

Filtrage collaboratif - Analyse de ml-1m

Biwei Cui, Claudia Delgado, Mehdi Miah et Ulrich Mpelé Mpelé

Contents

Préambule	3
Ouverture des fichiers	3
Détermination du volume des données	3
A partir de la base des notes : data.Ratings	3
Vérification avec les autres bases : data.Movies et data.Users	3
Nombre de films	3
Statistique univariée	3
Sur les notes	3
Sur les utilisateurs	4
Sur les films	8
Statistique bivariable	8
Statistique par utilisateur	8
Statistique par film	10
Corrélations entre les utilisateurs et les films	13
Corrélation entre note des films et nombre de visionnage de ce film	13
Corrélation entre note attribuée et nombre de films vus	14
Corrélation entre âge et note moyenne	15
Corrélation entre âge et nombre de films notés	16
Statistique d'ordre	16
Sur les films	16
Films les plus notés	16
Films les moins notés	17
Films n'ayant reçu qu'une seule note	17
Films ayant reçu moins de n votes	17
Films ayant les meilleurs notes	18
Nombre de films ayant une note moyenne de 5 (note maximale)	19
Films ayant les meilleurs notes et dépassant un seuil de visionnage	19
Films ayant les pires notes	19

Films ayant une note moyenne de 1	20
Films ayant les pires notes et dépassant un seuil de visionnage	20
Sur les utilisateurs	20
Utilisateurs ayant le plus vote	20
Utilisateurs ayant le moins vote	20
Utilisateurs n'ayant vote moins que 20 fois	21
Utilisateurs ayant vote moins de n fois	21
Utilisateurs ayant donne les plus hautes notes	21
Utilisateurs ayant donne les meilleurs notes et dépassant un seuil de participation	21
Utilisateurs ayant donne les plus faibles notes	22
Utilisateurs ayant donne les pires notes et dépassant un seuil de participation	22
Etude par age	24
Etude par sexe	25
Etude par genre de films	26
Nombre de films en commun	26
Etude sur les similarités	26
Corrélation de Pearson	26
Distance euclidienne	26
Distance manhattan	26

Préambule

```
## Clean up  
rm(list=ls())  
cat("\014")
```

Ouverture des fichiers

Détermination du volume des données

A partir de la base des notes : data.Ratings

```
## La base de données des notes contient exactement 1000209 notes.  
  
## La base de données des notes contient exactement 6040 utilisateurs uniques.  
  
## La base de données contient exactement 3706 films uniques.  
  
%### Nettoyage de la base data.Movies (à mettre dans clean)  
  
## Le taux de complétion de la matrice des notes Y est de 0.04 %
```

Vérification avec les autres bases : data.Movies et data.Users

```
## La base de données des utilisateurs contient bien 6040 utilisateurs.
```

Nombre de films

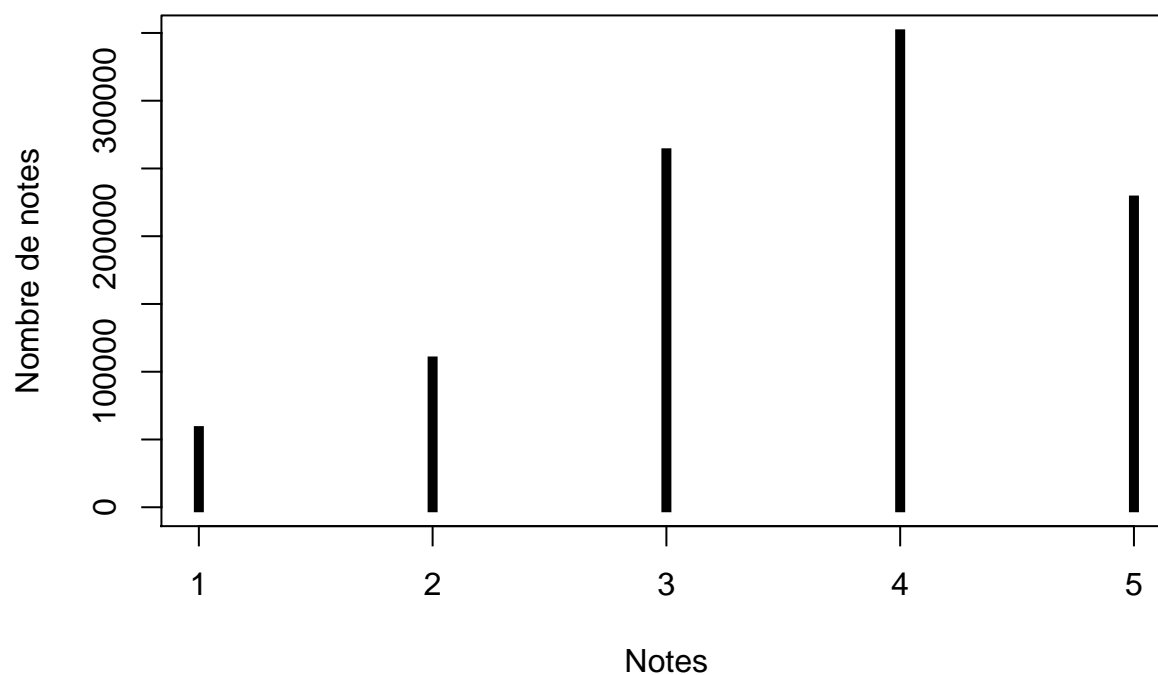
```
## La base de données contient bien 3706 films.
```

Statistique univariée

Sur les notes

```
## La moyenne des 1000209 notes est de 3.58 ; l'écart-type est de 1.12.
```

Distribution des notes



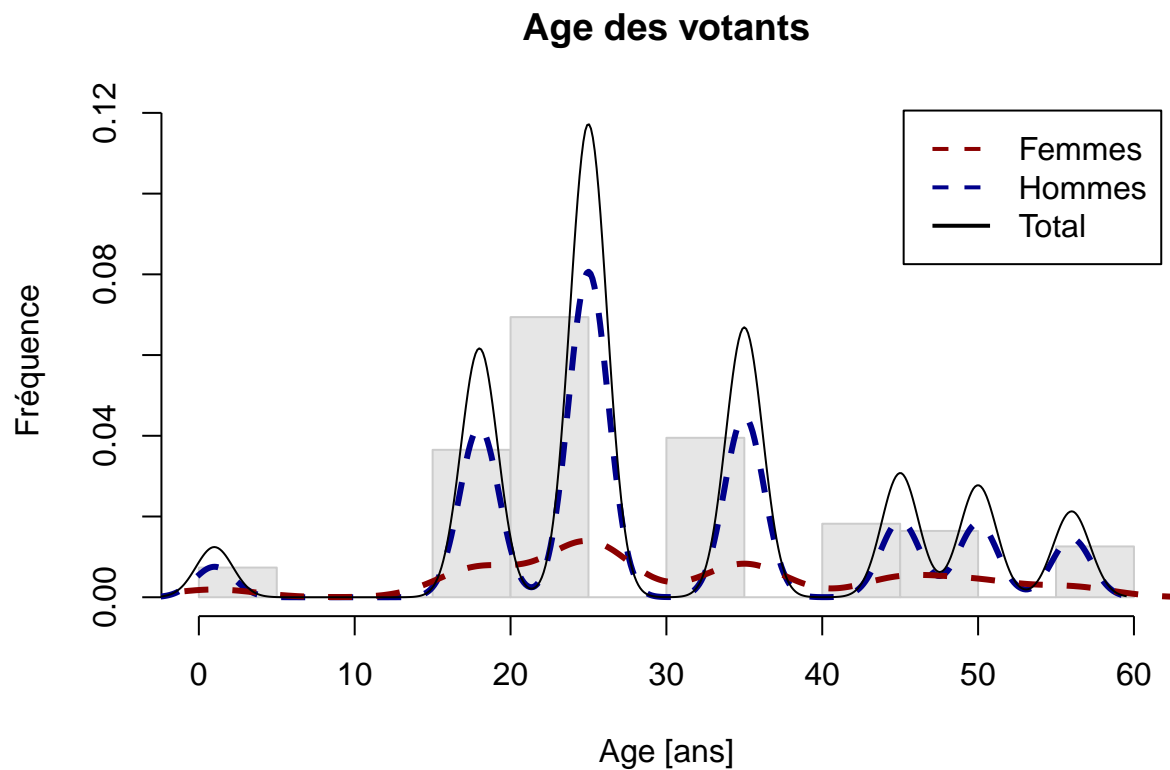
Effectifs par notes :

```
##
##      1      2      3      4      5
## 56174 107557 261197 348971 226310
```

Sur les utilisateurs

```
## La moyenne des âges est 30.64 ans avec un écart-type de 12.90 ans.
```

```
## La base de données est composée de 4331 hommes, soit 71.71% et 1709 femmes, soit 28.29%.
```



La répartition des âges est étrange : il y a des données incorrectes (1 an) et par classes (20-25 ans sur-représentés et 26-30 ans sous-représentés)

Proportions dans les métiers :

```
##      summary(data.Users$occupation)
## 0      711
## 1      528
## 10     195
## 11     129
## 12     388
## 13     142
## 14     302
## 15     144
## 16     241
## 17     502
## 18      70
## 19      72
## 2      267
## 20     281
## 3      173
## 4      759
## 5      112
## 6      236
## 7      679
## 8        17
## 9        92
```

Remarque : ces catégories professionnelles sont étranges, ce n'est pas une catégorisation CSP traditionnelle.

