

Uma análise dos 250 melhores filmes do IMDB

Caroline Neves

2023-04-09

Os favoritos da crítica

O IMDB é uma das maiores base de dados online de informação sobre cinema TV, música e games. Seus usuários podem avaliar os conteúdos com os quais interagiu, tendo a possibilidade de deixar comentários à respeito de sua avaliação. O conjunto de dados “IMDB Top 250 Movies Dataset”, disponível na plataforma Kaggle, foram extraídos do site do IMDB, no ano de 2021. No gráfico abaixo, podemos observar os gêneros que mais aparecem no top 250 do IMDB.

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.1      ✓ readr      2.1.4
## ✓ forcats    1.0.0      ✓ stringr    1.5.0
## ✓ ggplot2     3.4.1      ✓ tibble     3.2.1
## ✓ lubridate  1.9.2      ✓ tidyr      1.3.0
## ✓ purrr       1.0.1
## — Conflicts — tidyverse_conflicts() —
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to be
come errors
```

```
library(dplyr)
library(janitor)
```

```
##
## Attaching package: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##   chisq.test, fisher.test
```

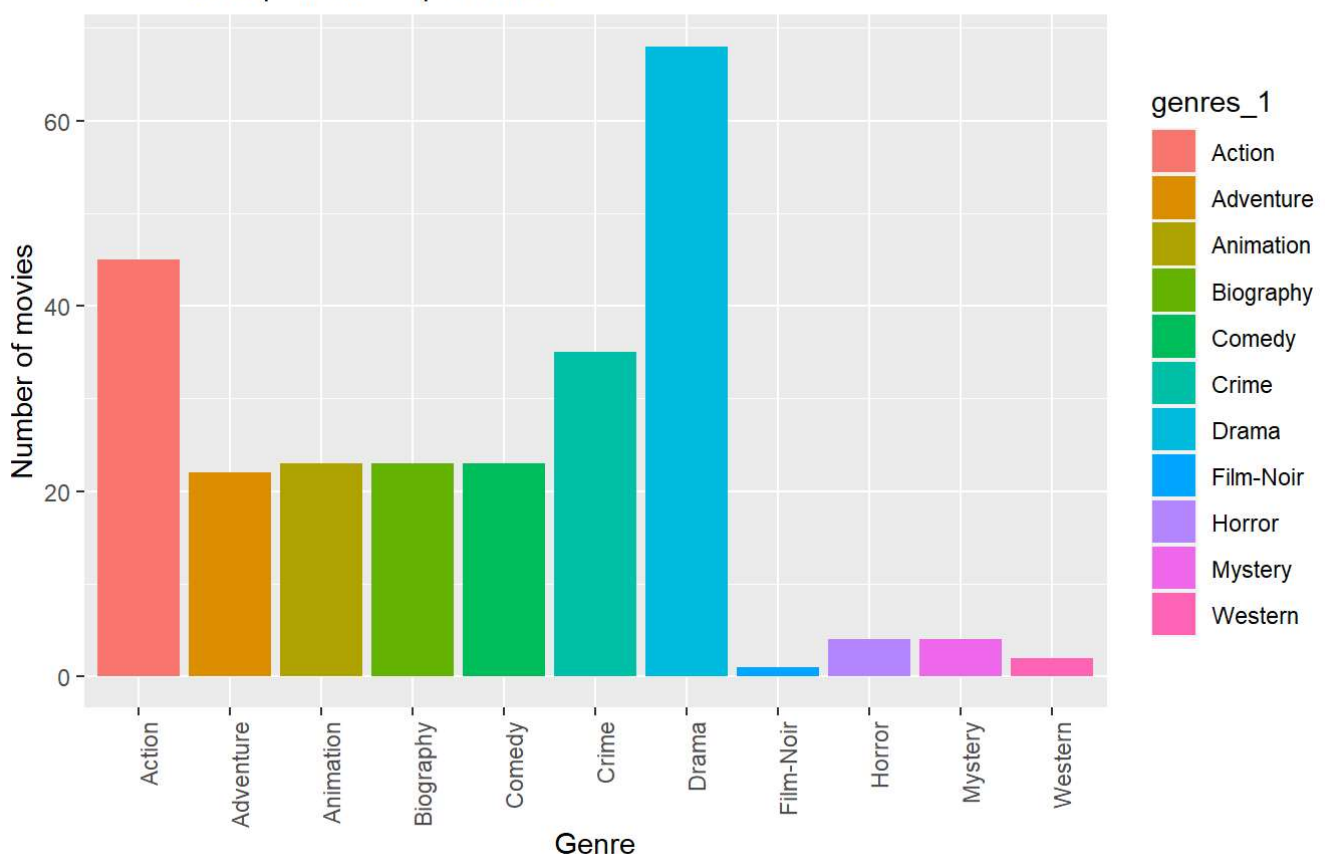
```
library(lubridate)
library(readr)
library(readr)
imdb_top_movies <- read_csv("C:/Users/Caroline/Downloads/imdb_analise - IMDB Top 250 Movies.csv")
```

