

RWorksheet#5_group9

2023-11-30

```
#JIRRAINE OCTAVIANO  
#BARBIE JOY OBAS  
#MEIGELYN JOY MATIAS  
#ANDREA PATRICE PINEDA
```

Web Scraping IMDB TV Shows

```
install.packages("tidyverse")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
install.packages("dplyr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
install.packages("rvest")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
install.packages("httr")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
install.packages("polite")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
install.packages("ggplot2")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
install.packages("kableExtra")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'  
## (as 'lib' is unspecified)
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
```

```
## v dplyr      1.1.4      v readr      2.1.4  
## v forcats    1.0.0      v stringr    1.5.1  
## v ggplot2    3.4.4      v tibble     3.2.1  
## v lubridate  1.9.3      v tidyr      1.3.0  
## v purrr      1.0.2
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```

## x dplyr::filter() masks stats::filter()
## x dplyr::lag() masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(dplyr)
library(rvest)

##
## Attaching package: 'rvest'
##
## The following object is masked from 'package:readr':
##
##     guess_encoding

library(httr)
library(polite)
library(ggplot2)
library(kableExtra)

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

polite::use_manners(save_as = 'polite_scrape.R')

## v Setting active project to '/cloud/project'

1. Each group needs to extract the top 50 tv shows in Imdb.com. It will include the rank, the title of the
   tv show, tv rating, the number of people who voted, the number of episodes, the year it was released.

#creating objects for the dataset
rank <- character(50)
title <- character(50)
peoples_vote <- character(50)
episodes <- character(50)
year <- character(50)

url <- "https://www.imdb.com/chart/toptv/?ref_=nv_tv_250"
session <- bow(url, user_agent= "Educational Purposes")
session

## <polite session> https://www.imdb.com/chart/toptv/?ref_=nv_tv_250
##     User-agent: Educational Purposes
##     robots.txt: 35 rules are defined for 3 bots
##     Crawl delay: 5 sec
##     The path is scrapable for this user-agent

#Ranks
rank <- scrape(session) %>%
  html_nodes('h3.ipc-title__text') %>%
  html_text()
rank[1:51]

## [1] "IMDb Charts"
## [2] "1. Breaking Bad"
## [3] "2. Planet Earth II"

```

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

```
#Titles
```

```
titles <- scrape(session) %>%
  html_nodes("h3.ipc-title__text") %>%
  html_text()
titles50 <- titles[2:51]
```

```

#Ratings
ratings <- scrape(session) %>%
  html_nodes("span.ipc-rating-star.ipc-rating-star--base.ipc-rating-star--imdb.ratingGroup--imdb-rating")
  html_text()

cleaned_ratings <- substr(ratings, 1, 3)
ratings50 <- cleaned_ratings[1:50]

#Peoples vote
people_vote <- scrape(session) %>%
  html_nodes("span.ipc-rating-star--voteCount") %>%
  html_text()

people_vote50 <- people_vote[1:50]
cleaned_vote <- gsub("\\(|\\)", "", people_vote50)

#Episodes
episodes <- scrape(session) %>%
  html_nodes("span.sc-43986a27-8:nth-of-type(2)") %>%
  html_text()
episodes50 <- episodes[1:50]

#Year
year <- scrape(session) %>%
  html_nodes("span.sc-43986a27-8:nth-of-type(1)") %>%
  html_text()
year50 <- year[1:50]

#Dataframe
TopTv_df <- data.frame(
  Ranks = rank[1:50], # Correct variable name
  Titles = titles50,
  Ratings = ratings50,
  People_Votes = cleaned_vote,
  Episodes = episodes50,
  Year = year50
)
TopTv_df

```

```

##                               Ranks
## 1                               IMDb Charts
## 2                             1. Breaking Bad
## 3                             2. Planet Earth II
## 4                             3. Planet Earth
## 5                             4. Band of Brothers
## 6                             5. Chernobyl
## 7                             6. The Wire
## 8                             7. Avatar: The Last Airbender
## 9                             8. Blue Planet II
## 10                            9. The Sopranos
## 11                            10. Cosmos: A Spacetime Odyssey
## 12                            11. Cosmos
## 13                            12. Our Planet

