

Análise IMDB

Katerine Witkoski

Meu caderno de instruções nº 1

Vamos fazer nossa primeira análise de dados, e eu Katerine vou tentar auxiliar vocês nessa jornada. Nosso objetivo é impressionar dois produtores de filmes e auxiliar eles a identificar qual seria o melhor filme para esse ano obtendo sucesso financeiro e reconhecimento da crítica. Vamos ajuda-los a concorrer o Oscar de 2021!!



O nosso primeiro passo é importar os dados, no meu caso minha base se encontrava na pasta `data_raw` no formato `.rds`.

Importando base

```
imdb <- readr::read_rds("data_raw/imdb.rds")
```

Aqui nos importamos usando o pacote `readr` utilizando a expressão `::` com a função `readr::read_rds` específica para importar arquivos no formato `rds`.

Vamos carregar os pacotes para esse projeto:

```
library(tidyverse)
library(ggplot2)
library(RColorBrewer)
library(ggthemes)
```

Usamos a função `glimpse` uma função do pacote `tidyverse` para visualizar todas as informações da base, ela mostra as suas variáveis e suas características. Com essas informações será mais fácil chegar a uma tomada de decisão na sua análise. Portanto temos que observar os seguintes pontos quais são as melhores notas, melhores atores (eles estão relacionados), melhores diretores e o gênero que mais teve notas altas.

```
glimpse(imdb)
```

```
## Rows: 3,713
## Columns: 15
## $ titulo      <chr> "Avatar", "Pirates of the Caribbean: At World's End"...
## $ ano         <int> 2009, 2007, 2012, 2012, 2007, 2010, 2015, 2006...
## $ diretor     <chr> "James Cameron", "Gore Verbinski", "Christopher Nola...
## $ duracao     <int> 178, 169, 164, 132, 156, 100, 141, 183, 169, 151, 15...
## $ cor         <chr> "Color", "Color", "Color", "Color", "Color", "Color"...
```

```
## $ generos      <chr> "Action|Adventure|Fantasy|Sci-Fi", "Action|Adventure...
## $ pais        <chr> "USA", "USA", "USA", "USA", "USA", "USA", "USA", "US...
## $ classificacao <chr> "A partir de 13 anos", "A partir de 13 anos", "A par...
## $ orcamento   <int> 237000000, 300000000, 250000000, 263700000, 25800000...
## $ receita      <int> 760505847, 309404152, 448130642, 73058679, 336530303...
## $ nota_imdb    <dbl> 7.9, 7.1, 8.5, 6.6, 6.2, 7.8, 7.5, 6.9, 6.1, 7.3, 6....
## $ likes_facebook <int> 33000, 0, 164000, 24000, 0, 29000, 118000, 197000, 0...
## $ ator_1       <chr> "CCH Pounder", "Johnny Depp", "Tom Hardy", "Daryl Sa...
## $ ator_2       <chr> "Joel David Moore", "Orlando Bloom", "Christian Bale...
## $ ator_3       <chr> "Wes Studi", "Jack Davenport", "Joseph Gordon-Levitt..."
```

Entendendo a problemática

Conhecendo as características da base podemos chegar nas seguintes conclusões:

1. Quais são os Top 10 filmes?
2. Qual é relação entre a Receita e o Orçamento do filme? Houve lucro?
3. Lucro dos filmes com as melhores notas?
4. Qual são os 5 maiores Diretores? Qual é a relação de lucro entre eles?
5. Qual é a duração de um filme de sucesso?
6. Quais são os 10 melhores atores? Qual é a nota média deles?
7. Quais são os gêneros mais badalados?

Explorando a base e manipulando dados

Antes de tudo criamos uma lógica em que *receita* - *orçamento* poderá me dar o lucro.