```
## New names:
## Rows: 250 Columns: 22
## — Column specification
## _____ Delimiter: "," chr
## (15): name, genre, certificate, run_time, tagline, budget, box_office, c... dbl
## (3): rank, year, rating lgl (4): ...14, ...15, ...21, ...22
## i Use `spec()` to retrieve the full column specification for this data. i
## Specify the column types or set `show_col_types = FALSE` to quiet this message.
## • `` -> `...14`
## • `` -> `...15`
## • `` -> `...21`
## • `` -> `...22`
```

```
genres_count <- imdb_top_movies %>%
  select(rank, name, year, genres_1) %>%
  group_by(genres_1) %>%
  summarize(n = n()) %>%
  arrange(desc(n))
View(genres_count)
#um gráfico geral dos gêneros, nos dando a percepção dos gêneros favoritos da crítica
ggplot(genres_count, aes(x = genres_1, y = n, fill = genres_1)) +
  geom_bar(stat = "identity") +
  labs(title = "Top movie genres on IMDB",
       subtitle = "Based on the platform's top 250 movies",
       x = "Genre",
       y = "Number of movies") +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```

Top movie genres on IMDB

Based on the platform's top 250 movies



Emoção e adrenalina!

A primeira instância, podemos perceber que os filmes de ação e drama são os preferidos e mais bem avaliados, entre os usuários da plataforma. No gráfico abaixo, podemos ver a ocorrência dos filmes destes gêneros de forma mais clara. Dentre os 250 filmes, mais de 60 são filmes de drama e mais de 40 são filmes de ação!

