RWorksheet5(Narra,Pelaez,Saria)

Agatha Hazel D. Narra, Riza Angelique F. Pelaez, Christine Pauline Saria

```r
library(polite)
library(tidyverse)
library(httr)
library(rvest)
library(dplyr)
library(stringr)
library(magrittr)
library(ggplot2)
```

1. Extracting TV Shows

```r
url <- "https://www.imdb.com/chart/toptv/?sort=rank%2Casc"

#1
#get the ranks and titles
title_list <- read_html(url) %>%
  html_nodes('.ipc-title__text') %>%
  html_text()

#Clean extracted text
title_list_sub <- as.data.frame(title_list[3:27], stringsAsFactors = FALSE)
colnames(title_list_sub) <- "ranks"

split_df <- strsplit(as.character(title_list_sub$ranks), "\\.", fixed = FALSE)
split_df <- data.frame(do.call(rbind, split_df), stringsAsFactors = FALSE)

colnames(split_df) <- c("rank", "title")
split_df <- split_df %>% select(rank, title)

split_df$title <- trimws(split_df$title)

rank_title <- split_df

#get tv rating, the number of people who voted, the number of episodes, and the year it was released.
rating_ls <- read_html(url) %>%
  html_nodes('.ipc-rating-star--rating') %>%
  html_text()

voter_ls <- read_html(url) %>%
  html_nodes('.ipc-rating-star--voteCount') %>%
  html_text()
clean_votes <- gsub('[()]', '', voter_ls)
```

```r
#get the number of episodes
eps_ls <- read_html(url) %>%
  html_nodes('span.sc-5bc66c50-6.OOdsw.cli-title-metadata-item:nth-of-type(2)') %>%
  html_text()
clean_eps <- gsub('[eps]', '', eps_ls)
num_eps <- as.numeric(clean_eps)

#get year released
years <- read_html(url) %>%
  html_nodes('span.sc-5bc66c50-6.OOdsw.cli-title-metadata-item:nth-of-type(1)') %>%
  html_text()

# Debugging: Check lengths of each component
print(length(rank_title$rank))
```

```
## [1] 25
```

```r
print(length(rank_title$title))
```

```
## [1] 25
```

```r
print(length(rating_ls))
```

```
## [1] 25
```

```r
print(length(clean_votes))
```

```
## [1] 25
```

```r
print(length(num_eps))
```

```
## [1] 0
```

```r
print(length(years))
```

```
## [1] 0
```

```r
# Maximum length of all components
max_length <- max(
  length(rank_title$rank), length(rank_title$title), length(rating_ls),
  length(clean_votes), length(num_eps), length(years)
)

# Pad shorter components with NA to match the max length
rank_title$rank <- c(rank_title$rank, rep(NA, max_length - length(rank_title$rank)))
rank_title$title <- c(rank_title$title, rep(NA, max_length - length(rank_title$title)))
rating_ls <- c(rating_ls, rep(NA, max_length - length(rating_ls)))
clean_votes <- c(clean_votes, rep(NA, max_length - length(clean_votes)))
num_eps <- c(num_eps, rep(NA, max_length - length(num_eps)))
```

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

# Create the data frame
top_tv_shows <- data.frame(
  Rank = rank_title$rank,
  Title = rank_title$title,
  Rating = rating_ls,
  Voters = clean_votes,
  Episodes = num_eps,
  Year = years,
  stringsAsFactors = FALSE
)

# Number of user reviews
home_link <- 'https://www.imdb.com/chart/toptv/'
main_page <- read_html(home_link)

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

#get link of each show's page
show_data <- lapply(links, function(link) {
  complete_link <- paste0("https://imdb.com", link)

  #get the link for user review page
  usrv_link <- read_html(complete_link)
  usrv_link_page <- usrv_link %>%
    html_nodes('a.isReview') %>%
    html_attr("href")

  #get critic reviews
  critic <- usrv_link %>%
              html_nodes("span.score") %>%
              html_text()
  critic_df <- data.frame(Critic_Reviews = critic[2], stringsAsFactors = FALSE)

  #get pop rating
  pop_rating <- usrv_link %>%
              html_nodes('[data-testid="hero-rating-bar__popularity__score"]') %>%
              html_text()

  #get user reviews of each shows
  usrv <- read_html(paste0("https://imdb.com", usrv_link_page[1]))
  usrv_count <- usrv %>%
    html_nodes('[data-testid="tturv-total-reviews"]') %>%
    html_text()

  return(data.frame(Show_Link = complete_link, User_Reviews = usrv_count, Critic = critic_df, Popularity
})

show_url_df <- do.call(rbind, show_data)
show_url_df
```

```
##                                                Show_Link  User_Reviews Critic_Reviews Popularity_Ratin
## 1    https://imdb.com/title/tt0903747/?ref_=chttvtp_t_1 5,108 reviews            175                2
## 2    https://imdb.com/title/tt0903747/?ref_=chttvtp_t_1 5,108 reviews            175                2
## 3    https://imdb.com/title/tt5491994/?ref_=chttvtp_t_2   158 reviews              6              99
## 4    https://imdb.com/title/tt5491994/?ref_=chttvtp_t_2   158 reviews              6              99
## 5    https://imdb.com/title/tt0795176/?ref_=chttvtp_t_3   111 reviews             10             1,77
## 6    https://imdb.com/title/tt0795176/?ref_=chttvtp_t_3   111 reviews             10             1,77
## 7    https://imdb.com/title/tt0185906/?ref_=chttvtp_t_4 1,058 reviews             34               15
## 8    https://imdb.com/title/tt0185906/?ref_=chttvtp_t_4 1,058 reviews             34               15
## 9    https://imdb.com/title/tt7366338/?ref_=chttvtp_t_5 3,534 reviews             88               14
## 10   https://imdb.com/title/tt7366338/?ref_=chttvtp_t_5 3,534 reviews             88               14
## 11   https://imdb.com/title/tt0306414/?ref_=chttvtp_t_6   787 reviews             77               10
## 12   https://imdb.com/title/tt0306414/?ref_=chttvtp_t_6   787 reviews             77               10
## 13   https://imdb.com/title/tt0417299/?ref_=chttvtp_t_7 1,003 reviews             57               35
## 14   https://imdb.com/title/tt0417299/?ref_=chttvtp_t_7 1,003 reviews             57               35
## 15   https://imdb.com/title/tt6769208/?ref_=chttvtp_t_8    53 reviews              9             4,12
## 16   https://imdb.com/title/tt6769208/?ref_=chttvtp_t_8    53 reviews              9             4,12
## 17   https://imdb.com/title/tt0141842/?ref_=chttvtp_t_9   966 reviews             93                3
## 18   https://imdb.com/title/tt0141842/?ref_=chttvtp_t_9   966 reviews             93                3
## 19  https://imdb.com/title/tt2395695/?ref_=chttvtp_t_10   205 reviews             12             1,57
## 20  https://imdb.com/title/tt2395695/?ref_=chttvtp_t_10   205 reviews             12             1,57
## 21  https://imdb.com/title/tt0081846/?ref_=chttvtp_t_11    80 reviews              8             3,64
## 22  https://imdb.com/title/tt0081846/?ref_=chttvtp_t_11    80 reviews              8             3,64
## 23  https://imdb.com/title/tt9253866/?ref_=chttvtp_t_12   245 reviews             15             2,40
## 24  https://imdb.com/title/tt9253866/?ref_=chttvtp_t_12   245 reviews             15             2,40
## 25  https://imdb.com/title/tt0944947/?ref_=chttvtp_t_13 5,907 reviews            368                1
## 26  https://imdb.com/title/tt0944947/?ref_=chttvtp_t_13 5,907 reviews            368                1
## 27  https://imdb.com/title/tt7678620/?ref_=chttvtp_t_14   369 reviews              4               37
## 28  https://imdb.com/title/tt7678620/?ref_=chttvtp_t_14   369 reviews              4               37
## 29  https://imdb.com/title/tt0071075/?ref_=chttvtp_t_15   126 reviews              5             2,53
## 30  https://imdb.com/title/tt0071075/?ref_=chttvtp_t_15   126 reviews              5             2,53
## 31  https://imdb.com/title/tt1355642/?ref_=chttvtp_t_16   468 reviews             16               47
## 32  https://imdb.com/title/tt1355642/?ref_=chttvtp_t_16   468 reviews             16               47
## 33  https://imdb.com/title/tt2861424/?ref_=chttvtp_t_17   910 reviews             94               12
## 34  https://imdb.com/title/tt2861424/?ref_=chttvtp_t_17   910 reviews             94               12
## 35  https://imdb.com/title/tt1533395/?ref_=chttvtp_t_18    12 reviews              9             3,05
## 36  https://imdb.com/title/tt1533395/?ref_=chttvtp_t_18    12 reviews              9             3,05
## 37  https://imdb.com/title/tt8420184/?ref_=chttvtp_t_19   542 reviews             28             1,40
## 38  https://imdb.com/title/tt8420184/?ref_=chttvtp_t_19   542 reviews             28             1,40
## 39  https://imdb.com/title/tt0052520/?ref_=chttvtp_t_20   214 reviews             85               33
## 40  https://imdb.com/title/tt0052520/?ref_=chttvtp_t_20   214 reviews             85               33
## 41  https://imdb.com/title/tt1877514/?ref_=chttvtp_t_21   175 reviews             13             1,73
## 42  https://imdb.com/title/tt1877514/?ref_=chttvtp_t_21   175 reviews             13             1,73
## 43  https://imdb.com/title/tt1475582/?ref_=chttvtp_t_22 1,098 reviews            121               10
## 44  https://imdb.com/title/tt1475582/?ref_=chttvtp_t_22 1,098 reviews            121               10
## 45  https://imdb.com/title/tt2560140/?ref_=chttvtp_t_23 2,370 reviews             64                5
## 46  https://imdb.com/title/tt2560140/?ref_=chttvtp_t_23 2,370 reviews             64                5
## 47  https://imdb.com/title/tt0103359/?ref_=chttvtp_t_24   219 reviews             25               40
## 48  https://imdb.com/title/tt0103359/?ref_=chttvtp_t_24   219 reviews             25               40
## 49 https://imdb.com/title/tt11126994/?ref_=chttvtp_t_25 2,217 reviews             59
## 50 https://imdb.com/title/tt11126994/?ref_=chttvtp_t_25 2,217 reviews             59
```

