R Analytics

2023-12-13

#GROUP 3 #Members: #Laurence Aguas #Dianah Marie Canonicato #Ann Margareth Camayodo #Paula Mae Salvador

install.packages("dplyr")   install.packages("rvest")   install.packages("httr")   install.packages("polite") install.packages("ggplot2")

library(rvest) library(httr) library(dplyr) library(polite) library(ggplot2)

```r
#1.
library(polite)
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
```

```r
link <- "https://www.imdb.com/chart/toptv/?ref_=nv_tvv_250"
session <- bow(link, user_agent= "Educational Purposes")
session
```

```
## <polite session> https://www.imdb.com/chart/toptv/?ref_=nv_tvv_250
##     User-agent: Educational Purposes
##     robots.txt: 34 rules are defined for 2 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
#rank of the Top 50 Tv Shows
rank <- scrape(session)  %>% html_nodes('h3.ipc-title__text') %>% html_text()
rank [2:51]
```

```
##  [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. The World at War"
## [15] "15. Bluey"
## [16] "16. Rick and Morty"
## [17] "17. Fullmetal Alchemist: Brotherhood"
## [18] "18. The Last Dance"
## [19] "19. Life"
## [20] "20. The Twilight Zone"
## [21] "21. Sherlock"
## [22] "22. The Vietnam War"
## [23] "23. Batman: The Animated Series"
## [24] "24. Attack on Titan"
## [25] "25. Scam 1992: The Harshad Mehta Story"
## [26] "26. The Office"
## [27] "27. Arcane"
## [28] "28. The Blue Planet"
## [29] "29. Better Call Saul"
## [30] "30. Human Planet"
## [31] "31. Firefly"
## [32] "32. Frozen Planet"
## [33] "33. Clarkson's Farm"
## [34] "34. Death Note"
## [35] "35. Only Fools and Horses"
## [36] "36. Hunter x Hunter"
## [37] "37. The Civil War"
## [38] "38. True Detective"
## [39] "39. Seinfeld"
## [40] "40. The Beatles: Get Back"
## [41] "41. The Decalogue"
## [42] "42. Persona"
## [43] "43. Fargo"
## [44] "44. Cowboy Bebop"
## [45] "45. Gravity Falls"
## [46] "46. Nathan for You"
## [47] "47. Last Week Tonight with John Oliver"
## [48] "48. When They See Us"
## [49] "49. Succession"
## [50] "50. Apocalypse: The Second World War"
```

```r
#Titles of the Top 50 Tv Shows
title <- scrape(session) %>%
    html_nodes('h3.ipc-title__text') %>%
    html_text()
title [2:51]
```

```
##  [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. The World at War"
## [15] "15. Bluey"
## [16] "16. Rick and Morty"
## [17] "17. Fullmetal Alchemist: Brotherhood"
## [18] "18. The Last Dance"
## [19] "19. Life"
## [20] "20. The Twilight Zone"
## [21] "21. Sherlock"
## [22] "22. The Vietnam War"
## [23] "23. Batman: The Animated Series"
## [24] "24. Attack on Titan"
## [25] "25. Scam 1992: The Harshad Mehta Story"
## [26] "26. The Office"
## [27] "27. Arcane"
## [28] "28. The Blue Planet"
## [29] "29. Better Call Saul"
## [30] "30. Human Planet"
## [31] "31. Firefly"
## [32] "32. Frozen Planet"
## [33] "33. Clarkson's Farm"
## [34] "34. Death Note"
## [35] "35. Only Fools and Horses"
## [36] "36. Hunter x Hunter"
## [37] "37. The Civil War"
## [38] "38. True Detective"
## [39] "39. Seinfeld"
## [40] "40. The Beatles: Get Back"
## [41] "41. The Decalogue"
## [42] "42. Persona"
## [43] "43. Fargo"
## [44] "44. Cowboy Bebop"
## [45] "45. Gravity Falls"
## [46] "46. Nathan for You"
## [47] "47. Last Week Tonight with John Oliver"
## [48] "48. When They See Us"
## [49] "49. Succession"
## [50] "50. Apocalypse: The Second World War"
```

```r
#Rating of the Top 50 Tv Shows
rating <- scrape(session) %>%
    html_nodes('span.ipc-rating-star--imdb') %>%
    html_text()
rating [2:51]
```

```
##  [1] "9.5 (155K)" "9.4 (218K)" "9.4 (508K)" "9.3 (835K)" "9.3 (367K)"
##  [6] "9.3 (349K)" "9.3 (45K)"  "9.2 (450K)" "9.3 (127K)" "9.3 (43K)"
## [11] "9.3 (49K)"  "9.2 (2.2M)" "9.2 (28K)"  "9.4 (22K)"  "9.1 (582K)"
## [16] "9.1 (191K)" "9.1 (146K)" "9.1 (42K)"  "9.1 (91K)"  "9.1 (977K)"
## [21] "9.1 (27K)"  "9.0 (113K)" "9.1 (478K)" "9.3 (154K)" "9.0 (682K)"
## [26] "9.0 (253K)" "9.0 (42K)"  "9.0 (620K)" "9.0 (28K)"  "9.0 (278K)"
```

```
## [31] "9.0 (33K)"  "9.0 (52K)"  "8.9 (365K)" "9.0 (56K)"  "9.0 (126K)"
## [36] "9.0 (18K)"  "8.9 (618K)" "8.9 (343K)" "9.0 (27K)"  "9.0 (27K)"
## [41] "9.0 (45K)"  "8.9 (398K)" "8.9 (134K)" "8.9 (129K)" "8.9 (37K)"
## [46] "8.9 (94K)"  "8.9 (134K)" "8.9 (247K)" "9.0 (14K)"  "8.9 (1.1M)"
```

```r
#Numbers Of Voters of the Top 50 Tv Shows
numberofvoters <- scrape(session) %>%
    html_nodes('span.ipc-rating-star--imdb') %>%
    html_text()
numberofvoters [2:51]
```

```
##  [1] "9.5 (155K)" "9.4 (218K)" "9.4 (508K)" "9.3 (835K)" "9.3 (367K)"
##  [6] "9.3 (349K)" "9.3 (45K)"  "9.2 (450K)" "9.3 (127K)" "9.3 (43K)"
## [11] "9.3 (49K)"  "9.2 (2.2M)" "9.2 (28K)"  "9.4 (22K)"  "9.1 (582K)"
## [16] "9.1 (191K)" "9.1 (146K)" "9.1 (42K)"  "9.1 (91K)"  "9.1 (977K)"
## [21] "9.1 (27K)"  "9.0 (113K)" "9.1 (478K)" "9.3 (154K)" "9.0 (682K)"
## [26] "9.0 (253K)" "9.0 (42K)"  "9.0 (620K)" "9.0 (28K)"  "9.0 (278K)"
## [31] "9.0 (33K)"  "9.0 (52K)"  "8.9 (365K)" "9.0 (56K)"  "9.0 (126K)"
## [36] "9.0 (18K)"  "8.9 (618K)" "8.9 (343K)" "9.0 (27K)"  "9.0 (27K)"
## [41] "9.0 (45K)"  "8.9 (398K)" "8.9 (134K)" "8.9 (129K)" "8.9 (37K)"
## [46] "8.9 (94K)"  "8.9 (134K)" "8.9 (247K)" "9.0 (14K)"  "8.9 (1.1M)"
```