```

## 14	13. Game of Thrones			
## 15	14. The World at War			
## 16	15. Bluey			
## 17	16. Rick and Morty			
## 18	17. Fullmetal Alchemist: Brotherhood			
## 19	18. The Last Dance			
## 20	19. Life			
## 21	20. The Twilight Zone			
## 22	21. Sherlock			
## 23	22. The Vietnam War			
## 24	23. Batman: The Animated Series			
## 25	24. Attack on Titan			
## 26	25. Scam 1992: The Harshad Mehta Story			
## 27	26. The Office			
## 28	27. Arcane			
## 29	28. The Blue Planet			
## 30	29. Better Call Saul			
## 31	30. Human Planet			
## 32	31. Firefly			
## 33	32. Frozen Planet			
## 34	33. Clarkson's Farm			
## 35	34. Death Note			
## 36	35. Only Fools and Horses			
## 37	36. Hunter x Hunter			
## 38	37. The Civil War			
## 39	38. True Detective			
## 40	39. Seinfeld			
## 41	40. The Beatles: Get Back			
## 42	41. The Decalogue			
## 43	42. Persona			
## 44	43. Fargo			
## 45	44. Cowboy Bebop			
## 46	45. Gravity Falls			
## 47	46. Nathan for You			
## 48	47. Last Week Tonight with John Oliver			
## 49	48. When They See Us			
## 50	49. Succession			
##	Titles	Ratings	People_Votes	Episodes
## 1	1. Breaking Bad	9.5	2.1M	62 eps
## 2	2. Planet Earth II	9.5	155K	6 eps
## 3	3. Planet Earth	9.4	218K	11 eps
## 4	4. Band of Brothers	9.4	509K	10 eps
## 5	5. Chernobyl	9.3	837K	5 eps
## 6	6. The Wire	9.3	368K	60 eps
## 7	7. Avatar: The Last Airbender	9.3	350K	62 eps
## 8	8. Blue Planet II	9.3	45K	7 eps
## 9	9. The Sopranos	9.2	451K	86 eps
## 10	10. Cosmos: A Spacetime Odyssey	9.3	127K	13 eps
## 11	11. Cosmos	9.3	43K	13 eps
## 12	12. Our Planet	9.3	49K	12 eps
## 13	13. Game of Thrones	9.2	2.2M	73 eps
## 14	14. The World at War	9.2	28K	26 eps
## 15	15. Bluey	9.4	22K	171 eps
## 16	16. Rick and Morty	9.1	584K	74 eps

## 17	17. Fullmetal Alchemist: Brotherhood	9.1	192K	68 eps
## 18	18. The Last Dance	9.1	146K	10 eps
## 19	19. Life	9.1	42K	11 eps
## 20	20. The Twilight Zone	9.1	91K	156 eps
## 21	21. Sherlock	9.1	978K	15 eps
## 22	22. The Vietnam War	9.1	27K	10 eps
## 23	23. Batman: The Animated Series	9.0	114K	85 eps
## 24	24. Attack on Titan	9.1	481K	98 eps
## 25	25. Scam 1992: The Harshad Mehta Story	9.3	154K	10 eps
## 26	26. The Office	9.0	684K	188 eps
## 27	27. Arcane	9.0	254K	10 eps
## 28	28. The Blue Planet	9.0	42K	8 eps
## 29	29. Better Call Saul	9.0	622K	63 eps
## 30	30. Human Planet	9.0	28K	8 eps
## 31	31. Firefly	9.0	278K	14 eps
## 32	32. Frozen Planet	9.0	33K	10 eps
## 33	33. Clarkson's Farm	9.0	53K	18 eps
## 34	34. Death Note	8.9	366K	37 eps
## 35	35. Only Fools and Horses	9.0	56K	64 eps
## 36	36. Hunter x Hunter	9.0	127K	148 eps
## 37	37. The Civil War	9.0	18K	9 eps
## 38	38. True Detective	8.9	619K	30 eps
## 39	39. Seinfeld	8.9	344K	173 eps
## 40	40. The Beatles: Get Back	9.0	27K	3 eps
## 41	41. The Decalogue	8.9	27K	10 eps
## 42	42. Persona	9.0	46K	20 eps
## 43	43. Fargo	8.9	400K	51 eps
## 44	44. Cowboy Bebop	8.9	134K	26 eps
## 45	45. Gravity Falls	8.9	129K	41 eps
## 46	46. Nathan for You	8.9	37K	32 eps
## 47	47. Last Week Tonight with John Oliver	8.9	94K	342 eps
## 48	48. When They See Us	8.9	134K	4 eps
## 49	49. Succession	8.9	249K	39 eps
## 50	50. Apocalypse: The Second World War	9.0	14K	6 eps
##	Year			
## 1	2008-2013			
## 2	2016			
## 3	2006			
## 4	2001			
## 5	2019			
## 6	2002-2008			
## 7	2005-2008			
## 8	2017			
## 9	1999-2007			
## 10	2014			
## 11	1980			
## 12	2019-2023			
## 13	2011-2019			
## 14	1973-1974			
## 15	2018-			
## 16	2013-			
## 17	2009-2010			
## 18	2020			
## 19	2009			

```
## 20 1959-1964
## 21 2010-2017
## 22      2017
## 23 1992-1995
## 24 2013-2023
## 25      2020
## 26 2005-2013
## 27      2021-
## 28      2001
## 29 2015-2022
## 30      2011
## 31 2002-2003
## 32 2011-2012
## 33      2021-
## 34 2006-2007
## 35 1981-2003
## 36 2011-2014
## 37      1990
## 38      2014-
## 39 1989-1998
## 40      2021
## 41 1989-1990
## 42      2018-
## 43 2014-2024
## 44 1998-1999
## 45 2012-2016
## 46 2013-2017
## 47      2014-
## 48      2019
## 49 2018-2023
## 50      2009
```

