A3: Project Proposal

Alison Yao (yy2564) | Fatima Nadeem (fn2093) | Min Jie Kim (mjk9913) | Jun Bin Ho (jbh439)

1. Motivation/problem statement

We will attempt to assess the comprehensiveness of the Bechdel Test, a simple and popular measure of feminism in works of fiction. The rules of the test were first defined back in 1985 in a comic strip featuring two queer women who could not find a movie that:

1) Has at least two women in it,
2) who talk to each other,
3) about something other than a man.

With US adults spending 6 hours a day consuming video-based media, media representation of social groups has an increasingly profound impact on the formation of prejudices and opinions about social groups. In her influential article Gendered Media, Julia T. Wood defines three ways in which media "perpetuate unrealistic, stereotypical, and limiting perceptions" about women:

1) Women are underrepresented, which falsely implies that men are the cultural standard and women are unimportant or invisible.
2) Second, men and women are portrayed in stereotypical ways that reflect and sustain socially endorsed views of gender.
3) Third, depictions of relationships between men and women emphasize traditional roles and normalize violence against women.

The degree to which each of these factors occurs in the media is difficult to measure, but The Bechdel Test offers us a very useful measure of how women and their agency are sidelined in fiction (Theme 1). Indeed, feminist media is difficult to measure because the second and third themes describe highly nuanced and layered relations between the genders and such relations are difficult to define with simple conditionals. Considering this, it is easy to see how the Bechdel test offers a limited measure of feminism in fiction, even if it has been widely applied due to its simplicity. Further, the meaning of Feminism more broadly is constantly progressing, varying over time, geography, and cultural context. First coined in legal scholarship by Kimberlé

Cranshaw, the concept of intersectionality in feminism has been brought into the spotlight in recent years, highlighting the "conceptual limitations of ... single-issue analyses". An intersectional framework helps us move beyond the unidimensional view that sexism is a problem faced by all women in the same way, and consider how it interplays with other obstacles like racism and homophobia to predispose individuals to vastly varying experiences. We would need many more tests or more detailed ones to measure varying representations of women of color or queer women, and is limited as a tool from an intersectional feminist perspective.

More practically, there are few other tools (gender representation in directors, producers, crew members, etc.) to measure gender equality in media, and it can be a useful endeavor to expand on the Bechdel test to account for these limitations.

We hope to learn about methods or proxies to measure the human-centered issue of representation, which requires us to think about how we can *qualify* how movies depict women. We also hope to explore how we can leverage other quantitative benchmarks as a tool to "fill the gaps" of the Bechdel test.

2. Research questions and/or hypotheses

Our primary goal is to find links between the Bechdel Test and other measurements of representation of female agency in movies, thereby minimizing the limitations of each in an attempt to construct a more comprehensive test. This leads us to the following research questions:

- How has performance on the Bechdel test evolved over the years?
- What is the relationship between the Bechdel test and other metrics of female representation (e.g. female directors, producers, screenwriters)?

In this vein, how can we further incorporate qualitative analysis/thick data into this test, for instance, incorporate critical film theory?

- A strength of the Bechdel test is its simplicity, which makes it easy to categorize movies into those that pass and fail. What are the limitations of this simple set of conditionals and are they still relevant?

- Can we convert critical film theory into a more defined set of conditionals for an improved test?

3. Background and/or Related Work

Gender portrayals of women in film have been frequently criticized by the media, especially as societal movements related to feminism have started to spark more frequently. The first spark regarding feminism started in the "late nineteenth and early twentieth century", and aligning with this timeline, the Bechdel Test was created to measure "the representation of women in fiction" in 1985. The test was widely adopted and commonly praised by critics, as it was viewed as "a good way to ensure inclusivity and a strong, dominant female presence rather than the typical passiveness that requires a female's character to depend on a male character for development." As a result, the test has been used in various social studies, such as investigating the correlation between the profit of movies and whether the respective movies have passed the Bechdel Test. Using Bechdel Test as the forefront of this analysis, studies have concluded that movies that passed the Bechdel Test are more likely to be financially successful, and feminist groups have used this reason to push for improved portrayals of women in the media.

However, the Bechdel Test was not safe from receiving negative criticism. Due to the nature of the test, which arguably only looks for extremely simple, quantitative data points, the test has been criticized for ignoring how characters "should be validated by the content of their character and their drive to accomplish clear, female-centric goals." Additionally, with feminism being a wide area, critics have commented that "the test's utility "has been elevated way beyond the original intention", where the Bechdel test only provides the degree of women's presence in a film or media source.

Encountering such mixed reviews of the Bechdel Test, we wanted to design our research questions to acknowledge the success and ups of the Bechdel Test, while also navigating the shortcomings of the test.

4. Data used

The Bechdel test website allows access to the raw data using their API. Through this, we have access to the Bechdel Test database that we plan to use when analyzing the effectiveness and the shortcomings of the test. This dataset is highly relevant to our research in that it contains human-labeled Bechdel Test scores. If the score is 3, the movie passes the Bechdel Test. Otherwise, it fails. Any movie viewer can sign up to this website and give scores to any movie as long as they provide reasonable comments to justify their scores. Other users can re-examine a movie and overturn the score if they can specify which characters or scenes of a movie violate the Bechdel Test. We can use an API call to save the data as a CSV file. There are four methods to get the data: (1) get movies by IMDb id, (2) get movies by title, (3) get all movie ids, and (4) get all movies. Since we are analyzing all movies available, we will use method (4) to get the data on the release year, rating (Bechdel score), imdbid, id (unique website id), and title of all movies rated on this website.

	year	rating	imdbid	id	title
0	1874	0	3155794	9602	Passage de Venus
1	1877	0	14495706	9804	La Rosace Magique
2	1878	0	2221420	9603	Sallie Gardner at a Gallop
3	1878	0	12592084	9806	Le singe musicien
4	1881	0	7816420	9816	Athlete Swinging a Pick

