Alison Silldorff

Computer Science One-Semester Independent Work Final Paper

Spring 2024

Advisor: Brian Kernighan

Female Representation at the Academy Awards

Contents:

Abstract

This project analyzes Academy Award nomination data to analyze female representation in Oscar-nominated films since the start of the awards in 1929. Female representation is measured using Bechdel Test scores, gender breakdown of the cast, gender breakdown of above the line crew positions, and gender breakdown of the first-listed actor for a film. The results reinforce that women have been historically underrepresented in Oscar-nominated films. In particular, women comprise about half as much of the casts of Oscar-nominated films as men, and this ratio has changed very little over each year of the Academy Awards. It is also found that

films directed by women have better female representation (according to these metrics) than films directed by men.

1. Introduction

In this project, I sought to quantify female representation in Academy Award-nominated films, both in terms of on-screen and off-screen representation. I wanted to use questions and problems posed in film scholarship, particularly feminist film scholarship, and find ways to express these problems through a quantitative lens. Although we see increasing numbers of female-centric stories and female filmmakers in the spotlight, both the historical and current states of gender equality in the film industry are far from ideal. This project serves as an opportunity to give greater voice to these stark gender inequalities, and provide data-based analyses to further probe how the male-domination of the film industry manifests in various quantitative measures of representation.

Since their inception in 1929, the Academy Awards (also known as the Oscars) have been one of the most influential entertainment award ceremonies in the world. An Academy Award nomination leads not only to financial benefits for the film's cast and crew, but also permanently increases the work's cultural weight. Given these impacts, who is nominated for them, as well as what is being depicted in these films, also carries great weight.

As an example of this gender gap at the Oscars, out of the 491 nominations for Best Director across all 96 years of the Academy Awards, only 9 of those nominations have gone to women. That is less than 2%. Beyond this evident exclusion, women in the film industry have faced poor treatment since the silent film era. Today, not yet a decade after the #MeToo movement, which was re-ignited in 2017 by the repeated sexual assaults of women by

Hollywood mogul Harvey Weinstein, mistreatment of women in the film industry is a more visible issue. However, it is unclear whether this movement will eventually bring about change to the male-dominated power structure of the film industry.

Given the nebulous definition of what actually constitutes "good" female representation, or what true gender equality in the film industry might look like, in this project I also wanted to try several different measurements of representation, including Bechdel Test data, gender breakdown of casts of Oscar nominated films, and gender breakdown of above the line versus below the line crew positions on Oscar-nominated films.

See Appendix A for background on the Academy Awards

2. Motivation

Although gender equality is improving in the film industry, it is still far from providing equal opportunity and representation to women. Some efforts are being made to remedy the industry's historical maltreatment and poor representation of women, but this change has only happened because of continual pushes from filmmakers to reach these new standards. Furthermore, arguments made regarding the lack of representation are often discredited, making it ever more important to continue to examine where these inequalities lie. Feminist film scholar Janet McCabe states in her book *Feminist Film Studies* [29]:

> feminist film theory can often appear hysterical as a result of being constructed from methodologies that negate female subjectivity and consign women's stories to the category of fantasy… Given the current resistance toward theory and the suspect accorded to the term 'feminism' it is not too surprising that women still face difficulties when attempting to speak under the sway of patriarchy. The more feminist film theory gains respectability within the academy, the more its methodological

differences/difficulties are revealed as problems of legitimacy and credibility in speaking from inside the discipline (119-120)

This provides insight into the importance of pushing back against patriarchal dismissal of feminist film theory and efforts to create better representation for women in cinema. This project is motivated by a desire to give further credibility to those who use their voices to speak against gender inequality in the film industry, but are met with weak solutions and minimal change.

3. Related Work

3.1 Quantitative Studies of Representation in Film

There are several studies that approach the question of female representation in a quantitative way. Most similar is the research by the Center for the Study of Women in Television and Film at San Diego State University. From this center, *The Celluloid Ceiling* annually tracks the number of women working in production positions in the film industry for top grossing films. “It’s a Man’s (Celluloid) World,” is another annual report which analyzes on-screen female representation for the year’s top grossing releases [32]. These have happened annually since 2002. In their analysis of crew positions, they include Director, Writer, Executive Producer, Producer, Editor, and Cinematographer, and look at how many women per year hold those roles for top grossing films [33]. To analyze on-screen representation, the report for 2023 quantified the number of female protagonists, major characters, and speaking roles. They also measure those three categories in terms of race and age. Some of the most creative measures I found in this study were the marital status and occupation of female characters [34].

The following studies more generally approach the question of representation in film in an interesting computational way: A 2019 study at the Claremont Colleges studies the gender

wage gaps in the film industry, concluding that the impact of gender on an actor's salary is greater than any other factor [35]. The study notes that, due to the large influence of the film industry in the United States, poor representation in Hollywood sets a poor standard for treatment and representation of women across the country. 2020 study "Measuring Female Representation and Impact in Films Over Time" again takes an economic perspective, examining how female representation affects a movie's monetary success. This study's quantification of female representation uses Bechdel Test scores and the percentage of women in the cast, and finds that female screenwriters are particularly instrumental to better representation. However, there does not seem to be demand from moviegoers for better female representation because movies with poor representation continue to perform well in the box office [36]. "Identifying Missing Component in the Bechdel Test Using Principal Component Analysis Method" points out some flaws of the Bechdel Test, and adds several parameters and uses principal component analysis to measure the content of the female dialogue [37]. "Quantifying the Global Film Festival Circuit: Networks, Diversity, and Public Value Creation" models the international film festival circuit as a network connected by overlapping films, and studies gender diversity, thematic diversity, geographic diversity, and linguistic diversity at these festivals [38].

3.2 Humanistic Studies of Female Film Representation

There is a large body of research surrounding qualitative and humanistic measures of female representation, and I will outline some of the key pieces of scholarship used for this paper. *Feminist Film Theory* by Anneke Smelik outlines key components of feminist film theory, noting the cultural impact of film on views of women throughout history, and how films have been shaped by the feminist movement [39]. It provides strong backing on why this

representation is important, and what constitutes good representation. “Film History and the Two Presents of Feminist Film Theory” goes deeper into women’s role in the film industry over time, reconstructing the narrative of the extent of their involvement [40], and Laura Mulvey’s “Visual Pleasure and Narrative Cinema” discusses the visual objectification of women in cinema and the impact of the male gaze [41]. “What ‘The Bechdel Test’ Doesn’t Tell Us: Examining Women’s Verbal and Vocal (Dis)Empowerment in Cinema” points out specific flaws of the Bechdel Test, especially relating to which women are represented, and the many limitations of this test [42]. “Unmanageable Inequalities: Sexism in the Film Industry” explains gender inequality in the film industry in New Zealand, in particular for above the line and below the line crew positions [43].

4. Approach

My approach for this project is to use motivations from feminist film scholarship to guide data analysis of female representation of Oscar-nominated films. Because the meaning of “good” female representation is not clearly or uniformly defined, I wanted to use a blend of qualitative and quantitative research methods in order to draw better-informed conclusions. Qualitative analyses from feminist film theory can give a more zoomed-in lens on films throughout history, while data analysis can point to larger trends over time. With feminist scholarship guiding this data analysis, I hope to zoom out from specific scholarly arguments and see how they apply more broadly.

Additionally, my approach narrows its focus to Oscar-nominated films. Oscar-nominated films are a particular subset of typically American, popular, and/or critically-acclaimed films. This subset, as opposed to the studies at the Center for the Study of Women in Television & Film

[32] [33] [34] which instead analyze top grossing films, allows for more specific conclusions regarding female representation in these films.

5. Data Collection

5. 1 Academy Award Nominees and Winners Data

To collect data on Academy Award wins and nominees, I used WikiData [1]. I made a call to their query service [2] in SPARQL, originally retrieving the award edition (which ceremony it was presented at); the award label; the nominated film and its director, IMDB id [5], and TMDB id [4]; and the person who won and their gender [3]. I adapted a sample query from WikiData's query service guide [7] and well as a sample query from OpenStackExchange [6], then exported the query results into a JSON file. I did the same for Oscar winners.

