

Statpadders

Toma Shigaki-Than, CJ Frederickson, Camden Reeves, Sam Kakarla

2025-03-17

Introduction

Online review platforms have transformed how audiences evaluate and choose movies. Between professional critics and everyday viewers, these platforms host a broad spectrum of opinions that increasingly shape the trajectory of films, from box office performance to streaming recommendations. However, these two groups often evaluate films through very different lenses: critics tend to emphasize artistic merit, narrative structure, and technical execution, while audiences may prioritize entertainment, emotional resonance, and accessibility. As a result, it is commonplace for a film to receive mixed signals across platforms. A film may be highly rated by audiences but poorly reviewed by critics, or vice versa.

This divergence can be confusing for consumers, who must navigate these conflicting assessments to make viewing decisions. For example, a film might thrive on word-of-mouth and accumulate high audience ratings on IMDb but perform poorly with critics or fail to receive awards recognition.

Motivation and Importance

Understanding these differences is essential not just for filmgoers but also for industry stakeholders. Movie studios, marketers, and streaming platforms rely on both critical acclaim and mass appeal to determine promotional strategies, content investments, and recommendation algorithms. Prior research has shown that user-driven ratings have powerful word-of-mouth effects, often extending a film's relevance and commercial success (Moon, Bergey, Iacobucci, 2010). Further, the typical consumer does not evaluate films with the same lens or criteria as professional critics.

By analyzing these patterns, we aim to identify the film characteristics, such as decade of release, runtime, gross earnings, and censorship rating, that are most predictive of audiences liking a film more than critics. These insights have direct implications for consumer behavior research, content recommendation systems, and media marketing strategies in an increasingly data-driven entertainment landscape.

Given the observed divergence between critic and audience evaluations, we pose the following research question:

What factors contribute to audiences liking a movie more than critics, and how can we use these factors to predict the likelihood of a film performing better with audiences than critics?

In this study, we focus on modeling the odds that audience scores exceed critic scores. By doing so, we shift attention from understanding average film quality to identifying conditions that foster fan-favorite films — those that resonate with general audiences even when they fail to win over the critics.

Data

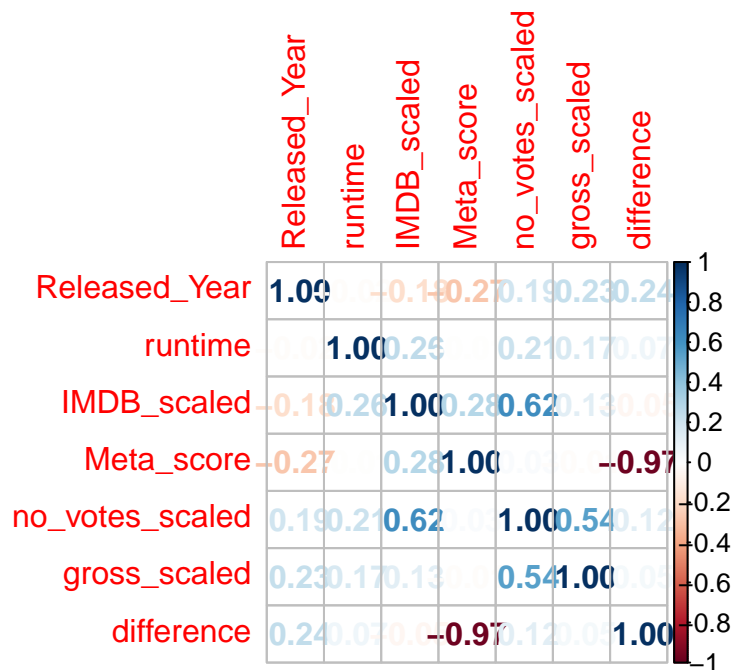
In cleaning our data, we first had to deal with the NA values in our data. Some observations had NA values in their Gross Revenues. After examining these observations, there were no discernible patterns or connections between the NA values; they were random. As such, we were able to drop these values without compromising our data set or losing important observations. We also created the variable **difference**, whose value indicates the difference between IMDB Score and MetaScore, scaled so that they can be compared. A negative value indicates that the audience score is lower than that of critics, and positive is vice versa. Because our dataset includes older and international films with various outdated or uncommon censorship ratings, we consolidated certificates into modern, widely recognizable categories: G, PG, PG-13, R, and Other, based on their closest equivalents in the current U.S. rating system.

Predictors:

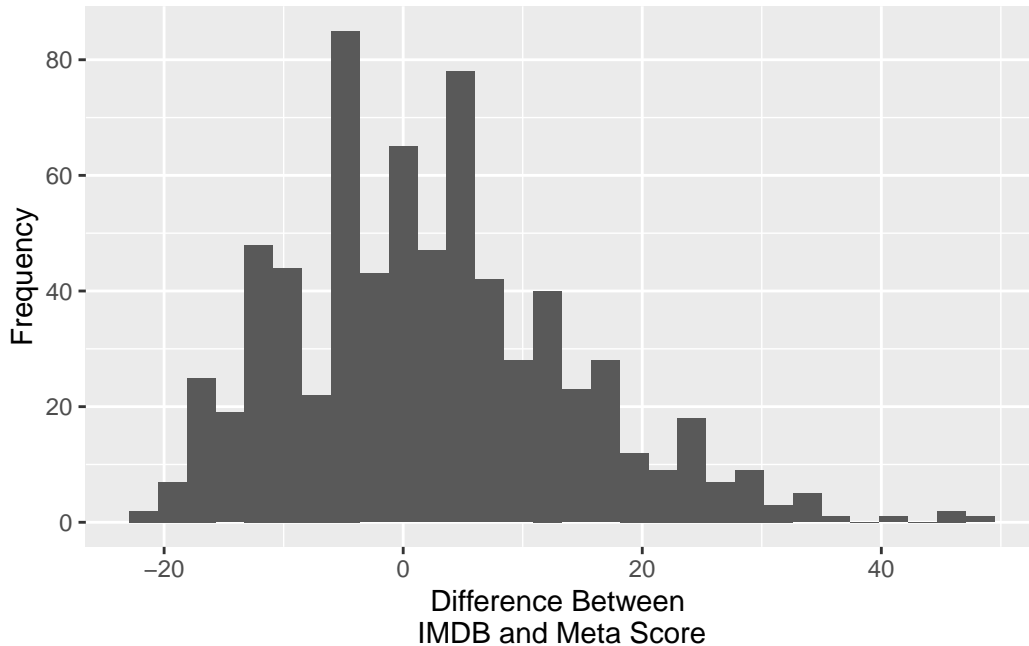
- Runtime (numerical)
- Gross Revenue (numerical)
- Censorship Certificate (categorical)
- Decade Released (categorical)
- Number of Votes (numerical)

Response Variables:

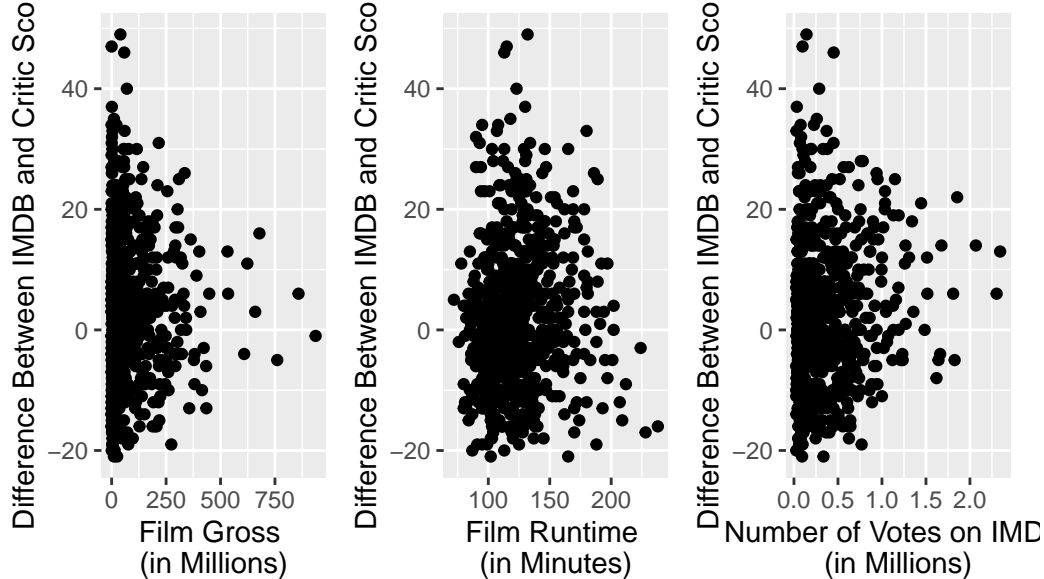
- Difference: *The quantity that the MetaScore (aggregate score of critics' ratings) differs from the IMDB score (score given by audience/fans on IMDB, scaled to match MetaScore)*



We used a correlation matrix to check for multicollinearity. As expected, difference is highly correlated with IMDB_scaled and meta_score, since it's derived from them. IMDB_scaled and no_votes_scaled also show strong correlation (0.62), so we avoid including both in the same model.



Film Gross, Runtime, and IMDB Votes, vs. Score Difference



In terms of the `difference` between the two scores, it seems that the values are almost normally distributed, with a slight left skew. This suggests that it is nearly equally common for a IMDB Score to be either higher or lower than the MetaScore score, though slightly more often lower.

Additionally, when plotting our numerical predictors, there appear to be no linear relationships between each predictor and our response.