LUCRO = receita - orçamento

```
imdb <- imdb %>% mutate(lucro = receita - orcamento)
```

1. Quais são os Top 10 Filmes?

Posição	Filme	Lucro
1	Avatar	\$523,505,847
2	Jurassic World	\$502,177,271
3	Titanic	\$458,672,302
4	Star Wars: Episode IV - A New Hope	\$449,935,665
5	E.T. the Extra-Terrestrial	\$424,449,459
6	The Avengers	\$403,279,547
7	The Lion King	\$377,783,777
8	Star Wars: Episode I - The Phantom Menace	\$359,544,677
9	The Dark Knight	\$348,316,061
10	The Hunger Games	\$329,999,255

Observamos que pela análise temos como o melhor classificado o Avatar, escrito e dirigido por James Cameron, seu lucro foi de 523505,847 dólares. Avatar foi nomeado em nove categorias do Oscar, incluído Melhor Filme e Melhor Diretor, porém foi premiado apenas em três categorias: Melhor Fotografia, Melhores Efeitos Visuais e Melhor Direção de Arte. Chique demais, quem não pegou aquela sessão 3D que era mega novidade para assistir o lançamento?? Foi emocionante!



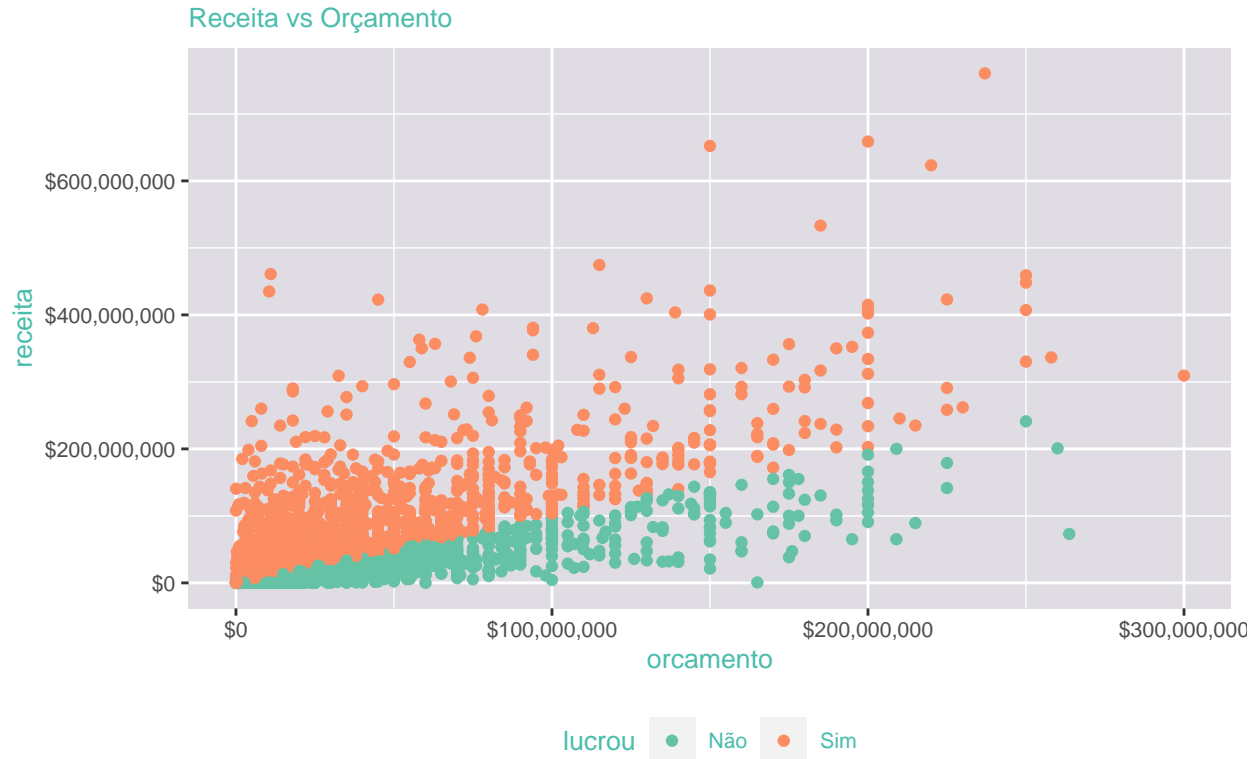
Criando uma função de tema

Para padronizar os gráficos eu criei um tema base:

```
tema_analise <- function() {  
  theme(  
    text = element_text(  
      color = "#4ABDAC",  
      size = 10),  
    plot.title = element_text(  
      colour = "#FC4A1A",  
      hjust = 0.5,  
      size = 20),  
    panel.background = element_rect(fill = "#DFDCE3"),  
    legend.position = "bottom")  
}
```

