Are Films Getting Worse?

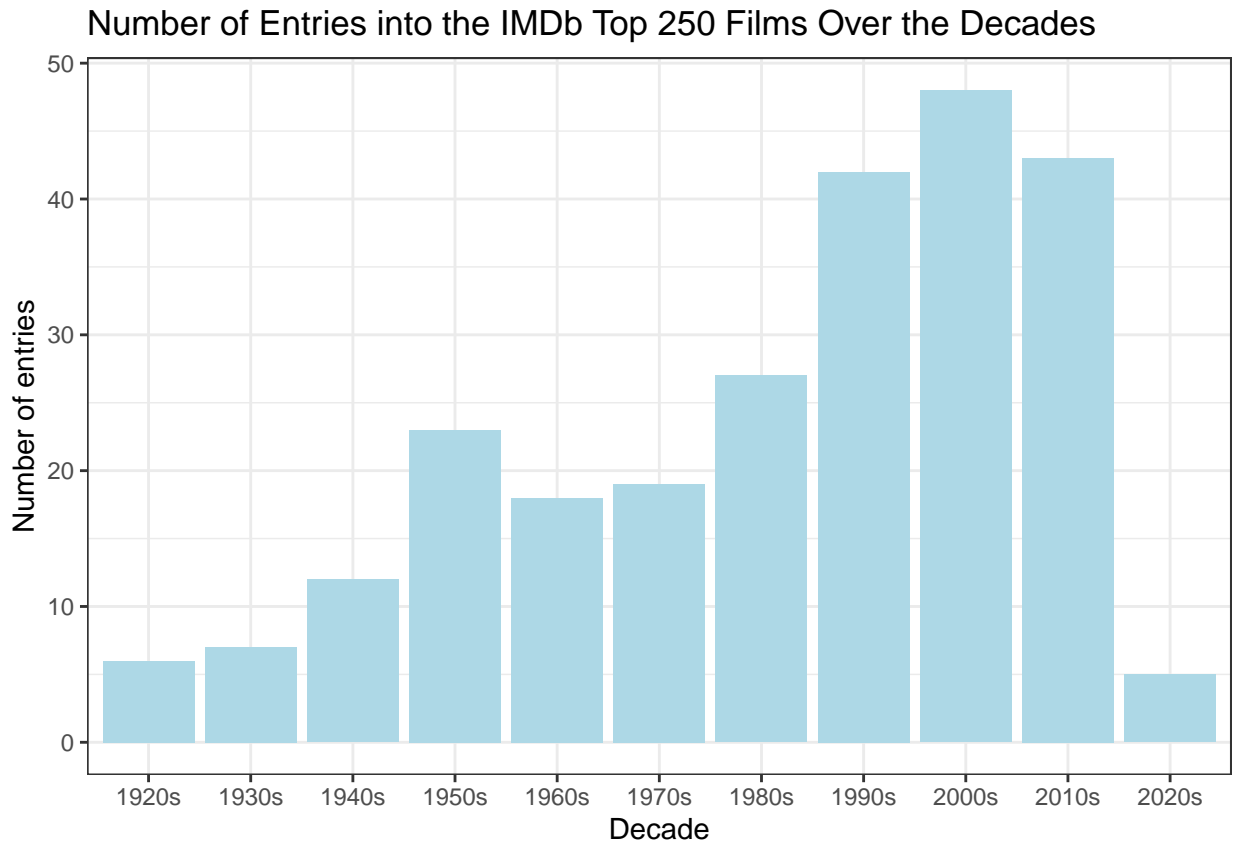Conor Cassidy

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/.
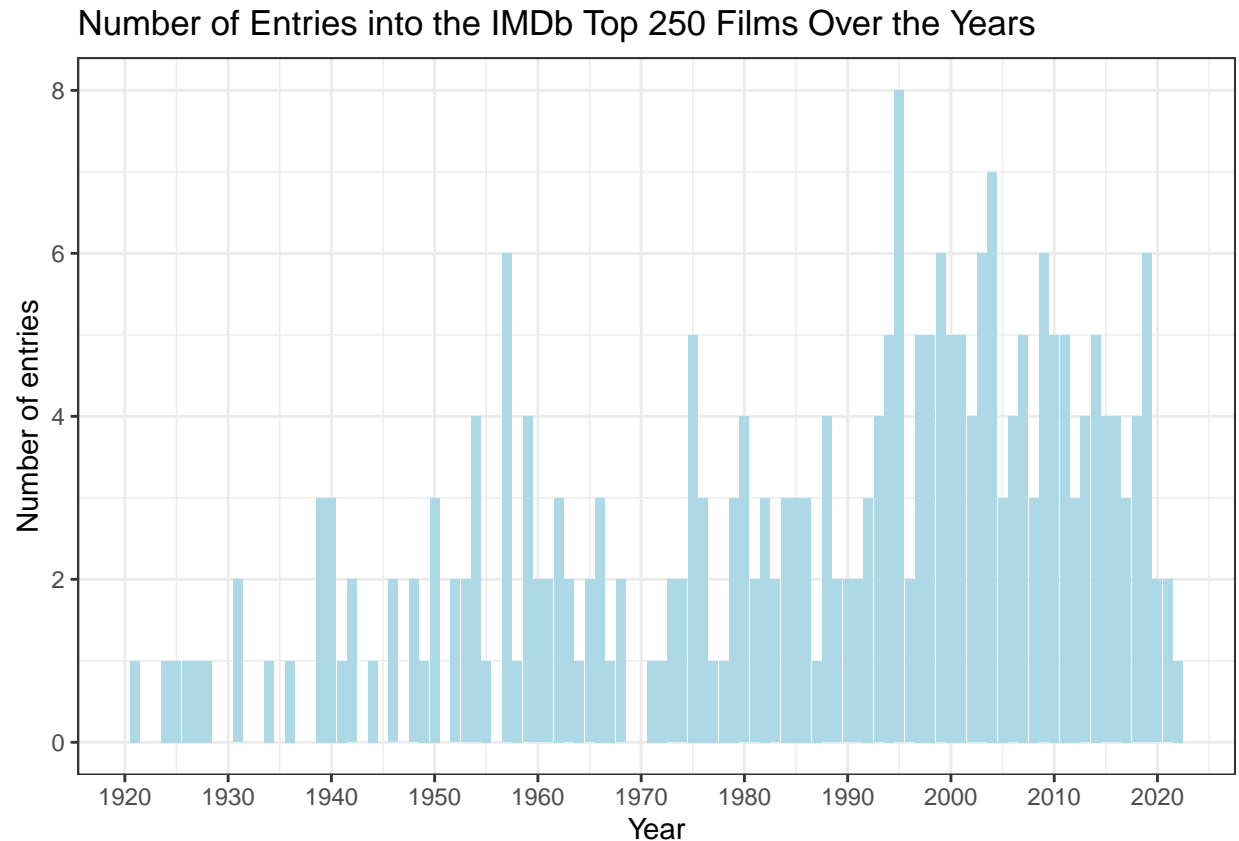
The Data

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.

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
