

Filtrage collaboratif - Analyse de ml-100k

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 que 20 fois (ou moins ...)	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	27
Corrélation de Pearson	27
Distance euclidienne	28
Distance manhattan	29

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 99684 notes.  
  
## La base de données des notes contient exactement 943 utilisateurs uniques.  
  
## La base de données contient exactement 1663 films uniques.  
  
## Le taux de complétion de la matrice des notes Y est de 0.06 %
```

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

```
## La base de données des utilisateurs contient bien 943 utilisateurs.
```

Nombre de films

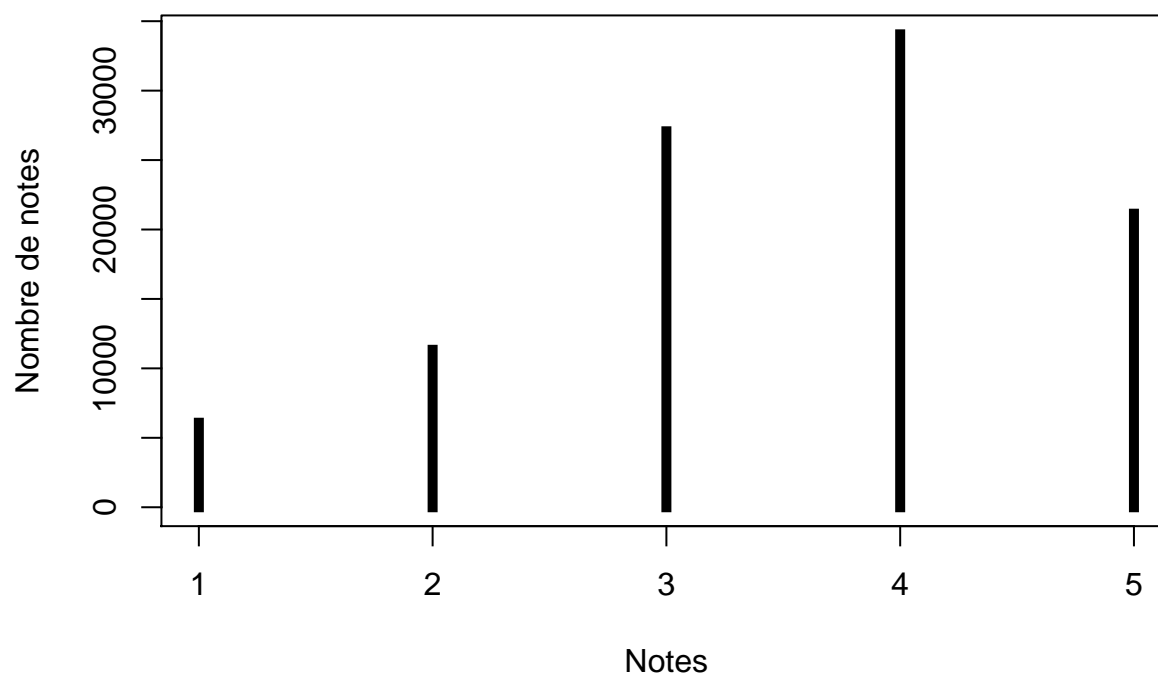
```
## La base de données contient bien 1663 films.
```

Statistique univariée

Sur les notes

```
## La moyenne des 99684 notes est de 3.53 ; l'écart-type est de 1.13.
```

Distribution des notes



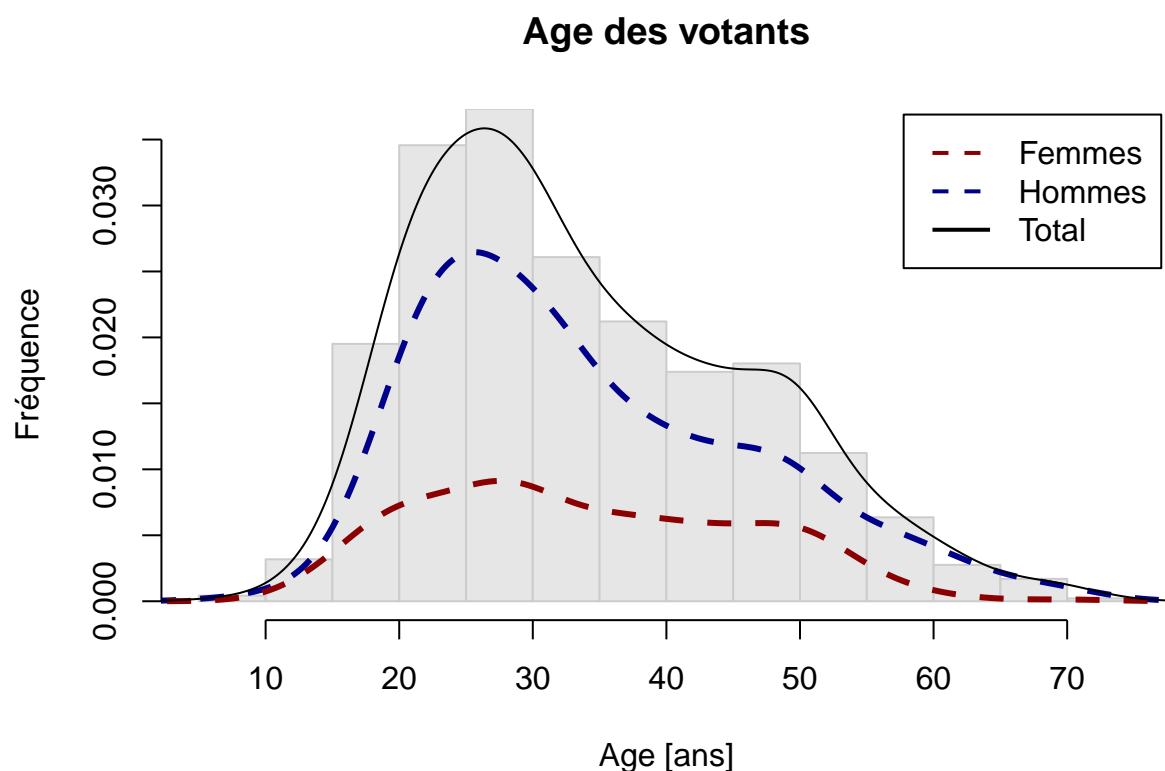
Effectifs par notes :

```
##
##      1      2      3      4      5
## 6085 11337 27072 34053 21137
```

Sur les utilisateurs

```
## La moyenne des âges est 34.05 ans avec un écart-type de 12.19 ans.
```

```
## La base de données est composée de 670 hommes, soit 71.05% et 273 femmes, soit 28.95%.
```



Proportions dans les métiers :

```
##          summary(data.Users$occupation)
## administrator          79
## artist                 28
## doctor                 7
## educator              95
## engineer              67
## entertainment         18
## executive             32
## healthcare            16
## homemaker             7
## lawyer                12
## librarian            51
## marketing             26
## none                  9
## other                105
## programmer            66
## retired              14
## salesman             12
## scientist             31
## student             196
## technician           27
## writer               45
```

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

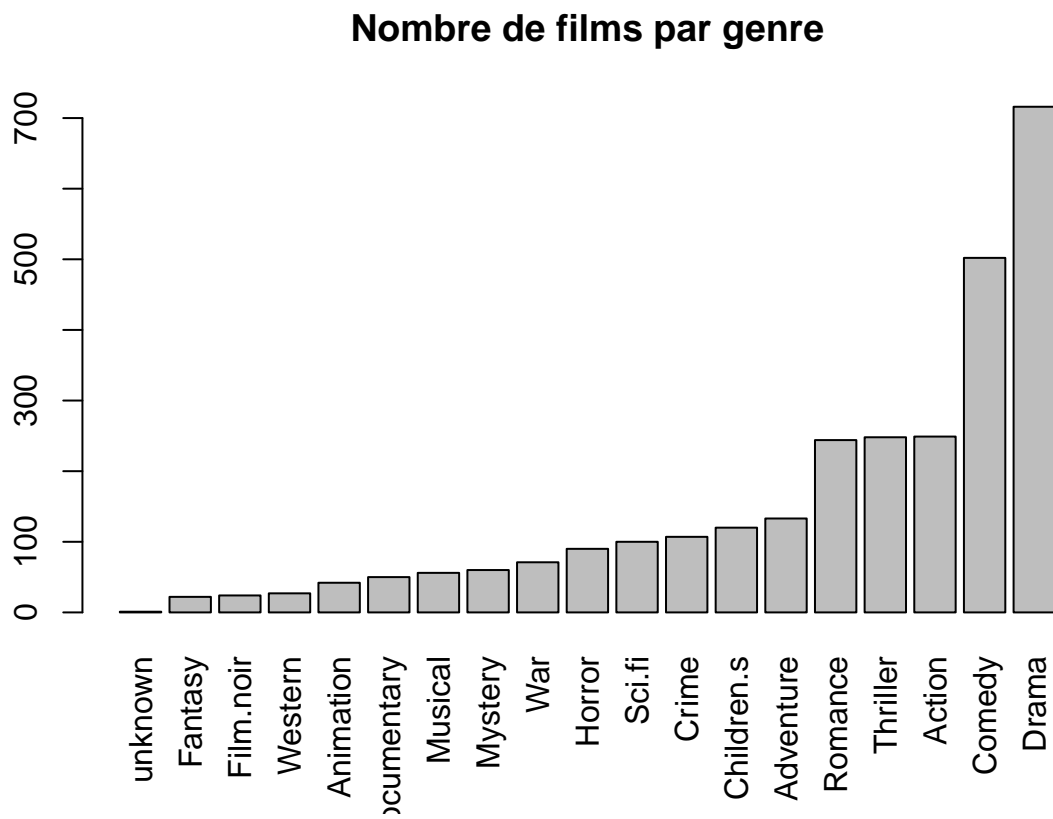
Proportions dans les codes postales :

