An Inferential Study of Movie Characteristics and Female Representation in the Film Industry

STAT 634 SPRING 2021

TEAM MATRIX

The Bechdel test was created by cartoonist Alison Bechdel in this 1985 cartoon.
Source: https://dykestowatchoutfor.com/the-rule/

1 Introduction

Hollywood movies and TV shows have historically lacked on-screen female and racial minority representation. In 2020, UCLA published its annual Hollywood Diversity Report that quantifies the representation of these groups in films and TV shows. They found that in 2020, minorities represented 33% of total actors and females represented 40% of total actors in the film industry. Although these numbers are increasing over time, there is still a large disparity between the representation of these groups in the population compared to their representation in the film industry (Hunt, 2020). One challenge is that determining if films are "diverse" is somewhat subjective and datasets with movies' diversity characteristics are rare. However, in 1985 cartoonist Allison Bechdel coined the Bechdel test, a heuristic for quantifying the level of female representation in movies. The Bechdel test is a simple 4 level rating system with a movie either receiving a score of 0, 1, 2, or 3 as follows (bechdeltest.com, n.d.):

0. The movie does *not* have at least two named women in it
1. The movie has at least two named women in it, but they do not talk to each other.
2. The two named women talk to each other, but they talk about a man.
3. The two named women talk to each other about something other than a man. (i.e., The movie passes the Bechdel Test)

The Bechdel test seems like a low threshold for movies, but a substantial portion of movies do not pass the test. In this study, we infer which movie characteristics are most associated with passing the Bechdel test based on data from Bechdeltest.com and IMDb.com. With the acknowledgement that the Bechdel test does not assess racial diversity nor full female representation, we aim to study if and how certain genres, release years, and IMDb ratings are associated with the outcome of the Bechdel test.

Given this purpose, our analysis methods are explanatory and inferential in nature. We focus on modeling the true generation of our data rather than making a prediction. We consider logistic regression which is ideal for our aim of explanatory modeling with variable selection. In many disciplines, there is near-exclusive use of statistical modeling for causal explanation and the assumption that models with high explanatory power are inherently of high predictive power. The conflation between explanation and prediction is common; however, the type of uncertainty associated with explanation is different than that associated with prediction (Shmueli, 2010). So, we add a random forest model with feature ranking by importance to compare the performance against our logistic regression model. We also explore potential temporal trends and changes of the relationships between our variables with the Bechdel test outcome. For example, if the IMDb rating is trending down overall, does that mean people over time are getting stricter in the rating, and how this trend compares with the Bechdel test rating? It could be interesting to look into these as there may be trends within the features themselves that indicate score improvements over the years.

In the following sections we describe the data detailing the sources, our preprocessing steps, and an exploratory data analysis. Next, we present the methods and models used throughout the study with the corresponding results. Finally, we discuss our conclusions, limitations of the study, and potential future research directions.

2 Data

The dataset we use in this analysis is derived from three publicly available datasets, one from Bechdeltest.com (bechdeltest.com, n.d.) and two from IMDb.com (IMDb, n.d.). The Bechdeltest.com dataset contains the Bechdel Test ratings, and the IMDb datasets contain all other movie characteristics. In total 8,200 movies are common between the three datasets. The datasets are joined by a common identifying field, IMDb ID. Table 1 shows the relevant variables from the joined dataset with descriptions.

Table 1 Relevant variables from the joined dataset.

Variables	Explanation
IMDb ID	IMDb ID of the movie; a unique identifier of the movie
Title	Title of the movie
(Bechdel) Rating	Bechdel rating; an integer from 0 to 3 corresponding to the results of the Bechdel test
Year	Year the movie was released (according to IMDb)
Runtime (Minutes)	Primary runtime of the title, in minutes
Genres	List of up to three genres associated with the movie
Average (IMDb) Rating	Weighted average* of all the individual user ratings; a continuous numeric value from 0 - 10
Number of Votes	The number of votes (ratings) the movie has received on IMDb

*IMDb does not disclose the exact method used to generate the Average Rating noting, "…although we accept and consider all votes received by users, not all votes have the same impact (or '*weight*') on the final rating" (Rating FAQ, n.d.).