```
# Combine the data frames
shows <- cbind(top_tv_shows, show_url_df)
```

```
shows
```

```
##    Rank                            Title Rating Voters Episodes Year
## 1     1                     Breaking Bad    9.5   2.2M       NA <NA>   https://imdb.com/title/tt0903
## 2     2                  Planet Earth II    9.5   163K       NA <NA>   https://imdb.com/title/tt0903
## 3     3                     Planet Earth    9.4   224K       NA <NA>   https://imdb.com/title/tt5491
## 4     4                 Band of Brothers    9.4   547K       NA <NA>   https://imdb.com/title/tt5491
## 5     5                        Chernobyl    9.3   910K       NA <NA>   https://imdb.com/title/tt0795
## 6     6                         The Wire    9.3   392K       NA <NA>   https://imdb.com/title/tt0795
## 7     7       Avatar: The Last Airbender    9.3   391K       NA <NA>   https://imdb.com/title/tt0185
## 8     8                   Blue Planet II    9.3    49K       NA <NA>   https://imdb.com/title/tt0185
## 9     9                     The Sopranos    9.2   500K       NA <NA>   https://imdb.com/title/tt7366
## 10   10      Cosmos: A Spacetime Odyssey    9.2   132K       NA <NA>   https://imdb.com/title/tt7366
## 11   11                           Cosmos    9.3    46K       NA <NA>   https://imdb.com/title/tt0306
## 12   12                       Our Planet    9.2    54K       NA <NA>   https://imdb.com/title/tt0306
## 13   13                  Game of Thrones    9.2   2.4M       NA <NA>   https://imdb.com/title/tt0417
## 14   14                            Bluey    9.3    34K       NA <NA>   https://imdb.com/title/tt0417
## 15   15                 The World at War    9.2    32K       NA <NA>   https://imdb.com/title/tt6769
## 16   16 Fullmetal Alchemist: Brotherhood    9.1   210K       NA <NA>   https://imdb.com/title/tt6769
## 17   17                   Rick and Morty    9.1   629K       NA <NA>   https://imdb.com/title/tt0141
## 18   18                             Life    9.1    44K       NA <NA>   https://imdb.com/title/tt0141
## 19   19                   The Last Dance    9.0   160K       NA <NA>  https://imdb.com/title/tt23956
## 20   20                The Twilight Zone    9.0    97K       NA <NA>  https://imdb.com/title/tt23956
## 21   21                  The Vietnam War    9.1    30K       NA <NA>  https://imdb.com/title/tt00818
## 22   22                         Sherlock    9.1     1M       NA <NA>  https://imdb.com/title/tt00818
## 23   23                  Attack on Titan    9.1   565K       NA <NA>  https://imdb.com/title/tt92538
## 24   24      Batman: The Animated Series    9.0   123K       NA <NA>  https://imdb.com/title/tt92538
## 25   25                           Arcane    9.0   332K       NA <NA>  https://imdb.com/title/tt09449
## 26    1                     Breaking Bad    9.5   2.2M       NA <NA>  https://imdb.com/title/tt09449
## 27    2                  Planet Earth II    9.5   163K       NA <NA>  https://imdb.com/title/tt76786
## 28    3                     Planet Earth    9.4   224K       NA <NA>  https://imdb.com/title/tt76786
## 29    4                 Band of Brothers    9.4   547K       NA <NA>  https://imdb.com/title/tt00710
## 30    5                        Chernobyl    9.3   910K       NA <NA>  https://imdb.com/title/tt00710
## 31    6                         The Wire    9.3   392K       NA <NA>  https://imdb.com/title/tt13556
## 32    7       Avatar: The Last Airbender    9.3   391K       NA <NA>  https://imdb.com/title/tt13556
## 33    8                   Blue Planet II    9.3    49K       NA <NA>  https://imdb.com/title/tt28614
## 34    9                     The Sopranos    9.2   500K       NA <NA>  https://imdb.com/title/tt28614
## 35   10      Cosmos: A Spacetime Odyssey    9.2   132K       NA <NA>  https://imdb.com/title/tt15333
## 36   11                           Cosmos    9.3    46K       NA <NA>  https://imdb.com/title/tt15333
## 37   12                       Our Planet    9.2    54K       NA <NA>  https://imdb.com/title/tt84201
## 38   13                  Game of Thrones    9.2   2.4M       NA <NA>  https://imdb.com/title/tt84201
## 39   14                            Bluey    9.3    34K       NA <NA>  https://imdb.com/title/tt00525
## 40   15                 The World at War    9.2    32K       NA <NA>  https://imdb.com/title/tt00525
## 41   16 Fullmetal Alchemist: Brotherhood    9.1   210K       NA <NA>  https://imdb.com/title/tt18775
## 42   17                   Rick and Morty    9.1   629K       NA <NA>  https://imdb.com/title/tt18775
## 43   18                             Life    9.1    44K       NA <NA>  https://imdb.com/title/tt14755
## 44   19                   The Last Dance    9.0   160K       NA <NA>  https://imdb.com/title/tt14755
## 45   20                The Twilight Zone    9.0    97K       NA <NA>  https://imdb.com/title/tt25601
## 46   21                  The Vietnam War    9.1    30K       NA <NA>  https://imdb.com/title/tt25601
## 47   22                         Sherlock    9.1     1M       NA <NA>  https://imdb.com/title/tt01033
## 48   23                  Attack on Titan    9.1   565K       NA <NA>  https://imdb.com/title/tt01033
## 49   24      Batman: The Animated Series    9.0   123K       NA <NA> https://imdb.com/title/tt111269
## 50   25                           Arcane    9.0   332K       NA <NA> https://imdb.com/title/tt111269
```

```r
#2.
# Define URL for Breaking Bad
BreakingBad_urls <- "https://www.imdb.com/title/tt0903747/reviews/?ref_=tt_ov_urv"

# Initialize list to store data frames
df <- list()
df_names <- "Breaking_Bad"

# Read HTML session for the current URL
session <- read_html(BreakingBad_urls)

# Scrape reviewer names
reviewer_name <- session %>%
  html_nodes(".ipc-link.ipc-link--base") %>%
  html_text() %>%
  head(20)

# Scrape review dates
review_date <- session %>%
  html_nodes(".ipc-inline-list__item.review-date") %>%
  html_text() %>%
  head(20)

# Scrape user ratings (update CSS selector)
user_rating <- session %>%
  html_nodes(".ipc-rating-star--rating") %>%  # Example selector, verify it in the HTML
  html_text() %>%
  head(20)

# Scrape reviews' titles
review_title <- session %>%
  html_nodes(".ipc-title__text") %>%
  html_text() %>%
  head(20)

# Scrape helpful reviews
helpful_reviews <- session %>%
  html_nodes(".ipc-voting__label__count.ipc-voting__label__count--up") %>%
  html_text() %>%
  head(20)

# Scrape not helpful reviews
not_helpful_reviews <- session %>%
  html_nodes(".ipc-voting__label__count.ipc-voting__label__count--down") %>%
  html_text() %>%
  head(20)

# Scrape text reviews
text_reviews <- session %>%
  html_nodes(".ipc-html-content-inner-div") %>%
  html_text() %>%
  head(20)
```

```r
# Ensure each column has exactly 20 entries, filling with NA if fewer than 20 were scraped
reviewer_name <- c(reviewer_name, rep(NA, 20 - length(reviewer_name)))[1:20]
review_date <- c(review_date, rep(NA, 20 - length(review_date)))[1:20]
user_rating <- c(user_rating, rep(NA, 20 - length(user_rating)))[1:20]
review_title <- c(review_title, rep(NA, 20 - length(review_title)))[1:20]
helpful_reviews <- c(helpful_reviews, rep(NA, 20 - length(helpful_reviews)))[1:20]
not_helpful_reviews <- c(not_helpful_reviews, rep(NA, 20 - length(not_helpful_reviews)))[1:20]
text_reviews <- c(text_reviews, rep(NA, 20 - length(text_reviews)))[1:20]

# Create a temporary data frame for the current URL
dfTemp <- data.frame(
  reviewer_name = reviewer_name,
  review_date = review_date,
  user_rating = user_rating,
  review_title = review_title,
  helpful_reviews = helpful_reviews,
  not_helpful_reviews = not_helpful_reviews,
  text_reviews = text_reviews,
  stringsAsFactors = FALSE

)

# Append the temporary data frame to the list with a custom name
df[[df_names]] <- dfTemp

# View the data frame for "Breaking Bad"
print(df$Breaking_Bad)
```

```
##         reviewer_name  review_date user_rating
## 1             FiRE010  Jul 3, 2021          10
## 2           Permalink  Mar 6, 2019          10
## 3          bruhperson Jul 29, 2021          10
## 4           Permalink Feb 18, 2020          10
## 5        KinoKoopaKid  Nov 8, 2021          10
## 6           Permalink May 30, 2019          10
## 7         jehuschultz  Dec 8, 2022          10
## 8           Permalink Nov 15, 2019          10
## 9      Supermanfan-13 Jul 17, 2021          10
## 10          Permalink  Dec 8, 2022          10
## 11  manishsingh-03299 Sep 28, 2024           7
## 12          Permalink Feb 14, 2021           5
## 13            Rob1331  Mar 5, 2014          10
## 14          Permalink Nov 12, 2017          10
## 15           xpinerhd Jan 11, 2014          10
## 16          Permalink  Nov 8, 2021          10 If you mix Scarface, Robin Hood and maybe Tyler Durde
## 17 dhanushreddy-14919 Aug 11, 2021          10
## 18          Permalink May 19, 2019          10
## 19        tushv-31482  May 4, 2021          10
## 20          Permalink Jun 23, 2021          10
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13 'Breaking Bad' is one of the most popular rated shows on IMDb, is one of those rarities where eve
## 14
## 15
## 16
## 17
## 18
## 19
## 20
```