```
##          summary(data.Users$zip.code)
## 55414          9
## 55105          6
## 10003          5
## 20009          5
## 55337          5
## 27514          4
## 55408          4
## 55454          4
## 02215          3
## 10021          3
## 11217          3
## 14216          3
## 22902          3
## 22903          3
## 48103          3
## 55104          3
## 55106          3
## 55108          3
## 55113          3
## 55409          3
## 60201          3
## 61801          3
## 61820          3
## 62901          3
## 63108          3
## 80525          3
## 94043          3
## 97301          3
## 00000          2
## 01720          2
## 01810          2
## 02139          2
## 02146          2
## 02154          2
## 10016          2
## 10019          2
## 10022          2
## 10025          2
## 10960          2
## 11231          2
## 14627          2
## 20003          2
## 20755          2
## 21114          2
## 21218          2
## 22206          2
## 22911          2
## 30067          2
## 32301          2
## 33884          2
## 38115          2
```

## 43212	2
## 44212	2
## 44224	2
## 45218	2
## 46260	2
## 48118	2
## 48197	2
## 49512	2
## 50613	2
## 53211	2
## 53706	2
## 53711	2
## 55116	2
## 55346	2
## 55369	2
## 55406	2
## 55422	2
## 60089	2
## 60402	2
## 60630	2
## 61755	2
## 68767	2
## 77042	2
## 77845	2
## 78212	2
## 78704	2
## 78741	2
## 80123	2
## 80526	2
## 83702	2
## 83814	2
## 85282	2
## 85711	2
## 90034	2
## 90210	2
## 91201	2
## 91206	2
## 91344	2
## 91711	2
## 92626	2
## 92629	2
## 93555	2
## 94086	2
## 94306	2
## 94533	2
## 94618	2
## 98006	2
## 98027	2
## (Other)	699

Remarque : le code postal n'est pas du tout pertinent (699 other)

Sur les films

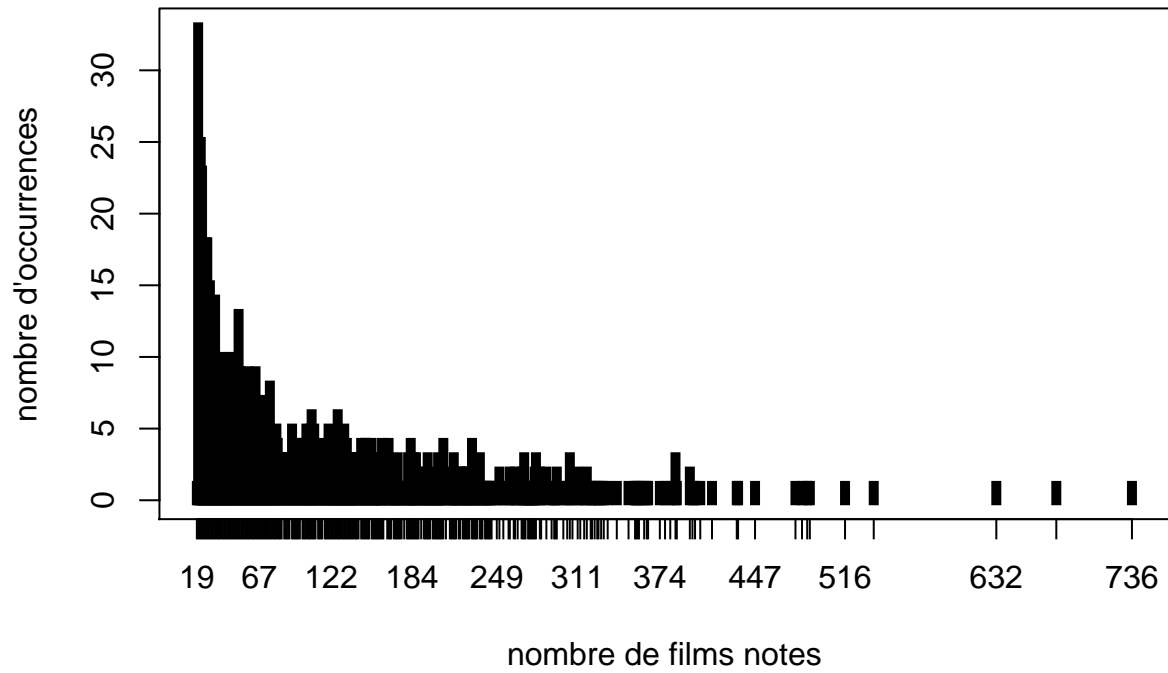


Statistique bivariée

Statistique par utilisateur

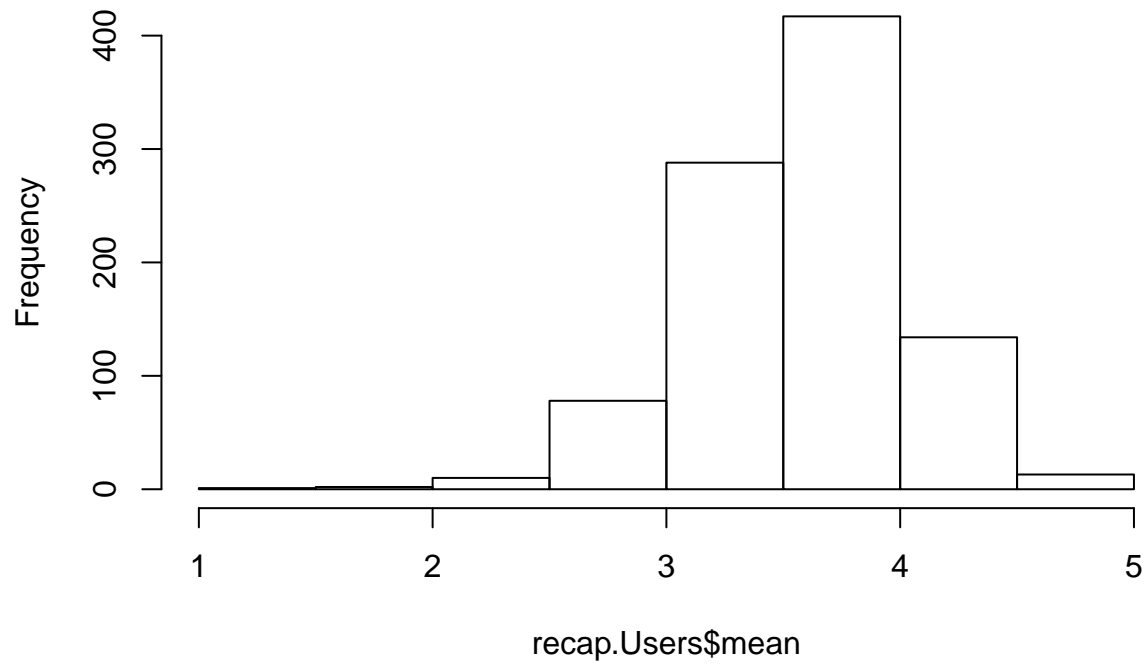
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	19.0	33.0	64.0	105.7	147.5	736.0

Nombre de films notes par utilisateur



##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.490	3.320	3.620	3.588	3.865	4.870

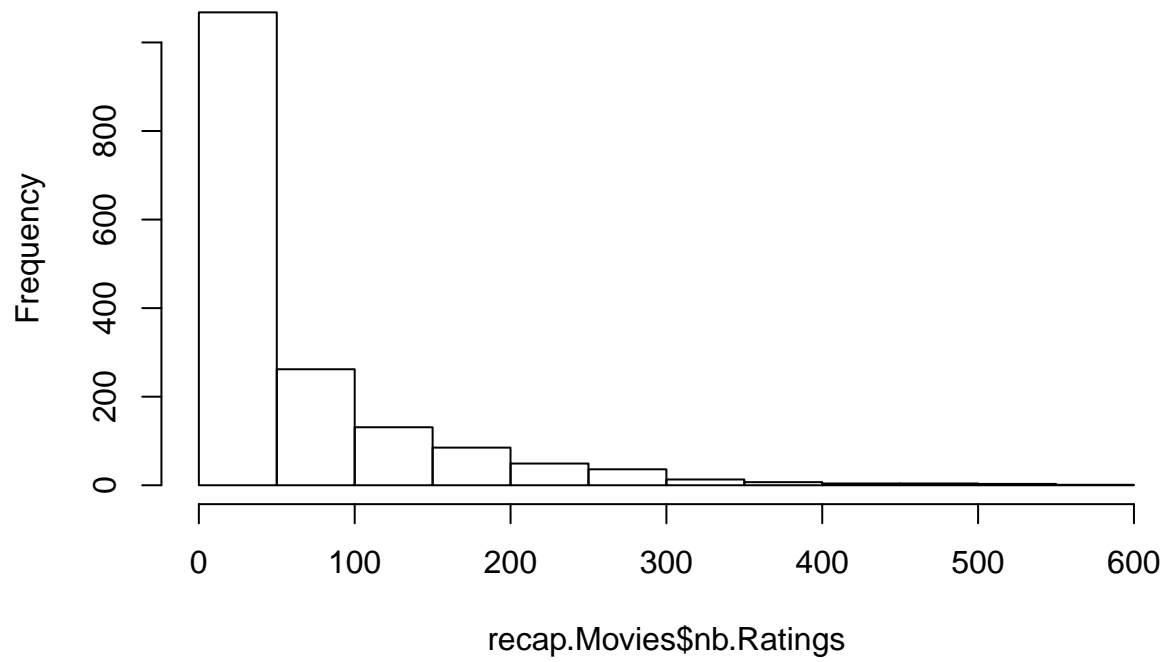
Note moyenne donnee par utilisateur



Statistique par film

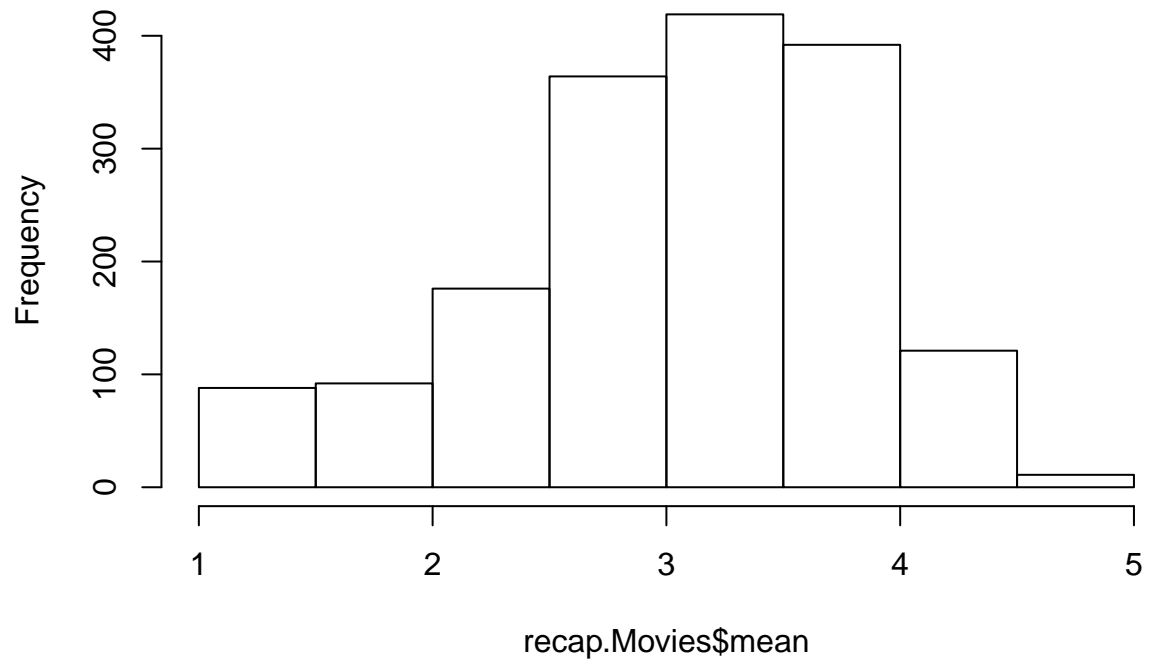
##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	1.00	7.00	27.00	59.94	80.00	583.00

Nombre de notes par film



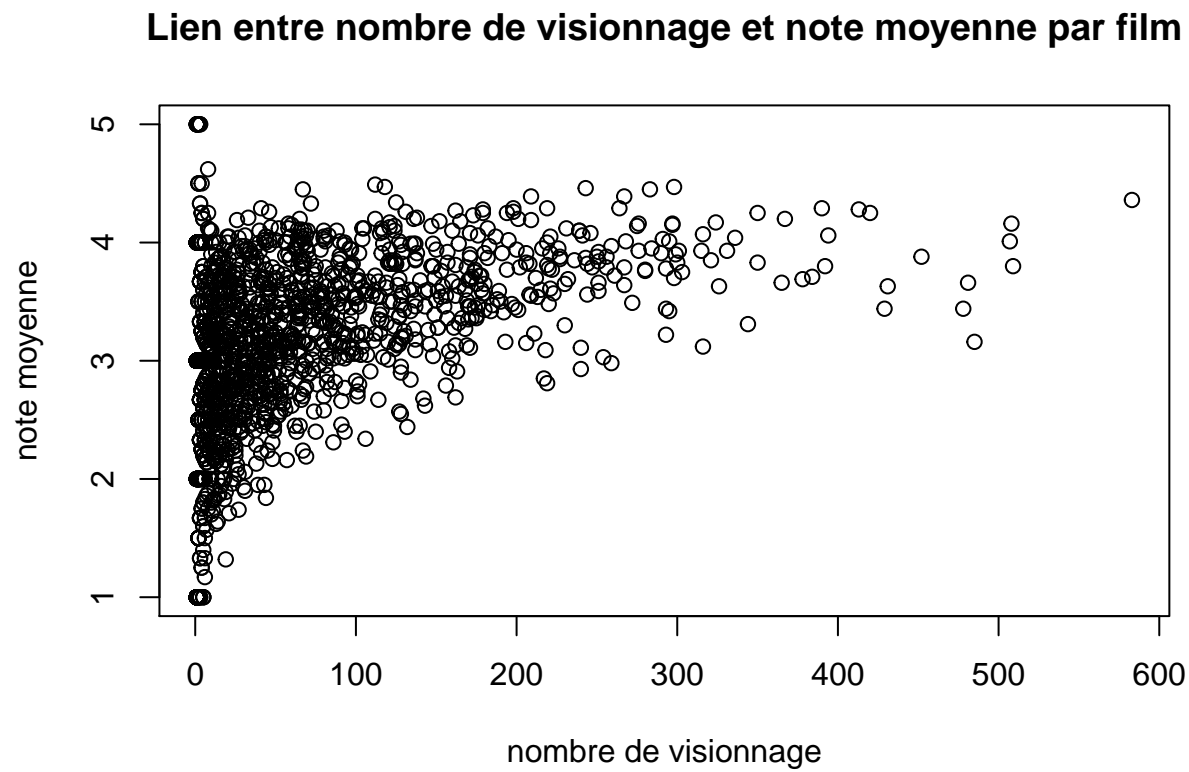
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  1.000   2.665   3.160   3.077   3.650   5.000
```