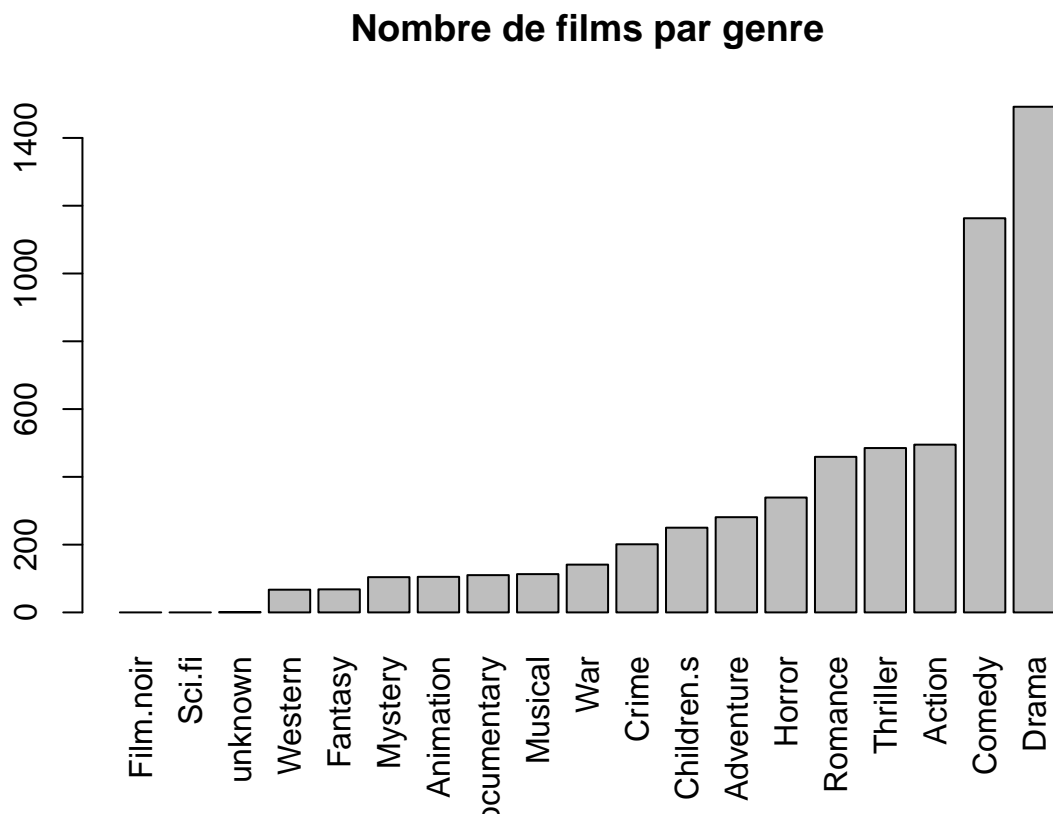
Proportions dans les codes postaux :

```
##          summary(data.Users$zip.code)
## 48104                      19
## 22903                      18
## 55104                      17
## 94110                      17
## 10025                      16
## 55105                      16
## 55455                      16
## 02138                      15
## 55408                      15
## 94114                      15
## 90024                      14
## 94117                      14
## 98103                      14
## 10016                      13
## 37922                      13
## 48103                      13
## 60640                      13
## 90210                      13
## 55406                      12
## 02139                      11
## 10003                      11
## 48105                      11
## 55113                      11
## 55405                      11
## 02143                      10
## 10011                      10
## 10021                      10
## 10023                      10
## 13210                      10
## 55112                      10
## 55116                      10
## 60201                      10
## 60657                      10
## 97330                      10
## 98115                      10
## 10014                      9
## 10019                      9
## 14850                      9
## 27514                      9
## 48197                      9
## 53706                      9
## 55117                      9
## 55410                      9
## 55421                      9
## 60611                      9
## 77005                      9
## 94109                      9
## 94115                      9
## 95616                      9
## 98102                      9
```

## 02115	8
## 02135	8
## 07030	8
## 11215	8
## 43201	8
## 55414	8
## 60614	8
## 60618	8
## 78705	8
## 91505	8
## 92103	8
## 94043	8
## 94121	8
## 94122	8
## 94306	8
## 97401	8
## 98105	8
## 01545	7
## 04240	7
## 10022	7
## 10027	7
## 20009	7
## 22030	7
## 43402	7
## 48073	7
## 53711	7
## 55126	7
## 60622	7
## 90069	7
## 92122	7
## 94025	7
## 94107	7
## 94133	7
## 94530	7
## 95008	7
## 01060	6
## 02144	6
## 02215	6
## 06074	6
## 06511	6
## 08904	6
## 10012	6
## 10013	6
## 10024	6
## 10036	6
## 11217	6
## 24060	6
## 30030	6
## 37923	6
## (Other)	5111

Remarque : le code postal n'est pas du tout pertinent (5111 other)