```
# Define URL for Planet Earth II
PlanetEarthII_urls <- "https://www.imdb.com/title/tt5491994/reviews/?ref_=tt_ov_urv"

# Initialize list to store data frames
df <- list()
df_names <- "Planet_Earth_II"

# Read HTML session for the current URL
session <- read_html(PlanetEarthII_urls)

# Scrape reviewer names
reviewer_name <- session %>%
  html_nodes(".ipc-link.ipc-link--base") %>%
  html_text() %>%
  head(20)

# Scrape review dates
review_date <- session %>%
  html_nodes(".ipc-inline-list__item.review-date") %>%
  html_text() %>%
  head(20)

# Scrape user ratings (update CSS selector)
# First, inspect the correct selector for user rating from the page structure.
user_rating <- session %>%
  html_nodes(".ipc-rating-star--rating") %>%  # Adjust this selector if needed (check the page source)
  html_text() %>%
  head(20)

# Scrape reviews' titles
review_title <- session %>%
  html_nodes(".ipc-title__text") %>%
  html_text() %>%
  head(20)

# Scrape helpful reviews
```

```r
helpful_reviews <- session %>%
  html_nodes(".ipc-voting__label__count.ipc-voting__label__count--up") %>%
  html_text() %>%
  head(20)

# Scrape not helpful reviews
not_helpful_reviews <- session %>%
  html_nodes(".ipc-voting__label__count.ipc-voting__label__count--down") %>%
  html_text() %>%
  head(20)

# Scrape text reviews
text_reviews <- session %>%
  html_nodes(".ipc-html-content-inner-div") %>%
  html_text() %>%
  head(20)

# Handle case where some elements might be missing, ensuring we have exactly 20 entries
reviewer_name <- c(reviewer_name, rep(NA, 20 - length(reviewer_name)))[1:20]
review_date <- c(review_date, rep(NA, 20 - length(review_date)))[1:20]
user_rating <- c(user_rating, rep(NA, 20 - length(user_rating)))[1:20]
review_title <- c(review_title, rep(NA, 20 - length(review_title)))[1:20]
helpful_reviews <- c(helpful_reviews, rep(NA, 20 - length(helpful_reviews)))[1:20]
not_helpful_reviews <- c(not_helpful_reviews, rep(NA, 20 - length(not_helpful_reviews)))[1:20]
text_reviews <- c(text_reviews, rep(NA, 20 - length(text_reviews)))[1:20]

# Create a temporary data frame for the current URL
dfTemp <- data.frame(
  reviewer_name = reviewer_name,
  review_date = review_date,
  user_rating = user_rating,
  review_title = review_title,
  helpful_reviews = helpful_reviews,
  not_helpful_reviews = not_helpful_reviews,
  text_reviews = text_reviews,
  stringsAsFactors = FALSE
)

# Append the temporary data frame to the list with a custom name
df[[df_names]] <- dfTemp

# View the data frame for "Planet Earth II"
print(df$Planet_Earth_II)
```

```
##         reviewer_name  review_date user_rating
## 1        arjanhylkema  Nov 7, 2016          10
## 2           Permalink  Nov 5, 2016          10
## 3            Wentloog  Nov 5, 2016          10
## 4           Permalink  Nov 9, 2016          10                                                    Y
## 5       john-m-madsen  Nov 5, 2016          10
## 6           Permalink  Nov 8, 2016          10
## 7        thespookybuz Nov 17, 2016          10
## 8           Permalink Nov 13, 2016          10
## 9         pjdickinson  Nov 6, 2016          10
## 10          Permalink Dec 31, 2016          10
## 11           dbijis33 Nov 19, 2016          10
## 12          Permalink Dec 28, 2016           7
## 13     dhanrajjughead May 19, 2019          10
## 14          Permalink Oct 20, 2018          10
## 15        NeilBarnett Sep 29, 2017          10
## 16          Permalink Nov 22, 2016          10
## 17      salmanu-27386 Oct 12, 2017          10
## 18          Permalink  Dec 4, 2016          10 Like the first 'Planet Earth', does for nature and ou
## 19 panagiotiskatsanos Apr 23, 2020          10
## 20          Permalink  Jan 5, 2017          10
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17 Absolutely adore the first 'Planet Earth' from 2007, one of the best documentaries ever made and
## 18
## 19
## 20
it is one of the great achievements in any area of culture from this or any other year. Little else sho
```

```r
# Define URL for Planet Earth
PlanetEarth_urls <- "https://www.imdb.com/title/tt0795176/reviews/?ref_=tt_ov_urv"

# Initialize list to store data frames
df <- list()
df_names <- "Planet_Earth"

# Read HTML session for the current URL
session <- read_html(PlanetEarth_urls)

# Scrape reviewer names
reviewer_name <- session %>%
  html_nodes(".ipc-link.ipc-link--base") %>%
  html_text() %>%
  head(20)

# Scrape review dates
review_date <- session %>%
  html_nodes(".ipc-inline-list__item.review-date") %>%
  html_text() %>%
  head(20)

# Scrape user ratings (corrected CSS selector)
user_rating <- session %>%
  html_nodes(".ipc-rating-star--rating") %>%  # Adjust this selector if needed (inspect page for correc
  html_text() %>%
  head(20)

# Scrape reviews' titles
review_title <- session %>%
  html_nodes(".ipc-title__text") %>%
  html_text() %>%
  head(20)

# Scrape helpful reviews
helpful_reviews <- session %>%
  html_nodes(".ipc-voting__label__count.ipc-voting__label__count--up") %>%
  html_text() %>%
  head(20)

# Scrape not helpful reviews
not_helpful_reviews <- session %>%
  html_nodes(".ipc-voting__label__count.ipc-voting__label__count--down") %>%
  html_text() %>%
  head(20)

# Scrape text reviews
text_reviews <- session %>%
  html_nodes(".ipc-html-content-inner-div") %>%
  html_text() %>%
  head(20)


# Handle case where some elements might be missing, ensuring we have exactly 20 entries
reviewer_name <- c(reviewer_name, rep(NA, 20 - length(reviewer_name)))[1:20]
review_date <- c(review_date, rep(NA, 20 - length(review_date)))[1:20]
user_rating <- c(user_rating, rep(NA, 20 - length(user_rating)))[1:20]
review_title <- c(review_title, rep(NA, 20 - length(review_title)))[1:20]
helpful_reviews <- c(helpful_reviews, rep(NA, 20 - length(helpful_reviews)))[1:20]
not_helpful_reviews <- c(not_helpful_reviews, rep(NA, 20 - length(not_helpful_reviews)))[1:20]
text_reviews <- c(text_reviews, rep(NA, 20 - length(text_reviews)))[1:20]

# Create a temporary data frame for the current URL
dfTemp <- data.frame(
  reviewer_name = reviewer_name,
  review_date = review_date,
  user_rating = user_rating,
  review_title = review_title,
  helpful_reviews = helpful_reviews,
  not_helpful_reviews = not_helpful_reviews,
  text_reviews = text_reviews,
  stringsAsFactors = FALSE
)

# Append the temporary data frame to the list with a custom name
df[[df_names]] <- dfTemp

# View the data frame for "Planet Earth"
print(df$Planet_Earth)
```

```
##      reviewer_name  review_date user_rating
## 1     robert-kamer  Feb 8, 2007          10
## 2        Permalink Nov 19, 2008          10
## 3         jim-1409  Jan 4, 2009          10                                                    A masterpiece of a
## 4        Permalink Dec 15, 2006          10                                                                In A W
## 5  ccthemovieman-1  Sep 1, 2007          10    The most amazing achievement in natural history TV ha
## 6        Permalink Aug 27, 2006          10                                                              Simply p
## 7         cmcoveos Apr 30, 2006          10                                             An amazing trip around our beaut
## 8        Permalink Jun 29, 2015           9 A visually impressive and memorable look at the world tha
## 9         Loordssm Jul 20, 2006          10                                                 Is it real? I mean, actu
## 10       Permalink Jan 28, 2009          10
## 11        ultimorn  Jun 1, 2015           7                                                         Are you kiddin
## 12       Permalink  Oct 8, 2020           3                                                 It doesn't get any bette
## 13     bob the moo  Dec 4, 2007          10                                                        Only 4 Eps can to
## 14       Permalink Jan 15, 2007          10                                              Should be called "BBC - Yeah, a
## 15           alfeu Jul 30, 2008          10                                                         Brilliant Docume
## 16       Permalink Dec 25, 2017           9                                               Explanation to those low-ratin
## 17         Cabrone Sep 14, 2009          10                                                                    Truly
## 18       Permalink May 31, 2020          10                                                           The Greatest
## 19         berndt65 Jul 27, 2014           9
## 20       Permalink Sep 20, 2020           9                                                             Absolutely
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7  As the influence of man expands across the globe, fewer and fewer truly untouched wilderness exist
## 8
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
```

```r
# Define URL for Band Of Brothers
BandOfBrothers_urls <- "https://www.imdb.com/title/tt0185906/reviews/?ref_=tt_ov_urv"

# Initialize list to store data frames
df <- list()
df_names <- "Band_Of_Brothers"

# Read HTML session for the current URL
session <- read_html(BandOfBrothers_urls)

# Scrape reviewer names
reviewer_name <- session %>%
  html_nodes(".ipc-link.ipc-link--base") %>%
  html_text() %>%
  head(20)

# Scrape review dates
review_date <- session %>%
  html_nodes(".ipc-inline-list__item.review-date") %>%
  html_text() %>%
  head(20)

# Scrape user ratings (corrected CSS selector)
user_rating <- session %>%
  html_nodes(".ipc-rating-star--rating") %>%
  html_text() %>%
  head(20)

