Boom Shakalaka

By: Lindsay Beaman, Kaylee Lemas Santos, Alexis Broderick, and Alyson Barnes







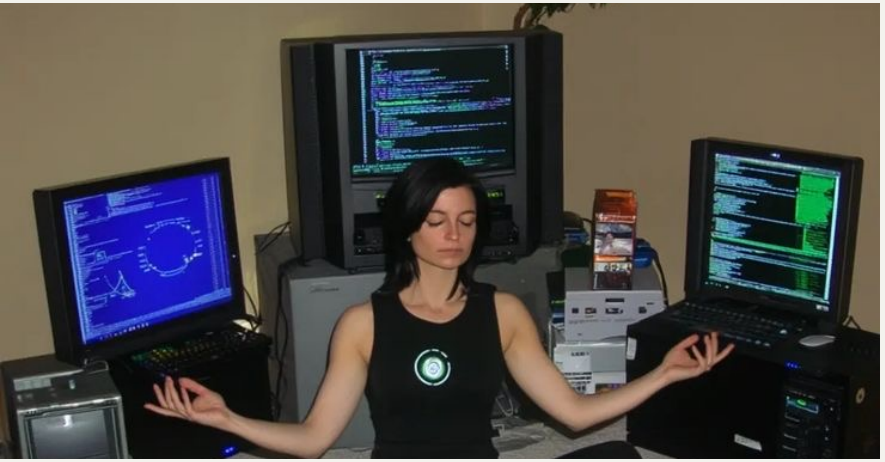
Intro of Topic + Motivation

- Explore what makes a movie highly rated.
- Using IMDb's Top 250 Movies dataset.
- Goal: Identify patterns that influence movie ratings.