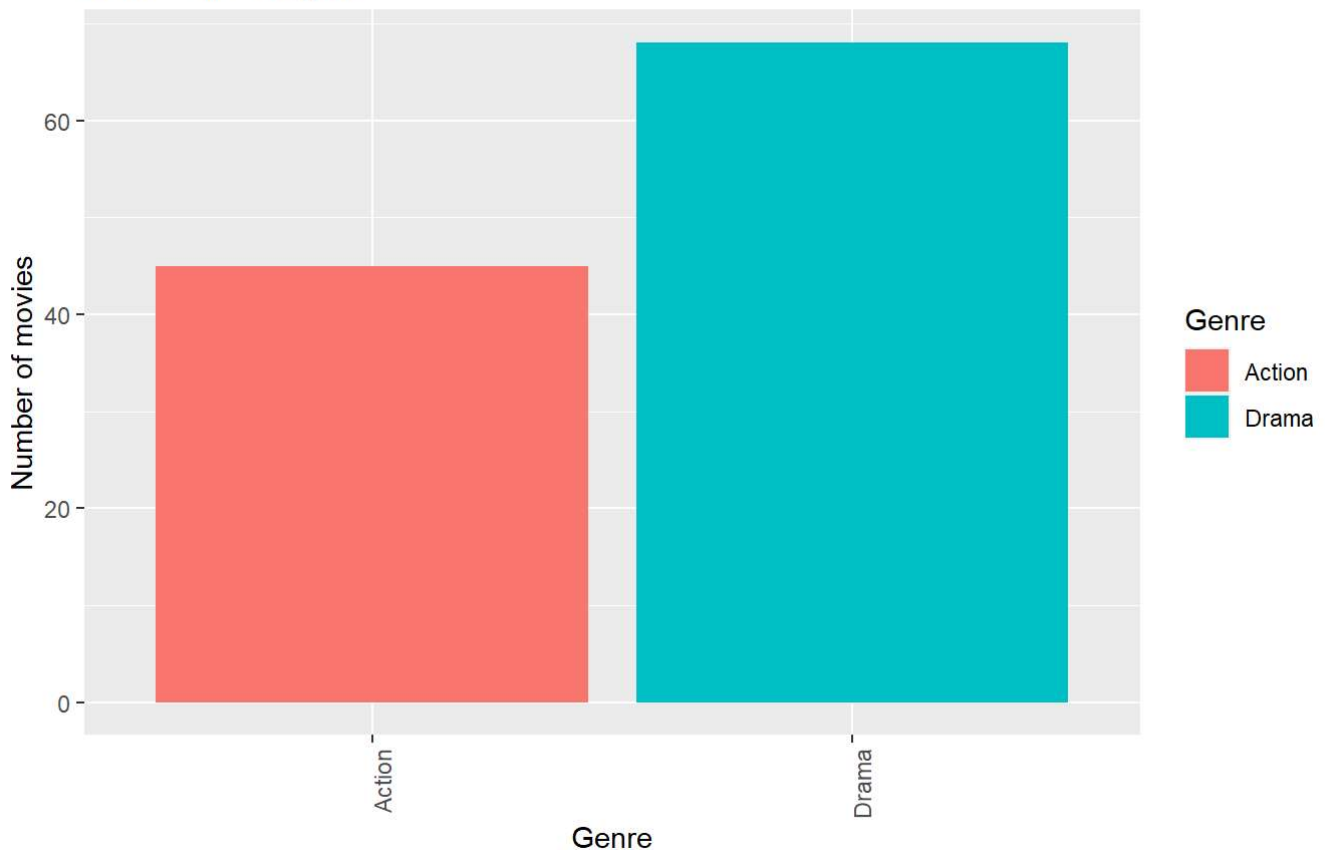
```
drama_action_films <- imdb_top_movies %>%  
  filter(genres_1 == "Drama" | genres_1 == "Action")  
  
##Verificar a quantidade de filmes selecionados  
nrow(drama_action_films)
```

```
## [1] 113
```

```
##contagem de filmes por gênero  
genre_count <- drama_action_films %>%  
  group_by(genres_1) %>%  
  summarise(count = n()) %>%  
  arrange(desc(count))  
##gráfico de barras  
ggplot(genre_count, aes(x = genres_1, y = count, fill = genres_1)) +  
  geom_bar(stat = "identity") +  
  labs(title = "Occurrence of Action and Drama films", subtitle= "Genres and Subgenres",  
        x = "Genre",  
        y = "Number of movies") +  
  theme(axis.text.x = element_text(angle = 90, hjust = 1)) +  
  scale_fill_discrete(name = "Genre")
```