```r
#This shows on how many Episodes does the Top 50 Tv Shows have
episodes <- scrape(session) %>%
    html_nodes('span.ipc-rating-star--voteCount') %>%
    html_text()
episodes [2:51]
```

```
##  [1] " (155K)" " (218K)" " (508K)" " (835K)" " (367K)" " (349K)" " (45K)"
##  [8] " (450K)" " (127K)" " (43K)"  " (49K)"  " (2.2M)" " (28K)"  " (22K)"
## [15] " (582K)" " (191K)" " (146K)" " (42K)"  " (91K)"  " (977K)" " (27K)"
## [22] " (113K)" " (478K)" " (154K)" " (682K)" " (253K)" " (42K)"  " (620K)"
## [29] " (28K)"  " (278K)" " (33K)"  " (52K)"  " (365K)" " (56K)"  " (126K)"
## [36] " (18K)"  " (618K)" " (343K)" " (27K)"  " (27K)"  " (45K)"  " (398K)"
## [43] " (134K)" " (129K)" " (37K)"  " (94K)"  " (134K)" " (247K)" " (14K)"
## [50] " (1.1M)"
```

```r
# The Year it was released (The Top 50 Tv Shows)
year <- scrape(session) %>%
    html_nodes('span.sc-43986a27-8:nth-of-type(1)') %>%
    html_text()
year [2:51]
```

```
##  [1] "2016"      "2006"      "2001"      "2019"      "2002-2008" "2005-2008"
##  [7] "2017"      "1999-2007" "2014"      "1980"      "2019-2023" "2011-2019"
## [13] "1973-1974" "2018- "    "2013- "    "2009-2010" "2020"      "2009"
## [19] "1959-1964" "2010-2017" "2017"      "1992-1995" "2013-2023" "2020"
## [25] "2005-2013" "2021- "    "2001"      "2015-2022" "2011"      "2002-2003"
## [31] "2011-2012" "2021- "    "2006-2007" "1981-2003" "2011-2014" "1990"
## [37] "2014- "    "1989-1998" "2021"      "1989-1990" "2018- "    "2014-2024"
## [43] "1998-1999" "2012-2016" "2013-2017" "2014- "    "2019"      "2018-2023"
## [49] "2009"      "1994-2004"
```

```r
library(polite)
library(rvest)
library(httr)
library(dplyr)
```

```
#Reviewers Data (Name, Date, Rating, Title of the Tv show, and the comment of each reviewers)

firstrevlink <- "https://www.imdb.com/title/tt0903747/reviews?ref_=tt_urv"
session2 <- bow(firstrevlink, user_agent= "Educational Purposes")
session2

## <polite session> https://www.imdb.com/title/tt0903747/reviews?ref_=tt_urv
##      User-agent: Educational Purposes
##      robots.txt: 34 rules are defined for 2 bots
##    Crawl delay: 5 sec
##    The path is scrapable for this user-agent

ReviewersNametv1 <- scrape(session2) %>%
    html_nodes('.display-name-link a') %>%
    html_text()
ReviewersNametv1

##  [1] "FiRE010"            "Supermanfan-13"      "TheLittleSongbird"
##  [4] "KinoKoopaKid"       "jehuschultz"         "gogoschka-1"
##  [7] "bruhperson"         "DiCaprioFan13"       "Leofwine_draca"
## [10] "FishDrowned"        "dhanushreddy-14919"  "EVON1TY"
## [13] "napierslogs"        "valis1949"           "Leofwine_draca"
## [16] "Quinoa1984"         "otnememento-2"       "lerrom"
## [19] "manishsingh-03299"  "alanbenfieldjr"      "xpinerhd"
## [22] "FeastMode"          "fatcat-73450"        "Ltufano23"
## [25] "anthonymichael-93585"

ReviewsDatetv1 <- scrape(session2) %>%
    html_nodes('.review-date') %>%
    html_text()
ReviewsDatetv1

##  [1] "4 July 2021"        "9 November 2021"    "13 November 2017"
##  [4] "30 July 2021"       "19 February 2020"   "12 January 2014"
##  [7] "7 March 2019"       "9 December 2022"    "5 May 2021"
## [10] "9 November 2021"    "18 July 2021"       "3 March 2023"
## [13] "17 May 2010"        "26 February 2009"   "7 December 2020"
## [16] "8 February 2015"    "21 February 2020"   "11 October 2022"
## [19] "31 May 2019"        "27 June 2017"       "16 November 2019"
## [22] "18 August 2019"     "3 February 2023"    "19 September 2023"
## [25] "29 October 2022"

UserRatingtv1 <- scrape(session2) %>%
    html_nodes('.rating-other-user-rating') %>%
    html_text()
UserRatingtv1

##  [1] "\n          10/10\n          " "\n          10/10\n          "
##  [3] "\n          10/10\n          " "\n          10/10\n          "
##  [5] "\n          10/10\n          " "\n          10/10\n          "
##  [7] "\n          10/10\n          " "\n          10/10\n          "
##  [9] "\n          10/10\n          " "\n          10/10\n          "
## [11] "\n          10/10\n          " "\n          10/10\n          "
## [13] "\n          10/10\n          " "\n          10/10\n          "
## [15] "\n          10/10\n          " "\n          10/10\n          "
## [17] "\n          10/10\n          " "\n          10/10\n          "
```

```
## [19] "\n          10/10\n        " "\n          10/10\n        "
## [21] "\n          10/10\n        " "\n          10/10\n        "
## [23] "\n          10/10\n        " "\n          10/10\n        "
## [25] "\n          10/10\n        "
```

```r
RevTitletv1 <-scrape(session2) %>%
    html_nodes('.parent a') %>%
    html_text()
RevTitletv1
```

```
## [1] "Breaking Bad"
```

```r
TextRevtv1 <-scrape(session2) %>%
    html_nodes('.collapsable div.text') %>%
    html_text()
TextRevtv1
```

