Anderson Amorim

https://www.kaggle.com/datasets/octopusteam/imdb-top-1000-movies/data

```r
a <- ggplot(filmes, aes(x = averageRating)) +
  geom_density(fill = "cyan4", alpha = .5, adjust = 1.5, color = NA) +
  labs(x = "Avaliação Média", y = "Densidade") +
  theme_tufte() +
  theme(axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=0.5)) +
  geom_vline(aes(xintercept = mean(averageRating)), color="gray", linetype="dashed", size=1)

b <- ggplot(filtered_filmes, aes(x = numVotes, y = averageRating)) +
  geom_point(aes(color = ifelse(grepl("Drama", genres), "Drama", "Outros"),
                 alpha = ifelse(grepl("Drama", genres), 0.3, 0.5))) +
  labs(x = "Número de Votos",
       y = "Avaliação Média") +
  scale_x_log10(labels = scales::comma) +
  scale_color_manual(values = c("Drama" = "orange", "Outros" = "cyan4")) +
  scale_alpha_identity() +
  theme_tufte() +
  theme(legend.position = "none")

c <- filmes %>%
  group_by(releaseYear) %>%
  summarise(numMovies = n()) %>%
  ggplot(aes(x = releaseYear, y = numMovies)) +
  geom_line(color = "cyan4", alpha = .5) +
  labs(x = "Ano de Lançamento", y = "Número de Filmes") +
  theme_tufte()

d <- filmes %>%
  separate_rows(genres, sep = ",") %>%
  mutate(genres = trimws(genres)) %>%
  group_by(genres) %>%
  summarise(numMovies = n(), avgRating = mean(averageRating)) %>%
  arrange(desc(numMovies)) %>%
  slice_head(n = 10) %>%
  ggplot(aes(x = reorder(genres, numMovies), y = numMovies, fill = genres == "Drama")) +
  geom_bar(stat = "identity", alpha = 0.5) +
  coord_flip() +
  labs(x = "Gênero", y = "Número de Filmes") +
  scale_fill_manual(values = c("FALSE" = "cyan4", "TRUE" = "orange")) +
  theme_tufte() +
  theme(legend.position = "none")
```

1000 filmes mais bem avaliados - IMDB