Sur les films

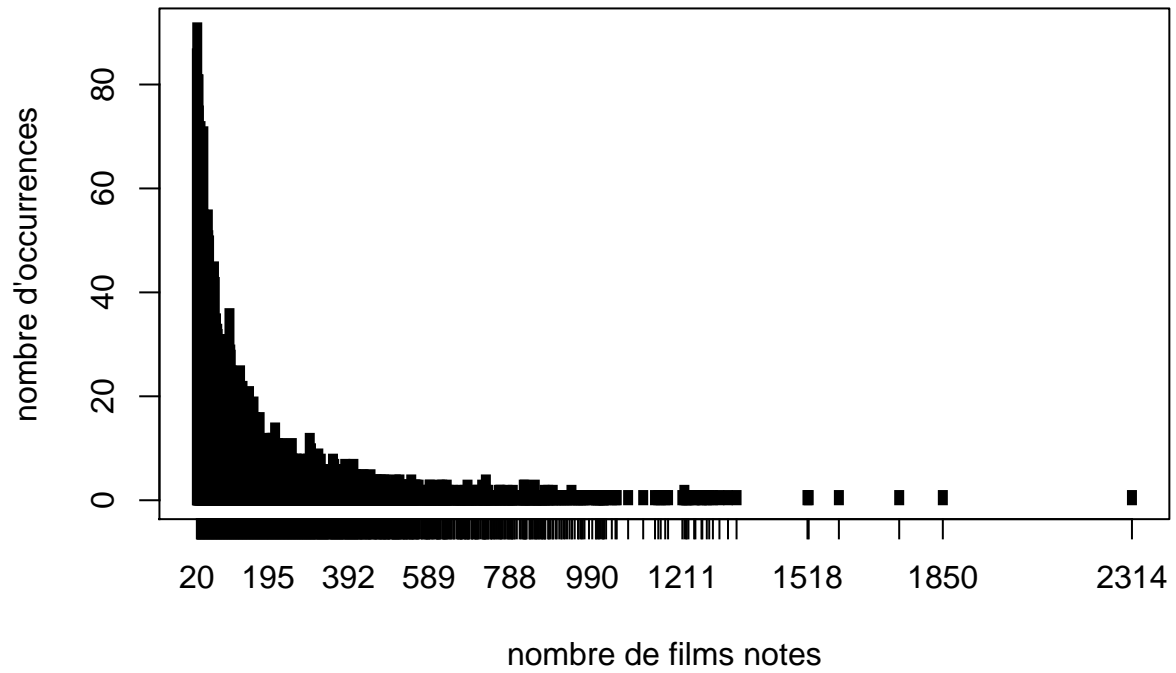


Statistique bivariée

Statistique par utilisateur

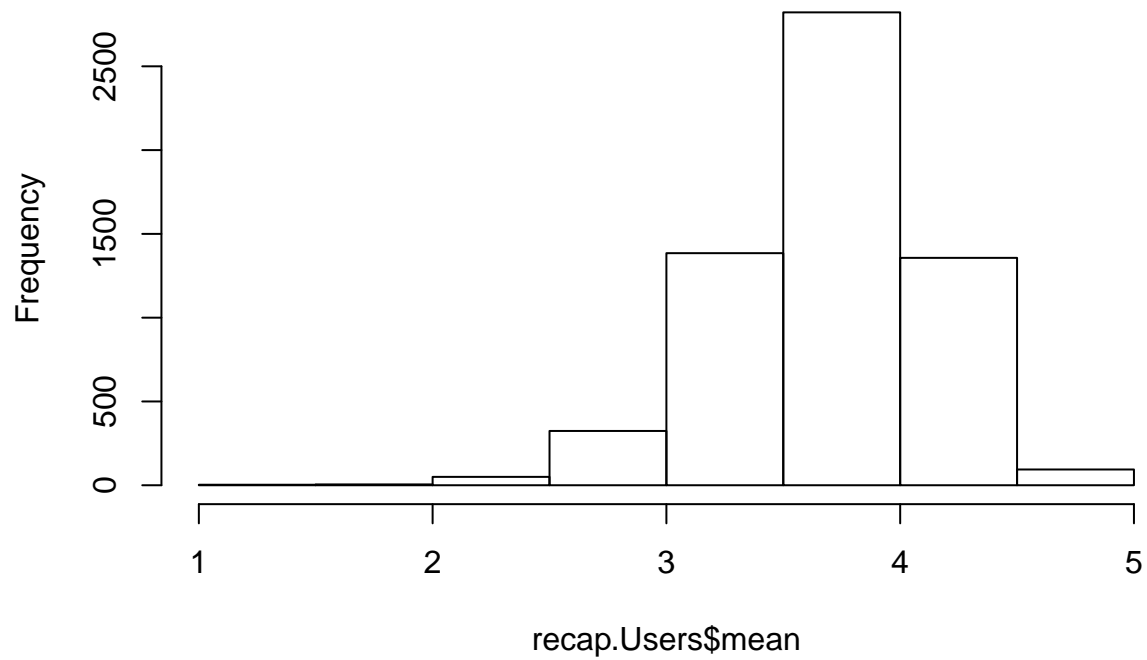
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	20.0	44.0	96.0	165.6	208.0	2314.0

Nombre de films notes par utilisateur



##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.020	3.440	3.740	3.703	4.000	4.960

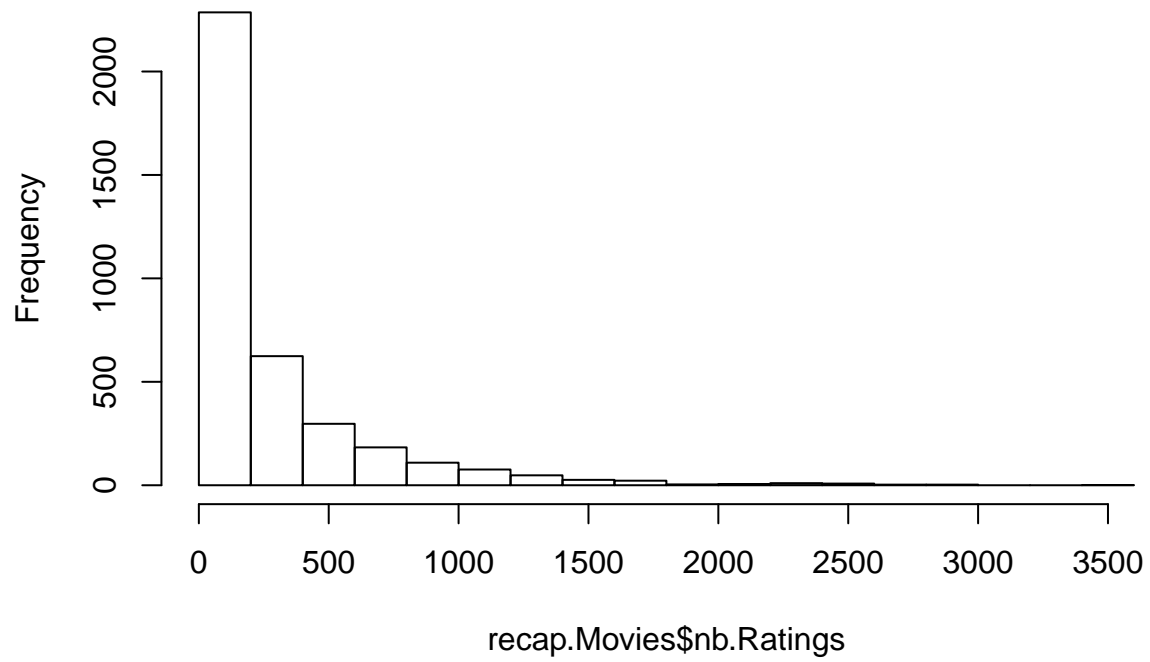
Note moyenne donnee par utilisateur



Statistique par film

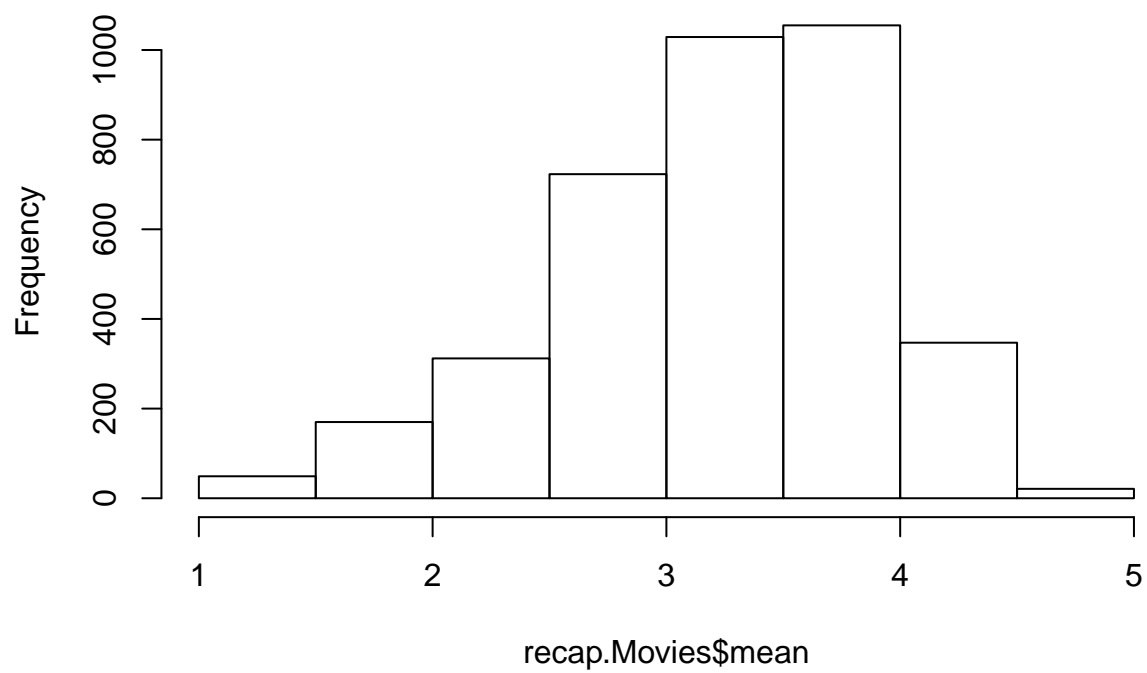
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.0	33.0	123.5	269.9	350.0	3428.0

