RWorksheet#5

Noblezada_Camasa_Cabia

2024-11-06

IMDB

```
library(rvest)
library(httr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(polite)
library(kableExtra)
```

```
##
## Attaching package: 'kableExtra'
```

```
## The following object is masked from 'package:dplyr':
##
##     group_rows
```

```
library(stringr)


url <- 'https://www.imdb.com/chart/toptv'

page <- read_html(url)

session <- bow(url, user_agent = "Educational")

session
```

```
## <polite session> https://www.imdb.com/chart/toptv
##     User-agent: Educational
##     robots.txt: 35 rules are defined for 3 bots
##   Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
name <- page %>% html_nodes(".ipc-title__text") %>% html_text()
name1 <- name[!grepl("Top 250 TV Shows|IMDb Charts|Recently viewed|More to explore", name, ignore.case =
name1
```

```
##  [1] "1. Breaking Bad"
##  [2] "2. Planet Earth II"
##  [3] "3. Planet Earth"
##  [4] "4. Band of Brothers"
##  [5] "5. Chernobyl"
##  [6] "6. The Wire"
##  [7] "7. Avatar: The Last Airbender"
##  [8] "8. Blue Planet II"
##  [9] "9. The Sopranos"
## [10] "10. Cosmos: A Spacetime Odyssey"
## [11] "11. Cosmos"
## [12] "12. Our Planet"
## [13] "13. Game of Thrones"
## [14] "14. Bluey"
## [15] "15. The World at War"
## [16] "16. Fullmetal Alchemist: Brotherhood"
## [17] "17. Rick and Morty"
## [18] "18. Life"
## [19] "19. The Last Dance"
## [20] "20. The Twilight Zone"
## [21] "21. The Vietnam War"
## [22] "22. Sherlock"
## [23] "23. Attack on Titan"
## [24] "24. Batman: The Animated Series"
## [25] "25. Arcane"
```

```r
rank <- str_extract(name1, "^\\d+\\.")
rank
```

```
##  [1] "1."  "2."  "3."  "4."  "5."  "6."  "7."  "8."  "9."  "10." "11." "12."
## [13] "13." "14." "15." "16." "17." "18." "19." "20." "21." "22." "23." "24."
## [25] "25."
```

```r
title <- str_replace(name, "^\\d+\\.", "")
title
```

```
##  [1] "IMDb Charts"                    "Top 250 TV Shows"
##  [3] " Breaking Bad"                  " Planet Earth II"
##  [5] " Planet Earth"                  " Band of Brothers"
##  [7] " Chernobyl"                     " The Wire"
##  [9] " Avatar: The Last Airbender"    " Blue Planet II"
## [11] " The Sopranos"                  " Cosmos: A Spacetime Odyssey"
## [13] " Cosmos"                        " Our Planet"
## [15] " Game of Thrones"               " Bluey"
## [17] " The World at War"              " Fullmetal Alchemist: Brotherhood"
## [19] " Rick and Morty"                " Life"
## [21] " The Last Dance"                " The Twilight Zone"
## [23] " The Vietnam War"               " Sherlock"
## [25] " Attack on Titan"               " Batman: The Animated Series"
## [27] " Arcane"                        "Recently viewed"
```

```r
year = page %>% html_nodes(".sc-5bc66c50-6.00dsw.cli-title-metadata-item") %>% html_text()
years <- str_extract_all(year, "\\b\\d{4}(?:-\\d{4})?\\b") %>% unlist()
years
```

```
## NULL
```

```r
ratings = page %>% html_nodes(".ipc-rating-star--rating") %>% html_text()
ratings
```

```
##  [1] "9.5" "9.5" "9.4" "9.4" "9.3" "9.3" "9.3" "9.3" "9.2" "9.2" "9.3" "9.2"
## [13] "9.2" "9.3" "9.2" "9.1" "9.1" "9.1" "9.0" "9.0" "9.1" "9.1" "9.1" "9.0"
## [25] "9.0"
```

```r
episode <- page %>% html_nodes(".sc-5bc66c50-6.00dsw.cli-title-metadata-item") %>%
html_text()
episodes <- str_extract_all(episode, "\\b\\d+ eps\\b") %>% unlist()
episodes
```

```
## NULL
```

```r
votes = page %>% html_nodes(".ipc-rating-star--voteCount") %>% html_text()
votes
```

```
##  [1] " (2.2M)" " (162K)" " (224K)" " (546K)" " (909K)" " (391K)" " (390K)"
##  [8] " (49K)"  " (500K)" " (131K)" " (46K)"  " (54K)"  " (2.4M)" " (34K)"
## [15] " (31K)"  " (209K)" " (628K)" " (44K)"  " (160K)" " (97K)"  " (29K)"
## [22] " (1M)"   " (564K)" " (123K)" " (324K)"
```

```r
urls <- c("https://www.imdb.com/title/tt0903747/?ref_=chttvtp_i_1",
          "https://www.imdb.com/title/tt5491994/?ref_=chttvtp_i_2",
          "https://www.imdb.com/title/tt0795176/?ref_=chttvtp_i_3",
          "https://www.imdb.com/title/tt0185906/?ref_=chttvtp_i_4",
          "https://www.imdb.com/title/tt7366338/?ref_=chttvtp_i_5",
          "https://www.imdb.com/title/tt0306414/?ref_=chttvtp_i_6",
          "https://www.imdb.com/title/tt0417299/?ref_=chttvtp_i_7",
          "https://www.imdb.com/title/tt6769208/?ref_=chttvtp_i_8",
          "https://www.imdb.com/title/tt0141842/?ref_=chttvtp_i_9",
          "https://www.imdb.com/title/tt2395695/?ref_=chttvtp_i_10",
          "https://www.imdb.com/title/tt0081846/?ref_=chttvtp_i_11",
          "https://www.imdb.com/title/tt9253866/?ref_=chttvtp_i_12",
          "https://www.imdb.com/title/tt0944947/?ref_=chttvtp_i_13",
          "https://www.imdb.com/title/tt7678620/?ref_=chttvtp_i_14",
          "https://www.imdb.com/title/tt0071075/?ref_=chttvtp_i_15",
          "https://www.imdb.com/title/tt1355642/?ref_=chttvtp_i_16",
          "https://www.imdb.com/title/tt2861424/?ref_=chttvtp_i_17",
          "https://www.imdb.com/title/tt1533395/?ref_=chttvtp_i_18",
          "https://www.imdb.com/title/tt8420184/?ref_=chttvtp_i_19",
          "https://www.imdb.com/title/tt0052520/?ref_=chttvtp_i_20",
          "https://www.imdb.com/title/tt1877514/?ref_=chttvtp_i_21",
          "https://www.imdb.com/title/tt1475582/?ref_=chttvtp_i_22",
          "https://www.imdb.com/title/tt2560140/?ref_=chttvtp_i_23",
          "https://www.imdb.com/title/tt0103359/?ref_=chttvtp_i_24",
          "https://www.imdb.com/title/tt0386676/?ref_=chttvtp_i_25")

userReviews <- vector("numeric", length(urls))
critiqueReviews <- vector("numeric", length(urls))
```

```r
popularity <- vector("numeric", length(urls))

for (i in seq_along(urls)) {

  session <- bow(urls[i], user_agent = "Educational")

  webpage <- scrape(session)

  popularityText <- webpage %>% html_nodes(".sc-39d285cf-1.dxqvqi") %>% html_text()
  popularity[i] <- as.numeric(gsub(",", "", popularityText[1]))

  reviews <- webpage %>% html_nodes(".score") %>% html_text()


  if (length(reviews) >= 2) {

    userReviews[i] <- ifelse(grepl("K", reviews[1]),
                             as.numeric(gsub("K", "", reviews[1])) * 1000,
                             as.numeric(reviews[1]))
    critiqueReviews[i] <- as.numeric(reviews[2])
  } else {
    userReviews[i] <- NA
    critiqueReviews[i] <- NA
  }
}
```

```r
userReviews
```

```
##  [1] 5100  158  111 1000 3500  787 1000   53  966  205   80  245 5900  368  126
## [16]  467  910   12  542  213  175 1000 2300  219 1700
```

```r
critiqueReviews
```

```
##  [1] 175    6   10   34   88   77   57    9   93   12    8   15  368    4    5   16   94    9   28
## [20]  85   13  121   64   25   76
```

```r
popularity
```

```
##  [1]   20  999 1778  153  146  104  357 4122   31 1571 3645 2401   16  373 2532
## [16]  474  125 3057 1403  337 1739  160   50  463   51
```

```r
max_length <- max(length(rank), length(title), length(year), length(ratings), length(episodes), length(v

rank <- c(rank, rep(NA, max_length - length(rank)))

title <- c(title, rep(NA, max_length - length(title)))

years <- c(years, rep(NA, max_length - length(years)))

ratings <- c(ratings, rep(NA, max_length - length(ratings)))

episodes <- c(episodes, rep(NA, max_length - length(episodes)))

votes <- c(votes, rep(NA, max_length - length(votes)))

userReviews <- c(userReviews, rep(NA, max_length - length(userReviews)))
```

```r
critiqueReviews <- c(critiqueReviews, rep(NA, max_length - length(critiqueReviews)))

popularity <- c(popularity, rep(NA, max_length - length(popularity)))
max_length
```

```
## [1] 28
```

```r
movies = data.frame(rank, title, years, ratings, episodes, votes, userReviews, critiqueReviews, populari

print(head(movies, 20))
```

```
##    rank                              title years ratings episodes   votes
## 1    1.                        IMDb Charts    NA     9.5       NA  (2.2M)
## 2    2.                     Top 250 TV Shows    NA     9.5       NA  (162K)
## 3    3.                       Breaking Bad    NA     9.4       NA  (224K)
## 4    4.                     Planet Earth II    NA     9.4       NA  (546K)
## 5    5.                       Planet Earth    NA     9.3       NA  (909K)
## 6    6.                     Band of Brothers    NA     9.3       NA  (391K)
## 7    7.                           Chernobyl    NA     9.3       NA  (390K)
## 8    8.                           The Wire    NA     9.3       NA   (49K)
## 9    9.          Avatar: The Last Airbender    NA     9.2       NA  (500K)
## 10  10.                     Blue Planet II    NA     9.2       NA  (131K)
## 11  11.                       The Sopranos    NA     9.3       NA   (46K)
## 12  12.       Cosmos: A Spacetime Odyssey    NA     9.2       NA   (54K)
## 13  13.                           Cosmos    NA     9.2       NA  (2.4M)
## 14  14.                         Our Planet    NA     9.3       NA   (34K)
## 15  15.                     Game of Thrones    NA     9.2       NA   (31K)
## 16  16.                             Bluey    NA     9.1       NA  (209K)
## 17  17.                   The World at War    NA     9.1       NA  (628K)
## 18  18.  Fullmetal Alchemist: Brotherhood    NA     9.1       NA   (44K)
## 19  19.                     Rick and Morty    NA     9.0       NA  (160K)
## 20  20.                               Life    NA     9.0       NA   (97K)
##    userReviews critiqueReviews popularity
## 1         5100             175         20
## 2          158               6        999
## 3          111              10       1778
## 4         1000              34        153
## 5         3500              88        146
## 6          787              77        104
## 7         1000              57        357
## 8           53               9       4122
## 9          966              93         31
## 10         205              12       1571
## 11          80               8       3645
## 12         245              15       2401
## 13        5900             368         16
## 14         368               4        373
## 15         126               5       2532
## 16         467              16        474
## 17         910              94        125
## 18          12               9       3057
## 19         542              28       1403
## 20         213              85        337
```