Occurrence of Action and Drama films

Genres and Subgenres



Muita ação e aventura, com uma pitada de risadas

Uma outra forma de conhecer esses telespectadores é dando uma olhada nos subgêneros preferidos. No gráfico abaixo, podemos perceber que a galera curte muito ação com aventura, mas quando o assunto é ação e ficção-científica, parece que a indústria deixa um pouco a desejar.

```
#dentre os filmes de ação, quais são os subgêneros preferidos?
imdb_top_movies %>% dplyr::count(subgenres_1=="Action_Adventure") #2
```

```
## # A tibble: 2 × 2
##   `subgenres_1 == "Action_Adventure"`      n
##   <lg1>                                <int>
## 1 FALSE                                231
## 2 TRUE                                 19
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Action_Biography") #3
```

```
## # A tibble: 2 × 2
##   `subgenres_1 == "Action_Biography"`      n
##   <lg1>                                <int>
## 1 FALSE                                247
## 2 TRUE                                 3
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Action_Comedy") #2
```

```
## # A tibble: 2 × 2
##   `subgenres_1 == "Action_Comedy"`      n
##   <lgl>                                <int>
## 1 FALSE                                248
## 2 TRUE                                 2
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Action_Crime") #5
```

```
## # A tibble: 2 × 2
##   `subgenres_1 == "Action_Crime"`      n
##   <lgl>                                <int>
## 1 FALSE                                245
## 2 TRUE                                 5
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Action_Drama") #12
```

```
## # A tibble: 2 × 2
##   `subgenres_1 == "Action_Drama"`      n
##   <lgl>                                <int>
## 1 FALSE                                238
## 2 TRUE                                 12
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Action_Sci-fi") #0
```

```
## # A tibble: 1 × 2
##   `subgenres_1 == "Action_Sci-fi"`      n
##   <lgl>                                <int>
## 1 FALSE                                250
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Action_Thriller") #1
```

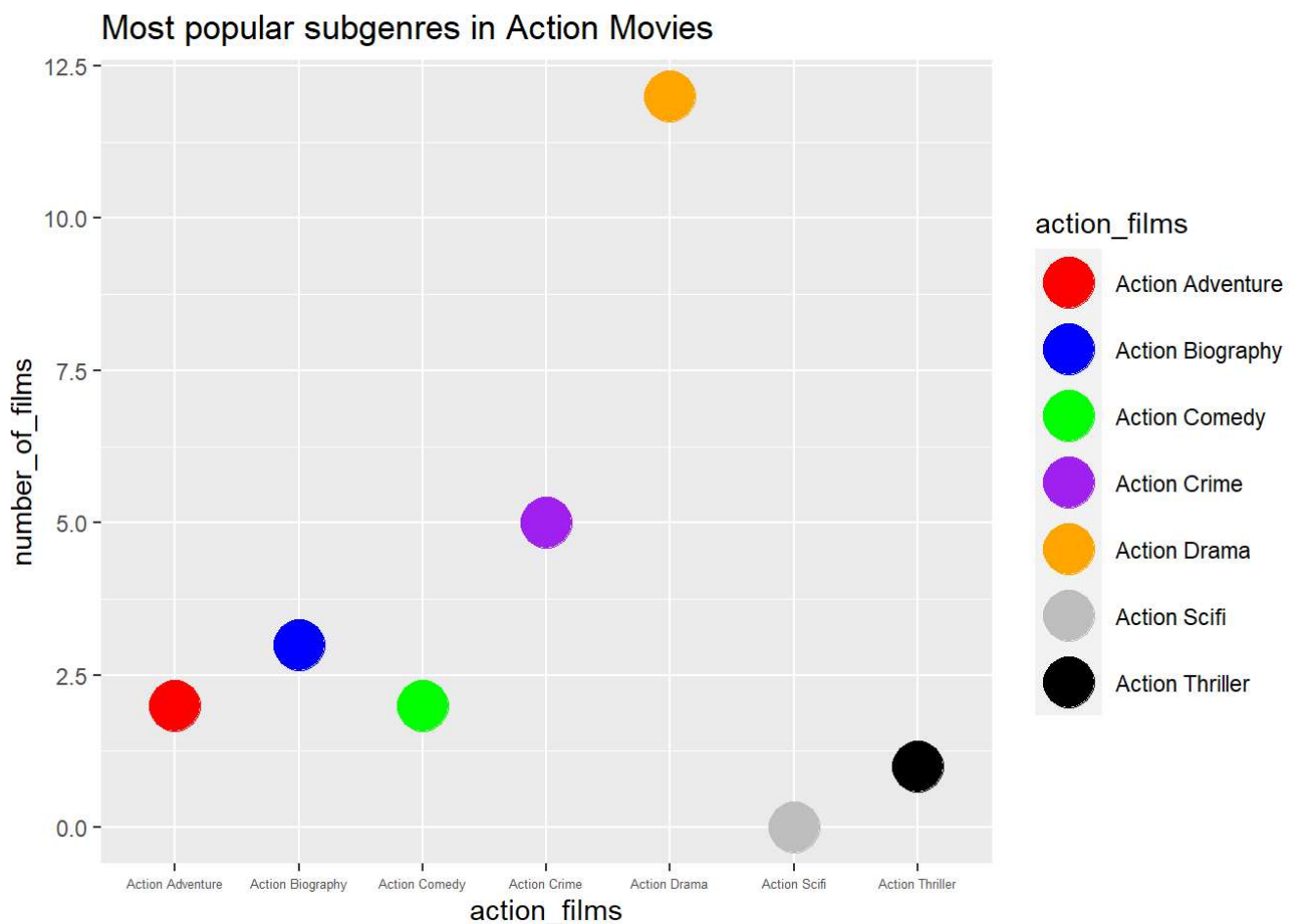
```
## # A tibble: 2 × 2
##   `subgenres_1 == "Action_Thriller"`      n
##   <lgl>                                <int>
## 1 FALSE                                249
## 2 TRUE                                 1
```