Nombre de notes par film



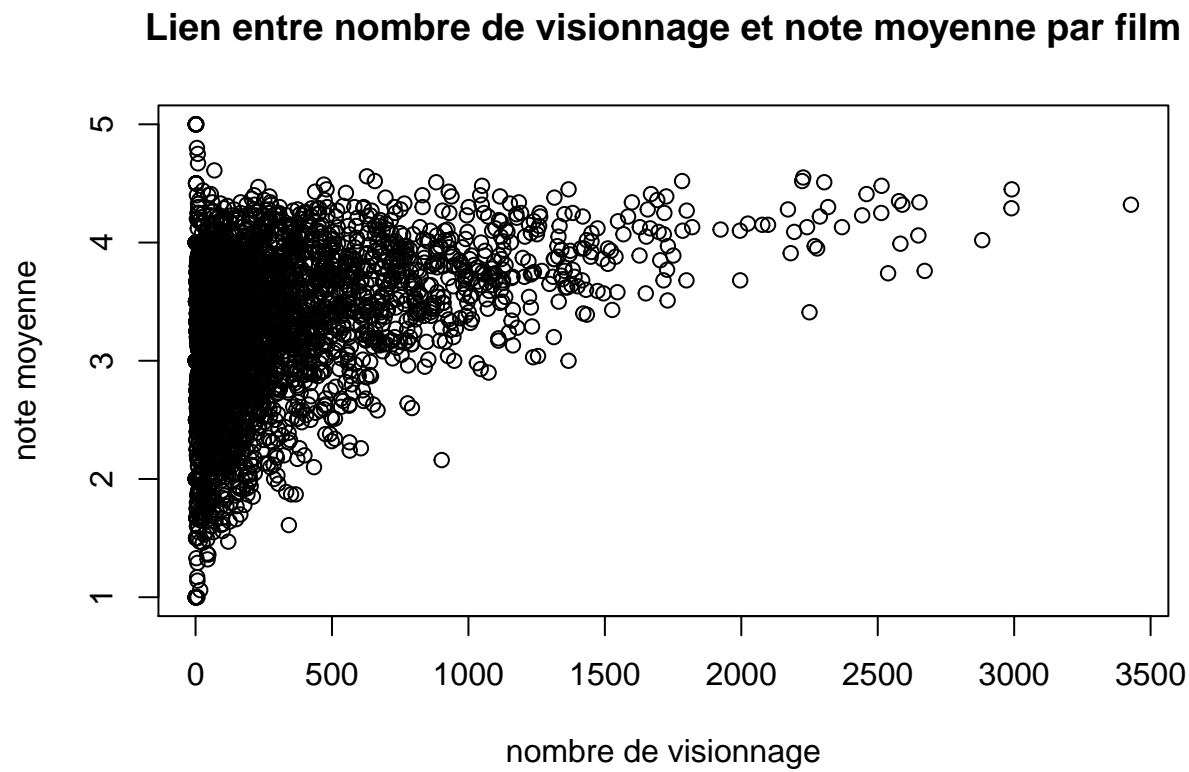
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.000	2.820	3.330	3.239	3.740	5.000

Note moyenne des films



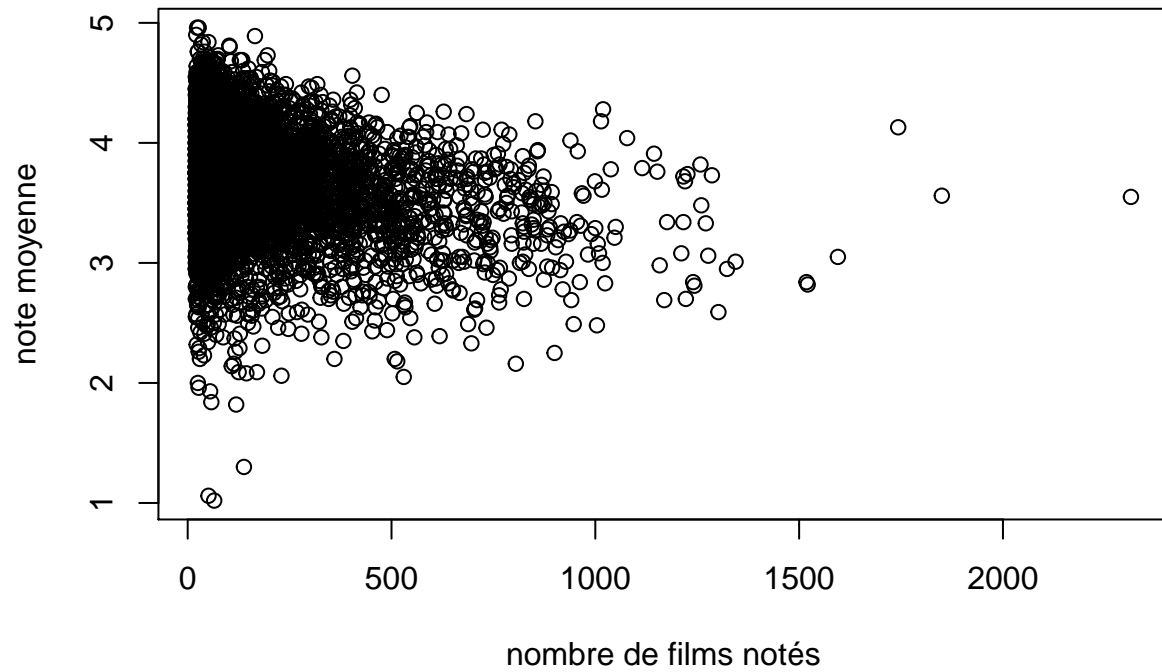
Corrélations entre les utilisateurs et les films