Note moyenne des films



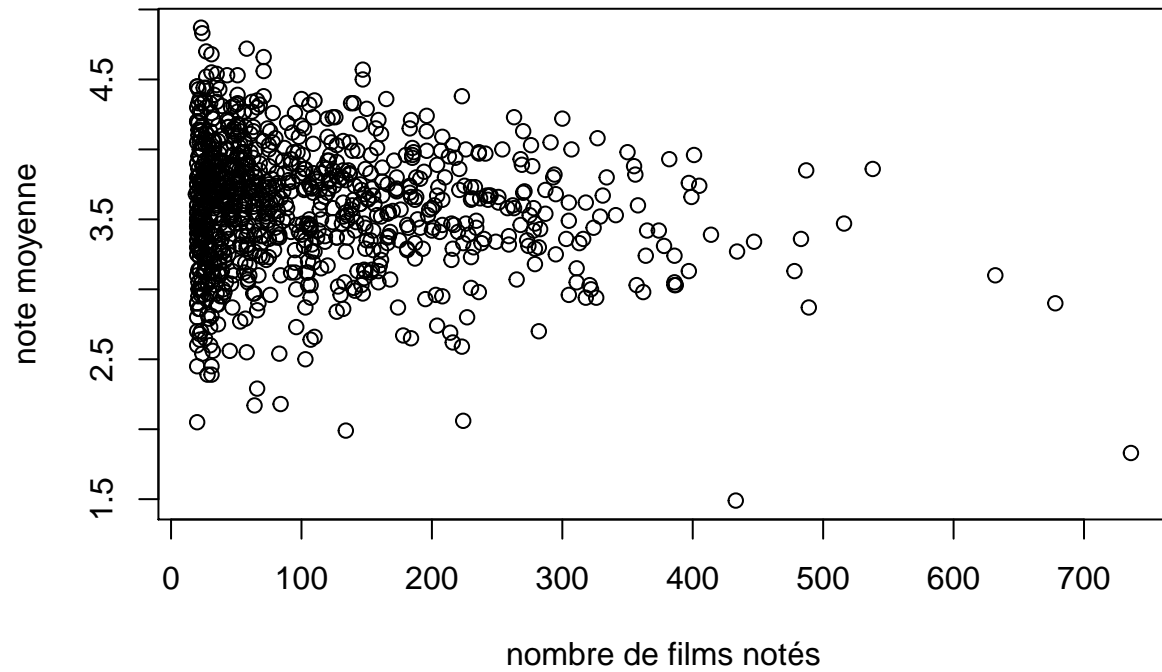
Corrélations entre les utilisateurs et les films