```

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

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

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...
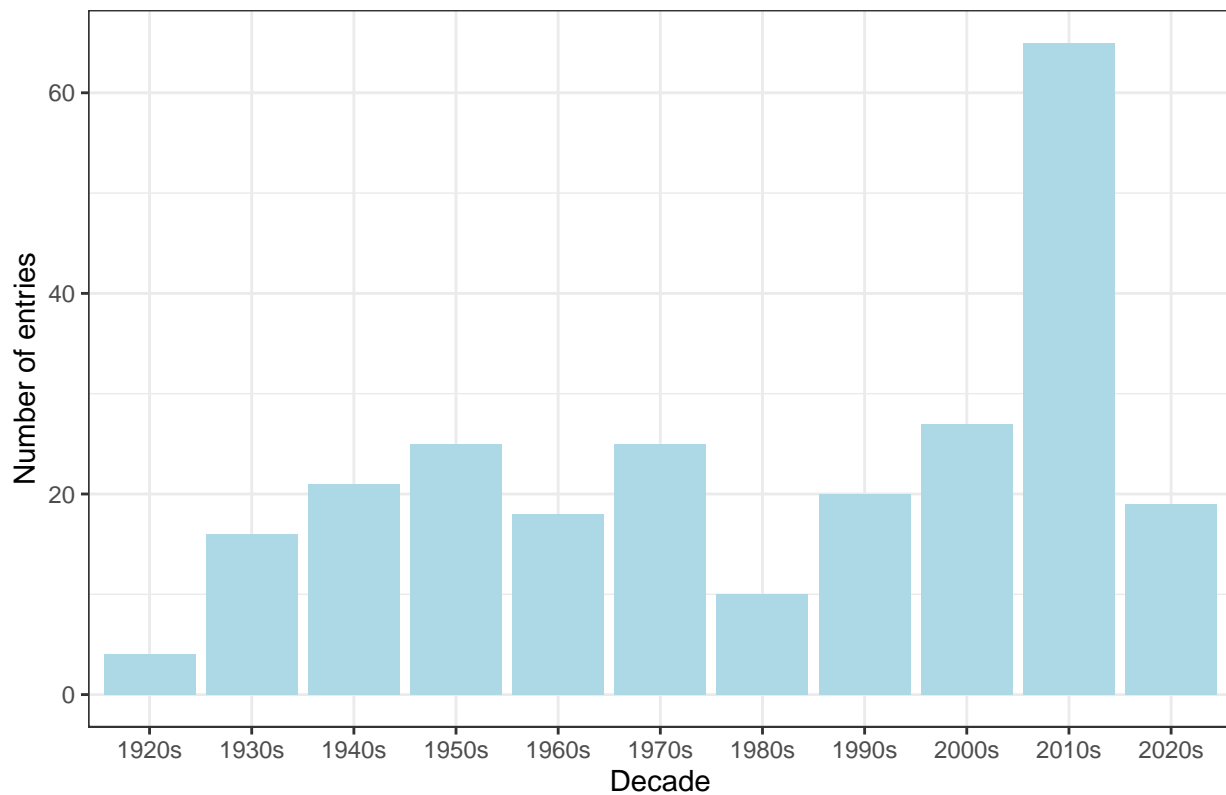
```
meta_250 %>%
  count(year, sort = TRUE) %>%
  as_tibble()
```