In addition to the raw variables, we also derive several variables to use in our analyses. First, we derive the binary variables “Full Pass” and “Mostly Pass” based on the Bechdel test rating. The variable “Full Pass” is “True” if the movie has a Bechdel rating of 3, otherwise it is “False”. The variable “Mostly Pass” is “True” if the Bechdel rating is 2 or 3, otherwise it is “False”.

Secondly, since we are highly interested in exploring which genres might help infer the Bechdel test result, we derive 10 binary variables from the original variable “Genres”. From Table 1, the original “Genres” variable is a list of up to 3 genres in alphabetical order (i.e., the ordering is not meaningful). We parse this value to create 10 new variables named for each of the 10 most prevalent movie genres: Action, Adventure, Comedy, Crime, Drama, Fantasy, Horror, Mystery, Romance, and Thriller. We create an 11th variable named “Other Genres” for movies that are associated with other genres not in the top 10 (e.g., biographies, musicals). These 11 binary variables are labeled “True” if the movie is associated with the respective genre, otherwise it is “False”.

Finally, based on a threshold of 30 movies per release year, we feel that there is not a representative sample of data for years prior to 1953 and for 2021. We remove these years of data which leaves 7,630 movies from 1953 – 2020. No movies from between 1953 and 2020 were excluded.

3 Exploratory Data Analysis

Exploratory data analysis is an important first step in any analytical endeavor and is essential for building a better understanding of important data features and characteristics, uncovering data-driven evidence that supports scientific hypotheses of interest (or generates new hypotheses to consider), and determining the most appropriate statistical procedures for further analysis. This section provides several charts where we aim to achieve an initial understanding of the variables and make an initial hypothesis about the relationships of variables in our dataset.

3.1 Bechdel test results over time

The proportion of movies that fully pass the Bechdel test each year can reflect long term trends in movie diversity. We first calculate the proportion of movies that fully pass the Bechdel test from 1953 to 2020 (shown in Figure 1). We see that although the proportion fluctuates, sometimes sharply, the overall trend of the proportion is rising. In other words, the diversity of movies by measure of the Bechdel test, is generally increasing each year. The red dot in Figure 1 shows the proportion of movies that completely pass the Bechdel test in 1993. This year is notable because from 1993-2020 at least 50% of the movies fully pass the Bechdel test.

This may coincide with film diversity and potentially the presence of female actors receiving more attention starting in the 1990s. Other notable years are 1966 where the proportion of films that completely pass the Bechdel Test is lowest at about 28%, and 2020 where the proportion is highest, at 80%. Movies in years with a higher pass proportion may also be more likely to achieve gender parity in terms of speaking roles.