```
##  [1] "I have never watched a show that is as consistently genuine and engaging as Breaking Bad. This
##  [2] "Breaking Bad is absolutely, without a doubt, one of the greatest tv shows ever created...it's
##  [3] "'Breaking Bad' is one of the most popular rated shows on IMDb, is one of those rarities where
##  [4] "One of the greatest shows ever, the pacing is excellent. The characters are well developed and
##  [5] "I cannot stress enough how good this show is. I've watched a lot of TV in my life and this sho
##  [6] "If you are among the few who haven't seen it yet: believe the hype, it really is THAT good. Br
##  [7] "Re-Watched it 7 times and counting. I guess I liked it."
##  [8] "Breaking Bad is every bit as good as everyone says it is. It's easily one of my favorites show
##  [9] "Breaking Bad feels like fever dream. A really REALLY good fever dream. The directing is absolu
## [10] "Direction- PERFECT!!Screenplay-PERFECT!!Writing-PERFECT!!Score-PERFECT!!Cinematography-PERFECT
## [11] "This serie is unique. This is somehow realistic and amazing also. Even if you're going to watc
## [12] "BREAKING BAD:Season One explodes like a sucker punch to the gut, and is nothing short of mind-
## [13] "It's hard for me to be super objective about this show. Over the course of 62 episodes, right
## [14] "The best series I have seen in my whole life, even better than the wire."
## [15] "Where do I even begin? The writing is near perfect, the actors and characters are all amazing,
## [16] "When you finish the show you'll never be the same..I guarantee you"
## [17] "Drug wars, meth, the lot. I thought no thank you. I kept hearing how good it was and I kept sa
## [18] "I wanna delete my brain and watch it again like I never knew it."
## [19] "PERFECTSeason 1: 8 stars. Too awesome (3 viewings)Season 2: 9 stars. Amazing (3 viewings)Seaso
## [20] "I wish I'd never seen it so I could see it for the first time again.It's the story of an undera
## [21] "Often hailed as one of the greatest TV ever made, Breaking Bad lives up to the hype. You will
## [22] "Breaking Bad is the greatest TV show of all time, and you can tell very easily because it's lo
```

```r
#
lengths1 <- c(
  ReviewerName = length(ReviewersNametv1),
  ReviewDate = length(ReviewsDatetv1),
  UserRating = length(UserRatingtv1),
  Title = length(RevTitletv1),
  Comment = length(TextRevtv1)
)


max_length <- max(lengths1)

ReviewersNametv1 <- c(ReviewersNametv1, rep(NA, max_length - length(ReviewersNametv1)))
ReviewsDatetv1 <- c(ReviewsDatetv1, rep(NA, max_length - length(ReviewsDatetv1)))
UserRatingtv1 <- c(UserRatingtv1, rep(NA, max_length - length(UserRatingtv1)))
RevTitletv1 <- c(RevTitletv1, rep(NA, max_length - length(RevTitletv1)))
```

```r
TextRevtv1 <- c(TextRevtv1, rep(NA, max_length - length(TextRevtv1)))

tv_show1 <- data.frame(
  ReviewerName = ReviewersNametv1,
  ReviewDate = ReviewsDatetv1,
  UserRating = UserRatingtv1,
  Title = RevTitletv1,
  Comment = TextRevtv1
)



#Second Data
secondrevlink <- "https://www.imdb.com/title/tt5491994/reviews?ref_=tt_urv"
session3 <- bow(secondrevlink, user_agent= "Educational Purposes")
session3
```

```
## <polite session> https://www.imdb.com/title/tt5491994/reviews?ref_=tt_urv
##     User-agent: Educational Purposes
##     robots.txt: 34 rules are defined for 2 bots
##   Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
ReviewersNametv2 <- scrape(session3) %>%
    html_nodes('.display-name-link a') %>%
    html_text()
ReviewersNametv2
```

```
##  [1] "Wentloog"           "john-m-madsen"      "thespookybuz"
##  [4] "pjdickinson"        "NeilBarnett"        "arjanhylkema"
##  [7] "dbijis33"           "dhanrajjughead"     "TheLittleSongbird"
## [10] "farshidkarimi"      "grantss"            "ianrobo"
## [13] "vibhus-17780"       "andrewchristianjr"  "rtoac1"
## [16] "areatw"             "panagiotiskatsanos" "salmanu-27386"
## [19] "BobFillmore"        "megannc"            "fierceeagle-40009"
## [22] "myersei-165-4350"   "RahulM007"          "cgtam"
## [25] "chubarova"
```

```r
ReviewsDatetv2 <- scrape(session3) %>%
    html_nodes('.review-date') %>%
    html_text()
ReviewsDatetv2
```

```
##  [1] "6 November 2016"   "6 November 2016"   "10 November 2016"
##  [4] "6 November 2016"   "14 November 2016"  "8 November 2016"
##  [7] "9 November 2016"   "18 November 2016"  "13 October 2017"
## [10] "23 November 2016"  "13 January 2023"   "20 November 2016"
## [13] "9 April 2023"      "20 October 2021"   "3 April 2022"
## [16] "29 December 2016"  "1 January 2017"    "7 November 2016"
## [19] "30 September 2017" "20 April 2023"     "24 April 2020"
## [22] "5 December 2016"   "31 December 2020"  "12 February 2020"
## [25] "8 January 2021"
```

```r
UserRatingtv2 <- scrape(session3) %>%
    html_nodes('.rating-other-user-rating') %>%
    html_text()
```

```
UserRatingtv2
```

```
##  [1] "\n            10/10\n        " "\n            10/10\n        "
##  [3] "\n            10/10\n        " "\n            10/10\n        "
##  [5] "\n            10/10\n        " "\n            10/10\n        "
##  [7] "\n            10/10\n        " "\n            10/10\n        "
##  [9] "\n            10/10\n        " "\n            10/10\n        "
## [11] "\n            10/10\n        " "\n            10/10\n        "
## [13] "\n            10/10\n        " "\n            10/10\n        "
## [15] "\n            10/10\n        " "\n            10/10\n        "
## [17] "\n            10/10\n        " "\n            10/10\n        "
## [19] "\n            10/10\n        " "\n            10/10\n        "
## [21] "\n            10/10\n        " "\n            10/10\n        "
## [23] "\n            10/10\n        " "\n            10/10\n        "
```