# Scrape reviews' titles
review_title <- session %>%
  html_nodes(".ipc-title__text") %>%
  html_text() %>%
  head(20)

# Scrape helpful reviews
helpful_reviews <- session %>%
  html_nodes(".ipc-voting__label__count.ipc-voting__label__count--up") %>%
  html_text() %>%
  head(20)

# Scrape not helpful reviews
not_helpful_reviews <- session %>%
  html_nodes(".ipc-voting__label__count.ipc-voting__label__count--down") %>%
  html_text() %>%
  head(20)

# Scrape text reviews
text_reviews <- session %>%
  html_nodes(".ipc-html-content-inner-div") %>%
  html_text() %>%
  head(20)
```

```r
# Handle case where some elements might be missing, ensuring we have exactly 20 entries
reviewer_name <- c(reviewer_name, rep(NA, 20 - length(reviewer_name)))[1:20]
review_date <- c(review_date, rep(NA, 20 - length(review_date)))[1:20]
user_rating <- c(user_rating, rep(NA, 20 - length(user_rating)))[1:20]
review_title <- c(review_title, rep(NA, 20 - length(review_title)))[1:20]
helpful_reviews <- c(helpful_reviews, rep(NA, 20 - length(helpful_reviews)))[1:20]
not_helpful_reviews <- c(not_helpful_reviews, rep(NA, 20 - length(not_helpful_reviews)))[1:20]
text_reviews <- c(text_reviews, rep(NA, 20 - length(text_reviews)))[1:20]

# Create a temporary data frame for the current URL
dfTemp <- data.frame(
  reviewer_name = reviewer_name,
  review_date = review_date,
  user_rating = user_rating,
  review_title = review_title,
  helpful_reviews = helpful_reviews,
  not_helpful_reviews = not_helpful_reviews,
  text_reviews = text_reviews,
  stringsAsFactors = FALSE
)

# Append the temporary data frame to the list with a custom name
df[[df_names]] <- dfTemp

# View the data frame for "band of brothers"
print(df$Band_Of_Brothers)
```

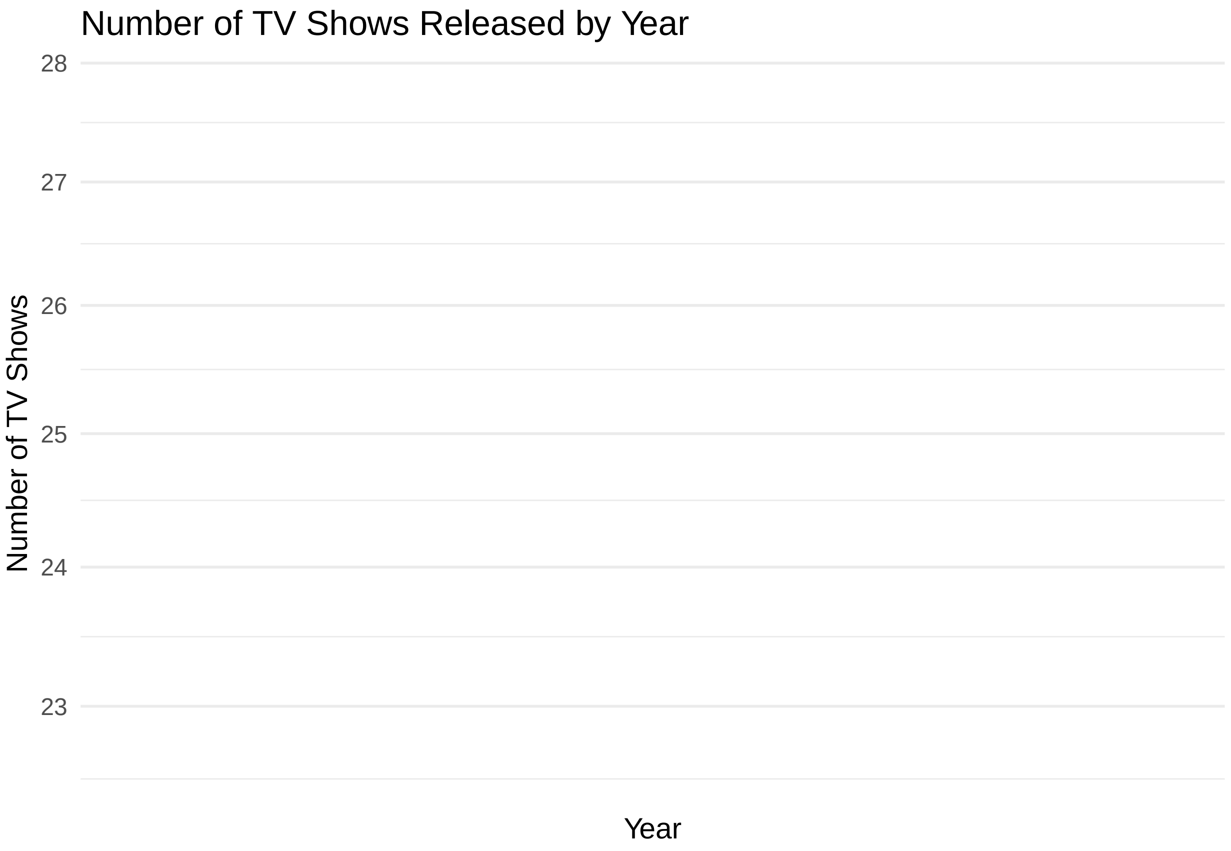
```
##         reviewer_name  review_date user_rating
## 1             Rob1331 Sep 27, 2022          10
## 2           Permalink Oct 14, 2001          10
## 3        sanderson777 Jan 18, 2002          10                          Possibly the finest 10 hours
## 4           Permalink Apr 18, 2004          10                             One of the best war movie
## 5         wildcatt268 Feb 13, 2003          10
## 6           Permalink Jan 23, 2005          10
## 7             arjay24 Sep 16, 2004          10                    One of, if not the best, mini seri
## 8           Permalink  May 6, 2022          10            This series is so unbelievably realistic, 
## 9           rbverhoef  Nov 4, 2019          10                          One of the best mini-series 
## 10          Permalink  Nov 5, 2001          10                                             Probably 
## 11         yodaschoda Aug 25, 2004          10                           Realistic WWII Drama With Wa
## 12          Permalink May 30, 2015           7                                                   wa
## 13 philip_vanderveken Apr 10, 2021           5                                            You can't 
## 14          Permalink  May 2, 2006          10
## 15     Supermanfan-13  Jun 3, 2019          10                                          Not very real
## 16          Permalink Jan 26, 2005          10                  Without Doubt, the Best Mini-Series 
## 17          thiagoutp  May 3, 2022          10                                                   Grea
## 18          Permalink Oct 24, 2018           9 A series like this won't be made again (see below), s
## 19         bsmith5552  Dec 7, 2002          10                                           Share With 
## 20          Permalink Nov 25, 2002          10                                              Best Min
##
## 1
## 2
## 3
## 4
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13
## 14 Lots of people applaud this series for its realism, but I can't really agree. I think there is st
## 15
## 16
## 17
## 18
## 19
## 20
```

```r
# Define URL for Chernobyl
Chernobyl_urls <- "https://www.imdb.com/title/tt7366338/reviews/?ref_=tt_ov_urv"

# Initialize list to store data frames
df <- list()
df_names <- "Chernobyl"

# Read HTML session for the current URL
session <- read_html(Chernobyl_urls)

# Scrape reviewer names
reviewer_name <- session %>%
  html_nodes(".ipc-link.ipc-link--base") %>%
  html_text() %>%
  head(20)

# Scrape review dates
review_date <- session %>%
  html_nodes(".ipc-inline-list__item.review-date") %>%
  html_text() %>%
  head(20)

# Scrape user ratings (corrected CSS selector)
user_rating <- session %>%
  html_nodes(".ipc-rating-star--rating") %>%
  html_text() %>%
  head(20)

# Scrape reviews' titles
review_title <- session %>%
  html_nodes(".ipc-title__text") %>%
  html_text() %>%
  head(20)

# Scrape helpful reviews
helpful_reviews <- session %>%
  html_nodes(".ipc-voting__label__count.ipc-voting__label__count--up") %>%
  html_text() %>%
  head(20)

# Scrape not helpful reviews
not_helpful_reviews <- session %>%
  html_nodes(".ipc-voting__label__count.ipc-voting__label__count--down") %>%
  html_text() %>%
  head(20)

# Scrape text reviews
text_reviews <- session %>%
  html_nodes(".ipc-html-content-inner-div") %>%
  html_text() %>%
  head(20)

# Handle case where some elements might be missing, ensuring we have exactly 20 entries
reviewer_name <- c(reviewer_name, rep(NA, 20 - length(reviewer_name)))[1:20]
review_date <- c(review_date, rep(NA, 20 - length(review_date)))[1:20]
user_rating <- c(user_rating, rep(NA, 20 - length(user_rating)))[1:20]
review_title <- c(review_title, rep(NA, 20 - length(review_title)))[1:20]
helpful_reviews <- c(helpful_reviews, rep(NA, 20 - length(helpful_reviews)))[1:20]
not_helpful_reviews <- c(not_helpful_reviews, rep(NA, 20 - length(not_helpful_reviews)))[1:20]
text_reviews <- c(text_reviews, rep(NA, 20 - length(text_reviews)))[1:20]

# Create a temporary data frame for the current URL
dfTemp <- data.frame(
  reviewer_name = reviewer_name,
  review_date = review_date,
  user_rating = user_rating,
  review_title = review_title,
  helpful_reviews = helpful_reviews,
  not_helpful_reviews = not_helpful_reviews,
  text_reviews = text_reviews,
  stringsAsFactors = FALSE
)