From here, I began examining this data in OpenRefine to confirm where I was missing data, comparing the true number of winners for each category over time with the number of winners in my dataset for each category. In this process, I discovered peculiarities of some award categories over time, as well as common missing data points, from which I decided to narrow my research to only the following categories:

Best Picture	Best Actor	Best Supporting Actor
Best Director	Best Actress	Best Supporting Actress

Going forward, when I refer to "Oscar nominees" or "Academy Award Nominees," I am referring to these six categories alone. This decision was supported not only by the cleaner nature

of the data in these categories, but also by the fact that these categories are typically the ones that receive the most attention.

To clean the data both for nominees and winners in these categories, I exported my OpenRefine projects to Google Sheets. This allowed me to easily insert and delete rows, which is not possible in OpenRefine. To clean this data, I manually went through the Academy Awards Database [8] and my data year-by-year, adding any rows that were missing in both my Oscar Winners and Oscar Nominees datasets. In cleaning, I made the following decisions about how I was handling some less-standard entries:

1. For Best Director nominations/wins shared by more than one person (e.g. Daniel Kwon and Daniel Schienert for *Everything Everywhere All at Once* in 2023), the nomination or win will be represented by a row per person.
2. For Oscar winners: For ties in any category, there will be a row for each person who won (e.g. Barbara Streisand and Katharine Hepburn for Best Actress in 1969)
3. For the first three Academy Awards ceremonies, Best Actor and Best Actress nominations "could honor work in one or more films" [8]. In other words, Best Actor nominations were each associated with a single actor, but multiple films. In my dataset, I count these nominations with a row per film.
4. At the 1st Academy Awards, Best Director was divided into categories of Comedy Picture and Drama Picture. In my dataset, I have renamed these awards to both just be "Academy Award for Best Director", omitting the distinction between comedy and drama.

Upon adding in all missing rows according to the above principles, I exported my Google Sheets for wins and nominations, respectively, into CSV files, and back into OpenRefine. The dataset of nominations contains 2,800 rows, and 1,402 distinct nominated films.

5.2 TMDB Credits Data

The Movie Database (TMDB) is a movie and television database containing 919,353 movies [9]. TMDB gives free access to its API [10], and houses a support forum to answer questions about the API [11]. Data on TMDB is 100% user contributed [18]. I decided to use TMDB rather than the Internet Movie Database (IMDb) because TMDB has a more extensive free API than IMDb, where many features of the API are only available for use at a lofty price. In particular, TMDB's free API contains access to credits data for all films.

An API call to get a movie's credits is formatted:

`https://api.themoviedb.org/3/movie/{tmdbid}/credits?api_key={api_key}`

where {api_key} is your personal API key, and {tmdbid} is the film's TMDB ID. A TMDB ID is the unique identifier for a film on the TMDB site. TMDB IDs are distinct from IMDb IDs, though both are used in this project.

The above call returns a JSON object with entries "id", "cast", and "crew". "cast" and "crew" each contain a list of every cast and crew member, respectively, that TMDB has associated with the film. Among the information in each person's entry is their gender, their character name if they are in the cast, and their role on the production if they are on the crew. Gender is listed as a one digit number, with the following meanings [12]:

0: not set / not specified
1: female
2: male
3: non-binary

I would like to note here the dataset's seeming confusion between gender and sex. The true meanings of "male" and "female" refer to sex, while "non-binary" is a gender identity. Below, in graphs that use this credits data, I will compare between the representation of women, men, and non-binary people, making all three categories refer to gender identity rather than sex, as I believe the data intends.

There is not a way to get credits data in bulk, so I iterated through each unique film in my dataset of Oscar nominees, and, using the entries TMDB ID, made a call to the API to retrieve credits data and exported it to a JSON file. No changes were made to the API's JSON output for each call, aside from my output file using utf-8 encoding.

Examining the data, it seemed like film casts were listed in some order of most important to least important, though it was unclear whether this corresponded to the actual order of the credits in a film or TMDB's judgment of importance. I found a few Support Forum questions from contributors regarding cast ordering that indicated the suggested way to order the cast [14] [15]. TMDB's Contribution Bible has a page that gives the following guidelines to uploading cast information [13]:

> Listing the main actors of a movie first is recommended. The main cast should not be sorted by credit order, order of appearance or in alphabetical order.
>
> The major roles should always be credited before the small parts, no matter the celebrity status. e.g. If Danny Trejo only has one scene in a movie, he should not be credited before actors with bigger parts.
>
> When all the main parts are listed, we recommend using the credits order instead of a random order for the remaining credits.
>
> (uncredited) credits should always be last.

This shows the guidelines that contributors are meant to follow, which includes a mix of listing actors in terms of importance, listing them in credit order, and in alphabetical order. However, I

did not find resources on TMDB or elsewhere that indicate a standard way that credits are consistently ordered in the database.

For the ordering of the crews, I similarly found very little information documenting TMDB's procedure. Unlike the cast, on TMDB's Contribution Bible page for the crew, there was no information on the suggested ordering [16]. On the Support Forum, I found a thread whose reply indicated that there is not any standardized ordering of the crew [19].

To gain a better understanding of how crews were currently ordered, I looked at a few random API calls. For *Barbie* (2023), the order of the crew is: Executive Producer, Producers, then Director. For *Oppenheimer*, however, the order is: Director, then Producers. Between these two examples, it is unclear if this difference in ordering has any significant indication.

For *Creed* (2015), the API call lists the Music Supervisor first in the credits ordering, followed by the Set Decorator, then Casting, then Director. This seems to have even less significance than the previous two examples, as Set Decorator and Music Supervisor are positions much lower down in a film's hierarchy than the Director [17].

Taking these three examples, which are all films from the past 10 years, I felt confident that the rest of the dataset would also lack significant consistency in the ordering of crew members. Even for films where it seems the most important positions are listed first, such as *Barbie* and *Oppenheimer*, there is neither consistency in what TMDB deems to be the most important role, nor documentation anywhere clarifying if the ordering gives some other indication, such as billing order in the film's ending credits.

From this, I concluded that I could somewhat reliably use the ordering of cast credits to indicate importance, but I could use the ordering of crew credits to draw any conclusions.

5.3 Bechdel Test Data

The Bechdel Test is a test created by cartoonist Alison Bechdel in a comic strip in 1985. A movie passes if it meets the following three criteria: (1) The film contains at least two named female characters (2) who have a conversation with each other (3) about something other than a man. The test has become "a convenient shorthand for measuring representation in movies" [20].

bechdeltest.com maintains a crowdsourced database of over 10,200 films and their Bechdel Test score. On their statistics page, they cite 57.1% of all of the films in their database passing the test [22]. Note that, rather than using a binary metric of "pass/fail," bechdeltest.com rates films on a score of 0-3. A 3 indicates that a film passed all 3 criteria listed above (thus the film passes), and a 0 indicates that 0 of the above criteria are met. 0, 1, and 2 are all failing scores.

The site's API allows you to make calls based on IMDb ID or title, or to fetch all movies and their scores [21]. For my project, I used a curl command from the command line to retrieve all Bechdel Test scores and their corresponding IMDb IDs. I put these into a JSON file, and created an OpenRefine project from this JSON file. This allowed me to then add a "bechdeltest_rating" column to my oscar nominees and oscar wins OpenRefine projects.

To do so, I first edited the IMDb IDs from the Bechdel Test data to have a prefix of "tt," which indicates a film. Because the entire database is films, they do not include the prefix, but my data from WikiData does. Once this prefix was added, I was able to use the "add a column based on this column" feature within both the Oscar Nominees and Oscar Wins projects, and then add a bechdeltest_rating column to those projects using the matching IMDb IDs.

Out of my 1,402 distinct nominated films from 6 nomination categories, 534 distinct nominated films were missing a Bechdel Test score, simply because there was not data for these

films on bechdeltest.com . About 21% of these missing films were nominated in the years 1929-1939 alone, while only 33 films (~6%) nominated between 2000-2024 were missing a score.

With the intention of filling in some of these missing entries myself, I exported the list of nominated films missing a score to a Google Sheet. I then organized this sheet by decade, and shared the sheet with some friends to see if they knew the scores of any of these films. From this, I received about 5 additional scores, which I manually updated in my OpenRefine project. If I had more time to work on this project, I would have liked to source more of these missing scores. However, making this spreadsheet was a start to being able to do that.

Because this data is crowdsourced, there are certainly some erroneous scores. For the most part, when I looked at the list of films I knew and their scores, they seemed to align with my knowledge of the film. I noticed a few incorrect entries myself, but did not go through the entire dataset manually altering entries based on my own knowledge. I decided to stick with bechdeltest.com ‘s ratings, aside fro the scores I added manually as mentioned above. With more time, I would like to more rigorously check existing entries to ensure they have accurate scores.