Methodology

A tibble: 20 x 5

term	estimate	std.error	statistic	p.value
<chr>	<dbl>	<dbl>	<dbl>	<dbl>
1 (Intercept)	-3.73	1.52	-2.45	0.0141
2 gross_cent	-0.00402	0.00157	-2.56	0.0105
3 runtime_cent	0.00180	0.00331	0.545	0.586
4 decade1940s	-14.2	413.	-0.0344	0.973
5 decade1950s	0.402	1.46	0.276	0.782
6 decade1960s	2.96	1.54	1.91	0.0558
7 decade1970s	3.09	1.55	2.00	0.0455
8 decade1980s	3.85	1.54	2.51	0.0122
9 decade1990s	4.26	1.53	2.79	0.00535

10 decade2000s	3.98	1.53	2.60	0.00920
11 decade2010s	3.75	1.53	2.45	0.0141
12 certificateOther	-2.87	8.69	-0.330	0.742
13 certificatePG	-0.0853	0.243	-0.351	0.725
14 certificatePG-13	-0.00206	0.456	-0.00451	0.996
15 certificateR	0.0964	0.218	0.443	0.658
16 no_votes_scaled	0.221	0.291	0.760	0.447
17 gross_cent:certificateOther	-0.0759	0.120	-0.632	0.527
18 gross_cent:certificatePG	0.00525	0.00183	2.88	0.00401
19 gross_cent:certificatePG-13	0.00198	0.00658	0.302	0.763
20 gross_cent:certificateR	0.00691	0.00249	2.77	0.00560

We selected a logistic regression model for two primary reasons. First, none of the numerical predictors exhibited a clear linear relationship with the response variable difference. Second, our goal was to model the odds that audience scores exceed those of critics. In our initial model, most levels of the certificate variable were not statistically significant. Additionally, although we hypothesized an interaction effect between gross revenue and censorship rating, the corresponding interaction terms were not significant and thus were excluded from the next model tested.

term	df.residual	residual.deviance	df.residual	deviance	p-value
difference_binary ~ gross_cent + runtime_cent + decade + no_votes_scaled	702	928.711	NA	NA	NA
difference_binary ~ gross_cent + runtime_cent + decade + certificate + no_votes_scaled + gross_cent * certificate	694	909.063	8	19.648	0.012

A drop-in-deviance test comparing a reduced model (excluding certificate and interaction terms) to a full model (including certificate and the gross \times certificate interaction) yielded a statistically significant improvement in fit ($\chi^2 = 19.648$, $df = 8$, $p = 0.012$). This suggests that certificate and its interaction with gross earnings contribute meaningfully to predicting whether audience ratings exceed those of critics.

Results

Appendix

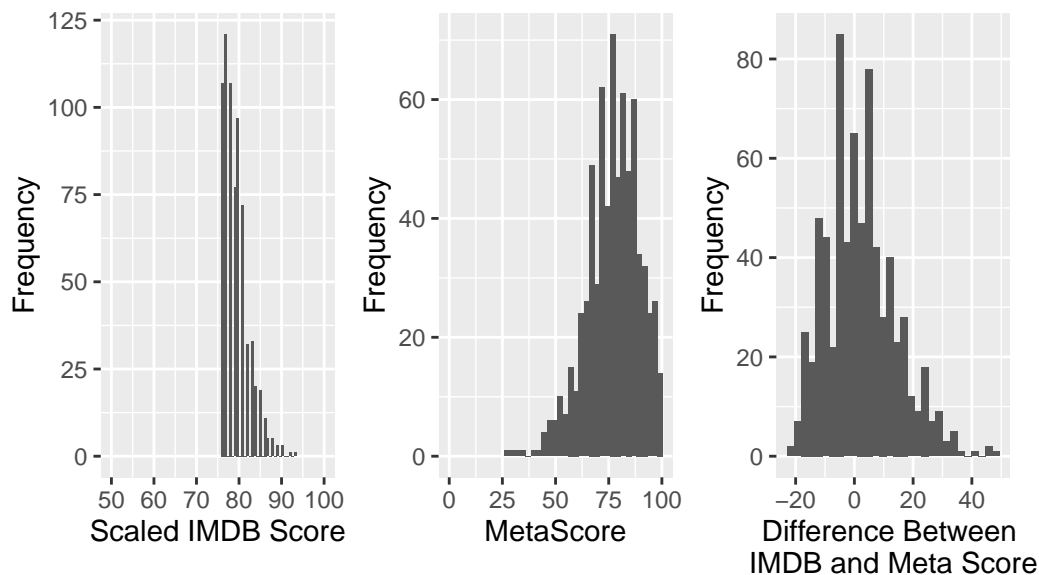
To begin our EDA, we first had to deal with the NA values in our data. Some observations had NA values in their Gross Revenues. After examining these observations, there were no discernible patterns or connections between the NA values; they were random. As such, we

were able to drop these values without compromising our data set or losing important observations. We also created the variable **difference**, whose value indicates the difference between MetaScore and IMDB Score, scaled so that they can be compared. A negative value indicates that the MetaScore is lower than IMDB Score, and a positive value indicates that it is higher. Further, we turned our year predictor into a categorical variable by creating a new variable: **decade**. Since there is a very wide range of values in **Released_Year** for the movies selected, that variable itself is not particularly useful for our analysis. Not many observations even had the same released year, and the differences between one unit in that variable were arbitrary for some movies (for example a movie released in 1966 vs 1967 does not give much insight). For data cleaning and to improve clarity and interpretability, we changed this variable into a categorical variable **decade**, where all of the years released are grouped into decades (i.e. 1950s, 1960s, etc.). This categorical approach gives better interpretability; grouping movies into decades creates a better identifier than simply using individual years.

