



UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA - IME
DEPARTAMENTO DE ESTATÍSTICA – DEST
Disciplina: Estatística Básica A – MAT B59

ANÁLISE DA INDICAÇÃO E PREMIAÇÃO DO OSCAR DE 1927- 2020

DANILO CADENA LIMA E LUCA FERREIRA BARBOZA

SALVADOR
2024

UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA - IME
DEPARTAMENTO DE ESTATÍSTICA – DEST
Disciplina: Estatística Básica A – MAT B59

ANÁLISE DA INDICAÇÃO E PREMIAÇÃO DO OSCAR DE 1927- 2020

DANILO CADENA LIMA E LUCA FERREIRA BARBOZA

Relatório apresentado como requisito de avaliação
para disciplina Estatística Básica A – MAT B59,
ofertada no semestre 2024.1
Docente: Nívea Bispo

SALVADOR

2024

1. INTRODUÇÃO

A base de dados selecionada é a Oscar Best Picture Movies, que contém todos os filmes nomeados e ganhadores da premiação de melhor filme do ano, desde 1927 até a premiação de 2020.

O Oscar é uma premiação que ocorre todos os anos e dentre todas as suas premiações a principal é a de melhor filme do ano, que dentre todos os filmes nomeados, têm um único ganhador anual. Para concorrer ao melhor filme do ano, os longas passam por um processo de votação em que aproximadamente 10000 pessoas da área votam para decidir o melhor filme.

O objetivo da análise é observar quais as principais mudanças e características dos filmes que participaram do Oscar com o passar dos anos, ou seja, o que os filmes que participaram da premiação têm em comum? As características principais mudaram com o passar do tempo ou seguem um padrão?

2. METODOLOGIA

Foram selecionados todos os filmes indicados aos Oscar a partir de 1927 até 2020, totalizando 568 filmes. As variáveis selecionadas para análise foram: nome do filme, estúdio ou produtor do filme, se foi nomeado ou ganhou o prêmio, tempo de duração do filme em minutos, gênero (romance, drama, comédia, etc), ano de lançamento do filme, e as notas dos sites IMDB e Rotten Tomatoes, sendo o primeiro média das notas dadas pelo público ao filme e o segundo a média da nota dos críticos sobre o filme. Vale ressaltar que cada filme recebeu até 3 classificações de gênero.

Para se ter uma melhor visualização das informações e para fazer comparações ao longo do tempo, separamos a base em 4 partes, sendo elas: filmes de 1927 até 1950, filmes entre 1950 até 1975, filmes de 1976 a 2000 e filmes de 2001 até 2020. Posteriormente, as medidas resumo tais como mínimo, máximo e mediana foram calculadas para todas as variáveis quantitativas, visando observar se houve perda ou ganho em qualidade e se os filmes ficaram mais curtos ou mais longos no período estudado. Boxplots foram construídos para as variáveis tempo de duração do filme (em minutos), nota do IMDB e nota do Rotten Tomatoes. Foi também construído o gráfico de dispersão entre as notas do Rotten Tomatoes e IMDB, e calculado o coeficiente de correlação linear de Pearson. Além disso, foi construído um gráfico de linhas que relaciona a média das notas do Rotten Tomatoes e do IMDB com o seu respectivo ano, para isso foi considerada a mesma escala para ambas as notas, de 0 a 10.

Em relação à variável qualitativa gênero foi calculada a frequência em que cada gênero aparece no Oscar, tanto no contexto geral como dividido em períodos, sendo esses de gênero dos filmes de 1927 até 1950, de 1950 até 1975, de 1976 a 2000 e de 2001 até 2020. Também foi calculada a frequência em que os estúdios e produtores foram indicados e ganharam o Oscar, separando as 10 maiores frequências, para, desse modo, tentar estabelecer algum padrão. Com o objetivo de analisar as frequências das variáveis qualitativas foram construídos gráficos de coluna.

3. RESULTADOS E CONSIDERAÇÕES FINAIS

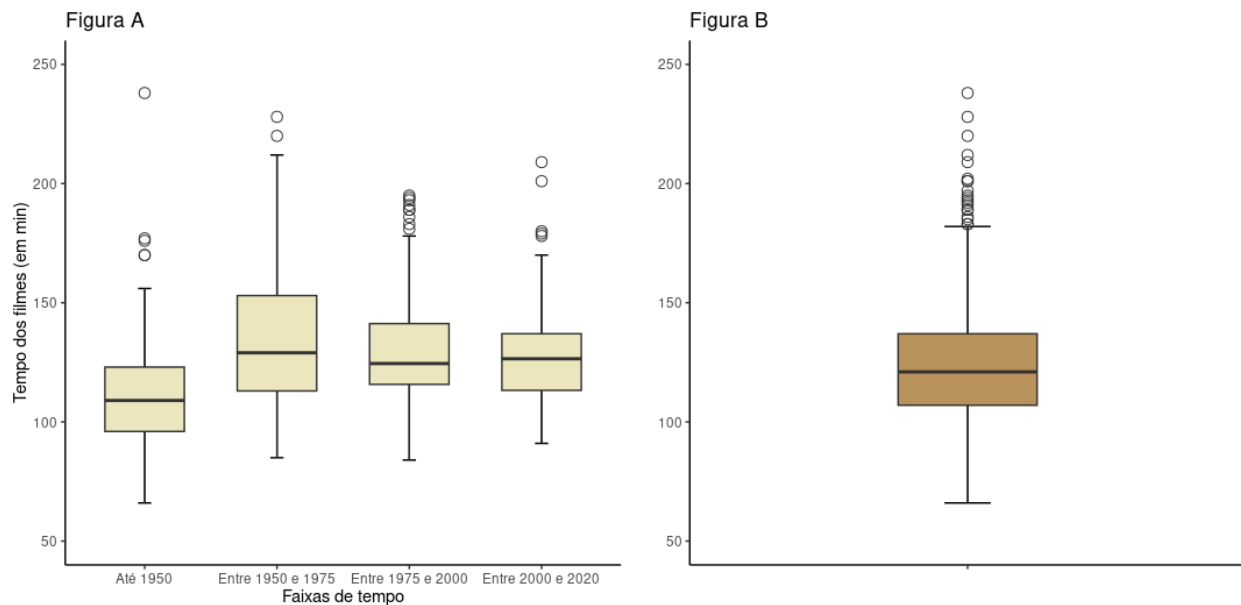


Figura 1. Distribuição do tempo de duração dos filmes por faixa (Figura A) e pelo total (Figura B).

Na Figura 1 estão sendo apresentadas as distribuições dos tempos de duração dos filmes em minutos. Observa-se que os filmes que apresentaram o menor e o maior tempo de duração foram lançados entre 1927 até 1950, com 1h6min e 3h58min, respectivamente. Além disso, observa-se que a amplitude do tempo de duração foi diminuindo com o passar dos anos.

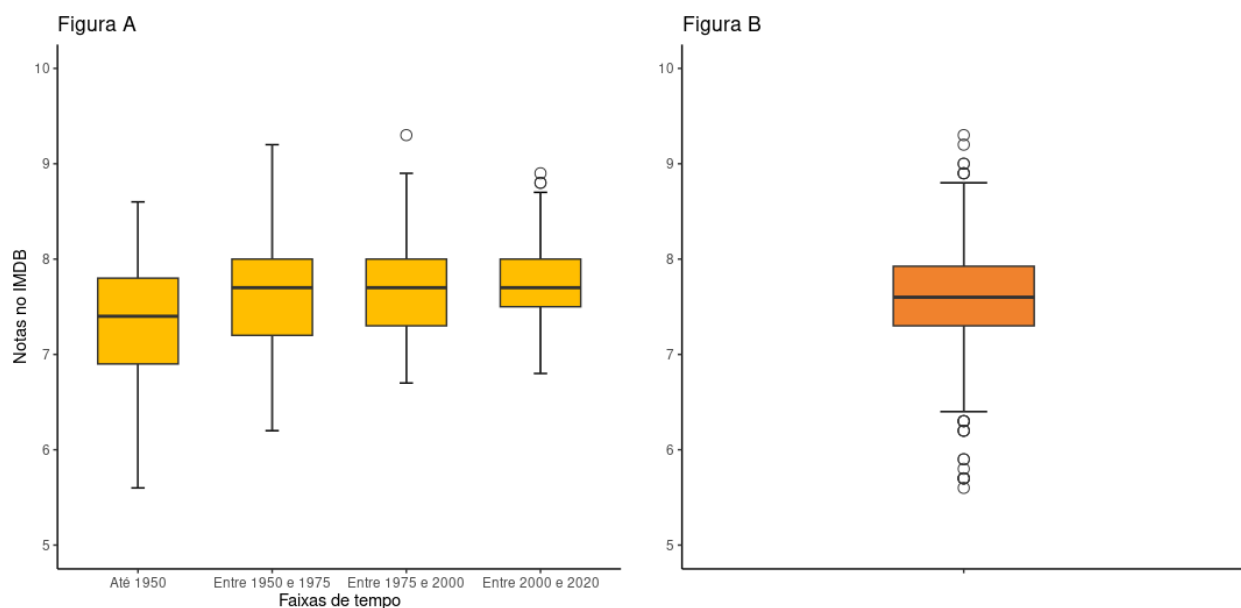


Figura 2. Distribuição das notas pelo IMDB dos filmes por faixa (Figura A) e pelo total (Figura B).