6. Data Processing and Evaluation

6.1 Female Representation Overall

First, I would like to show the findings over all of my data for several different measures of female representation. Figures 1-8 display these four measures across all distinct nominated films:

1. Bechdel Test Scores
2. Women listed first in the cast
3. Percent of women in the cast

4. Percent of women in Above the Line crew positions (see below for definition)

6.1.1 Bechdel Test Scores

As stated above, Bechdel Test scores are on a scale of 0-3, with a 3 being a passing score. This test is meant to give one angle on the *quality* of female representation in a film, as it measures what the female characters do in a movie, and whether their function relates to anything other than the film's male characters.

From Figure 2, it can be seen that, although there seems to be an upward trend in the percent of films that pass the Bechdel Test, the amount of missing data before ~1980 makes it difficult to determine if there is a true upward trend in Bechdel Test passes. Based on the typical male-centric content of films in the early-mid 20th Century, I would hypothesize that, if collected, this data would contain mostly failing Bechdel Test scores.

Figure 1 shows us that, from the Bechdel Test data that we *do* have, 36.1% pass the test, and 26.9% fail. As mentioned above, about 21% of the missing Bechdel Test scores come from the years 1929-1939 alone, which would lead me to hypothesize that finding these and other scores from early Oscar-nominated films would shift this ratio to have a larger percentage of failing scores.

Fig 1: Bechdel Test Scores of all Oscar-Nominated Films

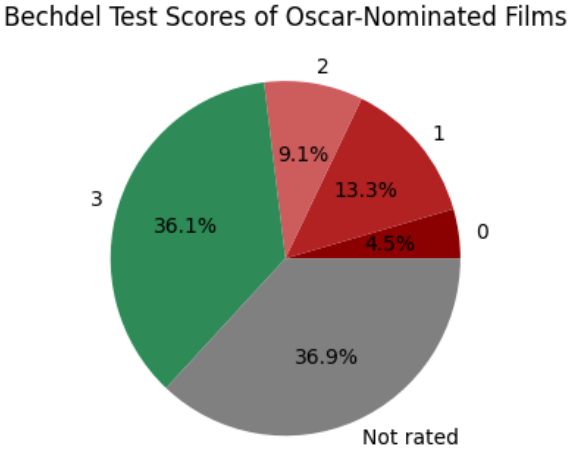
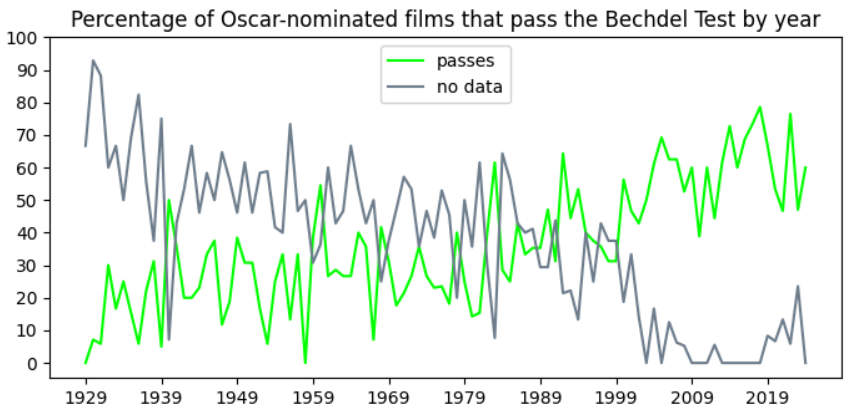


Fig 2: Percentage of Oscar-nominated films that pass the Bechdel Test by year



Looking at the first-listed cast member in the cast ordering from TMDB can give us an estimate of the gender of a film's protagonist. This lends to examining the function of women in a film, and whether a film is female-centric.

Figure 3 shows firstly that we have gender data for 99.9% of the first listed actors in all of our Oscar-nominated films. Secondly, it shows that men are listed first almost twice as frequently as women, and non-binary people are listed first for only 0.1% of these films.

Figure 4 shows a higher average number of women listed first from the years 1929-1970 than from 1970-2010. The average for 1929-1970 is 39.27%, and the average for 1970-2010 is 31.87%. In Figure 4 we also see that many consecutive years have a large difference in this percent

Fig 3: Gender Breakdown of the First Listed Actor in Oscar-Nominated Films

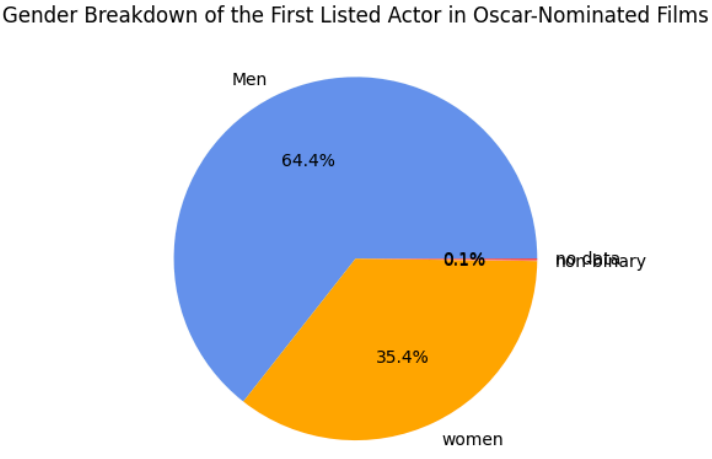
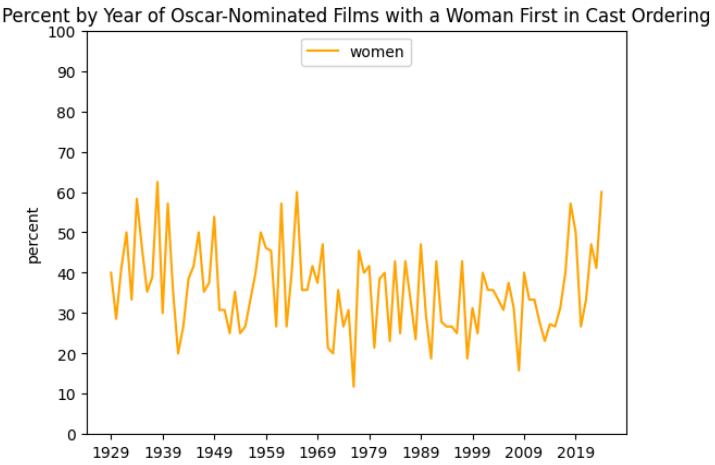


Fig 4: Percent by Year of Oscar-Nominated Films with a Woman First in the Cast Ordering



6.1.3 Percent of women in the cast

Using the same cast ordering from TMDB, we can measure the percent of women in each cast. For this measure, I have chosen to only use up to the first 35 cast members for each film. I

did this because, as the cast list goes on, TMDB also lists cast members such as extras and stunt doubles, many of whom do not have gender data. Therefore, by capping the cast at 35 people (the median cast size for all nominated films in my dataset), I hope to isolate only main members of the cast, and draw conclusions on gender breakdown from this. Additionally, for cast data, I have chosen to exclude from all graphs cast members whose gender identity data is missing. This makes this assumption that missing entries would maintain roughly the same ratio between genders.

Figure 5 shows the gender breakdown of the first 35 cast members (including actors whose gender we are missing), showing more than twice as many men as women in the casts of Oscar-nominated films.

Figure 6 shows that, over time, these gender breakdowns have had very little change since the 1930s, and the gap between the number of men and number of women in a cast has only recently begun to shrink. Additionally, the number of non-binary actors has only recently begun to increase from 0.