rank	title	years	ratings	episodes	votes	userReviews	critiqueReviews	popularity
1.	IMDb Charts	NA	9.5	NA	(2.2M)	5100	175	20
2.	Top 250 TV Shows	NA	9.5	NA	(162K)	158	6	999
3.	Breaking Bad	NA	9.4	NA	(224K)	111	10	1778
4.	Planet Earth II	NA	9.4	NA	(546K)	1000	34	153
5.	Planet Earth	NA	9.3	NA	(909K)	3500	88	146
6.	Band of Brothers	NA	9.3	NA	(391K)	787	77	104
7.	Chernobyl	NA	9.3	NA	(390K)	1000	57	357
8.	The Wire	NA	9.3	NA	(49K)	53	9	4122
9.	Avatar: The Last Airbender	NA	9.2	NA	(500K)	966	93	31
10.	Blue Planet II	NA	9.2	NA	(131K)	205	12	1571
11.	The Sopranos	NA	9.3	NA	(46K)	80	8	3645
12.	Cosmos: A Spacetime Odyssey	NA	9.2	NA	(54K)	245	15	2401
13.	Cosmos	NA	9.2	NA	(2.4M)	5900	368	16
14.	Our Planet	NA	9.3	NA	(34K)	368	4	373
15.	Game of Thrones	NA	9.2	NA	(31K)	126	5	2532
16.	Bluey	NA	9.1	NA	(209K)	467	16	474
17.	The World at War	NA	9.1	NA	(628K)	910	94	125
18.	Fullmetal Alchemist: Brotherhood	NA	9.1	NA	(44K)	12	9	3057
19.	Rick and Morty	NA	9.0	NA	(160K)	542	28	1403
20.	Life	NA	9.0	NA	(97K)	213	85	337
21.	The Last Dance	NA	9.1	NA	(29K)	175	13	1739
22.	The Twilight Zone	NA	9.1	NA	(1M)	1000	121	160
23.	The Vietnam War	NA	9.1	NA	(564K)	2300	64	50
24.	Sherlock	NA	9.0	NA	(123K)	219	25	463
25.	Attack on Titan	NA	9.0	NA	(324K)	1700	76	51
NA	Batman: The Animated Series	NA	NA	NA	NA	NA	NA	NA
NA	Arcane	NA	NA	NA	NA	NA	NA	NA
NA	Recently viewed	NA	NA	NA	NA	NA	NA	NA

```r
movies %>%
  kable("latex", booktabs = TRUE) %>%
  kable_styling(latex_options = "scale_down")
```

```r
reviewLinks <- c("https://www.imdb.com/title/tt0944947/reviews",
        "https://www.imdb.com/title/tt0903747/reviews",
        "https://www.imdb.com/title/tt4574334/reviews",
        "https://www.imdb.com/title/tt0108778/reviews",
        "https://www.imdb.com/title/tt1520211/reviews")

Name <- vector("list", length(reviewLinks))

DateOfTheReview <- vector("list", length(reviewLinks))

userRating <- vector("list", length(reviewLinks))

titleOfTheReview <- vector("list", length(reviewLinks))

helpful <- vector("list", length(reviewLinks))

notHelpful <- vector("list", length(reviewLinks))

textReview <- vector("list", length(reviewLinks))

reviewDf <- vector("list", length(reviewLinks))
```

```r
for (i in seq_along(reviewLinks)) {
  session <- bow(reviewLinks[i], user_agent = "Educational")

  webpage <- scrape(session)

  names <- webpage %>% html_nodes(".ipc-link.ipc-link--base") %>% html_text(trim = TRUE) %>% head(40)
  Name[[i]] <- names[!grepl("Permalink", names, ignore.case = TRUE)]

  DateOfTheReview[[i]] <- webpage %>% html_nodes(".ipc-inline-list__item.review-date") %>% html_text(tri

  userRating[[i]] <- webpage %>% html_nodes(".ipc-rating-star--rating") %>% html_text(trim = TRUE) %>% 

  titleOf <- webpage %>% html_nodes(".ipc-title__text") %>% html_text(trim = TRUE) %>% head(21)

  titleOfTheReview[[i]] <- titleOf[!grepl("User reviews|More from this title|More to explore|Recently v

  textReview[[i]] <- webpage %>% html_nodes(".ipc-html-content-inner-div") %>% html_text(trim = TRUE) %
}
```

```r
gameofthrones <- data.frame(Name[[1]],
    Date = DateOfTheReview[[1]],
    Rating = userRating[[1]],
    Title = titleOfTheReview[[1]],
    stringsAsFactors = FALSE
  )
```

```r
print(head(gameofthrones, 20))
```

```
##                 Name..1..         Date Rating
## 1             danielkpkp May 11, 2020      9
## 2     samxxx-671-826221 May 24, 2019      8
## 3              slowcando May 20, 2019      8
## 4               SaifOVGU  Apr 8, 2020     10
## 5      jacobnoble-02524  May 9, 2019      9
## 6         heavenacceptme Aug 22, 2022     10
## 7         Supermanfan-13  Dec 9, 2023     10
## 8         Dan_W_Reviews  Feb 5, 2023      9
## 9              tweaknhoe May 25, 2019      8
## 10   TheLittleSongbird  Nov 8, 2017     10
## 11              kunkell  Nov 1, 2019      6
## 12     akhil-marsonya27 May 20, 2019      1
## 13              Rob1331 Nov 10, 2023     10
## 14     alshamari-marwa May 18, 2020      9
## 15         tobiascramon  Nov 8, 2022      9
## 16 adamheritage-15333  May 5, 2019      9
## 17           LASTRONOME  Jun 2, 2020     10
## 18     SpitOnAStranger  May 9, 2019      8
## 19         psypeterson Apr 16, 2011     10
## 20             el-absy May 19, 2019      7
##
##                                                          Title
## 1                    It could have been the best TV series ever made...
## 2                A perfect example of: Falling in Love with the Wrong Guy
## 3            Seasons 1-6: outstanding. 7: daft but good. 8: disappointing
## 4                    Can you just make the remake the season finale?
```