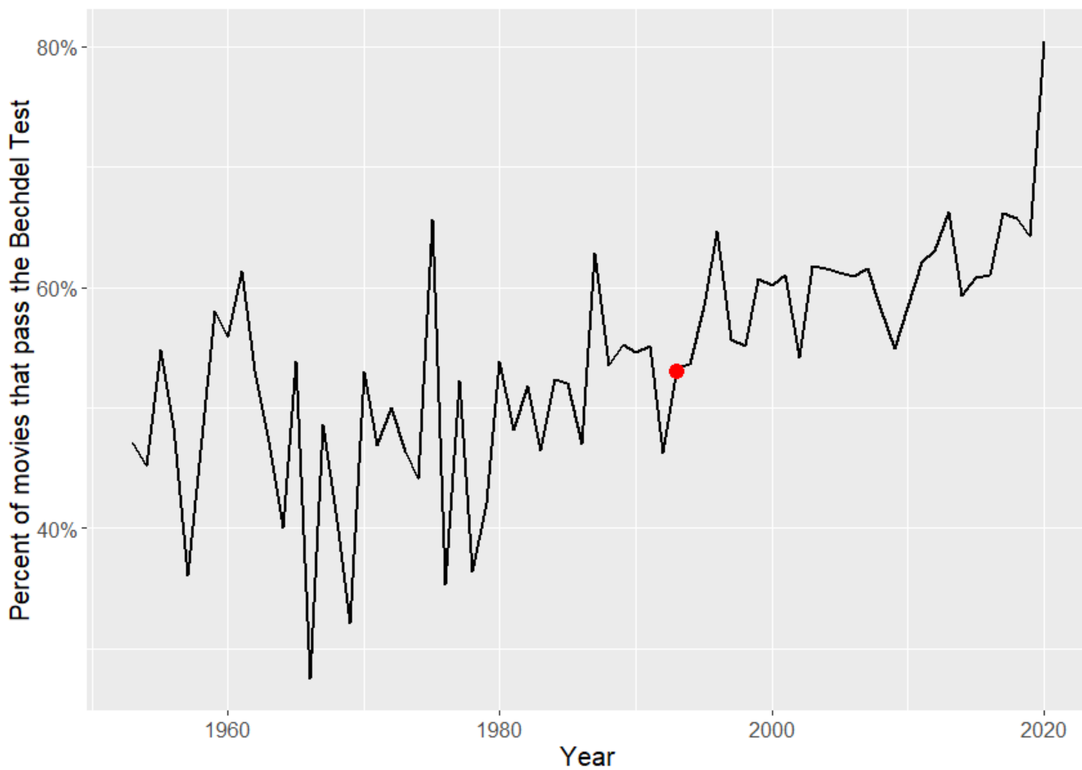

Figure 1 Percent of movies that fully pass the Bechdel test each year showing a long-term upward trend

3.2 IMDb rating trends

Under the assumption that IMDb movie ratings substantially reflect the attitudes of movies from the general population, we show how these IMDb ratings change over time in Figure 2. We see that the average IMDb rating, grouped by Bechdel test rating, over time shows a downward trend for all groups. Before 1990, the average IMDb rating of movies that do not include two named female characters (Bechdel rating of 0) experience a plateau and maintain a high average rating during this period. In fact, from 1965 to 2004, the average IMDb rating for movies that do not include two named female characters is highest among all Bechdel test groups. We surmise that film teams that produce high-quality and highly rated movies may be more inclined to maintain their habits of minimal female representation, which slows the improvement of gender diversity in movies. After about 2010, the average IMDb ratings plummet for the movies with Bechdel ratings of 0, suggesting movies with fewer than two female characters may no longer be as popular with audiences. However, the average rating of all movies, regardless of the Bechdel rating, is on a downward trajectory.

