

Statpadders

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

2025-03-17

Introduction

The proliferation of online review platforms has significantly influenced consumer decision-making across various industries, but particularly in entertainment. In the film industry, consumers rely on both professional critics and amateur audiences to gauge the quality of movies before making viewing decisions (Moon, Bergey, & Iacobucci, 2010). While professional critics evaluate films based on artistic and technical merit, amateur audiences may assess them based on personal enjoyment, accessibility, and entertainment value. This dual review system has created a dynamic where movies may receive different evaluations from critics and general audiences, raising questions about the relationship between the two and the factors influencing their divergence. Given this context, our research seeks to address the following question: What factors in a film influence IMDb user ratings and critic MetaScores; how do differences in these scores relate to movie characteristics such as gross earnings, number of votes, decade released, runtime, and certificate?

Motivation and Importance

Understanding the discrepancies between audience and critic ratings is critical for multiple stakeholders in the film industry. For movie studios and marketers, aligning promotional strategies with audience preferences while maintaining critical appeal can be a determinant of box office success and long-term profitability (Moon, Bergey, & Iacobucci, 2010). Platforms like Netflix and IMDb utilize rating-based recommendation systems to enhance user satisfaction and engagement, making it essential to refine these systems based on nuanced insights into rating behaviors. Moreover, prior research suggests that word-of-mouth (WOM) effects can sustain movie revenues over time, with online user ratings playing a crucial role in influencing later viewership trends. Understanding these dynamics is essential for optimizing film production, marketing investments, and recommendation algorithms in an era where consumer feedback is widely accessible and highly influential.

Hypotheses and Theoretical Considerations

We propose several key hypotheses regarding the divergence between IMDb ratings (representing audience scores) and critic MetaScores:

The Effect of Gross Earnings and Number of Votes

Are movies with higher gross earnings and a greater number of votes more likely to have higher IMDb ratings due to their broad audience reach, (while critics may evaluate them more critically, leading to a potential divergence in scores)? Conversely, could lower-grossing films receive higher MetaScores due to stronger artistic value but may not resonate as widely with general audiences, leading to lower IMDb ratings?

Impact of Decade Released and Runtime

Could older movies have higher MetaScores due to their established reputation and critical re-evaluation over time, while IMDb ratings could fluctuate based on contemporary audience preferences? What about runtime? Could longer runtime movies receive higher MetaScores as they are often associated with more complex storytelling, while audiences may rate them lower due to attention-span and pacing concerns?

The Role of Certificate (Censorship Rating)

R-rated movies may receive higher IMDb ratings due to mature content attracting a dedicated audience, whereas critics may assess them more rigorously depending on the execution of themes. Family-friendly movies may receive higher MetaScores due to broader accessibility but could have lower IMDb ratings if audiences find them less engaging compared to other categories. By investigating these hypotheses using this dataset that includes IMDb ratings, critic MetaScores, and various film characteristics, this study aims to provide a comprehensive analysis of the factors influencing rating discrepancies. Ultimately, our findings will offer valuable insights into how different audience segments perceive film quality, with implications for movie marketing, recommendation algorithms, and consumer behavior in the digital entertainment landscape.

In summary the following is our primary research question: how do different variable predictors in genre, censorship, runtime, and gross earnings predict how critics and fans will rate movies comparatively?

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>

Potential Predictors:

- Runtime (numerical)
- Gross Revenue (numerical)
- Certificate (categorical)
- Decade Released (categorical)

- Number of Votes (numerical)

Response Variables:

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

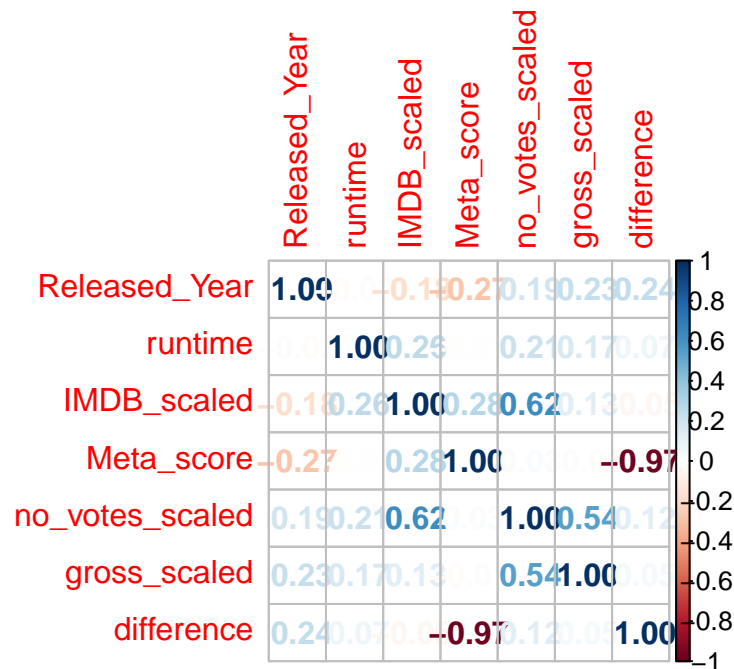
Exploratory Data Analysis

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.

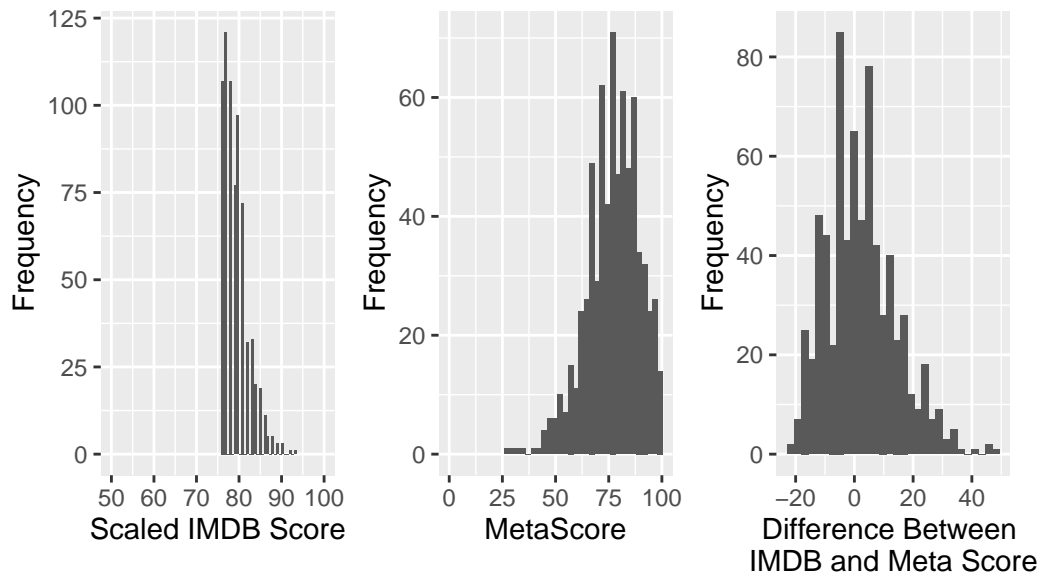
Potential Multicollinearity



Here, we created a correlation matrix to see which of our numerical predictors may be highly correlated, thus indicating potential multicollinearity. There is high correlation (0.97 correlation coefficient) between the variable `difference` and the variables `IMDB_scaled` and `meta_score`, but that is to be expected - the `difference` variable is mutated from both of those variables, to show the difference between them. `IMDB_scaled` and `no_votes_scaled` are also highly correlated with a coefficient of 0.62, but that is also to be expected - of course the votes on IMDB are correlated with an IMDB score. We would not fit a model with both of those variables.

