Netflix.R

nisrin

Thu Jan 10 00:16:43 2019

echo= TRUE   
eval= TRUE

movies = read.table("movieLens.txt", header=FALSE, sep="|",quote="\"")  
str(movies)

## 'data.frame': 1682 obs. of 24 variables:  
## $ V1 : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ V2 : Factor w/ 1664 levels "'Til There Was You (1997)",..: 1525 618 555 594 344 1318 1545 111 391 1240 ...  
## $ V3 : Factor w/ 241 levels "","01-Aug-1997",..: 71 71 71 71 71 71 71 71 71 182 ...  
## $ V4 : logi NA NA NA NA NA NA ...  
## $ V5 : Factor w/ 1661 levels "","http://us.imdb.com/M/title-exact/Independence%20(1997)",..: 1431 565 505 543 310 1661 1453 103 357 1183 ...  
## $ V6 : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ V7 : int 0 1 0 1 0 0 0 0 0 0 ...  
## $ V8 : int 0 1 0 0 0 0 0 0 0 0 ...  
## $ V9 : int 1 0 0 0 0 0 0 0 0 0 ...  
## $ V10: int 1 0 0 0 0 0 0 1 0 0 ...  
## $ V11: int 1 0 0 1 0 0 0 1 0 0 ...  
## $ V12: int 0 0 0 0 1 0 0 0 0 0 ...  
## $ V13: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ V14: int 0 0 0 1 1 1 1 1 1 1 ...  
## $ V15: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ V16: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ V17: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ V18: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ V19: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ V20: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ V21: int 0 0 0 0 0 0 1 0 0 0 ...  
## $ V22: int 0 1 1 0 1 0 0 0 0 0 ...  
## $ V23: int 0 0 0 0 0 0 0 0 0 1 ...  
## $ V24: int 0 0 0 0 0 0 0 0 0 0 ...

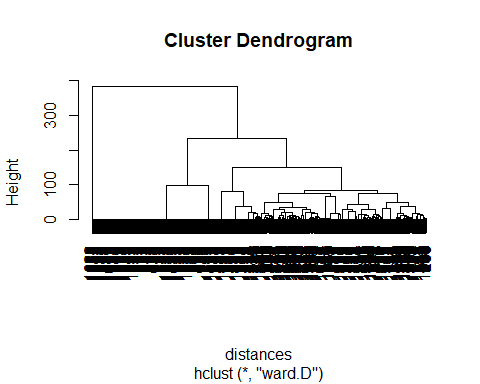
# Add column names  
colnames(movies) = c("ID", "Title", "ReleaseDate", "VideoReleaseDate", "IMDB", "Unknown", "Action", "Adventure", "Animation", "Childrens", "Comedy", "Crime", "Documentary", "Drama", "Fantasy", "FilmNoir", "Horror", "Musical", "Mystery", "Romance", "SciFi", "Thriller", "War", "Western")  
str(movies)

## 'data.frame': 1682 obs. of 24 variables:  
## $ ID : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ Title : Factor w/ 1664 levels "'Til There Was You (1997)",..: 1525 618 555 594 344 1318 1545 111 391 1240 ...  
## $ ReleaseDate : Factor w/ 241 levels "","01-Aug-1997",..: 71 71 71 71 71 71 71 71 71 182 ...  
## $ VideoReleaseDate: logi NA NA NA NA NA NA ...  
## $ IMDB : Factor w/ 1661 levels "","http://us.imdb.com/M/title-exact/Independence%20(1997)",..: 1431 565 505 543 310 1661 1453 103 357 1183 ...  
## $ Unknown : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Action : int 0 1 0 1 0 0 0 0 0 0 ...  
## $ Adventure : int 0 1 0 0 0 0 0 0 0 0 ...  
## $ Animation : int 1 0 0 0 0 0 0 0 0 0 ...  
## $ Childrens : int 1 0 0 0 0 0 0 1 0 0 ...  
## $ Comedy : int 1 0 0 1 0 0 0 1 0 0 ...  
## $ Crime : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ Documentary : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Drama : int 0 0 0 1 1 1 1 1 1 1 ...  
## $ Fantasy : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ FilmNoir : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Horror : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Musical : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Mystery : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Romance : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ SciFi : int 0 0 0 0 0 0 1 0 0 0 ...  
## $ Thriller : int 0 1 1 0 1 0 0 0 0 0 ...  
## $ War : int 0 0 0 0 0 0 0 0 0 1 ...  
## $ Western : int 0 0 0 0 0 0 0 0 0 0 ...

# Remove unnecessary and duplicates variables  
movies$ID = NULL  
movies$ReleaseDate = NULL  
movies$VideoReleaseDate = NULL  
movies$IMDB = NULL  
movies = unique(movies)  
str(movies)

## 'data.frame': 1664 obs. of 20 variables:  
## $ Title : Factor w/ 1664 levels "'Til There Was You (1997)",..: 1525 618 555 594 344 1318 1545 111 391 1240 ...  
## $ Unknown : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Action : int 0 1 0 1 0 0 0 0 0 0 ...  
## $ Adventure : int 0 1 0 0 0 0 0 0 0 0 ...  
## $ Animation : int 1 0 0 0 0 0 0 0 0 0 ...  
## $ Childrens : int 1 0 0 0 0 0 0 1 0 0 ...  
## $ Comedy : int 1 0 0 1 0 0 0 1 0 0 ...  
## $ Crime : int 0 0 0 0 1 0 0 0 0 0 ...  
## $ Documentary: int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Drama : int 0 0 0 1 1 1 1 1 1 1 ...  
## $ Fantasy : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ FilmNoir : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Horror : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Musical : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Mystery : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ Romance : int 0 0 0 0 0 0 0 0 0 0 ...  
## $ SciFi : int 0 0 0 0 0 0 1 0 0 0 ...  
## $ Thriller : int 0 1 1 0 1 0 0 0 0 0 ...  
## $ War : int 0 0 0 0 0 0 0 0 0 1 ...  
## $ Western : int 0 0 0 0 0 0 0 0 0 0 ...

