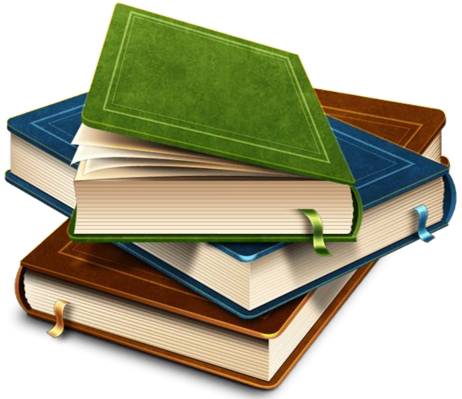


# Predicting the Ratings of Book to Film Adaptations

Amy Zhou

# Background

- Books - great
- Film adaptations - mixed bag

# Problem

- Which films to watch or avoid?

# Data Collection

- 2008-2017
- 216 points after cleaning

IMDb

- Rating
- Runtime
- Director
- Studio
- Genre

Box Office Mojo

- Avg gross of Director(s)
- Studio market share

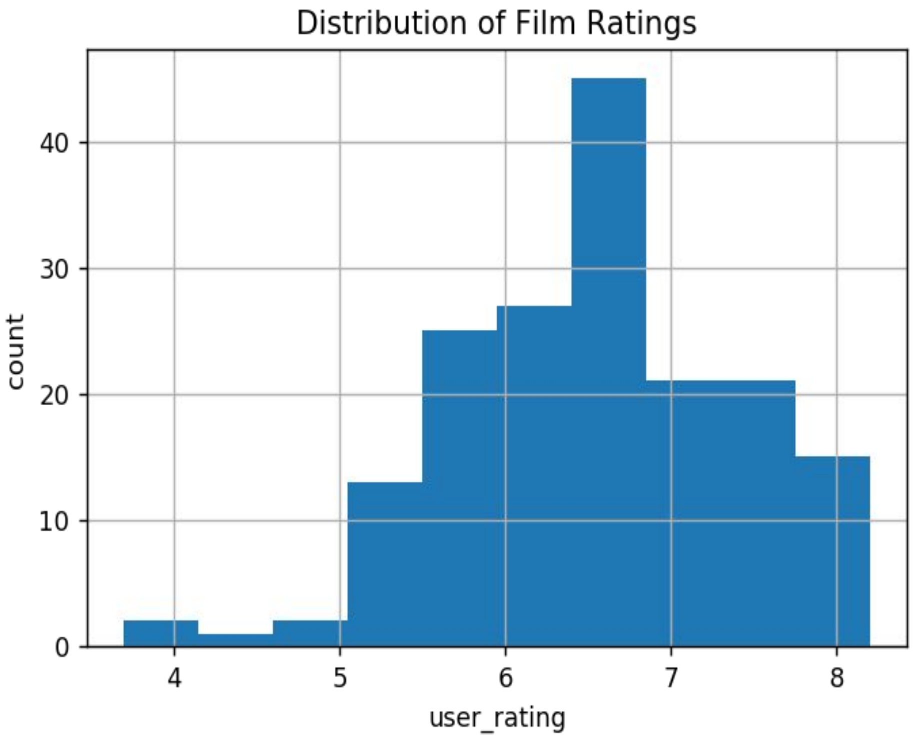
Wikipedia

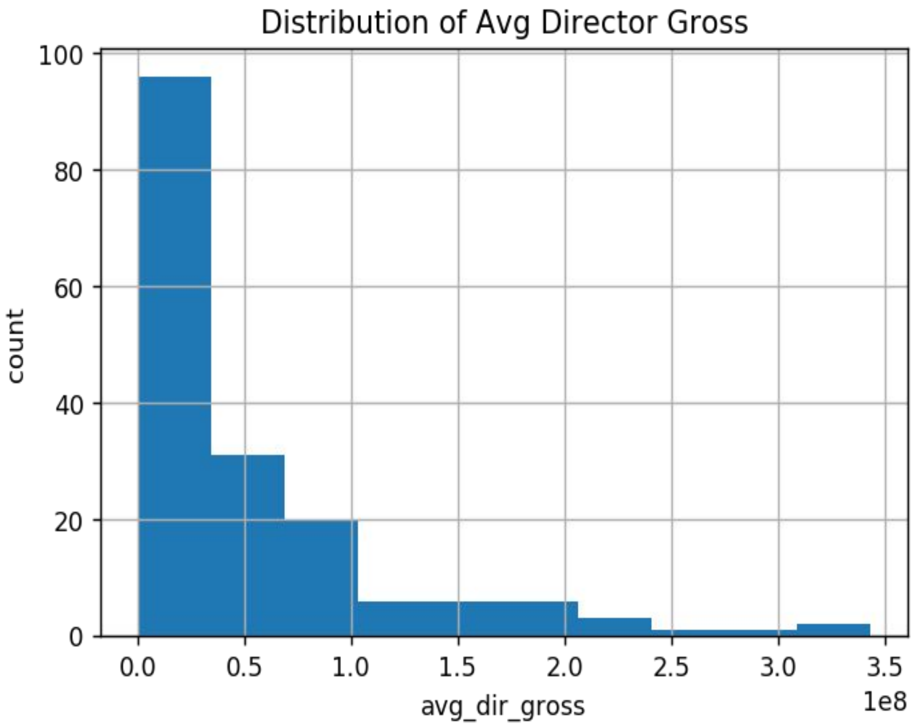
(Film to book link)