# Append the temporary data frame to the list with a custom name
df[[df_names]] <- dfTemp

# View the data frame for "Chernobyl"
print(df$Chernobyl)
```

```
##        reviewer_name  review_date user_rating
## 1    curiosityonmars May 23, 2019         10
## 2          Permalink May 10, 2019         10                                                    They
## 3           stelmakh  May 9, 2019         10                                                 Goosebu
## 4          Permalink May 14, 2019         10                                        I highly recomme
## 5       natashapekar  May 7, 2019         10                                      No hero wakes up
## 6          Permalink May 20, 2019         10                                               So far l
## 7         m-porpaczi  May 6, 2019         10
## 8          Permalink May 13, 2019         10                           Bleak, Unsettling, Haunting
## 9           Lladerat  May 6, 2019         10
## 10         Permalink Nov 27, 2019         10                                                    HBO
## 11          jfirebug May 23, 2019          5
## 12         Permalink Jun 15, 2019          8
## 13           thegldt May 20, 2019         10 The movie is far from thuth. A lot of fake info to crea
## 14         Permalink May 30, 2019         10                                              Emotional
## 15 alexander-phoenix  Jun 7, 2019          9                                                   Just
## 16         Permalink Sep 27, 2022         10                             Now you look like the mini
## 17       wmeduardowm  May 6, 2019          9
## 18         Permalink Jul 10, 2022          9
## 19    Leofwine_draca May 26, 2019         10
## 20         Permalink May 15, 2019          7              It is hard to overestimate the importance
##
## 1
## 2
## 3
## 4  As my mother tells it, the weather was quite nice, the sky was clear without any sign of clouds i
## 5
## 6
## 7
## 8
## 9
## 10
## 11
## 12
## 13
## 14
## 15
## 16
## 17
## 18
## 19
## 20
```

```r
#3.

# Convert the 'Year' column to numeric if it isn't already
top_tv_shows$Year <- as.numeric(top_tv_shows$Year)

# Group the data by Year and count the number of shows per year
shows_by_year <- top_tv_shows %>%
  group_by(Year) %>%
  summarise(Count = n())

# Plot the number of shows released by year
ggplot(shows_by_year, aes(x = Year, y = Count)) +
  geom_line(color = "skyblue", size = 1) +
  geom_point(color = "pink", size = 2) +
  labs(title = "Number of TV Shows Released by Year",
       x = "Year",
       y = "Number of TV Shows") +
  scale_y_log10() +  # Use log scale for y-axis
  theme_minimal()
```

```
## Warning: Removed 1 row containing missing values or values outside the scale range (`geom_line()`).
## Warning: Removed 1 row containing missing values or values outside the scale range (`geom_point()`).
```



```r
# Find the year with the most TV shows released
most_shows_year <- shows_by_year %>%
  filter(Count == max(Count))

# Print the year with the most releases
print(most_shows_year)
```

```
## # A tibble: 1 x 2
##    Year Count
##   <dbl> <int>
## 1    NA    25
```

2. Extracting Amazon Product Reviews

```r
#4. URLs
urls <- c('https://www.amazon.com/s?k=Laptop&crid=2ZK2KVH3ZE2VN&sprefix=laptop%2Caps%2C343&ref=nb_sb_nos
  'https://www.amazon.com/s?k=lamp&crid=RZQJTQH3PPFC&sprefix=flower+lamp%2Caps%2C334&ref=nb_sb_noss_1',
  'https://www.amazon.com/s?k=Flower&crid=25Z6A4YCCPMFC&sprefix=flower%2Caps%2C404&ref=nb_sb_noss_1',
  'https://www.amazon.com/s?k=cologne&crid=2JGLYDYZR9C1B&sprefix=colognr%2Caps%2C400&ref=nb_sb_ss_ts-doa
  'https://www.amazon.com/s?k=sweatshirt&crid=YTFRIOG9NESF&sprefix=sweatsh%2Caps%2C503&ref=nb_sb_ss_w_hi
```

```r
#5
df <- list()

for (i in seq_along(urls)) {

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

  product_name <- scrape(session) %>%
    html_nodes('h2.a-size-mini') %>%
    html_text() %>%
    head(30)


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


  product_rating <- scrape(session) %>%
    html_nodes('span.a-icon-alt') %>%
    html_text() %>%
    head(30)
  ratings <- as.numeric(str_extract(product_rating, "\\d+\\.\\d"))


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


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


  dfTemp <- data.frame(Product_Name = product_name[1:30],
                       Description = product_description[1:30],
                       Rating = ratings[1:30],
                       Price = price[1:30],
                       stringsAsFactors = FALSE)

  df[[i]] <- dfTemp
}

print(df[[1]])
```

```
##
## 1  HP Stream 14" HD BrightView Laptop, Intel Celeron N4120, 16GB RAM, 288GB Storage (128GB eMMC + 16
## 2                  15.6" FHD IPS Business Laptop, 32GB RAM 1TB SSD, AMD Ryzen 5(Beat i3-1215U, 4 Cor
## 3                                                                                 ASUS 15.6" Vivobook
## 4            acer Aspire 3 A315-24PT-R0UX Slim Laptop, 15.6" Full HD IPS Touch Display, AMD Ryzen 5
## 5                                             HP 14 inch Laptop, HD Display, 12th Generation Intel C
## 6                                                                       Lenovo V15 G2 Laptop, 15.6"
## 7     Lenovo V15 Newest 15.6" FHD Business Laptop Computer, Intel Dual-Core Processor, 32GB DDR4 RAM
## 8         Apple 2024 MacBook Air 13-inch Laptop with M3 chip: Built for Apple Intelligence, 13.6-inc
## 9            HP Newest 255 G10 Laptop for Home or Work, 16GB RAM, 1TB SSD, 15.6" Full HD, Ryzen 3 73
## 10                                                                     ASUS Chromebook CX1, 14" FHD,
## 11                                                         HP Lastest 255 G10 15.6" FHD Business Lap
## 12                                              ASUS Chromebook CX1, 15.6" FHD NanoEdge Display, Inte
## 13      HP Flagship Touchscreen Laptop Computer • Win 11 Pro • 2024 Office Lifetime • 15.6" Full HD
## 14         Acer Aspire Go 15 Slim Laptop | 15.6" Full HD IPS 1080P Display | Intel Core i3-N305| Inte
## 15                                     HP 17 Laptop, 17.3" HD+ Touchscreen Display, 12th Gen Intel C
## 16                                                   15.6 FHD IPS Laptop Computer Student, 16GB RAM
## 17
## 18
## 19
## 20
## 21
## 22
## 23
## 24
## 25
## 26
## 27
## 28
## 29
## 30
##     Price
## 1  262.65
## 2  399.00
## 3  379.00
## 4  499.99
## 5  219.99
## 6  457.00
## 7  519.99
## 8  249.99
## 9  318.00
## 10 331.55
## 11 469.00
## 12 289.60
## 13 359.00
## 14 844.00
## 15  99.00
## 16 399.72
## 17 599.99
## 18 156.50
## 19 209.99
## 20 799.99
## 21 799.98
## 22 169.99
## 23 249.99
## 24 499.99
## 25 799.99
## 26 276.88
## 27 299.99
## 28 847.99
## 29 999.00
## 30 299.99
```

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

```
##
## 1  Govee RGBIC Floor Lamp, LED Corner Lamp Works with Alexa, 1000 Lumen Smart Modern Floor Lamp with
## 2                               Modern Floor Lamp Simple Design with White Shade, Foot Pedal Switch,
## 3                                              SUNMORY Floor Lamp with Shelves, Modern Square Standi
## 4              ROTTOGOON Floor Lamp for Living Room with 3 Color Temperatures LED Bulb, Standing Lam
## 5                    Bedside Table Lamp for Bedroom - 3 Way Dimmable Touch Lamp USB C Charging Ports
## 6       Ambimall 60" Floor Lamps for Living Room with 3 Color Temperatures, Standing Lamp Tall with
## 7    Govee Floor Lamp 2 with Matter, RGBIC Warm Cool White LED Floor Lamp Works with Alexa, 1725lm C
## 8                          Small Table Lamp for Bedroom - Bedside Lamps for Nightstand, Minimalist N
## 9                                        AKASUKI Table Lamp for Bedroom, 3 Way Dimmable Touch Lamp w
## 10   Small Bedside Table lamp for Bedroom - 3 Color Temperatures Bedside Lamps with USB C and A Ports
## 11            Upgraded  Dimmable Floor Lamp, 1000 Lumens LED Edison Bulb Included, Arc Floor Lamps fo
## 12
## 13  luckystyle Floor Lamp,Super Bright Dimmable LED Lamps for Living Room, Custom Color Temperature
## 14                                                      white crown LED Desk Lamp Dimmable Table Lam
## 15                 Philips Hue Go Smart Portable Dimmable Table Lamp, White - White and Color Ambian
## 16         Set of 2 Industrial Table Lamps with 2 USB Port, Gold Fully Stepless Dimmable Bedside Lam
## 17             Touch Bedside Table Lamp, [Sleek Design & RGB Mode] 3 Way Dimmable Small Lamp for Bed
## 18                                          Dott Arts Table Lamp for Bedroom, 3-Color Bedside Lamps
## 19                     addlon Floor Lamps for Living Room Bright Lighting with Glass lampshade, Mode
## 20                                                              addlon Floor Lamp with Shelves, 4-Ti
## 21     Verilux SmartLight Full Spectrum LED Modern Floor Lamp with Adjustable Brightness, Flexible G
## 22                              PESRAE Floor Lamp with Table, End Table with Charging Station(USB &
## 23                                                 BRIGHTWORLD Moon Lamp Galaxy Lamp 5.9 inch 16 Col
## 24                    Philips Hue Signe Smart Table Lamp, Black - White and Color Ambiance LED Color
## 25                                                  Philips Hue Bloom White and Color Ambiance Smart
## 26      Govee RGBIC Smart Table Lamp 2, Touch Bedside Small Table Lamp Work with Matter, Alexa, LED
## 27                           Floor Lamp, 15w/1000lm Bright LED Floor Lamp with Stepless Adjustable 3
## 28                     Small Bedside Night Table Lamp for Bedroom, Minimalist Nightstand Lamp with R
## 29  Frideko 2 Pack Floor Lamps for Living Room, Dimmable Colors Temperature & Brightness, 68.11" Mod
## 30       Govee Floor Lamp Pro, RGBIC LED Lamp with Bluetooth Speaker and White Noise, 2100 Lumens St
##    Rating  Price
## 1     4.5  35.54
## 2     4.5  42.99
## 3     4.6  39.99
## 4     4.5  59.99
## 5     4.6  49.99
## 6     4.5  54.99
## 7     4.6  99.99
## 8     4.6  19.99
## 9     4.5  23.59
## 10    4.6  35.99
## 11    4.4  45.99
## 12    4.5  29.98
## 13    4.5  69.99
## 14    4.3  19.19
## 15    4.6  23.39
## 16    4.6  29.98
## 17    4.6  38.99
## 18    4.5  99.99
## 19    4.5 149.99
## 20    4.6   9.48
## 21    4.6  15.99
## 22    4.6  18.16
## 23    4.5  22.99
## 24    4.5  17.59
## 25    4.6  21.99
## 26    4.5  39.98
## 27    4.6  79.99
## 28    4.5  16.99
## 29    4.2  29.99
## 30    4.5  33.99
```

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

