Are Films Getting Worse?

Conor Cassidy

2023-01-07

Cinematic Slump. . .

I recently went to the cinema with a friend to see the new Black Panther film (I wouldn't recommend. . .)
which sparked an interesting conversation afterwards - are films getting worse? We were trying to rack our
brains of films that have came out in the last few years that were among our favourites. After we struggled
to think of many, we took to IMDb to check if we were just being forgetful or if there genuinely were not
that many great films released recently. I decided to do a little analysis of the Top 250 films, according to
IMDb, over the years to see if they are indeed getting worse. The list can be found here via the following
url: https://www.imdb.com/chart/top/.

The Data

Feel free to skip over this part - it just explains the technicalities of how I scraped the data for this analysis.
It was obtained from the IMDb website's Top 250 Films List - given in url given above. The code for this is
shown below, mostly using the tools from the rvest package. Once I had scraped the data, I just put it into
the neat format provided by a 'tibble' table.

```
library(tibble)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.4.0      v dplyr   1.0.10
## v tidyr   1.2.1      v stringr 1.5.0
## v readr   2.1.3      v forcats 0.5.2
## v purrr   1.0.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(robotstxt)
library(dplyr)
library(rvest)
```

```
##
## Attaching package: 'rvest'
##
## The following object is masked from 'package:readr':
##
##     guess_encoding
```

```
page <- read_html("https://www.imdb.com/chart/top/")

paths_allowed("https://www.imdb.com/chart/top/")
```