goodreads

- Rating
- Reviewer count
- Years between publication and film

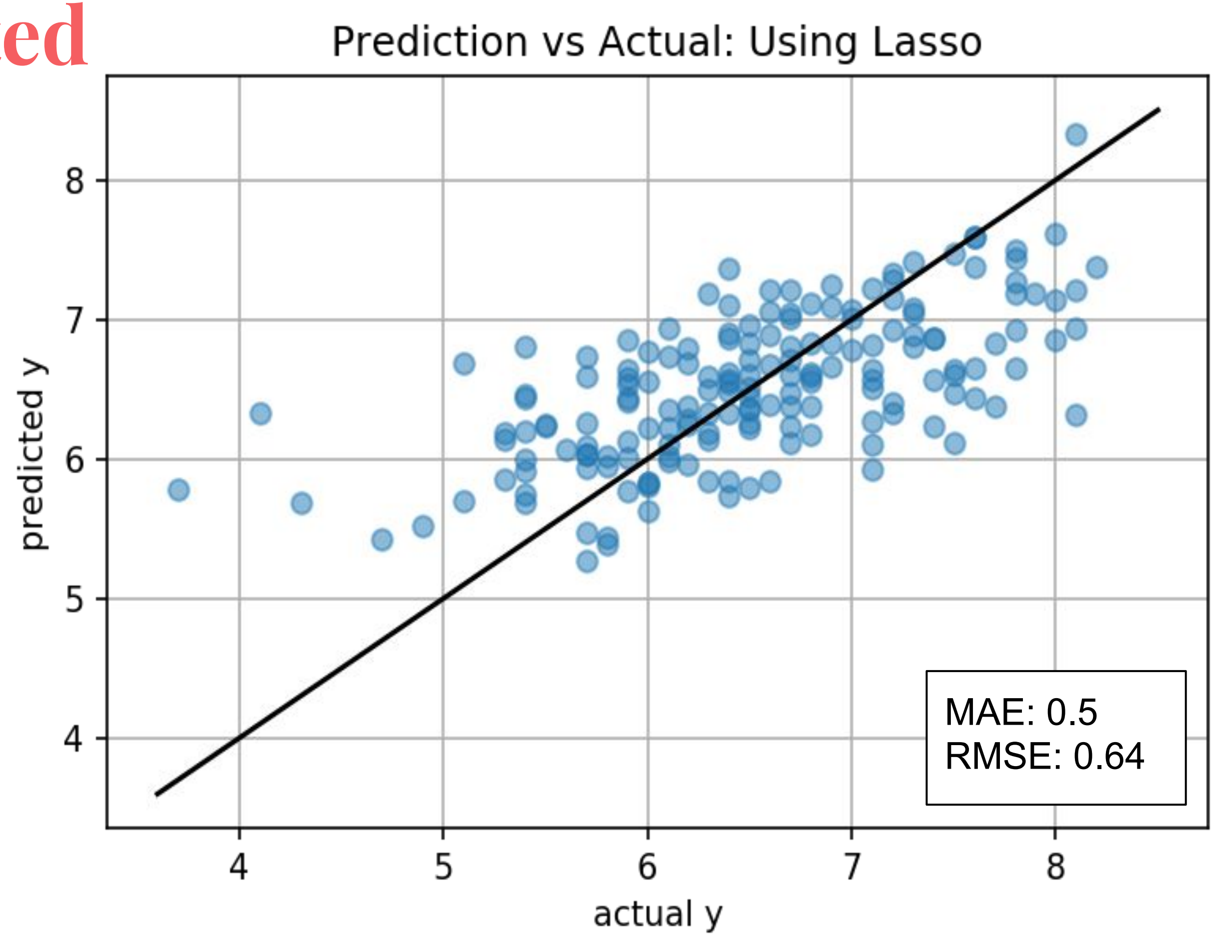
# Data Cleaning/Transformation

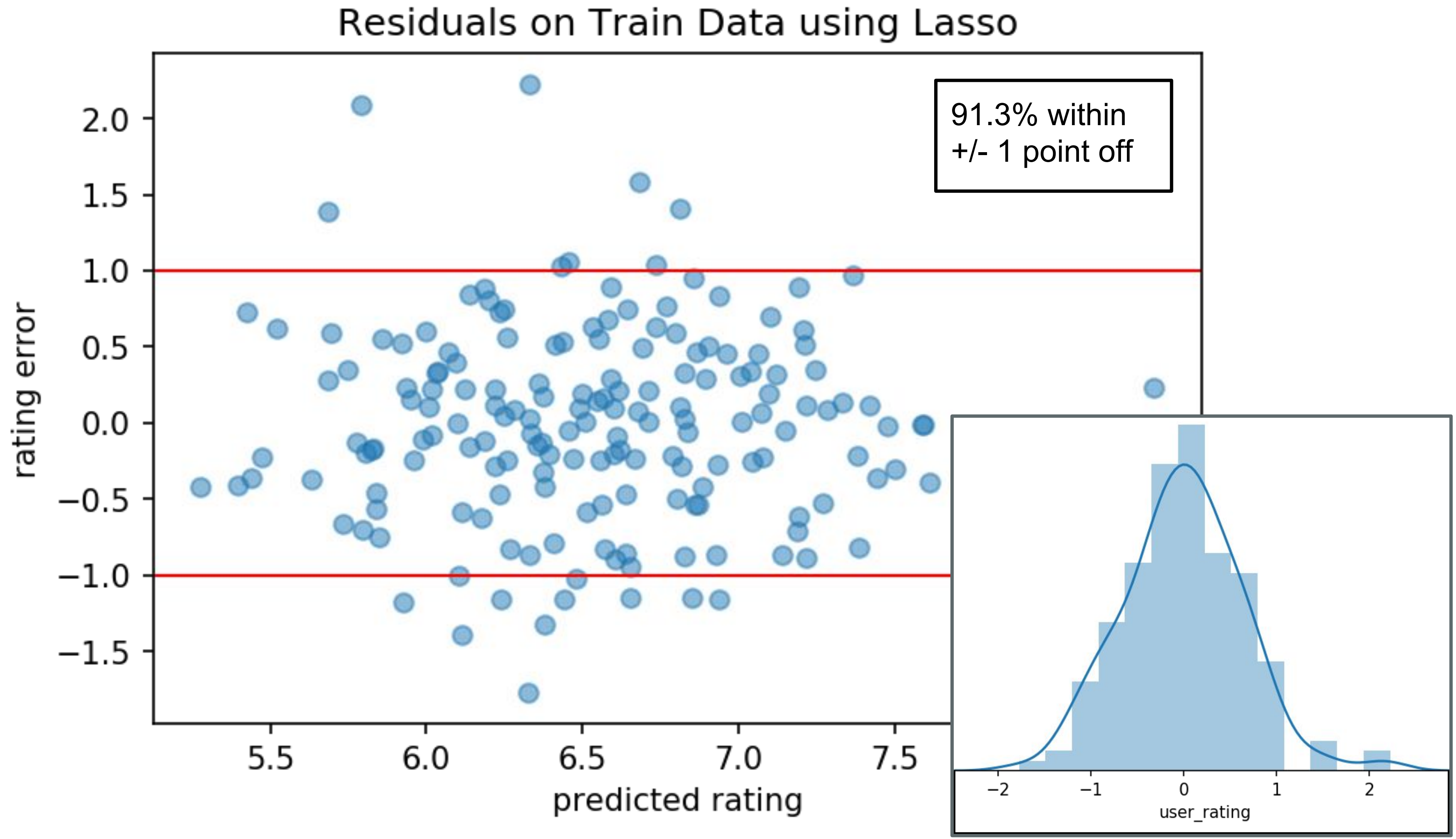


# Models Attempted

- Linear Regression
- Linear with Lasso
- Polynomial Features with Lasso
- Polynomial Features with Ridge
- Linear with Ridge
- Lasso after removing features