Name..1..	Date	Rating	Title
danielkpkp	May 11, 2020	9	It could have been the best TV series ever made...
samxxx-671-826221	May 24, 2019	8	A perfect example of: Falling in Love with the Wrong Guy
slowcando	May 20, 2019	8	Seasons 1-6: outstanding. 7: daft but good. 8: disappointing
SaifOVGU	Apr 8, 2020	10	Can you just make the remake the season finale?
jacobnoble-02524	May 9, 2019	9	Game of Thrones
heavenacceptme	Aug 22, 2022	10	Despite the final season, Game of Thrones remains an all time classic
Supermanfan-13	Dec 9, 2023	10	One of the best shows ever
Dan_W_Reviews	Feb 5, 2023	9	Captivating and Gripping but a Disappointing Final Season
tweaknhoe	May 25, 2019	8	Why just why? This show could have been the best ever.
TheLittleSongbird	Nov 8, 2017	10	This is a television show?
kunkell	Nov 1, 2019	6	Imagine an Ice Cream Shop
akhil-marsonya27	May 20, 2019	1	A Message to Dan and Dave
Rob1331	Nov 10, 2023	10	Amazing
alshamari-marwa	May 18, 2020	9	Extraordinary untill season 8
tobiascramon	Nov 8, 2022	9	one the best shows ever made
adamheritage-15333	May 5, 2019	9	Can we please just restart season 8, perhaps 7 as well, but mainly 8
LASTRONOME	Jun 2, 2020	10	almost perfect
SpitOnAStranger	May 9, 2019	8	This was an 10/10 until S08E03
psypeterson	Apr 16, 2011	10	Excellent adaptation.
el-absy	May 19, 2019	7	Great Beginning, WORST Ending!

```
## 5                                                      Game of Thrones
## 6   Despite the final season, Game of Thrones remains an all time classic
## 7                                                 One of the best shows ever
## 8             Captivating and Gripping but a Disappointing Final Season
## 9                 Why just why? This show could have been the best ever.
## 10                                                This is a television show?
## 11                                                Imagine an Ice Cream Shop
## 12                                                A Message to Dan and Dave
## 13                                                                  Amazing
## 14                                            Extraordinary untill season 8
## 15                                             one the best shows ever made
## 16   Can we please just restart season 8, perhaps 7 as well, but mainly 8
## 17                                                           almost perfect
## 18                                          This was an 10/10 until S08E03
## 19                                                    Excellent adaptation.
## 20                                           Great Beginning, WORST Ending!
```

```r
gameofthrones %>%
  kable("latex", booktabs = TRUE) %>%
  kable_styling(latex_options = "scale_down")
```

```r
breakingbad <- data.frame(Name[[2]],
    Date = DateOfTheReview[[2]],
    Rating = userRating[[2]],
    Title = titleOfTheReview[[2]],
    stringsAsFactors = FALSE
  )
```

```r
print(head(breakingbad, 20))
```

```
##               Name..2..        Date Rating
## 1               FiRE010  Jul 3, 2021     10
## 2             bruhperson  Mar 6, 2019     10
```

```
## 3         KinoKoopaKid Jul 29, 2021     10
## 4           jehuschultz Feb 18, 2020     10
## 5        Supermanfan-13  Nov 8, 2021     10
## 6     manishsingh-03299 May 30, 2019     10
## 7               Rob1331  Dec 8, 2022     10
## 8              xpinerhd Nov 15, 2019     10
## 9    dhanushreddy-14919 Jul 17, 2021     10
## 10    TheLittleSongbird Nov 12, 2017     10
## 11              dyarutd Sep 28, 2024      7
## 12            reebokroot Feb 14, 2021      5
## 13          alpierce1991 May 18, 2014     10
## 14           tushv-31482  Dec 8, 2022     10
## 15          gogoschka-1 Jan 11, 2014     10
## 16           FishDrowned  Nov 8, 2021     10
## 17            joegalgano Aug 11, 2021     10
## 18          agatt-87232 May 19, 2019     10
## 19        Leofwine_draca  May 4, 2021     10
## 20          dristysultana Jun 23, 2021     10
##
## 1
## 2
## 3
## 4
## 5
## 6                                                                              Those days a
## 7                                                                                          I
## 8                                                                                          A
## 9                                                         Among the best and most ac
## 10                                                                       My Review For
## 11
## 12
## 13                                                        Easily the most overrated televis
## 14                                                                                     Once
## 15 If you mix Scarface, Robin Hood and maybe Tyler Durden with enough meth - you'll get a mean cockta
## 16                                                                       By far the greatest
## 17                                                                               in a ca
## 18                                                        Since GOT is over, this is Officially the G
## 19                                                                            Every bit a
## 20
```

```r
breakingbad %>%
  kable("latex", booktabs = TRUE) %>%
  kable_styling(latex_options = "scale_down")
```

```r
strangerthings <- data.frame(Name[[3]],
    Date = DateOfTheReview[[3]],
    Rating = userRating[[3]],
    Title = titleOfTheReview[[3]],
    stringsAsFactors = FALSE
  )

print(head(strangerthings, 20))
```

```
##             Name..3..        Date Rating
## 1       Supermanfan-13 Jan 22, 2023      9
```