2. From the 50 tv shows, select at least 5 tv shows to scrape the user reviews that will include the reviewer's name, date of reviewed, user rating, title of the review, and text reviews.

1st Link

```
pad_with_na <- function(vec, target_length) {
  if (length(vec) < target_length) {
    return(c(vec, rep(NA, target_length - length(vec))))
  } else {
    return(vec)
  }
}

OP_Link <- "https://www.imdb.com/title/tt9253866/reviews?ref_=tt_urv"
session2 <- bow(OP_Link, user_agent = "Educational Purposes")

OP_Title <- scrape(session2) %>%
  html_nodes("a[itemprop='url']") %>%
  html_text()

OP_RevName <- scrape(session2) %>%
```

```

html_nodes(".display-name-link a") %>%
html_text()

OP_RevDate <- scrape(session2) %>%
html_nodes("span.review-date") %>%
html_text()

OP_UserRating <- scrape(session2) %>%
html_nodes("span.rating-other-user-rating") %>%
html_text() %>% str_trim()

OP_TitleRev <- scrape(session2) %>%
html_nodes("a.title") %>%
html_text() %>%
gsub("\n ", "", .)

OP_TextRev <- scrape(session2) %>%
html_nodes("div.text") %>%
html_text()

max_length <- max(
  length(OP_Title),
  length(OP_RevName),
  length(OP_RevDate),
  length(OP_UserRating),
  length(OP_TitleRev),
  length(OP_TextRev)
)

OP_Title <- pad_with_na(OP_Title, max_length)
OP_RevName <- pad_with_na(OP_RevName, max_length)
OP_RevDate <- pad_with_na(OP_RevDate, max_length)
OP_UserRating <- pad_with_na(OP_UserRating, max_length)
OP_TitleRev <- pad_with_na(OP_TitleRev, max_length)
OP_TextRev <- pad_with_na(OP_TextRev, max_length)

OP_Df <- data.frame(
  Tv_Shows = OP_Title,
  Reviewer_Name = OP_RevName,
  Reviewer_Date = OP_RevDate,
  Reviewer_Rating = OP_UserRating,
  Title_Review = OP_TitleRev,
  Text_Review = OP_TextRev
)

head(OP_Df)

```

```

##      Tv_Shows      Reviewer_Name Reviewer_Date Reviewer_Rating
## 1 Our Planet      tester-84146 11 April 2019      10/10
## 2      <NA>      AnthonyBehan   5 April 2019      10/10
## 3      <NA>      janoffpeter   8 April 2019      10/10
## 4      <NA> firballblaze-69689   9 April 2019      10/10

```



```
## 5      <NA> TheLittleSongbird    2 May 2019      10/10
## 6      <NA> FrenchEddieFelson   9 April 2019     9/10
##              Title_Review
## 1              Wow\n
## 2 Done before, But not like this\n
## 3   Simply jaw dropping wonders!\n
## 4              Fantastic!\n
## 5              Shock and awe\n
## 6   Make our planet great again\n
##
## 1
## 2
## 3
## 4
## 5 Do enjoy watching film and television on Netflix, but when it comes to the Netflix originals the o
## 6
```