Fig 5: Cast Gender Breakdown of Oscar-Nominated Films

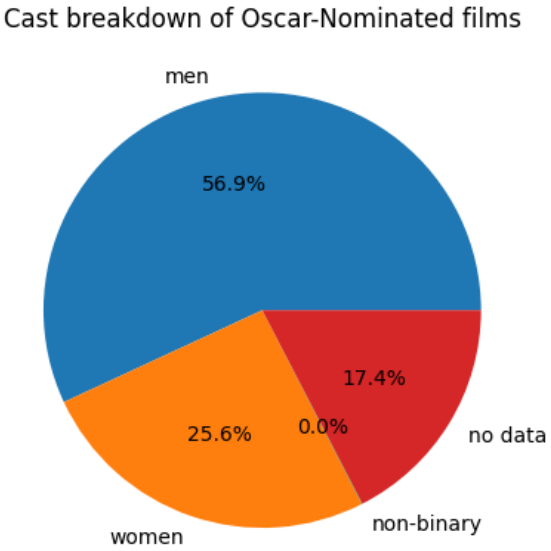
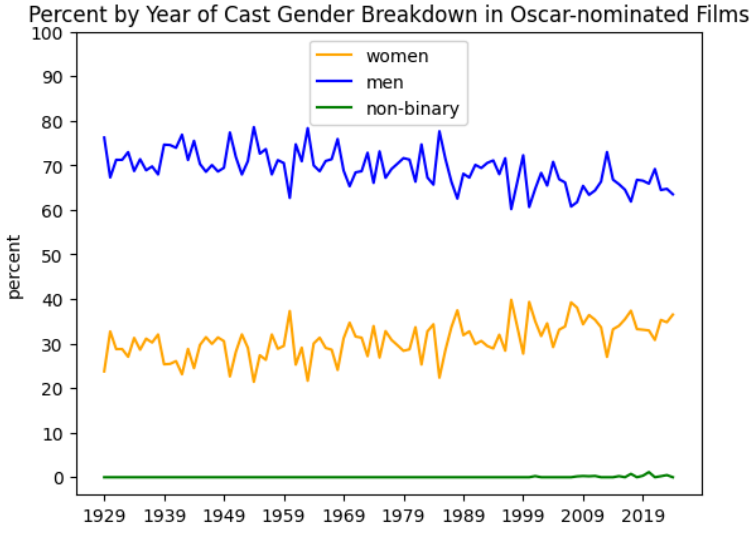


Fig 6: Percent by Year of Cast Gender Breakdown in Oscar-Nominated Films



6.1.4 Percent of women in Above the Line crew positions

Above the Line (ATL) crew positions is a term that literally comes from film budgets, and which individuals are listed "above the line" versus "below the line" on those budgets. ATL positions control the creative vision for a film, and are also generally present for the entire film's process. Which positions fall into this category is slightly different from film to film, but for my study the following positions are considered ATL: Producer, Executive Producer, Director, Director of Photography, Screenwriter. Below the Line (BTL) includes all other positions [44] [45].

As a final measure of female representation, in this study I measure how many women hold ATL positions in Oscar-nominated film crews. This can give an indication of whether women are holding creative power in these influential films.

Figure 7 gives insight into ATL crew positions (excluding missing data points), showing about 10 men for every woman in an ATL position on Oscar-nominated films. Figure 8, which also only includes ATL positions for which we have gender data, shows that the number of women in these positions has been steadily increasing since the 1980s. Nevertheless, the maximum for this ratio in the past five years is still only 30.0% of ATL positions held by women.

Fig 7: Gender Breakdown of ATL Positions in Oscar-Nominated Films

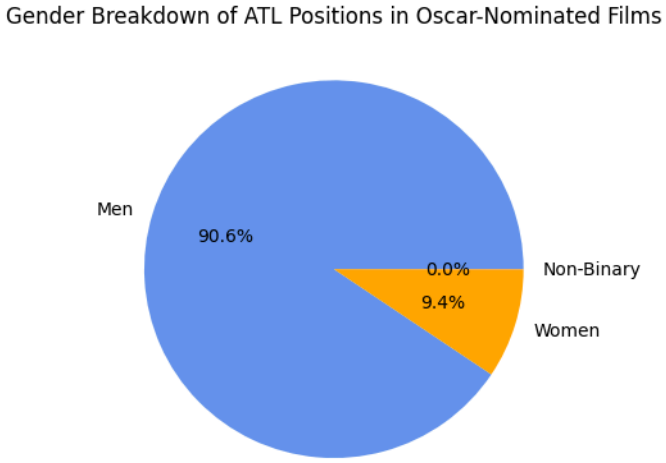
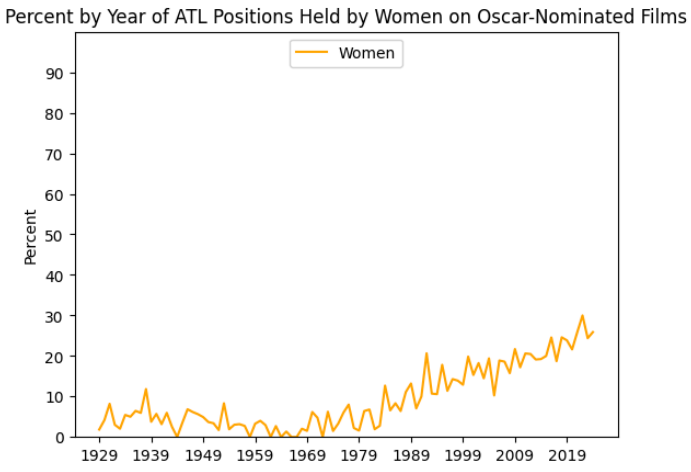


Fig 8: Percent by Year of ATL Positions Held by Women on Oscar-Nominated



6.2 How do different categories compare in terms of the gender of the first-listed actor?

To discover further possible insights in my data, I read through feminist film scholarship to motivate the next step of data processing. From *Hollywood's Second Sex* by Aubrey Malone [46]:

> "There are still quality parts being written for women today but not as many as for men and that will probably never change…The disquieting reality is that women will probably continue to be 'Hollywood's second sex' for as long as movies are made. They'll continue to play demeaning roles like the 'best friend' or the 'glamorous granny' when they reach a certain age and continue to be aware that in the Hollywood jungle they will have to do everything twice as well as men to survive" (7)

The above section of *Hollywood's Second Sex* made me interested in how different performances and stories were being valued. Namely, I wanted to view this by contrasting the gender of the first-listed actor across different nomination categories.

For this section, I opted to include the entries whose gender data was missing. Also note that when I use the term "first listed actor" I mean "actor" in its a gender-neutral meaning.

Films nominated for Best Picture, Best Director, and Best Supporting Actor have similar demographics for the first-listed cast member:

> **For Best Picture:** 71.7% men, 27.8% women, 0.3% non-binary, 0.2% missing data.
> **For Best Director:** 72.5% men, 26.9% women, 0.4% non-binary, 0.2% missing data
> **For Best Supporting Actor:** 77.9% men, 21.8% women, 0.2% non-binary, 0.0% missing data

See Appendix B for visualizations of these data

All of these categories are more male-dominated than our overall:

> (From Fig 5) **Overall**: 56.9% men, 25.6% women, 0% non-binary, 17.4% no data

The category of Best Supporting Actress shows similar ratios to the Overall:

Best Supporting Actress: 58.0% men, 41.7% women, 0.2% non-binary, 0.0% no data

If we assume that the "first listed actor" on TMDB gives some indication of the protagonist of a film, we can conclude that Best Picture and Best Director nominations are given a large majority of the time to male-centric stories. For me, this raises questions about whose stories are considered impactful enough to drive a Best Picture nomination.

For Best Actor and Best Actress:

Fig 9: Gender Breakdown of the First Listed Actor in Best Actor Nominees

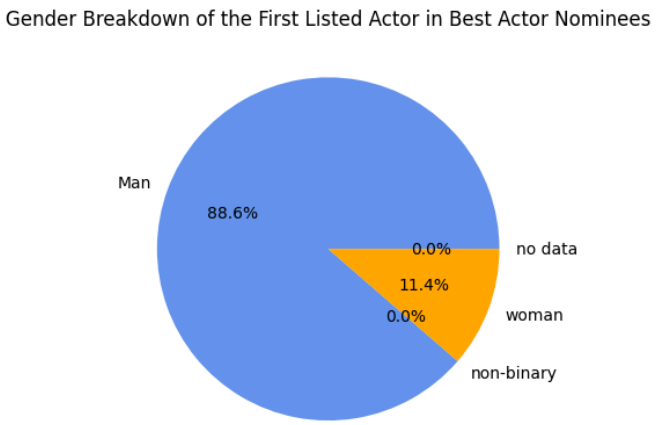
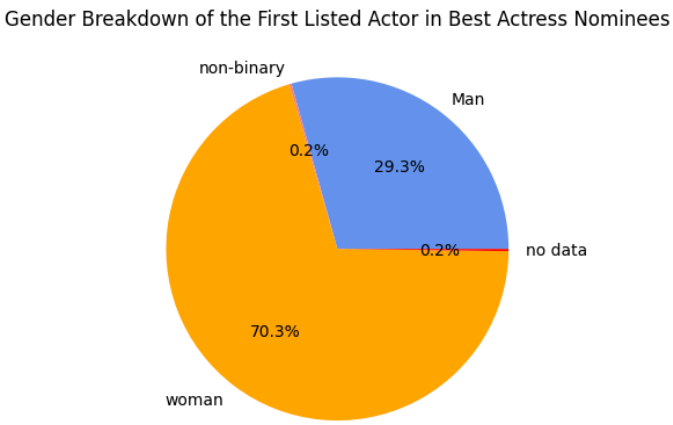


Fig 10: Gender Breakdown of the First Listed Actor in Best Actress Nominees



As shown by Figure 9 and 10, Best Actor/Actress nominations tend to correspond to the gender of the first-listed actor in a film. However, for Best Actress-nominated films, women are listed first 18% less of the time than men are for Best Actor-nominated films.