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



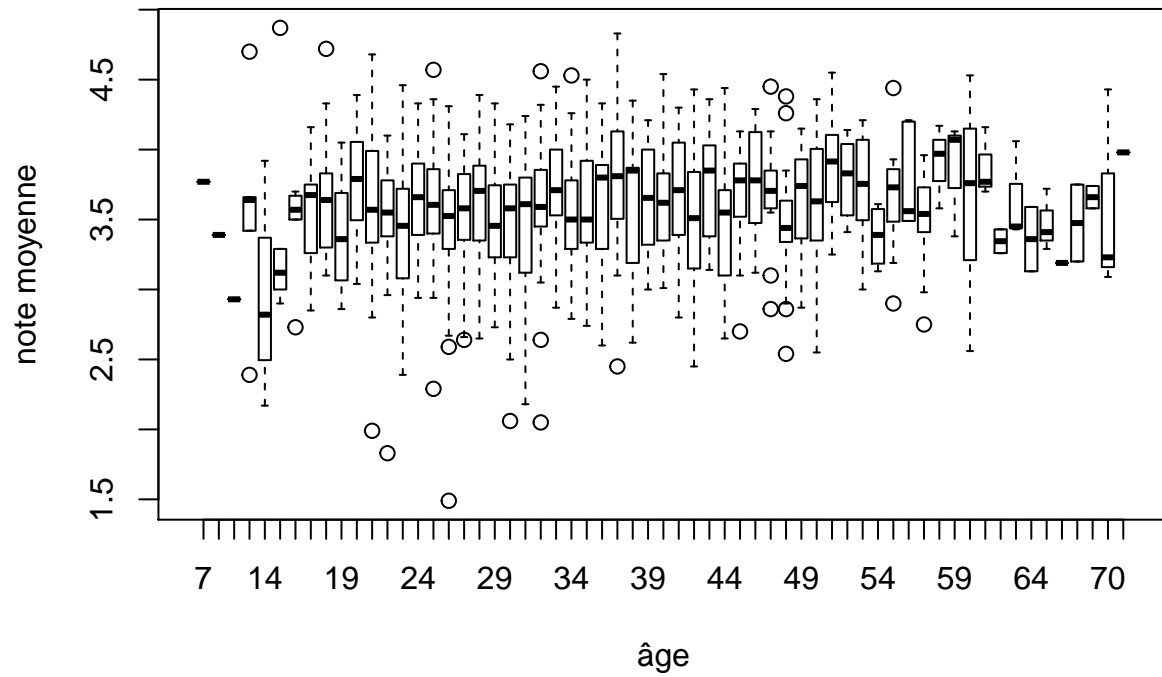
Corrélation entre note attribuée et nombre de films vus