Unfortunately, if we attempt to get all movies, the information on each movie is not going to be as abundant as getting each movie by IMDb id. Method (1) tells us if the movie has been approved, its title, year of release, Imdb ID, unique website ID, the date of entry into the dataset, ID of the submitter, the actual Bechdel score, and whether the submitter considered the rating is dubious or not. We can technically loop through all the movie IMDb id and expand the current dataset. However, the website owner warns us not to send a lot of queries in a short period of time, because this website is running on a shared hosting plan and generating too much traffic might crash the website. Therefore, we will try our best to retrieve more data on each movie one by one, but in the worst-case scenario, the current dataset already has the most important information - the Bechdel Test score. This ready-made dataset is under the Attribution-Noncommercial 3.0 Unported (CC BY-NC 3.0) license that allows sharing and adaptation. However, the comments available on site are property of the commenters and these would not be a part of the analysis for this project. The data collection and distribution measures

have been fair since the beginning so there seem to be no ethical issues involved. But, FiveThirtyEight points out a potential "feminist-leaning" bias that manifests as people more inclined to submit movies passing the Bechdel Test to the database.

Additionally, we are planning to acquire freely available datasets that include basic information about movies. For our analysis on the relation between the gender diversity of the movie crew members - such as writers, producers, and directors - and the Bechdel Test score for the movie, we are looking to find datasets with information on the crew members and their gender identity. This Kaggle data fits our needs. It concludes several datasets, one of which lists cast and crew names and gender information. Using such datasets, we plan to run our own test (comparing information and visualization) and analyze the relationship between the aforementioned criteria. This Kaggle dataset is collected from TMDB and Movielens and is under CC0: Public Domain license, so we can use it for our analysis. The only inconvenience with this dataset is that it was uploaded 4 years ago, so we might lose a lot of important information regarding movies released in recent years.

For the quantitative analysis, our group plans to watch a set of movies ourselves and analyze them. This way, we will be able to obtain our own ethically researched database - though small. The movies themselves are the data that can be analyzed in many ways.

5. Unknowns and dependencies

Factors outside of our control are the usability of certain data and immeasurable time and possibilities of qualitative analysis.

- Although we have a few ready-made datasets at our disposal, we are not sure what will happen when we join them together and extract the information we need. There might be a lot of empty values or missing entries. We cannot say for sure what our finalized dataset will look like before we start to write some codes.
- We have no control over how long it will take for us to watch and analyze each movie quantitatively. Therefore, our sample size for quantitative analysis will depend largely on the time we have.

- We are not sure where the qualitative analysis will take us. There is no limit on how much we can explore on this endeavor to improve on the Bechdel Test or even create a more representative metric for female empowerment in movies.

6. Methodology

We will employ both quantitative and qualitative methods to first map out Bechdel Test's plausibility and then question its effectiveness.

As mentioned in section 4, there are plenty of ready-made datasets that we can utilize, so our quantitative analysis already has a lot to build upon. As for the data for qualitative analysis, we will focus on some of the outliers in our quantitative analysis, controversial movies on media, and movies that stand out to us in the Bechdel Test.

For quantitative analysis, we will center around the Bechdel test data and join with complementary datasets such as director gender and gender ratio for the cast and crew, etc. We plan to first perform data visualization for descriptive and statistical analyses. More specifically, data visualization will include bar charts, pie charts, and scatter plots. Visualizing information is crucial for exploratory data analysis and helpful for understanding the trend in Hollywood over the years. After having a good understanding of the data, we will perform an independent two-sample t-test for the male and female director groups and compare means of the Bechdel Test scores to see if there is a statistical significance between their differences. The test here needs to be an independent t-test because the male and female director groups are not matched or paired. And we are using a t-test instead of a z test because we do not know the population mean or standard deviation. Moreover, once we have the gender ratio of the cast and the crew, we will have more continuous variables as our features. Then, we can perform a regression analysis using ordinary least squares on, for instance, the Bechdel test scores, gender proportion of cast members and gender proportion of crew members and so on to answer our research questions. Given that these features are continuous variables (ratio) or categorical (ordinal), regression analysis will probably be the most appropriate. But if we are using the Bechdel test scores, which is technically on an ordinal scale, we will resort to a one-way ANOVA analysis. Otherwise, if all features are continuous, we will resort to a linear regression model.

As for qualitative analysis, we will pay special attention to certain outliers, widely-known controversial movies or movies of our choice as our Thick Data. We will watch these movies and take notes of details regarding female representation. In our background information, we've noted criticisms that female representation "should be validated by the content of their character and their drive to accomplish clear, female-centric goals." Therefore, our qualitative analysis will examine the character and development arc of characters and break down the plot of the movies we select, in order to determine their goals and see if they align to be female-centric. We will then compare our findings to the Bechdel test to subjectively criticize if the Bechdel test is appropriate in each scenario.

Finally, this project will unite both quantitative and qualitative analysis to provide a comprehensive result on our investigation of the Bechdel test. Advantages, limitations and suggestions for improvement will be highlighted. We will present our research outcome in the format of a project report and a presentation.

7. Timeline

We will have 5 weeks to complete the final project. Since each step builds on the previous one, for example, we will look for outliers in the quantitative analysis to conduct the qualitative analysis, we will work on each part of the project one by one.

- 1 to 7 Nov: Data cleaning & organization - 1 week
- 8 to 14 Nov: Descriptive analysis, statistical analysis, regression - 1 week
- 15 to 29 Nov: Conduct qualitative analysis, improve on Bechdel test - 2 week
- 30 Nov - 7 Dec: Report & presentation - 1 week