2nd Link

```
pad_with_na <- function(vec, target_length) {
  if (length(vec) < target_length) {
    return(c(vec, rep(NA, target_length - length(vec))))
  } else {
    return(vec)
  }
}

GT_Link <- "https://www.imdb.com/title/tt1865718/reviews?ref_=tt_urv"
session2 <- bow(GT_Link, user_agent = "Educational Purposes")

GT_Title <- scrape(session2) %>%
  html_nodes(".parent a") %>%
  html_text()

GT_RevName <- scrape(session2) %>%
  html_nodes(".display-name-link") %>%
  html_text()

GT_RevDate <- scrape(session2) %>%
  html_nodes("span.review-date") %>%
  html_text()

GT_UserRating <- scrape(session2) %>%
  html_nodes("span.rating-other-user-rating") %>%
  html_text() %>% str_trim()

GT_TitleRev <- scrape(session2) %>%
  html_nodes("a.title") %>%
  html_text() %>%
  gsub("\n ", "", .)

GT_TextRev <- scrape(session2) %>%
  html_nodes("div.text") %>%
```

```

html_text()

max_length <- max(
  length(GT_Title),
  length(GT_RevName),
  length(GT_RevDate),
  length(GT_UserRating),
  length(GT_TitleRev),
  length(GT_TextRev)
)

GT_Title <- pad_with_na(GT_Title, max_length)
GT_RevName <- pad_with_na(GT_RevName, max_length)
GT_RevDate <- pad_with_na(GT_RevDate, max_length)
GT_UserRating <- pad_with_na(GT_UserRating, max_length)
GT_TitleRev <- pad_with_na(GT_TitleRev, max_length)
GT_TextRev <- pad_with_na(GT_TextRev, max_length)

GT_Df <- data.frame(
  Tv_Shows = GT_Title,
  Reviewer_Name = GT_RevName,
  Reviewer_Date = GT_RevDate,
  Reviewer_Rating = GT_UserRating,
  Title_Review = GT_TitleRev,
  Text_Review = GT_TextRev
)

head(GT_Df)

```

```

##      Tv_Shows      Reviewer_Name      Reviewer_Date Reviewer_Rating
## 1 Gravity Falls      planktonrules      27 July 2014           9/10
## 2      <NA> OnjiMooteDaMarle      22 July 2022           8/10
## 3      <NA>      doomedmac 14 December 2019           9/10
## 4      <NA>      Animany94 18 December 2019          10/10
## 5      <NA> newkidontheblock  1 February 2013           9/10
## 6      <NA>      berg-74532 25 January 2020          10/10
##
##                                     Title_Review
## 1                      Amazingly strange...and funny.\n
## 2                      The True Multiverse of Madness\n
## 3                      One of my all-time favorite shows\n
## 4 Expertly written with so much care put to creating interesting stories.\n
## 5                      Gravity Falls: a Diamond in the Disney Channel Rough\n
## 6                      I want MORE!!\n
##
## 1
## 2
## 3
## 4
## 5 There are few things I hate more on television than the Disney Channel. This factory of mediocrity
## 6

```

3rd Link

```

pad_with_na <- function(vec, target_length) {
  if (length(vec) < target_length) {
    return(c(vec, rep(NA, target_length - length(vec))))
  } else {
    return(vec)
  }
}

DN_Link <- "https://www.imdb.com/title/tt0877057/reviews?ref_=tt_urv"
session2 <- bow(DN_Link, user_agent = "Educational Purposes")

DN_Title <- scrape(session2) %>%
  html_nodes(".parent a") %>%
  html_text()

DN_RevName <- scrape(session2) %>%
  html_nodes(".display-name-link") %>%
  html_text()
DN_RevDate <- scrape(session2) %>%
  html_nodes("span.review-date") %>%
  html_text()

DN_UserRating <- scrape(session2) %>%
  html_nodes("span.rating-other-user-rating") %>%
  html_text() %>% str_trim()

DN_TitleRev <- scrape(session2) %>%
  html_nodes("a.title") %>%
  html_text() %>%
  gsub("\n ", "", .)

DN_TextRev <- scrape(session2) %>%
  html_nodes("div.text") %>%
  html_text()

max_length <- max(
  length(DN_Title),
  length(DN_RevName),
  length(DN_RevDate),
  length(DN_UserRating),
  length(DN_TitleRev),
  length(DN_TextRev)
)

DN_Title <- pad_with_na(DN_Title, max_length)
DN_RevName <- pad_with_na(DN_RevName, max_length)
DN_RevDate <- pad_with_na(DN_RevDate, max_length)
DN_UserRating <- pad_with_na(DN_UserRating, max_length)
DN_TitleRev <- pad_with_na(DN_TitleRev, max_length)
DN_TextRev <- pad_with_na(DN_TextRev, max_length)