Name..2..	Date	Rating	Title
FiRE010	Jul 3, 2021	10	Really Great
bruhperson	Mar 6, 2019	10	It's ok I guess
KinoKoopaKid	Jul 29, 2021	10	99.1% pure
jehuschultz	Feb 18, 2020	10	The Best
Supermanfan-13	Nov 8, 2021	10	Damn near perfect!
manishsingh-03299	May 30, 2019	10	Those days ain't gonna come back..
Rob1331	Dec 8, 2022	10	Absolutely Brilliant!!
xpinerhd	Nov 15, 2019	10	Best TV show ever made.
dhanushreddy-14919	Jul 17, 2021	10	ABSOLUTE PERFECTION!!!!
TheLittleSongbird	Nov 12, 2017	10	Among the best and most addictive shows there is
dyarutd	Sep 28, 2024	7	My Review For The Best TV Show Ever
reebokroot	Feb 14, 2021	5	mmmmm
alpierce1991	May 18, 2014	10	Easily the most overrated television show in history...
tushv-31482	Dec 8, 2022	10	Once in a life time series
gogoschka-1	Jan 11, 2014	10	If you mix Scarface, Robin Hood and maybe Tyler Durden with enough meth - you'll get a mean cocktail called 'Heisenberg'
FishDrowned	Nov 8, 2021	10	By far the greatest show I've ever watched
joegalgano	Aug 11, 2021	10	in a category all on its own.
agatt-87232	May 19, 2019	10	Since GOT is over, this is Officially the Greatest show ever made
Leofwine_draca	May 4, 2021	10	Every bit as good as you've heard
dristysultana	Jun 23, 2021	10	Masterpiece

```
## 2       Sleepin_Dragon  Jan 8, 2018      8
## 3              cherold Jul 21, 2016      9
## 4               Rob1331 Mar 10, 2023      9
## 5          planktonrules Nov 22, 2016     10
## 6               EVON1TY Mar 12, 2023      9
## 7           gogoschka-1 Jan 22, 2018     10
## 8         magnoliacream Jun 21, 2022      9
## 9          personalhonor Sep 22, 2020     10
## 10       chiragrathod19 Jan 22, 2023     10
## 11           pnxrxvgcfg  Jul 3, 2022     10
## 12   From_the_Future Jul 14, 2016      7
## 13   JarekReanimated  Jul 3, 2019      5
## 14           cdanielabr Jul 16, 2019     10
## 15               cc0077  Jul 9, 2022     10
## 16           slayerjmk95 Jul 14, 2016      9
## 17             TurnWolf Jul 14, 2016     10
## 18   abbykitkatyomama Jul 23, 2021      8
## 19          ChrystineA74 Jul 14, 2016     10
## 20 thisisreallychris May 30, 2022     10
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7  A nostalgic callback to the stories from my youth - but also a beautifully shot supernatural tale
## 8
## 9
## 10
## 11
## 12
## 13
## 14                                                                                              I rated it 9 l
## 15
## 16
## 17
## 18
```

Name..3..	Date	Rating	Title
Supermanfan-13	Jan 22, 2023	9	Lives up to all the hype
Sleepin_Dragon	Jan 8, 2018	8	Never judge a book by its cover.
cherold	Jul 21, 2016	9	Captures that '80s magic
Rob1331	Mar 10, 2023	9	One of the best supernatural shows ever
planktonrules	Nov 22, 2016	10	Strange, that's for sure...but also very, very good.
EVON1TY	Mar 12, 2023	9	Gets even Stranger in every season!
gogoschka-1	Jan 22, 2018	10	A nostalgic callback to the stories from my youth - but also a beautifully shot supernatural tale with charming performances and great production values
magnoliacream	Jun 21, 2022	9	Never fails to satisfy
personalhonor	Sep 22, 2020	10	One of my favourites
chiragrathod19	Jan 22, 2023	10	One of the best Netflix show.
pnxrxvgcfg	Jul 3, 2022	10	Best series I have ever watched
From__the__Future	Jul 14, 2016	7	Netflix Nails it again.
JarekReanimated	Jul 3, 2019	5	Season 3 feels like a parody of the show.
cdanielabr	Jul 16, 2019	10	I rated it 9 before, but after 3rd season I had to downgrade it to 5
cc0077	Jul 9, 2022	10	I get the hype now
slayerjmk95	Jul 14, 2016	9	Steven Spielberg meets Stephen King
TurnWolf	Jul 14, 2016	10	A spectacular mash of E.T, X-Files and 80's.
abbykitkatyomama	Jul 23, 2021	8	THE BEST
ChrystineA74	Jul 14, 2016	10	The Perfect Mash-Up of the Best 80's Sci-Fi!
thisisreallychris	May 30, 2022	10	WHAT THE F...

```
## 19
## 20
```

```
strangerthings %>%
  kable("latex", booktabs = TRUE) %>%
  kable_styling(latex_options = "scale_down")

friends <- data.frame(Name[[4]],
    Date = DateOfTheReview[[4]],
    Rating = userRating[[4]],
    Title = titleOfTheReview[[4]],
    stringsAsFactors = FALSE
  )

print(head(friends, 20))
```

```
##                        Name..4..          Date Rating
## 1                         jose990 Oct 28, 2023      9
## 2                       dee-regan Oct 28, 2023     10
## 3                         Rob1331 Sep 28, 2022     10
## 4                   Supermanfan-13 May 15, 2022     10
## 5               salokurdovanidze Nov 11, 2020     10
## 6                            midz Aug 11, 1998     10
## 7                   Zeitgeist1313 Jul 25, 2005     10
## 8                      alizafatyma Sep 26, 2022     10
## 9  Mangalassery_Neelakandan Aug 23, 2022      9
## 10                   Moviebility Oct 17, 2022     10
## 11                   manojmagecha Mar 24, 2020      6
## 12                     sofiaboneva Mar 12, 2020      5
## 13                 julian_abadia Nov 30, 2008     10
## 14                     Fabrizio03 Feb 19, 2005     10
## 15                       RahulM007 Oct 24, 2019     10
## 16                  yaxleycratler  Aug 5, 2018     10
## 17                     ramtrishan Sep 11, 2019     10
## 18                 patricia_abiog Jan 12, 2009     10
## 19                        embless Aug 12, 2018     10
## 20                 elsavanswieten  Jun 8, 2018     10
##                                                Title
## 1                                       RIP Chandler
## 2                             Goodnight Chandler Bing
## 3             One of the best sitcoms of all-time!
```

Name..4..	Date	Rating	Title
jose990	Oct 28, 2023	9	RIP Chandler
dee-regan	Oct 28, 2023	10	Goodnight Chandler Bing
Rob1331	Sep 28, 2022	10	One of the best sitcoms of all-time!
Supermanfan-13	May 15, 2022	10	Just as funny today
salokurdovanidze	Nov 11, 2020	10	Can this show be any better?!
midz	Aug 11, 1998	10	My Friends on T.V.
Zeitgeist1313	Jul 25, 2005	10	The best American TV has to offer
alizafatyma	Sep 26, 2022	10	MAKES ME FEEL HAPPY AND CALM
Mangalassery_Neelakandan	Aug 23, 2022	9	FINE WINE
Moviebility	Oct 17, 2022	10	Go-To show when you're sad
manojmagecha	Mar 24, 2020	6	A timeless and endless classic
sofiaboneva	Mar 12, 2020	5	The most overrated show I have ever watched
julian_abadia	Nov 30, 2008	10	The idea was good, but...
Fabrizio03	Feb 19, 2005	10	The Best Show Ever
RahulM007	Oct 24, 2019	10	The One that started it all!!!!!!!!
yaxleycratler	Aug 5, 2018	10	They'll be there for you!!!
ramtrishan	Sep 11, 2019	10	G.R.E.A.T.
patricia_abiog	Jan 12, 2009	10	A series No one will EVER forget
embless	Aug 12, 2018	10	Just finished watching whole series on Netflix
elsavanswieten	Jun 8, 2018	10	I still watch friends 4 times a week

```
## 4                            Just as funny today
## 5                      Can this show be any better?!
## 6                            My Friends on T.V.
## 7                The best American TV has to offer
## 8                  MAKES ME FEEL HAPPY AND CALM
## 9                                      FINE WINE
## 10                    Go-To show when you're sad
## 11                A timeless and endless classic
## 12    The most overrated show I have ever watched
## 13                        The idea was good, but...
## 14                            The Best Show Ever
## 15          The One that started it all!!!!!!!!
## 16                    They'll be there for you!!!
## 17                                    G.R.E.A.T.
## 18              A series No one will EVER forget
## 19 Just finished watching whole series on Netflix
## 20            I still watch friends 4 times a week
```

```r
friends %>%
  kable("latex", booktabs = TRUE) %>%
  kable_styling(latex_options = "scale_down")
```

```r
twd <- data.frame(Name[[5]],
    Date = DateOfTheReview[[5]],
    Rating = userRating[[5]],
    Title = titleOfTheReview[[5]],
    stringsAsFactors = FALSE
  )
```

```r
print(head(twd, 20))
```

```
##                 Name..5..        Date Rating
## 1             A2IOS08 Oct 22, 2024      9
## 2           mickelking Oct 14, 2024      9
## 3       Supermanfan-13 Oct 30, 2024      9
## 4             atm21669 Oct 31, 2024     10
## 5           sanilkuttan  Nov 4, 2024     10
## 6             gavdurlam Oct 19, 2024     10
## 7          aaliyahoneil  Nov 8, 2024      9
## 8              ReloadZ_ Oct 13, 2024      8
## 9         rosschiasson  Oct 6, 2024     10
## 10  marissaalda-31761  Oct 7, 2024      6
## 11 jmamede-942-353517  Mar 2, 2023      8
## 12          carro-20412  Oct 9, 2024      6
## 13    Silentship-93758  Jan 5, 2021      5
## 14              niriath Jan 29, 2022      8
## 15            kapilnhai Oct 21, 2024      9
## 16              Rob1331 Nov 28, 2022      8
## 17            balder777 Jun 12, 2021     10
## 18          thatmediaguy  Feb 1, 2018     10
## 19           tinytiffiny Sep 30, 2024     10
## 20          noblooakkit Sep 27, 2024      6
##                                                                              Title
## 1                                                                        My Fav Show
## 2                         Wow.. So much more than just a horror zombie show. 8.2 out of 10
## 3                                      I miss it even though it was prob time for it to end
## 4   Imagine how good a 10 episode reunion/ finale series where the whole group gets back together.
## 5                                                                       A masterpiece.
## 6                                          This show just needs a comic book replication show
## 7                                                                        Undying Tales
## 8                                             One of the best modern zombie classics !
## 9                                                                        Rollercoaster
## 10                                                                     Really good show.
## 11              It got me hooked for 6 to 7 seasons then stupidity grew too much to bear
## 12                                                                   Could've been a 10
## 13                                                                Until season 6 amazing
## 14                                                                         A real shame
## 15                                       S1-S5 awesome, not watching the rest because....
## 16                                                        The early seasons were classic!
## 17                                                           Amazing if you binge-watch.
## 18         Love Show BUT The Writers/Producers Need To Get Back To War Against Walkers!!!
## 19                                                                       best show ever
## 20                                                                         Amazing show
```

```r
twd %>%
  kable("latex", booktabs = TRUE) %>%
  kable_styling(latex_options = "scale_down")
```

```r
library(ggplot2)

movies$year <- as.numeric(movies$years)

yearCount <- movies %>%
  filter(!is.na(years)) %>%
  count(years)
```

