

# RWorksheet#5

Andica, Benedicto, Cautivar

2024-11-06

Extracting IMDB. 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.

```
library(polite)

url <- 'https://www.imdb.com/chart/toptv/?ref_=nv_tv_250'

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

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

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

title_list <- scrape(session) %>%
  html_nodes('h3.ipc-title__text') %>%
  html_text

title_list_sub <- as.data.frame(title_list[2:26])

colnames(title_list_sub) <- "Ranks"
split_df <- strsplit(as.character(title_list_sub$Ranks), ".", fixed = TRUE)
split_df <- data.frame(do.call(rbind, split_df))

colnames(split_df) <- c("Ranks", "Title")
```

```

titleAndRank <- data.frame(split_df)

ratings <- scrape(session) %>%
  html_nodes('span.ipc-rating-star--rating') %>%
  html_text

ratingsDf <- data.frame(ratings)

numberOfPeopleVoted <- scrape(session) %>%
  html_nodes('span.ipc-rating-star--voteCount') %>%
  html_text

cleanedVotes <- gsub('[(\)]', '', numberOfPeopleVoted)

cleanedVotesDf <- data.frame(cleanedVotes)

numEpisodes <- scrape(session) %>%
  html_nodes('span.sc-5bc66c50-6.00dsw.cli-title-metadata-item:nth-of-type(2)') %>%
  html_text

numEpisodesDf <- data.frame(numEpisodes)

Year <- scrape(session) %>%
  html_nodes('span.sc-5bc66c50-6.00dsw.cli-title-metadata-item:nth-of-type(1)') %>%
  html_text

YearDf <- data.frame(Year)

topShows <- cbind(titleAndRank, ratingsDf, cleanedVotesDf, numEpisodesDf, YearDf)

topShows

```

##	Ranks	Title	ratings	cleanedVotes	numEpisodes
## 1	1	Breaking Bad	9.5	2.2M	62 eps
## 2	2	Planet Earth II	9.5	162K	6 eps
## 3	3	Planet Earth	9.4	224K	11 eps
## 4	4	Band of Brothers	9.4	545K	10 eps
## 5	5	Chernobyl	9.3	907K	5 eps
## 6	6	The Wire	9.3	391K	60 eps
## 7	7	Avatar: The Last Airbender	9.3	390K	62 eps
## 8	8	Blue Planet II	9.3	49K	7 eps
## 9	9	The Sopranos	9.2	498K	86 eps
## 10	10	Cosmos: A Spacetime Odyssey	9.2	132K	13 eps
## 11	11	Cosmos	9.3	46K	13 eps
## 12	12	Our Planet	9.2	54K	12 eps
## 13	13	Game of Thrones	9.2	2.4M	74 eps
## 14	14	Bluey	9.3	34K	194 eps
## 15	15	The World at War	9.2	31K	26 eps
## 16	16	Fullmetal Alchemist: Brotherhood	9.1	209K	68 eps
## 17	17	Rick and Morty	9.1	627K	78 eps
## 18	18	Life	9.1	44K	11 eps
## 19	19	The Last Dance	9.1	160K	10 eps
## 20	20	The Twilight Zone	9.0	97K	156 eps

## 21	21	The Vietnam War	9.1	30K	10 eps
## 22	22	Sherlock	9.1	1M	15 eps
## 23	23	Attack on Titan	9.1	561K	98 eps
## 24	24	Batman: The Animated Series	9.0	122K	85 eps
## 25	25	Arcane	9.0	302K	18 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	2018-				
## 15	1973-1974				
## 16	2009-2010				
## 17	2013-				
## 18	2009				
## 19	2020				
## 20	1959-1964				
## 21	2017				
## 22	2010-2017				
## 23	2013-2023				
## 24	1992-1995				
## 25	2021-2024				

1. PART II It will also include the number of user reviews and the number of critic reviews, as well as the popularity rating for each tv shows.

```

homePage <- 'https://www.imdb.com/chart/toptv/'
mainPage <- read_html(homePage)

links <- mainPage %>%
  html_nodes("a.ipc-title-link-wrapper") %>%
  html_attr("href")

showInfo <- lapply(links, function(link) {
  fullLink <- paste0("https://imdb.com", link)

  userRevLink <- read_html(fullLink)
  userRevPageLink <- userRevLink %>%
    html_nodes('a.isReview') %>%
    html_attr("href")

  criticRev <- userRevLink %>%
    html_nodes("span.score") %>%
    html_text()

  criticDf <- data.frame(Critic_Reviews = criticRev[2], stringsAsFactors = FALSE)

```

```

popularityRating <- userRevLink %>%
  html_nodes('[data-testid="hero-rating-bar__popularity__score"]') %>%
  html_text()

userRev <- read_html(paste0("https://imdb.com", userRevPageLink[1]))
userRevCount <- userRev %>%
  html_nodes('[data-testid="tturv-total-reviews"]') %>%
  html_text()

return(data.frame(User_Reviews = userRevCount, Critic = criticDf, Popularity_Rating = popularityRating))
})

showUrlDf <- do.call(rbind, showInfo)
showUrlDf