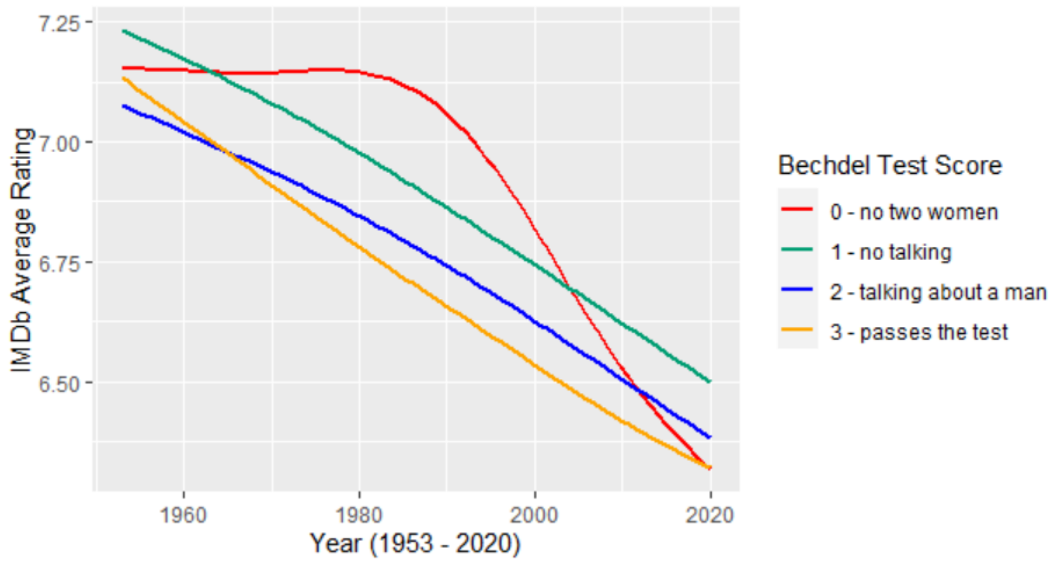

Figure 2 Average IMDb ratings by year grouped by Bechdel test results show an overall downward trend in average IMDb ratings

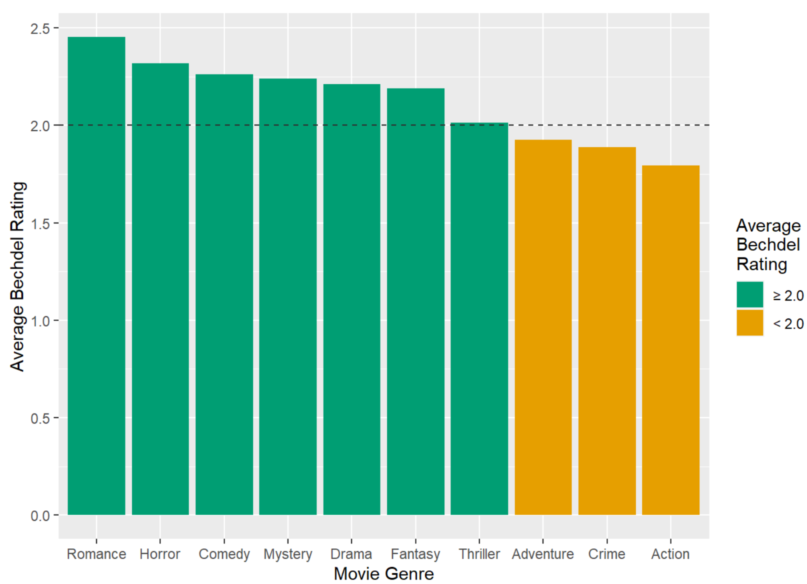

Figure 3 Average Bechdel rating by top 10 most prevalent genres. Nearly 99% of movies in our dataset are associated with at least 1 of these genres.

Figure 3 shows the average Bechdel rating for the top 10 most prevalent genres in our dataset from 1953 to 2020. The average ratings are centered around 2, which is associated with movies that have at least 2 named women in them who talk to each other. The chart is arranged from left to right in descending order of average Bechdel rating for different genres. The genre with the highest average Bechdel rating is romance, compared to the lowest rated genre, action. Even the top-rated genre does not round up to 3, or fully passing the Bechdel test. Although adventure, crime, and action films are rated below 2.0 on average, their ratings still round up. Nearly 99% of the movies in our dataset are associated with at least one of these 10 genres. Also, a movie can be associated with up to 3 genres so for example, a single movie could be included in the average Bechdel rating of romance, comedy, and drama genres.

4 Analysis

In this section we consider several statistical methodologies to infer the associations between movie characteristics and the Bechdel test rating. We begin with a binary explanatory variable setting with penalized logistic regression, then expand to a multi-class explanatory variable setting using random forest methods.

4.1 Inference of movie characteristics on fully passing the Bechdel test

We first consider the case of movies that fully pass the Bechdel test versus those that fail or partially pass the Bechdel test. This scenario is framed as a binary classification problem where movies with a rating of 3 are considered to fully pass the Bechdel test, otherwise they do not.

Our first model is a logistic regression model with a binary response variable, fully passing or not, considering all explanatory variables, i.e., Year, Run Time, Average Rating, Number of Votes, and the 11 binary variables encoding the genres. This full logistic regression model yields statistically significant coefficients (p-value < 0.05) for Year, Run Time, Average Rating, Action, Adventure, Crime, Horror, Romance, Thriller, and Other Genres. On the other hand, the coefficients associated with Number of Votes, Comedy, Drama, Fantasy, and Mystery are not statistically significant.

