

RWorksheet#5_group(Corvera, Paclibar, Sabarillo)

Rotciv Corvera, Jhon Albert Paclibar, Kirk Axl Dend Sabarillo

2024-11-11

```
library(rvest)
library(magrittr)
library(stringr)

url <- "https://www.imdb.com/chart/toptv/?ref_=nv_tvv_250"

page <- read_html(url)

titles <- page %>%html_nodes(".ipc-title-link-wrapper") %>%html_text()
ranking <- 1:25

head(titles, 50)

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

rating <- page %>%html_nodes(".ipc-rating-star--rating") %>%html_text()

rating
```

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

episodes <- page %>% html_nodes("span.sc-5bc66c50-6.00dsw.cli-title-metadata-item") %>% html_text()
eps_clean <- str_extract(episodes, "\\d+(?= eps)")
number_of_episodes <- eps_clean[!is.na(eps_clean)]
print(number_of_episodes)
```

```
## [1] "62" "6" "11" "10" "5" "60" "62" "7" "86" "13" "13" "12"
## [13] "74" "194" "26" "68" "78" "11" "10" "156" "10" "15" "98" "85"
## [25] "188"
```

```
episodes <- page %>% html_nodes("span.sc-5bc66c50-6.00dsw.cli-title-metadata-item") %>% html_text()

year_release <- str_extract(episodes, "\\d{4}")

year_release <- year_release[!is.na(year_release)]

year_release <- as.numeric(year_release)

print(year_release)
```

```
## [1] 2008 2016 2006 2001 2019 2002 2005 2017 1999 2014 1980 2019 2011 2018 1973
## [16] 2009 2013 2009 2020 1959 2017 2010 2013 1992 2005
```

```
df <- data.frame(ranking, titles, number_of_episodes, year_release, rating)
df
```

##	ranking	titles	number_of_episodes	year_release
## 1	1	1. Breaking Bad	62	2008
## 2	2	2. Planet Earth II	6	2016
## 3	3	3. Planet Earth	11	2006
## 4	4	4. Band of Brothers	10	2001
## 5	5	5. Chernobyl	5	2019
## 6	6	6. The Wire	60	2002
## 7	7	7. Avatar: The Last Airbender	62	2005
## 8	8	8. Blue Planet II	7	2017
## 9	9	9. The Sopranos	86	1999
## 10	10	10. Cosmos: A Spacetime Odyssey	13	2014
## 11	11	11. Cosmos	13	1980
## 12	12	12. Our Planet	12	2019
## 13	13	13. Game of Thrones	74	2011
## 14	14	14. Bluey	194	2018
## 15	15	15. The World at War	26	1973
## 16	16	16. Fullmetal Alchemist: Brotherhood	68	2009
## 17	17	17. Rick and Morty	78	2013
## 18	18	18. Life	11	2009
## 19	19	19. The Last Dance	10	2020
## 20	20	20. The Twilight Zone	156	1959
## 21	21	21. The Vietnam War	10	2017
## 22	22	22. Sherlock	15	2010
## 23	23	23. Attack on Titan	98	2013

## 24	24	24. Batman: The Animated Series	85	1992
## 25	25	25. The Office	188	2005
##	rating			
## 1	9.5			
## 2	9.5			
## 3	9.4			
## 4	9.4			
## 5	9.3			
## 6	9.3			
## 7	9.3			
## 8	9.3			
## 9	9.2			
## 10	9.2			
## 11	9.3			
## 12	9.2			
## 13	9.2			
## 14	9.3			
## 15	9.2			
## 16	9.1			
## 17	9.1			
## 18	9.1			
## 19	9.1			
## 20	9.0			
## 21	9.1			
## 22	9.1			
## 23	9.1			
## 24	9.0			
## 25	9.0			