Additionally, the variable **Runtime** listed the runtime of each observation as a string with the number of minutes followed by the word “mins.” For example, a movie 90 minutes long would be listed as the string “90 mins” instead of the number 90. As such, this made **Runtime** a categorical variable. We changed this by removing the “mins” label and refactoring it as numeric, thus making the **Runtime** into a numerical variable.

Univariate EDA

Distribution of Potential Response Variables



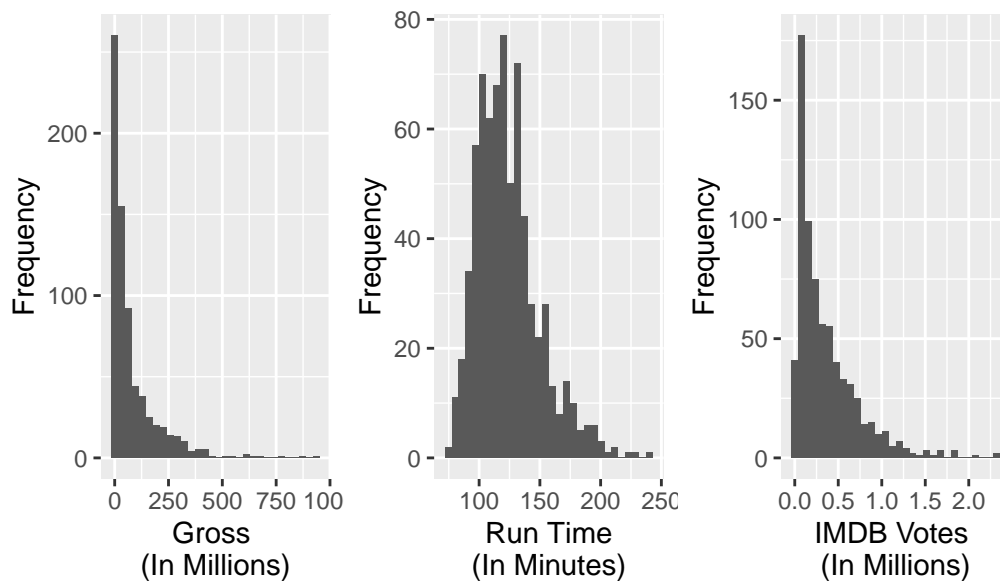
```
# A tibble: 3 x 7
  Variable    mean    med    sd   IQR   min   max
  <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 IMDB_Score 79.4    79  2.93    4    76    93
2 MetaScore  77.2    78 12.4   16    28   100
3 Difference  2.21     1 11.9   15   -21    49
```

As we can see from our response variables, scaled IMDB Score seems to be skewed right for these films, and scores tend to trend between 76 and 93. Both the mean score and median score are about 79, standard deviation of about 3, IQR of 4, and a range of 17.

The distribution for MetaScore seems to be skewed right, with a mean of about 77, a median about 78, a standard deviation of about 12, an IQR of 16 and a range of 72.

Furthermore in terms of the **difference** between the two scores, it seems that the values are almost normally distributed, with a slight left skew. This suggests that it is nearly equally common for a MetaScore to be either higher or lower than the IMDB score, though slightly more often lower. There is an outlier when MetaScore is about 49 points lower than IMDB score (-49). The mean **difference** is when about MetaScore is about two points lower than IMDB score (-2) and median at 1 point lower (-1). There is standard deviation of 12 points, IQR of about 15 points, and a range of 70 points.

Distribution of Key Numerical Predictors



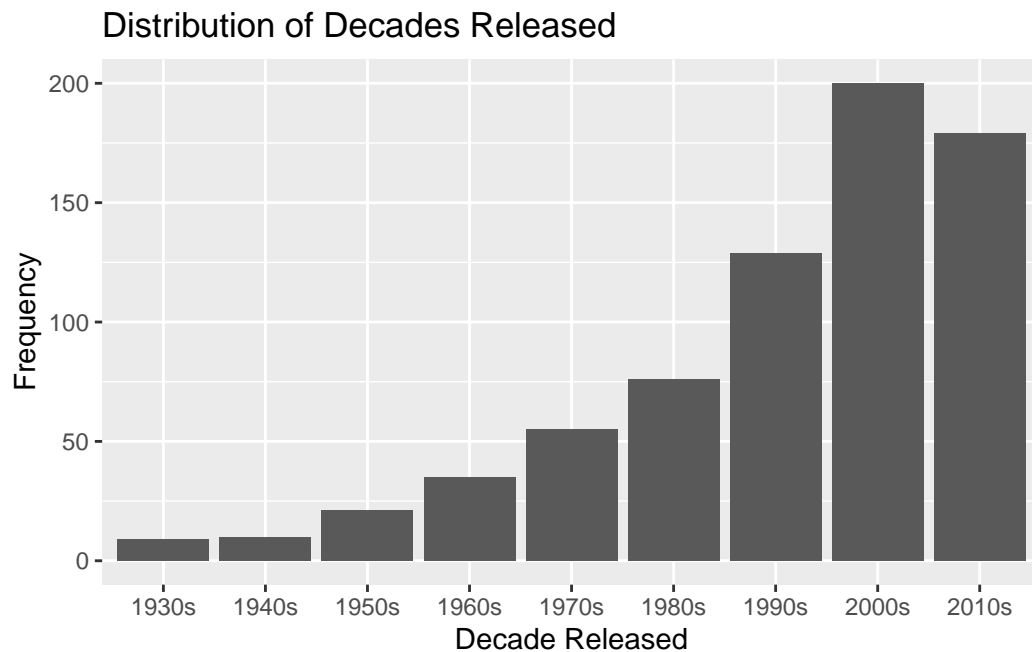
```
# A tibble: 3 x 7
  Variable    mean    med    sd   IQR   min   max
```