Na Figura 2 estão sendo apresentadas as distribuições das notas pelo IMDB dos filmes. Observa-se que o filme que apresentou a menor nota pelo público ocorreu entre os filmes lançados de 1927 a 1950, com nota de 5,6; a maior nota do público foi observada para filmes lançados entre 1975 a 2000, com nota 9,3. Além disso, observa-se que a amplitude das notas foi diminuindo com o passar dos anos. Considerando todos os anos, observa-se que 50% das notas pelo público foram maiores que 7,6.

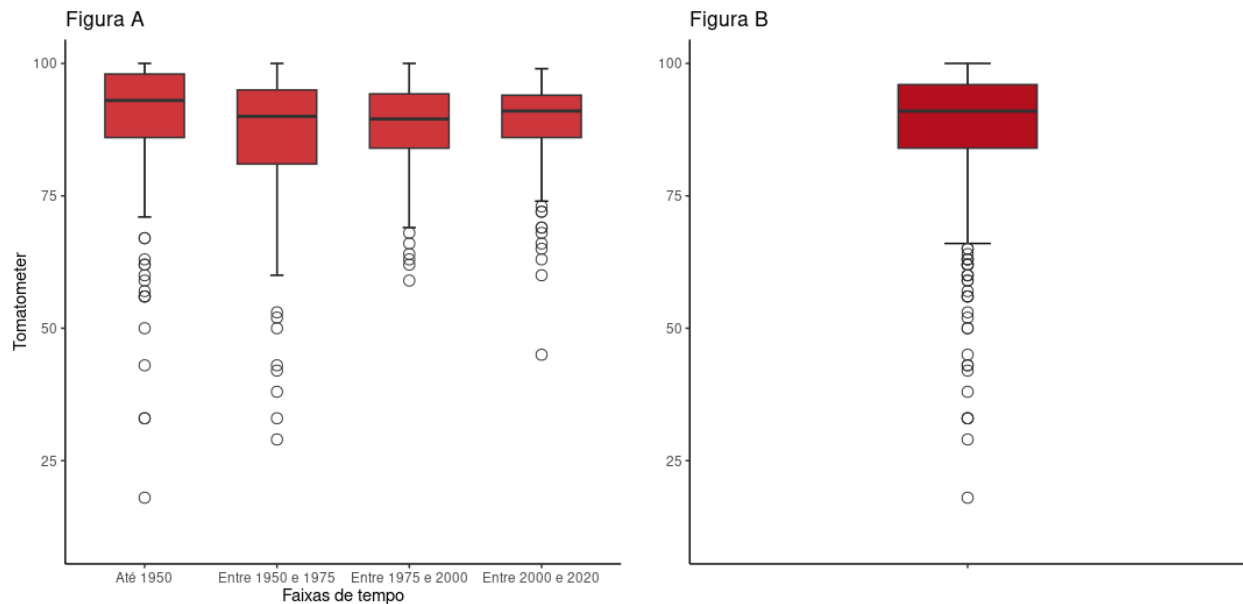


Figura 3. Distribuição das notas pelo Rotten Tomatoes dos filmes por faixa (Figura A) e pelo total (Figura B).

Na Figura 3 estão sendo apresentadas as distribuições das notas pelo Rotten Tomatoes dos filmes. Observa-se que o filme que apresentou a menor nota pelos críticos ocorreu entre os filmes lançados de 1927 a 1950, com nota de 18 e em todas as faixas notas iguais a 100 foram observadas; 50% dos filmes de 1927 a 1950 tiveram nota superior a 93, acima das notas de todos os outros anos. Além disso, percebe-se que em geral a variabilidade das notas tem diminuído ao longo dos anos.

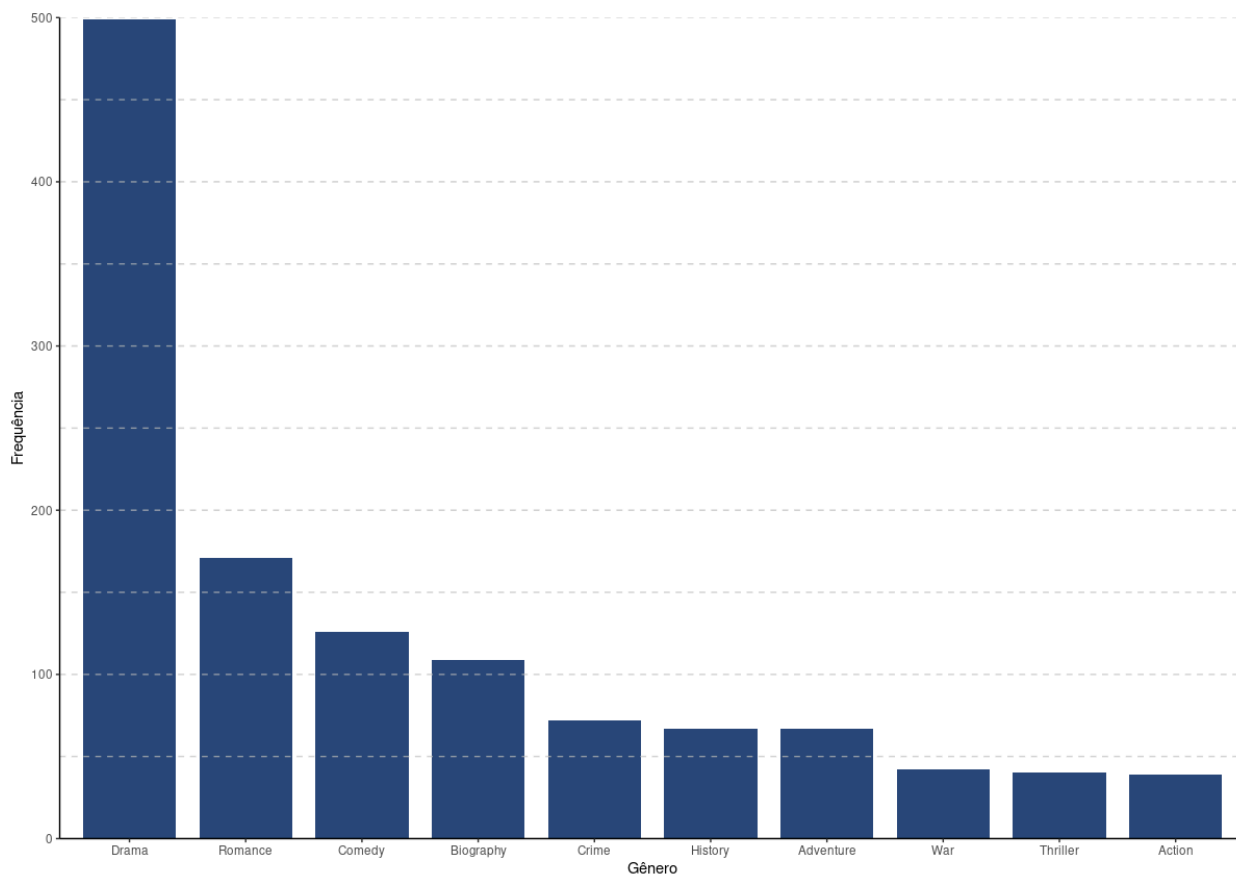


Figura 4. Frequência dos gêneros de filmes que foram nomeados ao Oscar.

Na Figura 4 está sendo apresentada a frequência dos gêneros das películas que concorreram ao prêmio de melhor filme do ano. É possível perceber que a maioria dos longas foram de drama seguido de romance, comédia e filmes biográficos. De modo geral observa-se que normalmente os filmes de drama possuem um maior apelo nesta competição.

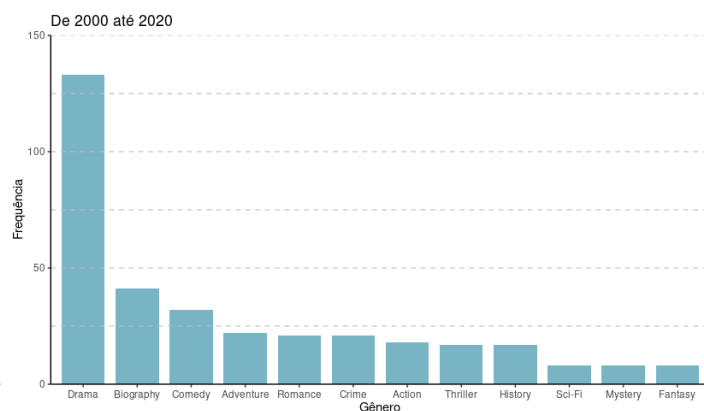
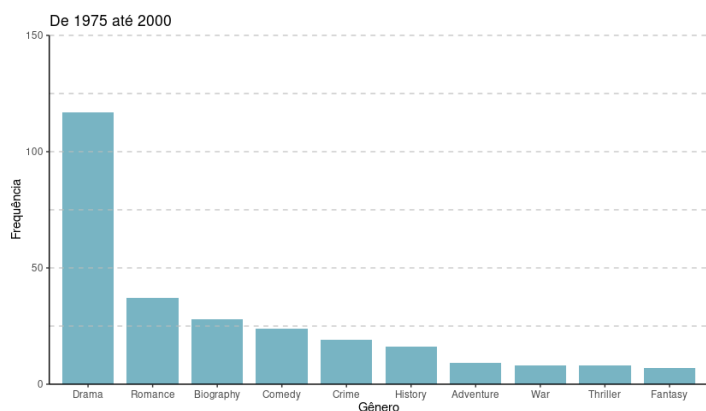
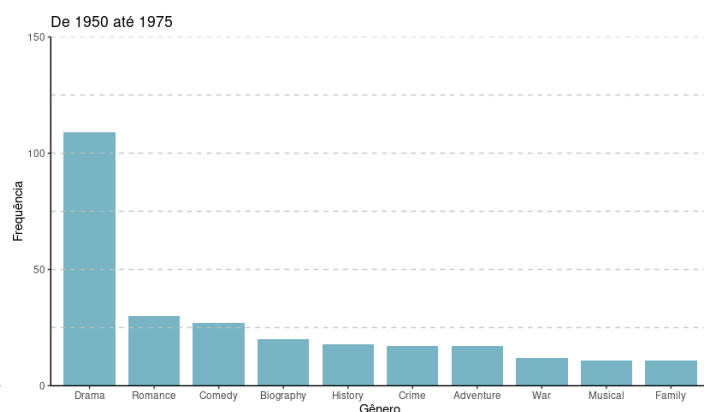
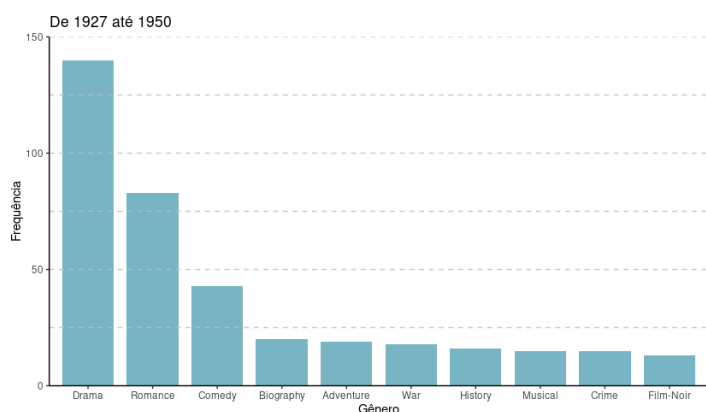