```

##	User_Reviews	Critic_Reviews	Popularity_Rating
## 1	5,091 reviews	175	20
## 2	5,091 reviews	175	20
## 3	158 reviews	6	1,121
## 4	158 reviews	6	1,121
## 5	111 reviews	10	2,011
## 6	111 reviews	10	2,011
## 7	1,056 reviews	34	171
## 8	1,056 reviews	34	171
## 9	3,533 reviews	88	173
## 10	3,533 reviews	88	173
## 11	787 reviews	77	108
## 12	787 reviews	77	108
## 13	998 reviews	57	373
## 14	998 reviews	57	373
## 15	53 reviews	9	4,415
## 16	53 reviews	9	4,415
## 17	963 reviews	93	33
## 18	963 reviews	93	33
## 19	205 reviews	12	1,499
## 20	205 reviews	12	1,499
## 21	80 reviews	8	3,866
## 22	80 reviews	8	3,866
## 23	245 reviews	15	2,765
## 24	245 reviews	15	2,765
## 25	5,899 reviews	368	14
## 26	5,899 reviews	368	14
## 27	367 reviews	4	411
## 28	367 reviews	4	411
## 29	126 reviews	5	2,627
## 30	126 reviews	5	2,627
## 31	466 reviews	16	508
## 32	466 reviews	16	508
## 33	911 reviews	94	137
## 34	911 reviews	94	137
## 35	12 reviews	9	3,455
## 36	12 reviews	9	3,455

```
## 37 541 reviews 28 1,521
## 38 541 reviews 28 1,521
## 39 213 reviews 85 354
## 40 213 reviews 85 354
## 41 175 reviews 13 2,022
## 42 175 reviews 13 2,022
## 43 1,096 reviews 121 172
## 44 1,096 reviews 121 172
## 45 2,360 reviews 64 60
## 46 2,360 reviews 64 60
## 47 219 reviews 25 527
## 48 219 reviews 25 527
## 49 1,952 reviews 50 15
## 50 1,952 reviews 50 15
```

```
allShows <- cbind(topShows, showUrlDf)
allShows
```

##	Ranks	Title	ratings	cleanedVotes	numEpisodes
## 1	1	Breaking Bad	9.5	2.2M	62 eps
## 2	2	Planet Earth II	9.5	162K	6 eps
## 3	3	Planet Earth	9.4	224K	11 eps
## 4	4	Band of Brothers	9.4	545K	10 eps
## 5	5	Chernobyl	9.3	907K	5 eps
## 6	6	The Wire	9.3	391K	60 eps
## 7	7	Avatar: The Last Airbender	9.3	390K	62 eps
## 8	8	Blue Planet II	9.3	49K	7 eps
## 9	9	The Sopranos	9.2	498K	86 eps
## 10	10	Cosmos: A Spacetime Odyssey	9.2	132K	13 eps
## 11	11	Cosmos	9.3	46K	13 eps
## 12	12	Our Planet	9.2	54K	12 eps
## 13	13	Game of Thrones	9.2	2.4M	74 eps
## 14	14	Bluey	9.3	34K	194 eps
## 15	15	The World at War	9.2	31K	26 eps
## 16	16	Fullmetal Alchemist: Brotherhood	9.1	209K	68 eps
## 17	17	Rick and Morty	9.1	627K	78 eps
## 18	18	Life	9.1	44K	11 eps
## 19	19	The Last Dance	9.1	160K	10 eps
## 20	20	The Twilight Zone	9.0	97K	156 eps
## 21	21	The Vietnam War	9.1	30K	10 eps
## 22	22	Sherlock	9.1	1M	15 eps
## 23	23	Attack on Titan	9.1	561K	98 eps
## 24	24	Batman: The Animated Series	9.0	122K	85 eps
## 25	25	Arcane	9.0	302K	18 eps
## 26	1	Breaking Bad	9.5	2.2M	62 eps
## 27	2	Planet Earth II	9.5	162K	6 eps
## 28	3	Planet Earth	9.4	224K	11 eps
## 29	4	Band of Brothers	9.4	545K	10 eps
## 30	5	Chernobyl	9.3	907K	5 eps
## 31	6	The Wire	9.3	391K	60 eps
## 32	7	Avatar: The Last Airbender	9.3	390K	62 eps
## 33	8	Blue Planet II	9.3	49K	7 eps
## 34	9	The Sopranos	9.2	498K	86 eps
## 35	10	Cosmos: A Spacetime Odyssey	9.2	132K	13 eps
## 36	11	Cosmos	9.3	46K	13 eps

## 37	12	Our Planet	9.2	54K	12 eps
## 38	13	Game of Thrones	9.2	2.4M	74 eps
## 39	14	Bluey	9.3	34K	194 eps
## 40	15	The World at War	9.2	31K	26 eps
## 41	16	Fullmetal Alchemist: Brotherhood	9.1	209K	68 eps
## 42	17	Rick and Morty	9.1	627K	78 eps
## 43	18	Life	9.1	44K	11 eps
## 44	19	The Last Dance	9.1	160K	10 eps
## 45	20	The Twilight Zone	9.0	97K	156 eps
## 46	21	The Vietnam War	9.1	30K	10 eps
## 47	22	Sherlock	9.1	1M	15 eps
## 48	23	Attack on Titan	9.1	561K	98 eps
## 49	24	Batman: The Animated Series	9.0	122K	85 eps
## 50	25	Arcane	9.0	302K	18 eps

##	Year	User_Reviews	Critic_Reviews	Popularity_Rating
## 1	2008-2013	5,091 reviews	175	20
## 2	2016	5,091 reviews	175	20
## 3	2006	158 reviews	6	1,121
## 4	2001	158 reviews	6	1,121
## 5	2019	111 reviews	10	2,011
## 6	2002-2008	111 reviews	10	2,011
## 7	2005-2008	1,056 reviews	34	171
## 8	2017	1,056 reviews	34	171
## 9	1999-2007	3,533 reviews	88	173
## 10	2014	3,533 reviews	88	173
## 11	1980	787 reviews	77	108
## 12	2019-2023	787 reviews	77	108
## 13	2011-2019	998 reviews	57	373
## 14	2018-	998 reviews	57	373
## 15	1973-1974	53 reviews	9	4,415
## 16	2009-2010	53 reviews	9	4,415
## 17	2013-	963 reviews	93	33
## 18	2009	963 reviews	93	33
## 19	2020	205 reviews	12	1,499
## 20	1959-1964	205 reviews	12	1,499
## 21	2017	80 reviews	8	3,866
## 22	2010-2017	80 reviews	8	3,866
## 23	2013-2023	245 reviews	15	2,765
## 24	1992-1995	245 reviews	15	2,765
## 25	2021-2024	5,899 reviews	368	14
## 26	2008-2013	5,899 reviews	368	14
## 27	2016	367 reviews	4	411
## 28	2006	367 reviews	4	411
## 29	2001	126 reviews	5	2,627
## 30	2019	126 reviews	5	2,627
## 31	2002-2008	466 reviews	16	508
## 32	2005-2008	466 reviews	16	508
## 33	2017	911 reviews	94	137
## 34	1999-2007	911 reviews	94	137
## 35	2014	12 reviews	9	3,455
## 36	1980	12 reviews	9	3,455
## 37	2019-2023	541 reviews	28	1,521
## 38	2011-2019	541 reviews	28	1,521
## 39	2018-	213 reviews	85	354