```

```
DN_Df <- data.frame(
  Tv_Shows = DN_Title,
  Reviewer_Name = DN_RevName,
  Reviewer_Date = DN_RevDate,
  Reviewer_Rating = DN_UserRating,
  Title_Review = DN_TitleRev,
  Text_Review = DN_TextRev
)
```

```
head(DN_Df)
```

```
##      Tv_Shows  Reviewer_Name    Reviewer_Date Reviewer_Rating
## 1 Death Note  NicolasTheWolf 14 September 2021          9/10
## 2      <NA> loncarkristina  6 September 2019          9/10
## 3      <NA> premnigamkarpnk    3 June 2019          9/10
## 4      <NA> eirenightsade     6 January 2007         10/10
## 5      <NA>      dj_vlado     8 August 2019          9/10
## 6      <NA>      kojisuzuki  30 October 2006         10/10
```

```
##                                     Title_Review
## 1                                     Genius\n
## 2                                I'm glad Death Note was my first anime.\n
## 3 Brilliant first half, and the rest of it was just okay.\n
## 4                                Clever, Slick, and Suspenseful\n
## 5 Masterpiece but latter half ruined the series slightly\n
## 6                                let them eat cake!\n
```

```
## 1
## 2 The first time I heard of Death Note was from a youtube channel Hello Future Me and his storytelling
## 3
## 4
## 5
## 6
```

4th Link

```
pad_with_na <- function(vec, target_length) {
  if (length(vec) < target_length) {
    return(c(vec, rep(NA, target_length - length(vec))))
  } else {
    return(vec)
  }
}
```

```
A_Link <- "https://www.imdb.com/title/tt11126994/reviews?ref_=tt_urv"
session2 <- bow(A_Link, user_agent = "Educational Purposes")
```

```
A_Title <- scrape(session2) %>%
  html_nodes(".parent a") %>%
  html_text()
```

```
A_RevName <- scrape(session2) %>%
  html_nodes(".display-name-link") %>%
  html_text()
```

```

A_RevDate <- scrape(session2) %>%
  html_nodes("span.review-date") %>%
  html_text()

A_UserRating <- scrape(session2) %>%
  html_nodes("span.rating-other-user-rating") %>%
  html_text() %>% str_trim()

A_TitleRev <- scrape(session2) %>%
  html_nodes("a.title") %>%
  html_text() %>%
  gsub("\n ", "", .)

A_TextRev <- scrape(session2) %>%
  html_nodes("div.text") %>%
  html_text()

max_length <- max(
  length(A_Title),
  length(A_RevName),
  length(A_RevDate),
  length(A_UserRating),
  length(A_TitleRev),
  length(A_TextRev)
)

A_Title <- pad_with_na(A_Title, max_length)
A_RevName <- pad_with_na(A_RevName, max_length)
A_RevDate <- pad_with_na(A_RevDate, max_length)
A_UserRating <- pad_with_na(A_UserRating, max_length)
A_TitleRev <- pad_with_na(A_TitleRev, max_length)
A_TextRev <- pad_with_na(A_TextRev, max_length)

A_Df <- data.frame(
  Tv_Shows = A_Title,
  Reviewer_Name = A_RevName,
  Reviewer_Date = A_RevDate,
  Reviewer_Rating = A_UserRating,
  Title_Review = A_TitleRev,
  Text_Review = A_TextRev
)

head(A_Df)

```

##	Tv_Shows	Reviewer_Name	Reviewer_Date	Reviewer_Rating
## 1	Arcane	nancyldraper	22 November 2021	9/10
## 2	<NA>	stevebondi	10 November 2021	10/10
## 3	<NA>	Pairic	19 November 2021	9/10
## 4	<NA>	djurepower	7 November 2021	9/10
## 5	<NA>	Tactrix	7 November 2021	10/10
## 6	<NA>	garabedian123	8 November 2021	10/10

```
##                                     Title_Review
## 1                                     Visually stunning\n
## 2  Coolest animation style I have seen to date in any TV series or movie! :-)\n
## 3                                     Great Steampunk Adventure\n
## 4                                     Excellent, also for non league fans\n
## 5                                     This is ART!\n
## 6                                     LOL is a video game right?\n
##
## 1
## 2
## 3
## 4 I watched it with my parents who don't know anything about league's universe. I have been playing
## 5
## 6
```