Figura 5. Frequência dos gêneros de filmes que foram nomeados ao Oscar.

Na Figura 5 está mostrando a frequência dos gêneros de filmes por período. É possível perceber que o gênero de drama foi o mais frequente para todas as faixas de tempo e ele ainda aparece superando em muito a frequência do segundo gênero. Outro ponto a ser destacado é que o gênero de romance que aparece como segundo mais frequente em 1950 apresenta diminuição de frequência logo após esse período, ficando como quinto mais frequente no período de 2001 até 2020. Também é possível observar que o gênero biográfico apresenta aumento desde 1927 e o de comédia sempre aparece entre os mais frequentes.

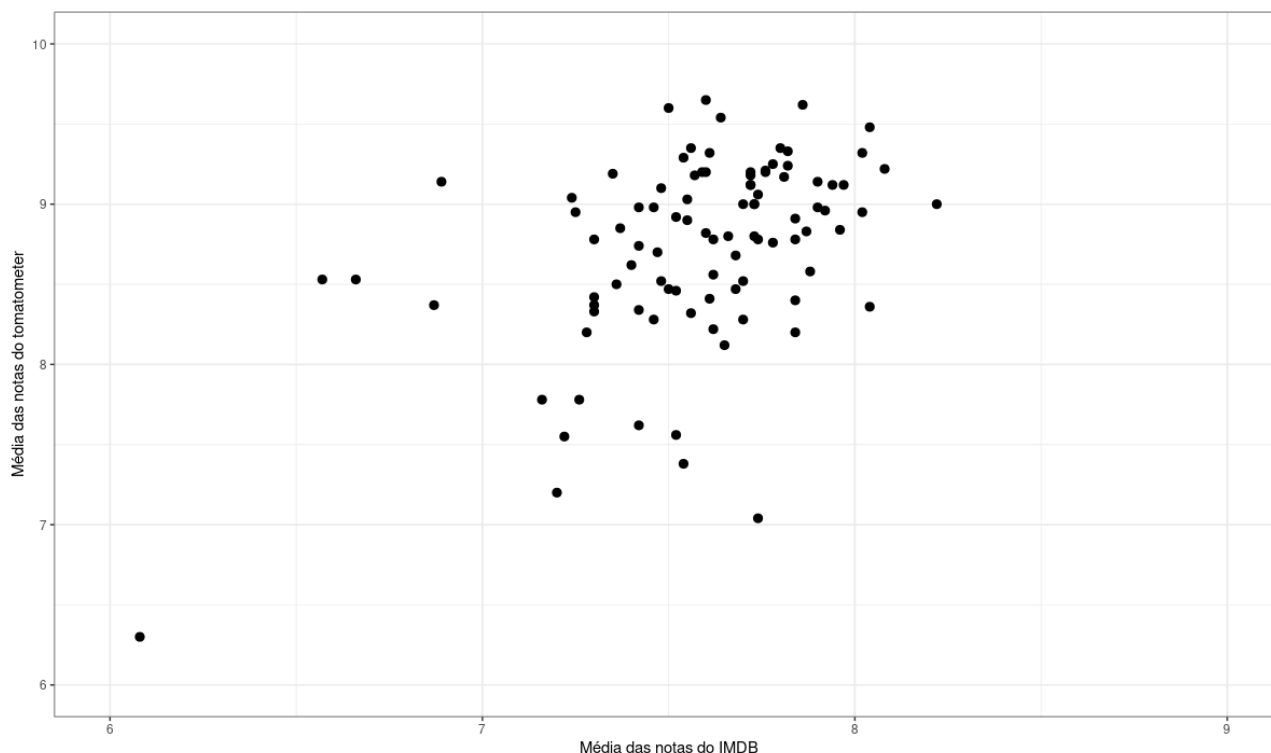


Figura 6. Diagrama de dispersão entre as médias das notas do IMDB e Rotten Tomatoes.

Na Figura 6, pode-se observar que, em média, os críticos tendem a dar notas mais altas pelos filmes nomeados ao Oscar, enquanto o público tende a dar, em média, notas mais baixas, evidenciando assim uma baixa associação linear (correlação de 0,48 aproximadamente). É possível observar essa relação mais claramente na Figura 7, onde se compara as médias das notas em cada ano.

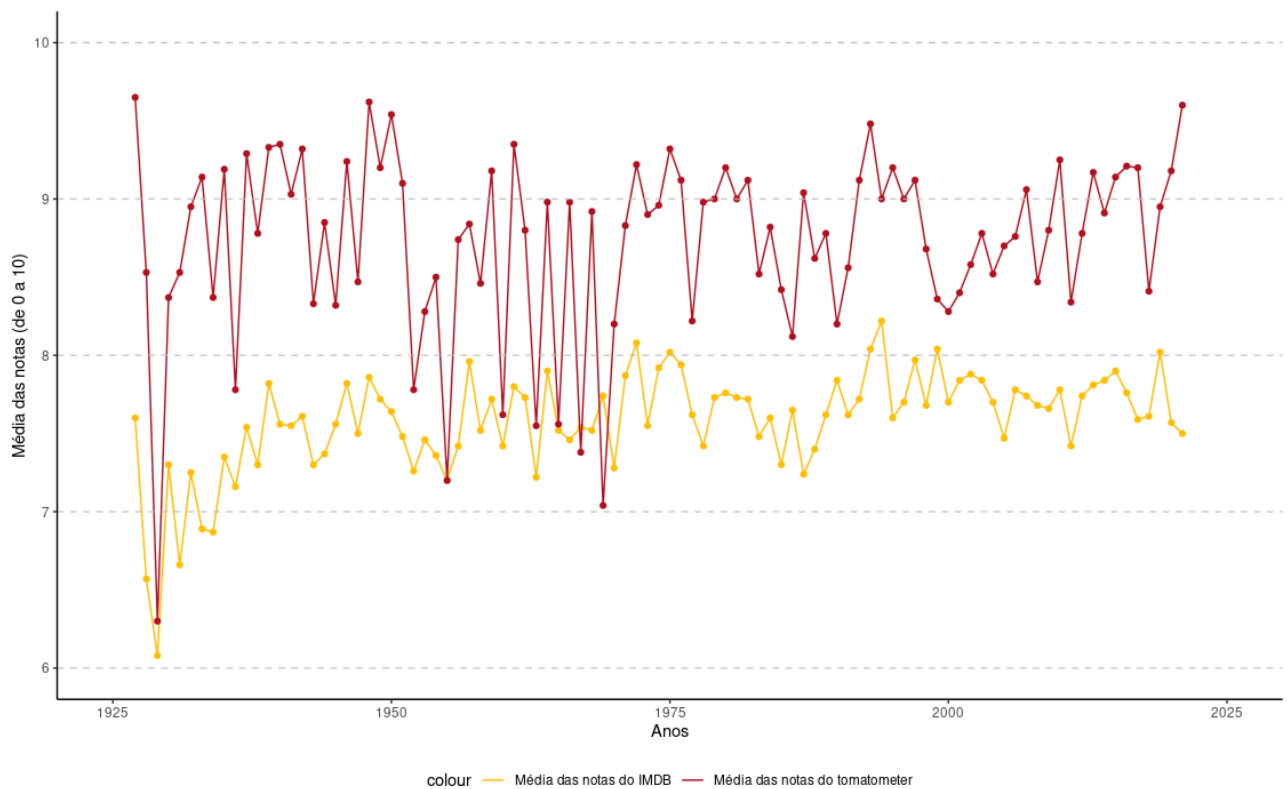


Figura 7. Notas do público e dos críticos ao longo do tempo.