Unfortunately, the full logistic regression model may include variables that are strongly correlated with each other. This collinearity affects the coefficient estimates from the logistic regression model fit and may lead us to draw incorrect conclusions regarding the associations of variables with fully passing the Bechdel test. For this reason, we are interested in fitting a model with as few explanatory variables as possible, while still sufficiently modeling the probability of passing the Bechdel test. To accomplish model reduction, we perform a LASSO regression with the binomial family and refer to this as the penalized logistic regression model. The LASSO regression method adds a penalty term to the log-likelihood function used in the standard logistic regression. The penalty term is the sum of coefficient magnitudes in the model multiplied by a positive factor, λ. The goal of the LASSO method is to minimize the penalized negative log-likelihood at different values of λ, which consequently forces the coefficients to 0, effectively eliminating them from the model.

We perform the LASSO regression method considering all explanatory variables and use the value of λ that gives the simplest model that is within 1 standard error of the best model (i.e., model with minimum mean cross-validated error) (Hastie, 2014). Using this value of λ, known as the “1 SE rule”, generally yields a simpler model than the model with the lowest error overall. It generally has lower predictive accuracy, however, since we are interested in associations, not predictions, we elect to base our conclusions on simpler models over those with the best predictive capabilities. Table 2 shows the remaining exponentiated coefficient estimates after the LASSO regression method, as they relate to the odds of a movie fully passing the Bechdel test.

Table 2: LASSO regression coefficient estimates in the model for the probability of fully passing the Bechdel test.

Explanatory Variable	Odds ratio of fully passing the Bechdel test
Year	1.009
Average Rating	0.879
Action	0.545
Adventure	0.964
Crime	0.692
Horror	1.050
Romance	1.304
Thriller	0.956
Other Genre	0.866

We see that the penalized logistic regression model includes 9 explanatory variables, 6 less than the full logistic regression model. The explanatory variables with odds ratios greater than 1 are associated with an increased probability that a movie fully passes the Bechdel test. Romance movies, which are most associated with an increased probability of passing the Bechdel test, have about 30% higher odds of passing than non-romance movies. Similarly, horror movies, have about 0.5% higher odds of passing the test than non-horror movies.

Additionally, Year is associated with an increased probability that a movie fully passes the Bechdel test. For each additional year in which the movie is released, the odds of a movie passing the Bechdel test increases by 0.9%. This may seem negligible, but over time becomes more meaningful. For example, a movie released in 2020 would have 18% higher odds of passing the test compared to a movie released in 2000.

Conversely, explanatory variables with odds ratios less than 1 are associated with a decreased probability that a movie fully passes the Bechdel test. The strong negatively associated variables are action, adventure, crime, thriller, and movies in "other genres". The average IMDb rating of a movie is also associated with a decreased probability of fully passing the test. Perhaps not surprisingly, action and crime films have 46% and 31% lower odds of passing the Bechdel test, respectively, than movies that do not fall into these genres.

We also perform the LASSO regression on the same full set of explanatory variables except with Year removed. The intent is to infer the associations of passing the Bechdel test without a temporal component. The reduced model includes the 7 explanatory variables: Average Rating and the genre variables associated with action, adventure, crime, romance, thriller, and "other genres". Compared to the LASSO regression on the full variable set, the genre variable for horror is eliminated. The remaining coefficient estimates, and interpretations remain approximately the same.

4.2 Inference of movie characteristics on mostly or fully passing the Bechdel test

In our dataset, about 58% of the movies fully pass the Bechdel test and about 10% of the movies have a Bechdel rating of 2 (have two named women who talk to each other, but they talk about a man). These additional movies we consider to "mostly pass" the Bechdel test. We next fit another LASSO regression model but with a binary response variable of movies that mostly or fully pass the test vs. movies that do not (i.e., rating of 0 or 1 vs. rating of 2 or 3). The results are nearly identical to that of the LASSO regression with not passing vs. fully passing as the response variable. The LASSO method reduces the model to having the same explanatory variables and similar coefficient estimates. The only meaningful difference is that the odds of a romance movie mostly passing or fully passing the Bechdel test is about 75% higher than a non-romance movie. This is much larger than the previous findings of a 30% increase in odds for romance movies fully passing the Bechdel test.