Lien entre nombre de films notés et note moyenne par utilisateur

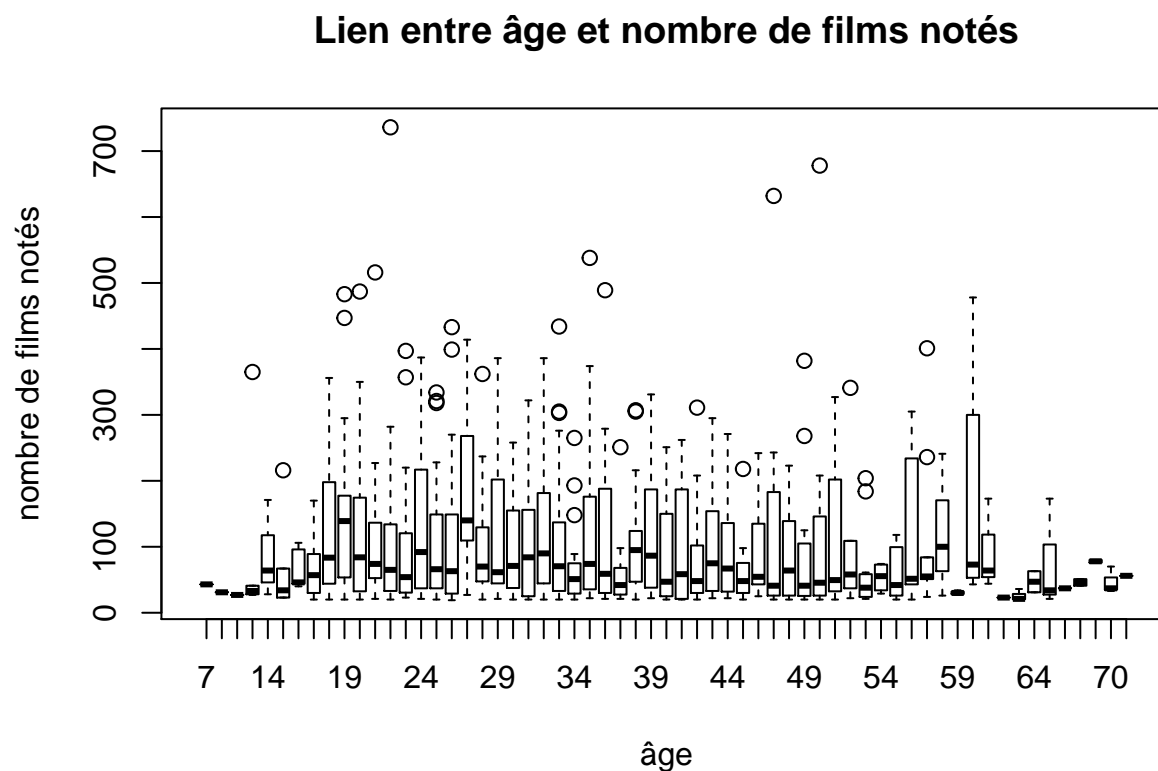


Corrélation entre âge et note moyenne

Lien 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 50, Star Wars (1977), a reçu 583 notes.  
## Le film d'ID 258, Contact (1997), a reçu 509 notes.  
## Le film d'ID 100, Fargo (1996), a reçu 508 notes.  
## Le film d'ID 181, Return of the Jedi (1983), a reçu 507 notes.  
## Le film d'ID 294, Liar Liar (1997), a reçu 485 notes.  
## Le film d'ID 286, English Patient, The (1996), a reçu 481 notes.  
## Le film d'ID 288, Scream (1996), a reçu 478 notes.  
## Le film d'ID 1, Toy Story (1995), a reçu 452 notes.  
## Le film d'ID 300, Air Force One (1997), a reçu 431 notes.  
## Le film d'ID 121, Independence Day (ID4) (1996), a reçu 429 notes.  
## Le film d'ID 174, Raiders of the Lost Ark (1981), a reçu 420 notes.  
## Le film d'ID 127, Godfather, The (1972), a reçu 413 notes.  
## Le film d'ID 56, Pulp Fiction (1994), a reçu 394 notes.  
## Le film d'ID 7, Twelve Monkeys (1995), a reçu 392 notes.
```


Le film d'ID 98, Silence of the Lambs, The (1991), a reçu 390 notes.
 ## Le film d'ID 237, Jerry Maguire (1996), a reçu 384 notes.
 ## Le film d'ID 117, Rock, The (1996), a reçu 378 notes.
 ## Le film d'ID 172, Empire Strikes Back, The (1980), a reçu 367 notes.
 ## Le film d'ID 222, Star Trek: First Contact (1996), a reçu 365 notes.
 ## Le film d'ID 204, Back to the Future (1985), a reçu 350 notes.
 ## Le film d'ID 313, Titanic (1997), a reçu 350 notes.
 ## Le film d'ID 405, Mission: Impossible (1996), a reçu 344 notes.
 ## Le film d'ID 79, Fugitive, The (1993), a reçu 336 notes.
 ## Le film d'ID 210, Indiana Jones and the Last Crusade (1989), a reçu 331 notes.
 ## Le film d'ID 151, Willy Wonka and the Chocolate Factory (1971), a reçu 326 notes.

Films les moins notés

Les 25 films ayant reçus le moins de notes :

Le film d'ID 599, Police Story 4: Project S (Chao ji ji hua) (1993), a reçu 1 note(s).
 ## Le film d'ID 677, Fire on the Mountain (1996), a reçu 1 note(s).
 ## Le film d'ID 814, Great Day in Harlem, A (1994), a reçu 1 note(s).
 ## Le film d'ID 830, Power 98 (1995), a reçu 1 note(s).
 ## Le film d'ID 852, Bloody Child, The (1996), a reçu 1 note(s).
 ## Le film d'ID 857, Paris Was a Woman (1995), a reçu 1 note(s).
 ## Le film d'ID 1122, They Made Me a Criminal (1939), a reçu 1 note(s).
 ## Le film d'ID 1130, Jupiter's Wife (1994), a reçu 1 note(s).
 ## Le film d'ID 1156, Cyclo (1995), a reçu 1 note(s).
 ## Le film d'ID 1201, Marlene Dietrich: Shadow and Light (1996) , a reçu 1 note(s).
 ## Le film d'ID 1235, Big Bang Theory, The (1994), a reçu 1 note(s).
 ## Le film d'ID 1236, Other Voices, Other Rooms (1997), a reçu 1 note(s).
 ## Le film d'ID 1309, Very Natural Thing, A (1974), a reçu 1 note(s).
 ## Le film d'ID 1310, Walk in the Sun, A (1945), a reçu 1 note(s).
 ## Le film d'ID 1320, Homage (1995), a reçu 1 note(s).
 ## Le film d'ID 1325, August (1996), a reçu 1 note(s).
 ## Le film d'ID 1329, Low Life, The (1994), a reçu 1 note(s).
 ## Le film d'ID 1339, Stefano Quantestorie (1993), a reçu 1 note(s).
 ## Le film d'ID 1340, Crude Oasis, The (1995), a reçu 1 note(s).
 ## Le film d'ID 1341, Hedd Wyn (1992), a reçu 1 note(s).
 ## Le film d'ID 1343, Lotto Land (1995), a reçu 1 note(s).
 ## Le film d'ID 1348, Every Other Weekend (1990), a reçu 1 note(s).
 ## Le film d'ID 1349, Mille bolle blu (1993), a reçu 1 note(s).
 ## Le film d'ID 1352, Shadow of Angels (Schatten der Engel) (1976), a reçu 1 note(s).
 ## Le film d'ID 1363, Leopard Son, The (1996), a reçu 1 note(s).

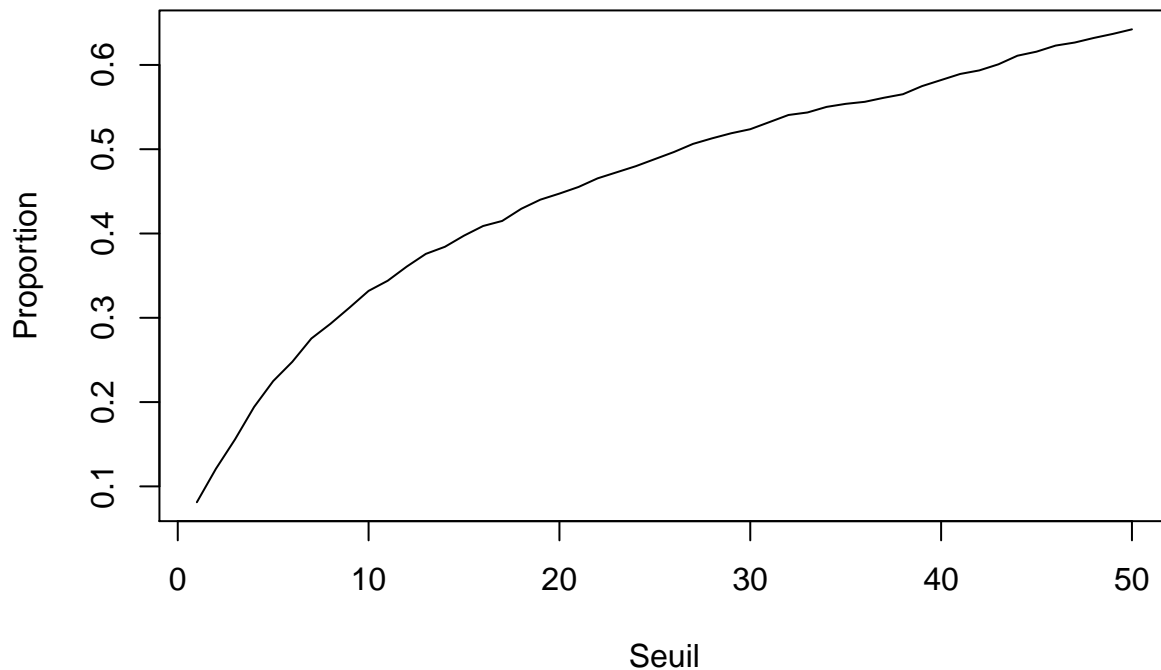
Films n'ayant reçu qu'une seule note

Il y a 135 films qui n'ont reçu qu'une seule note.

Films ayant reçu moins de n votes

Il y a 552 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 814, Great Day in Harlem, A (1994), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 1122, They Made Me a Criminal (1939), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 1189, Prefontaine (1997), a une note moyenne de 5.00 sur 3 vote(s).
Le film d'ID 1201, Marlene Dietrich: Shadow and Light (1996) , a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 1293, Star Kid (1997), a une note moyenne de 5.00 sur 3 vote(s).
Le film d'ID 1467, Saint of Fort Washington, The (1993), a une note moyenne de 5.00 sur 2 vote(s).
Le film d'ID 1500, Santa with Muscles (1996), a une note moyenne de 5.00 sur 2 vote(s).
Le film d'ID 1536, Aiqing wansui (1994), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 1599, Someone Else's America (1995), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 1653, Entertaining Angels: The Dorothy Day Story (1996), a une note moyenne de 5.00 sur 1 vote(s).
Le film d'ID 1449, Pather Panchali (1955), a une note moyenne de 4.62 sur 8 vote(s).
Le film d'ID 119, Maya Lin: A Strong Clear Vision (1994), a une note moyenne de 4.50 sur 4 vote(s).
Le film d'ID 1398, Anna (1996), a une note moyenne de 4.50 sur 2 vote(s).
Le film d'ID 1594, Everest (1998), a une note moyenne de 4.50 sur 2 vote(s).
Le film d'ID 1642, Some Mother's Son (1996), a une note moyenne de 4.50 sur 2 vote(s).
Le film d'ID 408, Close Shave, A (1995), a une note moyenne de 4.49 sur 112 vote(s).
Le film d'ID 169, Wrong Trousers, The (1993), a une note moyenne de 4.47 sur 118 vote(s).
Le film d'ID 318, Schindler's List (1993), a une note moyenne de 4.47 sur 298 vote(s).
Le film d'ID 483, Casablanca (1942), a une note moyenne de 4.46 sur 243 vote(s).
Le film d'ID 64, Shawshank Redemption, The (1994), a une note moyenne de 4.45 sur 283 vote(s).
Le film d'ID 114, Wallace & Gromit: The Best of Aardman Animation (1996), a une note moyenne de 4.45 sur 118 vote(s).

Le film d'ID 12, Usual Suspects, The (1995), a une note moyenne de 4.39 sur 267 vote(s).
Le film d'ID 603, Rear Window (1954), a une note moyenne de 4.39 sur 209 vote(s).
Le film d'ID 50, Star Wars (1977), a une note moyenne de 4.36 sur 583 vote(s).
Le film d'ID 178, 12 Angry Men (1957), a une note moyenne de 4.34 sur 125 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 408, Close Shave, A (1995), a une note moyenne de 4.49 sur 112 votes.
Le film d'ID 169, Wrong Trousers, The (1993), a une note moyenne de 4.47 sur 118 votes.
Le film d'ID 318, Schindler's List (1993), a une note moyenne de 4.47 sur 298 votes.
Le film d'ID 483, Casablanca (1942), a une note moyenne de 4.46 sur 243 votes.
Le film d'ID 64, Shawshank Redemption, The (1994), a une note moyenne de 4.45 sur 283 votes.
Le film d'ID 114, Wallace & Gromit: The Best of Aardman Animation (1996), a une note moyenne de 4.45 sur 114 votes.
Le film d'ID 12, Usual Suspects, The (1995), a une note moyenne de 4.39 sur 267 votes.
Le film d'ID 603, Rear Window (1954), a une note moyenne de 4.39 sur 209 votes.
Le film d'ID 50, Star Wars (1977), a une note moyenne de 4.36 sur 583 votes.
Le film d'ID 178, 12 Angry Men (1957), a une note moyenne de 4.34 sur 125 votes.
Le film d'ID 513, Third Man, The (1949), a une note moyenne de 4.33 sur 72 votes.
Le film d'ID 98, Silence of the Lambs, The (1991), a une note moyenne de 4.29 sur 390 votes.
Le film d'ID 134, Citizen Kane (1941), a une note moyenne de 4.29 sur 198 votes.
Le film d'ID 357, One Flew Over the Cuckoo's Nest (1975), a une note moyenne de 4.29 sur 264 votes.
Le film d'ID 427, To Kill a Mockingbird (1962), a une note moyenne de 4.29 sur 219 votes.
Le film d'ID 963, Some Folks Call It a Sling Blade (1993), a une note moyenne de 4.29 sur 41 votes.
Le film d'ID 127, Godfather, The (1972), a une note moyenne de 4.28 sur 413 votes.
Le film d'ID 480, North by Northwest (1959), a une note moyenne de 4.28 sur 179 votes.
Le film d'ID 285, Secrets & Lies (1996), a une note moyenne de 4.27 sur 162 votes.
Le film d'ID 251, Shall We Dance? (1996), a une note moyenne de 4.26 sur 46 votes.
Le film d'ID 272, Good Will Hunting (1997), a une note moyenne de 4.26 sur 198 votes.
Le film d'ID 657, Manchurian Candidate, The (1962), a une note moyenne de 4.26 sur 131 votes.
Le film d'ID 174, Raiders of the Lost Ark (1981), a une note moyenne de 4.25 sur 420 votes.
Le film d'ID 313, Titanic (1997), a une note moyenne de 4.25 sur 350 votes.
Le film d'ID 474, Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1963), a une note moyenne de 4.25 sur 174 votes.

Films ayant les pires notes

Les 10 pires films :

Le film d'ID 314, 3 Ninjas: High Noon At Mega Mountain (1998), a une note moyenne de 1.00 sur 5 vote(s).
Le film d'ID 437, Amityville 1992: It's About Time (1992), a une note moyenne de 1.00 sur 5 vote(s).
Le film d'ID 439, Amityville: A New Generation (1993), a une note moyenne de 1.00 sur 5 vote(s).
Le film d'ID 599, Police Story 4: Project S (Chao ji ji hua) (1993), a une note moyenne de 1.00 sur 5 vote(s).
Le film d'ID 784, Beyond Bedlam (1993), a une note moyenne de 1.00 sur 2 vote(s).
Le film d'ID 830, Power 98 (1995), a une note moyenne de 1.00 sur 1 vote(s).
Le film d'ID 852, Bloody Child, The (1996), a une note moyenne de 1.00 sur 1 vote(s).
Le film d'ID 858, Amityville: Dollhouse (1996), a une note moyenne de 1.00 sur 3 vote(s).
Le film d'ID 1308, Babyfever (1994), a une note moyenne de 1.00 sur 2 vote(s).
Le film d'ID 1309, Very Natural Thing, A (1974), a une note moyenne de 1.00 sur 1 vote(s).

Films ayant une note moyenne de 1

Il y a 69 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 424, Children of the Corn: The Gathering (1996), a une note moyenne de 1.32 sur 19 votes.
Le film d'ID 669, Body Parts (1991), a une note moyenne de 1.62 sur 13 votes.
Le film d'ID 440, Amityville II: The Possession (1982), a une note moyenne de 1.64 sur 14 votes.
Le film d'ID 1087, Bloodsport 2 (1995), a une note moyenne de 1.70 sur 10 votes.
Le film d'ID 758, Lawnmower Man 2: Beyond Cyberspace (1996), a une note moyenne de 1.71 sur 21 votes.
Le film d'ID 1274, Robocop 3 (1993), a une note moyenne de 1.73 sur 11 votes.
Le film d'ID 457, Free Willy 3: The Rescue (1997), a une note moyenne de 1.74 sur 27 votes.
Le film d'ID 1336, Kazaam (1996), a une note moyenne de 1.80 sur 10 votes.
Le film d'ID 1254, Gone Fishin' (1997), a une note moyenne de 1.82 sur 11 votes.
Le film d'ID 545, Vampire in Brooklyn (1995), a une note moyenne de 1.83 sur 12 votes.

Sur les utilisateurs

Utilisateurs ayant le plus vote

Les 10 utilisateurs ayant note le plus de films :

L'utilisateur d'ID 405 a vote 736 fois.
L'utilisateur d'ID 655 a vote 678 fois.
L'utilisateur d'ID 13 a vote 632 fois.
L'utilisateur d'ID 450 a vote 538 fois.
L'utilisateur d'ID 276 a vote 516 fois.
L'utilisateur d'ID 537 a vote 489 fois.
L'utilisateur d'ID 416 a vote 487 fois.
L'utilisateur d'ID 303 a vote 483 fois.
L'utilisateur d'ID 234 a vote 478 fois.
L'utilisateur d'ID 393 a vote 447 fois.

Utilisateurs ayant le moins vote

Les 15 utilisateurs ayant note le moins de films:

L'utilisateur d'ID 300 n'a vote que 19 fois.
L'utilisateur d'ID 19 n'a vote que 20 fois.
L'utilisateur d'ID 34 n'a vote que 20 fois.
L'utilisateur d'ID 36 n'a vote que 20 fois.
L'utilisateur d'ID 93 n'a vote que 20 fois.
L'utilisateur d'ID 143 n'a vote que 20 fois.
L'utilisateur d'ID 147 n'a vote que 20 fois.
L'utilisateur d'ID 166 n'a vote que 20 fois.
L'utilisateur d'ID 202 n'a vote que 20 fois.
L'utilisateur d'ID 242 n'a vote que 20 fois.
L'utilisateur d'ID 302 n'a vote que 20 fois.