```
## # A tibble: 89 x 2
##    year      n
##    <chr> <int>
##  1 2018     11
##  2 2015     10
```

```
##  3 2013       8
##  4 2019       8
##  5 2021       8
##  6 2016       7
##  7 2017       7
##  8 1956       6
##  9 2020       6
## 10 1940       5
## # ... with 79 more rows
```

```r
# by decades
ggplot(meta_250, aes(x=decade, fill = decade)) +
  geom_bar(stat = "count", fill = "light blue") +
  labs(
    x = "Decade",
    y = "Number of entries",
    title = "Number of Entries into the Metacritic Top 250 Films Over the Decades") +
  theme(legend.position = "none") +
  theme_bw()
```

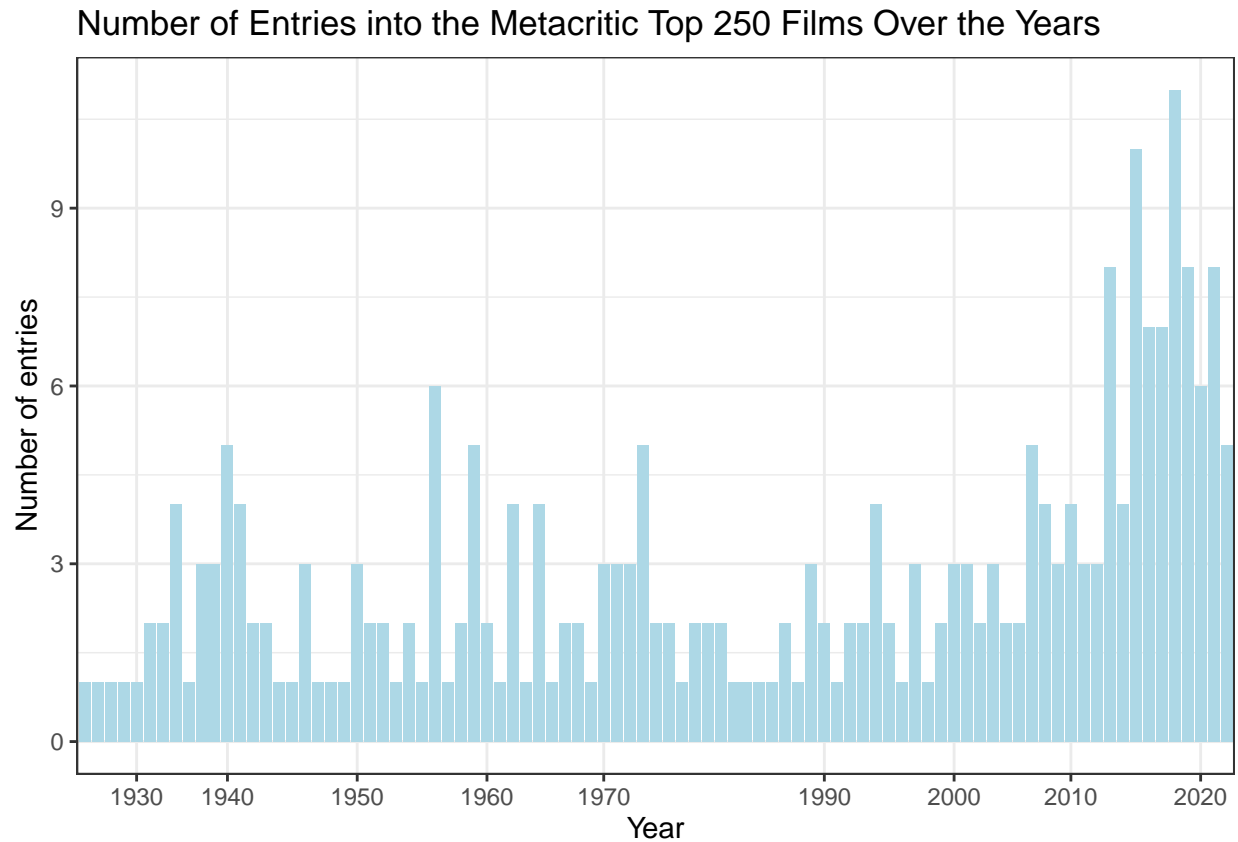## Number of Entries into the Metacritic Top 250 Films Over the Decades