One limitation of logistic regression is that it does not extend to multi-class explanatory variable scenarios well. Consequently, with the logistic regression and LASSO regression methods in this section, we are limited to modeling the probability of passing the test, or the probability of mostly passing the test. Because the response variable rating contains more information than just passing or failing, we use the random forest method to extend our analysis to a multi-class scenario and fully utilize the information in our response variable.

4.3 Inference of movie characteristics on the true Bechdel test rating

We now try to obtain more insight into the relationships between the actual Bechdel test rating (0, 1, 2, and 3) and the explanatory variables. Random Forest is robust in predictive modeling with variable ranking by importance. Since the main goal of our study is to model the true generation of our data rather than make a prediction, we expect Random Forest to be useful for ranking variable importance and to compare with the performance of our LASSO model. After obtaining the variable selection and explanatory variables that rank results from both LASSO and Random Forest, we compare the variables and the top ranked variables to see if there are any similar results.

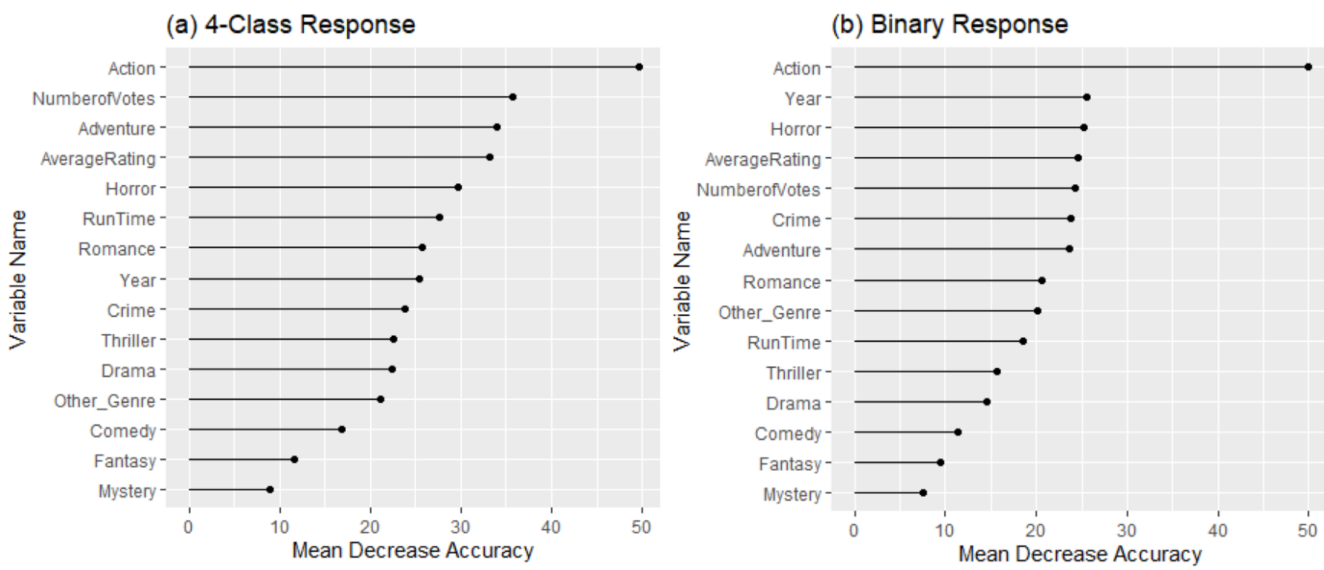

Figure 4: The importance of explanatory variables in the Random Forest models. These Mean Decrease Accuracy plots express how much accuracy the model losses by excluding each variable. The more the accuracy suffers, the more important the variable is for the successful classification. The variables are presented in descending importance. Plot (a) is for the Random Forest model that uses a 4-class response and plot (b) is for the Random Forest model that uses a binary response.