```
## L'utilisateur d'ID 309 n'a vote que 20 fois.  
## L'utilisateur d'ID 341 n'a vote que 20 fois.  
## L'utilisateur d'ID 364 n'a vote que 20 fois.  
## L'utilisateur d'ID 418 n'a vote que 20 fois.
```

Remarque : la suppression de la double notation (due aux films en doublons) a fait qu'il existe dans la base un utilisateur avec 19 notes attribuées !

Utilisateurs n'ayant vote que 20 fois (ou moins ...)

```
## Il y a 34 utilisateurs qui ont voté moins que 20 fois.
```

Utilisateurs ayant vote moins de n fois

```
## Il y a 381 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 849 a donne en moyenne une note de 4.87 pour 23 notes.  
## L'utilisateur d'ID 688 a donne en moyenne une note de 4.83 pour 24 notes.  
## L'utilisateur d'ID 507 a donne en moyenne une note de 4.72 pour 58 notes.  
## L'utilisateur d'ID 628 a donne en moyenne une note de 4.70 pour 27 notes.  
## L'utilisateur d'ID 928 a donne en moyenne une note de 4.68 pour 31 notes.  
## L'utilisateur d'ID 118 a donne en moyenne une note de 4.66 pour 71 notes.  
## L'utilisateur d'ID 907 a donne en moyenne une note de 4.57 pour 147 notes.  
## L'utilisateur d'ID 686 a donne en moyenne une note de 4.56 pour 71 notes.  
## L'utilisateur d'ID 427 a donne en moyenne une note de 4.55 pour 31 notes.  
## L'utilisateur d'ID 565 a donne en moyenne une note de 4.54 pour 35 notes.  
## L'utilisateur d'ID 469 a donne en moyenne une note de 4.53 pour 43 notes.  
## L'utilisateur d'ID 850 a donne en moyenne une note de 4.53 pour 51 notes.  
## L'utilisateur d'ID 225 a donne en moyenne une note de 4.52 pour 27 notes.  
## L'utilisateur d'ID 330 a donne en moyenne une note de 4.50 pour 147 notes.  
## L'utilisateur d'ID 477 a donne en moyenne une note de 4.46 pour 35 notes.  
## L'utilisateur d'ID 242 a donne en moyenne une note de 4.45 pour 20 notes.  
## L'utilisateur d'ID 636 a donne en moyenne une note de 4.45 pour 20 notes.  
## L'utilisateur d'ID 583 a donne en moyenne une note de 4.44 pour 27 notes.  
## L'utilisateur d'ID 810 a donne en moyenne une note de 4.44 pour 25 notes.  
## L'utilisateur d'ID 252 a donne en moyenne une note de 4.43 pour 21 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 507 a donne en moyenne une note de 4.72 pour 58 notes.  
## L'utilisateur d'ID 118 a donne en moyenne une note de 4.66 pour 71 notes.  
## L'utilisateur d'ID 907 a donne en moyenne une note de 4.57 pour 147 notes.  
## L'utilisateur d'ID 686 a donne en moyenne une note de 4.56 pour 71 notes.  
## L'utilisateur d'ID 850 a donne en moyenne une note de 4.53 pour 51 notes.
```