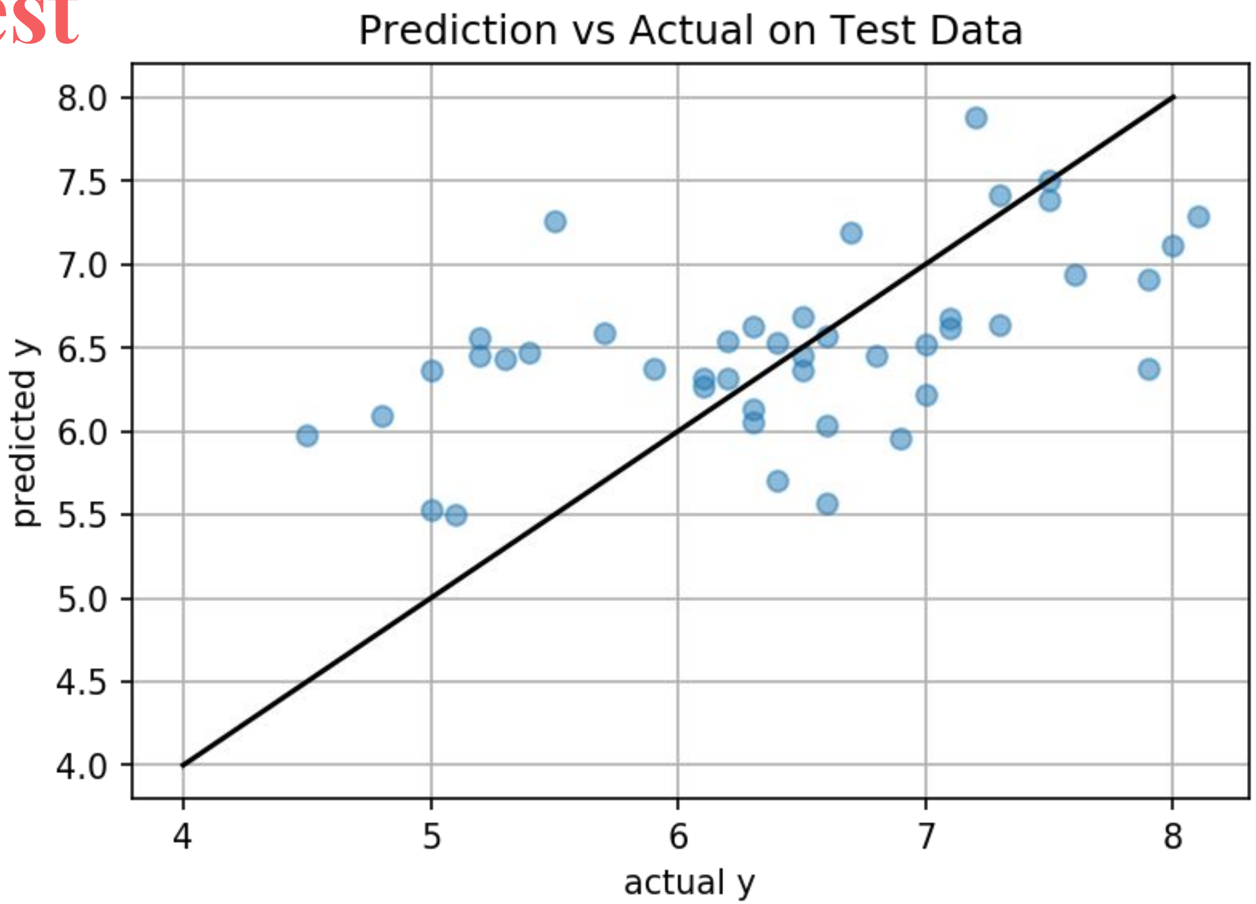


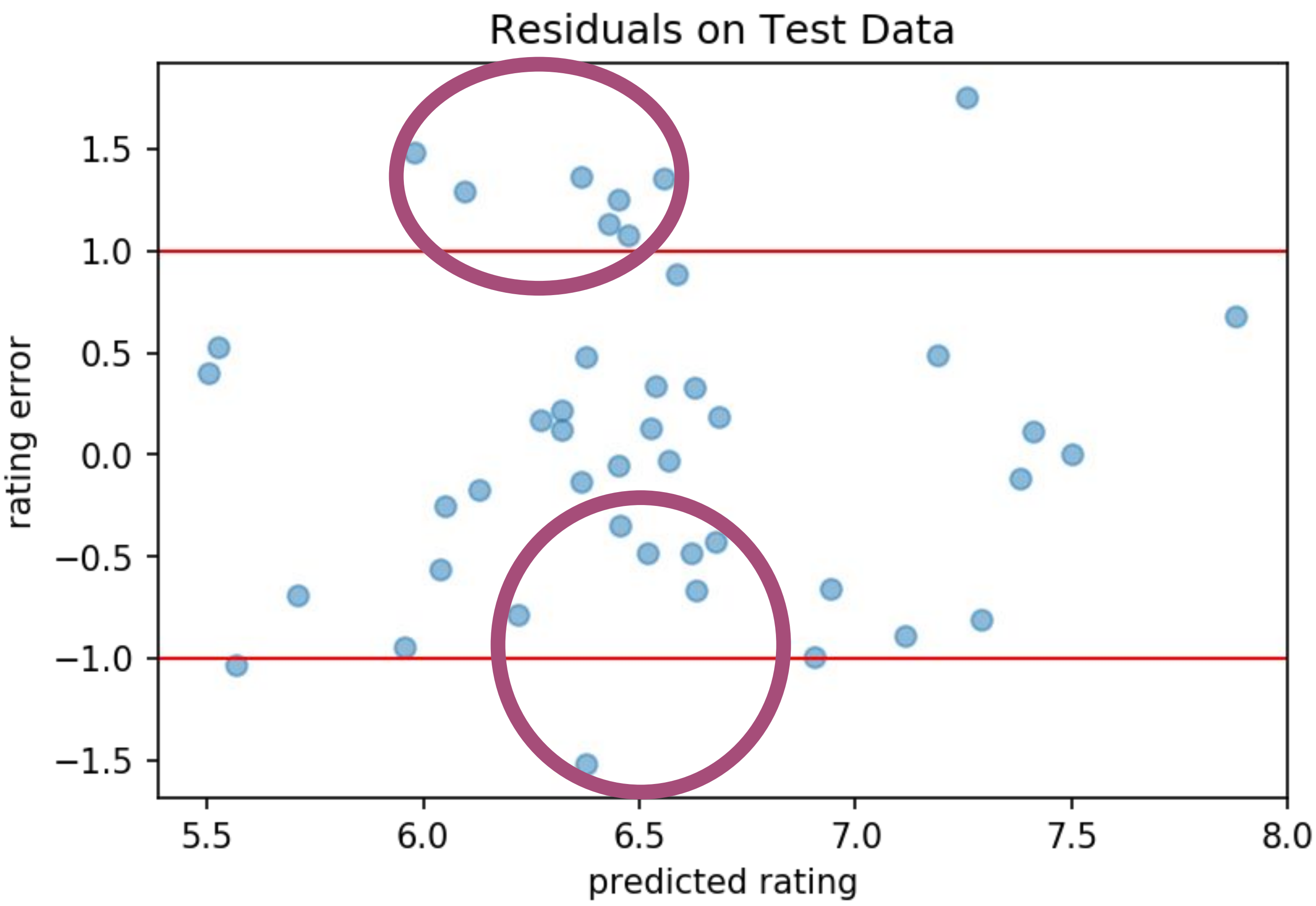
Residuals on Train Data using Lasso
91.3% within
+/- 1 point off
rating error
predicted rating
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
5.5
6.0
6.5
7.0
7.5
−2
−1
0
1
2
user_rating

# Applying to test

- 44 datapoints
- 9 outside the +/- 1 point tolerance

- MAE: 0.63
  Compare to 0.5
- RMSE: 0.78
  Compare to 0.64




Residuals on Test Data
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
rating error
5.5
6.0
6.5
7.0
7.5
8.0
predicted rating