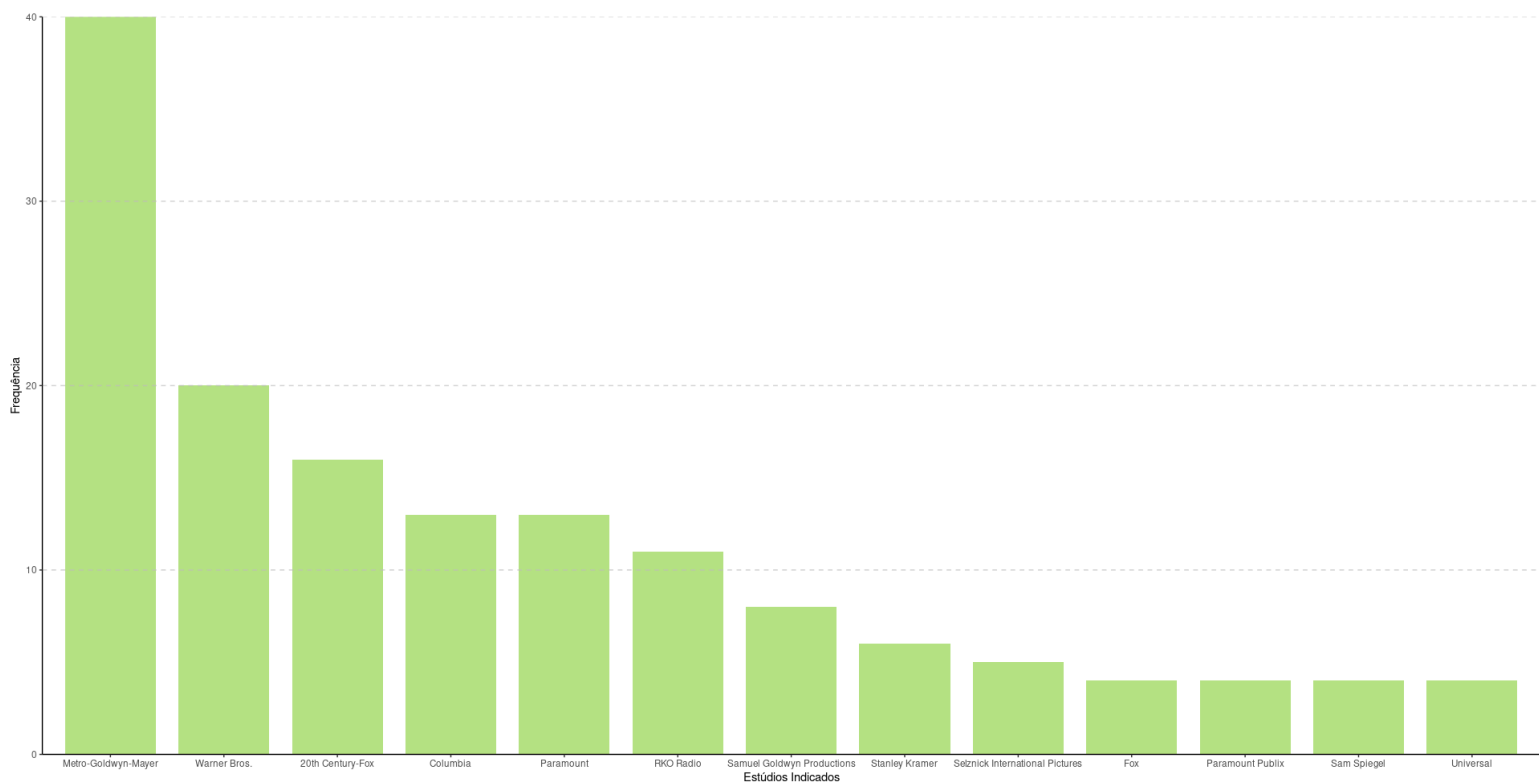


Figura 8. Frequência dos 10 estudios que mais tiveram seus filmes indicados ao Oscar.

Pela Figura 8 é possível observar que o Metro-Goldwyn Meyer foi o mais frequente, com 40 indicações, seguido do estúdio Warner Bros e 20th Century-Fox, 20 e 16 indicações, respectivamente. Porém, pode-se observar que, pela Figura 9, o estúdio Metro-Goldwyn Meyer tem menos vitórias relativas que o 20th Century-Fox, já que o primeiro tem 40 indicações e 5 prêmios e o segundo contendo 16 indicações e 3 prêmios.

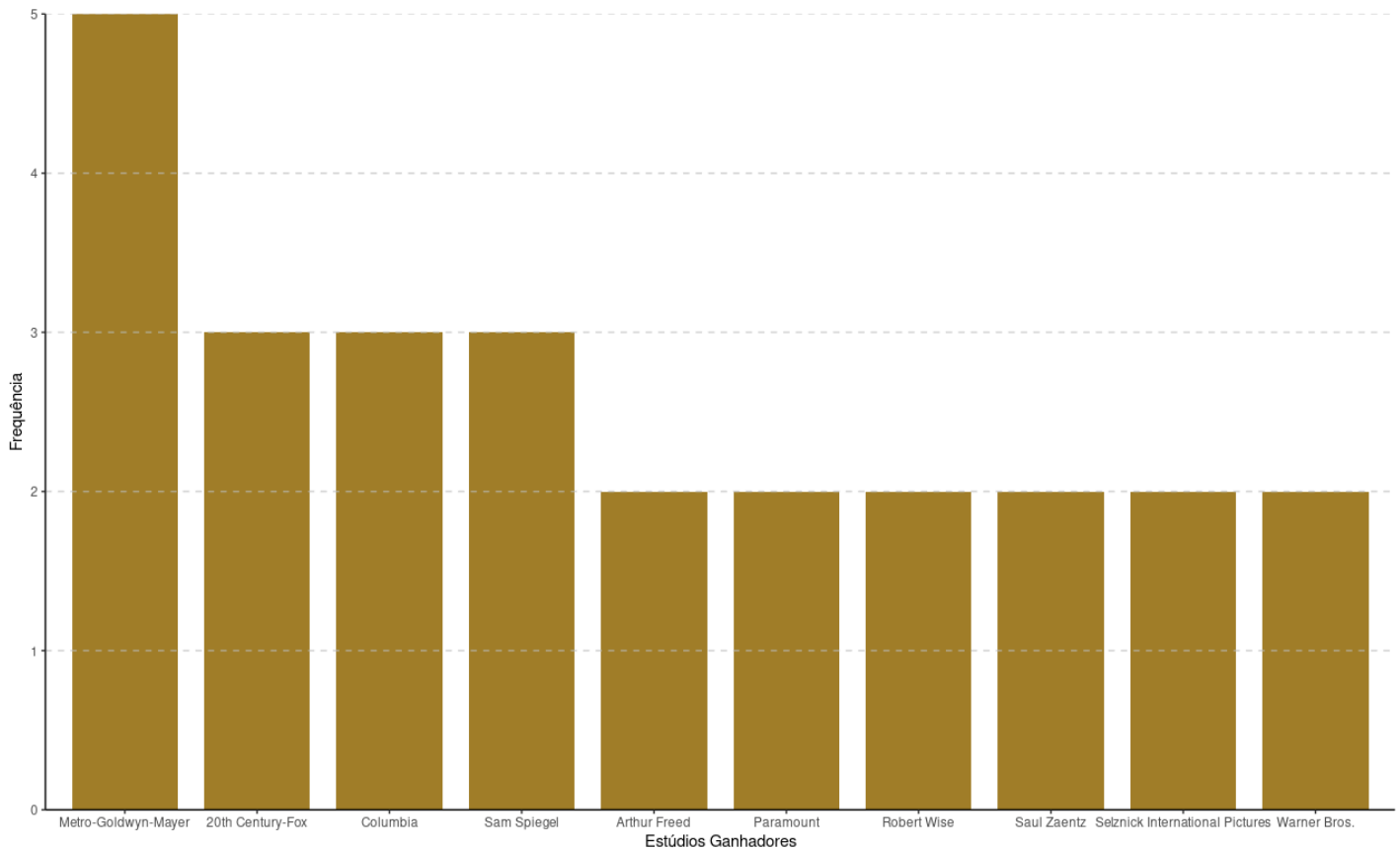


Figura 9. Frequência dos 10 estúdios que mais ganharam o Oscar de melhor filme.

Com base nas análises realizadas foi possível ter uma completa visualização do que ocorreu com o passar dos quase 100 anos da indústria de filmes (1927-2020) e podemos concluir que, para as variáveis quantitativas, embora o tempo de filme seja a variável que mais venha se alterando, de modo geral, as variáveis não tiveram uma alteração tão perceptível e por isso parecem seguir um padrão. Já nas variáveis qualitativas é possível observar que, excluindo drama que é o gênero mais frequente em todos os períodos, os outros gêneros parecem estar se alterando ao longo do tempo. Além disso, por mais que um estúdio receba muitas indicações ao Oscar, não é garantia de receber a premiação.

