# Dataset information and variable description

## Bechdel dataset

(files: bechdel_allmovies.json and bechdel_allmovies.csv)
http://bechdeltest.com/
The Bechdel Test, sometimes called the Mo Movie Measure or Bechdel Rule is a simple test which names the following three criteria: (1) it has to have at least two women in it, who (2) who talk to each other, about (3) something besides a man. The test was popularized by Alison Bechdel's comic Dykes to Watch Out For, in a 1985 strip called The Rule. For a nice video introduction to the subject please check out The Bechdel Test for Women in Movies on feministfrequency.com.

7372 movies
years: 1892 – 2017

**Variable description**

| **name** | **type** | **description** |
|---|---|---|
| imdbid | numeric | The IMDb id. |
| title | text | The title of the movie. Any weird characters are HTML encoded (so Brüno is returned as "Brüno"). |
| id | numeric | The bechdeltest.com unique id. |
| year | numeric | The year this movie was released (according to IMDb). |
| rating | numeric | The actual score. Number from 0 to 3 (0 means no two women, 1 means no talking, 2 means talking about a man, 3 means it passes the test). |

## IMDB dateset

files: imdb.json and imdb.csv
https://www.kaggle.com/deepmatrix/imdb-5000-movie-dataset
IMDB 5000 Movie Dataset
5000+ movie data scraped from IMDB website
28 variables for 5043 movies and 4906 posters (998MB), spanning across 100 years in 66 countries. There are 2399 unique director names, and thousands of actors/actresses.

5043 movies
years: 1916 – 2016

**Variable description**

| name | type | description |
|---|---|---|
| color | String | Color/Black and White |
| director_name | String | The directors name |
| num_critic_for_reviews | Numeric | number of reviews |
| duration | DateTime | Movie duration in minutes |
| director_facebook_likes | Numeric | Director facebook likes |
| actor_3_facebook_likes | Numeric | Third actor facebook likes |
| actor_2_name | String | Second actor name |
| actor_1_facebook_likes | Numeric | First actor facebook likes |
| gross | Numeric | Gross income of movie. Comment: Please be mindful that, analyzing currency related attributes, such as "gross" or "budget", is actually more complicated than it seems. For a really thorough and accurate analysis (EDA or prediction), we may want to do some feature engineering on those attributes in a systematic way. For example, one US dollar in 1920 is different from that of 2010. So we need to take inflation factors across years into consideration, and normalize all US dollars into one basis (a certain year). |
| genres | String | Action, Adventure, Drama, Romance, Western, Animation etc. Even combinations of these allowed eg. Action \| Comedy |
| actor_1_name | String | First actor name |
| movie_title | String | Movie title |
| num_voted_users | Numeric | Number of users voted |
| cast_total_facebook_likes | Numeric | Total cast facebook likes |
| actor_3_name | String | Third actor's name |
| facenumber_in_poster | Numeric | Number of face in poster |
| plot_keywords | String | Plot keywords |
| movie_imdb_link | String | Movie IMDB link |
| num_user_for_reviews | Numeric | Number of user reviews |
| language | String | Language |
| country | String | Country |
| content_rating | String | Content rating (eg. PG, PG-13 etc) |
| budget | Numeric | Movie budget |
| title_year | Numeric | Year of movie |

| | | |
|---|---|---|
| actor_2_facebook_likes | Numeric | Second actor facebook likes |
| imdb_score | Numeric | IMDB score for movie |
| aspect_ratio | Numeric | Aspect ration |
| movie_facebook_likes | Numeric | Movie facebook likes |

## Merged IMDB and Bechdel test datasets

(files: merged_imdb_bechdel.json and merged_imdb_bechdel.csv)
Combination of two previous with a reduced number of the imdb dataset variables and added bechdel test results to all matched movies.

2746 movies
years: 1916 – 2016

**Variable description**

| **name** | **type** | **description** |
|---|---|---|
| counter | numeric | counter id for each of the movies |
| imdbid | | |
| movie_title | | |
| title_year | | |
| director_name | | |
| actor_1_name | | |
| actor_2_name | | |
| actor_3_name | | |
| language | | |
| country | | |
| imdb_score | | |
| movie_imdb_link | | |
| bechdel_id | | |
| bechdel_rating | | |
| budget | | |
| movie_facebook_likes | | |