# Conclusion

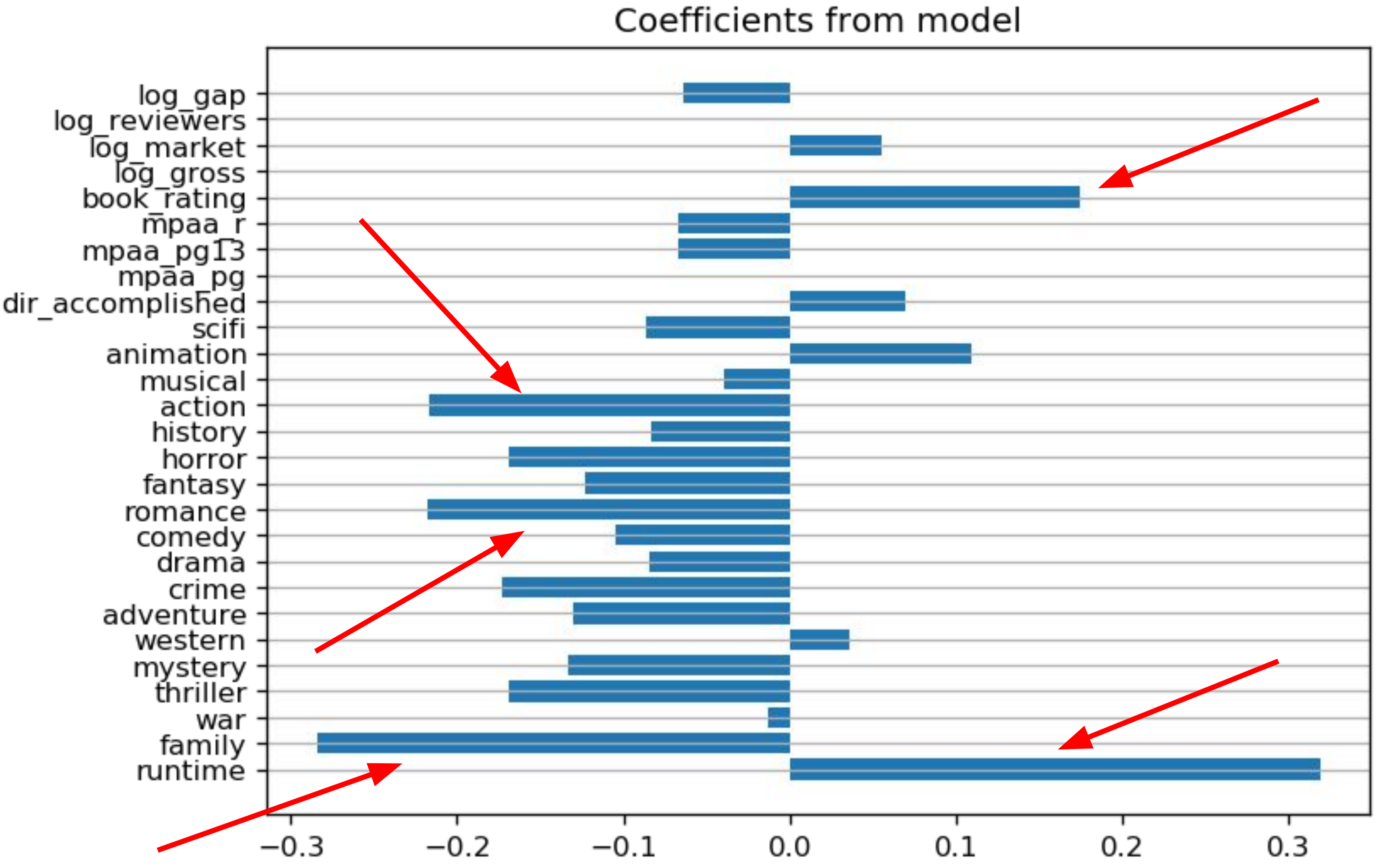
- Model identifies bad movies, most good movies
- Avoid disappointment!