2. Qual é relação entre a Receita e o Orçamento do filme?

Houve Lucro ou Não?



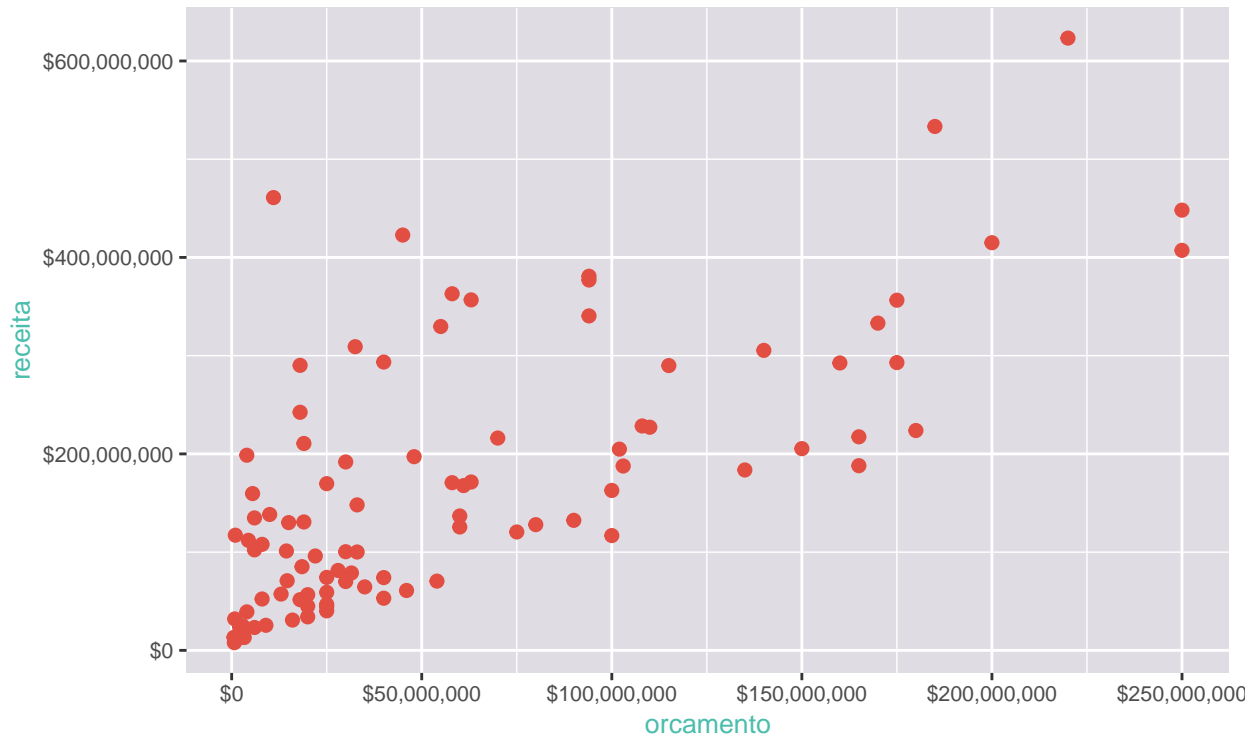
Para interpretar esse parâmetro eu utilizei o método do gráfico de dispersão, observando a correlação entre os eixos *receita* e *orçamento*. Note-se que a maioria se encontra em um orçamento abaixo de 100.000,00 dolares e a receita de R\$ 200.000,00 dolares. Nesse gráfico vimos alguns pontos de outliers, excedendo dos valores esperados.