Univariate EDA

Distribution of Potential Response Variables



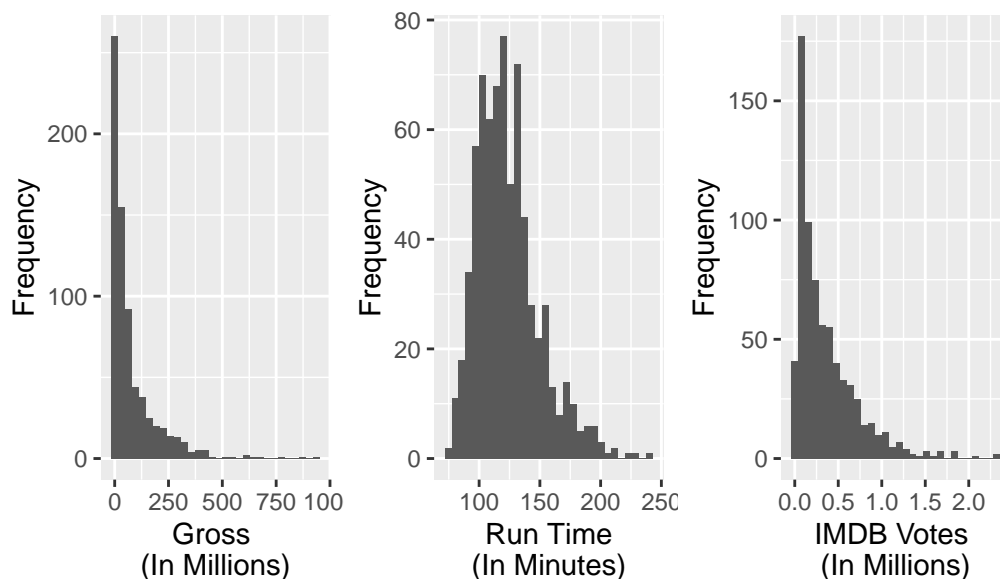
```
# A tibble: 3 x 7
  Variable    mean    med    sd   IQR   min   max
  <chr>      <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 IMDB_Score 79.4    79  2.93    4    76    93
2 MetaScore  77.2    78 12.4   16    28   100
3 Difference  2.21     1 11.9   15   -21    49
```

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

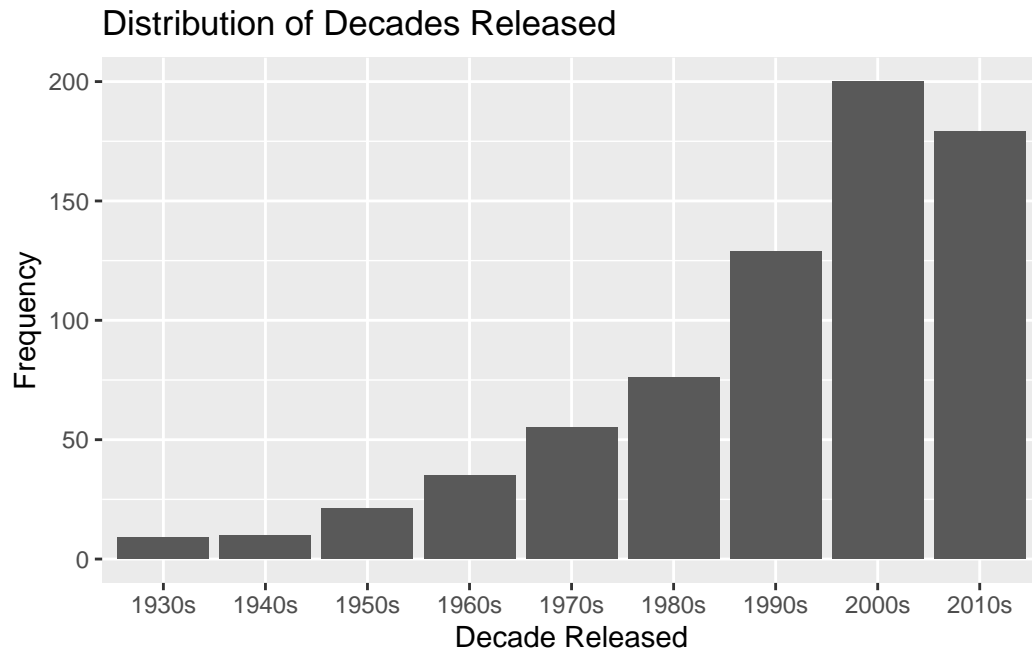


```
# A tibble: 3 x 7
  Variable    mean    med    sd   IQR    min    max
  <chr>      <dbl>  <dbl> <dbl> <dbl>  <dbl> <dbl>
1 gross      78.5   34.9  115.  96.3   0.00130 937.
2 runtime   124.   120   25.9  31.8    72     238
3 votes     0.356  0.237  0.354  0.412  0.0252  2.34
```