```
RevTitletv2 <-scrape(session3) %>%
    html_nodes('.parent a') %>%
    html_text()
RevTitletv2
```

```
## [1] "Planet Earth II"
```

```
TextRevtv2 <-scrape(session3) %>%
    html_nodes('.collapsable div.text') %>%
    html_text()
TextRevtv2
```

```
##  [1] "I have just finished watching the first episode on BBC 1 and I am amazed to find that Planet E
##  [2] "What to say about Planet Earth II which has not already been said about so many other fantasti
##  [3] "The original Planet Earth back in 2006 was an astounding and refreshing series that tore down
##  [4] "The professionalism of everyone involved in this glorious production is evident in every frame
##  [5] "This is without doubt the best thing I have seen on television for years. The photography, nar
##  [6] "As Planet Earth I was released about ten years ago, a visual improvement could have been expec
##  [7] "Hello this is my first written review. Note that I haven't watched the first season (2006) so
##  [8] "By far the greatest and best nature documentaries ever made. Everyone needs to watch this. I h
##  [9] "Absolutely adore the first 'Planet Earth' from 2007, one of the best documentaries ever made a
## [10] "This is my first comment and I hope I am not bungling it.I have been watching nature documenta
## [11] "A documentary series on the wildlife found on Earth. Each episode covers a different habitat:
## [12] "Plant Earth 2 was everything you would expect and more. Every scene, every vista shows that hu
## [13] "Spectacular Experience\nActually watched it out of curiosity why it's highly rated before gett
## [14] "The sequel of the 2006 documentary is a much more advanced and technical spectacle -- with jus
## [15] "I would call this a must watch for children and adults alike.. If we want to genuine empathy i
## [16] "I never thought the original 'Planet Earth' could be topped, but 10 years on and Attenborough
## [17] "One of the best documentaries I 've ever seen. Every episode is a new wonderful experience,get
## [18] "The opening image of Planet Earth II contained the promise of everything to come. It was specta
## [19] "This is probably the most intimate documentary I have seen of our fellow species on earth. Awe
## [20] "This documentary continues what the first one did so well. Using excellent and high-quality ca
## [21] "Let me say the score, the music, the scenery, the animals and all the detail is like nothing y
## [22] "It's a shame I can only give this breathtaking masterpiece 10 out of 10, it really deserves mo
## [23] "The documentary titled \"Planet Earth: Part II\" is a TV Mini-Series showcasing life on planet
## [24] "David Attenborough's narration completes the fascinating cinematography, and dire need for imme
## [25] "I don't want to say too much. This documentary is amazing, outstanding, awesome and other syno
```

```
#
lengths2 <- c(
  ReviewerName = length(ReviewersNametv2),
```

```r
  ReviewDate = length(ReviewsDatetv2),
  UserRating = length(UserRatingtv2),
  Title = length(RevTitletv2),
  Comment = length(TextRevtv2)
)

# Find the maximum length
max_length2 <- max(lengths2)

ReviewersNametv2 <- c(ReviewersNametv2, rep(NA, max_length - length(ReviewersNametv2)))
ReviewsDatetv2 <- c(ReviewsDatetv2, rep(NA, max_length - length(ReviewsDatetv2)))
UserRatingtv2 <- c(UserRatingtv2, rep(NA, max_length - length(UserRatingtv2)))
RevTitletv2 <- c(RevTitletv2, rep(NA, max_length - length(RevTitletv2)))
TextRevtv2 <- c(TextRevtv2, rep(NA, max_length - length(TextRevtv2)))

tv_show2 <- data.frame(
  ReviewerName = ReviewersNametv2,
  ReviewDate = ReviewsDatetv2,
  UserRating = UserRatingtv2,
  Title = RevTitletv2,
  Comment = TextRevtv2
)


#third
thirdrevlink <- "https://www.imdb.com/title/tt0185906/reviews?ref_=tt_urv"
session4 <- bow(thirdrevlink, user_agent= "Educational Purposes")
session4
```

```
## <polite session> https://www.imdb.com/title/tt0185906/reviews?ref_=tt_urv
##     User-agent: Educational Purposes
##      robots.txt: 34 rules are defined for 2 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
ReviewersNametv3 <- scrape(session4) %>%
    html_nodes('.display-name-link a') %>%
    html_text()
ReviewersNametv3
```

```
##  [1] "rbverhoef"          "philip_vanderveken" "DiCaprioFan13"
##  [4] "bsmith5552"         "planktonrules"      "theshape79"
##  [7] "yodaschoda"         "SnoopyStyle"        "grahamsj3"
## [10] "mickman91-1"        "Libretio"           "erwan_ticheler"
## [13] "wildcatt268"        "Supermanfan-13"     "mhorg2018"
## [16] "arjay24"            "faded_english_monkey" "kipmcmillan"
## [19] "TusharViv"          "bob the moo"        "MovieCriticDave"
## [22] "jazmodo"            "paulimiles"         "paul_haakonsen"
```

```r
ReviewsDatetv3 <- scrape(session4) %>%
    html_nodes('.review-date') %>%
    html_text()
ReviewsDatetv3
```

```
##  [1] "14 February 2003"  "17 September 2004" "28 September 2022"
##  [4] "6 November 2001"   "31 May 2015"       "5 November 2001"
```

```
##  [7] "24 January 2005"    "5 December 2015"   "26 November 2002"
## [10] "4 May 2022"         "4 May 2005"        "8 December 2002"
## [13] "19 January 2002"    "7 May 2022"        "6 June 2019"
## [16] "19 April 2004"      "26 August 2004"    "25 October 2018"
## [19] "2 August 2022"      "7 February 2016"   "13 April 2005"
## [22] "4 June 2019"        "25 July 2010"      "1 December 2012"
```