```r
# by years
ggplot(meta_250, aes(x= years2, fill = years2)) +
  geom_bar(stat = "count", fill = "light blue") +
  labs(
    x = "Year",
```

```
    y = "Number of entries",
    title = "Number of Entries into the Metacritic Top 250 Films Over the Years") +
theme(legend.position = "none") +
scale_x_discrete(breaks=seq(1920,2025,by=10)) +
theme_bw()
```

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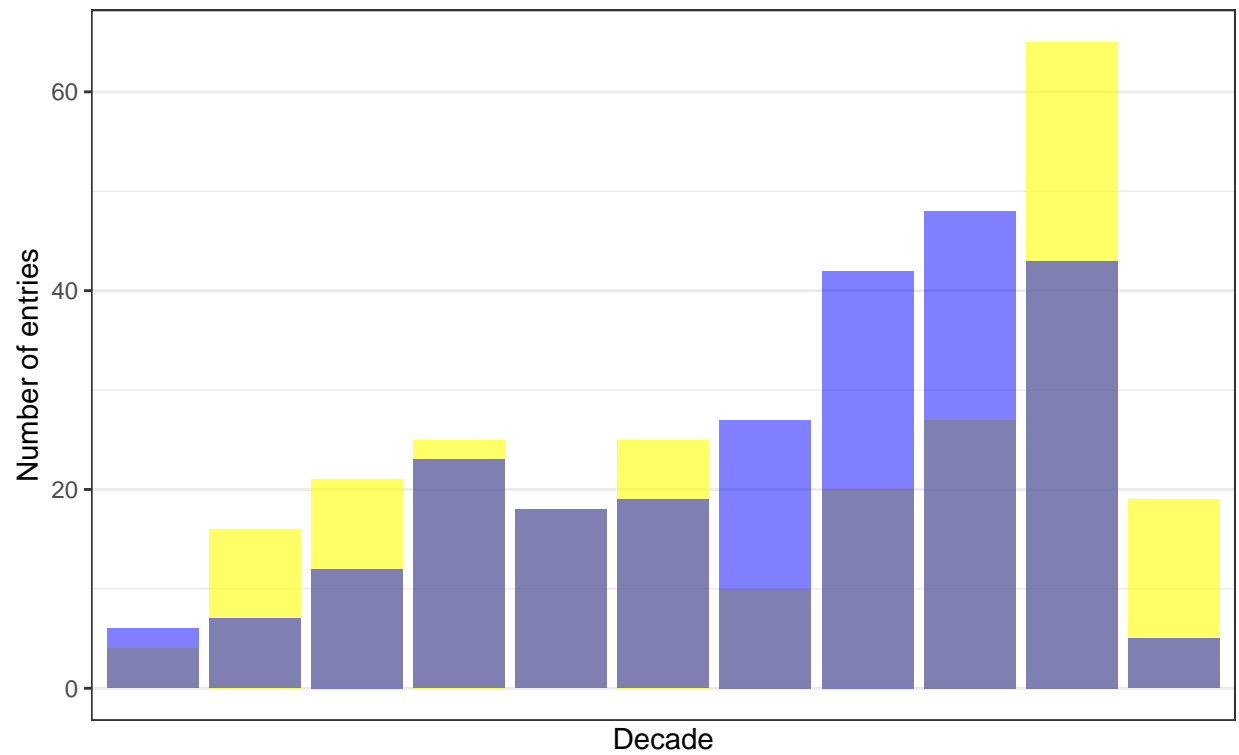


```
ggplot(NULL, aes(x = decade)) +
  geom_bar(data = meta_250, stat = "count", fill = "yellow", alpha = .6) +
  geom_bar(data = imdb_top_250, stat = "count", fill = "blue", alpha = .5) +
  labs(
    x = "Decade",
    y = "Number of entries",
    title = "Number of Entries into the Top 250 Films Over the Decades",
    subtitle = "Yellow represents the Metacritic data while blue represents IMDb") +
  scale_fill_manual(values=c("lightblue", "pink")) +
  scale_x_discrete(breaks=seq(1920,2025,by=10)) +
  theme_bw()
```

Number of Entries into the Top 250 Films Over the Decades

Yellow represents the Metacritic data while blue represents IMDb


```
## no meta on there and no legend ??
```