Name..5..	Date	Rating	Title
A2I0S08	Oct 22, 2024	9	My Fav Show
mickelking	Oct 14, 2024	9	Wow.. So much more than just a horror zombie show. 8.2 out of 10
Supermanfan-13	Oct 30, 2024	9	I miss it even though it was prob time for it to end
atm21669	Oct 31, 2024	10	Imagine how good a 10 episode reunion/ finale series where the whole group gets back together.
sanilkuttan	Nov 4, 2024	10	A masterpiece.
gavdurlam	Oct 19, 2024	10	This show just needs a comic book replication show
aaliyahoneil	Nov 8, 2024	9	Undying Tales
ReloadZ__	Oct 13, 2024	8	One of the best modern zombie classics !
rosschiasson	Oct 6, 2024	10	Rollercoaster
marissaalda-31761	Oct 7, 2024	6	Really good show.
jmamede-942-353517	Mar 2, 2023	8	It got me hooked for 6 to 7 seasons then stupidity grew too much to bear
carro-20412	Oct 9, 2024	6	Could've been a 10
Silentship-93758	Jan 5, 2021	5	Until season 6 amazing
niriath	Jan 29, 2022	8	A real shame
kapilnhai	Oct 21, 2024	9	S1-S5 awesome, not watching the rest because....
Rob1331	Nov 28, 2022	8	The early seasons were classic!
balder777	Jun 12, 2021	10	Amazing if you binge-watch.
thatmediaguy	Feb 1, 2018	10	Love Show BUT The Writers/Producers Need To Get Back To War Against Walkers!!!
tinytiffiny	Sep 30, 2024	10	best show ever
noblooakkit	Sep 27, 2024	6	Amazing show

```r
ggplot(yearCount, aes(x = years, y = n)) +
  geom_line(color = "blue") +
  geom_point(color = "red") +
  labs(title = "Number of TV Shows Released by Year",
       x = "Year",
       y = "Number of TV Shows") +
  theme_minimal() + theme( axis.text.x = element_text(angle = 45, hjust = 1) )
```

Number of TV Shows Released by Year

Number of TV Shows

Year

```
most_releases <- yearCount[which.max(yearCount$n), ]
print(most_releases)

## [1] years n
## <0 rows> (or 0-length row.names)
```