www.imdb.com

[1] TRUE

```
typeof(page)
```

[1] "list"

```
class(page)
```

[1] "xml_document" "xml_node"

```
titles <- page %>%
  html_nodes(".titleColumn a") %>%
  html_text()

years <- page %>%
  html_nodes(".secondaryInfo") %>%
  html_text() %>%
  str_remove("\\(") %>%
  str_remove("\\)") %>%
  as.numeric()

ratings <- page %>%
  html_nodes("strong") %>%
  html_text() %>%
  as.numeric()

imdb_top_250 <- tibble(
  title = titles,
  year = years,
  rating = ratings
)

imdb_top_250 <- imdb_top_250 %>%
  mutate(rank = 1:nrow(imdb_top_250)) %>%
  relocate(rank)
```
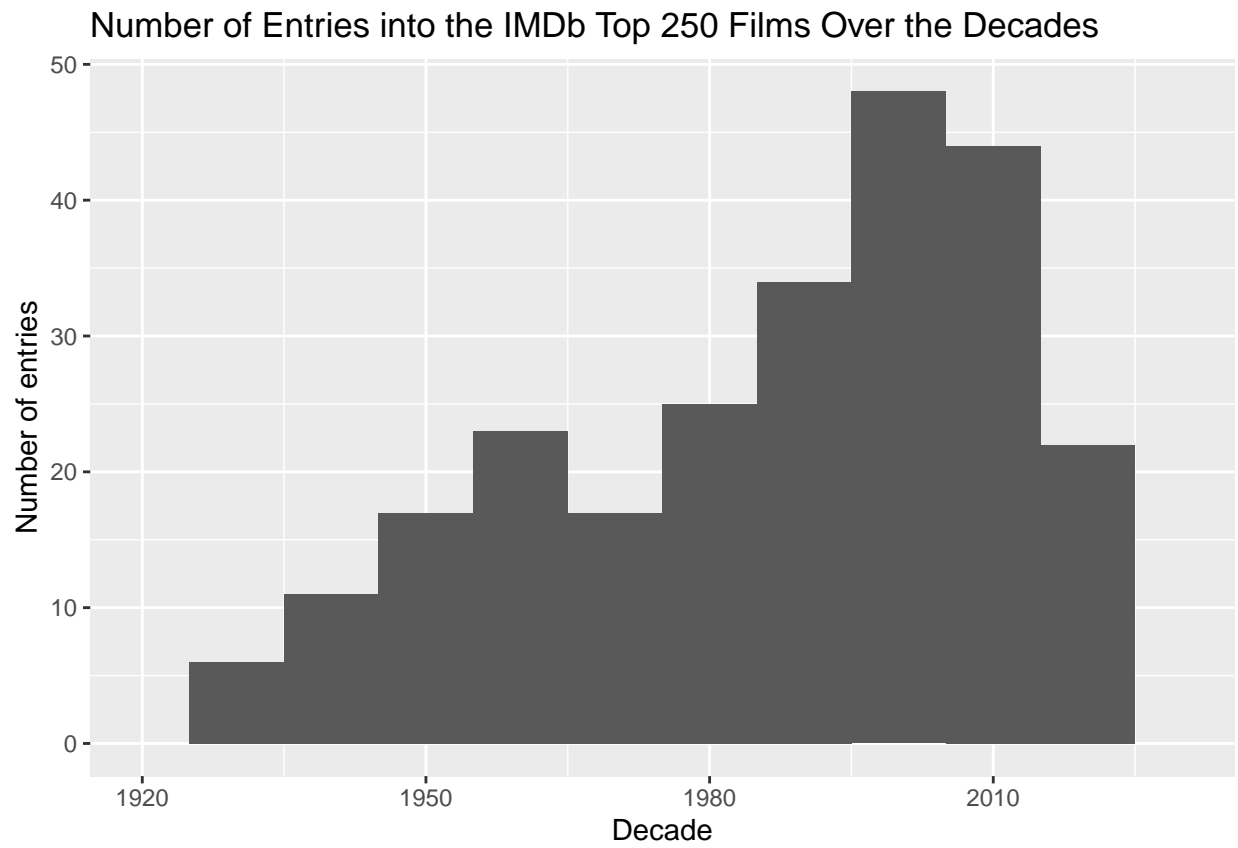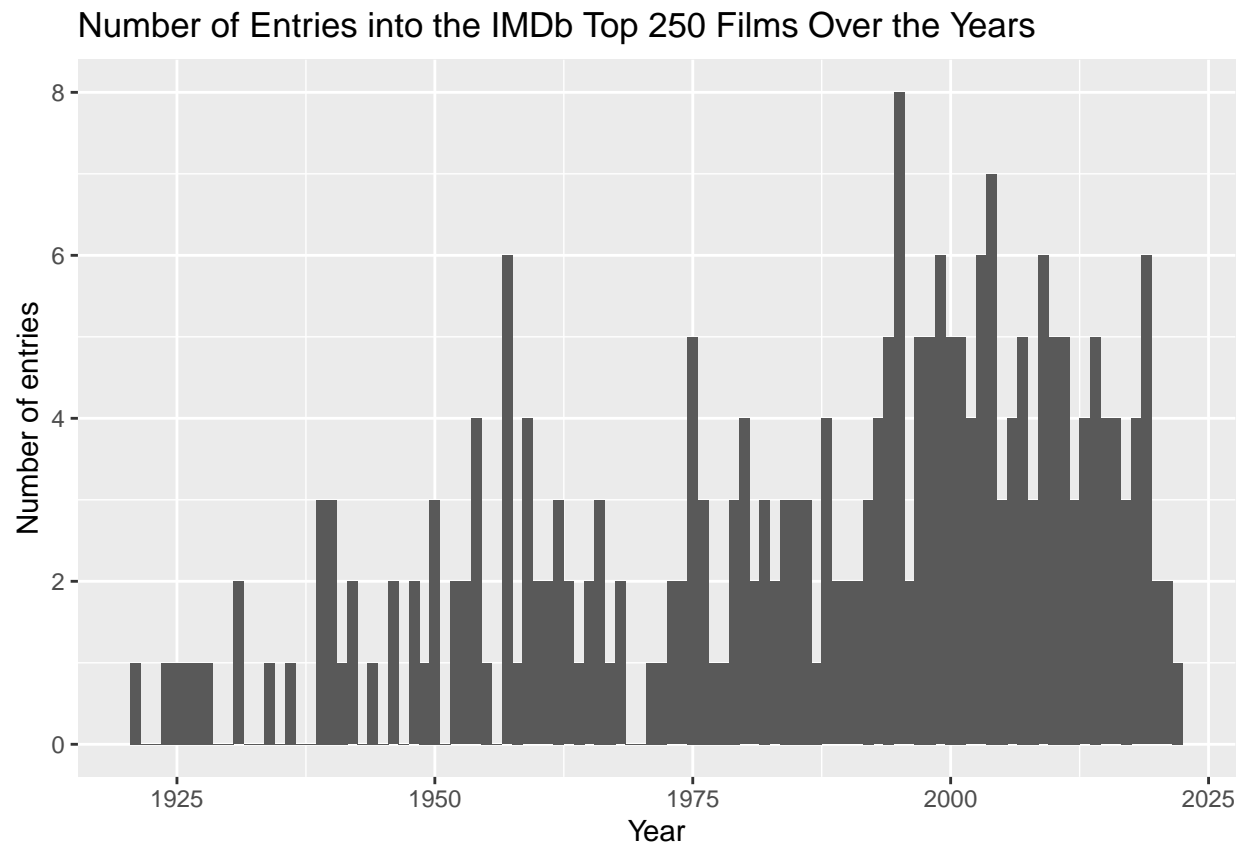
Which years produced the best films?

```
## # A tibble: 86 x 2
##     year     n
##    <dbl> <int>
## 1  1995     8
## 2  2004     7
## 3  1957     6
```
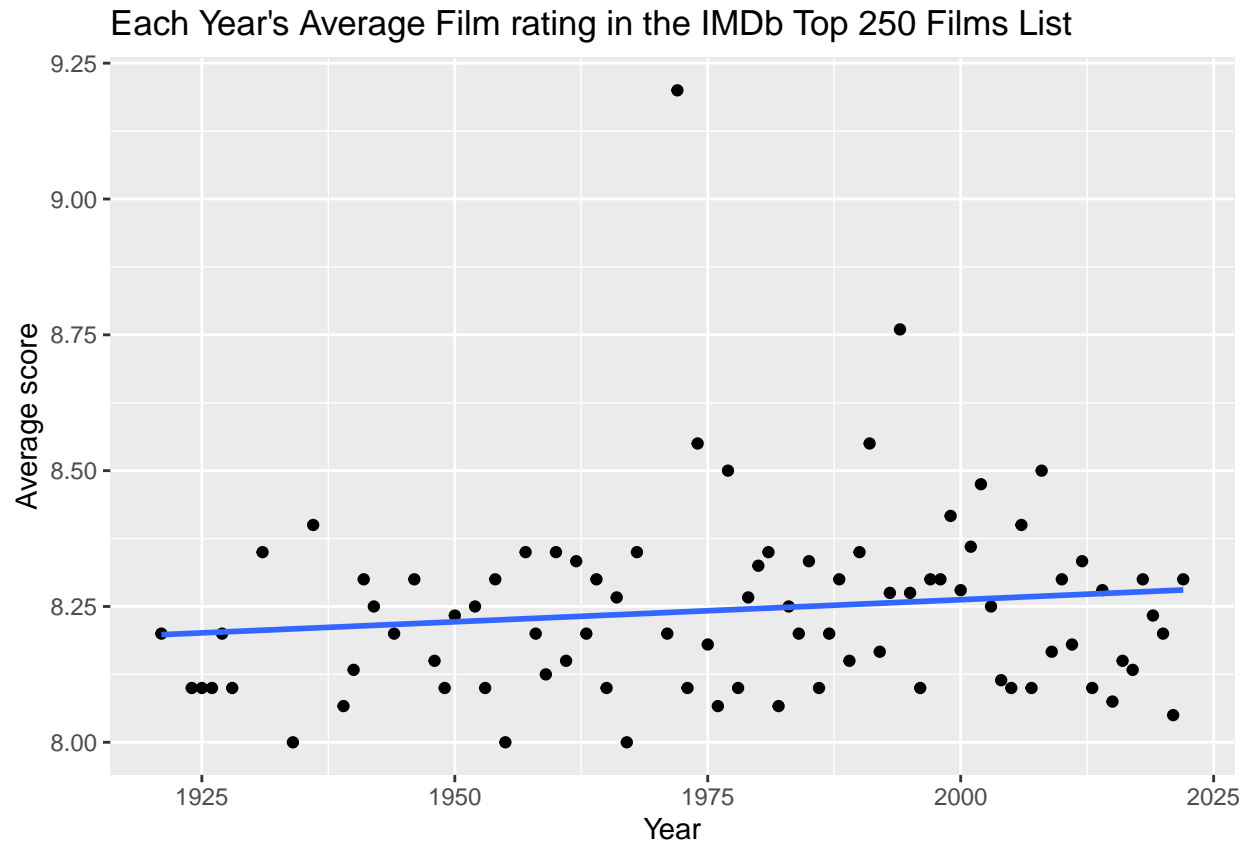
```
##  4  1999     6
##  5  2003     6
##  6  2009     6
##  7  2019     6
##  8  1975     5
##  9  1994     5
## 10  1997     5
## # ... with 76 more rows
```

```
## Warning: Removed 2 rows containing missing values (`geom_bar()`).
```

Number of Entries into the IMDb Top 250 Films Over the Decades

Number of Entries into the IMDb Top 250 Films Over the Years


```
## 'geom_smooth()' using formula = 'y ~ x'
```

Each Year's Average Film rating in the IMDb Top 250 Films List

We can also calculate the average number of films in the top 250 from each year: $250/(2022 - 1921) \approx 2.5$. As a result, we can say that any year with 3 or more films in the top 250 is a pretty good year for films. Likewise, if a year has no film or only 1 film in the top 250, it was a fairly bad year for films.

We can see that the mid-to-late 90s and early-mid 2000s were great years for films - contributing many of the top 250 films. Other great years include 1957, 2019, and 1975.

The Verdict

Taking all of the analysis into consideration, I think it's fair to conclude that there hasn't been a 'good' year for films since 2019. The last time it wasn't a 'good' year for films before this was over 20 years ago - in 1996. So yes, me and my friend seem vindicated in the view that the cinema seems to be in a bit of a slump at the minute. . .