## 40	1973-1974	213 reviews	85	354
## 41	2009-2010	175 reviews	13	2,022
## 42	2013-	175 reviews	13	2,022
## 43	2009	1,096 reviews	121	172
## 44	2020	1,096 reviews	121	172
## 45	1959-1964	2,360 reviews	64	60
## 46	2017	2,360 reviews	64	60
## 47	2010-2017	219 reviews	25	527
## 48	2013-2023	219 reviews	25	527
## 49	1992-1995	1,952 reviews	50	15
## 50	2021-2024	1,952 reviews	50	15

2. From the 50 tv shows, select at least 5 tv shows to scrape 20 user reviews that will include the reviewer's name, date of reviewed, user rating, title of the review, the numbers for "is helpful" and "is not helpful", and text reviews.

```
library(rvest)
library(dplyr)

urlsOfFiveShows <- c(
  "https://www.imdb.com/title/tt0903747/reviews/?ref=ttexr_q1_2",
  "https://www.imdb.com/title/tt5491994/reviews/?ref=tt_ov_q1_2",
  "https://www.imdb.com/title/tt0185906/reviews/?ref=tt_ov_q1_2",
  "https://www.imdb.com/title/tt7366338/reviews/?ref=tt_ov_q1_2",
  "https://www.imdb.com/title/tt0944947/reviews/?ref=tt_ov_q1_2"
)

fiveShowsUrlDf <- data.frame(
  Title = c(
    "Breaking Bad",
    "Planet Earth II",
    "Band of Brothers",
    "Chernobyl",
    "Game of Thrones"
  ),
  URLs = urlsOfFiveShows
)

scrapeReviews <- function(show_url) {
  page <- read_html(show_url)

  userNames <- page %>%
    html_nodes('[data-testid="author-link"]') %>%
    html_text()

  reviewDates <- page %>%
    html_nodes('li.review-date') %>%
    html_text()

  userRating <- page %>%
    html_nodes('span.ipc-rating-star--rating') %>%
    html_text()

  revTitle <- page %>%
    html_nodes('h3.ipc-title__text') %>%
```

```

html_text()

helpfulRev <- page %>%
html_nodes('span.count--up') %>%
html_text()

notHelpful <- page %>%
html_nodes('span.count--down') %>%
html_text()

data.frame(Usernames = head(userNames, 20), Dates = head(reviewDates, 20), userRating = head(userRating, 20))
}

reviews_data <- lapply(fiveShowsUrlDf$URLs, scrapeReviews)
names(reviews_data) <- fiveShowsUrlDf$Title
reviews_data[["Breaking Bad"]]

```

```

##           Usernames      Dates userRating
## 1           FiRE010  Jul 3, 2021          10
## 2         bruhperson  Mar 6, 2019          10
## 3       KinoKoopakid  Jul 29, 2021          10
## 4         jehuschultz Feb 18, 2020          10
## 5       Supermanfan-13 Nov 8, 2021          10
## 6   manishsingh-03299 May 30, 2019          10
## 7           xpinerhd Nov 15, 2019          10
## 8           Rob1331  Dec 8, 2022          10
## 9   dhanushreddy-14919 Jul 17, 2021          10
## 10  TheLittleSongbird Nov 12, 2017          10
## 11           JaxTL   Aug 5, 2022           7
## 12         reebokroot Feb 14, 2021           5
## 13 cursedlovesherbestfriend Feb 20, 2021          10
## 14           tushv-31482 Dec 8, 2022          10
## 15           gogoschka-1 Jan 11, 2014          10
## 16         FishDrowned Nov 8, 2021          10
## 17           joegalgano Aug 11, 2021          10
## 18           agatt-87232 May 19, 2019          10
## 19   Leofwine_draca  May 4, 2021          10
## 20   dristysultana  Jun 23, 2021          10

```

```

##
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13
## 14
## 15 If you mix Scarface, Robin Hood and maybe Tyler Durden with enough meth - you'll get a mean cockt

```

Those days a

Among the best and most a

Slow

Once



```
## 16 By far the greatest
## 17 in a c
## 18 Since GOT is over, this is Officially the C
## 19 Every bit a
## 20
```

```
reviews_data[["Planet Earth II"]]
```