Extracting Amazon Product Reviews

4. Select 5 categories from Amazon and select 30 products from each category.

5. Extract the price, description, ratings and reviews of each product.

```
library(rvest)
library(httr)
library(dplyr)
library(polite)
library(kableExtra)
library(rmarkdown)

makeup_url <- 'https://www.amazon.com/s?k=BLUSH+ON&rh=n%3A11058281&ref=nb_sb_noss'

session1 <- bow(makeup_url,
                user_agent = "Educational")
session1

## <polite session> https://www.amazon.com/s?k=BLUSH+ON&rh=n%3A11058281&ref=nb_sb_noss
##      User-agent: Educational
##       robots.txt: 138 rules are defined for 5 bots
##     Crawl delay: 5 sec
##    The path is scrapable for this user-agent
library(rvest)

amazondf1 <- data.frame()

page1 <- scrape(session1)

price1 <- page1 %>%
  html_nodes('.a-price .a-offscreen') %>%
  html_text(trim = TRUE) %>%
  gsub("[^0-9\\.]", "", .)

description1 <- page1 %>%
  html_nodes('.a-color-base.a-text-normal') %>%
  html_text(trim = TRUE) %>%
  gsub("\n", "", .)

ratings1 <- page1 %>%
  html_nodes('.a-icon-alt') %>%
  html_text() %>%
```

```r
  gsub(" out of 5 stars", "", .)

reviews1 <- page1 %>%
  html_nodes('div.a-row.a-spacing-small.review-data') %>%
  html_text()

price1 <- price1[1:min(30, length(price1))]
description1 <- description1[1:min(30, length(description1))]
ratings1 <- ratings1[1:min(30, length(ratings1))]
reviews1 <- reviews1[1:min(30, length(reviews1))]

price1 <- c(price1, rep(NA, 30 - length(price1)))
description1 <- c(description1, rep(NA, 30 - length(description1)))
ratings1 <- c(ratings1, rep(NA, 30 - length(ratings1)))
reviews1 <- c(reviews1, rep(NA, 30 - length(reviews1)))


amazondf1 <- data.frame(
  Prices_amazon = price1,
  Descriptions_amazon = description1,
  Ratings_amazon = ratings1,
  Reviewss_amazon = reviews1,
  stringsAsFactors = FALSE
)

print(amazondf1)
```

```
##    Prices_amazon
## 1           5.64
## 2          17.09
## 3          11.99
## 4           6.73
## 5          20.39
## 6          11.99
## 7           4.47
## 8          13.55
## 9          11.99
## 10          3.28
## 11         15.62
## 12          3.79
## 13          5.98
## 14         37.38
## 15          7.49
## 16          5.00
## 17         32.26
## 18          7.00
## 19         14.89
## 20          9.00
## 21          8.97
## 22         74.75
## 23          5.94
## 24         33.00
## 25          6.99
## 26          5.97
```

```
## 27            20.59
## 28             9.97
## 29            58.65
## 30            11.05
##
## 1                                                                wet n wild Color Icon Blush, Effortless Matte Natural She
## 2                                                                                              Maybelline Fit Me Bl
## 3                                                                                     e.l.f. Monochromatic Mul
## 4
## 5                                                                Milani Baked Blush - Petal Primavera (0
## 6
## 7                                            COVERGIRL Instant Cheekbones Contouring Blush Peach Perfection, 
## 8                                                                NYX PROFESSIONAL MAKEUP Buttermelt Po
## 9                                                                L'Oreal Paris Infallible Up to 24H Fresh Wea
## 10                                                                             e.l.f. Camo Liquid Blush, Long-lasti
## 11                 essence | Baby Got Blush (30 | Rosé All Day) | Easy to Apply & Blend Pigmented Crea
## 12
## 13                                                                Revlon Blush Stick, PhotoReady Insta-Blush Face Makeup 
## 14
## 15                                                                                                                     W
## 16                                                                e.l.f. Primer-Infused Matte Blush, Long-Lasting, 
## 17                                            Maybelline Cheek Heat Gel-Cream Blush Makeup, Lightweight, 
## 18                                                                                             wet n w
## 19
## 20                                                                e.l.f. Halo Glow Blush Beauty Wand, Liqu
## 21                                                                Maybelline Super Stay 24, 2-Step Liquid 
## 22                 Multi-Use Soft Creamy Blush Stick,2-in-1 On-the-Go Blush for Cheek and Lip 
## 23
## 24                                                                                              Maybelline 
## 25                                                                LAURA GELLER NEW YORK 
## 26 FOCALLURE 3 Pcs Cream Contour Sticks,Shades with Highlighter & Bronzer & Blush,Non-greasy Long-wea
## 27                                                                Multi Glow Balm, Cream Blush for Cheeks, Blush Balm Fac
## 28                                                                Pixi On-the-Glow Blush - Fleur, Hydrating Tinted Moistu
## 29                                                                LAURA GELLER NEW YORK Seru
## 30
##     Ratings_amazon Reviewss_amazon
## 1             4.4           <NA>
## 2             4.4           <NA>
## 3             4.3           <NA>
## 4             4.5           <NA>
## 5             4.5           <NA>
## 6             4.5           <NA>
## 7             4.5           <NA>
## 8             4.5           <NA>
## 9             4.4           <NA>
## 10            4.6           <NA>
## 11            4.2           <NA>
## 12            4.4           <NA>
## 13            4.4           <NA>
## 14            4.6           <NA>
## 15            4.4           <NA>
## 16            4.6           <NA>
## 17            4.0           <NA>
## 18            4.3           <NA>
```

```
## 19            4.5            <NA>
## 20            4.4            <NA>
## 21            4.5            <NA>
## 22            4.3            <NA>
## 23            4.4            <NA>
## 24            4.1            <NA>
## 25            4.5            <NA>
## 26            4.0            <NA>
## 27            4.2            <NA>
## 28            4.6            <NA>
## 29            4.1            <NA>
## 30            4.3            <NA>
```

```r
skincare_url <- 'https://www.amazon.com/s?k=moisturizer&rh=n%3A11060451&ref=nb_sb_noss'

session2 <- bow(skincare_url,
                user_agent = "Educational")
session2
```

```
## <polite session> https://www.amazon.com/s?k=moisturizer&rh=n%3A11060451&ref=nb_sb_noss
##      User-agent: Educational
##      robots.txt: 138 rules are defined for 5 bots
##    Crawl delay: 5 sec
##    The path is scrapable for this user-agent
```

```r
library(rvest)

amazondf2 <- data.frame()

page2 <- scrape(session2)

price2 <- page2 %>%
  html_nodes('.a-price .a-offscreen') %>%
  html_text(trim = TRUE) %>%
  gsub("[^0-9\\.]", "", .)

description2 <- page2 %>%
  html_nodes('.a-color-base.a-text-normal') %>%
  html_text(trim = TRUE) %>%
  gsub("\n", "", .)

ratings2 <- page2 %>%
  html_nodes('.a-icon-alt') %>%
  html_text() %>%
  gsub(" out of 5 stars", "", .)

reviews2 <- page2 %>%
  html_nodes('div.a-row.a-spacing-small.review-data') %>%
  html_text()

price2 <- price2[1:min(30, length(price2))]
description2 <- description2[1:min(30, length(description2))]
ratings2 <- ratings2[1:min(30, length(ratings2))]
reviews2 <- reviews2[1:min(30, length(reviews2))]
```

```
price2 <- c(price2, rep(NA, 30 - length(price2)))
description2 <- c(description2, rep(NA, 30 - length(description2)))
ratings2 <- c(ratings2, rep(NA, 30 - length(ratings2)))
reviews2 <- c(reviews2, rep(NA, 30 - length(reviews2)))


amazondf2 <- data.frame(
  Prices_amazon = price2,
  Descriptions_amazon = description2,
  Ratings_amazon = ratings2,
  Reviewss_amazon = reviews2,
  stringsAsFactors = FALSE
)

print(amazondf2)
```

```