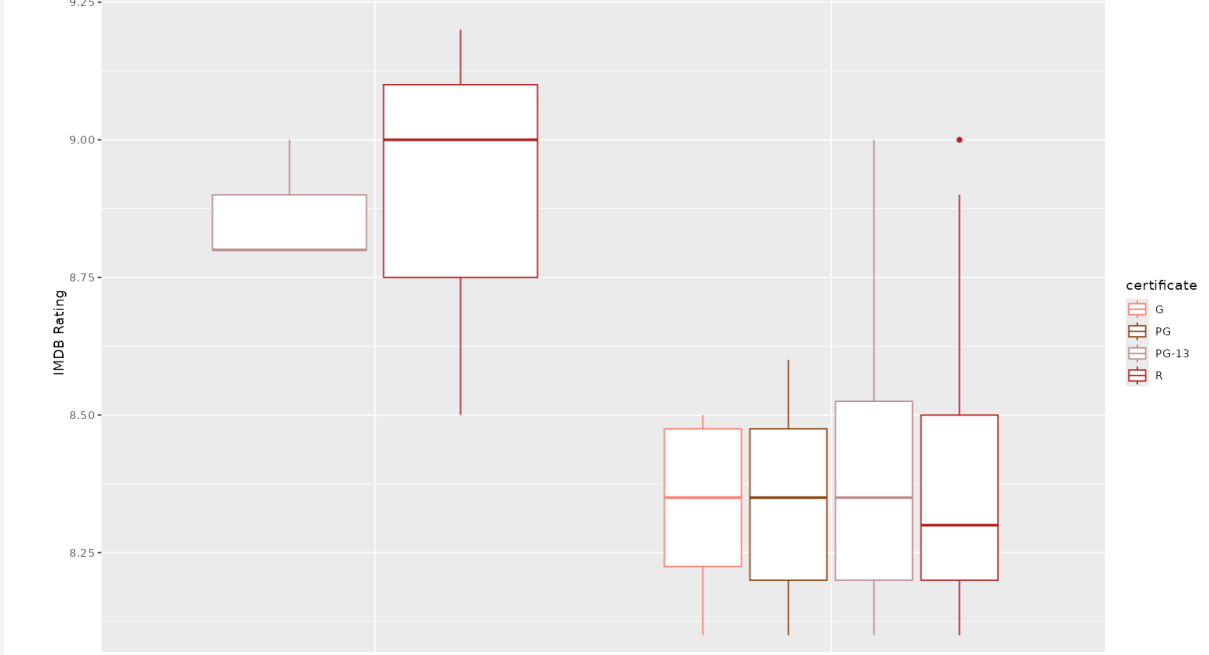


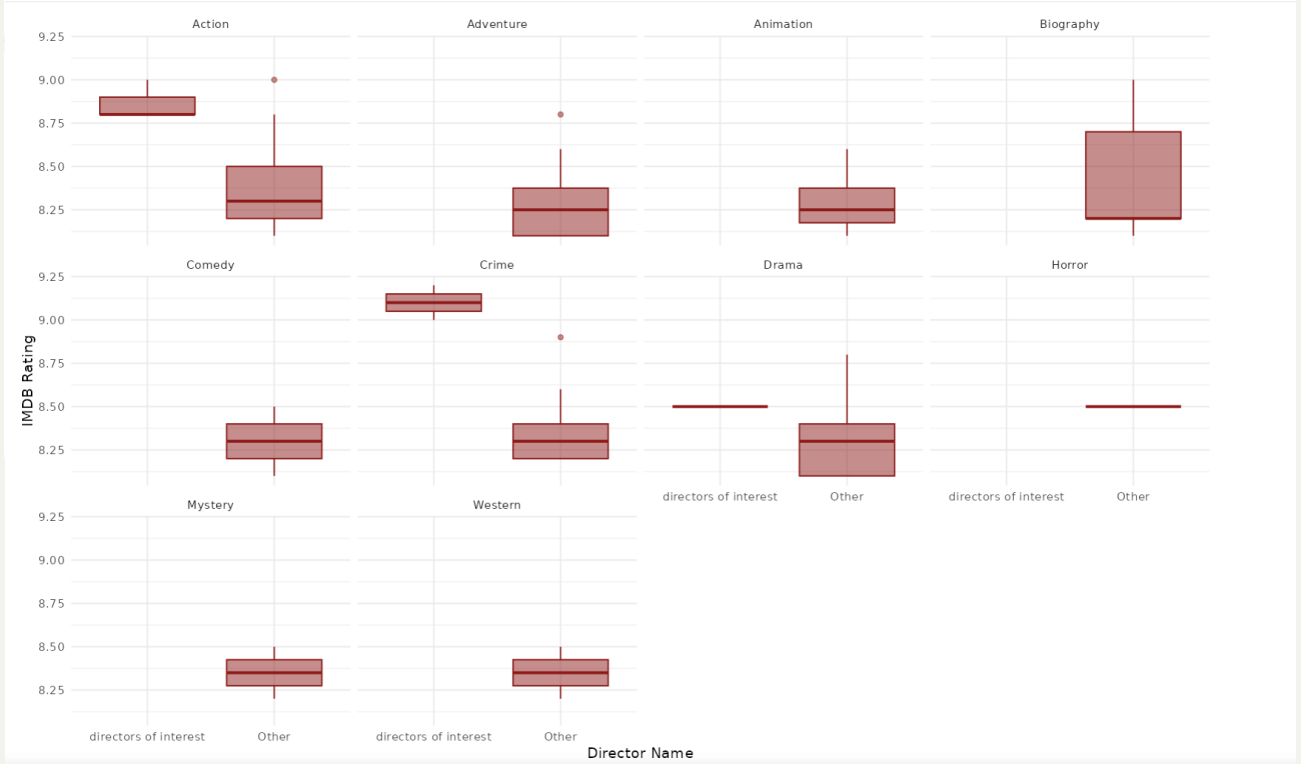
Introduce Data

- IMDb's top 250 movies data set
  - 1990-2026
  - 250 rows x 38 columns
- Genres, directors, year released, certificate, duration, IMDb votes and ratings, and IMDb rankings