From the importance plot shown in Figure 4 (a), the most important explanatory variable is the genre Action, followed by the Number of Votes, Adventure, and Average Rating.

We also fit the Random Forest using the binary response of whether movies pass or fail the Bechdel test (the same response variable we use in the LASSO regression). The variable importance plot is shown in Figure 4(b). In this situation, Action movies, Year, and Horror movies are the top 3 most important variables in this model. In both plots, Action is the most important variable. We also see gaps in importance rank between Action and Number of Votes in Figure 4(a) and Action and Year in Figure 4(b) which means that Action plays a more significantly important role in the Bechdel test rating than other variables. Comedy, Fantasy, and Mystery are the 3 least important variables.

We then fit the same explanatory variables but without Year using the Random Forest method to infer the ranking of variables without a temporal component. The results are similar to the model with Year included: Action is the most important variable, followed by Horror, Adventure, and Average Rating. Again, Comedy, Fantasy, and Mystery are the 3 least important variables.

5 Discussion and Limitations

In our analysis, we utilize two common statistical methods, penalized logistic regression (via LASSO) which is suitable for our goal of modeling the probability of passing the Bechdel test, and random forest with variable ranking by importance. The interpretations of the results differ, but we can consider them as two distinctive dimensions: explanatory power and predictive accuracy. In penalized logistic regression, the relationships between the probability of passing the Bechdel test and the movie characteristics are explained by the magnitude and sign of the regression coefficient estimates. In random forests, variable importance is not a measure of effect size but rather its contribution to predictive performance. In random forests, a variable can be important due to the way it interacts with other variables or due to the way it separates movies and their Bechdel test rating. Although we are not necessarily interested in predicting Bechdel test results with our random forest model, we can still use the results in an inferential manner. A benefit of this pairing of methods is that the random forest model can suggest improvements to our existing inferential models.

The primary results of the penalized logistic regression indicate that romance movies and release year are associated with an increased probability of mostly and fully passing the Bechdel test. Conversely action, crime, and “other” genres, as well as a high average IMDb rating are associated with a lower probability of mostly and fully passing the Bechdel test. Generally, the results of the random forest analyses agree with the logistic regression results (Table 3). In the 4-class setting with random forest, we identify the top 5 most important variables for determining the Bechdel rating, in order of most to least important, as action, number of votes, adventure, average IMDb rating, and horror. Notably, romance, “other genres”, and release year appear less important than in the penalized logistic regression and the number of votes appears to be more important. In the binary response setting for the random forest analysis, the most important variables are action, year, horror, average rating, and number of votes. These results align more with the results from the penalized logistic regression but still romance and “other genres” are not in the top 5 most important variables.

Another similarity between the penalized logistic regression and random forest analyses is that the genre variables for comedy, drama, fantasy, and thriller are not strongly associated with the probability of passing the Bechdel test or important for classifying the Bechdel test rating.

Table 3. Odds ratios of fully passing the Bechdel test by LASSO Regression and Variable Importance Ranking by Random Forest for the Variables Selected

Explanatory Variable	**Ranking of Variable in Random Forest**	**Odds ratio of fully passing the Bechdel test**
Action	1	0.545
Year	2	1.009
Horror	3	1.050
Average Rating	4	0.879
Crime	6	0.692
Adventure	7	0.964
Romance	8	1.304
Other Genre	9	0.866
Thriller	10	0.956

Clearly, specific genres (horror, romance, action, and adventure) are strongly associated with the probability of passing the Bechdel test, although in different directions. This suggests a lack of female representation in action and adventure movies and more equal representation in romance and horror films. We think this aligns with a general notion that romance movies usually involve both men and women, as do horror films. However, action and adventure movies seem to be dominated by male actors.

Our analysis highlights that higher Bechdel test ratings are observed in movies with more recent release dates, which is encouraging to see. This aligns with UCLA's 2020 Hollywood Diversity Report which notes "Hollywood executives have apparently gotten the memo that diversity on screen sells, as the minority and female shares of these critical roles have increased steadily over the course of this report series". Although female representation in movies has not reached the same representation of females in the U.S. population (about 50%), it appears we are headed in the right direction.