Women are listed first in Best Actress-nominated films approximately as often as men are listed first in Best Picture-nominated films. It could be argued, then, that the category of Best

Actress gives attention to more female-centric stories than would be given attention without the category. This attention then, can generate further monetary success for a film and those working on the film, which, in the case of Best Actress nominees, seems to often include a high-billed woman in the cast. This is not a definitive causal relationship, but rather a speculative chain of events based on the monetary benefits of Academy Awards and the difference in gender representation across these Oscar categories.

6.3 How does a director's gender correlate to measures of female representation?

In *Genre, Authorship and Contemporary Women Filmmakers*, author Katarzyna Paszkiewicz describes the controversy surrounding Kathryn Bigelow's Best Director win for *The Hurt Locker* in 2010. This was the first time a woman had ever won the award, yet Bigelow was criticized by the media, particularly from feminist circles, because her film was very male-centric [47]. This controversy over what stories female directors *ought* to tell led to to wonder how films directed by men and women compare in terms of female representation. Do many women directors reject expectations like Bigelow, or do women directors tend to correspond to better female representation?

Comparing the Bechdel Test scores for films directed by a woman versus by a man, we can see that films directed by a woman pass (score of 3) almost twice as often as films directed by men, as seen in Figures 11 and 12.

For films directed by women, there is a larger percentage of ATL crew that are women. It must be remembered that this does not reflect the current state of film crews employed by each gender of director, but rather the historical correlation between a director's gender and the gender identities of those ATL on the same film. See Figure 13 and 14.

Fig 11: Bechdel Test Scores of Oscar-Nominated Films Directed by a Woman

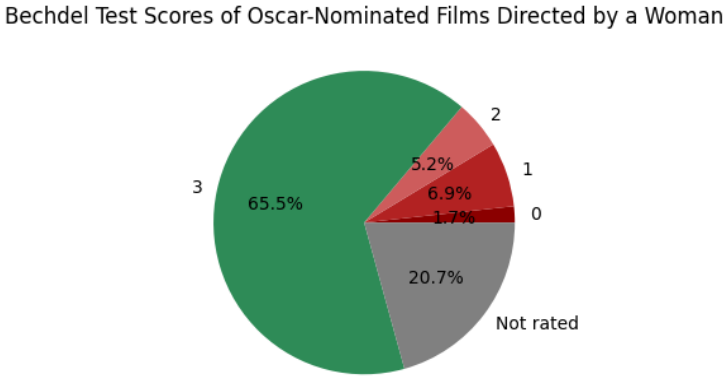


Fig 12: Bechdel Test Scores of Oscar-Nominated Films Directed by a Man

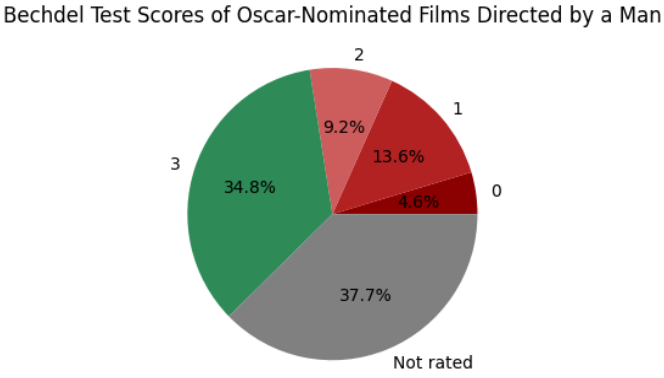


Figure 13: Gender Breakdown of ATL Positions in Oscar-Nominated Films Directed by a Woman

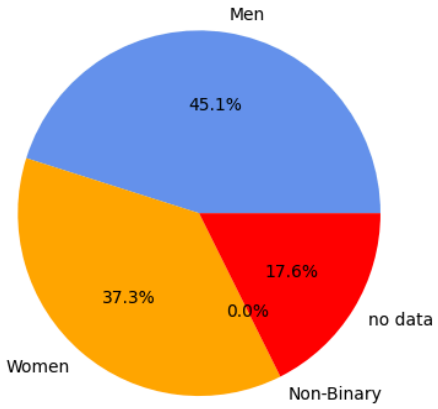
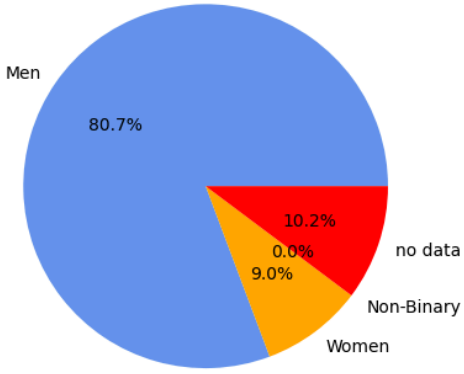


Figure 14: Gender Breakdown of ATL Positions in Oscar-Nominated Films Directed by a Man



For films directed by women, Oscar-nominated casts have ben comprised of 12% more women. This excludes missing data, and again cuts off the cast at 35 individuals as mentioned above. This can be seen in figures 15 and 16 below:

Figure 15: Cast Gender Breakdown of Oscar-Nominated Films Directed by Women

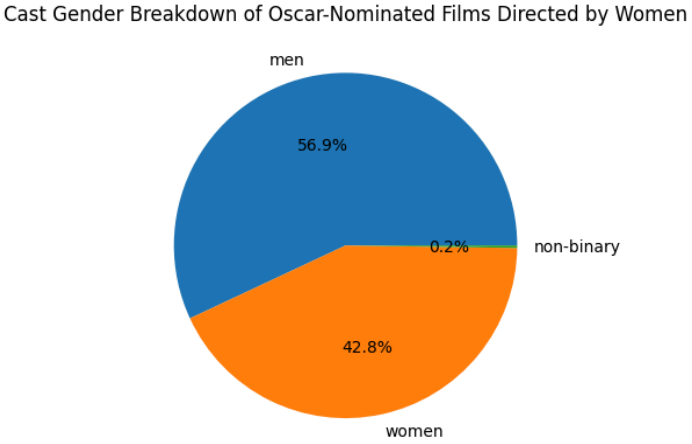
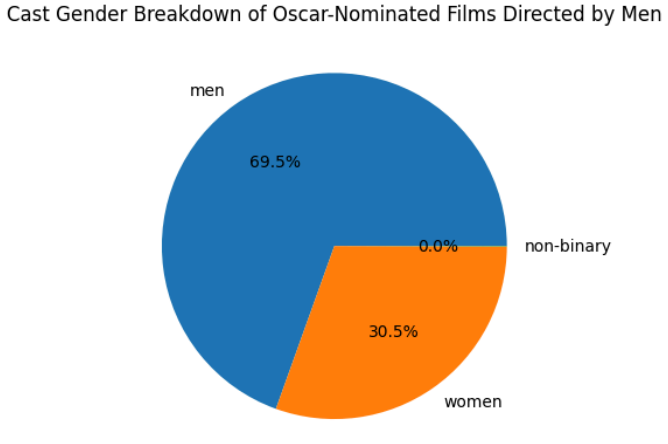


Figure 16: Cast Gender Breakdown of Oscar-Nominated Films Directed by Men



Films directed by a woman have a woman listed first in the cast almost twice as often as in a film directed by a man, as seen below in Figures 17 and 18:

Figure 17: Gender Breakdown of the First Listed Actor in Oscar-Nominated Films Directed by Women