Director's Names vs. IMDb Rating

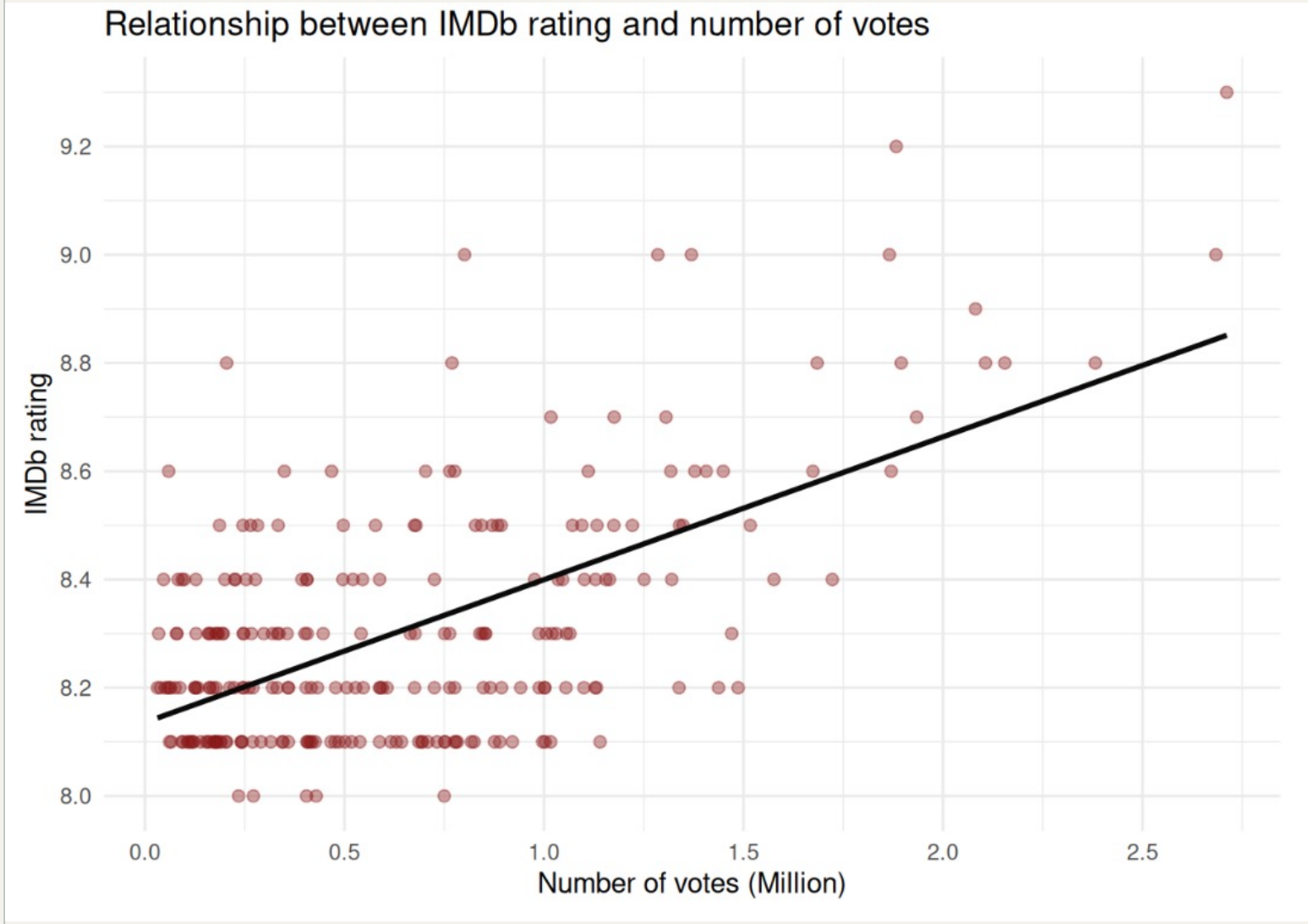
- Peter Jackson and Francis Ford Coppola
  - Both had higher Q1, median, and Q3
  - We created two boxplots, one with color mapped to certificate and one faceted for genre
  - Films were spread out across three genres and two certificates
  - Created a linear model rating to find an adjusted r squared of 0.26
  - Films by Jackson and Coppola tend to be associated with higher IMDB ratings.