ANEXO (código R)

```
library(readxl)
```

```
oscars_df <- read_excel("oscars_df.xlsx")
```

```
View(oscars_df)
```

```
attach(oscars_df)
```

```
install.packages("forcats")
```

```
install.packages("cowplot")
```

```
install.packages("fdth")
```

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

```
library(tidyverse)
```

```
library(magrittr)
```

```
library(fdth)
```

```
library(cowplot)
```

```
library(gganimate)
```

```
library(forcats)
```

```
oscar_1950 = oscars_df %>%
```

```
  filter(`Year of Release` <= 1950) %>%
```

```
  mutate(Ano_limite = "Até 1950")
```

```
view(oscar_1950)
```

```
oscar_1975 = oscars_df %>%
```

```
  filter(`Year of Release` > 1950 & `Year of Release` <= 1975) %>%
```

```
  mutate(Ano_limite = "Entre 1950 e 1975")
```

```
view(oscar_1975)
```

```
oscar_2000 = oscars_df %>%
```

```
  filter(`Year of Release` > 1975 & `Year of Release` <=2000) %>%
```

```
  mutate(Ano_limite = "Entre 1975 e 2000")
```

```
view(oscar_2000)
```

```
oscar_2020 = oscars_df %>%
```

```
  filter(`Year of Release` > 2000) %>%
```

```
  mutate(Ano_limite = "Entre 2000 e 2020")
```

```
view(oscar_2020)
```

```
oscar_dados = rbind(oscar_1950, oscar_1975, oscar_2000, oscar_2020)
view(oscar_dados)
```

```
#Análise da base até 1950
```

```
#Tempo dos filmes
```

```
Freq_tempo_1950 <- fdt(oscar_1950$`Movie Time`)
```

```
print(Freq_tempo_1950, format=TRUE, col = c(1,2,4,5) , pattern = "%.2f ")
```

```
Sumario_tempo_1950 = summary(oscar_1950$`Movie Time`)
```

```
Sumario_tempo_1950
```

```
Media_tempo_1950 = round(mean(oscar_1950$`Movie Time`), 2)
```

```
Media_tempo_1950
```

```
Mediana_tempo_1950 = round(median(oscar_1950$`Movie Time`), 2)
```

```
Mediana_tempo_1950
```

```
Var_tempo_1950 = round(var(oscar_1950$`Movie Time`), 2)
```

```
Var_tempo_1950
```

```
Dp_tempo_1950 = round(sd(oscar_1950$`Movie Time`), 2)
```

```
Dp_tempo_1950
```

```
Dm_tempo_1950 = round(sum(abs(oscar_1950$`Movie Time` -
Mediana_tempo_1950))/length(oscar_1950$`Movie Time`), 2)
```

```
Dm_tempo_1950
```

```
cv_media_tempo_1950 = round((Dp_tempo_1950/Media_tempo_1950)*100, 2)
```

```
cv_media_tempo_1950
```

```
cv_mediana_tempo_1950 = round((Dm_tempo_1950/Mediana_tempo_1950)*100, 2)
```

```
cv_mediana_tempo_1950
```

```
#Notas IMDB
```

```
Freq_IMDB_1950 <- fdt(oscar_1950$`IMDB Rating`)
```

```
print(Freq_IMDB_1950, format=TRUE, col = c(1,2,4,5) , pattern = "%.2f ")
```

```
Sumario_IMDB_1950 = summary(oscar_1950$`IMDB Rating`)
```

```
Sumario_IMDB_1950
```

```
Media_IMDB_1950 = round(mean(oscar_1950$`IMDB Rating`), 2)
```

```
Media_IMDB_1950
```

```
Mediana_IMDB_1950 = round(median(oscar_1950$`IMDB Rating`), 2)
```

```
Mediana_IMDB_1950
```

```
Var_IMDB_1950 = round(var(oscar_1950$`IMDB Rating`), 2)
```

```
Var_IMDB_1950
```

```
Dp_IMDB_1950 = round(sd(oscar_1950$`IMDB Rating`), 2)
```

```
Dp_IMDB_1950
```

```
Dm_IMDB_1950 = round(sum(abs(oscar_1950$`IMDB Rating` -  
Mediana_IMDB_1950))/length(oscar_1950$`IMDB Rating`), 2)
```

```
Dm_IMDB_1950
```

```
cv_media_IMDB_1950 = round((Dp_IMDB_1950/Media_IMDB_1950)*100, 2)
```

```
cv_media_IMDB_1950
```

```
cv_mediana_IMDB_1950 = round((Dm_IMDB_1950/Mediana_IMDB_1950)*100, 2)
```

```
cv_mediana_IMDB_1950
```

```
#Notas Tomatometer
```

```
Freq_tomatometer_1950 <- fdt(oscar_1950$`Tomatometer Rating`)
```

```
print(Freq_tomatometer_1950, format=TRUE, col = c(1,2,4,5) , pattern = "%.2f ")
```

```
Sumario_tomatometer_1950 = summary(oscar_1950$`Tomatometer Rating`)
```

```
Sumario_tomatometer_1950
```

```
Media_tomatometer_1950 = round(mean(oscar_1950$`Tomatometer Rating`), 2)
Media_tomatometer_1950
```

```
Mediana_tomatometer_1950 = round(median(oscar_1950$`Tomatometer Rating`), 2)
Mediana_tomatometer_1950
```

```
Var_tomatometer_1950 = round(var(oscar_1950$`Tomatometer Rating`), 2)
Var_tomatometer_1950
```

```
Dp_tomatometer_1950 = round(sd(oscar_1950$`Tomatometer Rating`), 2)
Dp_tomatometer_1950
```

```
Dm_tomatometer_1950 = round(sum(abs(oscar_1950$`Tomatometer Rating` -
Mediana_tomatometer_1950))/length(oscar_1950$`Tomatometer Rating`), 2)
Dm_tomatometer_1950
```

```
cv_media_tomatometer_1950 =
round((Dp_tomatometer_1950/Media_tomatometer_1950)*100, 2)
cv_media_tomatometer_1950
```

```
cv_mediana_tomatometer_1950 =
round((Dm_tomatometer_1950/Mediana_tomatometer_1950)*100, 2)
cv_mediana_tomatometer_1950
```

```
#Estudios
```

```
Freq_estudios_1950 = oscar_1950 %>%
  group_by(oscar_1950$`Film Studio/Producer(s)`) %>%
  summarise(frequencia_absoluta = n()) %>%
  mutate(frequencia_relativa = round(frequencia_absoluta/sum(frequencia_absoluta) *
100, 2))
```

```
Freq_ganhadores_1950_0 = oscar_1950 %>%
  filter(oscar_1950$Award == "Winner")
```

```

Freq_ganhadores_1950 = Freq_ganhadores_1950_0 %>%
  group_by(Freq_ganhadores_1950_0$`Film Studio/Producer(s)`) %>%
  summarise(frequencia_absoluta = n()) %>%
  mutate(frequencia_relativa = round(frequencia_absoluta/sum(frequencia_absoluta) *
100, 2))

```

```

#Genero
genero_1950 = strsplit(oscar_1950$`Movie Genre`, ",")
un_gen_1950 = unlist(genero_1950)
table_gen_1950 = sort(table(un_gen_1950), decreasing = T)
df_gen_1950 = as.data.frame(table_gen_1950)
df_gen_1950 = df_gen_1950 %>%
  top_n(10, Freq)

```

```

# Analise da base de 1975

```

```

#Tempo dos filmes
Freq_tempo_1975 <- fdt(oscar_1975$`Movie Time`)
print(Freq_tempo_1975, format=TRUE, col = c(1,2,4,5) , pattern = "%.2f ")

```

```

Sumario_tempo_1975 = summary(oscar_1975$`Movie Time`)
Sumario_tempo_1975

```

```

Media_tempo_1975 = round(mean(oscar_1975$`Movie Time`), 2)
Media_tempo_1975

```

```

Mediana_tempo_1975 = round(median(oscar_1975$`Movie Time`), 2)
Mediana_tempo_1975

```

```

Var_tempo_1975 = round(var(oscar_1975$`Movie Time`), 2)
Var_tempo_1975

```

```

Dp_tempo_1975 = round(sd(oscar_1975$`Movie Time`), 2)
Dp_tempo_1975

```

```

Dm_tempo_1975 = round(sum(abs(oscar_1975$`Movie Time` -

```

```
Mediana_tempo_1975))/length(oscar_1975$`Movie Time`), 2)
```

```
Dm_tempo_1975
```

```
cv_media_tempo_1975 = round((Dp_tempo_1975/Media_tempo_1975)*100, 2)
```

```
cv_media_tempo_1975
```

```
cv_mediana_tempo_1975 = round((Dm_tempo_1975/Mediana_tempo_1975)*100, 2)
```

```
cv_mediana_tempo_1975
```

```
#Notas IMDB
```

```
Freq_IMDB_1975 <- fdt(oscar_1975$`IMDB Rating`)
```

```
print(Freq_IMDB_1975, format=TRUE, col = c(1,2,4,5) , pattern = "%.2f ")
```

```
Sumario_IMDB_1975 = summary(oscar_1975$`IMDB Rating`)
```

```
Sumario_IMDB_1975
```

```
Media_IMDB_1975 = round(mean(oscar_1975$`IMDB Rating`), 2)
```

```
Media_IMDB_1975
```

```
Mediana_IMDB_1975 = round(median(oscar_1975$`IMDB Rating`), 2)
```

```
Mediana_IMDB_1975
```

```
Var_IMDB_1975 = round(var(oscar_1975$`IMDB Rating`), 2)
```

```
Var_IMDB_1975
```

```
Dp_IMDB_1975 = round(sd(oscar_1975$`IMDB Rating`), 2)
```

```
Dp_IMDB_1975
```

```
Dm_IMDB_1975 = round(sum(abs(oscar_1975$`IMDB Rating` -  
Mediana_IMDB_1975))/length(oscar_1975$`IMDB Rating`), 2)
```

```
Dm_IMDB_1975
```

```
cv_media_IMDB_1975 = round((Dp_IMDB_1975/Media_IMDB_1975)*100, 2)
```

```
cv_media_IMDB_1975
```

```
cv_mediana_IMDB_1975 = round((Dm_IMDB_1975/Mediana_IMDB_1975)*100, 2)
cv_mediana_IMDB_1975
```

```
#Notas Tomatometer
```

```
Freq_tomatometer_1975 <- fdt(oscar_1975$`Tomatometer Rating`)
print(Freq_tomatometer_1975, format=TRUE, col = c(1,2,4,5) , pattern = "%.2f ")
```

```
Sumario_tomatometer_1975 = summary(oscar_1975$`Tomatometer Rating`)
Sumario_tomatometer_1975
```

```
Media_tomatometer_1975 = round(mean(oscar_1975$`Tomatometer Rating`), 2)
Media_tomatometer_1975
```

```
Mediana_tomatometer_1975 = round(median(oscar_1975$`Tomatometer Rating`), 2)
Mediana_tomatometer_1975
```

```
Var_tomatometer_1975 = round(var(oscar_1975$`Tomatometer Rating`), 2)
Var_tomatometer_1975
```

```
Dp_tomatometer_1975 = round(sd(oscar_1975$`Tomatometer Rating`), 2)
Dp_tomatometer_1975
```

```
Dm_tomatometer_1975 = round(sum(abs(oscar_1975$`Tomatometer Rating` -
Mediana_tomatometer_1975))/length(oscar_1975$`Tomatometer Rating`), 2)
Dm_tomatometer_1975
```

```
cv_media_tomatometer_1975 =
round((Dp_tomatometer_1975/Media_tomatometer_1975)*100, 2)
cv_media_tomatometer_1975
```

```
cv_mediana_tomatometer_1975 =
round((Dm_tomatometer_1975/Mediana_tomatometer_1975)*100, 2)
cv_mediana_tomatometer_1975
```

```
#Estudios
```



```

Freq_estudios_1975 = oscar_1975 %>%
  group_by(oscar_1975$`Film Studio/Producer(s)`) %>%
  summarise(frequencia_absoluta = n()) %>%
  mutate(frequencia_relativa = round(frequencia_absoluta/sum(frequencia_absoluta) *
100, 2))