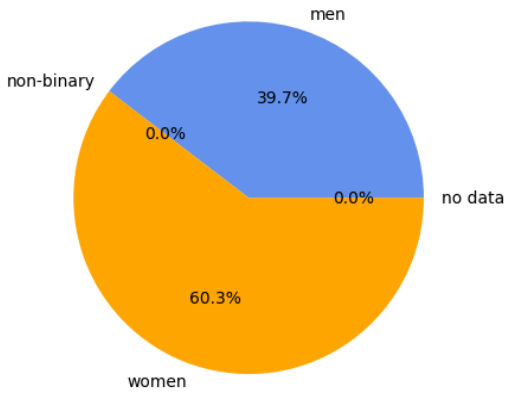
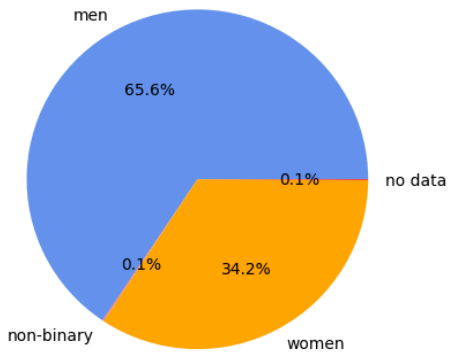


Figure 18: Gender Breakdown of the First Listed Actor in Oscar-Nominated Films Directed by Men



In all four of the above measures of female representation, the presence of a female director in Oscar-nominated films has historically corresponded to better female representation according to these metrics.

7. Conclusions

In this project I sought to use motivations from feminist film theory to guide the quantification of female representation in Oscar-nominated films. Looking at all Oscar-nominated films showed that men comprise approximately twice as much of Oscar-nominated casts, and appear almost doubly as often as the first-listed actor in TMDB's ordering. Furthermore, the gender breakdown of Oscar-nominated casts has had almost no change since the 1930s, meaning that women have not only been represented less on-screen, but fewer female actors have been employed by these high-profile films. Looking at above the line (ATL) film crew positions, men hold almost ten times as many of these positions as women across the body of Oscar-nominated films. The number of women in these roles has, however, been steadily increasing since the 1980s.

Comparing across award categories, I found that for Best Picture and Best Director, about 70% of nominated films have a man as the first-listed actor, indicating (assuming the TMDB credits ordering is reliable) that the large majority of contenders for these categories are male-centric stories. The category of Best Actress was the only category where the majority of nominated films had a woman listed first in the cast. This seems logical given that the category is specifically intended to honor a leading actress in a film.

Motivated by debate within feminist film theory of the obligation of female directors to provide more female representation, I also examined correlations between the gender of the

director and the female representation in the film according to my metrics. My data show that films directed by women correlate with greater female representation in all four of my metrics across all Academy-Award nominated films.

Overall, these results support other analytical findings as well as humanistic findings regarding gender inequality in the film industry. In short, women have been historically underrepresented in Oscar-nominated films, in some categories more than others. In some ways, this representation is improving, but certain areas are slower to improve than others.

8. Future Work

In the future, I would firstly like to collect and clean more data. I would like to collect missing Bechdel Test scores, as well as confirm the accuracy of my current Bechdel Test data. For credits data, I would like to find a way to get data that more reliably lists the cast in their billing order in the on-screen credits and/or in order of importance. This would allow me to more accurately equate the "first-listed actor" with either a film's protagonist or the highest billed actor. I would like to clean crew credits data in a similar way.

I would also be interested in gaining the data mentioned in Center for the Study of Women in Television and Film at San Diego State University's report that has more detailed information regarding on-screen female representation for all top grossing films since 2002 [34]. I would, for example, be interested in looking at the most common character occupations for Best Actor versus Best Actress nominees.

Lastly, I would like to expand this research to more subsets of films, such as other award ceremonies or film festivals. It would be interesting to see how different contexts alter the female representation according to these metrics.

9. Works Cited

[1] *Academy Awards*. https://www.wikidata.org/wiki/Q19020.
[2] *Wikidata Query Service*. https://query.wikidata.org/.
[3] *Sex or Gender*. https://www.wikidata.org/wiki/Property:P21.
[4] *TMDB Movie ID*. https://www.wikidata.org/wiki/Property:P4947.
[5] *IMDb ID*. https://www.wikidata.org/wiki/Property:P345.
[6] Kralin, Stanislav. "Answer to 'Is There an API for the Oscars/Academy Awards That Lists Past Winners as Well as Current Nominees?'" *Open Data Stack Exchange*, 5 Feb. 2018, https://opendata.stackexchange.com/a/12366.
[7] *Wikidata:SPARQL Query Service/Queries/Examples*. https://www.wikidata.org/wiki/Wikidata:SPARQL_query_service/queries/examples#Academy_award_data.
[8] *Home - Academy Awards Search | Academy of Motion Picture Arts & Sciences*. https://awardsdatabase.oscars.org/. Accessed 23 Feb. 2024.
[9] *About TMDB — The Movie Database (TMDB)*. https://www.themoviedb.org/about?language=en-US. Accessed 19 Apr. 2024.
[10] "Getting Started." *The Movie Database (TMDB)*, https://developer.themoviedb.org/docs/getting-started. Accessed 14 Apr. 2024.
[11] *The Movie Database Support*. https://www.themoviedb.org/talk/category/5047958519c29526b50017d6. Accessed 19 Apr. 2024.
[12] "Details." *The Movie Database (TMDB)*, https://developer.themoviedb.org/reference/person-details. Accessed 19 Apr. 2024.
[13] *Cast - Movie Bible — The Movie Database (TMDB)*. https://www.themoviedb.org/bible/movie/59f3b16d9251414f20000003#59f73ca49251416e7100000e. Accessed 19 Apr. 2024.
[14] *Question about Cast Sorting - Talk — The Movie Database (TMDB)*. https://www.themoviedb.org/talk/576f4f26c3a368473300326a. Accessed 19 Apr. 2024.
[15] *Cast Credits Initial Sort - Talk — The Movie Database (TMDB)*. https://www.themoviedb.org/talk/5bac13d3c3a3683aa202db9f. Accessed 19 Apr. 2024.
[16] *Crew - Movie Bible — The Movie Database (TMDB)*. https://www.themoviedb.org/bible/movie/59f3b16d9251414f20000004. Accessed 19 Apr. 2024.
[17] Assemble. *The Definitive Film Crew Hierarchy Chart*. https://blog.assemble.tv/the-definitive-film-crew-hierarchy-chart. Accessed 19 Apr. 2024.
[18] *General - General Bible — The Movie Database (TMDB)*. https://www.themoviedb.org/bible/general#59f793929251413e93000014. Accessed 19 Apr. 2024.
[19] *Order for Crew Listings (e.g. Writer) - Talk — The Movie Database (TMDB)*. https://www.themoviedb.org/talk/567c4c5d9251417def003f63. Accessed 19 Apr. 2024.
[20] Abdelfatah, Rund, and Ramtin Arablouei. "What Is the Bechdel Test? A Shorthand for Measuring Representation in Movies." *NPR*, 5 Apr. 2023. *NPR*, https://www.npr.org/2023/04/05/1168116147/what-is-the-bechdel-test-a-shorthand-for-measuring-representation-in-movies.
[21] API Documentation. https://bechdeltest.com/api/v1/doc. Accessed 23 Apr. 2024.