3. Lucro dos filmes com as melhores notas?

Vamos fazer um gráfico de dispersão da lucro maior de 50000 em filmes com nota imdb > 8 x orçamento *É considerado melhores notas aqueles filmes que tem nota acima de 8*

Lucro das Melhores Notas

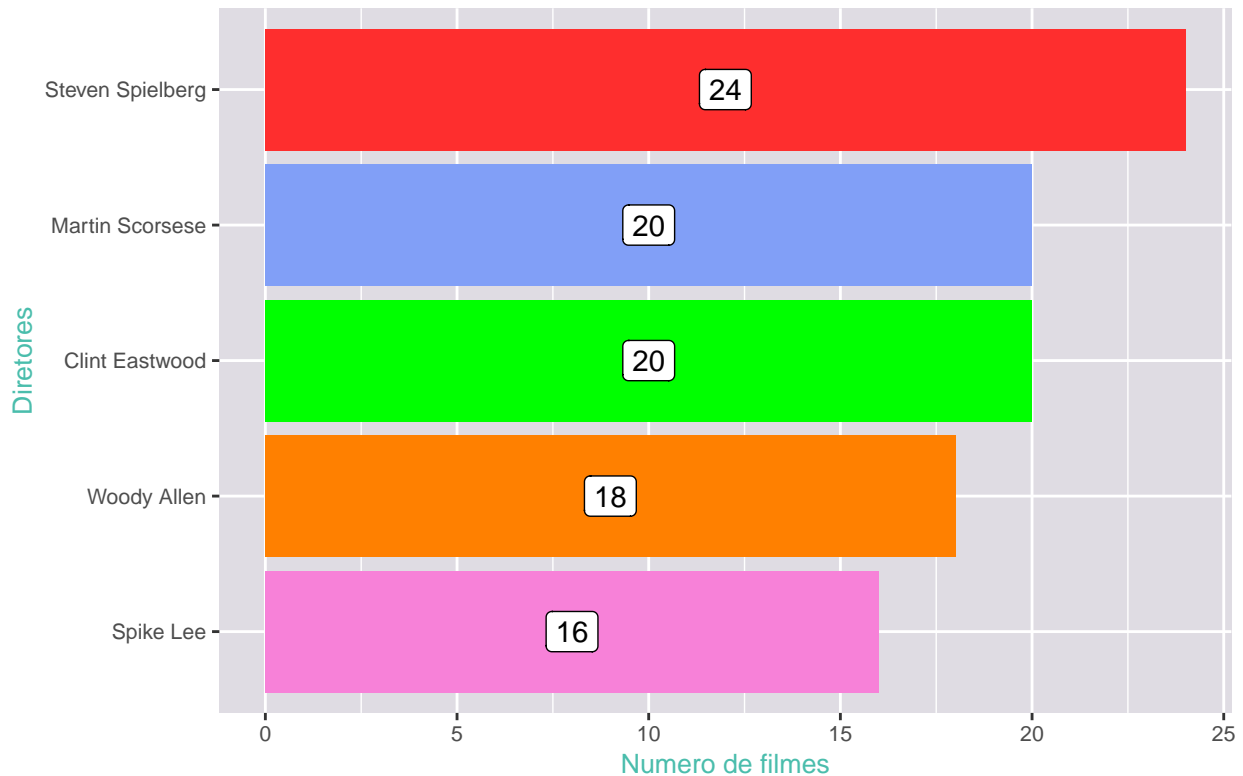
Lucro maior de 5000000 em filmes com nota imdb maiores de 8 vs Orçamento



Nesse gráfico observamos que a relação entre os pontos um crescimento linear, contendo uma distância relevante entre elas, a qual faz entender que existe uma possível correlação positiva entre as variáveis *receita* e *orçamento*, significando que a variável afeta positivamente a outra.