Corrélation entre note des films et nombre de visionnage de ce film

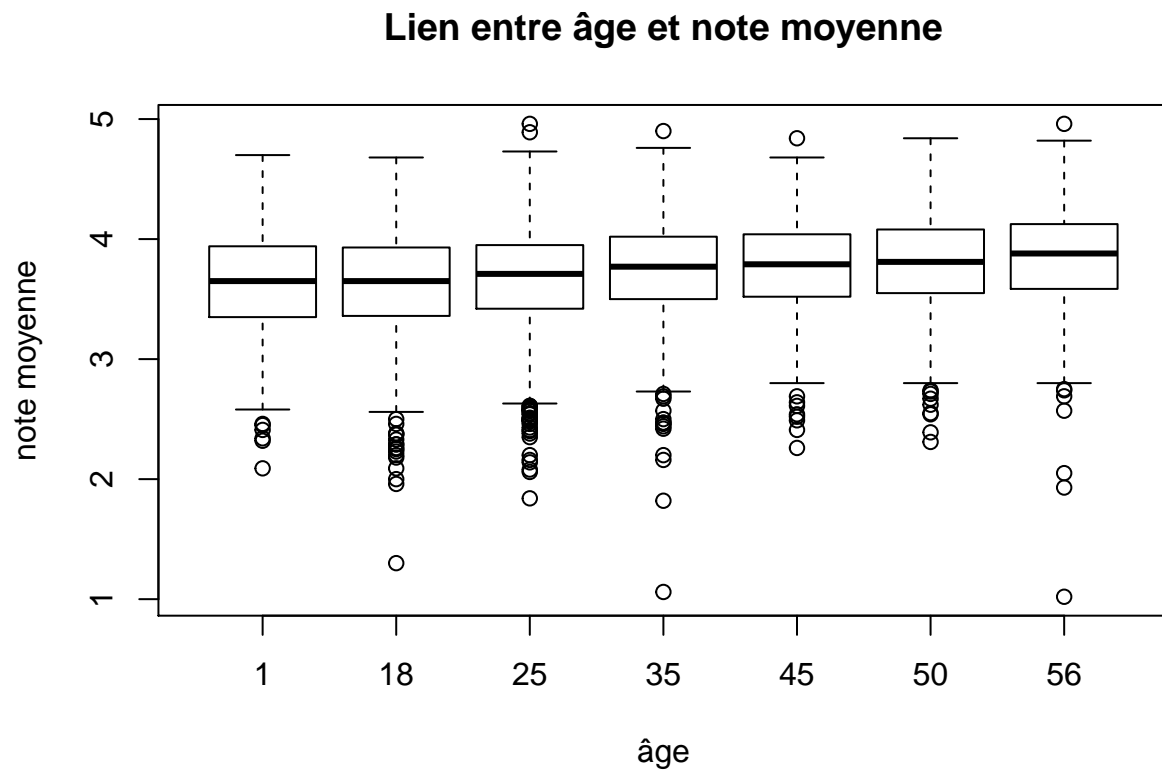


Corrélation entre note attribuée et nombre de films vus

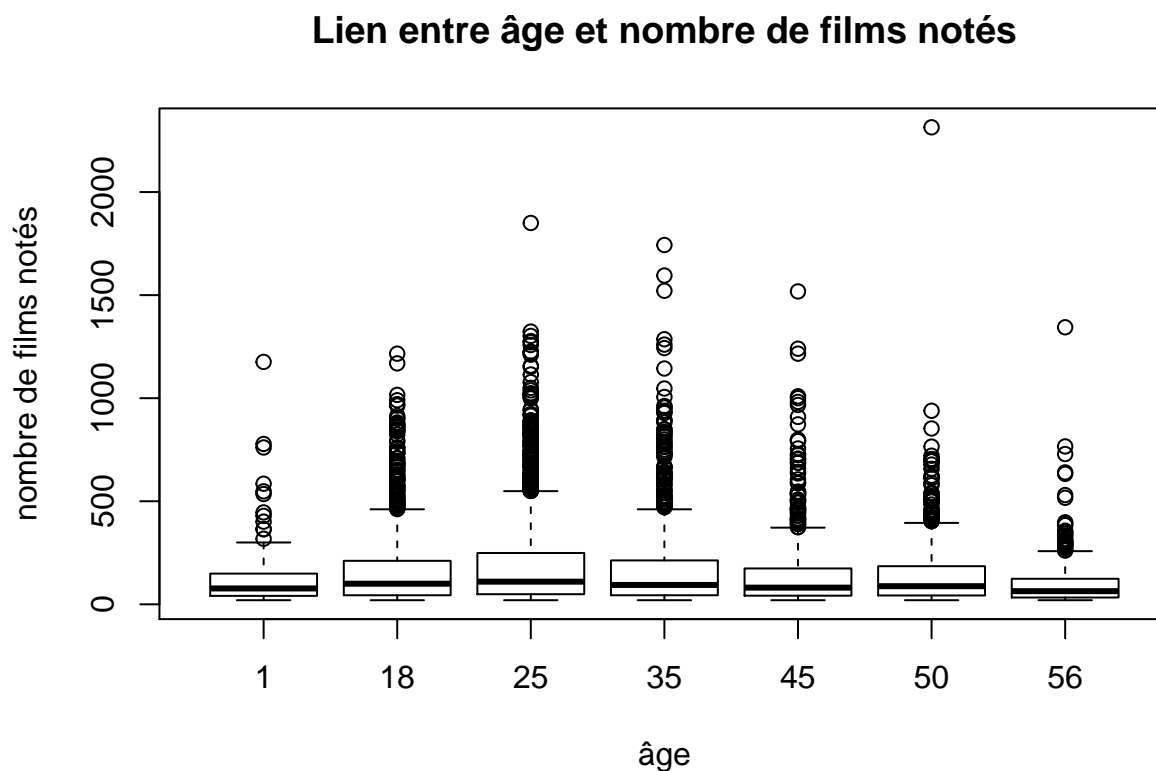
Lien entre nombre de films notés et note moyenne par utilisateur



Corrélation entre âge et note moyenne



Corrélation entre âge et nombre de films notés



Statistique d'ordre

Sur les films

Films les plus notés

Les 25 films ayant reçu le plus de notes :

```
## Le film d'ID 2858, American Beauty (1999), a reçu 3428 notes.
## Le film d'ID 260, Star Wars: Episode IV - A New Hope (1977), a reçu 2991 notes.
## Le film d'ID 1196, Star Wars: Episode V - The Empire Strikes Back (1980), a reçu 2990 notes.
## Le film d'ID 1210, Star Wars: Episode VI - Return of the Jedi (1983), a reçu 2883 notes.
## Le film d'ID 480, Jurassic Park (1993), a reçu 2672 notes.
## Le film d'ID 2028, Saving Private Ryan (1998), a reçu 2653 notes.
## Le film d'ID 589, Terminator 2: Judgment Day (1991), a reçu 2649 notes.
## Le film d'ID 2571, Matrix, The (1999), a reçu 2590 notes.
## Le film d'ID 1270, Back to the Future (1985), a reçu 2583 notes.
## Le film d'ID 593, Silence of the Lambs, The (1991), a reçu 2578 notes.
## Le film d'ID 1580, Men in Black (1997), a reçu 2538 notes.
## Le film d'ID 1198, Raiders of the Lost Ark (1981), a reçu 2514 notes.
## Le film d'ID 608, Fargo (1996), a reçu 2513 notes.
## Le film d'ID 2762, Sixth Sense, The (1999), a reçu 2459 notes.
```


Le film d'ID 110, Braveheart (1995), a reçu 2443 notes.
 ## Le film d'ID 2396, Shakespeare in Love (1998), a reçu 2369 notes.
 ## Le film d'ID 1197, Princess Bride, The (1987), a reçu 2318 notes.
 ## Le film d'ID 527, Schindler's List (1993), a reçu 2304 notes.
 ## Le film d'ID 1617, L.A. Confidential (1997), a reçu 2288 notes.
 ## Le film d'ID 1265, Groundhog Day (1993), a reçu 2278 notes.
 ## Le film d'ID 1097, E.T. the Extra-Terrestrial (1982), a reçu 2269 notes.
 ## Le film d'ID 2628, Star Wars: Episode I - The Phantom Menace (1999), a reçu 2250 notes.
 ## Le film d'ID 2997, Being John Malkovich (1999), a reçu 2241 notes.
 ## Le film d'ID 318, Shawshank Redemption, The (1994), a reçu 2227 notes.
 ## Le film d'ID 858, Godfather, The (1972), a reçu 2223 notes.

Films les moins notés

Les 25 films ayant reçus le moins de notes :

Le film d'ID 127, Silence of the Palace, The (Saint el Qusur) (1994), a reçu 1 note(s).
 ## Le film d'ID 133, Nueba Yol (1995), a reçu 1 note(s).
 ## Le film d'ID 139, Target (1995), a reçu 1 note(s).
 ## Le film d'ID 142, Shadows (Cienie) (1988), a reçu 1 note(s).
 ## Le film d'ID 226, Dream Man (1995), a reçu 1 note(s).
 ## Le film d'ID 286, Nemesis 2: Nebula (1995), a reçu 1 note(s).
 ## Le film d'ID 311, Relative Fear (1994), a reçu 1 note(s).
 ## Le film d'ID 396, Fall Time (1995), a reçu 1 note(s).
 ## Le film d'ID 398, Frank and Ollie (1995), a reçu 1 note(s).
 ## Le film d'ID 402, Open Season (1996), a reçu 1 note(s).
 ## Le film d'ID 545, Harlem (1993), a reçu 1 note(s).
 ## Le film d'ID 576, Fausto (1993), a reçu 1 note(s).
 ## Le film d'ID 579, Scorta, La (1993), a reçu 1 note(s).
 ## Le film d'ID 584, I Don't Want to Talk About It (De eso no se habla) (1993), a reçu 1 note(s).
 ## Le film d'ID 601, Wooden Man's Bride, The (Wu Kui) (1994), a reçu 1 note(s).
 ## Le film d'ID 624, Condition Red (1995), a reçu 1 note(s).
 ## Le film d'ID 641, Little Indian, Big City (Un indien dans la ville) (1994), a reçu 1 note(s).
 ## Le film d'ID 642, Roula (1995), a reçu 1 note(s).
 ## Le film d'ID 644, Happy Weekend (1996), a reçu 1 note(s).
 ## Le film d'ID 651, Superweib, Das (1996), a reçu 1 note(s).
 ## Le film d'ID 655, Mutters Courage (1995), a reçu 1 note(s).
 ## Le film d'ID 658, Billy's Holiday (1995), a reçu 1 note(s).
 ## Le film d'ID 684, Windows (1980), a reçu 1 note(s).
 ## Le film d'ID 701, Daens (1992), a reçu 1 note(s).
 ## Le film d'ID 717, Mouth to Mouth (Boca a boca) (1995), a reçu 1 note(s).