```
##
## 1                                    Benchmark Bouquets, Flowering Fields, Glass Vase Included, Gift
## 2     Birthday Gifts For Women,Rose Flower Gifts for Mom,Womens Gifts for Christmas,Colorful Rainbow
## 3                       Floroom Artificial Flowers 25pcs Real Looking Ivory Foam Fake Roses with Ste
## 4                     Benchmark Bouquets, 20 Stem Rainbow Mini Carnations, Glass Vase Included, Gift
## 5             lovenfold Flowers for Delivery Prime,Preserved Flowers Bouquets,Purple Rose Bouquets T
## 6                                   LEGO Roses - Building Toy for Kids, Girls & Boys, Ages 8+ - Arti
## 7                       Benchmark Bouquets, Charming Roses & Alstroemeria, Glass Vase Included, Gift
## 8                                 Benchmark Bouquets, Pink Roses & Lilies, Glass Vase Included, Gift
## 9         12 Assorted Roses and Greens with Vase - Holidays Collection - KaBloom Prime Next Day - Gi
## 10                                                                                      What Do You
## 11                                       Benchmark Bouquets, Red Elegance, Glass Vase Included, Gift
## 12           LEGO Icons Plum Blossom Building Set for Adults Ages 18+ - Artificial Flowers for Dinne
## 13 Preserved Flowers Bouquets, Purple Flowers, Flowers for Delivery Prime, Natural Real Long Lasting
## 14                                 Benchmark Bouquets, 10 Stem Sunflowers, Glass Vase Included, Gift
## 15                              Benchmark Bouquets, Orange Roses & Lilies, Glass Vase Included, Gift
## 16                                      Benchmark Bouquets, Pink Elegance, Glass Vase Included, Gift
## 17                              Benchmark Bouquets, 24 Stem Rainbow Roses, Glass Vase Included, Gift
## 18                                      Benchmark Bouquets, Joyful Wishes, Glass Vase Included, Gift
## 19                                   Benchmark Bouquets, Rays of Sunshine, Glass Vase Included, Gift
## 20                                Benchmark Bouquets, Life is Good Orange, Glass Vase Included, Gift
## 21                     Benchmark Bouquets, Signature Roses & Alstroemeria, Glass Vase Included, Gift
## 22                                  Benchmark Bouquets, 24 Stem Red Roses, Glass Vase Included, Gift
## 23                            Benchmark Bouquets, 25 Stem Peruvian Lilies, Glass Vase Included, Gift
## 24                      LEGO Cherry Blossom - Building Toy for Kids, Girls & Boys, Ages 8+ - Artifici
## 25                                         Benchmark Bouquets, Big Blooms, Glass Vase Included, Gift
## 26                      Benchmark Bouquets, Elegance Roses & Alstroemeria, Glass Vase Included, Gift
## 27                                                                         Pumpkin Spice | Fresh Fal
## 28                             Benchmark Bouquets, 20 Stem Rainbow Tulips, Glass Vase Included, Gift
## 29                                     Benchmark Bouquets, White Elegance, Glass Vase Included, Gift
## 30                     LEGO Icons Orchid Artificial Plant, Building Set with Flowers, Home Décor Gif
##    Price
## 1  43.95
## 2  43.95
## 3  11.04
## 4  12.99
## 5  12.33
## 6   0.49
## 7  15.99
## 8  35.95
## 9  35.95
## 10 36.79
## 11 45.99
## 12 13.49
## 13 14.99
## 14 43.95
## 15 43.95
## 16 43.95
## 17 38.21
## 18 38.21
## 19 19.99
## 20 40.95
## 21 40.95
## 22 23.99
## 23 29.99
## 24 29.56
## 25 36.95
## 26 31.95
## 27 31.95
## 28 43.95
## 29 43.95
## 30 40.95
```

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

```
## 
## 1                             Azzaro The Most Wanted Eau de Parfum Intense - Woody & Seductive Mens C
## 2                                               Azzaro The Most Wanted Parfum - Intense Mens Cologne
## 3                                               Voyage Eau De Toilette for Men - Fresh, Romantic, Fru
## 4                                                               Cupid Hypnosis 2.0 Cologne For Men (
## 5                                                                                       RawChemistry
## 6                                                                                  RawChemistry Midni
## 7                                               D.Sauvage for Men Eau De Parfum Spray,Spray Refreshin
## 8                                                                                 Armani Beauty- Code
## 9                             Azzaro The Most Wanted Eau de Parfum Intense - Woody & Seductive Mens C
## 10
## 11                                                         Ralph Lauren - Ralph's Club - Eau de Parfu
## 12
## 13
## 14                                                                       Ralph Lauren - Polo Blue -
## 15
## 16
## 17
## 18               Mercedes-Benz Cologne for Men - 4 Fl Oz - Eau De Toilette Perfume Spray for Man - Wo
## 19                                                               Cremo Cologne for Men - Bourbon & Oa
## 20 RASASI Hawas For Men - Eau de Parfum Spray 100ML (3.4 Oz), Premium Arabian Perfume for Men, A Sig
## 21                                                             Armani Beauty - Code - Eau de Parfum -
## 22
## 23
## 24                                                                                               LUC
## 25                                             NAUTICA Classic Eau de Toilette for Men - Citrusy and
## 26                                           Armani Beauty - Acqua di Giò - Eau de Toilette - Cologn
## 27
## 28                                                                                              Men'
## 29                                                          Armani Beauty - Stronger With You Intens
## 30                                                            Ralph Lauren - Polo Black - Eau de Toi
##     Price
## 1   84.00
## 2  120.00
## 3  100.00
## 4  125.00
## 5   12.52
## 6    3.79
## 7   14.74
## 8   20.79
## 9   12.23
## 10  25.99
## 11  15.60
## 12  45.88
## 13  19.50
## 14  15.60
## 15  45.88
## 16  19.50
## 17  70.13
## 18  20.63
## 19  87.66
## 20  73.50
## 21  29.40
## 22 105.00
## 23  70.00
## 24 100.00
## 25  22.98
## 26   6.76
## 27  45.50
## 28  45.50
## 29  65.00
## 30  25.33
```

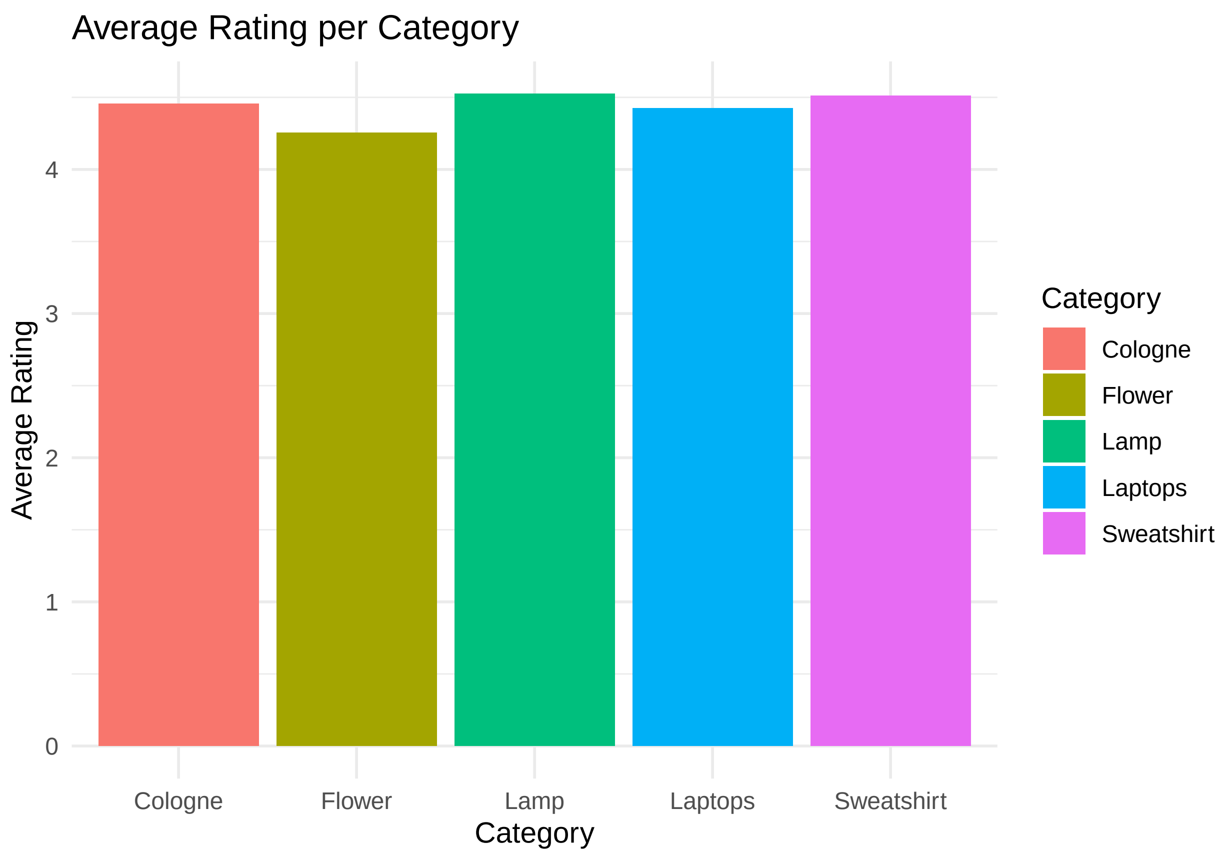
```r
print(df[[5]])
```