##     Prices_amazon
## 1           23.99
## 2            7.10
## 3           17.78
## 4            0.94
## 5           19.49
## 6           12.97
## 7            4.32
## 8           14.95
## 9           13.68
## 10           1.14
## 11          14.99
## 12          14.97
## 13           4.99
## 14          19.99
## 15           9.98
## 16           1.42
## 17          34.99
## 18          22.00
## 19           8.12
## 20          31.99
## 21          35.40
## 22          17.70
## 23          59.00
## 24          14.97
## 25           4.99
## 26          19.99
## 27          19.99
## 28           1.48
## 29          20.30
## 30           7.99
##
## 1                                   La Roche-Posay Toleriane Double Repair Face Moisturizer | Da
## 2           CeraVe Moisturizing Cream | Body and Face Moisturizer for Dry Skin | Body Cream with Hyal
## 3                                        Vanicream Daily Facial Moisturizer With Ceramides and Hy
## 4                     CeraVe Daily Moisturizing Lotion for Dry Skin | Body Lotion & Face Moi
## 5                                          CeraVe PM Facial Moisturizing Lotion | Night C
## 6      COSRX Snail Mucin 92% Moisturizer, Daily Repair Face Gel Cream Tube Type for Dry, Sensitive Sk
```

```
## 7    AESTURA ATOBARRIER365 Cream with Ceramide, Korean Moisturizer for Barrier Repair | 120-hour Lastin
## 8                    IT Cosmetics Confidence in a Cream Anti Aging Face Moisturizer - Visibly Reduces
## 9   CeraVe AM Facial Moisturizing Lotion with SPF 30 | Oil-Free Face Moisturizer with SPF | Formulate
## 10        La Roche-Posay Lipikar AP+ Triple Repair Moisturizing Cream | Face & Body Lotion For Dry
## 11                      Embryolisse Lait-Crème Concentré, Multifunction Daily Moisturizer,
## 12                                          EltaMD UV Daily SPF 40 Tinted Face Su
## 13                    COSRX Hyaluronic Acid Moisturizing Cream, Long-lasting Hydration, R
## 14                                            Beauty of Joseon Dynasty Cream
## 15                              COSRX Snail Mucin 92% Face Moisturizer, 3.52
## 16                          Vanicream Moisturizing Skin Cream with Pump Disp
## 17
## 18       Kiehl's Ultra Facial Cream, with 4.5% Squalane to Strengthen Skin's Moisture Barrier, Skin
## 19                                        e.l.f. SKIN Daily Hydration Moisturizer,
## 20            Neutrogena Hydro Boost Water Gel with Signature Fragrance, Hyaluronic Acid Facia
## 21
## 22    Illiyoon Ceramide Ato Concentrate Cream | Face Moisturizer for Dry & Sensitive Skin, 100hr La
## 23    BYOMA Moisturizing Gel Cream - Barrier Repair Cream/Gel Face Lotion with Niacinamide and Gre
## 24                                     Clinique Moisture Surge 100H Auto-Replenishing
## 25
## 26              Aveeno Calm + Restore Oat Gel Facial Moisturizer for Sensitive Skin, Fast-Absorbing
## 27                                  La Roche-Posay Toleriane Double Repair Matte Daily Face
## 28                    Pyunkang Yul [PKY] Calming Moisture Barrier Cream Instantly Sooth
## 29              Neutrogena Fragrance Free Daily Facial Moisturizer, Face Moisturizer for Sensit
## 30                                   Cetaphil Daily Hydrating Lotion for Face, With Hyaluron
##     Ratings_amazon Reviewss_amazon
## 1             4.6.            <NA>
## 2             4.3.            <NA>
## 3             4.6             <NA>
## 4             4.7             <NA>
## 5             4.6             <NA>
## 6             4.7             <NA>
## 7             4.7             <NA>
## 8             4.6             <NA>
## 9             4.6             <NA>
## 10            4.5             <NA>
## 11            4.5             <NA>
## 12            4.6             <NA>
## 13            4.5             <NA>
## 14            4.5             <NA>
## 15            4.6             <NA>
## 16            4.4             <NA>
## 17            4.5             <NA>
## 18            4.6             <NA>
## 19            4.7             <NA>
## 20            4.6             <NA>
## 21            4.6             <NA>
## 22            4.6             <NA>
## 23            4.5             <NA>
## 24            4.6             <NA>
## 25            4.2             <NA>
## 26            4.7             <NA>
## 27            4.5             <NA>
## 28            4.5             <NA>
## 29            4.5             <NA>
```

```
## 30              4.4             <NA>
```

```r
haircare_url <- 'https://www.amazon.com/s?k=shampoo&rh=n%3A11057241&ref=nb_sb_noss'

session3 <- bow(haircare_url,
                user_agent = "Educational")
session3
```

```
## <polite session> https://www.amazon.com/s?k=shampoo&rh=n%3A11057241&ref=nb_sb_noss
##      User-agent: Educational
##      robots.txt: 138 rules are defined for 5 bots
##    Crawl delay: 5 sec
##    The path is scrapable for this user-agent
```

```r
library(rvest)

amazondf3 <- data.frame()

page3 <- scrape(session3)

price3 <- page3 %>%
  html_nodes('.a-price .a-offscreen') %>%
  html_text(trim = TRUE) %>%
  gsub("[^0-9\\.]", "", .)

description3 <- page3 %>%
  html_nodes('.a-color-base.a-text-normal') %>%
  html_text(trim = TRUE) %>%
  gsub("\n", "", .)

ratings3 <- page3 %>%
  html_nodes('.a-icon-alt') %>%
  html_text() %>%
  gsub(" out of 5 stars", "", .)

reviews3 <- page3 %>%
  html_nodes('div.a-row.a-spacing-small.review-data') %>%
  html_text()

price3 <- price3[1:min(30, length(price3))]
description3 <- description3[1:min(30, length(description3))]
ratings3 <- ratings3[1:min(30, length(ratings3))]
reviews3 <- reviews3[1:min(30, length(reviews3))]

price3 <- c(price3, rep(NA, 30 - length(price3)))
description3 <- c(description3, rep(NA, 30 - length(description3)))
ratings3 <- c(ratings3, rep(NA, 30 - length(ratings3)))
reviews3 <- c(reviews3, rep(NA, 30 - length(reviews3)))


amazondf3 <- data.frame(
  Prices_amazon = price3,
  Descriptions_amazon = description3,
  Ratings_amazon = ratings3,
  Reviewss_amazon = reviews3,
```

```
  stringsAsFactors = FALSE
)

print(amazondf3)
```

```
##    Prices_amazon
## 1          20.98
## 2           1.31
## 3          29.99
## 4          33.94
## 5           1.06
## 6          59.99
## 7          11.88
## 8           2.97
## 9          19.99
## 10         21.00
## 11          2.47
## 12         30.00
## 13         20.98
## 14          1.31
## 15         29.99
## 16         37.00
## 17          4.11
## 18         26.00
## 19          2.57
## 20         42.00
## 21          1.24
## 22         15.00
## 23          1.48
## 24         20.00
## 25          9.97
## 26          0.83
## 27         26.11
## 28          0.41
## 29          7.49
## 30          0.29
##
## 1                                                        Olaplex No. 4 Bond Maintenance Shamp
## 2   PURA D'OR 16 Oz Original Gold Label Anti-Thinning Biotin Shampoo Natural Earthy Scent, CLINICALLY
## 3                                                          Pureology Hydrate Moisturizing Shampoo
## 4
## 5                     Biolage Color Last Shampoo | Helps Protect Hair & Maintain Vibrant Color | Fo
## 6
## 7                                                       Mielle Organics Rosemary Mint Str
## 8                                                       Head & Shoulders Classic Clean Dan
## 9                                                            Dove Ultra Care Shamp
## 10
## 11   Rosemary and Biotin Shampoo for Thinning Hair - Vegan Volumizing Shampoo for Fine Hair with Arga
## 12
## 13                                                       Nioxin Scalp + Hair Thickening
## 14                       Pantene Sulfate Free Rose Water Shampoo, Soothes, Replenishes Hydration
## 15                                                       OGX Renewing Argan Oil of
## 16               OGX Thick & Full + Biotin & Collagen Volumizing Shampoo, Nutrient-Infused Hair Sham
## 17                                                       Botanic Hearth Tea T
```