[22] *Stats and Graphs - Bechdel Test Movie List*. https://bechdeltest.com/statistics/. Accessed 23 Apr. 2024.
[23] *Academy Membership | Oscars.Org | Academy of Motion Picture Arts and Sciences*. 9 Feb. 2024, https://www.oscars.org/about/join-academy.
[24] Glitre, Kathrina. *Academy Awards*. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=66164c0a784744a8893c2aae6d6fdcca6cc91bb4.
[25] Nelson, Ra, et al. "What's an Oscar Worth?" *Economic Inquiry*, vol. 39, no. 1, 2001, pp. 1–6. *Wiley Online Library*, https://doi.org/10.1111/j.1465-7295.2001.tb00046.x.
[26] Gunter, Barrie. "Do Awards Make a Difference?" *Predicting Movie Success at the Box Office*, edited by Barrie Gunter, Springer International Publishing, 2018, pp. 195–207. *Springer Link*, https://doi.org/10.1007/978-3-319-71803-3_12.
[27] *2024 Oscar Ratings: Academy Awards Audience Rises to 21 Million*. https://variety.com/2024/tv/news/2024-oscar-ratings-academy-awards-1235933476/. Accessed 23 Apr. 2024.
[28] Zauzmer, Rebecca Keegan, Ben, et al. "Inside the Academy's Inclusion Drive." *The Hollywood Reporter*, 25 Mar. 2022, https://www.hollywoodreporter.com/movies/movie-features/the-academy-inclusion-data-1235116216/.
[29] McCabe, Janet. *Feminist Film Studies*. 2004.
[30] *ACADEMY ESTABLISHES REPRESENTATION AND INCLUSION STANDARDS FOR OSCARS® ELIGIBILITY | Oscars.Org | Academy of Motion Picture Arts and Sciences*. 9 Sept. 2020, https://www.oscars.org/news/academy-establishes-representation-and-inclusion-standards-oscars-eligibility.
[31] Peters, Jeremy W., and Brooks Barnes. "The Oscars Now Have D.E.I. Rules, but Some Say It's Just a Performance." *The New York Times*, 8 Mar. 2024. *NYTimes.com*, https://www.nytimes.com/2024/03/08/us/oscars-oppenheimer-dei-diversity.html.
[32] *About Us – Center for the Study of Women in Television & Film*. https://womenintvfilm.sdsu.edu/about-us/. Accessed 24 Apr. 2024.
[33] Lauzen, Martha. *The Celluloid Celing*. 2024, https://womenintvfilm.sdsu.edu/wp-content/uploads/2024/01/2023-Celluloid-Ceiling-Report.pdf.
[34] Lauzen, Martha. *It's a Man's (Celluloid) World*. https://womenintvfilm.sdsu.edu/wp-content/uploads/2024/02/2023-Its-a-Mans-Celluloid-World-Report.pdf.
[35] Milana, Taylor. "Gender Inequality in Hollywood: The Magnitude, Determinants, and Influence of the Gender Wage Gap in the Film Industry." *Scripps Senior Theses*, Jan. 2019,
[36] Yang, Luoying, et al. "Measuring Female Representation and Impact in Films over Time." *ACM/IMS Transactions on Data Science*, vol. 1, no. 4, Nov. 2020, p. 30:1-30:14. *ACM Digital Library*, https://doi.org/10.1145/3411213.
[37] Lakhotia, Raghav, et al. *Identifying Missing Component in the Bechdel Test Using Principal Component Analysis Method*. May 2019. *arXiv.org*, https://doi.org/10.5281/zenodo.3299335.
[38] Zemaityte, Vejune, et al. "Quantifying the Global Film Festival Circuit: Networks, Diversity, and Public Value Creation." *PLOS ONE*, vol. 19, no. 3, Mar. 2024, p. e0297404. *PLoS Journals*, https://doi.org/10.1371/journal.pone.0297404.

[39] Smelik, Anneke. “Feminist Film Theory.” *The Cinema Book*, chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.annekesmelik.nl/TheCinemaBook.pdf.
[40] Gaines, Jane M. “Film History and the Two Presents of Feminist Film Theory.” *Cinema Journal*, vol. 44, no. 1, 2004, pp. 113–19.
[41] Mulvey, Laura. “Visual Pleasure and Narrative Cinema.” *Feminism and Film Theory*, Routledge, 1988.
[42] O’Meara, Jennifer. “What ‘The Bechdel Test’ Doesn’t Tell Us: Examining Women’s Verbal and Vocal (Dis)Empowerment in Cinema.” *Feminist Media Studies*, vol. 16, no. 6, Nov. 2016, pp. 1120–23. *Taylor and Francis+NEJM*, https://doi.org/10.1080/14680777.2016.1234239.
[43] Jones, Deborah, and Judith K. Pringle. “Unmanageable Inequalities: Sexism in the Film Industry.” *The Sociological Review*, vol. 63, no. 1_suppl, May 2015, pp. 37–49. *SAGE Journals*, https://doi.org/10.1111/1467-954X.12239.
[44] *Above-the-Line vs. Below-the-Line Crew: What’s the Difference?* 8 Apr. 2022, https://www.backstage.com/magazine/article/above-the-line-vs-below-the-line-crew-differences-74969/.
[45] Assemble. *The Definitive Film Crew Hierarchy Chart*. https://blog.assemble.tv/the-definitive-film-crew-hierarchy-chart. Accessed 19 Apr. 2024.
[46] Malone, Aubrey. *Hollywood’s Second Sex*. 2015.
[47] Paszkiewicz, Katarzyna. *Genre, Authorship and Contemporary Women Filmmakers*. 2018.

10. Appendix

Appendix A: Background on the Academy Awards

The Academy Awards, otherwise known as the Oscars, are an annual award ceremony that is presented by the Academy of Motion Picture Arts and Sciences (AMPAS). The awards recognize film achievements in a range of categories, including acting, directing, and many technical and design-related categories. The first Academy Awards were presented in 1929, and the most recent Academy Awards took place on March 10, 2024, marking the 96th Academy Awards.

Films are nominated and voted on by members of the Academy. More specifically, nominations are produced by members of the branch of the Academy that the award is in (e.g. Acting branch for Acting awards), and then all members vote to determine the winners of each

award. There are currently 18 branches. Academy membership is automatically given to any nominee, and the only other way to receive membership is to be sponsored by two current members of the branch to into which the candidate wants to be admitted [23] [24].

As mentioned above, Oscar nominations provide significant financial benefits to a film. 2001 study "What's an Oscar Worth?" found that a nomination for Best Picture increases predicted box office revenues by almost $5,000,000 , and winning Best Picture increases predicted box office revenues by over $12,600,000 [25]. 2018 study "Do Awards Make a Difference," similarly finds that Oscar wins in main categories (Best Picture, Best Actor/Actress, Best Director, etc) provide confirmation of the quality and free publicity for a film, and in turn boost box office revenue [26]. The awards ceremonies themselves continue to have widespread viewership, with an estimated 21 million viewers watching the broadcast this year [27].

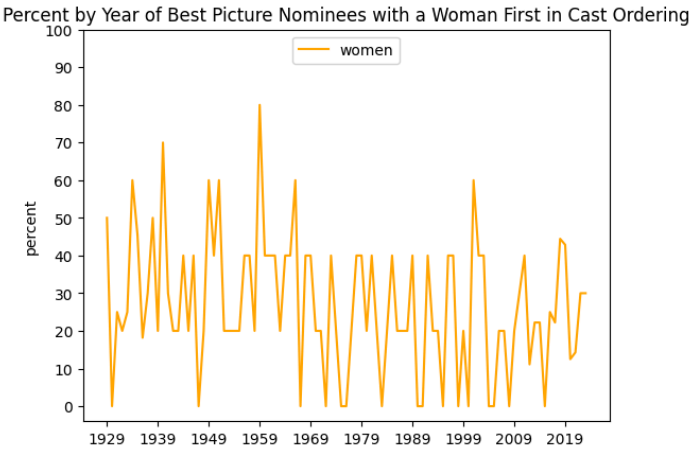
The Academy recently inducted many new members with the goal of evening racial and gender representation in their membership, which was previously dominated by white men [28]. Additionally, in 2020 the Academy established a new set of "representation and inclusion standards" that a film must meet in order to be nominated for Best Picture [30]. These standards went into effect for the most recent Academy Awards in March of 2024, and include criteria for on-screen and off-screen representation of underrepresented groups that a film must meet. Despite the Academy's pride in these new standards, many filmmakers are skeptical of the rules' real impact [31]:

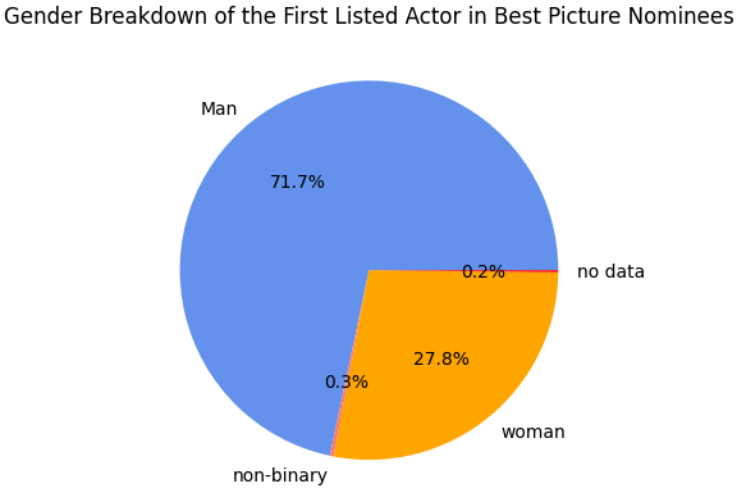
> The director Spike Lee, whose films often explore the country's tortured history with racism, has said that while he thinks the academy's "heart is in the right place," the standards contain "a lot of loopholes." Mr. Lee, who declined to comment further, has