```r
UserRatingtv3 <- scrape(session4) %>%
    html_nodes('.rating-other-user-rating') %>%
    html_text()
UserRatingtv3
```

```
##  [1] "\n            10/10\n        " "\n            10/10\n        "
##  [3] "\n            10/10\n        " "\n            10/10\n        "
##  [5] "\n            10/10\n        " "\n            10/10\n        "
##  [7] "\n            10/10\n        " "\n            10/10\n        "
##  [9] "\n            10/10\n        " "\n            10/10\n        "
## [11] "\n            9/10\n        "  "\n            10/10\n        "
## [13] "\n            10/10\n        " "\n            10/10\n        "
## [15] "\n            10/10\n        " "\n            10/10\n        "
## [17] "\n            10/10\n        " "\n            10/10\n        "
## [19] "\n            10/10\n        " "\n            10/10\n        "
## [21] "\n            10/10\n        "
```

```r
RevTitletv3 <-scrape(session4) %>%
    html_nodes('.parent a') %>%
    html_text()
RevTitletv3
```

```
## [1] "Band of Brothers"
```

```r
TextRevtv3 <-scrape(session4) %>%
    html_nodes('.collapsable div.text') %>%
    html_text()
TextRevtv3
```

```
##  [1] "This week I saw three things based on WW-II novels. The first was 'The Pianist' about the Warsa
##  [2] "There aren't much TV-series which have left such an impression on me as Band of Brothers did. I
##  [3] "Band of Brothers absolutely one of the best miniseries of all-time...period. Not only that but
##  [4] "\"Band of Brothers\" in a word is awesome. I couldn't wait to see each episode. Co-Executive P
##  [5] "\"Band of Brothers\" is a nearly 12 hour long show about the experiences of a group of soldier
##  [6] "Band of Brothers finds us following the exploits of Easy Company throughout their campaign in
##  [7] "This is the true story of Easy Company, 506th Parachute Infantry Regiment, 101st Airborne Divi
##  [8] "Judging by other comments, it seems that this miniseries struck a chord with many viewers. I a
##  [9] "After 20 years it remains rightfully one of the highest rated things on IMDB. This is because
## [10] "BAND OF BROTHERS Aspect ratio: 1.78:1Sound format: Dolby Digital(10 episodes)The trials and tr
## [11] "I have read virtually all of Ambrose's WWII books, and this mini-series faithfully follows one
## [12] "Band of Brothers absolutely one of the best miniseries of all-time...period. Not only that but
## [13] "Yes, I said it. Why? Because this is a true story. It doesn't (unlike the inferior Pacific) ho
## [14] "I am an 80 year old combat veteran (88th Inf. Div. Italy).I watched BOB on the History Channel
## [15] "I absolutely love this miniseries.As a keen amateur historian, I got sick and tired of books a
## [16] "This is an incredibly poignant and important show that everyone should watch once. It's so rea
## [17] "This is peak television. It's authentic and characters feel very real. It feels like you're th
## [18] "It has been some time since I watched this show on first broadcast - long before streaming and
## [19] "How do we perpetuate the honor, the memory, and the lesson of the nameless and faceless thousa
## [20] "As near to perfection as a TV Series can be, I cannot put into words the gravity & importance
```

```
## [21] "In some ways reviewing Band of Brothers ten years on is superfluous. Its already received every
## [22] "\"Band of Brothers\" is one of my all time favorite series to have been on TV, and of course st
```

```r
lengths3 <- c(
  ReviewerName = length(ReviewersNametv3),
  ReviewDate = length(ReviewsDatetv3),
  UserRating = length(UserRatingtv3),
  Title = length(RevTitletv3),
  Comment = length(TextRevtv3)
)

max_length3 <- max(lengths3)

ReviewersNametv3 <- c(ReviewersNametv3, rep(NA, max_length - length(ReviewersNametv3)))
ReviewsDatetv3 <- c(ReviewsDatetv3, rep(NA, max_length - length(ReviewsDatetv3)))
UserRatingtv3 <- c(UserRatingtv3, rep(NA, max_length - length(UserRatingtv3)))
RevTitletv3 <- c(RevTitletv3, rep(NA, max_length - length(RevTitletv3)))
TextRevtv3 <- c(TextRevtv3, rep(NA, max_length - length(TextRevtv3)))

# Create the data frame
tv_show3 <- data.frame(
  ReviewerName = ReviewersNametv3,
  ReviewDate = ReviewsDatetv3,
  UserRating = UserRatingtv3,
  Title = RevTitletv3,
  Comment = TextRevtv3
)
```

```r
#fourth
fourthrevlink <- "https://www.imdb.com/title/tt7366338/reviews?ref_=tt_urv"
session5 <- bow(fourthrevlink, user_agent= "Educational Purposes")
session5
```

```
## <polite session> https://www.imdb.com/title/tt7366338/reviews?ref_=tt_urv
##      User-agent: Educational Purposes
##      robots.txt: 34 rules are defined for 2 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
ReviewersNametv4 <- scrape(session5) %>%
    html_nodes('.display-name-link a') %>%
    html_text()
ReviewersNametv4
```

```
##  [1] "Leofwine_draca"        "jfirebug"
##  [3] "ahmetkozan"            "deepfrieddodo"
##  [5] "EVON1TY"               "emholberg"
##  [7] "Sleepin_Dragon"        "SnoopyStyle"
##  [9] "thegldt"               "Lladerat"
## [11] "classicsoncall"        "DimitrisPassas-TapTheLine"
## [13] "Vivkon"                "justahunch-70549"
## [15] "DiCaprioFan13"         "wmeduardowm"
## [17] "jazz1"                 "gregoryblanch88"
```

```
## [19] "anyakiss"              "m-porpaczi"
## [21] "dannehh"               "raggingbull"
## [23] "manga-th"              "claudio_carvalho"
## [25] "natashapekar"
```

```
ReviewsDatetv4 <- scrape(session5) %>%