	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	gross	78.5	34.9	115.	96.3	0.00130	937.
2	runtime	124.	120	25.9	31.8	72	238
3	votes	0.356	0.237	0.354	0.412	0.0252	2.34

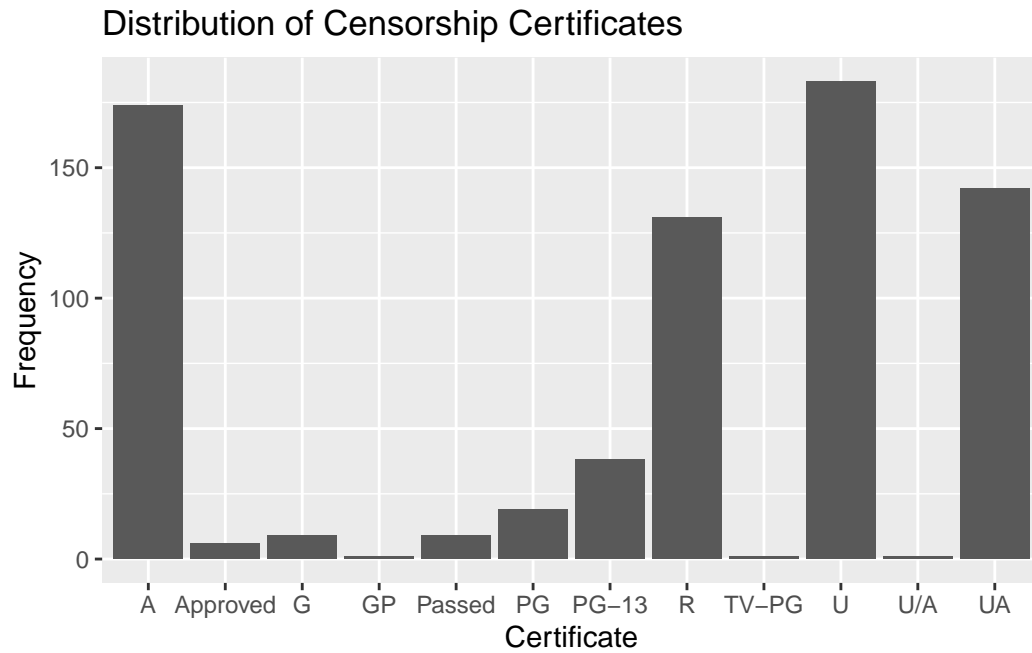
Exploring our numerical predictors, it seems that the distribution of gross revenue in millions of dollars seems to have a right skew, with most values being below \$500 million. It has a mean of \$78.514 million, median of \$34.850 million, standard deviation of about \$115 million, IQR of \$96.310 million, and a range of about \$937 million.

Run time seems fairly normal with a slight right skew. There is a potential outlier around 238 minutes. It has a mean of about 124 minutes, median of about 120 minutes, standard deviation of about 26 minutes, IQR of about 32 minutes, and a range of about 166 minutes.

Finally, number of votes has a right skew. With a potential outlier at about 2.34 million votes, it has a mean of about 356,000 votes, median of about 267,000 votes, standard deviation of about 354,000 votes, IQR of about 412,000 votes, and a range of about 2.32 million votes.



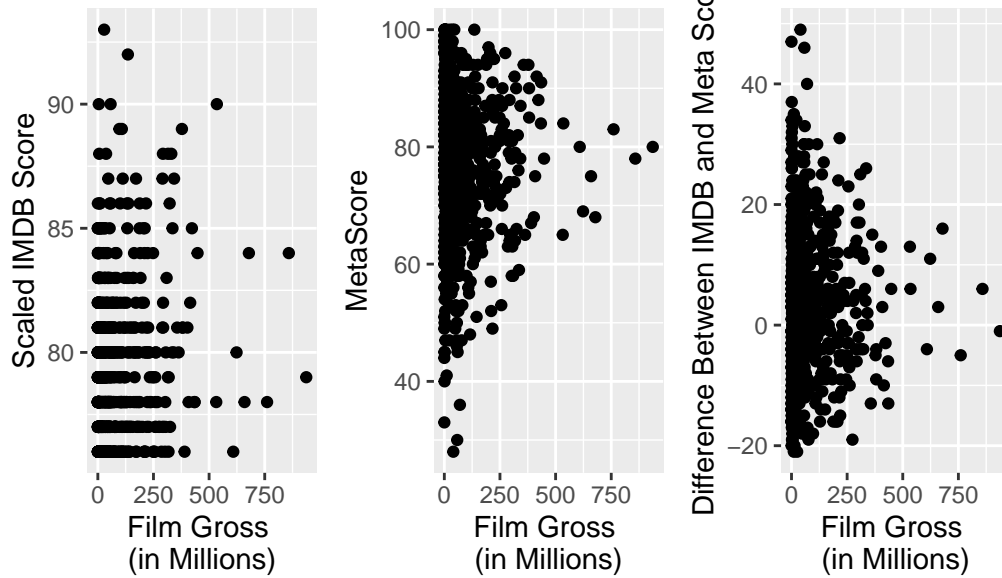
The distribution of **decade** seems to be skewed left. This is expected, as newer films are more likely to have been added to the internet in real time after release whereas older films are added retroactively.