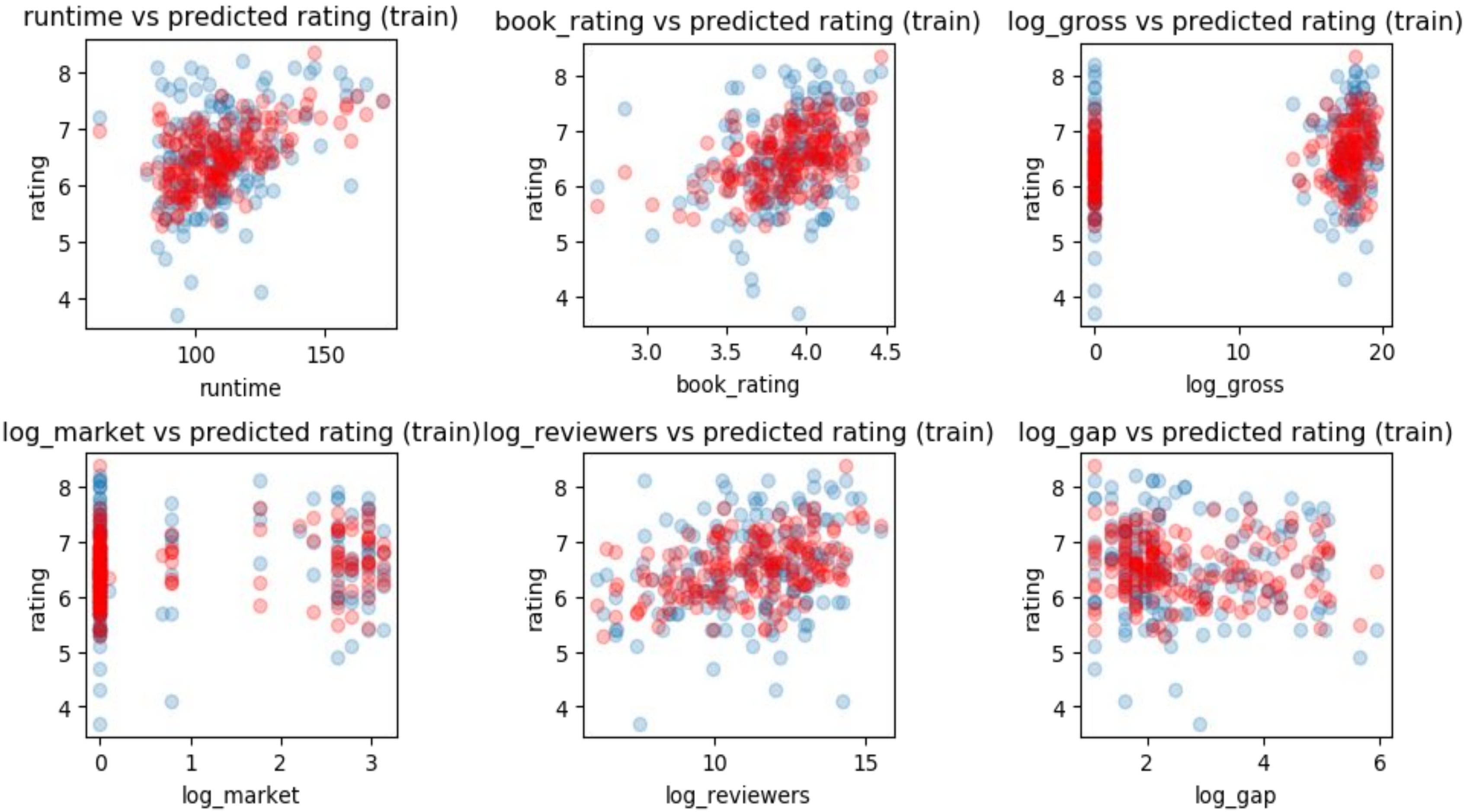
# Next Steps

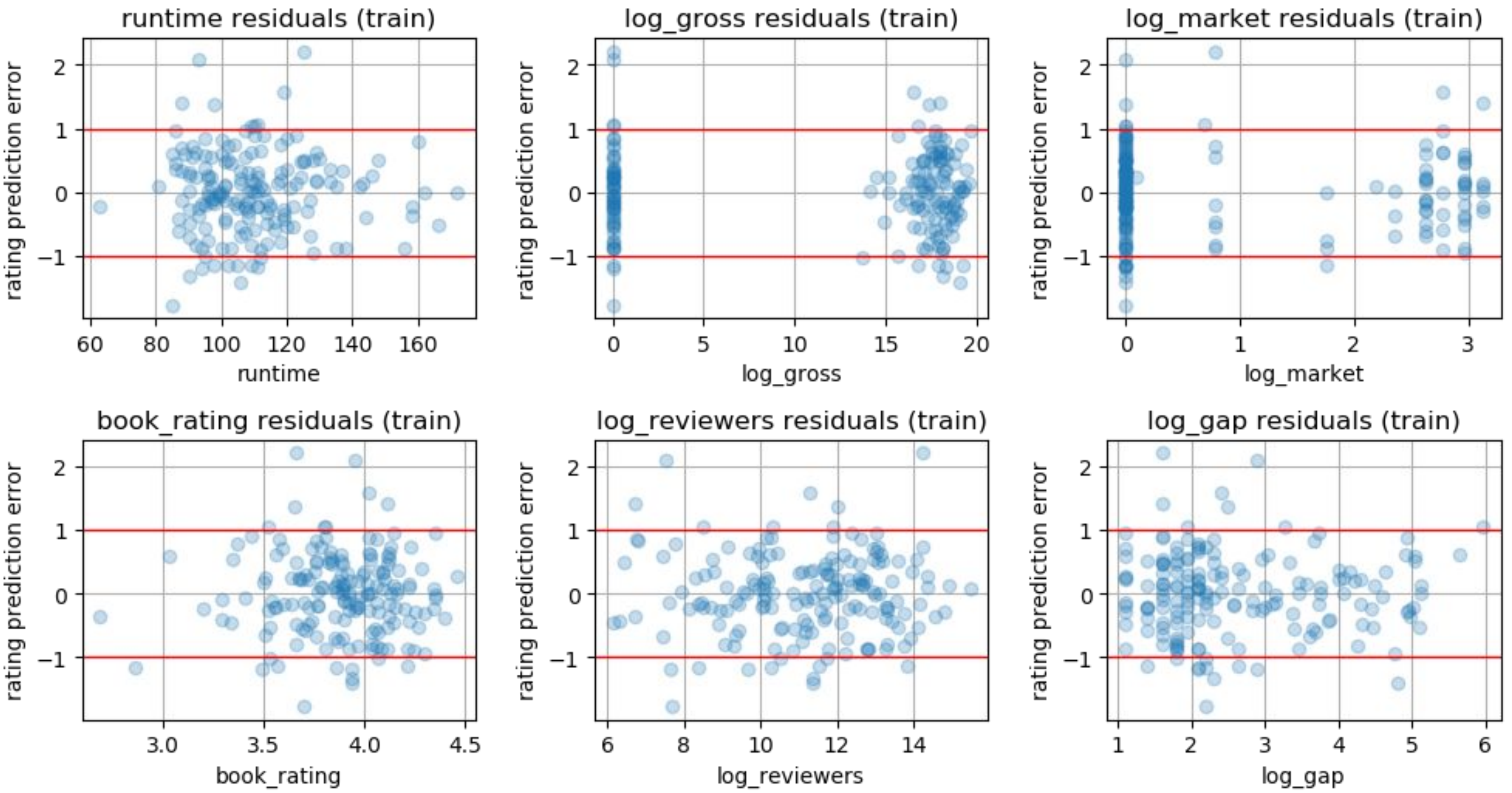
- Expand date range
- Different features/ feature engineering
- Explore per genre/clustering
- Set lower expectations

# Appendix



Coefficients from model
log_gap
log_reviewers
log_market
log_gross
book_rating
mpaa_r
mpaa_pg13
mpaa_pg
dir_accomplished
scifi
animation
musical
action
history
horror
fantasy
romance
comedy
drama
crime
adventure
western
mystery
thriller
war
family
runtime
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3


runtime vs predicted rating (train)
book_rating vs predicted rating (train)
log_gross vs predicted rating (train)
log_market vs predicted rating (train)
log_reviewers vs predicted rating (train)
log_gap vs predicted rating (train)
rating
runtime
book_rating
log_gross
log_market
log_reviewers
log_gap


runtime residuals (train)
log_gross residuals (train)
log_market residuals (train)
book_rating residuals (train)
log_reviewers residuals (train)
log_gap residuals (train)
rating prediction error
runtime
log_gross
log_market
book_rating
log_reviewers
log_gap