##		Username	Date	userRating
## 1	arjanhulkema	Nov 7, 2016	10	
## 2	Wentloog	Nov 5, 2016	10	
## 3	john-m-madsen	Nov 5, 2016	10	
## 4	thespookybuz	Nov 9, 2016	10	
## 5	pjdickinson	Nov 5, 2016	10	
## 6	dbijis33	Nov 8, 2016	10	
## 7	dhanrajjughead	Nov 17, 2016	10	
## 8	NeilBarnett	Nov 13, 2016	10	
## 9	salmanu-27386	Nov 6, 2016	10	
## 10	panagiotiskatsanos	Dec 31, 2016	10	
## 11	ianrobo	Nov 19, 2016	10	
## 12	adamonIMDb	Dec 28, 2016	7	
## 13	tinyfordst	May 19, 2019	10	
## 14	BobFillmore	Sep 29, 2017	10	
## 15	farshidkarimi	Nov 22, 2016	10	
## 16	TheLittleSongbird	Oct 12, 2017	10	
## 17	myerse-165-4350	Dec 4, 2016	10	
## 18	larask-21775	Oct 20, 2018	10	
## 19	fierceeagle-40009	Apr 23, 2020	10	
## 20	adam-whitmore	Jan 5, 2017	10	

```
##
## 1
## 2 At once awe-inspiring a
## 3 Yet another masterpiece from BBC Nature & Davi
## 4
## 5
## 6 Danger
## 7 Greatest documentary
## 8 Best thing on TV since la
## 9
## 10 One of the best documentaries
## 11 In times of climat
## 12
## 13 More irritated with IMDb for the bias than
## 14 Should be required view
## 15 What a Beautiful Planet
## 16 Like the first 'Planet Earth', does for nature and our planet as 'Walking with Dinosaurs' did with
## 17 This masterpiece deserve
## 18
## 19 Absol
## 20 Peerless evocation of nature a
```

```
reviews_data[["Band of Brothers"]]
```

##		Username	Date	userRating
## 1	Rob1331	Sep 27, 2022	10	

## 2	sanderson777	Oct 14, 2001	10
## 3	wildcatt268	Jan 18, 2002	10
## 4	arjay24	Apr 18, 2004	10
## 5	rbverhoef	Feb 13, 2003	10
## 6	yodaschoda	Jan 23, 2005	10
## 7	philip_vanderveken	Sep 16, 2004	10
## 8	Supermanfan-13	May 6, 2022	10
## 9	thiagoutp	Nov 4, 2019	10
## 10	bsmith5552	Nov 5, 2001	10
## 11	faded_english_monkey	Aug 25, 2004	10
## 12	planktonrules	May 30, 2015	7
## 13	stilonkostrzyn	Apr 10, 2021	5
## 14	faded_Glory	May 2, 2006	10
## 15	jazmodo	Jun 3, 2019	10
## 16	kait2007	Jan 26, 2005	10
## 17	mickman91-1	May 3, 2022	10
## 18	kipmcmillan	Oct 24, 2018	9
## 19	erwan_ticheler	Dec 7, 2002	10
## 20	grahamsj3	Nov 25, 2002	10

##	Review_Title
## 1	Incredible!!
## 2	Possibly the finest 10 hours ever created
## 3	One of the best war movies/series ever
## 4	Realistic
## 5	Excellent
## 6	One of, if not the best, mini series' ever made
## 7	This series is so unbelievably realistic, so authentic.
## 8	One of the best mini-series ever created!
## 9	Probably the best ever
## 10	Realistic WWII Drama With Warts Included
## 11	war, no frills
## 12	You can't beat this....
## 13	Overrated??
## 14	Not very realistic at all
## 15	Without Doubt, the Best Mini-Series Ever Recorded
## 16	Great Miniseries
## 17	A series like this won't be made again (see below), so treasure it
## 18	Share With Your Children
## 19	Best Mini series ever
## 20	A-1, TOPS, the BOMB what else can I say?

```
reviews_data[["Chernobyl"]]
```

##	Username	Date	userRating
## 1	curiosityonmars	May 23, 2019	10
## 2	stelmakh	May 10, 2019	10
## 3	natashapekar	May 9, 2019	10
## 4	m-porpaczi	May 14, 2019	10
## 5	Lladerat	May 7, 2019	10
## 6	jfirebug	May 20, 2019	10
## 7	thegldt	May 6, 2019	10
## 8	alexander-phoenix	May 13, 2019	10
## 9	wmeduardowm	May 6, 2019	10
## 10	Leofwine_draca	Nov 27, 2019	10
## 11	Jamie_Seaton	May 23, 2019	7

```

## 12      m-abtahi Jan 27, 2024      1
## 13      piter554 Jun 29, 2019      8
## 14      frimark May 20, 2019      10
## 15      krzysztof-18241 May 30, 2019      10
## 16      ahmetkozan Jun 7, 2019      10
## 17      stephenpdodds May 6, 2019      9
## 18      Rob1331 Sep 27, 2022      9
## 19      emholberg May 26, 2019      9
## 20      Supermanfan-13 Jul 10, 2022      10
##
##                                     Review_Title
## 1                                     They got it right
## 2                                     Goosebumps and tears
## 3                                     I highly recommend this film!
## 4                                     No hero wakes up wanting to die
## 5                                     So far looks excellent
## 6                                     Incredible
## 7      Bleak, Unsettling, Haunting All Throughout
## 8                                     Unbelievable
## 9                                     HBO did it again!
## 10                                    Exemplary
## 11                                    Amazing!
## 12      Unveiling Human Errors and Political Shadows
## 13                                    How cost the lie?
## 14                                    Emotionally drained...
## 15                                    Just watch it (!)
## 16      Now you look like the minister of coal!
## 17                                    Cracking.
## 18                                    Must Watch!
## 19 It is hard to overestimate the importance of this show.
## 20                                    Brilliant!

