsAAOSwvMhc0eMk/s-l225.jpg"  
## [19] "https://i.ebayimg.com/thumbs/images/g/q~sAAOSw3WxZ28Pk/s-l225.jpg"  
## [20] "https://i.ebayimg.com/thumbs/images/g/e7YAAOSwPlFgEMfE/s-l225.jpg"  
## [21] "https://i.ebayimg.com/thumbs/images/g/ilIAAOSwxp1f9OBE/s-l225.jpg"  
## [22] "https://i.ebayimg.com/thumbs/images/g/0a0AAOSw8GRgCIm3/s-l225.jpg"  
## [23] "https://i.ebayimg.com/thumbs/images/g/j8kAAOSwxeNgFypw/s-l225.jpg"  
## [24] "https://i.ebayimg.com/thumbs/images/g/ts4AAOSwgG5bsuEQ/s-l225.jpg"  
## [25] "https://i.ebayimg.com/thumbs/images/g/wJcAAOSwFRpf2TUk/s-l225.jpg"  
## [26] "https://i.ebayimg.com/thumbs/images/g/lLMAAOSwNqJbZIo3/s-l225.jpg"  
## [27] "https://i.ebayimg.com/thumbs/images/g/fFoAAOSw70FgE3Mb/s-l225.jpg"  
## [28] "https://i.ebayimg.com/thumbs/images/g/bh0AAOSw9j5gEr-3/s-l225.jpg"  
## [29] "https://i.ebayimg.com/thumbs/images/g/OYoAAOSwO29f~xDf/s-l225.jpg"  
## [30] "https://i.ebayimg.com/thumbs/images/g/ep8AAOSwDyZf-0xM/s-l225.jpg"  
## [31] "https://i.ebayimg.com/thumbs/images/g/e80AAOSwK7RgAga-/s-l225.jpg"  
## [32] "https://i.ebayimg.com/thumbs/images/g/7ggAAOSw7ZlgEZTw/s-l225.jpg"  
## [33] "https://i.ebayimg.com/thumbs/images/g/V88AAOSw3iRe7WDB/s-l225.jpg"  
## [34] "https://i.ebayimg.com/thumbs/images/g/qRAAAOSw2x5fxCD1/s-l225.jpg"  
## [35] "https://i.ebayimg.com/thumbs/images/g/aD8AAOSwLW5gC0p-/s-l225.jpg"  
## [36] "https://i.ebayimg.com/thumbs/images/g/J9cAAOSw6qdf3U-c/s-l225.jpg"  
## [37] "https://i.ebayimg.com/thumbs/images/g/mOUAAOSwb~xd7vgP/s-l225.jpg"  
## [38] "https://i.ebayimg.com/thumbs/images/g/1YMAAOSwyN5f5OrH/s-l225.jpg"  
## [39] "https://i.ebayimg.com/thumbs/images/g/K88AAOSw0FRfrrpp/s-l225.jpg"  
## [40] "https://i.ebayimg.com/thumbs/images/g/y0IAAOSwevReznwT/s-l225.jpg"  
## [41] "https://i.ebayimg.com/thumbs/images/g/d0kAAOSwsQxgBeDO/s-l225.jpg"  
## [42] "https://i.ebayimg.com/thumbs/images/g/ANsAAOSwgQNf-BTT/s-l225.jpg"  
## [43] "https://i.ebayimg.com/thumbs/images/g/eQYAAOSwdLpf4h4h/s-l225.jpg"  
## [44] "https://i.ebayimg.com/thumbs/images/g/AdwAAOSw~FZf~pt1/s-l225.jpg"  
## [45] "https://i.ebayimg.com/thumbs/images/g/7MoAAOSwf-dfMxOt/s-l225.jpg"  
## [46] "https://i.ebayimg.com/thumbs/images/g/jTMAAOSwzkVgEeL5/s-l225.jpg"  
## [47] "https://i.ebayimg.com/thumbs/images/g/qscAAOSwK05f~Bb0/s-l225.jpg"  
## [48] "https://i.ebayimg.com/thumbs/images/g/qdMAAOSwOtNgCkDf/s-l225.jpg"  
## [49] "https://i.ebayimg.com/thumbs/images/g/7bAAAOSwP9tgBdyx/s-l225.jpg"  
## [50] "https://i.ebayimg.com/thumbs/images/g/T1UAAOSw6ZZf8pkX/s-l225.jpg"  
## [51] "https://i.ebayimg.com/thumbs/images/g/230AAOSwUKRfTQoJ/s-l225.jpg"  
## [52] "https://i.ebayimg.com/thumbs/images/g/vXwAAOSw0jJfzVLi/s-l225.jpg"  
## [53] "https://i.ebayimg.com/thumbs/images/g/WS4AAOSwk0xfxdel/s-l225.jpg"  
## [54] "https://i.ebayimg.com/thumbs/images/g/E4QAAOSwdpRf82qH/s-l225.jpg"  
## [55] "https://i.ebayimg.com/thumbs/images/g/4rMAAOSwOl1ddCLP/s-l225.jpg"  
## [56] "https://i.ebayimg.com/thumbs/images/g/cWwAAOSwFgxgBLoW/s-l225.jpg"  
## [57] "https://i.ebayimg.com/thumbs/images/g/eb4AAOSwc-df468p/s-l225.jpg"  
## [58] "https://i.ebayimg.com/thumbs/images/g/8msAAOSwVu1f8UW7/s-l225.jpg"  
## [59] "https://i.ebayimg.com/thumbs/images/g/MA4AAOSwB~Bfa7Oi/s-l225.jpg"  
## [60] "https://i.ebayimg.com/thumbs/images/g/mF0AAOSwv-1ebPHg/s-l225.jpg"  
## [61] "https://i.ebayimg.com/thumbs/images/g/FdMAAOSwI2NdI-6i/s-l225.jpg"  
## [62] "https://i.ebayimg.com/thumbs/images/g/JWkAAOSwRodf6QYm/s-l225.jpg"  
## [63] "https://i.ebayimg.com/thumbs/images/g/Pu0AAOSwDPhc-f4K/s-l225.jpg"