```

```

Freq_ganhadores_1975_0 = oscar_1975 %>%
  filter(oscar_1975$Award == "Winner")

```

```

Freq_ganhadores_1975 = Freq_ganhadores_1975_0 %>%
  group_by(Freq_ganhadores_1975_0$`Film Studio/Producer(s)`) %>%
  summarise(frequencia_absoluta = n()) %>%
  mutate(frequencia_relativa = round(frequencia_absoluta/sum(frequencia_absoluta) *
100, 2))

```

```

#Genero
genero_1975 = strsplit(oscar_1975$`Movie Genre`, ",")
un_gen_1975 = unlist(genero_1975)
table_gen_1975 = sort(table(un_gen_1975), decreasing = T)
df_gen_1975 = as.data.frame(table_gen_1975)
df_gen_1975 = df_gen_1975 %>%
  top_n(10, Freq)

```

```

# Analise da base de 2000

```

```

#Tempo dos filmes
Freq_tempo_2000 <- fdt(oscar_2000$`Movie Time`)
print(Freq_tempo_2000, format=TRUE, col = c(1,2,4,5) , pattern = "%.2f ")

```

```

Sumario_tempo_2000 = summary(oscar_2000$`Movie Time`)
Sumario_tempo_2000

```

```

Media_tempo_2000 = round(mean(oscar_2000$`Movie Time`), 2)
Media_tempo_2000

```

```

Mediana_tempo_2000 = round(median(oscar_2000$`Movie Time`), 2)

```

Mediana_tempo_2000

Var_tempo_2000 = round(var(oscar_2000\$`Movie Time`), 2)

Var_tempo_2000

Dp_tempo_2000 = round(sd(oscar_2000\$`Movie Time`), 2)

Dp_tempo_2000

Dm_tempo_2000 = round((sum(abs(oscar_2000\$`Movie Time` -
Mediana_tempo_2000)))/length(oscar_2000\$`Movie Time`), 2)

Dm_tempo_2000

cv_media_tempo_2000 = round((Dp_tempo_2000/Media_tempo_2000)*100, 2)

cv_media_tempo_2000

cv_mediana_tempo_2000 = round((Dm_tempo_2000/Mediana_tempo_2000)*100, 2)

cv_mediana_tempo_2000

#Nota IMDB

Freq_IMDB_2000 <- fdt(oscar_2000\$`IMDB Rating`)

print(Freq_IMDB_2000, format=TRUE, col = c(1,2,4,5), pattern = "%.2f ")

Sumario_IMDB_2000 = summary(oscar_2000\$`IMDB Rating`)

Sumario_IMDB_2000

Media_IMDB_2000 = round(mean(oscar_2000\$`IMDB Rating`), 2)

Media_IMDB_2000

Mediana_IMDB_2000 = round(median(oscar_2000\$`IMDB Rating`), 2)

Mediana_IMDB_2000

Var_IMDB_2000 = round(var(oscar_2000\$`IMDB Rating`), 2)

Var_IMDB_2000

Dp_IMDB_2000 = round(sd(oscar_2000\$`IMDB Rating`), 2)

Dp_IMDB_2000

```
Dm_IMDB_2000 = round(sum(abs(oscar_2000$`IMDB Rating` -  
Mediana_IMDB_2000))/length(oscar_2000$`IMDB Rating`), 2)  
Dm_IMDB_2000
```

```
cv_media_IMDB_2000 = round((Dp_IMDB_2000/Media_IMDB_2000)*100, 2)  
cv_media_IMDB_2000
```

```
cv_mediana_IMDB_2000 = round((Dm_IMDB_2000/Mediana_IMDB_2000)*100, 2)  
cv_mediana_IMDB_2000
```

#Notas Tomatometer

```
Freq_tomatometer_2000 <- fdt(oscar_2000$`Tomatometer Rating`)  
print(Freq_tomatometer_2000, format=TRUE, col = c(1,2,4,5), pattern = "%.2f ")
```

```
Sumario_tomatometer_2000 = summary(oscar_2000$`Tomatometer Rating`)  
Sumario_tomatometer_2000
```

```
Media_tomatometer_2000 = round(mean(oscar_2000$`Tomatometer Rating`), 2)  
Media_tomatometer_2000
```

```
Mediana_tomatometer_2000 = round(median(oscar_2000$`Tomatometer Rating`), 2)  
Mediana_tomatometer_2000
```

```
Var_tomatometer_2000 = round(var(oscar_2000$`Tomatometer Rating`), 2)  
Var_tomatometer_2000
```

```
Dp_tomatometer_2000 = round(sd(oscar_2000$`Tomatometer Rating`), 2)  
Dp_tomatometer_2000
```

```
Dm_tomatometer_2000 = round(sum(abs(oscar_2000$`Tomatometer Rating` -  
Mediana_tomatometer_2000))/length(oscar_2000$`Tomatometer Rating`), 2)  
Dm_tomatometer_2000
```

```
cv_media_tomatometer_2000 =
round((Dp_tomatometer_2000/Media_tomatometer_2000)*100, 2)
cv_media_tomatometer_2000
```

```
cv_mediana_tomatometer_2000 =
round((Dm_tomatometer_2000/Mediana_tomatometer_2000)*100, 2)
cv_mediana_tomatometer_2000
```

```
#Estudios
Freq_estudios_2000 = oscar_2000 %>%
  group_by(oscar_2000$`Film Studio/Producer(s)`) %>%
  summarise(frequencia_absoluta = n()) %>%
  mutate(frequencia_relativa = round(frequencia_absoluta/sum(frequencia_absoluta) *
100, 2))
```

```
Freq_ganhadores_2000_0 = oscar_2000 %>%
  filter(oscar_2000$Award == "Winner")
```

```
Freq_ganhadores_2000 = Freq_ganhadores_2000_0 %>%
  group_by(Freq_ganhadores_2000_0$`Film Studio/Producer(s)`) %>%
  summarise(frequencia_absoluta = n()) %>%
  mutate(frequencia_relativa = round(frequencia_absoluta/sum(frequencia_absoluta) *
100, 2))
```

```
# Genero
genero_2000 = strsplit(oscar_2000$`Movie Genre`, ",")
un_gen_2000 = unlist(genero_2000)
table_gen_2000 = sort(table(un_gen_2000), decreasing = T)
df_gen_2000 = as.data.frame(table_gen_2000)
df_gen_2000 = df_gen_2000 %>%
  top_n(10, Freq)
```

```
# Analise da base de 2020
```

```
#Tempo dos filmes
Freq_tempo_2020 <- fdt(oscar_2020$`Movie Time`)
```

```
print(Freq_tempo_2020, format=TRUE, col = c(1,2,4,5) , pattern = "%.2f ")
```

```
Sumario_tempo_2020 = summary(oscar_2020$`Movie Time`)
```

```
Sumario_tempo_2020
```

```
Media_tempo_2020 = round(mean(oscar_2020$`Movie Time`), 2)
```

```
Media_tempo_2020
```

```
Mediana_tempo_2020 = round(median(oscar_2020$`Movie Time`), 2)
```

```
Mediana_tempo_2020
```

```
Var_tempo_2020 = round(var(oscar_2020$`Movie Time`), 2)
```

```
Var_tempo_2020
```

```
Dp_tempo_2020 = round(sd(oscar_2020$`Movie Time`) ,2)
```

```
Dp_tempo_2020
```

```
Dm_tempo_2020 = round(sum(abs(oscar_2020$`Movie Time` -  
Mediana_tempo_2020))/length(oscar_2020$`Movie Time`), 2)
```

```
Dm_tempo_2020
```

```
cv_media_tempo_2020 = round((Dp_tempo_2020/Media_tempo_2020)*100, 2)
```

```
cv_media_tempo_2020
```

```
cv_mediana_tempo_2020 = round((Dm_tempo_2020/Mediana_tempo_2020)*100, 2)
```

```
cv_mediana_tempo_2020
```

```
#Nota IMDB
```

```
Freq_IMDB_2020 <- fdt(oscar_2020$`IMDB Rating`)
```

```
print(Freq_IMDB_2020, format=TRUE, col = c(1,2,4,5) , pattern = "%.2f ")
```

```
Sumario_IMDB_2020 = summary(oscar_2020$`IMDB Rating`)
```

```
Sumario_IMDB_2020
```

```
Media_IMDB_2020 = round(mean(oscar_2020$`IMDB Rating`), 2)
```