The distribution of `certificates` does not exhibit much of a normal shape, but notably the highest distribution is of “U” movies - those with unrestricted audiences.

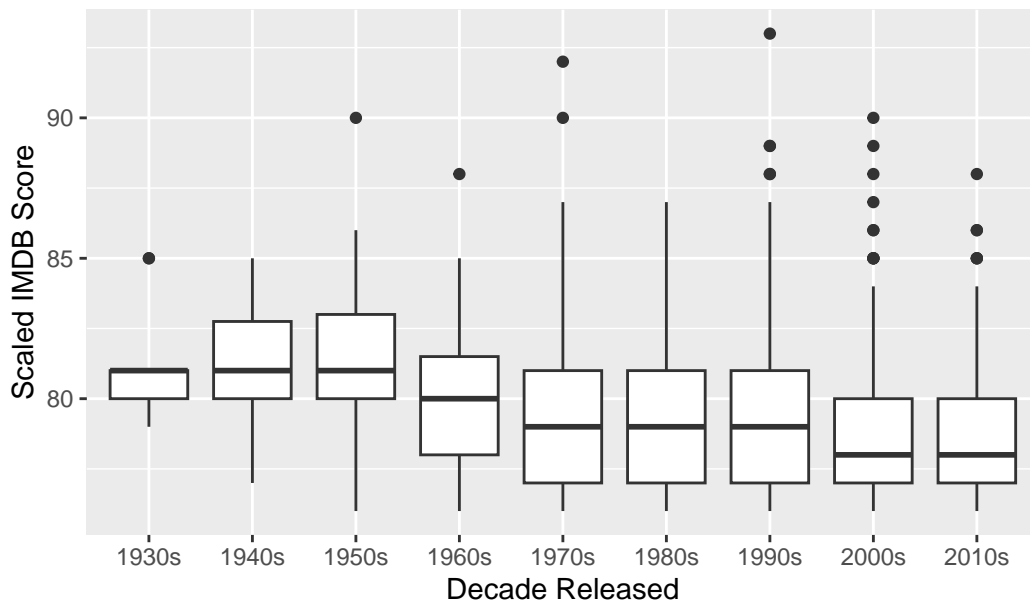
Bivariate EDA

Film Gross vs. IMDB Score, MetaScore, and Score Difference



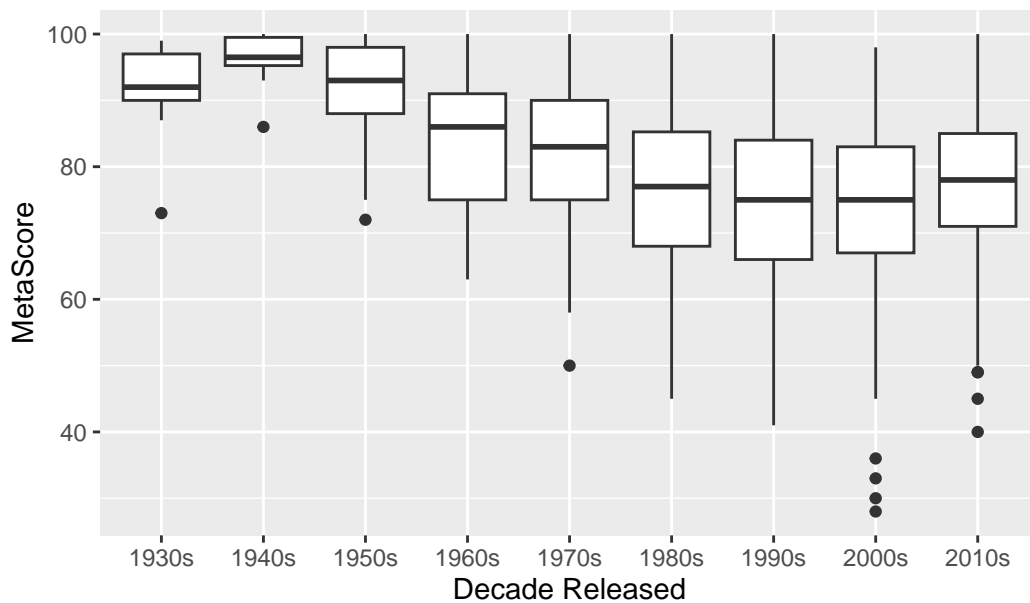
Upon initial Bivariate EDA, a clear linear relationship does not seem to appear between film's gross in millions and our three potential predictor variables. Perhaps later, to fit a model, we will need to find a variable transformation that gives us a promising model.