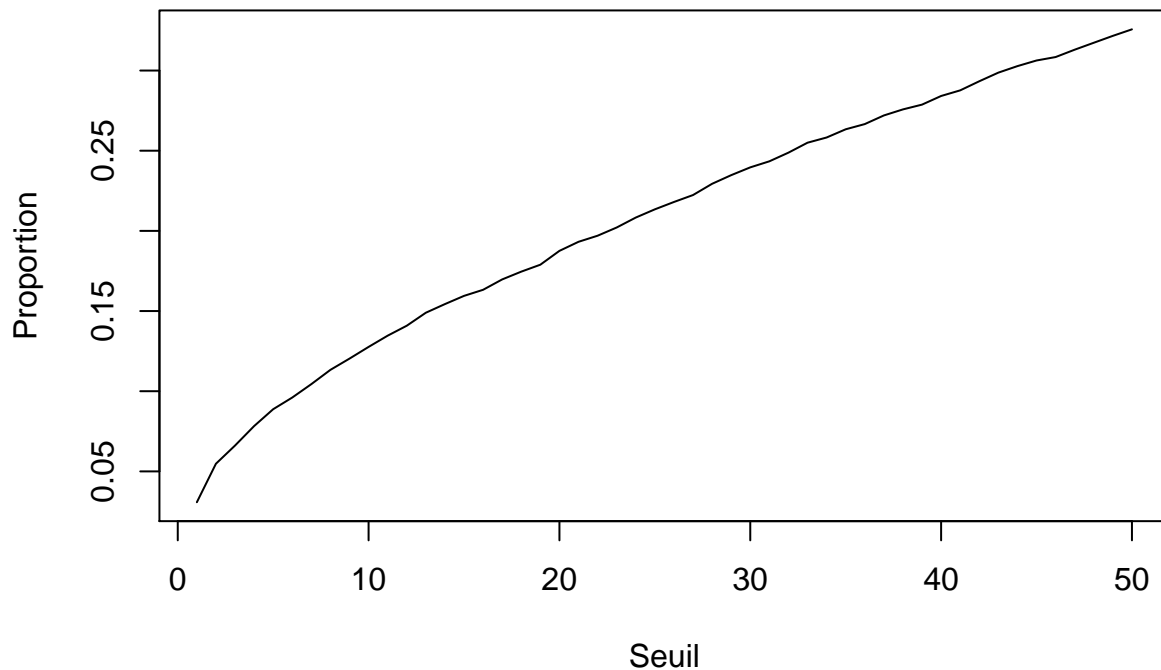
Films n'ayant reçu qu'une seule note

Il y a 114 films qui n'ont reçu qu'une seule note.

Films ayant reçu moins de n votes

Il y a 473 films qui ont reçu moins de 10 votes.

Proportion de films notés en dessous d'un certain seuil



Films ayant les meilleurs notes

Les 25 meilleurs films :

Le film d'ID 787, Gate of Heavenly Peace, The (1995), a une note moyenne de 5.00 sur 3 vote(s).
Le film d'ID 989, Schlafes Bruder (Brother of Sleep) (1995), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 1830, Follow the Bitch (1998), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 3172, Ulysses (Ulysse) (1954), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 3233, Smashing Time (1967), a une note moyenne de 5.00 sur 2 vote(s).
Le film d'ID 3280, Baby, The (1973), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 3382, Song of Freedom (1936), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 3607, One Little Indian (1973), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 3656, Lured (1947), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 3881, Bittersweet Motel (2000), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 3245, I Am Cuba (Soy Cuba/Ya Kuba) (1964), a une note moyenne de 4.80 sur 5 vote(s).
Le film d'ID 53, Lamerica (1994), a une note moyenne de 4.75 sur 8 vote(s).
Le film d'ID 2503, Apple, The (Sib) (1998), a une note moyenne de 4.67 sur 9 vote(s).
Le film d'ID 2905, Sanjuro (1962), a une note moyenne de 4.61 sur 69 vote(s).
Le film d'ID 2019, Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954), a une note m
Le film d'ID 318, Shawshank Redemption, The (1994), a une note moyenne de 4.55 sur 2227 vote(s).
Le film d'ID 50, Usual Suspects, The (1995), a une note moyenne de 4.52 sur 1783 vote(s).
Le film d'ID 745, Close Shave, A (1995), a une note moyenne de 4.52 sur 657 vote(s).
Le film d'ID 858, Godfather, The (1972), a une note moyenne de 4.52 sur 2223 vote(s).
Le film d'ID 527, Schindler's List (1993), a une note moyenne de 4.51 sur 2304 vote(s).
Le film d'ID 1148, Wrong Trousers, The (1993), a une note moyenne de 4.51 sur 882 vote(s).

Le film d'ID 439, Dangerous Game (1993), a une note moyenne de 4.50 sur 2 vote(s).
Le film d'ID 557, Mamma Roma (1962), a une note moyenne de 4.50 sur 2 vote(s).
Le film d'ID 578, Hour of the Pig, The (1993), a une note moyenne de 4.50 sur 2 vote(s).
Le film d'ID 1795, Callejón de los milagros, El (1995), a une note moyenne de 4.50 sur 2 vote(s).

Nombre de films ayant une note moyenne de 5 (note maximale)

Il y a 10 films qui ont reçu une note moyenne de 5.

Films ayant les meilleurs notes et dépassant un seuil de visionnage

Les 25 meilleurs films ayant reçu plus de 10 notes:

Le film d'ID 2905, Sanjuro (1962), a une note moyenne de 4.61 sur 69 votes.
Le film d'ID 2019, Seven Samurai (The Magnificent Seven) (Shichinin no samurai) (1954), a une note moyenne de 4.60 sur 100 votes.
Le film d'ID 318, Shawshank Redemption, The (1994), a une note moyenne de 4.55 sur 2227 votes.
Le film d'ID 50, Usual Suspects, The (1995), a une note moyenne de 4.52 sur 1783 votes.
Le film d'ID 745, Close Shave, A (1995), a une note moyenne de 4.52 sur 657 votes.
Le film d'ID 858, Godfather, The (1972), a une note moyenne de 4.52 sur 2223 votes.
Le film d'ID 527, Schindler's List (1993), a une note moyenne de 4.51 sur 2304 votes.
Le film d'ID 1148, Wrong Trousers, The (1993), a une note moyenne de 4.51 sur 882 votes.
Le film d'ID 922, Sunset Blvd. (a.k.a. Sunset Boulevard) (1950), a une note moyenne de 4.49 sur 470 votes.
Le film d'ID 904, Rear Window (1954), a une note moyenne de 4.48 sur 1050 votes.
Le film d'ID 1198, Raiders of the Lost Ark (1981), a une note moyenne de 4.48 sur 2514 votes.
Le film d'ID 1178, Paths of Glory (1957), a une note moyenne de 4.47 sur 230 votes.
Le film d'ID 260, Star Wars: Episode IV - A New Hope (1977), a une note moyenne de 4.45 sur 2991 votes.
Le film d'ID 750, Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963), a une note moyenne de 4.45 sur 1000 votes.
Le film d'ID 1212, Third Man, The (1949), a une note moyenne de 4.45 sur 480 votes.
Le film d'ID 3338, For All Mankind (1989), a une note moyenne de 4.44 sur 27 votes.
Le film d'ID 720, Wallace & Gromit: The Best of Aardman Animation (1996), a une note moyenne de 4.43 sur 100 votes.
Le film d'ID 1207, To Kill a Mockingbird (1962), a une note moyenne de 4.43 sur 928 votes.
Le film d'ID 3435, Double Indemnity (1944), a une note moyenne de 4.42 sur 551 votes.
Le film d'ID 670, World of Apu, The (Apu Sansar) (1959), a une note moyenne de 4.41 sur 56 votes.
Le film d'ID 912, Casablanca (1942), a une note moyenne de 4.41 sur 1669 votes.
Le film d'ID 2762, Sixth Sense, The (1999), a une note moyenne de 4.41 sur 2459 votes.
Le film d'ID 668, Pather Panchali (1955), a une note moyenne de 4.40 sur 47 votes.
Le film d'ID 913, Maltese Falcon, The (1941), a une note moyenne de 4.40 sur 1043 votes.
Le film d'ID 1204, Lawrence of Arabia (1962), a une note moyenne de 4.40 sur 831 votes.