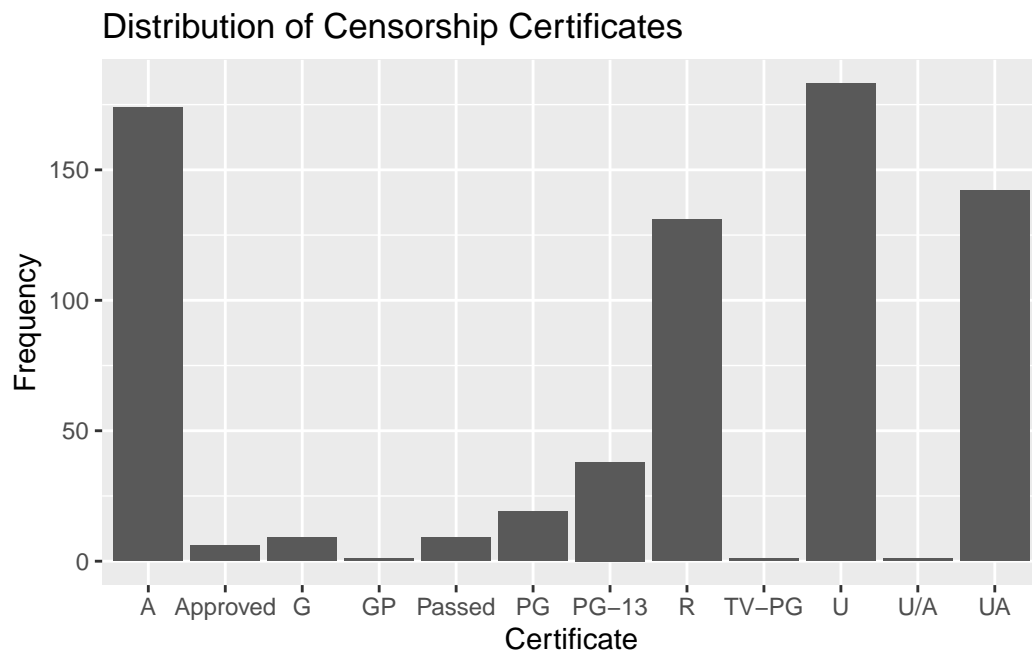
Exploring our numerical predictors, it seems that the distribution of gross revenue in millions of dollars seems to have a right skew, with most values being below \$500 million. It has a mean of \$78.514 million, median of \$34.850 million, standard deviation of about \$115 million, IQR of \$96.310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