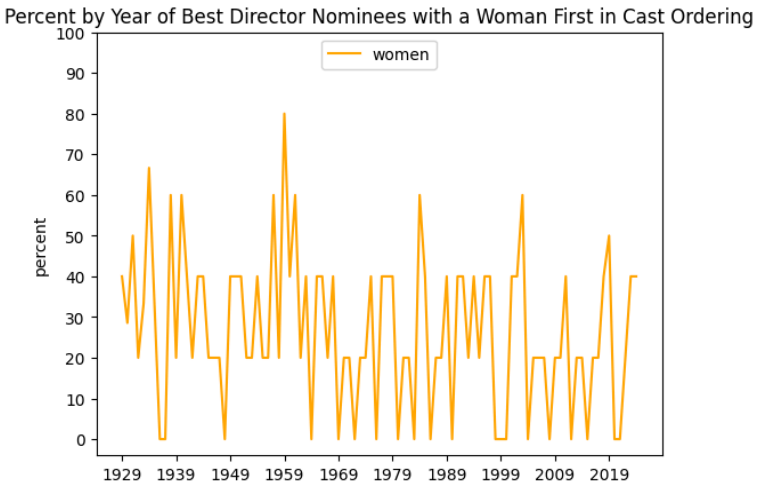
also said that nothing will change unless the studio gatekeepers who greenlight films come from more diverse backgrounds.

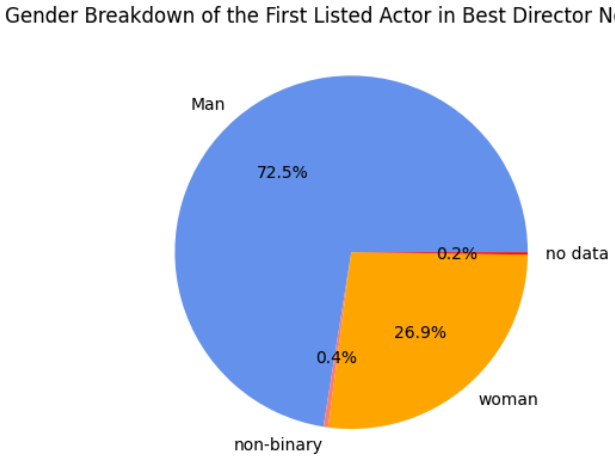
This same New York Times article takes *Oppenheimer* as an example of a film that, despite being centered on white men, managed to pass the diversity standards. These rules seem to be a step in the right direction for the Academy, but controversy surrounds whether these standards can actually promote real change in the industry.

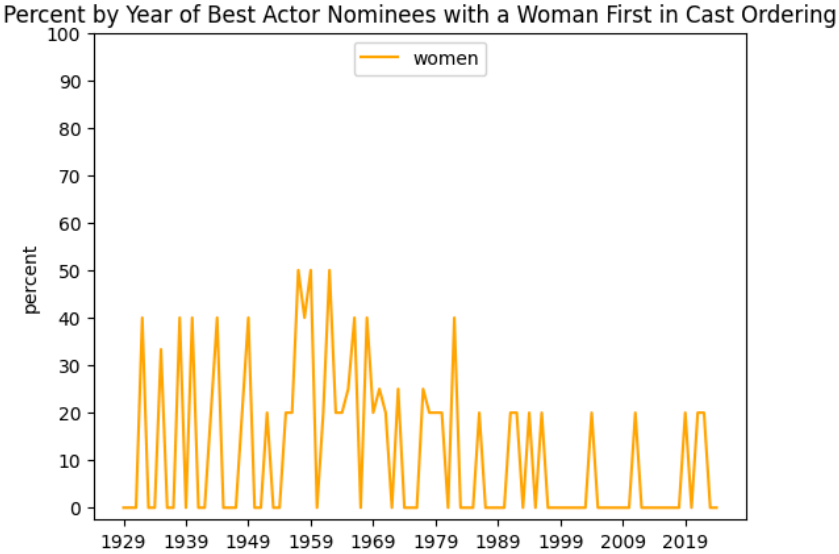
Appendix B: Extra Figures

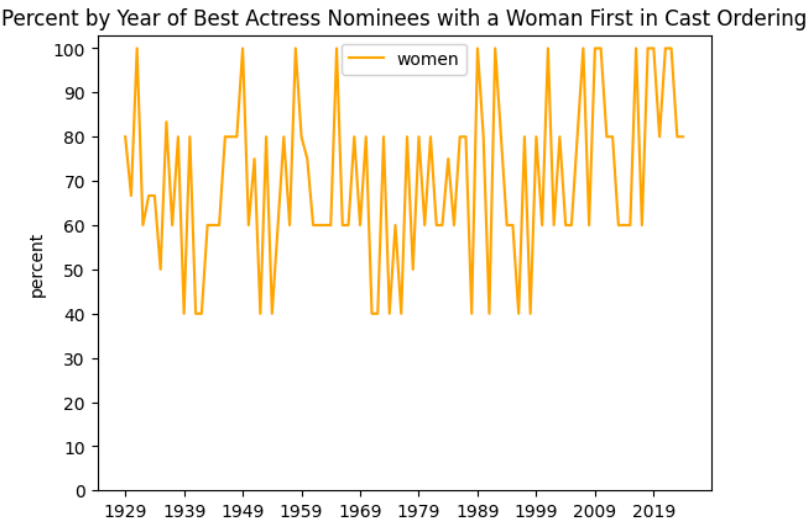


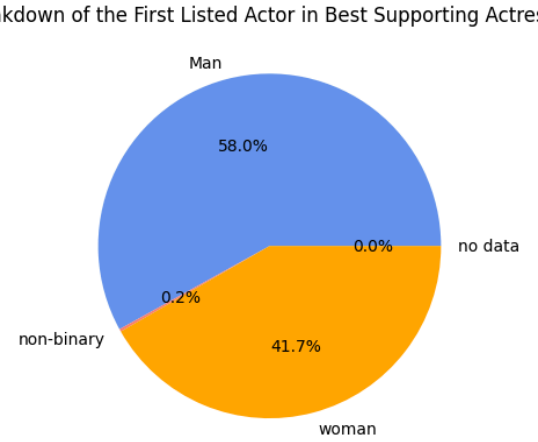


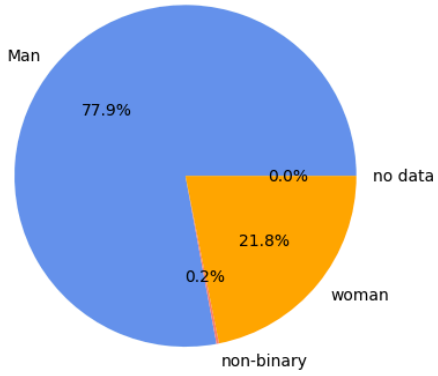

Percent by Year of Best Director Nominees with a Woman First in Cast Ordering
women
percent
100
90
80
70
60
50
40
30
20
10
0
1929
1939
1949
1959
1969
1979
1989
1999
2009
2019


Gender Breakdown of the First Listed Actor in Best Director Nominees
Man
72.5%
0.2%
no data
26.9%
0.4%
woman
non-binary


Percent by Year of Best Actor Nominees with a Woman First in Cast Ordering
women
percent
100
90
80
70
60
50
40
30
20
10
0
1929
1939
1949
1959
1969
1979
1989
1999
2009
2019


Percent by Year of Best Actress Nominees with a Woman First in Cast Ordering
women
percent
100
90
80
70
60
50
40
30
20
10
0
1929
1939
1949
1959
1969
1979
1989
1999
2009
2019


Gender Breakdown of the First Listed Actor in Best Supporting Actress Nominees
Man
58.0%
0.0%
no data
0.2%
non-binary
41.7%
woman


Gender Breakdown of the First Listed Actor in Best Supporting Actor Nominees
Man
77.9%
0.0%
no data
21.8%
0.2%
woman
non-binary

11. Honor Pledge

I pledge my honor that this paper represents my own work in accordance with University regulations

Alison Silldorff