#criando um dataframe com os dados obtidos para criação do gráfico

```
action_films <- c("Action Adventure", "Action Biography","Action Comedy","Action Crime","Action Drama","Action Scifi","Action Thriller")
number_of_films <- c(2,3,2,5,12,0,1)
action_subgenres <- data.frame(action_films <- c("Action Adventure", "Action Biography",
"Action Comedy","Action Crime","Action Drama","Action Scifi","Action Thriller"), number_of_films <- c(2,3,2,5,12,0,1))
View(action_subgenres)

colors <- c("red", "blue", "green", "purple", "orange", "gray", "black")

ggplot(data = action_subgenres) +
  geom_point(mapping = aes(x = action_films, y = number_of_films, color = action_films), size = 9) +
  scale_color_manual(values = colors) +
  labs(title = "Most popular subgenres in Action Movies",
        xlab = "Subgenre",
        ylab = "Number of films") +
  theme(axis.text.x = element_text(size = 5))
```



Drama, romance e mistérios...

Parece a receita perfeita para um dia frio e chuvoso, não é mesmo? Acho que os usuários concordam conosco, vejamos no gráfico abaixo quais são os subgêneros de drama que mais apareceram no rank.

```
drama_subgenres <- imdb_top_movies %>%  
  filter(grepl("Drama", genres_1)) %>%  
  summarise(total_movies = n())  
View(drama_subgenres)  
  
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_Family") #4
```

```
## # A tibble: 2 × 2  
##   `subgenres_1 == "Drama_Family"`      n  
##   <lgl>                                <int>  
## 1 FALSE                                246  
## 2 TRUE                                 4
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_Fantasy") #2
```

```
## # A tibble: 2 × 2  
##   `subgenres_1 == "Drama_Fantasy"`      n  
##   <lgl>                                <int>  
## 1 FALSE                                248  
## 2 TRUE                                 2
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_Film-Noir") #2
```

```
## # A tibble: 2 × 2  
##   `subgenres_1 == "Drama_Film-Noir"`      n  
##   <lgl>                                <int>  
## 1 FALSE                                248  
## 2 TRUE                                 2
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_Horror") #1
```

```
## # A tibble: 2 × 2  
##   `subgenres_1 == "Drama_Horror"`      n  
##   <lgl>                                <int>  
## 1 FALSE                                249  
## 2 TRUE                                 1
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_Music") #1
```

```
## # A tibble: 2 × 2  
##   `subgenres_1 == "Drama_Music"`      n  
##   <lgl>                                <int>  
## 1 FALSE                                249  
## 2 TRUE                                 1
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_Mystery") #10
```

```
## # A tibble: 2 × 2
##   `subgenres_1` == "Drama_Mystery" `      n
##   <lgl>                                <int>
## 1 FALSE                                240
## 2 TRUE                                 10
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_Romance") #11
```

```
## # A tibble: 2 × 2
##   `subgenres_1` == "Drama_Romance" `      n
##   <lgl>                                <int>
## 1 FALSE                                239
## 2 TRUE                                 11
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_Sci-Fi") #1
```

```
## # A tibble: 2 × 2
##   `subgenres_1` == "Drama_Sci-Fi" `      n
##   <lgl>                                <int>
## 1 FALSE                                249
## 2 TRUE                                 1
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_Sport") #2
```

```
## # A tibble: 2 × 2
##   `subgenres_1` == "Drama_Sport" `      n
##   <lgl>                                <int>
## 1 FALSE                                248
## 2 TRUE                                 2
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_Thriller") #5
```

```
## # A tibble: 2 × 2
##   `subgenres_1` == "Drama_Thriller" `      n
##   <lgl>                                <int>
## 1 FALSE                                245
## 2 TRUE                                 5
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_War") #8
```

```
## # A tibble: 2 × 2
##   `subgenres_1` == "Drama_War" `      n
##   <lgl>                                <int>
## 1 FALSE                                242
## 2 TRUE                                 8
```

```
imdb_top_movies %>% dplyr::count(subgenres_1=="Drama_Western") #2
```