Media_IMDB_2020

Mediana_IMDB_2020 = round(median(oscar_2020\$`IMDB Rating`), 2)

Mediana_IMDB_2020

Var_IMDB_2020 = round(var(oscar_2020\$`IMDB Rating`), 2)

Var_IMDB_2020

Dp_IMDB_2020 = round(sd(oscar_2020\$`IMDB Rating`), 2)

Dp_IMDB_2020

Dm_IMDB_2020 = round(sum(abs(oscar_2020\$`IMDB Rating` -
Mediana_IMDB_2020))/length(oscar_2020\$`IMDB Rating`), 2)

Dm_IMDB_2020

cv_media_IMDB_2020 = round((Dp_IMDB_2020/Media_IMDB_2020)*100, 2)

cv_media_IMDB_2020

cv_mediana_IMDB_2020 = round((Dm_IMDB_2020/Mediana_IMDB_2020)*100, 2)

cv_mediana_IMDB_2020

#Notas Tomatometer

Freq_tomatometer_2020 <- fdt(oscar_2020\$`Tomatometer Rating`)

print(Freq_tomatometer_2020, format=TRUE, col = c(1,2,4,5), pattern = "%.2f ")

Sumario_tomatometer_2020 = summary(oscar_2020\$`Tomatometer Rating`)

Sumario_tomatometer_2020

Media_tomatometer_2020 = round(mean(oscar_2020\$`Tomatometer Rating`), 2)

Media_tomatometer_2020

Mediana_tomatometer_2020 = round(median(oscar_2020\$`Tomatometer Rating`), 2)

Mediana_tomatometer_2020

Var_tomatometer_2020 = round(var(oscar_2020\$`Tomatometer Rating`), 2)

```
Var_tomatometer_2020
```

```
Dp_tomatometer_2020 = round(sd(oscar_2020$`Tomatometer Rating`), 2)
```

```
Dp_tomatometer_2020
```

```
Dm_tomatometer_2020 = round(sum(abs(oscar_2020$`Tomatometer Rating` -  
Mediana_tomatometer_2020))/length(oscar_2020$`Tomatometer Rating`), 2)
```

```
Dm_tomatometer_2020
```

```
cv_media_tomatometer_2020 =
```

```
round((Dp_tomatometer_2020/Media_tomatometer_2020)*100, 2)
```

```
cv_media_tomatometer_2020
```

```
cv_mediana_tomatometer_2020 =
```

```
round((Dm_tomatometer_2020/Mediana_tomatometer_2020)*100, 2)
```

```
cv_mediana_tomatometer_2020
```

```
#Estudios
```

```
Freq_estudios_2020 = oscar_2020 %>%
```

```
group_by(oscar_2020$`Film Studio/Producer(s)`) %>%
```

```
summarise(frequencia_absoluta = n()) %>%
```

```
mutate(frequencia_relativa = round(frequencia_absoluta/sum(frequencia_absoluta) *  
100, 2))
```

```
Freq_ganhadores_2020_0 = oscar_2020 %>%
```

```
filter(oscar_2020$Award == "Winner")
```

```
Freq_ganhadores_2020 = Freq_ganhadores_2020_0 %>%
```

```
group_by(Freq_ganhadores_2020_0$`Film Studio/Producer(s)`) %>%
```

```
summarise(frequencia_absoluta = n()) %>%
```

```
mutate(frequencia_relativa = round(frequencia_absoluta/sum(frequencia_absoluta) *  
100, 2))
```

```
# Genero
```

```
genero_2020 = strsplit(oscar_2020$`Movie Genre`, ",")
```

```

un_gen_2020 = unlist(genero_2020)
table_gen_2020 = sort(table(un_gen_2020), decreasing = T)
df_gen_2020 = as.data.frame(table_gen_2020)
df_gen_2020 = df_gen_2020 %>%
  top_n(10, Freq)

```

#Geral dos Oscars

```

#Tempo
Freq_tempo <- fdt(oscars_df$`Movie Time`)
print(Freq_tempo, format=TRUE, col = c(1,2,4,5) , pattern = "%.2f ")

```

```

Sumario_tempo = summary(oscars_df$`Movie Time`)
Sumario_tempo

```

```

Media_tempo = round(mean(oscars_df$`Movie Time`), 2)
Media_tempo

```

```

Mediana_tempo = round(median(oscars_df$`Movie Time`), 2)
Mediana_tempo

```

```

Var_tempo = round(var(oscars_df$`Movie Time`), 2)
Var_tempo

```

```

Dp_tempo = round(sd(oscars_df$`Movie Time`), 2)
Dp_tempo

```

```

Dm_tempo = round((sum(abs(oscars_df$`Movie Time` -
Mediana_tempo)))/length(oscars_df$`Movie Time`), 2)
Dm_tempo

```

```

cv_media_tempo = round((Dp_tempo/Media_tempo)*100, 2)
cv_media_tempo

```

```

cv_mediana_tempo = round((Dm_tempo/Mediana_tempo)*100, 2)
cv_mediana_tempo

```


#Notas IMDB

```
Freq_IMDB <- fdt(oscars_df$`IMDB Rating`)  
print(Freq_IMDB, format=TRUE, col = c(1,2,4,5), pattern = "%.2f ")
```

```
Sumario_IMDB = summary(oscars_df$`IMDB Rating`)  
Sumario_IMDB
```

```
Media_IMDB = round(mean(oscars_df$`IMDB Rating`), 2)  
Media_IMDB
```

```
Mediana_IMDB = round(median(oscars_df$`IMDB Rating`), 2)  
Mediana_IMDB
```

```
Var_IMDB = round(var(oscars_df$`IMDB Rating`), 2)  
Var_IMDB
```

```
Dp_IMDB = round(sd(oscars_df$`IMDB Rating`), 2)  
Dp_IMDB
```

```
Dm_IMDB = round(sum(abs(oscars_df$`IMDB Rating` -  
Mediana_IMDB))/length(oscars_df$`IMDB Rating`), 2)  
Dm_IMDB
```

```
cv_media_IMDB = round((Dp_IMDB/Media_IMDB)*100, 2)  
cv_media_IMDB
```

```
cv_mediana_IMDB = round((Dm_IMDB/Mediana_IMDB)*100, 2)  
cv_mediana_IMDB
```

#Notas Tomatometer

```
Freq_tomatometer <- fdt(oscars_df$`Tomatometer Rating`)  
print(Freq_tomatometer, format=TRUE, col = c(1,2,4,5), pattern = "%.2f ")
```

```
Sumario_tomatometer = summary(oscars_df$`Tomatometer Rating`)
```

Sumario_tomatometer

```
Media_tomatometer = round(mean(oscars_df$`Tomatometer Rating`), 2)
```

Media_tomatometer

```
Mediana_tomatometer = round(median(oscars_df$`Tomatometer Rating`), 2)
```

Mediana_tomatometer

```
Var_tomatometer = round(var(oscars_df$`Tomatometer Rating`), 2)
```

Var_tomatometer

```
Dp_tomatometer = round(sd(oscars_df$`Tomatometer Rating`), 2)
```

Dp_tomatometer

```
Dm_tomatometer = round(sum(abs(oscars_df$`Tomatometer Rating` -  
Mediana_tomatometer))/length(oscars_df$`Tomatometer Rating`), 2)
```

Dm_tomatometer

```
cv_media_tomatometer = round((Dp_tomatometer/Media_tomatometer)*100, 2)
```

cv_media_tomatometer

```
cv_mediana_tomatometer = round((Dm_tomatometer/Mediana_tomatometer)*100, 2)
```