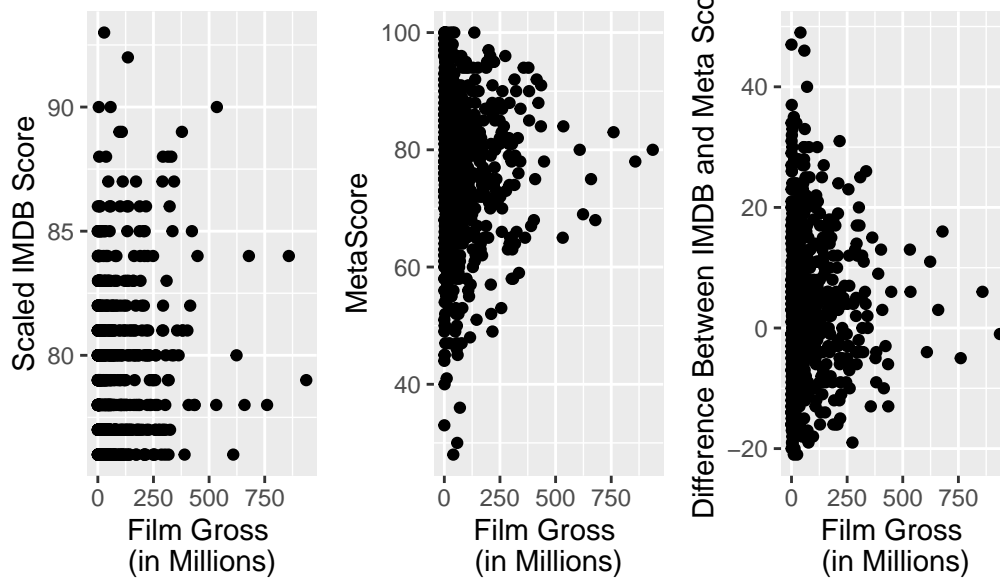
The distribution of **decade** seems to be skewed left. This is expected, as newer films are more likely to have been added to the internet in real time after release whereas older films are added retroactively.



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.