Scaled IMDB Score vs. Decade Released

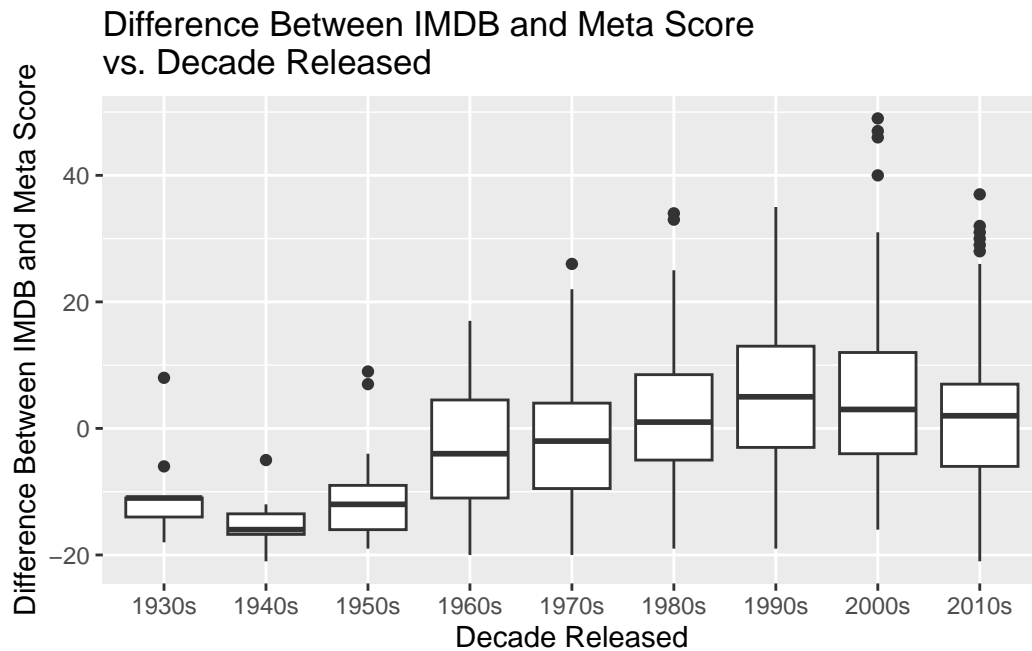


Judging from this initial bivariate EDA of decade released vs the scaled IMDB score, there seems to be a negative correlation between date and IMDB score; as movies are newer (coming out in more recent decades), the median scaled IMDB score tends to be lower.

MetaScore vs. Decade Released

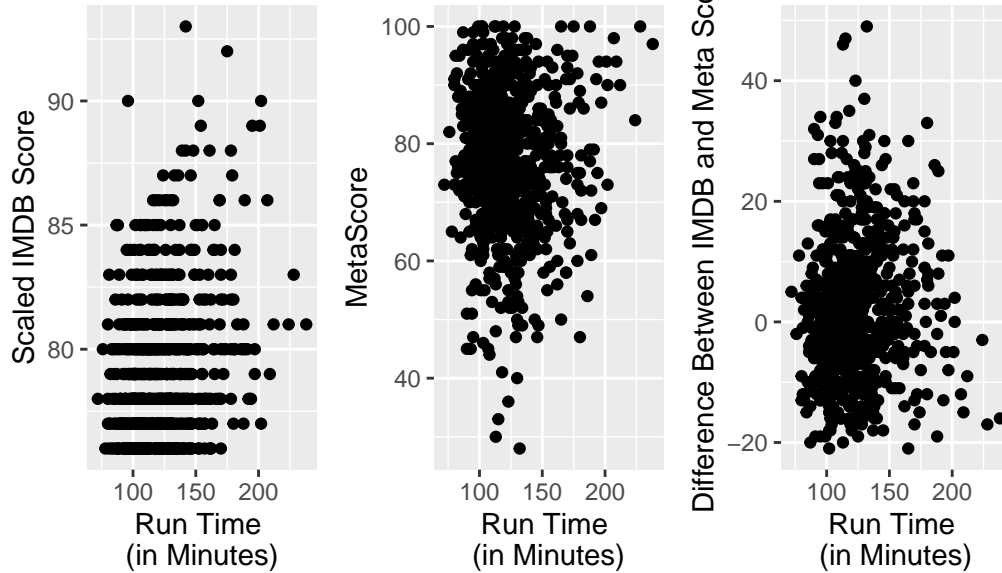


Similarly to IMDB score, the critics' median MetaScores also seem to be lower as movies are newer. In other words, the overall aggregated critic scores for films tend to be lower for movies in more recent decades.