Although the datasets from Bechdeltest.com and IMDb.com contain many movie characteristics, there are additional characteristics we would have liked to use in our analysis. These include financial information about the movie (budget, revenue, etc.), but we have not been able to identify a sufficiently large and reliable dataset to use. However, we feel that financial success is adequately captured by a combination of the variables we did include, namely IMDb rating and Number of Votes. There may also be other movie characteristics that influence the Bechdel test rating that we did not include in our study. For example, the race, ethnicity, and genders of directors or producers may be associated with the results of the Bechdel test.

Additionally, our dataset may be a biased sample of movies which affects the final conclusions. We only use movies available from both Bechdeltest.com and IMDb. The list of movies available from Bechdeltest.com may be biased as noted by a FiveThirtyEight article, “as with any film database that relies on user submissions to compile its film list, it’s in many ways a product of its user base, which in this case is feminist-leaning” (Hickey, 2014).

The most severe study limitation is with regards to how female diversity is measured. Our findings indicate that some genres like romance and horror are positively associated with the Bechdel test results, but is the female representation of good quality? For example, romance films often portray women in a stereotypical manner, e.g., mostly talking about men. Yet if they talk about something other than a man only once, then the movie would fully pass the Bechdel test. Another example comes from a 1987 horror film study where the authors found that audiences enjoy the victimization of women more than the victimization of men in horror movies (Tamborini, 1987). Perhaps these movies may pass the Bechdel test, but their quality female representation is still questionable.

6 Conclusion

In this study, we explored the gender diversity of movies based on the Bechdel test. An exploratory analysis showed an increasing proportion of movies tend to pass the Bechdel test over the past several decades. Upon a deep look with both a regression analysis and random forest analysis, we found specific genres like romance to be highly associated with passing the Bechdel test, and action to be highly associated with *not* passing the Bechdel test. This suggests some genres, not all, are associated with the Bechdel test results. Specifically, our study implied that action movies continue to lack female representation, even though overall movies are increasingly becoming more diverse. Altogether, these findings do indicate an improvement in gender diversity in the U.S. movie industry but reveal some potential underlying factors that affect gender diversity. Future studies and analyses can include additional movie characteristics, such as investment and revenue, which are not included in this study, to further study diversity in the film industry.

Citations

Bechdeltest.com (n.d.) [Data file] Retrieved on February 14, 2021 from http://bechdeltest.com/api/v1/getMoviesByTitle?title=

Hastie, T., Qian J. (2014) Glmnet vignette. Retrieved from https://web.stanford.edu/~hastie/glmnet/glmnet_alpha.html#log

Hickey, W. (2014). The dollar-and-cents case against Hollywood's exclusion of women. Retrieved from https://fivethirtyeight.com/features/the-dollar-and-cents-case-against-hollywoods-exclusion-of-women/

Hunt, D., Rámon A. C. (2020). Hollywood diversity report 2020: a tale of two Hollywood. Retrieved from https://socialsciences.ucla.edu/wp-content/uploads/2020/02/UCLA-Hollywood-Diversity-Report-2020-Film-2-6-2020.pdf

Internet Movie Database. (n.d.). title.basics.tsv and title.ratings.tsv [Data files]. Retrieved on February 14, 2021 from https://www.imdb.com/interfaces/

Rating FAQ (n.d.) Internet Movie Database. Retrieved from https://help.imdb.com/article/imdb/track-movies-tv/ratings-faq/G67Y87TFYYP6TWAV#

Shmueli, G. (2010). To Explain or to Predict? *Statistical Science. 25*(3), 289-310. doi:10.1214/10-STS330

Tamborni, R., Stiff, J., & Zillmann, D. (1987). Preference for graphic horror featuring male versus female victimization: Personality and past film viewing experiences. *Human Communication Research*, *13*(4), 529–552. https://doi.org/10.1111/j.1468-2958.1987.tb00117.x