General Information about the IMDB Movie dataset

(original data/documentation from Hadley Wickham, RStudio)

IMDB.com is a website devoted to collecting movie data supplied by studios and fans. It claims that it is the biggest movie database on the web and is run by Amazon. For this purpose, we look at a data set of movies with a known length and at least one IMDB user rating (58,788 movies total in 2006). For each movie, we have the following variables:

- **title**: title of the movie
- **year**: year of release
- **budget**: total budget in US dollars
- **length**: length in minutes
- **rating**: average IMDB user rating
- **votes**: number of IMDB users who rated the movie
- **r1, r2, r3, ..., r10**: A user can rate the movie on a scale of 1 to 10 (best=10).

  We have the vote distribution for each rating, to midpoint of nearest decile
  0 = no votes, 4.5 = 1-9% votes, 14.5 = 10-19% votes,..., 94.5 = 90-99% of votes.

  A movie with the distribution of: {4.5, 0, 4.5, 4.5, 24.5, 24.5, 14.5, 4.5, 4.5, 14.5} received 1-9% Rating 1 votes, zero Rating 2 votes, 1-9% Rating 3 votes, 1-9% Rating 4 votes, 20-29% Rating 5 votes, 20-29% Rating 6 votes, 10-19% Rating 7 votes, 1-9% Rating 8 votes, 1-9% Rating 9 votes, and 10-19% Rating 10 votes.

  The distribution may not sum to 100 just due to rounding errors.
- **mpaa**: Motion Picture Association of America rating: G, PG, PG-13, R, NC-17
- **Action, Animation, Comedy, Drama, Documentary, Romance, Short**: indicator genre variables (1 = Yes)