```

```
reviews_data[["Game of Thrones"]]
```

```

##      Usernames      Dates userRating
## 1      danielkpkp May 11, 2020      9
## 2      samxxx-671-826221 May 24, 2019      8
## 3      slowcando May 20, 2019      8
## 4      Saif0VGU Apr 8, 2020      10
## 5      jacobnoble-02524 May 9, 2019      9
## 6      heavenacceptme Aug 22, 2022      10
## 7      Supermanfan-13 Dec 9, 2023      10
## 8      Dan_W_Reviews Feb 5, 2023      9
## 9      tweaknhoe May 25, 2019      8
## 10     TheLittleSongbird Nov 8, 2017      10
## 11      kunkell Nov 1, 2019      6
## 12     akhil-marsonya27 May 20, 2019      1
## 13      Rob1331 Nov 10, 2023      10
## 14     alshamari-marwa May 18, 2020      9
## 15     tobiascramon Nov 8, 2022      9
## 16     adamheritage-15333 May 5, 2019      9
## 17      LASTRONOME Jun 2, 2020      10
## 18     SpitOnAStranger May 9, 2019      8
## 19     psypeterson Apr 16, 2011      10
## 20     el-absy May 19, 2019      7
##
##                                     Review_Title

```

```
## 1           It could have been the best TV series ever made...
## 2           A perfect example of: Falling in Love with the Wrong Guy
## 3           Seasons 1-6: outstanding. 7: daft but good. 8: disappointing
## 4           Can you just make the remake the season finale?
## 5           Game of Thrones
## 6 Despite the final season, Game of Thrones remains an all time classic
## 7           One of the best shows ever
## 8           Captivating and Gripping but a Disappointing Final Season
## 9           Why just why? This show could have been the best ever.
## 10          This is a television show?
## 11          Imagine an Ice Cream Shop
## 12          A Message to Dan and Dave
## 13          Amazing
## 14          Extraordinary untill season 8
## 15          one the best shows ever made
## 16 Can we please just restart season 8, perhaps 7 as well, but mainly 8
## 17          almost perfect
## 18          This was an 10/10 until S08E03
## 19          Excellent adaptation.
## 20          Great Beginning, WORST Ending!
```

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)
years <- substr(Year, 1,4)
years <- as.numeric(Year)
```

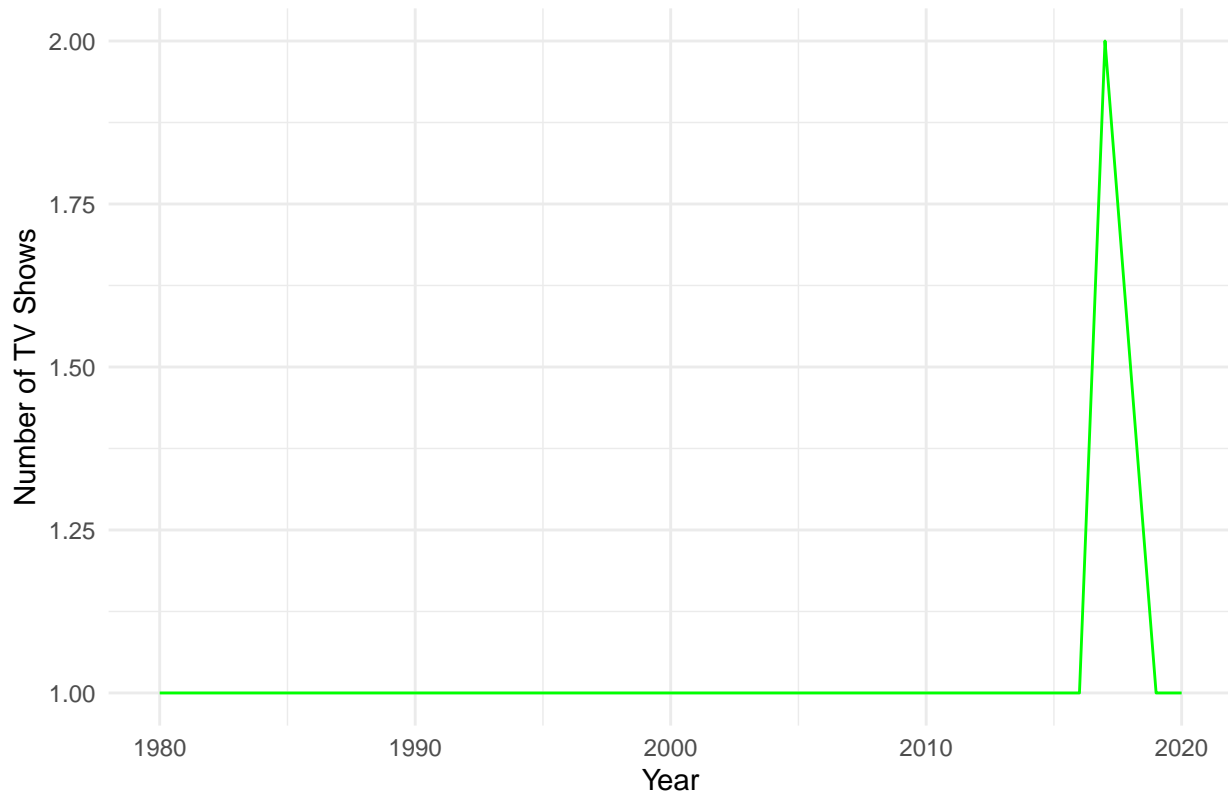
```
## Warning: NAs introduced by coercion
```

```
ggplot(data.frame(Year = years), aes(x = Year)) +
  geom_line(stat = "count", fill = "skyblue", color = "green") +
  labs(title = "Number of TV Shows Released by Year",
       x = "Year",
       y = "Number of TV Shows") +
  theme_minimal()
```

```
## Warning in geom_line(stat = "count", fill = "skyblue", color = "green"):
## Ignoring unknown parameters: `fill`
```

```
## Warning: Removed 15 rows containing non-finite outside the scale range
## (`stat_count()`).
```

## Number of TV Shows Released by Year



```
mostShowsYear <- as.data.frame(table(Year))
mostShowsYear <- mostShowsYear[which.max(mostShowsYear$Freq), ]
print(mostShowsYear)
```

```
##      Year Freq
## 19 2017     2
```

Extracting Amazon Product Reviews 4.

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