    html_nodes('.review-date') %>%
    html_text()
ReviewsDatetv4
```

```
##  [1] "28 November 2019"  "21 May 2019"       "8 June 2019"
##  [4] "6 September 2022"  "13 March 2023"     "27 May 2019"
##  [7] "2 June 2019"       "5 June 2019"       "7 May 2019"
## [10] "8 May 2019"        "6 February 2020"   "24 June 2019"
## [13] "7 May 2019"        "10 October 2022"   "28 September 2022"
## [16] "7 May 2019"        "30 May 2019"       "31 May 2019"
## [19] "7 May 2019"        "15 May 2019"       "4 June 2019"
## [22] "14 May 2019"       "12 May 2019"       "6 February 2020"
## [25] "10 May 2019"
```

```
UserRatingtv4 <- scrape(session5) %>%
    html_nodes('.rating-other-user-rating') %>%
    html_text()
UserRatingtv4
```

```
##  [1] "\n          10/10\n          " "\n          10/10\n          "
##  [3] "\n          10/10\n          " "\n          10/10\n          "
##  [5] "\n          10/10\n          " "\n          9/10\n          "
##  [7] "\n          10/10\n          " "\n          9/10\n          "
##  [9] "\n          10/10\n          " "\n          10/10\n          "
## [11] "\n          10/10\n          " "\n          10/10\n          "
## [13] "\n          10/10\n          " "\n          10/10\n          "
## [15] "\n          9/10\n          " "\n          10/10\n          "
## [17] "\n          10/10\n          " "\n          10/10\n          "
## [19] "\n          9/10\n          " "\n          10/10\n          "
## [21] "\n          10/10\n          " "\n          10/10\n          "
## [23] "\n          10/10\n          " "\n          9/10\n          "
## [25] "\n          10/10\n          "
```

```
RevTitletv4 <-scrape(session5) %>%
    html_nodes('.parent a') %>%
    html_text()
RevTitletv4
```

```
## [1] "Chernobyl"
```

```
TextRevtv4 <-scrape(session5) %>%
    html_nodes('.collapsable div.text') %>%
    html_text()
TextRevtv4
```

```
##  [1] "My husband grew up near Kyiv and his father drove one of the buses that evacuated the civilians
##  [2] "In terms of series expression, it moves on in a style that we can call \"documentary-drama\". I
##  [3] "An infamous world event captured in a riveting series. Chernobyl provides so much depth and wi
##  [4] "This TV Show is a life lesson. They could make a Documentary with a narrator, you'll know the
##  [5] "I was born in Ukraine, 1971. Parents - Soviet \"intelligentsia\": mother was a high school tea
##  [6] "Sometimes, you watch something, and it lives with you, when it's something that has happened,
```

```
##  [7] "This is an HBO mini-series about the 1986 nuclear power plant disaster. A series of human error
##  [8] "'Chernobyl' is scarier than most horror movies in that it is a dramatization of actual, real-li
##  [9] "Im ukrainian, born in 1988 and still live here. I want to give the authors of this show a big
## [10] "That was an amazing cinematic experience that one can relish in the small screen, a true master
## [11] "This is a heartbreaking story based on true events and people's memories of the 1986 nuclear di
## [12] "If there is anything wrong with this, I'm unaware of it, though I will say the final two episod
## [13] "Chernobyl is about the brave men and women who had to navigate the ridiculously scary nuclear a
## [14] "The first episode somehow surpassed my expectations for what this TV Show would do with the bi
## [15] "I grew up in soviet Russia, born in 1977 and this is incredibly accurate. Starting with the exp
## [16] "Captivating from the first moment, Chernobyl's Cherenkov radiation lights up the screen.The cra
## [17] "Oh God. It's stunning. I'm impressed. Good job HBO. Getting from Russia."
## [18] "As my mother tells it, the weather was quite nice, the sky was clear without any sign of clouds
## [19] "I don't like writing long reviews, but let me just say that this show is one of the best. From
## [20] "Only two episodes in, and I'm totally hooked. Hard to believe that these things actually took p
## [21] "After seeing the first episode of Chernobyl I feel contaminated by radiation. Stating, acting,
## [22] "In April 1986, it happened two events that I will never forget: my daughter was born and USSR
## [23] "Hi. I'm from Kiev, Ukraine. I was born in 1983 and I was 2 and a half years when the Chernobyl
```

```r
lengths4 <- c(
  ReviewerName = length(ReviewersNametv4),
  ReviewDate = length(ReviewsDatetv4),
  UserRating = length(UserRatingtv4),
  Title = length(RevTitletv4),
  Comment = length(TextRevtv4)
)

max_length4 <- max(lengths4)

ReviewersNametv4 <- c(ReviewersNametv4, rep(NA, max_length - length(ReviewersNametv4)))
ReviewsDatetv4 <- c(ReviewsDatetv4, rep(NA, max_length - length(ReviewsDatetv4)))
UserRatingtv4 <- c(UserRatingtv4, rep(NA, max_length - length(UserRatingtv4)))
RevTitletv4 <- c(RevTitletv4, rep(NA, max_length - length(RevTitletv4)))
TextRevtv4 <- c(TextRevtv4, rep(NA, max_length - length(TextRevtv4)))

tv_show4 <- data.frame(
  ReviewerName = ReviewersNametv4,
  ReviewDate = ReviewsDatetv4,
  UserRating = UserRatingtv4,
  Title = RevTitletv4,
  Comment = TextRevtv4
)




#fifth
fifthrevlink <- "https://www.imdb.com/title/tt2395695/reviews?ref_=tt_urv"
session5 <- bow(fifthrevlink, user_agent= "Educational Purposes")
session5
```

```
## <polite session> https://www.imdb.com/title/tt2395695/reviews?ref_=tt_urv
##     User-agent: Educational Purposes
##     robots.txt: 34 rules are defined for 2 bots
##    Crawl delay: 5 sec
##   The path is scrapable for this user-agent
```

```r