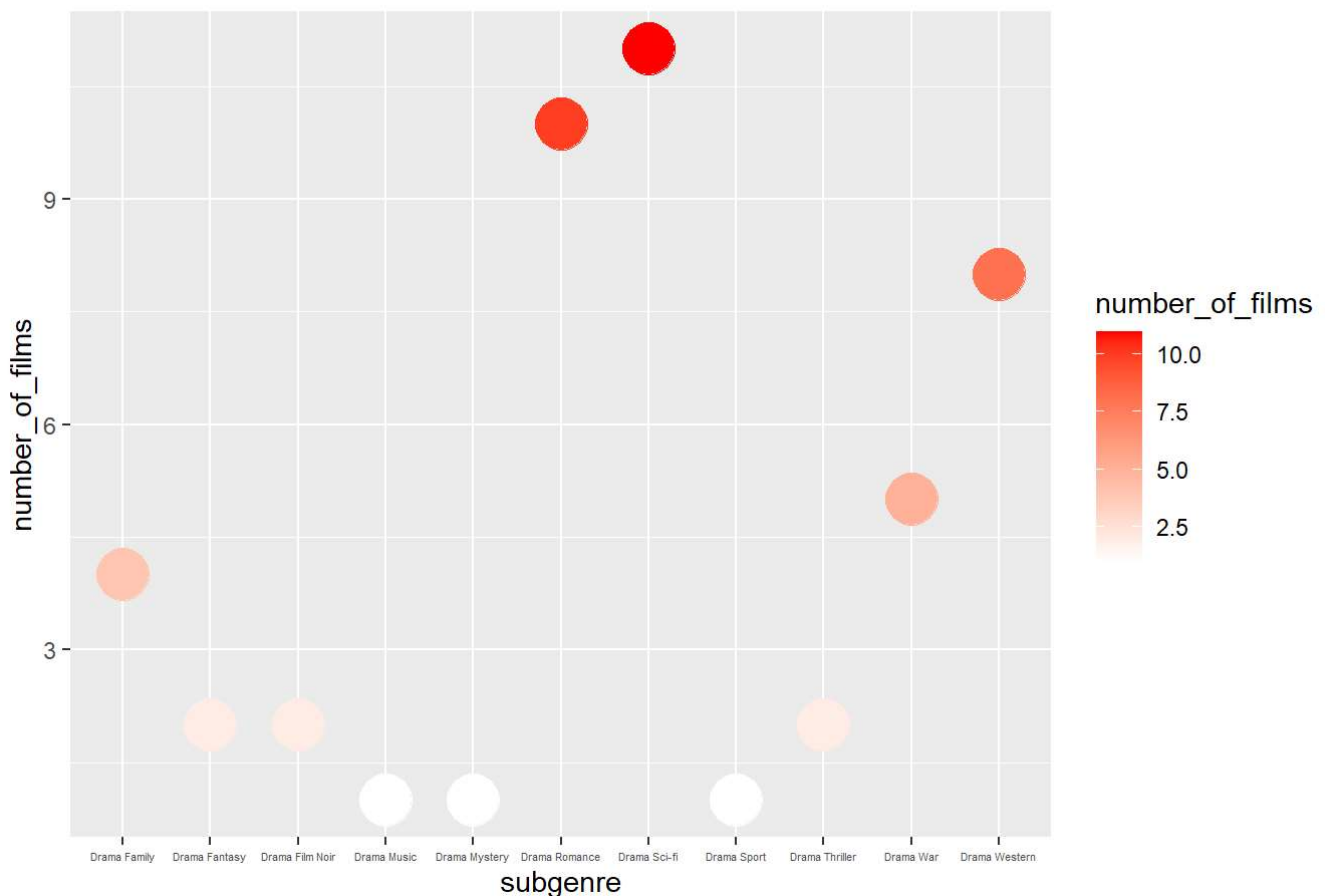


```
## # A tibble: 2 × 2
##   `subgenres_1 == "Drama_Western"`     n
##   <lgl>                               <int>
## 1 FALSE                               248
## 2 TRUE                                2
```

```
#criando um dataframe com os dados obtidos para criação do gráfico
drama_subgenres <- data.frame(subgenre = c("Drama Family", "Drama Fantasy",
                                           "Drama Film Noir", "Drama Music",
                                           "Drama Mystery", "Drama Romance", "Drama Sci-fi",
                                           "Drama Sport",
                                           "Drama Thriller", "Drama War", "Drama Western"),
                              number_of_films = c(4, 2, 2, 1, 1, 10, 11, 1, 2, 5, 8))

ggplot(data = drama_subgenres) +
  geom_point(mapping = aes(x = subgenre, y = number_of_films, color = number_of_films), size = 9) +
  scale_color_gradient(low = "white", high = "red") +
  labs(title = "Most popular subgenres in Drama Movies",
       xlab = "Subgenre",
       ylab = "Number of films") +
  theme(axis.text.x = element_text(size = 4))
```

Most popular subgenres in Drama Movies



Nada de clássicos!

E quando se trata dos anos de lançamentos, podemos ver uma tendência à melhores classificações para os filmes lançados nos ultimos 40 anos, principalmente quando se trata de filmes de ação.

```

action_movies <- imdb_top_movies %>% filter(genres_1 == "Action")
drama_movies <- imdb_top_movies %>% filter(genres_1 == "Drama")

ggplot()+
  geom_point(data = action_movies, aes(x = year, y= rank, color="Action"), alpha = 0.5, size
= 4)+
  geom_point(data = drama_movies, aes(x= year, y=rank, color="Drama"), alpha= 0.5, size = 4)+
  labs(title= "Occurrence os Action and Drama Films",
  subtitle= "A Comparison between release year and IMDB rank",
  x= "Release year",
  y= "IMDB Ranking",
  color= "Gênero")+
  scale_color_manual(values= c("Action"= "red", "Drama" = "blue"))

```

Occurrence os Action and Drama Films

A Comparison between release year and IMDB rank