```
##
## 1
## 2                                                                   Men's Loose Fit Midweight Logo Sl
## 3
## 4                                          EcoSmart Fleece, Cotton-Blend Pullover, Crewneck Sweatshir
## 5
## 6                      Womens Oversized Hoodies Fleece Sweatshirts Long Sleeve Sweaters Pullover Fall
## 7
## 8                                      Men's Eversoft Fleece Sweatshirts, Moisture Wicking & Breathab
## 9
## 10                                                         Unisex Adult Fleece Crewneck Sweatshirt,
## 11
## 12                                                         Men's Rain Defender Loose Fit Heavyweight
## 13
## 14             Womens Sweatshirts Half Zip Cropped Pullover Fleece Quarter Zipper Hoodies 2024 Fall I
## 15
## 16                                 Men's Eversoft Fleece Hoodies, Moisture Wicking & Breathable, Full
## 17
## 18                                            Powerblend, Fleece Comfortable Hoodie, Sweatshirt for Me
## 19
## 20 Womens Oversized Sweatshirts Hoodies Fleece Crewneck Sweaters Casual Tops Comfy Fall Fashion Outfi
## 21
## 22       Womens Hoodies Oversized Sweatshirts Pullover Fleece Sweaters Long Sleeve Winter Fall Outfi
## 23
## 24                  Womens Long Sleeve Crop Tops Crewneck Pullover Shirts Lightweight Sweatshirts 202
## 25
## 26                                                           Men's Hoodie, EcoSmart Fleece Hoodie, Hoo
## 27
## 28                                          Women's Ecosmart V-Notch Crewneck Sweatshirt, Fleece Pullove
## 29
## 30                              Oversized Hoodies for Women Pullover Fleece Fall Sweatshirts Fashion Y2
```

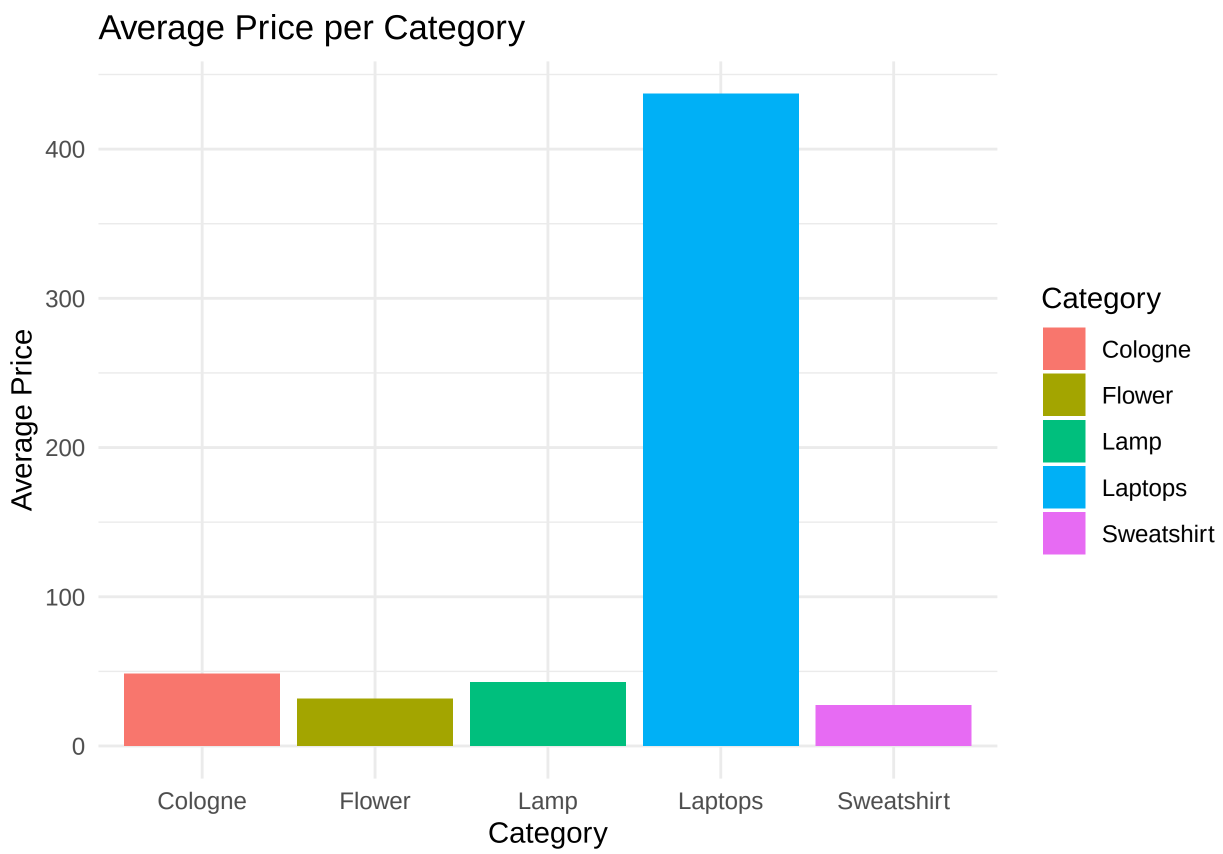
```r
#6.

#The code extracts data from Amazon product listing pages based on different search queries, such as "L
```

```r
#7

#This data can be used to compare product popularity, analyze price trends, examine the relationship be
```

```r
#8
combined_df <- do.call(rbind, df)
combined_df$Category <- rep(c("Laptops", "Lamp", "Flower", "Cologne", "Sweatshirt"), each = 30)

avg_rating <- combined_df %>%
  group_by(Category) %>%
  summarize(Average_Rating = mean(Rating, na.rm = TRUE))

ggplot(avg_rating, aes(x = Category, y = Average_Rating, fill = Category)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Rating per Category", x = "Category", y = "Average Rating") +
  theme_minimal()
```

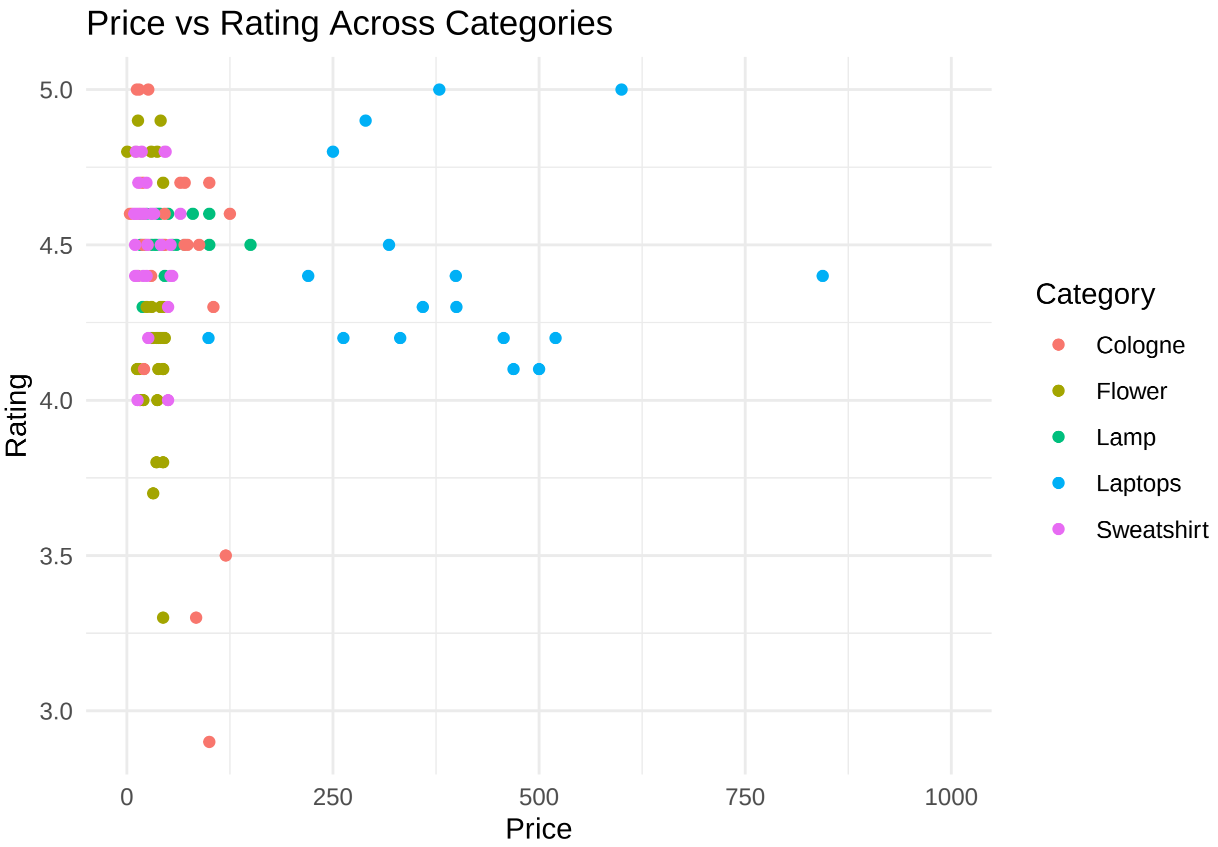