ReviewersNametv5 <- scrape(session5) %>%
    html_nodes('.display-name-link a') %>%
    html_text()
ReviewersNametv5
```

```
##  [1] "lavatch"             "willwri14"
##  [3] "brighterside11"      "rmax304823"
##  [5] "pvineet131"          "gannoncannon-944-122285"
##  [7] "gabkoost"            "FabledGentleman"
##  [9] "compcyantist"        "hauklanglo"
## [11] "jetmox-458-423688"   "melinazahalka"
## [13] "xscd"                "sckioskanderson"
## [15] "llltdesq"            "christian94"
## [17] "chatless46"          "Arnror_II"
## [19] "TheMarwood"          "mikecart1"
## [21] "ChandRath"           "Lewis01-228-458550"
## [23] "rzajac"              "kld0068"
## [25] "Defunct_mouse"
```

```r
ReviewsDatetv5 <- scrape(session5) %>%
    html_nodes('.review-date') %>%
    html_text()
ReviewsDatetv5
```

```
##  [1] "9 March 2014"       "10 March 2014"      "11 March 2014"
##  [4] "9 March 2014"       "9 March 2014"       "9 March 2014"
##  [7] "29 April 2014"      "7 June 2014"        "13 March 2014"
## [10] "13 March 2014"      "10 March 2014"      "11 March 2014"
## [13] "25 June 2014"       "10 March 2014"      "14 September 2014"
## [16] "10 February 2016"   "9 June 2014"        "30 March 2014"
## [19] "30 June 2014"       "3 September 2014"   "9 June 2014"
## [22] "18 May 2014"        "5 May 2014"         "23 April 2014"
## [25] "11 May 2014"
```

```r
UserRatingtv5 <- scrape(session5) %>%
    html_nodes('.rating-other-user-rating') %>%
    html_text()
UserRatingtv5
```

```
##  [1] "\n          10/10\n        " "\n          10/10\n        "
##  [3] "\n          10/10\n        " "\n          10/10\n        "
##  [5] "\n          10/10\n        " "\n          10/10\n        "
##  [7] "\n          10/10\n        " "\n          10/10\n        "
##  [9] "\n          10/10\n        " "\n          9/10\n        "
## [11] "\n          10/10\n        " "\n          10/10\n        "
## [13] "\n          10/10\n        " "\n          10/10\n        "
## [15] "\n          10/10\n        " "\n          10/10\n        "
## [17] "\n          10/10\n        " "\n          10/10\n        "
## [19] "\n          9/10\n        " "\n          10/10\n        "
## [21] "\n          10/10\n        " "\n          10/10\n        "
## [23] "\n          9/10\n        " "\n          10/10\n        "
## [25] "\n          10/10\n        "
```

```r
RevTitletv5 <-scrape(session5) %>%
    html_nodes('.parent a') %>%
    html_text()
```

```
RevTitletv5
```

```
## [1] "Cosmos: A Spacetime Odyssey"
```

```
TextRevtv5 <-scrape(session5) %>%
    html_nodes('.collapsable div.text') %>%
    html_text()
TextRevtv5
```

```
##  [1] "Absolutely blown away by Neil deGrasse Tyson's hosting of Cosmos. You can tell he enjoys tellin
##  [2] "Neil really takes the baton from Carl and does a fine job with the show. With Ann co-producing
##  [3] "i have been watching and loving films and TV shows for years now, and loving the study of the u
##  [4] "What can i say about this remake of an already epic TV/Documentary show? If it is not enough th
##  [5] "After seeing all 13 episodes of this, i can hands down say, that this is the best science progr
##  [6] "Neil deGrasse Tyson is a marvelous astrophysicist who really loves his field and you can see th
##  [7] "Disclaimer: This review is based exclusively on the first episode.First things first. Audience
##  [8] "Thank You Seth MacFarlane for having the foresight for bringing this back with your own money.
##  [9] "This was so much fun--and it is so amazing that Carl Sagan's wife, Seth McFarlane, and Tyson De
## [10] "We don't have satellite or cable TV, so we watched the over-the-air broadcasts of Cosmos, and
## [11] "This was an absolute blast to watch! I was excited from the moment that I found out that Cosmo
## [12] "This series is stupendous and even perhaps better than the original it pays homage to with a la
## [13] "I've watched this series from the beginning. Meaning Carl Sagan. And I've gotta tell you, Dr. S
## [14] "This is a program made for (and rightly so) children, and should be rated as such. Production:
## [15] "The always watchable Neil deGrasse Tyson hosts this 13 part series on all things science, as he
## [16] "\"Cosmos: A SpaceTime Odyssey\" is those rare Documentary that might appear twice in a human li
## [17] "Fox needs to make sure that Cosmos is on next season and for many more. It is the only show on
## [18] "It's the rare person who has the temperament to truly \"do\" science. It's a lot of hard work,
## [19] "Many of the concepts were a nice refresher of the astronomy course that I took in college. I pa
## [20] "Being a long time lurker, I decided to register with the site in order to review Cosmos: A Spac
```

```
lengths5 <- c(
  ReviewerName = length(ReviewersNametv5),
  ReviewDate = length(ReviewsDatetv5),
  UserRating = length(UserRatingtv5),
  Title = length(RevTitletv5),
  Comment = length(TextRevtv5)
)

max_length5 <- max(lengths5)

ReviewersNametv5 <- c(ReviewersNametv5, rep(NA, max_length - length(ReviewersNametv5)))
ReviewsDatetv5 <- c(ReviewsDatetv5, rep(NA, max_length - length(ReviewsDatetv5)))
UserRatingtv5 <- c(UserRatingtv5, rep(NA, max_length - length(UserRatingtv5)))
RevTitletv5 <- c(RevTitletv5, rep(NA, max_length - length(RevTitletv5)))
TextRevtv5 <- c(TextRevtv5, rep(NA, max_length - length(TextRevtv5)))

tv_show5 <- data.frame(
  ReviewerName = ReviewersNametv5,
  ReviewDate = ReviewsDatetv5,
  UserRating = UserRatingtv5,
  Title = RevTitletv5,
  Comment = TextRevtv5
)
```

#3.