cv_mediana_tomatometer

#Estudios

```
Freq_estudios = oscars_df %>%
```

```
  group_by(oscars_df$`Film Studio/Producer(s)`) %>%
```

```
  summarise(frequencia_absoluta = n()) %>%
```

```
  mutate(frequencia_relativa = round(frequencia_absoluta/sum(frequencia_absoluta) *  
100, 2))
```

```
Estudios_10 = Freq_estudios %>%
```

```
  top_n(10, frequencia_absoluta)
```

```
Freq_ganhadores_0 = oscars_df %>%
```

```
filter(oscars_df$Award == "Winner")
```

```
Freq_ganhadores = Freq_ganhadores_0 %>%  
  group_by(Freq_ganhadores_0$`Film Studio/Producer(s)`) %>%  
  summarise(frequencia_absoluta = n()) %>%  
  mutate(frequencia_relativa = round(frequencia_absoluta/sum(frequencia_absoluta) *  
100, 2))
```

```
Ganhadores_10 = Freq_ganhadores %>%  
  top_n(10, frequencia_absoluta)
```

```
#Genero
```

```
genero = strsplit(oscars_df$`Movie Genre`, ",")  
un_gen = unlist(genero)  
table_gen = sort(table(un_gen))  
df_gen = as.data.frame(table_gen)  
df_gen = df_gen %>%  
  top_n(10, Freq)
```

```
#Graficos
```

```
#Boxplot Tempo
```

```
bp_tempo_geral = ggplot(data = oscars_df, aes(y = oscars_df$`Movie Time`, x = "")) +  
  geom_errorbar(stat = "boxplot", width = 0.1) +  
  geom_boxplot(width = 0.3, fill = "#bd925a", outlier.shape = 1, outlier.size = 3) +  
  labs(title = "Figura B", y = "", x = "") +  
  ylim(50, 250)+  
  theme_classic()
```

```
bp_tempo_dividido = ggplot(data = oscar_dados, aes(y = oscar_dados$`Movie Time`, x =  
oscar_dados$Ano_limite)) +  
  geom_errorbar(stat = "boxplot", width = 0.1) +  
  geom_boxplot(width = 0.6, fill = "#edeac2", outlier.shape = 1, outlier.size = 3) +  
  labs(title = "Figura A", y = "Tempo dos filmes (em min)", x = "Faixas de tempo") +  
  ylim(50, 250)+  
  theme_classic()
```

```
plot_grid(bp_tempo_dividido, bp_tempo_geral)
```

```
#Boxplot IMDB
```

```
bp_IMDB_dividido = ggplot(data = oscar_dados, aes(y = oscar_dados$`IMDB Rating`, x =  
oscar_dados$Ano_limite)) +  
  geom_errorbar(stat = "boxplot", width = 0.1) +  
  geom_boxplot(width = 0.6, fill = "#FFC000", outlier.shape = 1, outlier.size = 3) +  
  labs(title = "Figura A", y = "Notas no IMDB", x = "Faixas de tempo") +  
  ylim(5, 10)+  
  theme_classic()
```

```
bp_IMDB_geral = ggplot(data = oscars_df, aes(y = oscars_df$`IMDB Rating`, x = "")) +  
  geom_errorbar(stat = "boxplot", width = 0.1) +  
  geom_boxplot(width = 0.3, fill = "#F38630FF", outlier.shape = 1, outlier.size = 3) +  
  labs(title = "Figura B", y = "", x = "") +  
  ylim(5, 10)+  
  theme_classic()
```

```
plot_grid(bp_IMDB_dividido, bp_IMDB_geral)
```

```
#Boxplot Tomatometer
```

```
bp_tomatometer_dividido = ggplot(data = oscar_dados, aes(y =  
oscar_dados$`Tomatometer Rating`, x = oscar_dados$Ano_limite)) +  
  geom_errorbar(stat = "boxplot", width = 0.1) +  
  geom_boxplot(width = 0.6, fill = "#D03840", outlier.shape = 1, outlier.size = 3) +  
  labs(title = "Figura A", y = "Tomatometer", x = "Faixas de tempo") +  
  ylim(10, 100)+  
  theme_classic()
```

```
bp_tomatometer_geral = ggplot(data = oscars_df, aes(y = oscars_df$`Tomatometer  
Rating`, x = "")) +  
  geom_errorbar(stat = "boxplot", width = 0.1) +  
  geom_boxplot(width = 0.3, fill = "#B40F20", outlier.shape = 1, outlier.size = 3) +  
  labs(title = "Figura B", y = "", x = "") +
```

```
ylim(10, 100)+  
theme_classic()
```

```
plot_grid(bp_tomatometer_dividido, bp_tomatometer_geral)
```

```
#Gráfico de dispersão  
ggplot(data= media_notas_ano, aes(x= media_notas_IMDB, y=  
media_notas_tomatometer)) +  
  geom_point(mapping = aes(x= media_notas_IMDB, y= media_notas_tomatometer),  
color="black", size=4,shape=20) +  
  xlab("Média das notas do IMDB") + ylab("Média das notas do tomatometer") +  
  theme_bw() +  
  xlim(6,9) +  
  ylim(6,10)
```

```
cor(media_notas_ano$media_notas_IMDB,  
media_notas_ano$media_notas_tomatometer)
```

```
#Gráfico de linhas Notas por Ano  
media_IMDB_ano = oscars_df %>%  
  group_by(`Year of Release`) %>%  
  summarize(media_notas_IMDB = round(mean(`IMDB Rating`), 2))
```

```
media_tomatometer_ano = oscars_df %>%  
  group_by(`Year of Release`) %>%  
  summarize(media_notas_tomatometer = round(mean(`Tomatometer Rating`)/10), 2))
```

```
media_notas_ano = cbind(media_IMDB_ano, media_tomatometer_ano[,  
!(names(media_tomatometer_ano) %in% "Year of Release")])  
view(media_notas_ano)
```

```
bivariada_notas = ggplot(data = media_notas_ano, aes(x = media_notas_ano$`Year of  
Release`, y = media_notas_ano$media_notas_IMDB)) +  
  geom_line(aes(col = "Média das notas do IMDB")) +  
  geom_point(colour = "#FFC000") +
```

```

  geom_line(aes(y = media_notas_ano$media_notas_tomatometer, col = "Média das notas
do tomatometer")) +
  geom_point(aes(y = media_notas_ano$media_notas_tomatometer), colour = "#B40F20")
+
  theme_classic() +
  labs(y = "Média das notas (de 0 a 10)", x = "Anos") +
  geom_hline(yintercept = seq(1, 10, by = 1), color = "gray75", linetype = "dashed", alpha =
0.8) +
  ylim(6,10) +
  xlim(1925,2025) +
  scale_color_manual(values = c("#FFC000", "#B40F20")) +
  theme(legend.position = "bottom") +
  transition_reveal(media_notas_ano$`Year of Release`)

animate(bivariada_notas, renderer = gifski_renderer(), nframes = 200)
anim_save("notas.gif")

```

#Gráfico de barras Gênero

```

ggplot(data = df_gen) +
  geom_bar(width = 0.8, aes(x = fct_rev(un_gen), y = Freq), stat = "identity", fill =
"#284878") +
  geom_hline(yintercept = seq(50, 500, by = 50), color = "gray75", linetype = "dashed",
alpha = 0.7) +
  ylim(0, 500) +
  theme_classic() +
  scale_y_continuous(expand = c(0, 0)) +
  labs(y = "Frequência", x = "Gênero")

```

```

bar_1950 = ggplot(data = df_gen_1950) +
  geom_bar(width = 0.8, aes(x = fct_rev(df_gen_1950$un_gen), y = df_gen_1950$Freq),
stat = "identity", fill = "#78B7C5") +
  geom_hline(yintercept = seq(25, 150, by = 25), color = "gray75", linetype = "dashed",
alpha = 0.8) +
  ylim(0, 150) +
  theme_classic() +

```

```
scale_y_continuous(expand = c(0, 0)) +  
labs(title = "De 1927 até 1950", y = "Frequência", x = "Gênero")
```

```
bar_1975 = ggplot(data = df_gen_1975) +  
  geom_bar(width = 0.8, aes(x = fct_rev(df_gen_1975$un_gen), y = df_gen_1975$Freq),  
stat = "identity", fill = "#78B7C5") +  
  geom_hline(yintercept = seq(25, 150, by = 25), color = "gray75", linetype = "dashed",  
alpha = 0.8) +  
  ylim(0, 150) +  
  theme_classic() +  
  scale_y_continuous(expand = c(0, 0)) +  
  labs(title = "De 1950 até 1975", y = "Frequência", x = "Gênero")
```

```
bar_2000 = ggplot(data = df_gen_2000) +  
  geom_bar(width = 0.8, aes(x = fct_rev(df_gen_2000$un_gen), y = df_gen_2000$Freq),  
stat = "identity", fill = "#78B7C5") +  
  geom_hline(yintercept = seq(25, 150, by = 25), color = "gray75", linetype = "dashed",  
alpha = 0.8) +  
  ylim(0, 150) +  
  theme_classic() +  
  scale_y_continuous(expand = c(0, 0)) +  
  labs(title = "De 1975 até 2000", y = "Frequência", x = "Gênero")
```

```
bar_2020 = ggplot(data = df_gen_2020) +  
  geom_bar(width = 0.8, aes(x = fct_rev(df_gen_2020$un_gen), y = df_gen_2020$Freq),  
stat = "identity", fill = "#78B7C5") +  
  geom_hline(yintercept = seq(25, 150, by = 25), color = "gray75", linetype = "dashed",  
alpha = 0.8) +  
  ylim(0, 150) +  
  theme_classic() +  
  scale_y_continuous(expand = c(0, 0)) +  
  labs(title = "De 2000 até 2020", y = "Frequência", x = "Gênero")
```

```
plot_grid(bar_1950, bar_1975, bar_2000, bar_2020)
```

#Gráfico de barras Estúdio/Produtores

```
ggplot(data = Estudios_10) +  
  geom_bar(width = 0.8, aes(x = fct_reorder(Estudios_10$`oscars_df$`Film  
Studio/Producer(s)``, Estudios_10$frequencia_absoluta, .desc = TRUE), y =  
Estudios_10$frequencia_absoluta), stat = "identity", fill = "#B8E186") +  
  geom_hline(yintercept = seq(10, 40, by = 10), color = "gray75", linetype = "dashed",  
alpha = 0.7) +  
  theme_classic() +  
  ylim(0,50) +  
  scale_y_continuous(expand = c(0, 0)) +  
  labs(y = "Frequência", x = "Estúdios Indicados")
```

```
ggplot(data = Ganhadores_10) +  
  geom_bar(width = 0.8, aes(x = fct_reorder(Ganhadores_10$`Freq_ganhadores_0$`Film  
Studio/Producer(s)``, Ganhadores_10$frequencia_absoluta, .desc = TRUE), y =  
Ganhadores_10$frequencia_absoluta), stat = "identity", fill = "#a07e28") +  
  geom_hline(yintercept = seq(1, 5, by = 1), color = "gray75", linetype = "dashed", alpha =  
0.7) +  
  theme_classic() +  
  ylim(0,5) +  
  scale_y_continuous(expand = c(0, 0)) +  
  labs(y = "Frequência", x = "Estúdios Ganhadores")
```