5th Link

```
pad_with_na <- function(vec, target_length) {
  if (length(vec) < target_length) {
    return(c(vec, rep(NA, target_length - length(vec))))
  } else {
    return(vec)
  }
}

FP_Link <- "https://www.imdb.com/title/tt2092588/reviews?ref_=tt_urv"
session2 <- bow(FP_Link, user_agent = "Educational Purposes")

FP_Title <- scrape(session2) %>%
  html_nodes(".parent a") %>%
  html_text()

FP_RevName <- scrape(session2) %>%
  html_nodes(".display-name-link") %>%
  html_text()

FP_RevDate <- scrape(session2) %>%
  html_nodes("span.review-date") %>%
  html_text()

FP_UserRating <- scrape(session2) %>%
  html_nodes("span.rating-other-user-rating") %>%
  html_text() %>% str_trim()

FP_TitleRev <- scrape(session2) %>%
  html_nodes("a.title") %>%
  html_text() %>%
  gsub("\n ", "", .)

FP_TextRev <- scrape(session2) %>%
  html_nodes("div.text") %>%
  html_text()
```

```

max_length <- max(
  length(FP_Title),
  length(FP_RevName),
  length(FP_RevDate),
  length(FP_UserRating),
  length(FP_TitleRev),
  length(FP_TextRev)
)

FP_Title <- pad_with_na(FP_Title, max_length)
FP_RevName <- pad_with_na(FP_RevName, max_length)
FP_RevDate <- pad_with_na(FP_RevDate, max_length)
FP_UserRating <- pad_with_na(FP_UserRating, max_length)
FP_TitleRev <- pad_with_na(FP_TitleRev, max_length)
FP_TextRev <- pad_with_na(FP_TextRev, max_length)

FP_Df <- data.frame(
  Tv_Shows = FP_Title,
  Reviewer_Name = FP_RevName,
  Reviewer_Date = FP_RevDate,
  Reviewer_Rating = FP_UserRating,
  Title_Review = FP_TitleRev,
  Text_Review = FP_TextRev
)

head(FP_Df)

```

```

##      Tv_Shows      Reviewer_Name      Reviewer_Date Reviewer_Rating
## 1 Frozen Planet      bob the moo 26 December 2011          10/10
## 2      <NA> roarz-198-320112 28 October 2011          10/10
## 3      <NA>      fikamugg 25 January 2013          10/10
## 4      <NA> meshed-2-285923 5 November 2011          10/10
## 5      <NA>      Bert45 12 April 2012           8/10
## 6      <NA> jboothmillard 21 November 2011          10/10
##                                     Title_Review
## 1      Amazingly filmed and engaging throughout\n
## 2      Superb, went far and beyond my expectations.\n
## 3                                     Simply astonishing!\n
## 4                                     Chilling!\n
## 5      Absolutely brilliant work\n
## 6      Frozen Planet\n
##
## 1 Looking spritely as ever David Attenborough returned to the BBC with yet another brilliant document
## 2
## 3
## 4
## 5
## 6

```

3. Create a time series graph for the tv shows released by year. Which year has the most number of tv shows released?

```

library(ggplot2)
library(dplyr)

```

```
ggplot(TopTv_df, aes(x = Year, fill = Year)) +
  geom_bar() +
  labs(title = "Number of TV Shows Released by Year",
       x = "Year",
       y = "Number of TV Shows") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text.x = element_text(angle = 44, hjust = 1))
```

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <93>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <93>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <93>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <93>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <93>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <93>
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014- ' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014- ' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014- ' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014- ' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014- ' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014- ' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014-2024' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014-2024' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014-2024' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014-2024' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014-2024' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014-2024' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014-2024' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2014-2024' in 'mbcsToSbcs': dot substituted for <80>
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2008-2013' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2008-2013' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2008-2013' in 'mbcsToSbcs': dot substituted for <93>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2008-2013' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2008-2013' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2008-2013' in 'mbcsToSbcs': dot substituted for <93>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2008-2013' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2008-2013' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2008-2013' in 'mbcsToSbcs': dot substituted for <93>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2009-2010' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2009-2010' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2009-2010' in 'mbcsToSbcs': dot substituted for <93>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2009-2010' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2009-2010' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2009-2010' in 'mbcsToSbcs': dot substituted for <93>

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1959-1964' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1973-1974' in 'mbcsToSbcs': dot substituted for <93>
```

[illegible]

[illegible]

[illegible]

[illegible]

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1989-1990' in 'mbscsToSbcs': dot substituted for <80>
```