22

```
## 18                                            OGX Nourishing + Coconut Milk Shampoo, Hydrating & Restoring Shar
## 19                     Nutrafol Shampoo, Cleanse and Hydrate Hair and Scalp, Improves Hair Volume, Str
## 20
## 21                                     Head & Shoulders Clinical Dandruff Shampoo Twin Pack, Presc:
## 22
## 23
## 24
## 25        Native Shampoo and Conditioner Contain Naturally Derived Ingredients | All Hair Type Co
## 26
## 27
## 28                                     L'Oreal Paris Elvive Hyaluron Plump Hydratir
## 29               OGX Nourishing + Coconut Milk Moisturizing Shampoo, Hydrating & Restoring Sh
## 30                     Neutrogena T/Sal Therapeutic Shampoo for Scalp Build-Up Control with Sa
##     Ratings_amazon Reviewss_amazon
## 1             4.6            <NA>
## 2             4.3            <NA>
## 3             4.6            <NA>
## 4             4.7            <NA>
## 5             4.7            <NA>
## 6             4.6            <NA>
## 7             4.5            <NA>
## 8             4.8            <NA>
## 9             4.6            <NA>
## 10            4.5            <NA>
## 11            4.3            <NA>
## 12            4.6            <NA>
## 13            4.7            <NA>
## 14            4.5            <NA>
## 15            4.5            <NA>
## 16            4.4            <NA>
## 17            4.3            <NA>
## 18            4.6            <NA>
## 19            4.3            <NA>
## 20            4.7            <NA>
## 21            4.6            <NA>
## 22            4.6            <NA>
## 23            4.5            <NA>
## 24            4.6            <NA>
## 25            4.4            <NA>
## 26            4.5            <NA>
## 27            4.6            <NA>
## 28            4.6            <NA>
## 29            4.6            <NA>
## 30            4.5            <NA>
```

```r
fragrance_url <- 'https://www.amazon.com/s?k=women&rh=n%3A11056591&ref=nb_sb_noss'

session4 <- bow(fragrance_url,
                user_agent = "Educational")
session4
```

```
## <polite session> https://www.amazon.com/s?k=women&rh=n%3A11056591&ref=nb_sb_noss
##     User-agent: Educational
##     robots.txt: 138 rules are defined for 5 bots
##     Crawl delay: 5 sec
```

```
##    The path is scrapable for this user-agent
library(rvest)

amazondf4 <- data.frame()

page4 <- scrape(session4)

price4 <- page4 %>%
  html_nodes('.a-price .a-offscreen') %>%
  html_text(trim = TRUE) %>%
  gsub("[^0-9\\.]", "", .)

description4 <- page4 %>%
  html_nodes('.a-color-base.a-text-normal') %>%
  html_text(trim = TRUE) %>%
  gsub("\n", "", .)

ratings4 <- page4 %>%
  html_nodes('.a-icon-alt') %>%
  html_text() %>%
  gsub(" out of 5 stars", "", .)

reviews4 <- page4 %>%
  html_nodes('div.a-row.a-spacing-small.review-data') %>%
  html_text()

price4 <- price4[1:min(30, length(price4))]
description4 <- description4[1:min(30, length(description4))]
ratings4 <- ratings4[1:min(30, length(ratings4))]
reviews4 <- reviews4[1:min(30, length(reviews4))]

price4 <- c(price4, rep(NA, 30 - length(price4)))
description4 <- c(description4, rep(NA, 30 - length(description4)))
ratings4 <- c(ratings4, rep(NA, 30 - length(ratings4)))
reviews4 <- c(reviews4, rep(NA, 30 - length(reviews4)))


amazondf4 <- data.frame(
  Prices_amazon = price4,
  Descriptions_amazon = description4,
  Ratings_amazon = ratings4,
  Reviewss_amazon = reviews4,
  stringsAsFactors = FALSE
)

print(amazondf4)

##      Prices_amazon
## 1            78.99
## 2            11.79
## 3           155.00
## 4            22.98
## 5             6.76
## 6            24.99
```

```
## 7            13.27
## 8             1.58
## 9            19.95
## 10           11.21
## 11            4.48
## 12           14.95
## 13           21.34
## 14            6.47
## 15           70.00
## 16          139.99
## 17           41.17
## 18          168.00
## 19           18.89
## 20            5.56
## 21           26.99
## 22           85.50
## 23          114.00
## 24           45.39
## 25           26.70
## 26           74.29
## 27           50.00
## 28           14.71
## 29          100.00
## 30           53.09
##
## 1
## 2
## 3
## 4                          Victoria's Secret Bombshell Mini Fragrance Mist, Notes of Purple Passion
## 5                                                                                               E
## 6
## 7            Eau De Parfum Set for Women (4x20 mL) with Date, Senorita, Glam, Rose | Vanilla, Floral, S
## 8   Calvin Klein Eternity Eau de Parfum - Floral Women's Perfume - With Notes of Bergamot, White Lily
## 9
## 10
## 11
## 12                              Women Pheromone Perfume - Long-lasting and Addictive Personal Ro
## 13                                                                Victoria's Secret Frag
## 14                                                                                       F
## 15                                        Ralph Lauren - Ralph - Eau de Toilette -
## 16
## 17
## 18                                        Victoria's Secret Tease Eau de Parfum, Women's Po
## 19                              Ariana Grande Cloud Eau de Parfum - Warm Gourmand F
## 20
## 21
## 22            Roll-On Pheromone Perfume for Women, rollerball perfume, travel pe
## 23
## 24
## 25
## 26
## 27
## 28
## 29                                                                              Clinique F
```

```
## 30              Calvin Klein Obsession Eau de Parfum - Ambery Women's Perfume - With Notes of Mandarir
##   Ratings_amazon Reviewss_amazon
## 1             4.6            <NA>
## 2             4.4            <NA>
## 3             4.6            <NA>
## 4             4.5            <NA>
## 5             4.6            <NA>
## 6             4.3            <NA>
## 7             4.1            <NA>
## 8             4.6            <NA>
## 9             4.7            <NA>
## 10            4.6            <NA>
## 11            4.5            <NA>
## 12            3.4            <NA>
## 13            4.6            <NA>
## 14            4.1            <NA>
## 15            4.7            <NA>
## 16            4.6            <NA>
## 17            4.7            <NA>
## 18            4.6            <NA>
## 19            4.6            <NA>
## 20            4.6            <NA>
## 21            4.5            <NA>
## 22            4.0            <NA>
## 23            4.6            <NA>
## 24            4.2            <NA>
## 25            4.3            <NA>
## 26            4.6            <NA>
## 27            4.4            <NA>
## 28            4.6            <NA>
## 29            4.7            <NA>
## 30            4.6            <NA>
```

```r
oralcare_url <- 'https://www.amazon.com/s?k=toothbrush&rh=n%3A10079992011&ref=nb_sb_noss'

session5 <- bow(oralcare_url,
                user_agent = "Educational")
session5
```

```
## <polite session> https://www.amazon.com/s?k=toothbrush&rh=n%3A10079992011&ref=nb_sb_noss
##     User-agent: Educational
##     robots.txt: 138 rules are defined for 5 bots
##   Crawl delay: 5 sec
##  The path is scrapable for this user-agent
```

```r
library(rvest)

amazondf5 <- data.frame()

page5 <- scrape(session5)

price5 <- page5 %>%
  html_nodes('.a-price .a-offscreen') %>%
  html_text(trim = TRUE) %>%
  gsub("[^0-9\\.]", "", .)
```

```r
description5 <- page5 %>%
  html_nodes('.a-color-base.a-text-normal') %>%
  html_text(trim = TRUE) %>%
  gsub("\n", "", .)

ratings5 <- page5 %>%
  html_nodes('.a-icon-alt') %>%
  html_text() %>%
  gsub(" out of 5 stars", "", .)

reviews5 <- page5 %>%
  html_nodes('div.a-row.a-spacing-small.review-data') %>%
  html_text()

price5 <- price5[1:min(30, length(price5))]
description5 <- description5[1:min(30, length(description5))]
ratings5 <- ratings5[1:min(30, length(ratings5))]
reviews5 <- reviews5[1:min(30, length(reviews5))]

price5 <- c(price5, rep(NA, 30 - length(price5)))
description5 <- c(description5, rep(NA, 30 - length(description5)))
ratings5 <- c(ratings5, rep(NA, 30 - length(ratings5)))
reviews5 <- c(reviews5, rep(NA, 30 - length(reviews5)))


amazondf5 <- data.frame(
  Prices_amazon = price5,
  Descriptions_amazon = description5,
  Ratings_amazon = ratings5,
  Reviewss_amazon = reviews5,
  stringsAsFactors = FALSE
)

print(amazondf5)
```

```
##     Prices_amazon
## 1           59.99
## 2           59.99
## 3           99.99
## 4           99.99
## 5           99.99
## 6          199.99
## 7          119.99
## 8          119.99
## 9          219.99
## 10          29.95
## 11          29.95
## 12          59.95
## 13          29.99
## 14          29.99
## 15          49.96
## 16          59.96
## 17          59.96
```

```
## 18          109.96
## 19           17.99
## 20            3.00
## 21            6.97
## 22            3.49
## 23            8.99
## 24           39.94
## 25           39.94
## 26           69.99
## 27           39.94
## 28           39.94
## 29           65.99
## 30            5.19
##
## 1                     Aquasonic Black Series Ultra Whitening Toothbrush - ADA Accepted Electric Toothbrush-
## 2                                          Philips Sonicare 4100 Rechargeable Electric Toothbrush, with Pressu
## 3  Philips Sonicare ProtectiveClean 5300 Rechargeable Electric Toothbrush, with Pressure Sensor, 3 Cl
## 4                 Oral-B Pro Health CrossAction All in One Soft Toothbrushes, Deep Plaque Fighter, Re
## 5
## 6
## 7
## 8
## 9                                                      Oral-B Advantage Vivid Dual Action Whitening
## 10                                                          Oral-B Pulsar, Battery Powered 3D
## 11                                              Colgate 360 Vibrate Deep Clean Battery Operated To
## 12
## 13                                              Colgate 360 Optic White Whitening Toothbrush, Ad
## 14
## 15  Oral-B iO Deep Clean Rechargeable Electric Powered Toothbrush, Black with iO Series 3 Limited, 2
## 16
## 17                                                                       Reach Essentials Toothl
## 18
## 19
## 20          7AM2M Sonic Electric Toothbrush with 6 Brush Heads for Adults&Kids, One Charge for 120 Day
## 21
## 22                                                       NIMBUS Extra Soft Toothbrushes (Regular Size Head)
## 23
## 24                                                          Biodegradable Bamboo Toothbrush
## 25
## 26
## 27
## 28
## 29
## 30
##     Ratings_amazon Reviewss_amazon
## 1             4.6             <NA>
## 2             4.4             <NA>
## 3             4.6             <NA>
## 4             4.7             <NA>
## 5             4.7             <NA>
## 6             4.6             <NA>
## 7             4.5             <NA>
## 8             4.5             <NA>
## 9             4.5             <NA>
```

```
## 10              4.6          <NA>
## 11              4.6          <NA>
## 12              4.6          <NA>
## 13              4.6          <NA>
## 14              4.5          <NA>
## 15              4.5          <NA>
## 16              4.5          <NA>
## 17              4.6          <NA>
## 18              4.5          <NA>
## 19              4.7          <NA>
## 20              4.3          <NA>
## 21              4.6          <NA>
## 22              4.8          <NA>
## 23              4.7          <NA>
## 24              4.5          <NA>
## 25   4 Stars & Up           <NA>
## 26             <NA>          <NA>
## 27             <NA>          <NA>
## 28             <NA>          <NA>
## 29             <NA>          <NA>
## 30             <NA>          <NA>
```