**Exercice 3 Etude du corpus**

1. Nous utilisons le corpus MovieLens 100k qui contient 1682 films et 943 utilisateurs.

2. A chaque film sont associées les caractéristiques suivantes : *movie id, movie title, release date, video release date, IMDb URL*, ainsi que 19 booléens (valeurs 1 ou 0) qui permettent de définir le genre du film : *unknown, Action, Adventure, Animation, Children's, Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery, Romance, Sci-Fi, Thriller, War, Western*

3. Chaque film a reçu un nombre moyen de 59 jugements. Chaque utilisateur a noté en moyenne 106 films. L’écart type du nombre de jugements par film est de 80.

4. La répartition des notes est la suivante :