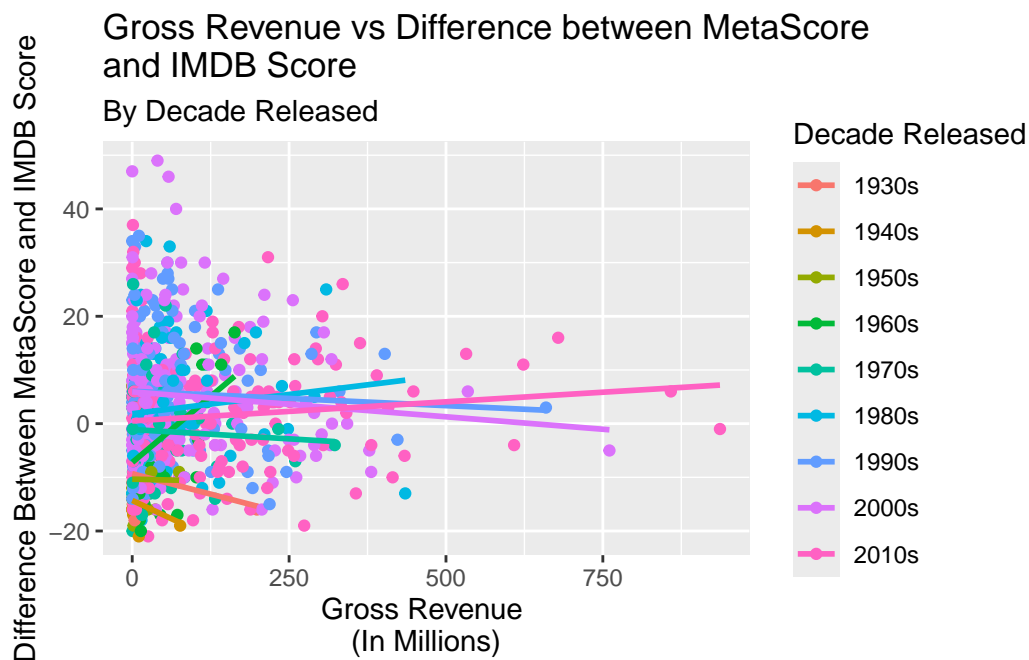
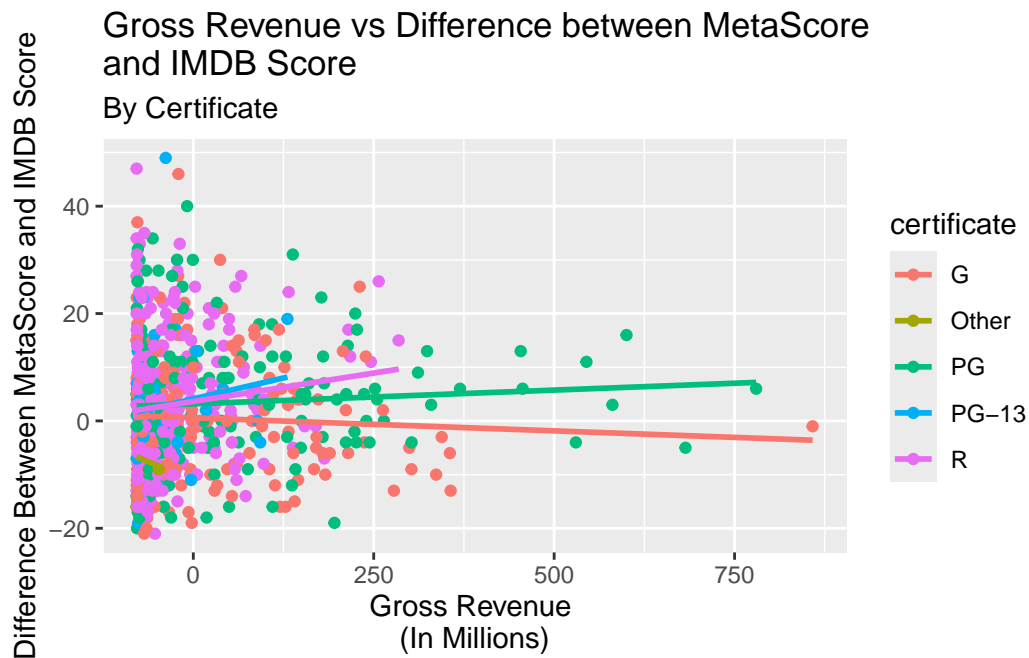
Judging from this EDA, the median difference between `Meta_score` and `IMDB_scaled` also tends to be lower as movies are newer. In other words, MetaScores tend to be lower than IMDB scores in more recent decades.

Run Time vs. IMDB Score, MetaScore, and Score Difference



Similar to our gross predictor, a clear linear relationship does not seem to appear between film's run time in minutes and our three potential predictor variables. Perhaps later, to fit a model, we will need to find a variable transformation here as well that gives us a promising model.

Interaction Effects



Here, we plotted gross revenue against the difference in scores (MetaScore minus scaled IMDB

score) to examine how these variables interact against our categorical variables. The visualizations reveal clear interaction effects in two key relationships:

Gross Revenue and Certificate

Gross Revenue and Decade Released

! Important

Before you submit, make sure your code chunks are turned off with `echo: false` and there are no warnings or messages with `warning: false` and `message: false` in the YAML.

Reference: Moon, S., Bergey, P. K., & Iacobucci, D. (2010). Dynamic Effects among Movie Ratings, Movie Revenues, and Viewer Satisfaction. *Journal of Marketing*, 74(1), 108-121. <https://doi.org/10.1509/jmkg.74.1.108>