```
urls <- c('https://www.amazon.com/s?k=backpacks&crd=35ZQ1H72MC3G9&sprefix=backpacks%2Caps%2C590&ref=nb_sb_noss_1',
          'https://www.amazon.com/s?k=laptops&crd=L7MQBW7MD4SX&sprefix=laptopb%2Caps%2C1304&ref=nb_sb_noss_1',
          'https://www.amazon.com/s?k=phone+case&dc&crd=1VPDCJ87S93TL&sprefix=phone+cas%2Caps%2C451&ref=nb_sb_noss_1',
          'https://www.amazon.com/s?k=mountain+bike&crd=1ZQR71S8XHZN6&sprefix=mountain+bik%2Caps%2C499&ref=nb_sb_noss_1',
          'https://www.amazon.com/s?k=tshirt&crd=2RQIP7MP6IYAW&sprefix=tshirt%2Caps%2C443&ref=nb_sb_noss_1')
```

```
df <- list()
```

```
for (i in seq_along(urls)) {
```

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

```
  productName <- scrape(sessions) %>%
    html_nodes('h2.a-size-mini') %>%
```

```

    html_text() %>%
    head(30)

productDescription <- scrape(sessions) %>%
  html_nodes('div.productDescription') %>%
  html_text() %>%
  head(30)

productRating <- scrape(sessions) %>%
  html_nodes('span.a-icon-alt') %>%
  html_text() %>%
  head(30)

ratings <- as.numeric(str_extract(productRating, "\\d+\\.\\d"))

productPrice <- scrape(sessions) %>%
  html_nodes('span.a-price') %>%
  html_text() %>%
  head(30)
price <- as.numeric(str_extract(productPrice, "\\d+\\.\\d+"))

productReview <- scrape(sessions) %>%
  html_nodes('div.review-text-content') %>%
  html_text() %>%
  head(30)

dfAma <- data.frame(Product_Name = productName[1:30],
                    Description = productDescription[1:30],
                    Rating = ratings[1:30],
                    Price = price[1:30],
                    stringsAsFactors = FALSE)

df[[i]] <- dfAma
}

print(df[[1]])

##
## 1 MATEIN Travel Laptop Backpack, Business Anti Theft Slim Sturdy Laptops Backpack with USB Charging
## 2 JanSport SuperBreak One Backpacks - Durable, Lightweight Bo
## 3 YOREPEK Travel Backpack, Extra Large 50L Laptop Backpacks for Men Women, Water Resistant Col
## 4
## 5
## 6 Lapsouno Travel Backpack, Large Carry on Backpack, 17 Inch Laptop Backpack, Big Backpack, Extra
## 7
## 8 Laptop Backpack,Business Travel Anti Theft Slim Durable Laptops Backpack with USB Charging
## 9
## 10
## 11 JanSport Right Pack Backpack - Durable Dayp

```

```

## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
## 25
## 26
## 27
## 28
## 29
## 30
##      Description Rating Price
## 1      <NA>      4.6 21.99
## 2      <NA>      4.6 39.96
## 3      <NA>      4.5 38.00
## 4      <NA>      4.7 27.99
## 5      <NA>      4.6 39.99
## 6      <NA>      4.7 92.00
## 7      <NA>      4.8 99.00
## 8      <NA>      4.6 23.99
## 9      <NA>      4.5 23.98
## 10     <NA>      4.7 49.99
## 11     <NA>      4.6 65.00
## 12     <NA>      3.7 29.99
## 13     <NA>      4.6  8.99
## 14     <NA>      4.7 12.00
## 15     <NA>      4.6 23.99
## 16     <NA>      4.8 45.00
## 17     <NA>      4.7 64.99
## 18     <NA>      4.8 20.25
## 19     <NA>      4.7 45.00
## 20     <NA>      NA 47.32
## 21     <NA>      NA 60.00
## 22     <NA>      NA 29.99
## 23     <NA>      NA 33.99
## 24     <NA>      NA 16.99
## 25     <NA>      NA 80.00
## 26     <NA>      NA  NA
## 27     <NA>      NA  NA
## 28     <NA>      NA  NA
## 29     <NA>      NA  NA
## 30     <NA>      NA  NA

```