[illegible]

[illegible]

[illegible]

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '1992-1995' in 'mbcsToSbc': dot substituted for <80>
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]


```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2006-2007' in 'mbsToSbs': dot substituted for <80>
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2009-2010' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2009-2010' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2009-2010' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2010-2017' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2010-2017' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2010-2017' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2010-2017' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2010-2017' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2010-2017' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2010-2017' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2010-2017' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2010-2017' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2010-2017' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2010-2017' in 'mbcsToSbcs': dot substituted for <80>
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013- ' in 'mbcsToSbcs': dot substituted for <93>
```

[illegible]

[illegible]

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2013-2017' in 'mbsToSbs': dot substituted for <80>
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2015-2022' in 'mbsToSbs': dot substituted for <80>
```

[illegible]

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <93>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <e2>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <80>  
  
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :  
## conversion failure on '2018- ' in 'mbcsToSbcs': dot substituted for <93>
```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2021- ' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2021- ' in 'mbcsToSbcs': dot substituted for <80>

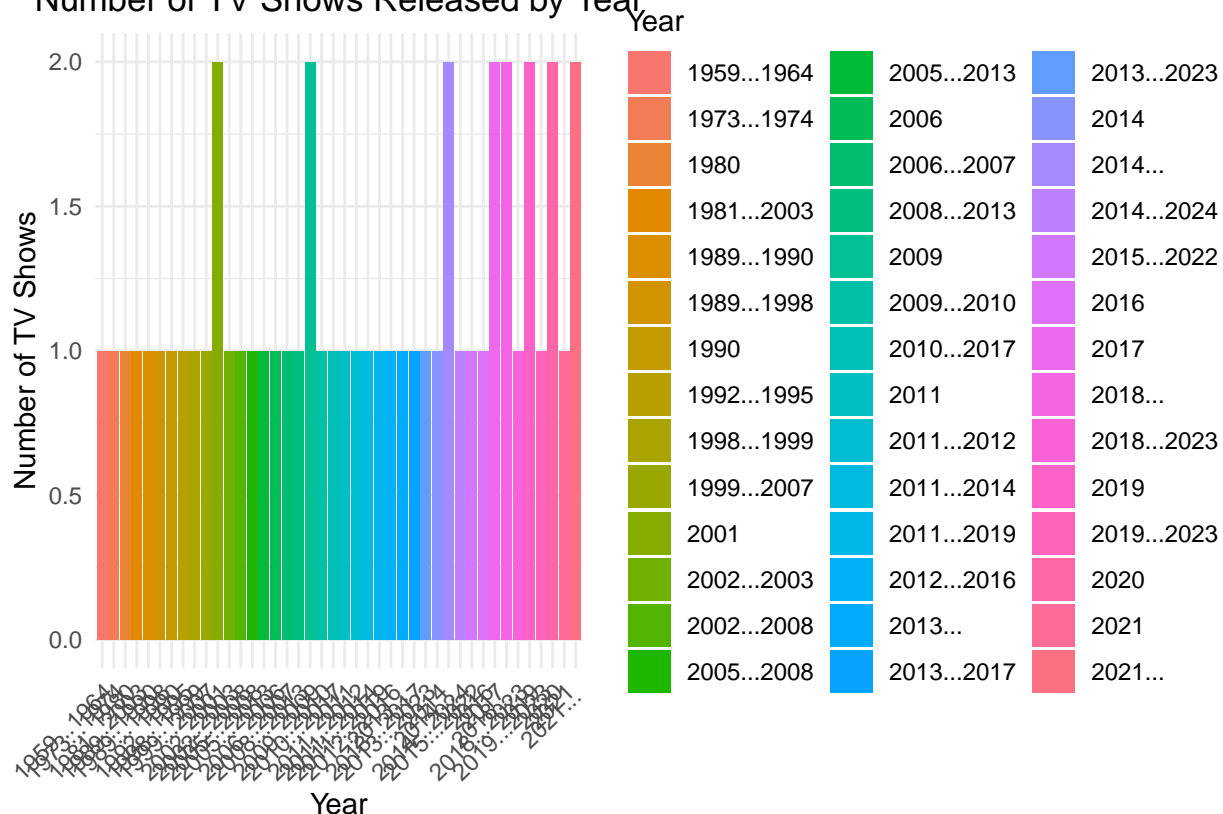
## Warning in grid.Call(C_textBounds, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2021- ' in 'mbcsToSbcs': dot substituted for <93>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2021- ' in 'mbcsToSbcs': dot substituted for <e2>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2021- ' in 'mbcsToSbcs': dot substituted for <80>

## Warning in grid.Call.graphics(C_text, as.graphicsAnnot(x$label), x$x, x$y, :
## conversion failure on '2021- ' in 'mbcsToSbcs': dot substituted for <93>
```

Number of TV Shows Released by Year



```
cat("The year with the most number of tv shows is", TopTv_df$Year[1])
```

```
## The year with the most number of tv shows is 2008-2013
```

Extracting Amazon Product Reviews

4. Select 3 products from Amazon of the same category. Extract the price, description, ratings and reviews of each product.