ebayDF <- tibble(title = title,   
 price = price,  
 condition = condition,  
 photo = photo)  
  
write\_csv(ebayDF, "ebayDFp1.csv")

## Exercise 2

This exercise was more challenging since you needed to loop through all pages.

ebayMP <- sprintf("https://www.ebay.com/sch/i.html?\_from=R40&\_nkw=mac+laptop&\_sacat=0&rt=nc&\_pgn=%d", 1:7) # wrapper allowing you to combine text with variable values  
  
# Loop through the links using the map function then  
tbl2 <- map(ebayMP, ~ {  
   
 # Read link  
 doc <- read\_html(.x)  
   
 # Get the item title   
 title <-  
 doc %>%  
 html\_nodes(".s-item\_\_title") %>%  
 html\_text()  
   
 # Get price  
 price <-  
 doc %>%  
 html\_nodes(".s-item\_\_detail--primary:nth-child(1) .s-item\_\_price") %>%  
 html\_text()   
   
 # Get condition  
 condition <-   
 doc %>%  
 html\_nodes(".SECONDARY\_INFO") %>%  
 html\_text()   
   
 # Get img link  
 photo\_link <-  
 doc %>%  
 html\_nodes(".s-item\_\_image-img") %>%  
 html\_attr("src")  
   
 # Create a tibble to store the data  
 tibble(title = title,   
 price = price,  
 condition = condition,  
 photo\_link = photo\_link)  
   
}  
) %>%  
   
 # Merge the data above with the page id  
 bind\_rows(.id = 'page\_no')  
  
  
write\_csv(tbl2, "ebayDFp17.csv")