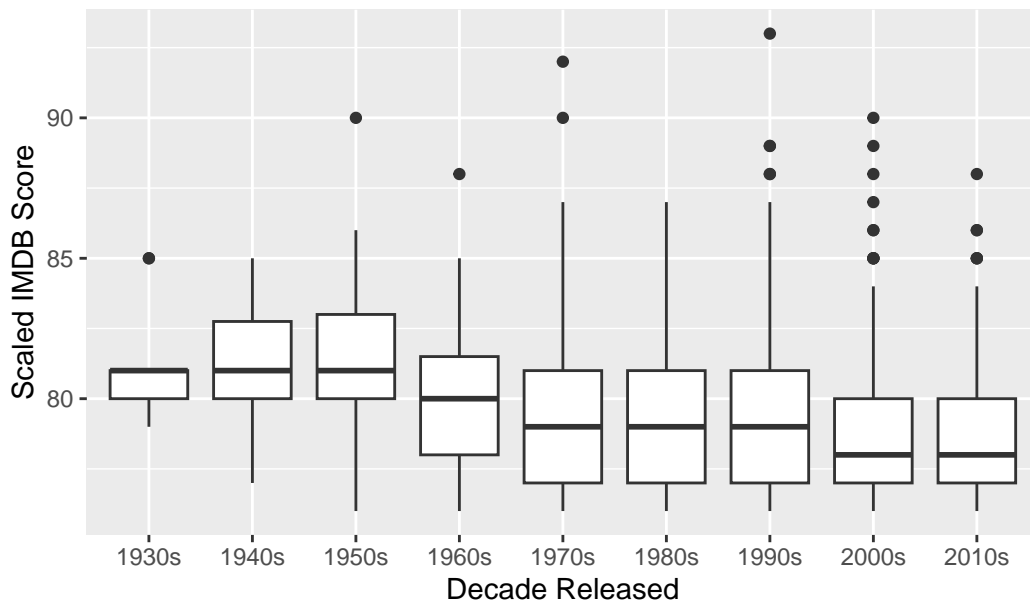
Bivariate EDA

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



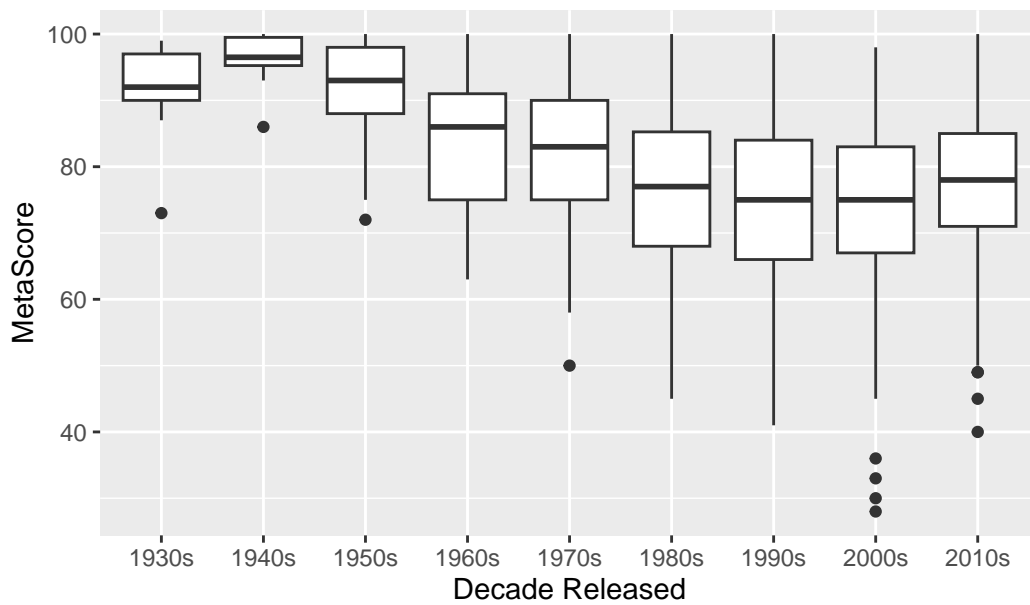
Upon initial Bivariate EDA, a clear linear relationship does not seem to appear between film's gross in millions and our three potential predictor variables. Perhaps later, to fit a model, we will need to find a variable transformation that gives us a promising model.

Scaled IMDB Score vs. Decade Released

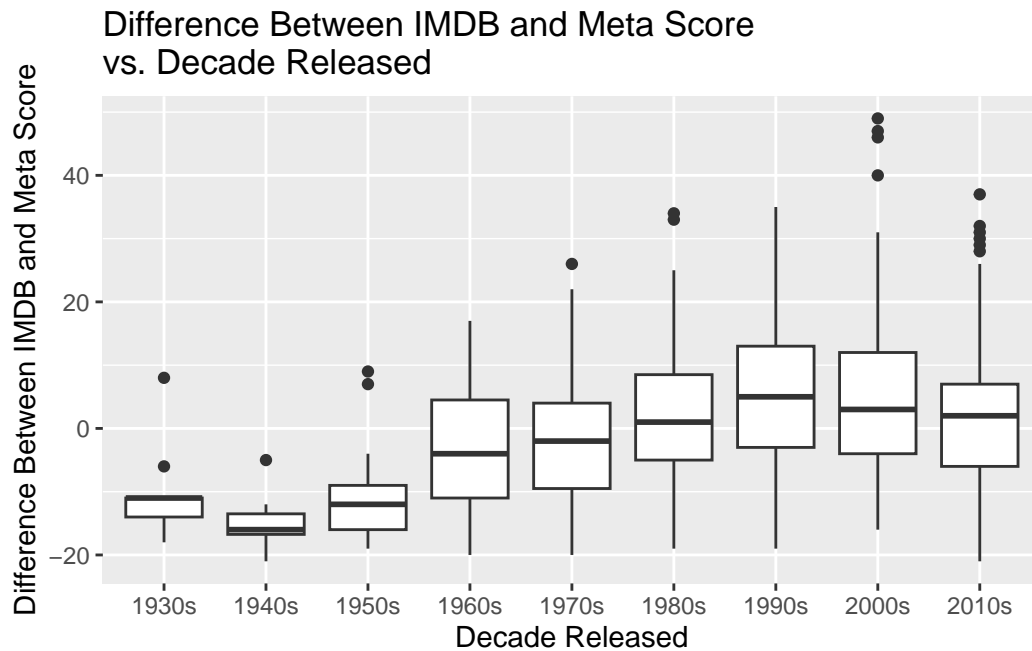


Judging from this initial bivariate EDA of decade released vs the scaled IMDB score, there seems to be a negative correlation between date and IMDB score; as movies are newer (coming out in more recent decades), the median scaled IMDB score tends to be lower.