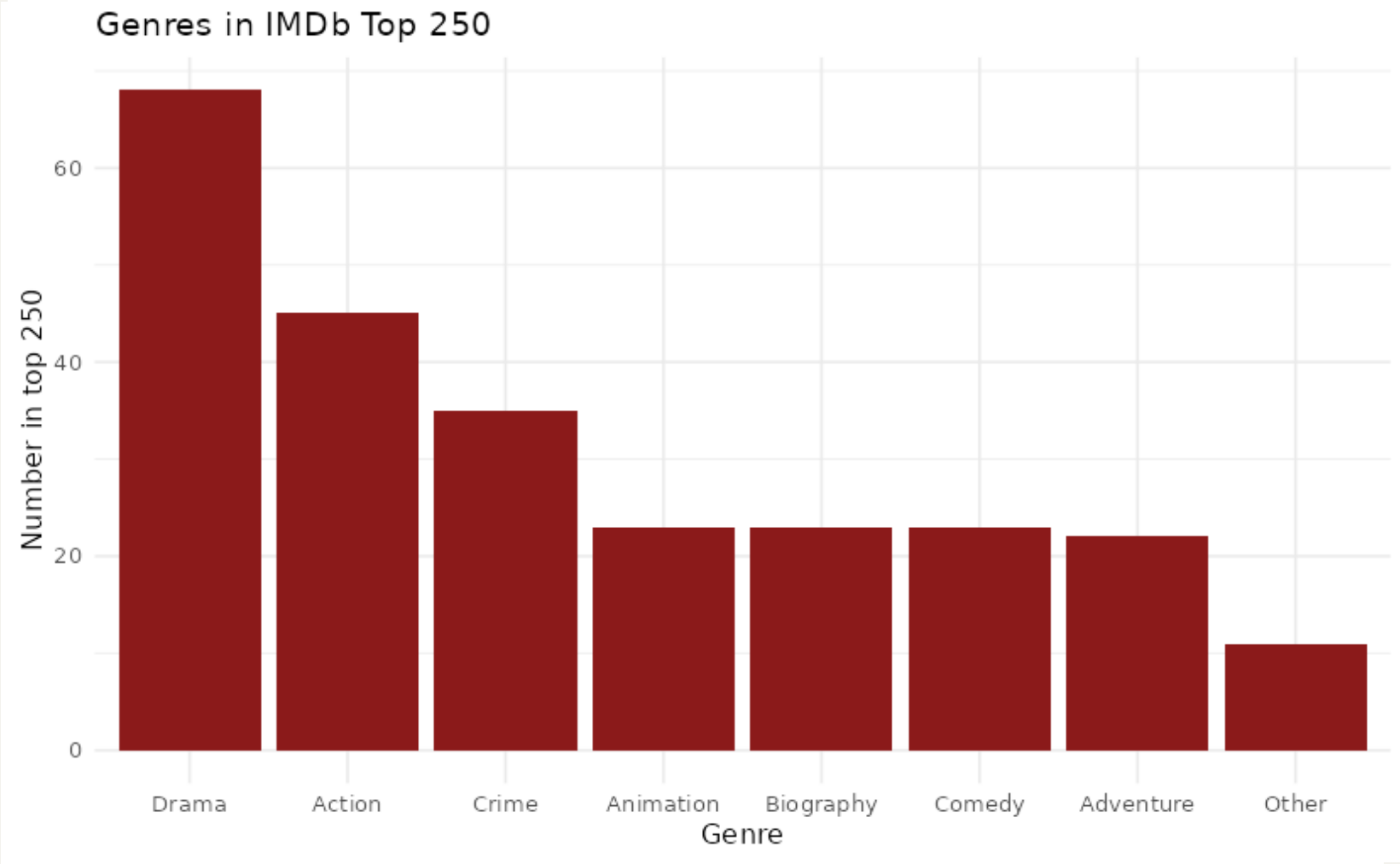
Votes vs rating

- Positive relation between number of votes and IMBDb rating
  - r= 0.603, moderate positive assiocation
- Linear regression showed a statistically significant relationship
  - R-squared of 0.364
- Overall, vote count is assiociated with IMDb ratings, but does not fully explain them