```
print(df[[2]])
```

```

##
## 1      Acer Aspire 3 A315-24P-R7VH Slim Laptop | 15.6" Full HD IPS Display | AMD Ryzen 3 7320U Quad-C

```

```

## 2 HP Notebook Laptop, 15.6" HD Touchscreen, Intel Core i3-1115G4
## 3 HP Newest 255 G10 Laptop for Home or Work, 16GB RAM, 1TB SSD, 15.6" Full HD, Ryzen 3
## 4 HP Portable Laptop, Student and Business, 14" HD Display, Intel Quad-Core
## 5 HP 14 Laptop, Intel Celeron N4020, 4 GB RAM, 64 GB Storage, 14-inch Micro-edge HD Display,
## 6 HP 17 Laptop, 17.3" HD+ Display,
## 7 HP 15.6" Portable Laptop (Include 1 Year Microsoft 365), HD Display, Intel Quad-Core
## 8
## 9 HP Newest 14" LED Business Laptop Computer, 16GB RAM 320GB Storage (64GB eMMC+256GB SD Card)
## 10 HP Newest Pavilion 15.6" HD Touchscreen Anti-Glare Laptop, 16GB RAM, 1TB SSD Storage, Intel
## 11 Lenovo 15.6" IdeaPad Essential Laptop Computer, Windows 11 Pro with Microsoft Office Lifetime
## 12 Apple 2024 MacBook Air 15-inch Laptop with M3 chip: Built for Apple Intelligence, 15.3-inch
## 13 HP 2024 Newest Laptop for Business and Student, 15.6" HD Touchscreen, Intel 6-Core i3-1215U Processor
## 14 ASUS Lightweight 15.5" Full HD Laptop, Windows 11 Home
## 15 HP Newest 14" Ultral Light Laptop for Students and Business, Intel Quad-Core N4120, 8GB
## 16 HP 15.6 FHD Display G9 Laptop • 32GB RAM • 1TB Storage (512GB SSD & 512GB HDD)
## 17 Apple 2024 MacBook Air 13-inch Laptop with M3 chip: Built for Apple Intelligence, 13.6-inch
## 18 HP Newest Pavilion 15.6" Touchscreen Laptop with 12 Month Microsoft • 40GB RAM • 1TB Storage (512GB SSD & 512GB HDD)
## 19 Magnetic Privacy Screen for MacBook Pro 14 Inch (2021, 2022, 2023, 2024)
## 20 280W ADP-280EB B Laptop Charger AC Adapter for Asus ROG Strix G15 G513 G513QM G17 G713QR G713QM G713Q
## 21 170W Laptop Charger Fit for Lenovo ThinkPad P73 P50 P51 P52 W540 W541 P15v P16v T14 T15 T16 T17 T18 T19
## 22
## 23
## 24
## 25
## 26
## 27
## 28
## 29
## 30
## Description Rating Price
## 1 <NA> 4.4 279.99
## 2 <NA> 4.1 4.46
## 3 <NA> 4.3 321.99
## 4 <NA> 4.1 404.82
## 5 <NA> 4.0 6.75
## 6 <NA> 4.2 431.99
## 7 <NA> 4.1 448.88
## 8 <NA> 4.1 599.00
## 9 <NA> 3.9 208.99
## 10 <NA> 4.1 299.00
## 11 <NA> 5.0 180.00
## 12 <NA> 4.7 209.99
## 13 <NA> 4.3 475.00
## 14 <NA> 4.1 267.95
## 15 <NA> 4.0 305.00
## 16 <NA> 4.4 146.00
## 17 <NA> 4.7 299.99
## 18 <NA> 4.3 399.99
## 19 <NA> 4.3 397.44
## 20 <NA> 5.0 6.62
## 21 <NA> 3.3 399.00
## 22 <NA> NA 999.00
## 23 <NA> NA 299.00
## 24 <NA> NA 399.00

```



```
## 25      <NA>      NA 169.99
## 26      <NA>      NA 249.99
## 27      <NA>      NA 419.58
## 28      <NA>      NA 499.00
## 29      <NA>      NA  99.00
## 30      <NA>      NA 299.00
```

```
print(df[[3]])
```

```
##
## 1          ESR for iPhone 14 Case/iPhone 13 Case, Compatible with MagSafe, Shockproof
## 2 BENTOBEN Magnetic for iPhone 13 Case & iPhone 14 Case [Compatible with Magsafe] Translucent Matte
## 3                                     OtterBox iPhone 15, iPh
## 4          Temdan for iPhone 16 Pro Case Clear, [Compatible with Magsafe][Anti-Y
## 5                                                         OtterBox :
## 6
## 7          ORNARTO Compatible with iPhone
## 8          FABSPARK Phone Case for iPhone 13 (
## 9                                                         Otter
## 10          OtterBox iPhone 13 Pro Max
## 11          OtterBox iPhone 12 & iPhone 12 Pro Commuter Series
## 12          elago Compatible with iPhone 14 Pro Max Case, Liquid Silicone Case, Full Bo
## 13          elago Compatible with iPhone 14 Pro Case, Liquid Silicone Case, Full Bo
## 14          elago Compatible with iPhone 14 Case, Liquid Silicone Case, Full Bo
## 15          elago Liquid Silicone Case Compatible with iPhone 13 Pro Case (6.1"), Premium Silic
## 16          TAURI 3 in 1 for iPhone 14 Case Clear, [Not Y
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
## 25
## 26
## 27
## 28
## 29
## 30
##      Description Rating Price
## 1      <NA>      3.7 14.99
## 2      <NA>      3.5 30.99
## 3      <NA>      3.7 12.98
## 4      <NA>      4.6 34.95
## 5      <NA>      4.5 39.95
## 6      <NA>      4.6  8.99
## 7      <NA>      4.6 35.06
## 8      <NA>      4.7 39.95
## 9      <NA>      3.9 44.95
## 10     <NA>      4.5  9.99
## 11     <NA>      4.3  9.99
## 12     <NA>      4.7 29.95
## 13     <NA>      4.6 29.95
## 14     <NA>      4.6 39.95
```

```
## 15      <NA>      4.5 24.99
## 16      <NA>      4.4 39.95
## 17      <NA>      4.5 12.99
## 18      <NA>      4.4 12.99
## 19      <NA>      4.7 12.99
## 20      <NA>      NA 12.99
## 21      <NA>      NA 12.74
## 22      <NA>      NA 16.99
## 23      <NA>      NA    NA
## 24      <NA>      NA    NA
## 25      <NA>      NA    NA
## 26      <NA>      NA    NA
## 27      <NA>      NA    NA
## 28      <NA>      NA    NA
## 29      <NA>      NA    NA
## 30      <NA>      NA    NA
```