MetaScore vs. Decade Released

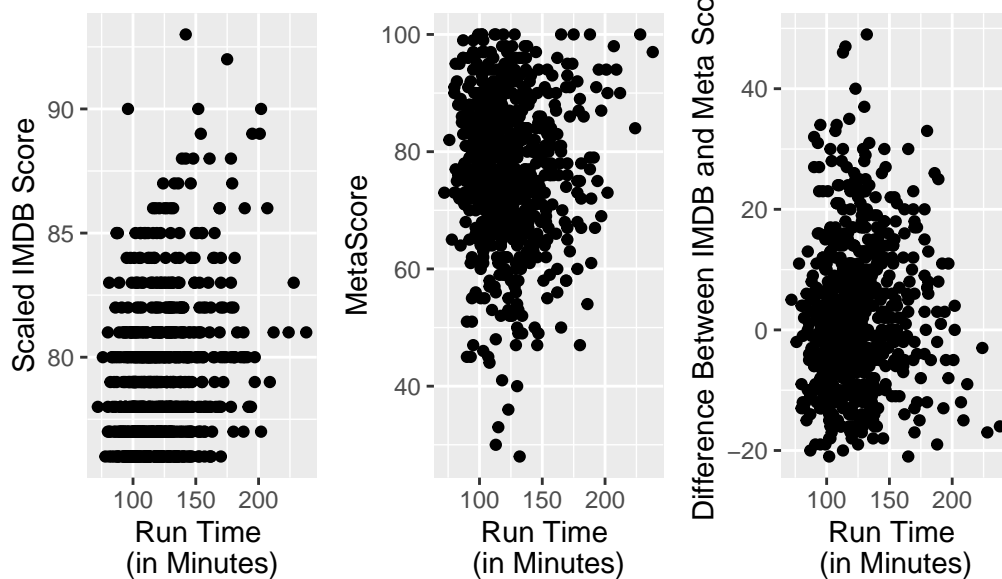


Similarly to IMDB score, the critics' median MetaScores also seem to be lower as movies are newer. In other words, the overall aggregated critic scores for films tend to be lower for movies in more recent decades.



