

A close-up, sepia-toned photograph of a film reel. The reel is on the left side of the frame, with a film strip extending from it and curving across the bottom of the image. The background is a plain, light-colored surface.

IMDB movies data

Hadley Wickham

Outline

- Data description
- Transformations
- Exploration

Data

- 58,788 movies from the IMDB website
- Title and year
- Length, budget
- Average rating, number of votes, voting breakdown (r1-r10)
- MPAA rating
- Genre (action, animation, comedy, drama, documentary...)

Transformation

- Some of the variables are very skewed (either low values or high values are more common)
- Sometimes it can be useful to transform the variables to be more symmetrical

Votes

Your turn

- Usual drill - explore and then report back.

Evaluation

- What do you like about the class? What would you like to see more of?
- What don't you like? What could I do better?
- What have you learned?