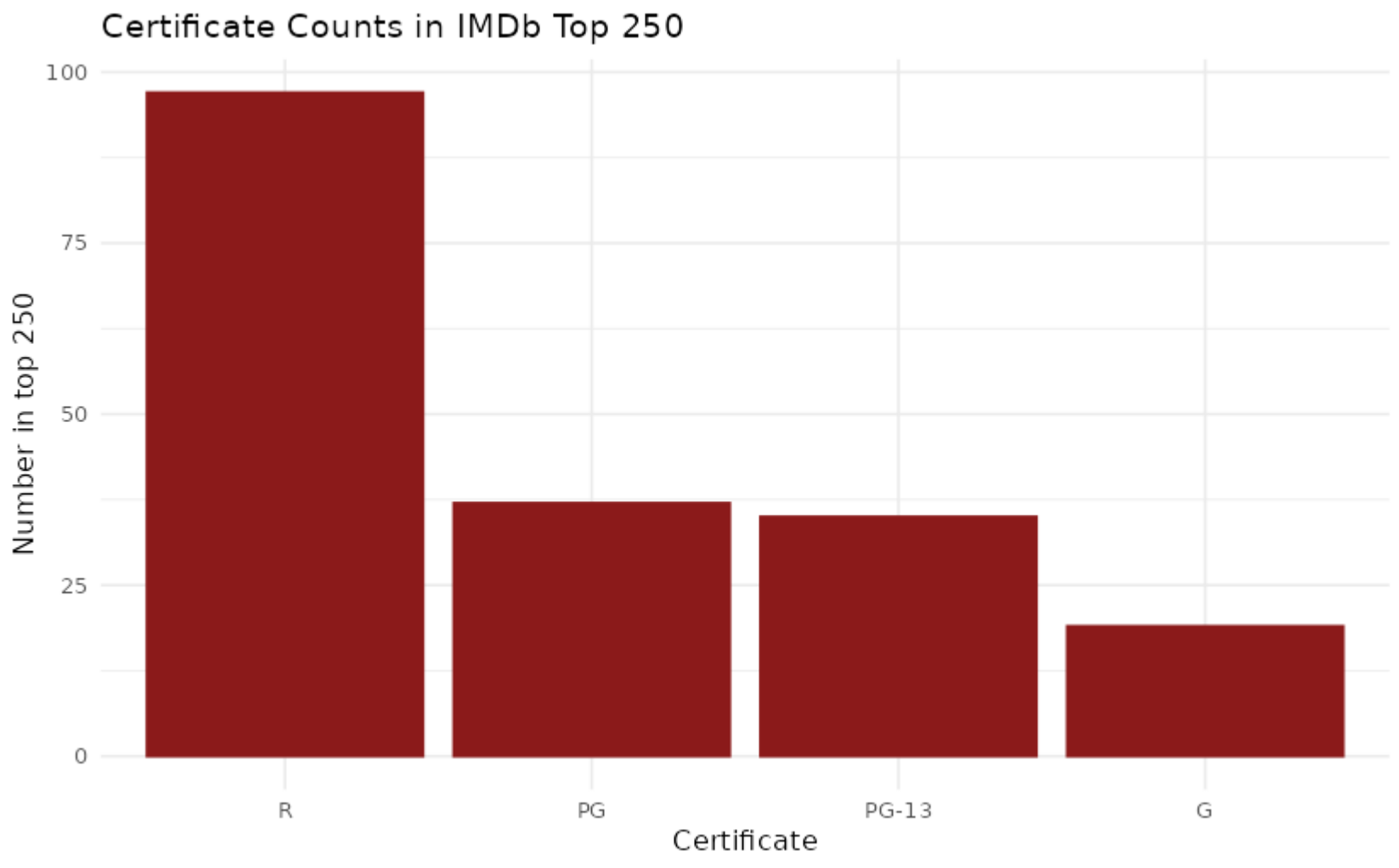


Genre

- Little variation among rating and genre
  - Median 8.2-8.3
  - Minimal variation in tidy summary
- Association observed in count of genres in top 250
  - Cannot prove causation