```

## L'utilisateur d'ID 330 a donne en moyenne une note de 4.50 pour 147 notes.
## L'utilisateur d'ID 12 a donne en moyenne une note de 4.39 pour 51 notes.
## L'utilisateur d'ID 312 a donne en moyenne une note de 4.38 pour 223 notes.
## L'utilisateur d'ID 383 a donne en moyenne une note de 4.38 pour 71 notes.
## L'utilisateur d'ID 523 a donne en moyenne une note de 4.36 pour 100 notes.
## L'utilisateur d'ID 862 a donne en moyenne une note de 4.36 pour 165 notes.
## L'utilisateur d'ID 233 a donne en moyenne une note de 4.35 pour 110 notes.
## L'utilisateur d'ID 324 a donne en moyenne une note de 4.35 pour 66 notes.
## L'utilisateur d'ID 821 a donne en moyenne une note de 4.34 pour 62 notes.
## L'utilisateur d'ID 16 a donne en moyenne une note de 4.33 pour 140 notes.
## L'utilisateur d'ID 270 a donne en moyenne une note de 4.33 pour 138 notes.
## L'utilisateur d'ID 272 a donne en moyenne une note de 4.33 pour 51 notes.
## L'utilisateur d'ID 24 a donne en moyenne une note de 4.32 pour 68 notes.
## L'utilisateur d'ID 152 a donne en moyenne une note de 4.32 pour 106 notes.
## L'utilisateur d'ID 350 a donne en moyenne une note de 4.32 pour 50 notes.

```

Utilisateurs ayant donnees les plus faibles notes

```

## Les 20 utilisateurs les plus severes :

```

```

## L'utilisateur d'ID 181 a donne en moyenne une note de 1.49 pour 433 notes.
## L'utilisateur d'ID 405 a donne en moyenne une note de 1.83 pour 736 notes.
## L'utilisateur d'ID 445 a donne en moyenne une note de 1.99 pour 134 notes.
## L'utilisateur d'ID 685 a donne en moyenne une note de 2.05 pour 20 notes.
## L'utilisateur d'ID 774 a donne en moyenne une note de 2.06 pour 224 notes.
## L'utilisateur d'ID 206 a donne en moyenne une note de 2.17 pour 64 notes.
## L'utilisateur d'ID 724 a donne en moyenne une note de 2.18 pour 84 notes.
## L'utilisateur d'ID 865 a donne en moyenne une note de 2.29 pour 66 notes.
## L'utilisateur d'ID 609 a donne en moyenne une note de 2.39 pour 28 notes.
## L'utilisateur d'ID 626 a donne en moyenne une note de 2.39 pour 31 notes.
## L'utilisateur d'ID 302 a donne en moyenne une note de 2.45 pour 20 notes.
## L'utilisateur d'ID 702 a donne en moyenne une note de 2.45 pour 31 notes.
## L'utilisateur d'ID 637 a donne en moyenne une note de 2.50 pour 103 notes.
## L'utilisateur d'ID 255 a donne en moyenne une note de 2.54 pour 83 notes.
## L'utilisateur d'ID 656 a donne en moyenne une note de 2.54 pour 24 notes.
## L'utilisateur d'ID 161 a donne en moyenne une note de 2.55 pour 58 notes.
## L'utilisateur d'ID 509 a donne en moyenne une note de 2.56 pour 32 notes.
## L'utilisateur d'ID 900 a donne en moyenne une note de 2.56 pour 45 notes.
## L'utilisateur d'ID 660 a donne en moyenne une note de 2.59 pour 223 notes.
## L'utilisateur d'ID 129 a donne en moyenne une note de 2.60 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 181 a donne en moyenne une note de 1.49 pour 433 notes.
## L'utilisateur d'ID 405 a donne en moyenne une note de 1.83 pour 736 notes.
## L'utilisateur d'ID 445 a donne en moyenne une note de 1.99 pour 134 notes.
## L'utilisateur d'ID 774 a donne en moyenne une note de 2.06 pour 224 notes.
## L'utilisateur d'ID 206 a donne en moyenne une note de 2.17 pour 64 notes.
## L'utilisateur d'ID 724 a donne en moyenne une note de 2.18 pour 84 notes.
## L'utilisateur d'ID 865 a donne en moyenne une note de 2.29 pour 66 notes.

```

L'utilisateur d'ID 637 a donne en moyenne une note de 2.50 pour 103 notes.
L'utilisateur d'ID 255 a donne en moyenne une note de 2.54 pour 83 notes.
L'utilisateur d'ID 161 a donne en moyenne une note de 2.55 pour 58 notes.
L'utilisateur d'ID 660 a donne en moyenne une note de 2.59 pour 223 notes.
L'utilisateur d'ID 102 a donne en moyenne une note de 2.62 pour 216 notes.
L'utilisateur d'ID 617 a donne en moyenne une note de 2.64 pour 107 notes.
L'utilisateur d'ID 933 a donne en moyenne une note de 2.65 pour 184 notes.
L'utilisateur d'ID 104 a donne en moyenne une note de 2.66 pour 110 notes.
L'utilisateur d'ID 21 a donne en moyenne une note de 2.67 pour 178 notes.
L'utilisateur d'ID 49 a donne en moyenne une note de 2.69 pour 214 notes.
L'utilisateur d'ID 653 a donne en moyenne une note de 2.70 pour 282 notes.
L'utilisateur d'ID 451 a donne en moyenne une note de 2.73 pour 96 notes.
L'utilisateur d'ID 843 a donne en moyenne une note de 2.74 pour 204 notes.

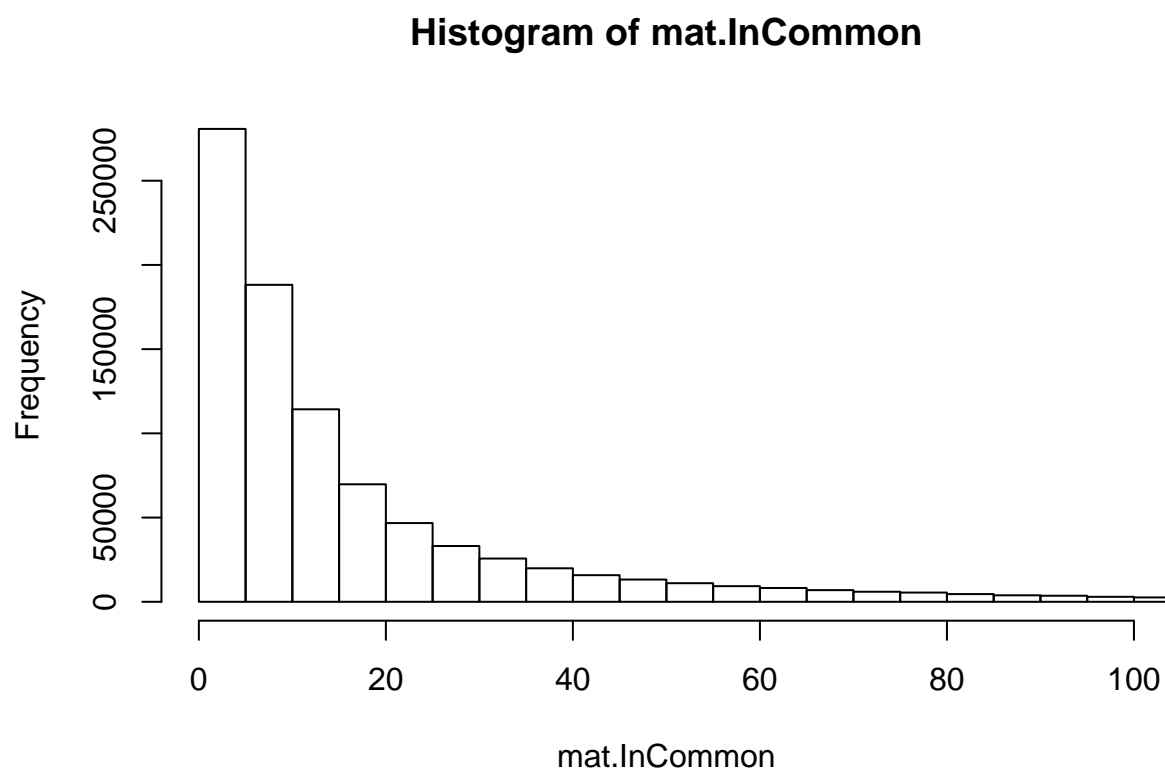
Etude par age

Etude par sexe