```r
avg_price <- combined_df %>%
  group_by(Category) %>%
  summarize(Average_Price = mean(Price, na.rm = TRUE))

ggplot(avg_price, aes(x = Category, y = Average_Price, fill = Category)) +
  geom_bar(stat = "identity") +
  labs(title = "Average Price per Category", x = "Category", y = "Average Price") +
  theme_minimal()
```



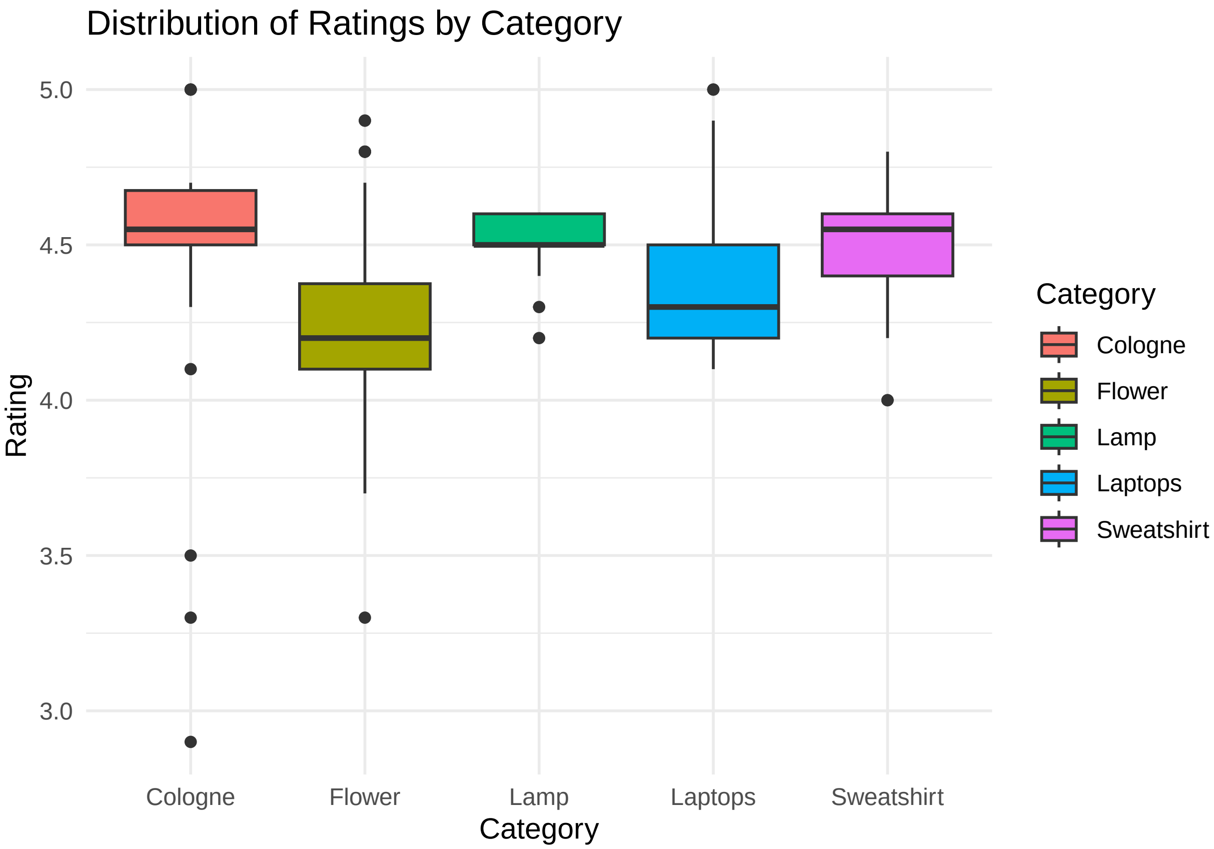
```r
ggplot(combined_df, aes(x = Price, y = Rating, color = Category)) +
  geom_point() +
  labs(title = "Price vs Rating Across Categories", x = "Price", y = "Rating") +
  theme_minimal()
```

```
## Warning: Removed 13 rows containing missing values or values outside the scale range (`geom_point()`
```

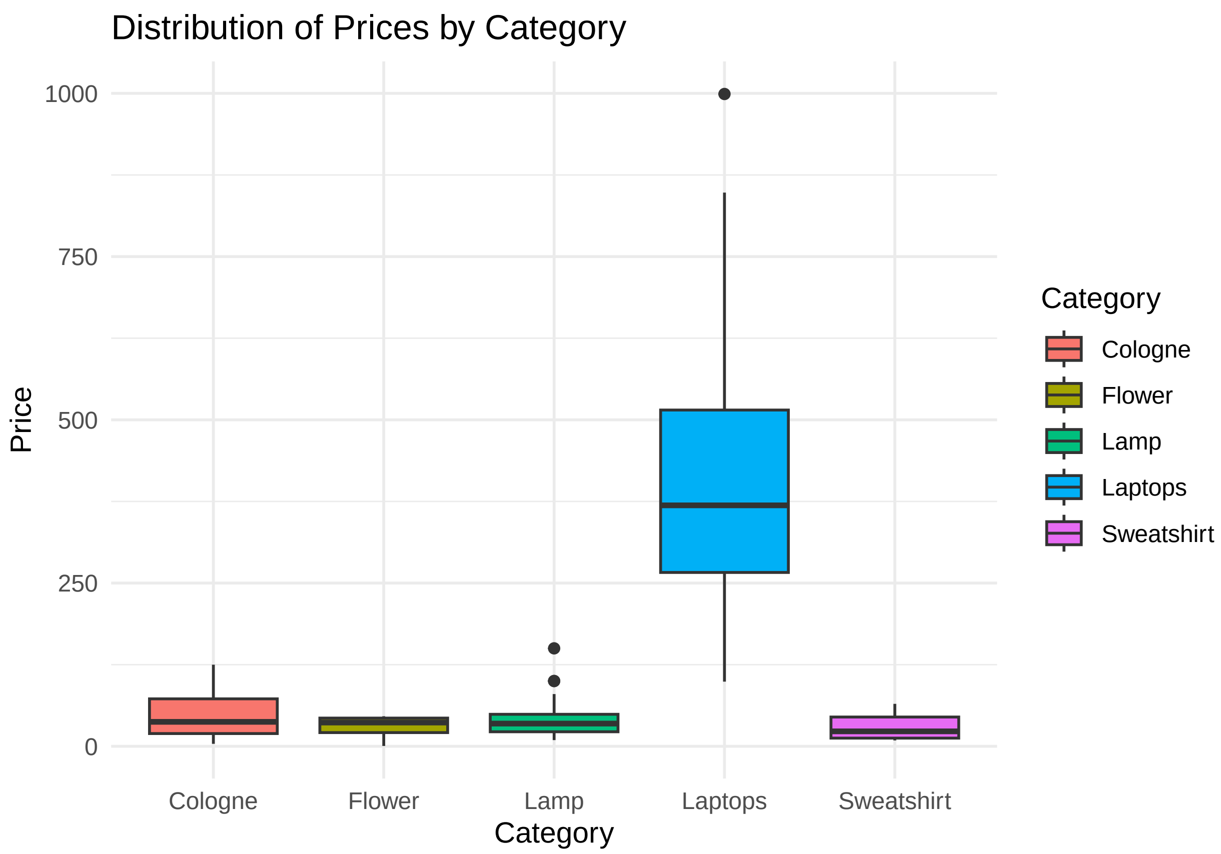


```
#9
ggplot(combined_df, aes(x = Category, y = Rating, fill = Category)) +
  geom_boxplot() +
  labs(title = "Distribution of Ratings by Category", x = "Category", y = "Rating") +
  theme_minimal()
```

```
## Warning: Removed 13 rows containing non-finite outside the scale range (`stat_boxplot()`).
```



```
ggplot(combined_df, aes(x = Category, y = Price, fill = Category)) +
  geom_boxplot() +
  labs(title = "Distribution of Prices by Category", x = "Category", y = "Price") +
  theme_minimal()
```



```
#10
ranked_data <- lapply(df, function(df_category) {
  df_category %>%
    arrange(desc(Rating), Price) %>%
    mutate(Rank = row_number()) %>%
    select(Rank, everything())
})

categories <- c("Laptops", "Lamp", "Flower", "Cologne", "Sweatshirt")
for (i in seq_along(ranked_data)) {
  ranked_data[[i]]$Category <- categories[i]
}

ranked_combined_df <- do.call(rbind, ranked_data)
ranked_combined_df <- ranked_combined_df %>%
  arrange(Category, Rank) %>%
  group_by(Category) %>%
  slice(1:5)

print(ranked_combined_df)
```

```
## # A tibble: 25 x 6
## # Groups:   Category [5]
##     Rank Product_Name
##    <int> <chr>
##  1     1 "Azzaro The Most Wanted Eau de Parfum Intense - Woody & Seductive Mens Cologne - Fougère, Ar
##  2     2 "D.Sauvage for Men Eau De Parfum Spray,Spray Refreshing & Warm Masculine Scent for Daily Use
##  3     3 "Guess Seductive Men Edt Spray, 3.4 Fl. Oz  "
##  4     4 "Ralph Lauren - Polo Blue - Eau de Toilette - Men's Cologne - Aquatic & Fresh - With Citrus
##  5     5 "Versace The Dreamer for Men 3.4 oz Eau de Toilette Spray  "
##  6     1 "LEGO Icons Plum Blossom Building Set for Adults Ages 18+ - Artificial Flowers for Dinner Ta
##  7     2 "LEGO Icons Orchid Artificial Plant, Building Set with Flowers, Home Décor Gift for Adults,
##  8     3 "LEGO Roses - Building Toy for Kids, Girls & Boys, Ages 8+ - Artificial, Fake Roses for Home
##  9     4 "Floroom Artificial Flowers 25pcs Real Looking Ivory Foam Fake Roses with Stems for DIY Wedd
## 10     5 "LEGO Cherry Blossom - Building Toy for Kids, Girls & Boys, Ages 8+ - Artificial Cherry Blos
## # i 15 more rows
```