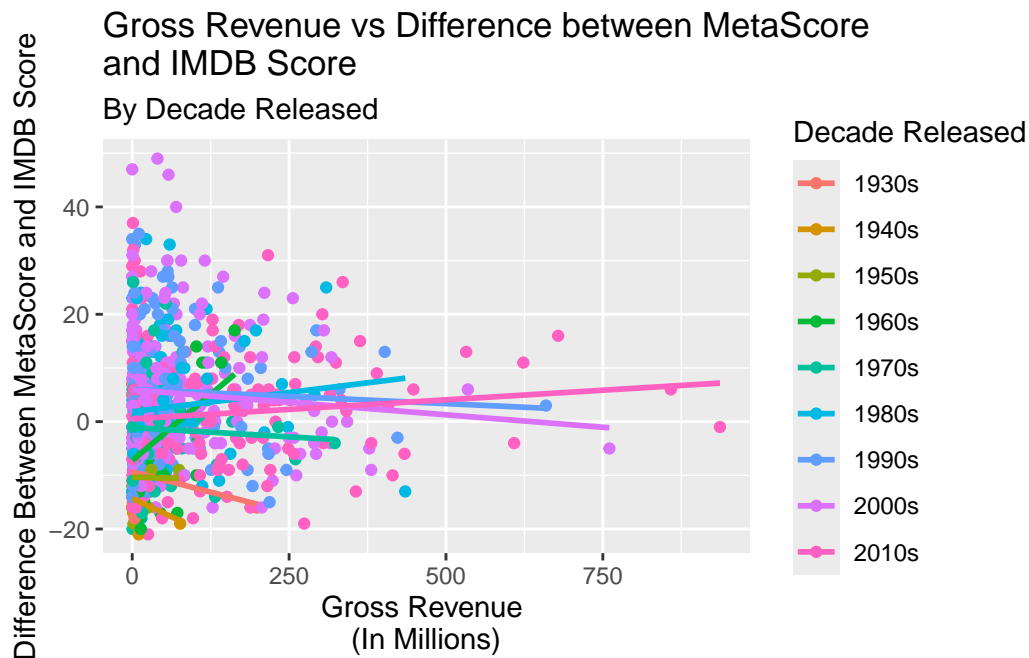
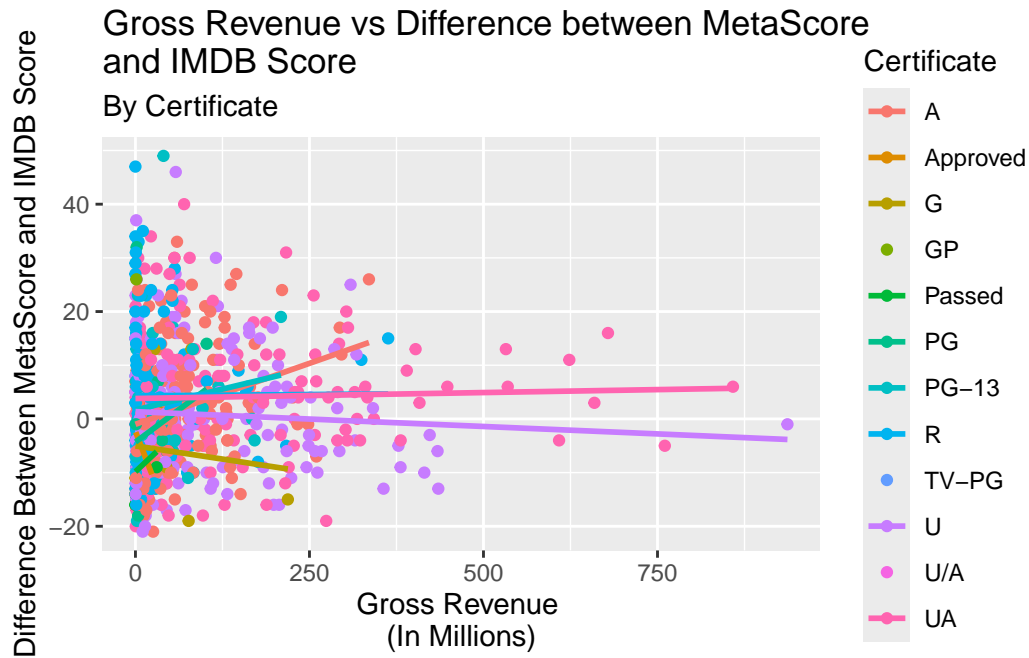
Judging from this EDA, the median difference between `Meta_score` and `IMDB_scaled` also tends to be lower as movies are newer. In other words, MetaScores tend to be lower than IMDB scores in more recent decades.

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.

```
# A tibble: 16 x 5
  term                estimate std.error statistic p.value
  <chr>              <dbl>    <dbl>    <dbl>    <dbl>
1 (Intercept)       -3.70      1.50     -2.46    0.0138
2 gross_cent        -0.000337  0.000862  -0.391   0.696
3 runtime_cent       0.00351    0.00324    1.08    0.278
4 decade1940s     -14.0      420.     -0.0334  0.973
5 decade1950s       0.274      1.43      0.192   0.848
6 decade1960s       2.90       1.52      1.91    0.0564
7 decade1970s       3.02       1.53      1.98    0.0480
8 decade1980s       3.76       1.52      2.48    0.0131
9 decade1990s       4.14       1.51      2.74    0.00612
10 decade2000s      3.84       1.51      2.55    0.0109
11 decade2010s      3.65       1.51      2.42    0.0155
12 certificateOther  2.75       1.22      2.26    0.0238
13 certificatePG     0.0551     0.234     0.236   0.813
14 certificatePG-13  0.167     0.379     0.439   0.660
15 certificateR      0.0593     0.206     0.288   0.773
16 no_votes_scaled  0.264     0.281     0.937   0.349
```