Films ayant les pires notes

Les 10 pires films :

Le film d'ID 127, Silence of the Palace, The (Saint el Qusur) (1994), a une note moyenne de 1.00 sur 1 vote(s).
Le film d'ID 133, Nueba Yol (1995), a une note moyenne de 1.00 sur 1 vote(s).
Le film d'ID 142, Shadows (Cienie) (1988), a une note moyenne de 1.00 sur 1 vote(s).
Le film d'ID 641, Little Indian, Big City (Un indien dans la ville) (1994), a une note moyenne de 1.00 sur 1 vote(s).
Le film d'ID 655, Muttters Courage (1995), a une note moyenne de 1.00 sur 1 vote(s).
Le film d'ID 684, Windows (1980), a une note moyenne de 1.00 sur 1 vote(s).
Le film d'ID 730, Low Life, The (1994), a une note moyenne de 1.00 sur 1 vote(s).
Le film d'ID 826, Diebinnen (1995), a une note moyenne de 1.00 sur 1 vote(s).
Le film d'ID 843, Lotto Land (1995), a une note moyenne de 1.00 sur 1 vote(s).
Le film d'ID 895, Venice/Venice (1992), a une note moyenne de 1.00 sur 2 vote(s).

Films ayant une note moyenne de 1

Il y a 31 films qui ont reçu une note moyenne de 1.

Films ayant les pires notes et dépassant un seuil de visionnage

Les 10 pires films ayant reçu plus de 10 notes:

Le film d'ID 3574, Carnosaur 3: Primal Species (1996), a une note moyenne de 1.06 sur 17 votes.
Le film d'ID 1495, Turbo: A Power Rangers Movie (1997), a une note moyenne de 1.32 sur 44 votes.
Le film d'ID 1739, 3 Ninjas: High Noon On Mega Mountain (1998), a une note moyenne de 1.36 sur 47 votes.
Le film d'ID 1323, Amityville 3-D (1983), a une note moyenne de 1.37 sur 43 votes.
Le film d'ID 3573, Carnosaur 2 (1995), a une note moyenne de 1.46 sur 26 votes.
Le film d'ID 810, Kazaam (1996), a une note moyenne de 1.47 sur 120 votes.
Le film d'ID 1989, Prom Night III: The Last Kiss (1989), a une note moyenne de 1.47 sur 15 votes.
Le film d'ID 3945, Digimon: The Movie (2000), a une note moyenne de 1.49 sur 43 votes.
Le film d'ID 2449, Garbage Pail Kids Movie, The (1987), a une note moyenne de 1.53 sur 32 votes.
Le film d'ID 1383, Adrenalin: Fear the Rush (1996), a une note moyenne de 1.55 sur 11 votes.

Sur les utilisateurs

Utilisateurs ayant le plus vote

Les 10 utilisateurs ayant note le plus de films :

L'utilisateur d'ID 4169 a vote 2314 fois.
L'utilisateur d'ID 1680 a vote 1850 fois.
L'utilisateur d'ID 4277 a vote 1743 fois.
L'utilisateur d'ID 1941 a vote 1595 fois.
L'utilisateur d'ID 1181 a vote 1521 fois.
L'utilisateur d'ID 889 a vote 1518 fois.
L'utilisateur d'ID 3618 a vote 1344 fois.
L'utilisateur d'ID 2063 a vote 1323 fois.
L'utilisateur d'ID 1150 a vote 1302 fois.
L'utilisateur d'ID 1015 a vote 1286 fois.

Utilisateurs ayant le moins vote

Les 15 utilisateurs ayant note le moins de films:

L'utilisateur d'ID 98 n'a vote que 20 fois.
L'utilisateur d'ID 160 n'a vote que 20 fois.
L'utilisateur d'ID 217 n'a vote que 20 fois.
L'utilisateur d'ID 250 n'a vote que 20 fois.
L'utilisateur d'ID 311 n'a vote que 20 fois.
L'utilisateur d'ID 341 n'a vote que 20 fois.
L'utilisateur d'ID 421 n'a vote que 20 fois.
L'utilisateur d'ID 572 n'a vote que 20 fois.
L'utilisateur d'ID 665 n'a vote que 20 fois.
L'utilisateur d'ID 703 n'a vote que 20 fois.
L'utilisateur d'ID 708 n'a vote que 20 fois.

```
## L'utilisateur d'ID 761 n'a vote que 20 fois.  
## L'utilisateur d'ID 932 n'a vote que 20 fois.  
## L'utilisateur d'ID 947 n'a vote que 20 fois.  
## L'utilisateur d'ID 1351 n'a vote que 20 fois.
```

Utilisateurs n'ayant vote moins que 20 fois

```
## Il y a 86 utilisateurs qui ont voté moins que 20 fois.
```

Utilisateurs ayant vote moins de n fois

```
## Il y a 1793 utilisateurs qui ont vote moins de 50 fois.
```

Utilisateurs ayant donne les plus hautes notes

```
## Les 20 utilisateurs les plus gentils :
```

```
## L'utilisateur d'ID 283 a donne en moyenne une note de 4.96 pour 27 notes.  
## L'utilisateur d'ID 2339 a donne en moyenne une note de 4.96 pour 23 notes.  
## L'utilisateur d'ID 3324 a donne en moyenne une note de 4.90 pour 21 notes.  
## L'utilisateur d'ID 3902 a donne en moyenne une note de 4.89 pour 165 notes.  
## L'utilisateur d'ID 446 a donne en moyenne une note de 4.84 pour 51 notes.  
## L'utilisateur d'ID 447 a donne en moyenne une note de 4.84 pour 37 notes.  
## L'utilisateur d'ID 4649 a donne en moyenne une note de 4.82 pour 33 notes.  
## L'utilisateur d'ID 4634 a donne en moyenne une note de 4.81 pour 102 notes.  
## L'utilisateur d'ID 1131 a donne en moyenne une note de 4.80 pour 103 notes.  
## L'utilisateur d'ID 4755 a donne en moyenne une note de 4.76 pour 25 notes.  
## L'utilisateur d'ID 4925 a donne en moyenne une note de 4.76 pour 42 notes.  
## L'utilisateur d'ID 5069 a donne en moyenne une note de 4.76 pour 25 notes.  
## L'utilisateur d'ID 682 a donne en moyenne une note de 4.73 pour 75 notes.  
## L'utilisateur d'ID 4801 a donne en moyenne une note de 4.73 pour 196 notes.  
## L'utilisateur d'ID 1670 a donne en moyenne une note de 4.71 pour 56 notes.  
## L'utilisateur d'ID 91 a donne en moyenne une note de 4.70 pour 44 notes.  
## L'utilisateur d'ID 3461 a donne en moyenne une note de 4.70 pour 37 notes.  
## L'utilisateur d'ID 5768 a donne en moyenne une note de 4.70 pour 74 notes.  
## L'utilisateur d'ID 5984 a donne en moyenne une note de 4.70 pour 30 notes.  
## L'utilisateur d'ID 953 a donne en moyenne une note de 4.69 pour 189 notes.
```

Utilisateurs ayant donnees les meilleurs notes et depassant un seuil de participation

```
## Les 20 utilisateurs ayant donne les meilleurs notes et ayant vote plus de 50 fois:
```

```
## L'utilisateur d'ID 3902 a donne en moyenne une note de 4.89 pour 165 notes.  
## L'utilisateur d'ID 446 a donne en moyenne une note de 4.84 pour 51 notes.  
## L'utilisateur d'ID 4634 a donne en moyenne une note de 4.81 pour 102 notes.  
## L'utilisateur d'ID 1131 a donne en moyenne une note de 4.80 pour 103 notes.  
## L'utilisateur d'ID 682 a donne en moyenne une note de 4.73 pour 75 notes.  
## L'utilisateur d'ID 4801 a donne en moyenne une note de 4.73 pour 196 notes.  
## L'utilisateur d'ID 1670 a donne en moyenne une note de 4.71 pour 56 notes.  
## L'utilisateur d'ID 5768 a donne en moyenne une note de 4.70 pour 74 notes.  
## L'utilisateur d'ID 953 a donne en moyenne une note de 4.69 pour 189 notes.
```

```

## L'utilisateur d'ID 1021 a donne en moyenne une note de 4.69 pour 131 notes.
## L'utilisateur d'ID 1856 a donne en moyenne une note de 4.69 pour 134 notes.
## L'utilisateur d'ID 2155 a donne en moyenne une note de 4.69 pour 71 notes.
## L'utilisateur d'ID 3596 a donne en moyenne une note de 4.69 pour 128 notes.
## L'utilisateur d'ID 101 a donne en moyenne une note de 4.68 pour 106 notes.
## L'utilisateur d'ID 2243 a donne en moyenne une note de 4.68 pour 60 notes.
## L'utilisateur d'ID 1940 a donne en moyenne une note de 4.67 pour 64 notes.
## L'utilisateur d'ID 3306 a donne en moyenne une note de 4.67 pour 51 notes.
## L'utilisateur d'ID 5542 a donne en moyenne une note de 4.66 pour 71 notes.
## L'utilisateur d'ID 661 a donne en moyenne une note de 4.65 pour 71 notes.
## L'utilisateur d'ID 5255 a donne en moyenne une note de 4.62 pour 149 notes.