```
print(df[[4]])
```

```
##
## 1  WEIZE Mountain Bike, 24/26/27.5 inch Outdoor Cycling Bike,18-Speed/High-Carbon Steel/Dual Full Su
## 2      Mongoose Malus Mens and Women Fat Tire Mountain Bike, 26-Inch Bicycle W
## 3      Racer Electric Bike for Adults - 21-Speed Mountain Lightweight Ebike with
## 4      Dynacraft Magna Echo Ridge Mountain Bike - Rugged and Durable Des
## 5      Mongoose Flatrock 21-Speed Hardtail Mountain Bike, 24 to 1
## 6      Ktaxon Mountain Bike 26/27.5/29 Inch Men &
## 7  Schwinn Traxion Mountain Bike for Adult Men Women, 29-Inch Wheels, Full Suspension, 24-Speed Shif
## 8      Mountain Bike 24/26/27.5 Inch, Dual Full Suspens
## 9      Schwinn High Timber Mountain Bike for Adult Youth Men Wome
## 10     Ecarpat 24/26/27.5 Inch Mountain Bike, Dual/Front Suspension Mens Mountain Bike with 21 Sp
## 11     Outroad 26 Inch Mountain Bike, 21-Speed/High-Carbon Steel/Aviation Grade Frame, 1
## 12     isinwheel M10 Electric Bike Adult 500W, 26" Commuting Electric Mountain Bike 20MPH Max Range 5
## 13     Dynacraft Vertical Alpine Eagle Mountain Bike - Rug
## 14     Schwinn Bonafide Men and Wor
## 15     Jasion EB5 Electric Bike for Adults with Peak 500W Brushless Motor, 40Miles 20MPH Commu
## 16     Outroad Fat Tire Mountain Bike with 26 inch
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
## 25
## 26
## 27
## 28
## 29
## 30
##      Description Rating  Price
## 1      <NA>      4.0 179.99
## 2      <NA>      4.4 199.99
## 3      <NA>      4.3 492.99
## 4      <NA>      4.0 519.99
```

```
## 5      <NA>      4.1 199.99
## 6      <NA>      4.2 158.94
## 7      <NA>      4.3 309.99
## 8      <NA>      4.3 199.99
## 9      <NA>      4.1 613.14
## 10     <NA>      4.4 674.99
## 11     <NA>      3.9 179.99
## 12     <NA>      4.3 499.99
## 13     <NA>      4.3 249.99
## 14     <NA>      4.2  99.98
## 15     <NA>      4.2 389.99
## 16     <NA>      4.0 613.26
## 17     <NA>      NA 399.99
## 18     <NA>      NA 349.99
## 19     <NA>      NA    NA
## 20     <NA>      NA    NA
## 21     <NA>      NA    NA
## 22     <NA>      NA    NA
## 23     <NA>      NA    NA
## 24     <NA>      NA    NA
## 25     <NA>      NA    NA
## 26     <NA>      NA    NA
## 27     <NA>      NA    NA
## 28     <NA>      NA    NA
## 29     <NA>      NA    NA
## 30     <NA>      NA    NA
```

```
print(df[[5]])
```

```
##
## 1
## 2      Men's Crew T-Shirts, Multipack, Style G2000, M
## 3
## 4      Men's Pocket Undershirt Pack, Cotton Crew Neck T-Shirt, Moisture Wicking Tee, Assorted Colors, Fruit Colors
## 5
## 6      Men's Eversoft Cotton Stay Tucked Crew Neck T-Shirt, Moisture Wicking, Odor Protection, UPF 30+, Small
## 7
## 8      Men's T-Shirt, Beefy-T Heavyweight Cotton Crewneck Tee, 1 or 2 Pack, Available in T-Shirts, Fruit Colors, Fruit Colors
## 9
## 10     Unisex Adult Ultra Cotton T-Shirt, Style G2000, M
## 11     T-Shirts, Fruit Colors, Fruit Colors
## 12     6 Pack, Men's Short Sleeve Crew Neck T-Shirt, Moisture Wicking, Odor Protection, UPF 30+, Small
## 13
## 14     Men's Cotton, Moisture-Wicking Crew Tee Undershirts, Multi-Packs A
## 15
## 16     Men's V-Neck T-Shirts, Multipack, Style G2000, M
## 17     Fruit Colors, Fruit Colors
## 18     Men's Eversoft Cotton Stay Tucked V-Neck T-Shirt, Moisture Wicking, Odor Protection, UPF 30+, Small
## 19     Russe
## 20     Men's Dri-Power Cotton Blend Short Sleeve Tees, Moisture Wicking, Odor Protection, UPF 30+, Small
## 21     Com
## 22     Adult Heavyweight Short Sleeve Tee, Style G1
## 23     Libe
## 24 5 Pack Men's Active Quick Dry Crew Neck T Shirts | Athletic Running Gym Workout Short Sleeve Tee
## 25     T-Shirts
```

## 26				Tees Men's Short Sleeve Crew Neck
## 27				
## 28				Men's Short Sleeve T-Shirt Pack, Essentials Crewneck Cotton T-Shirt, 4
## 29				
## 30				Men's Loose Fit Heavyweight Short-Sleeve Pocket
##	Description	Rating	Price	
## 1	<NA>	NA	21.97	
## 2	<NA>	4.4	21.98	
## 3	<NA>	4.6	21.48	
## 4	<NA>	4.4	6.85	
## 5	<NA>	4.6	12.00	
## 6	<NA>	4.5	13.50	
## 7	<NA>	4.5	99.99	
## 8	<NA>	4.3	15.40	
## 9	<NA>	4.5	22.00	
## 10	<NA>	4.5	19.79	
## 11	<NA>	4.6	21.99	
## 12	<NA>	4.5	17.42	
## 13	<NA>	4.5	18.49	
## 14	<NA>	4.6	7.48	
## 15	<NA>	4.0	10.99	
## 16	<NA>	4.4	11.48	
## 17	<NA>	4.6	14.99	
## 18	<NA>	4.4	33.95	
## 19	<NA>	4.4	59.95	
## 20	<NA>	4.5	56.99	
## 21	<NA>	4.5	89.97	
## 22	<NA>	4.5	17.58	
## 23	<NA>	4.4	26.00	
## 24	<NA>	4.3	19.99	
## 25	<NA>	4.3	68.40	
## 26	<NA>	4.3	76.00	
## 27	<NA>	4.3	43.99	
## 28	<NA>	4.3	9.69	
## 29	<NA>	4.5	10.99	
## 30	<NA>	4.5	18.75	

““