```
##      userID      age      sex      occupation      zip.code
## Min.   : 1.0   Min.   : 7.00   F: 0   student      :136   55414 : 6
## 1st Qu.:234.2  1st Qu.:25.00   M:670  educator      : 69   10003 : 5
## Median :467.5  Median :31.00           other      : 69   55454 : 4
## Mean   :468.2  Mean   :34.15           engineer   : 65   02215 : 3
## 3rd Qu.:695.8  3rd Qu.:42.00           programmer : 60   14216 : 3
## Max.   :943.0  Max.   :73.00           administrator: 43   22902 : 3
##                                     (Other)      :228   (Other):646
##      nb.Ratings      mean      sd      max
## Min.   : 20.0   Min.   :1.490   Min.   :0.480   Min.   :4.000
## 1st Qu.: 35.0   1st Qu.:3.350   1st Qu.:0.870   1st Qu.:5.000
## Median : 69.5   Median :3.620   Median :1.000   Median :5.000
## Mean   :110.5   Mean   :3.589   Mean   :1.011   Mean   :4.981
## 3rd Qu.:154.8   3rd Qu.:3.860   3rd Qu.:1.130   3rd Qu.:5.000
## Max.   :632.0   Max.   :4.700   Max.   :1.690   Max.   :5.000
##
##      min      med
## Min.   :1.000   Min.   :1.000
## 1st Qu.:1.000   1st Qu.:3.000
## Median :1.000   Median :4.000
## Mean   :1.258   Mean   :3.711
## 3rd Qu.:1.000   3rd Qu.:4.000
## Max.   :3.000   Max.   :5.000
##
##      userID      age      sex      occupation      zip.code
## Min.   : 2.0   Min.   :13.00   F:273  student      :60   20009 : 3
## 1st Qu.:240.0  1st Qu.:25.00   M: 0   administrator:36   22903 : 3
## Median :491.0  Median :32.00           other      :36   55105 : 3
## Mean   :481.4  Mean   :33.81           librarian   :29   55414 : 3
## 3rd Qu.:720.0  3rd Qu.:44.00           educator    :26   10016 : 2
## Max.   :942.0  Max.   :70.00           writer      :19   11217 : 2
##                                     (Other)      :67   (Other):257
##      nb.Ratings      mean      sd      max
## Min.   : 19.00   Min.   :1.830   Min.   :0.340   Min.   :3.000
## 1st Qu.: 29.00   1st Qu.:3.250   1st Qu.:0.890   1st Qu.:5.000
## Median : 54.00   Median :3.640   Median :1.020   Median :5.000
## Mean   : 93.96   Mean   :3.587   Mean   :1.041   Mean   :4.989
## 3rd Qu.:127.00   3rd Qu.:3.890   3rd Qu.:1.180   3rd Qu.:5.000
## Max.   :736.00   Max.   :4.870   Max.   :1.750   Max.   :5.000
##
##      min      med
## Min.   :1.000   Min.   :1.000
## 1st Qu.:1.000   1st Qu.:3.000
## Median :1.000   Median :4.000
## Mean   :1.282   Mean   :3.683
## 3rd Qu.:1.000   3rd Qu.:4.000
## Max.   :4.000   Max.   :5.000
##
```

Etude par genre de films

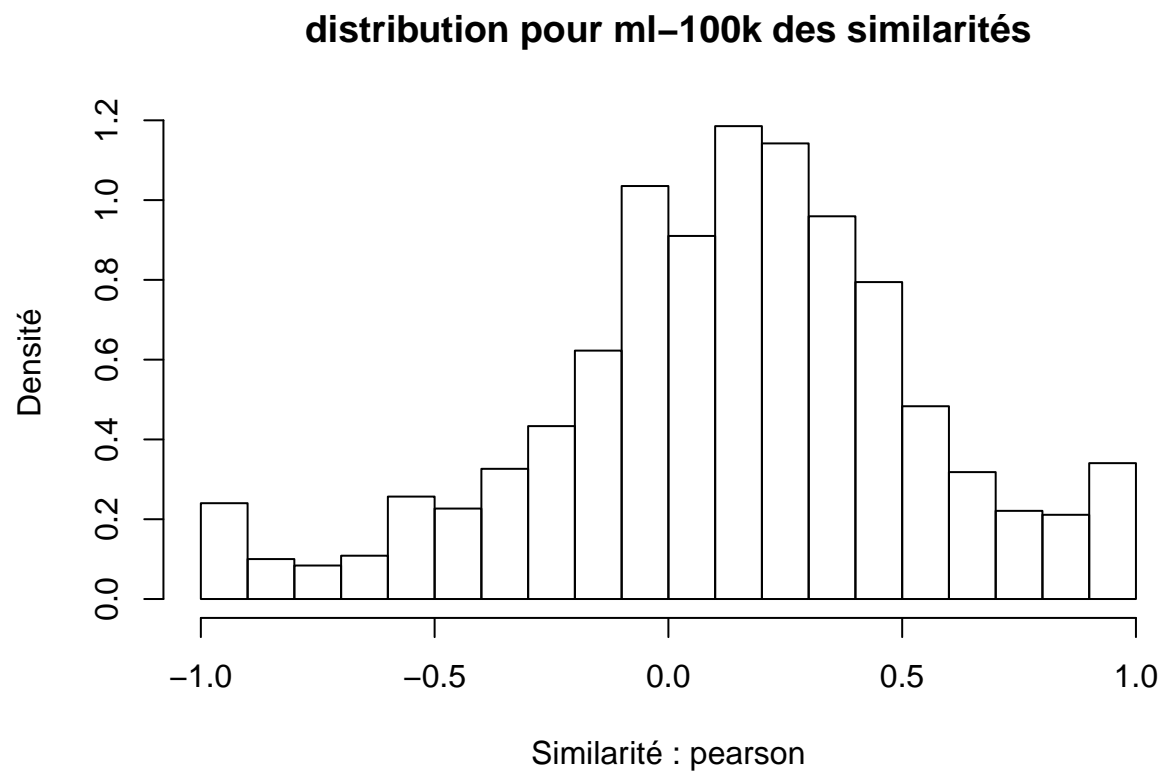
Nombre de films en commun



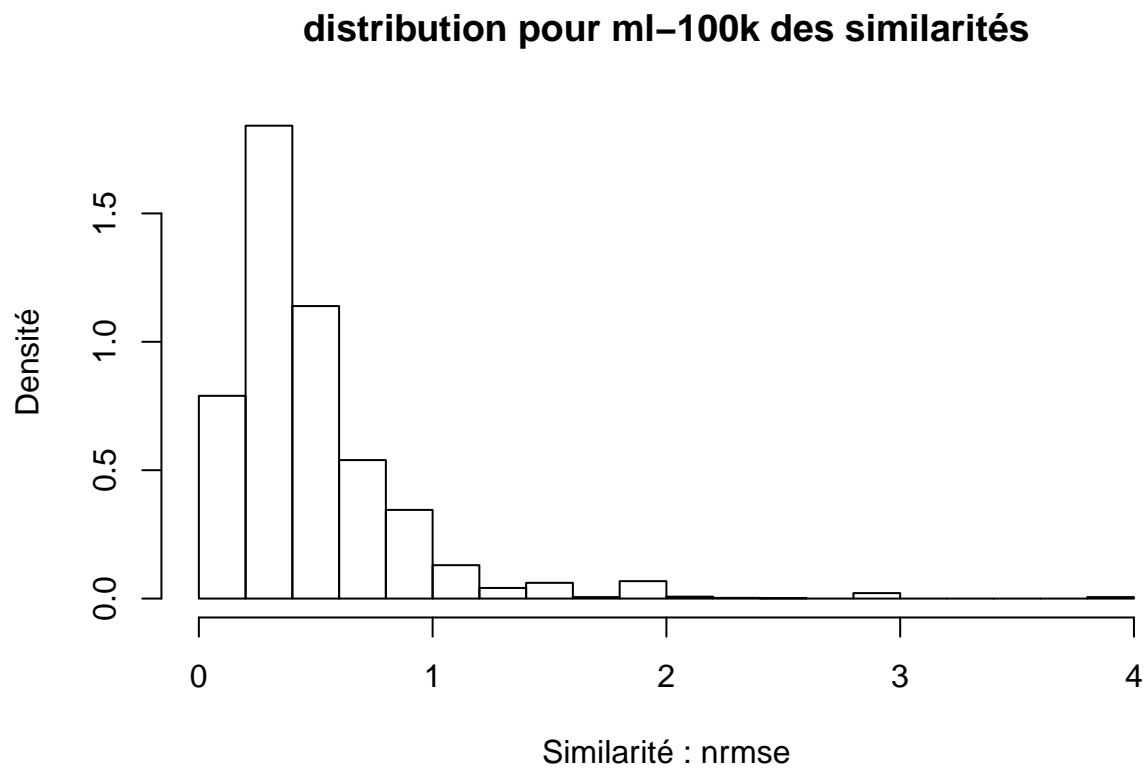
```
## mat.InCommon
##      0      1      2      3      4      5      6      7      8      9     10
## 29920 44640 52000 54046 51554 48620 45084 41198 37264 34032 30626
```

Etude sur les similarités

Corrélation de Pearson



Distance euclidienne



Distance manhattan