```

Utilisateurs ayant donnees les plus faibles notes

Les 20 utilisateurs les plus severes :

```

## L'utilisateur d'ID 3598 a donne en moyenne une note de 1.02 pour 65 notes.
## L'utilisateur d'ID 4486 a donne en moyenne une note de 1.06 pour 51 notes.
## L'utilisateur d'ID 2744 a donne en moyenne une note de 1.30 pour 138 notes.
## L'utilisateur d'ID 4539 a donne en moyenne une note de 1.82 pour 119 notes.
## L'utilisateur d'ID 5850 a donne en moyenne une note de 1.84 pour 58 notes.
## L'utilisateur d'ID 5334 a donne en moyenne une note de 1.93 pour 55 notes.
## L'utilisateur d'ID 4349 a donne en moyenne une note de 1.96 pour 27 notes.
## L'utilisateur d'ID 4636 a donne en moyenne une note de 2.00 pour 25 notes.
## L'utilisateur d'ID 5686 a donne en moyenne une note de 2.05 pour 530 notes.
## L'utilisateur d'ID 3209 a donne en moyenne une note de 2.06 pour 230 notes.
## L'utilisateur d'ID 1608 a donne en moyenne une note de 2.08 pour 144 notes.
## L'utilisateur d'ID 4575 a donne en moyenne une note de 2.09 pour 125 notes.
## L'utilisateur d'ID 4916 a donne en moyenne une note de 2.09 pour 170 notes.
## L'utilisateur d'ID 1747 a donne en moyenne une note de 2.14 pour 108 notes.
## L'utilisateur d'ID 1340 a donne en moyenne une note de 2.16 pour 805 notes.
## L'utilisateur d'ID 1761 a donne en moyenne une note de 2.16 pour 113 notes.
## L'utilisateur d'ID 163 a donne en moyenne une note de 2.18 pour 514 notes.
## L'utilisateur d'ID 1100 a donne en moyenne une note de 2.20 pour 508 notes.
## L'utilisateur d'ID 5039 a donne en moyenne une note de 2.20 pour 360 notes.
## L'utilisateur d'ID 5944 a donne en moyenne une note de 2.20 pour 30 notes.

```

Utilisateurs ayant donnees les pires notes et depassant un seuil de participation

Les 20 utilisateurs ayant donne les pires notes et ayant vote plus de 50 fois:

```

## L'utilisateur d'ID 3598 a donne en moyenne une note de 1.02 pour 65 notes.
## L'utilisateur d'ID 4486 a donne en moyenne une note de 1.06 pour 51 notes.
## L'utilisateur d'ID 2744 a donne en moyenne une note de 1.30 pour 138 notes.
## L'utilisateur d'ID 4539 a donne en moyenne une note de 1.82 pour 119 notes.
## L'utilisateur d'ID 5850 a donne en moyenne une note de 1.84 pour 58 notes.
## L'utilisateur d'ID 5334 a donne en moyenne une note de 1.93 pour 55 notes.
## L'utilisateur d'ID 5686 a donne en moyenne une note de 2.05 pour 530 notes.
## L'utilisateur d'ID 3209 a donne en moyenne une note de 2.06 pour 230 notes.
## L'utilisateur d'ID 1608 a donne en moyenne une note de 2.08 pour 144 notes.
## L'utilisateur d'ID 4575 a donne en moyenne une note de 2.09 pour 125 notes.
## L'utilisateur d'ID 4916 a donne en moyenne une note de 2.09 pour 170 notes.

```

L'utilisateur d'ID 1747 a donne en moyenne une note de 2.14 pour 108 notes.
L'utilisateur d'ID 1340 a donne en moyenne une note de 2.16 pour 805 notes.
L'utilisateur d'ID 1761 a donne en moyenne une note de 2.16 pour 113 notes.
L'utilisateur d'ID 163 a donne en moyenne une note de 2.18 pour 514 notes.
L'utilisateur d'ID 1100 a donne en moyenne une note de 2.20 pour 508 notes.
L'utilisateur d'ID 5039 a donne en moyenne une note de 2.20 pour 360 notes.
L'utilisateur d'ID 2106 a donne en moyenne une note de 2.25 pour 900 notes.
L'utilisateur d'ID 1630 a donne en moyenne une note de 2.26 pour 117 notes.
L'utilisateur d'ID 203 a donne en moyenne une note de 2.29 pour 127 notes.

Etude par age

Etude par sexe

```
##      userID      sex      age      occupation      zip.code
## Min.   : 2      F: 0      Min.   : 1.00      7      : 540      22903 : 16
## 1st Qu.:1504    M:4331    1st Qu.:25.00    4      : 525      94114 : 13
## Median :2984                Median :25.00    0      : 479      10025 : 12
## Mean   :2982                Mean   :30.55    17      : 450      48104 : 12
## 3rd Qu.:4436                3rd Qu.:35.00    12      : 338      55104 : 11
## Max.   :6040                Max.   :56.00    1      : 319      55105 : 11
##                                     (Other):1680 (Other):4256
##      nb.Ratings      mean      sd      max
## Min.   : 20      Min.   :1.020      Min.   :0.120      Min.   :2.000
## 1st Qu.: 46      1st Qu.:3.430      1st Qu.:0.870      1st Qu.:5.000
## Median :102      Median :3.730      Median :1.000      Median :5.000
## Mean   : 174      Mean   :3.692      Mean   :1.009      Mean   :4.994
## 3rd Qu.:219      3rd Qu.:3.990      3rd Qu.:1.130      3rd Qu.:5.000
## Max.   :2314      Max.   :4.960      Max.   :1.860      Max.   :5.000
##
##      min      med
## Min.   :1.000      Min.   :1.000
## 1st Qu.:1.000      1st Qu.:4.000
## Median :1.000      Median :4.000
## Mean   :1.267      Mean   :3.829
## 3rd Qu.:1.000      3rd Qu.:4.000
## Max.   :4.000      Max.   :5.000
##
```

```
##      userID      sex      age      occupation      zip.code
## Min.   : 1      F:1709    Min.   : 1.00      4      :234      02138 : 10
## 1st Qu.:1535    M: 0      1st Qu.:25.00    0      :232      14850 : 8
## Median :3096                Median :25.00    1      :209      48104 : 7
## Mean   :3118                Mean   :30.86    7      :139      94117 : 7
## 3rd Qu.:4771                3rd Qu.:45.00    6      :102      10022 : 6
## Max.   :6039                Max.   :56.00    3      :100      48103 : 6
##                                     (Other):693 (Other):1665
##      nb.Ratings      mean      sd      max
## Min.   : 20.0      Min.   :1.840      Min.   :0.210      Min.   :4.000
## 1st Qu.: 39.0      1st Qu.:3.490      1st Qu.:0.870      1st Qu.:5.000
## Median : 77.0      Median :3.760      Median :1.000      Median :5.000
## Mean   :144.2      Mean   :3.729      Mean   :1.014      Mean   :4.998
## 3rd Qu.:175.0      3rd Qu.:4.020      3rd Qu.:1.140      3rd Qu.:5.000
## Max.   :1302.0      Max.   :4.960      Max.   :1.770      Max.   :5.000
##
##      min      med
## Min.   :1.000      Min.   :1.000
## 1st Qu.:1.000      1st Qu.:4.000
## Median :1.000      Median :4.000
## Mean   :1.287      Mean   :3.872
## 3rd Qu.:1.000      3rd Qu.:4.000
## Max.   :4.000      Max.   :5.000
##
```

Etude par genre de films

Nombre de films en commun

Etude sur les similarités

Corrélation de Pearson

Distance euclidienne

Distance manhattan