4. Qual são os 5 maiores Diretores? Qual é a relação de lucro entre eles?

Os Melhores Diretores



Com esse gráfico de colunas, para os amantes de cinema, temos uma boa notícia. Classificamos os melhores Diretores:

- Steven Spilberg, diretor dos filmes que você provavelmente já assistiu 567864 vezes em alguma sessão na TV aberta como “E.T.: O Extraterrestre”, “Tubarão”, “Contatos Imediatos”, entre outros;
- Martin Scorsese, o diretor mais “gangster, com os filmes queridinhos do cinema como”Taxi- Driver“,”Os Infiltrados“,”O irlandês”, entre outros;
- Clint Eastwood, o diretor mais bang-bang da lista, produtor de filmes como “Os Imperdoáveis”, “Sniper Americano”, mas não podemos esquecer das interpretações como ator em filmes com a temática faroeste, o mais memorável “Três Homens em Comflito”. E acredita TODO filme de foroeste terá a famosa trilha sonora composta pelo Ennio Morricone, feita exclusivamente para seu tema principal;
- Woody Allen, aquele diretor mais meloso e carismático com suas comédias romanticas “Annie Hall” e “Manhattan”, para quem curte um momento fofura ele é uma ótima sugestão;
- Spike Lee, um diretor que tem discursos fortes contra o preconceito racial, ele trás reflexão em filmes como “Ela quer tudo”, “Malcom X”, “Quatro Meninas: Uma história real”, entre outros;

e vamos falar de lucro...

Lucro dos diretores



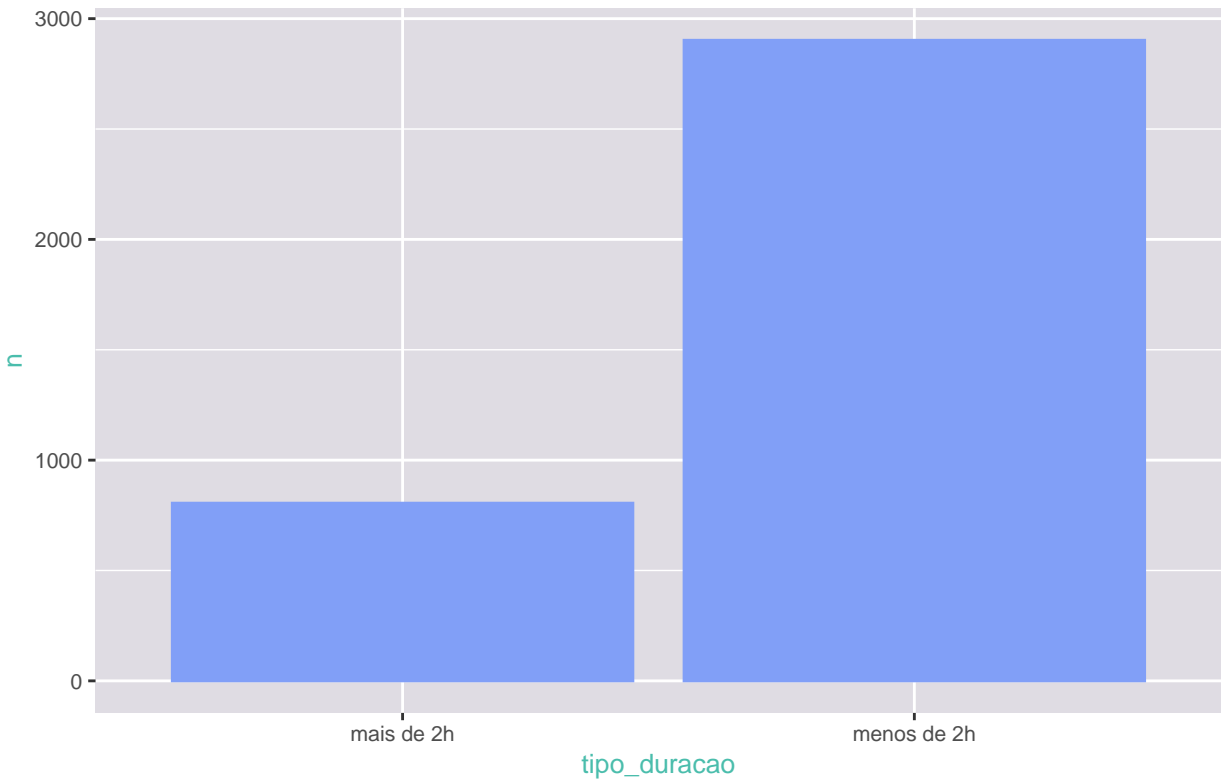
Observamos que entre os diretores Clint Eastwood e Steven Spielberg encontramos outliers, excedendo o valor esperado. Sendo o máximo 300.000,00 dolares e o mínimo 100.000,00.