```
install.packages("rvest")
```

```
## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)
```

```

install.packages("dplyr")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)

install.packages("polite")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)

install.packages("tidyverse")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)

library(rvest)
library(dplyr)
library(polite)
library(tidyverse)

#First Item
Item1 <- "www.amazon.com/KOLIGHT-Cosmetic-Eyeline-Eyebrow-Foundation/dp/B083GQMJS3/ref=pd_bxgy_img_d_s

session1 <- html_session(Item1, user_agent = "Educational Purposes")

## Warning: `html_session()` was deprecated in rvest 1.0.0.
## i Please use `session()` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

price1 <- session1 %>%
  html_nodes('span.a-offscreen') %>%
  html_text() %>%
  trimws()
price1[1]

## [1] "$6.99"

description1 <- session1 %>%
  html_nodes('span.a-size-large.product-title-word-break') %>%
  html_text() %>%
  trimws()
description1

## [1] "KOLIGHT® Pack of 20pcs Cosmetic Eye Shadow Sponge Eyeliner Eyebrow Lip Nose Foundation Powder M

prodrating1 <- session1 %>%
  html_nodes('.a-icon-star .a-icon-alt') %>%
  html_text() %>%
  trimws()
prodrating1[1]

## [1] "4.4 out of 5 stars"

prodreview1 <- session1 %>%
  html_nodes('#acrCustomerReviewText') %>%
  html_text() %>%

```



```

    trimws()
prodreview1[1]

## [1] "8,838 ratings"
#SECOND ITEM
Item2 <- "www.amazon.com/dp/B0CBRSM8Z7/ref=sspa_dk_detail_0?psc=1&pd_rd_i=B0CBRSM8Z7&pd_rd_w=g6DUy&cont

session2 <- html_session(Item2, user_agent = "Educational Purposes")

price2 <- session2 %>%
  html_nodes('span.a-offscreen') %>%
  html_text() %>%
  trimws()
price2[1]

## [1] "$25.99"
description2 <- session2 %>%
  html_nodes('span.a-size-large.product-title-word-break') %>%
  html_text() %>%
  trimws()
description2

## [1] "Color Nymph Girls Makeup Kits For Teens With Pink Retro Train Case Included Portable Matte Shim

prodrating2 <- session2 %>%
  html_nodes('.a-icon-star .a-icon-alt') %>%
  html_text() %>%
  trimws()
prodrating2[1]

## [1] "4.6 out of 5 stars"
prodreview2 <- session2 %>%
  html_nodes('#acrCustomerReviewText') %>%
  html_text() %>%
  trimws()
prodreview2[1]

## [1] "232 ratings"

```

THIRD ITEM

```

Item3 <- "www.amazon.com/dp/B0BWYFYMZ7X/ref=sspa_dk_detail_8?psc=1&pf_rd_p=eb7c1ac5-7c51-4df5-ba34-ca810

session3 <- html_session(Item3, user_agent = "Educational Purposes")

price3 <- session3 %>%
  html_nodes('span.a-offscreen') %>%
  html_text() %>%
  trimws()
price3[1]

## [1] "$32.99"

```

```
description3 <- session3 %>%
  html_nodes('span.a-size-large.product-title-word-break') %>%
  html_text() %>%
  trimws()
description3
```

```
## [1] "MISS ROSE M 148 Colors Makeup Pallet,Professional Makeup Kit for Women Full Kit,All in One Make"
```

```
prodrating3 <- session3 %>%
  html_nodes('.a-icon-star .a-icon-alt') %>%
  html_text() %>%
  trimws()
prodrating3[1]
```

```
## [1] "4.5 out of 5 stars"
```

```
prodreview3 <- session3 %>%
  html_nodes('#acrCustomerReviewText') %>%
  html_text() %>%
  trimws()
prodreview3[1]
```

```
## [1] "284 ratings"
```