```r
library(polite)
library(rvest)
library(httr)
library(dplyr)
library(ggplot2)

# Function to extract release year from IMDb URL
get_release_year <- function(url) {
  parts <- strsplit(url, "/")[[1]]
  index <- which(parts == "title")
  if (length(index) > 0 && (index + 1) <= length(parts)) {
    return(as.integer(parts[index + 1]))
  } else {
    return(NA)
  }
}

# Extract release year for each TV show
release_years <- sapply(c(firstrevlink, secondrevlink, thirdrevlink, fourthrevlink, fifthrevlink), get_
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```
## Warning in FUN(X[[i]], ...): NAs introduced by coercion
```

```r
# Combine TV show data frames
all_tv_shows <- rbind(tv_show1, tv_show2, tv_show3, tv_show4, tv_show5)

# Add release year to the combined data frame
all_tv_shows$ReleaseYear <- rep(release_years, each = nrow(all_tv_shows) / length(release_years))

# Convert ReleaseYear to a factor for plotting
all_tv_shows$ReleaseYear <- as.factor(all_tv_shows$ReleaseYear)

# Create a time series graph
ggplot(all_tv_shows, aes(x = ReleaseYear)) +
  geom_bar() +
  labs(title = "TV Shows Released Over Time", x = "Release Year", y = "Number of TV Shows") +
  theme_minimal()
```
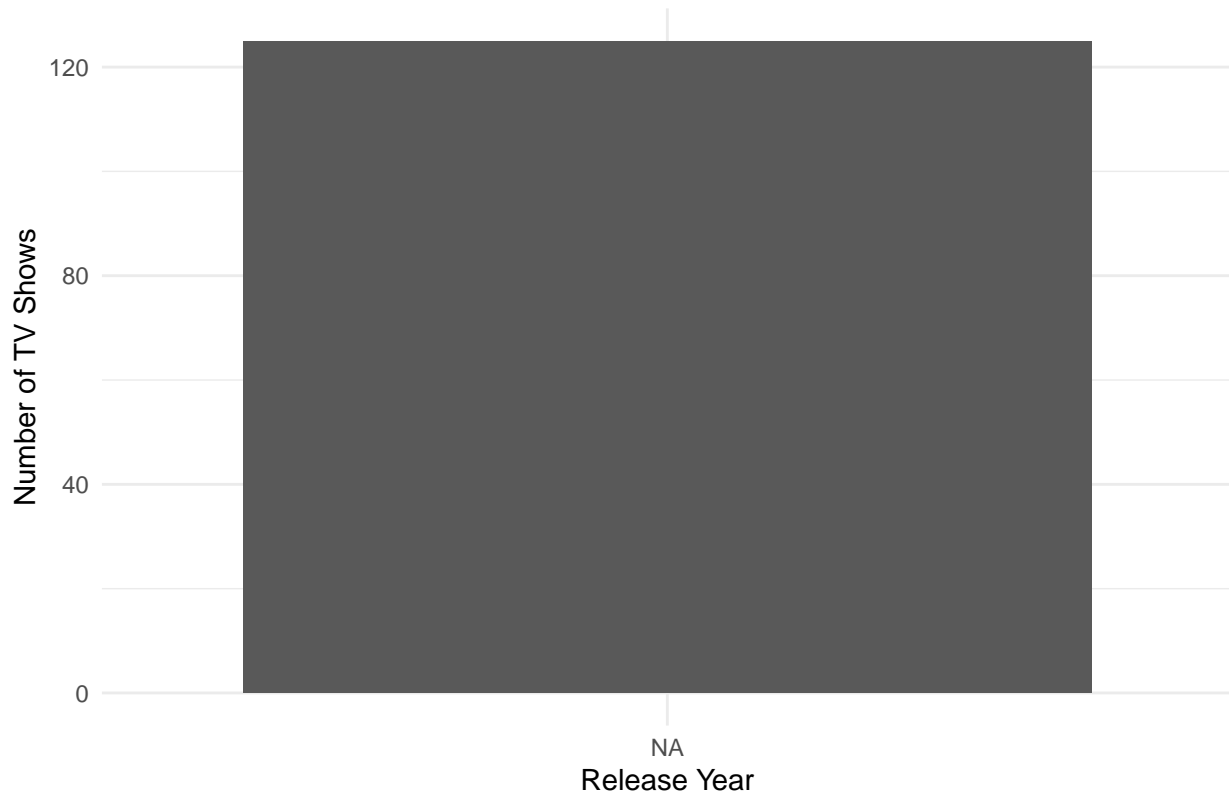
TV Shows Released Over Time

#4.

```r
library(polite)
library(rvest)
library(httr)
library(dplyr)
#1stproduct

amazonlink <- "https://www.amazon.com/Soundcore-Cancelling-Headphones-Wireless-Bluetooth/dp/B07NM3RSRQ/:
session6 <- bow(amazonlink, user_agent= "Educational Purposes")


price <-scrape(session6) %>%
    html_nodes('.reinventPricePriceToPayMargin span.a-price-whole') %>%
    html_text()
```

```
## No encoding supplied: defaulting to UTF-8.
```

```r
price
```

```
## character(0)
```

```r
description <-scrape(session6) %>%
    html_nodes('div.a-spacing-medium.a-spacing-top-small') %>%
    html_text()
description
```

```
## character(0)
```

```r
prodrating <-scrape(session6) %>%
    html_nodes('#acrPopover') %>%
    html_text()
prodrating
```

```
## character(0)
```

```r
prodreview <- scrape(session6) %>%
    html_nodes('#acrPopover') %>%
    html_text()
prodreview
```

```
## character(0)
```

```r
#2rdproduct
```

```r
amazonlink2nd <- "https://www.amazon.com/dp/B09S5MJLF6/ref=sspa_dk_detail_2?psc=1&pd_rd_i=B09S5MJLF6&pd_

session7 <- bow(amazonlink2nd, user_agent= "Educational Purposes")

price2 <-scrape(session7) %>%
    html_nodes('#corePriceDisplay_desktop_feature_div .a-price-whole') %>%
    html_text()
```

```
## No encoding supplied: defaulting to UTF-8.
```

```r
price2
```

```
## character(0)
```

```r
description2 <-scrape(session7) %>%
    html_nodes('#feature-bullets .a-list-item') %>%
    html_text()
description2
```

```
## character(0)
```

```r
prodrating2 <-scrape(session7) %>%
    html_nodes('#acrPopover') %>%
    html_text()
prodrating2
```

```
## character(0)
```

```r
prodreview2 <- scrape(session7) %>%
    html_nodes('#acrPopover') %>%
    html_text()
prodreview2
```

```
## character(0)
```

```r
#3rdproduct
```