5. Qual é a duração de um filme de sucesso?

Se há algum tempo apenas grandes épicos e filmes históricos costumavam passar das 2 horas e meia, agora não há mais um limite para gêneros, mas sabemos que para concorrer ao Oscar somente 40 minutos de filmagem já é o suficiente para ser indicado.

Portanto podemos analisar:

Filmes com Duração de 2 Horas



Entre os filmes da base observamos que a maioria contém uma duração de menos de 2 horas.

6. Quais são os 10 melhores atores? Qual é a nota média deles?

Um filmão precisa que um ator a altura, alguém versátil de uma atuação mais de tirar o folego e menos Kristen Stewart no filme “Crepúsculo”.



Realizei uma análise da média das notas entre os melhores atores relacionados a filmes com notas acima de 8:

```
## # A tibble: 640 x 4
## # Groups:   ator, nota_imdb [640]
##   ator          nota_imdb data          media_das_notas
##   <chr>          <dbl> <list>          <dbl>
## 1 Morgan Freeman    9.3 <tibble [1 x 0]>    9.3
## 2 Jeffrey DeMunn    9.3 <tibble [1 x 0]>    9.3
## 3 Bob Gunton        9.3 <tibble [1 x 0]>    9.3
## 4 Al Pacino         9.2 <tibble [1 x 0]>    9.2
## 5 Marlon Brando     9.2 <tibble [1 x 0]>    9.2
## 6 Robert Duvall     9.2 <tibble [1 x 0]>    9.2
## 7 Matthew Ziff      9.1 <tibble [1 x 0]>    9.1
## 8 T.J. Storm        9.1 <tibble [1 x 0]>    9.1
## 9 Sam Medina        9.1 <tibble [1 x 0]>    9.1
## 10 Christian Bale   9    <tibble [1 x 0]>    9
## # ... with 630 more rows
```

Concluindo os 5 melhores:

Atores	Media das notas
Morgan Freeman	9.3
Jeffrey DeMunn	9.3
Bob Gunton	9.3
Al Pacino	9.2
Marlon Brando	9.2

7. Quais são os gêneros mais badalados? Está na hora de decidir, tem muitos tipos de gêneros de filmes e cada um deles desperta uma sensação diferente na hora de assistir, e aqui podemos ver qual é o gênero mais queridinho:

generos	n
Comedy	185
Comedy Drama	152
Comedy Drama Romance	149
Comedy Romance	135
Drama	166

Observamos que o gênero que alcançou o coração do telespectador é o Comédia.

Mas vamos pensar no lucro:

Média de Lucro dos Filmes de Comédia



A média do lucro está perto do 25.000,000 e nesse gráfico podemos ver que boa parte dos filmes atingiram a

meta para o sucesso.

Conclusão

Para um filme de sucesso, além da criatividade dos produtores chegamos na conclusão que a fim de criar uma experiência incrível em meio desses tempos de pandemia. Faremos o pessoal rir um pouco com o gênero **Comédia**, na direção do nosso queridinho **Steven Spielberg**, de duração de pelo menos 2 horas ou menor, com os atores protagonistas **Morgan Freeman** e **Jeffrey DeMunn**.