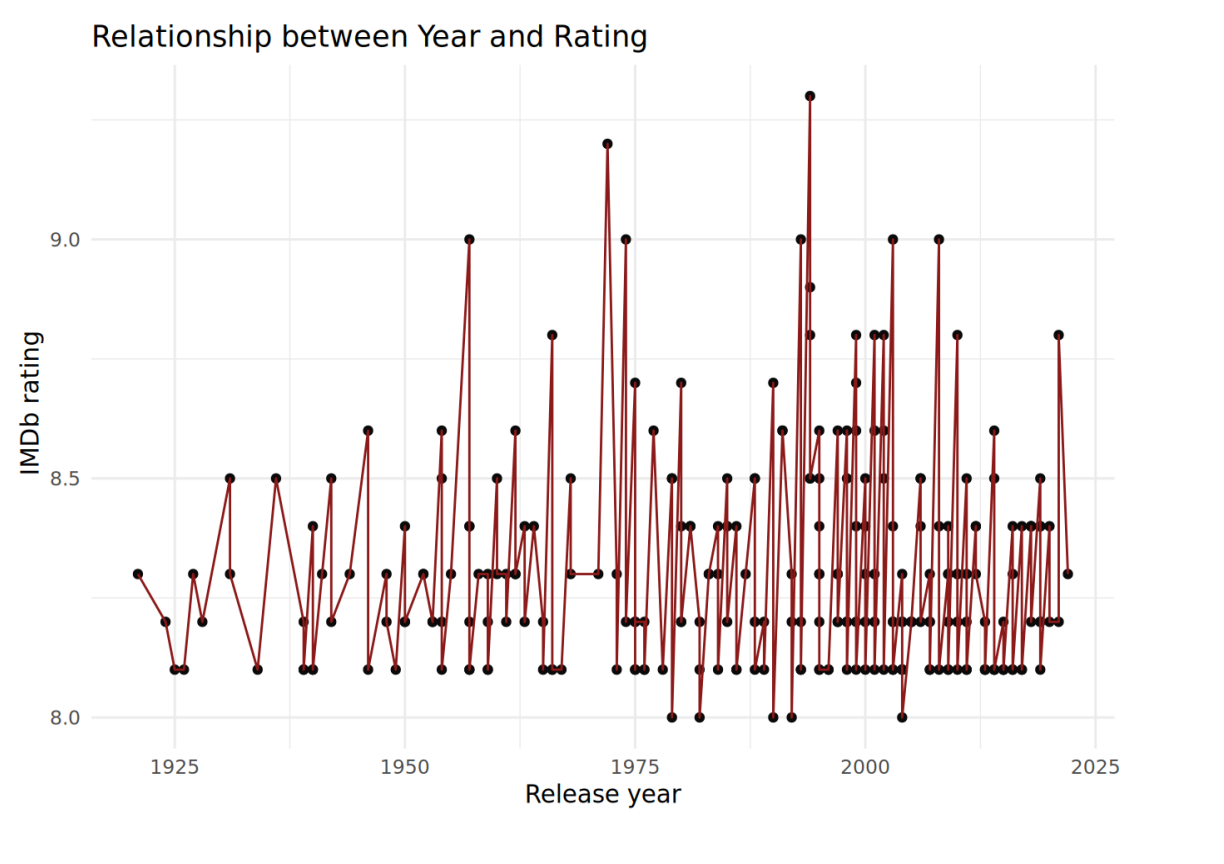
Certificate



- Little variation among certificate and rating
- Possible observed correlation in certificate counts in top 250
  - Cannot prove causation

Release year

- Created a scatterplot and fitted a linear model
  - r = 0.0315
- Line and point graph for further investigation
- Medians:
  - After 2000, 8.2
  - 1950-1975, 8.2
  - Before 1950, 8.3





Conclusion+ Discussion+ Critique

- Limitations
  - Purely correlation/no causal
  - Only top 250 missing large amount of data variety
    - Conclusions only apply to small sample
    - Minimal outliers due to minimal variability
    - None affecting data
- Looking forward
  - Explore more data
    - Larger data set
    - Variety with ratings
  - Explore possible correlation with genre and certificate