# distances  
distances = dist(movies[2:20], method = "euclidean")  
  
# clustering  
clusterMovies = hclust(distances, method = "ward.D")   
  
#dendrogram  
plot(clusterMovies)



# Assign points to clusters  
clusterGroups = cutree(clusterMovies, k = 10)  
  
  
#percentage of movies in each genre and cluster  
  
tapply(movies$Action, clusterGroups, mean)

## 1 2 3 4 5 6 7   
## 0.1784512 0.7839196 0.1238532 0.0000000 0.0000000 0.1015625 0.0000000   
## 8 9 10   
## 0.0000000 0.0000000 0.0000000

tapply(movies$Romance, clusterGroups, mean)

## 1 2 3 4 5 6   
## 0.10437710 0.04522613 0.03669725 0.00000000 0.00000000 1.00000000   
## 7 8 9 10   
## 1.00000000 0.00000000 0.00000000 0.00000000

# Create a new data set with just the movies from cluster 2  
cluster2 = subset(movies, clusterGroups==2)  
  
# Look at the first 10 titles in this cluster:  
cluster2$Title[1:10]

## [1] GoldenEye (1995)   
## [2] Bad Boys (1995)   
## [3] Apollo 13 (1995)   
## [4] Net, The (1995)   
## [5] Natural Born Killers (1994)   
## [6] Outbreak (1995)   
## [7] Stargate (1994)   
## [8] Fugitive, The (1993)   
## [9] Jurassic Park (1993)   
## [10] Robert A. Heinlein's The Puppet Masters (1994)  
## 1664 Levels: 'Til There Was You (1997) ... Zeus and Roxanne (1997)

clusterGroups = cutree(clusterMovies, k = 2)  
  
clusterGroups

## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15   
## 1 1 1 1 1 2 1 1 2 1 1 1 1 1 2   
## 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30   
## 1 1 2 2 1 1 1 1 1 1 1 1 1 1 2   
## 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45   
## 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1   
## 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60   
## 2 1 1 1 1 1 2 1 1 1 1 2 2 2 2   
## 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75   
## 2 1 1 2 1 1 1 1 1 1 1 1 1 1 1   
## 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90   
## 1 1 1 1 1 1 1 1 1 1 2 2 1 1 1   
## 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105   
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120   
## 1 2 1 1 1 1 1 2 1 1 1 1 1 1 1   
## 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135   
## 1 1 1 1 1 2 1 1 1 1 1 1 1 2 1   
## 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150   
## 2 2 1 1 1 1 1 1 1 1 2 1 1 2 1   
## 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165   
## 1 1 1 1 1 1 1 1 1 2 1 2 1 1 2   
## 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180   
## 2 1 1 1 1 1 1 1 1 1 1 1 2 1 1   
## 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195   
## 1 1 1 1 1 1 1 1 1 1 1 2 2 1 1   
## 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210   
## 2 1 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225   
## 1 2 1 1 2 1 1 1 1 1 2 1 1 2 1   
## 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240   
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255   
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 256 257 258 259 260 261 262 263 264 265 266 267 269 270 271   
## 1 1 1 1 1 1 2 1 1 1 1 1 1 1 1   
## 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286   
## 2 1 1 1 1 2 1 2 1 1 2 1 1 2 1   
## 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301   
## 2 1 1 1 1 2 1 1 1 2 2 1 1 1 1   
## 302 304 305 306 307 308 309 310 311 312 313 314 315 316 317   
## 1 1 2 1 1 1 1 2 1 1 1 1 1 1 2   
## 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332   
## 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1   
## 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347   
## 1 1 1 1 1 1 1 2 1 1 1 2 1 1 1   
## 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363   
## 1 1 1 1 1 1 1 1 2 1 1 1 1 1 1   
## 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378   
## 1 2 2 1 1 1 1 1 1 1 1 2 1 1 2   
## 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393   
## 1 1 1 1 1 1 1 1 2 1 1 1 1 2 1   
## 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408   
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423   
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438   
## 1 1 1 2 1 1 1 1 1 1 1 1 1 1 1   
## 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453   
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468   
## 2 1 1 1 2 2 2 2 1 1 2 1 1 2 2   
## 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483   
## 2 1 1 1 1 1 2 1 1 1 1 1 1 1 1   
## 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498   
## 1 1 1 1 1 1 1 1 2 1 1 1 2 1 1   
## 499 501 502 503 504 505 506 507 508 509 510 511 512 513 514   
## 2 1 1 2 1 1 2 2 2 2 1 1 1 1 1   
## 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529   
## 1 1 1 2 1 1 1 1 1 1 1 1 2 1 2   
## 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544   
## 1 1 1 1 2 1 2 2 1 1 1 1 1 1 1   
## 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559   
## 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1   
## 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574   
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589   
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604   
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619   
## 1 2 1 1 1 1 1 2 1 2 1 1 2 2 1   
## 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634   
## 2 1 1 1 1 1 1 2 1 1 1 1 2 2 1   
## 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649   
## 1 1 1 2 2 2 1 1 1 1 1 1 1 1 1   
## 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664   
## 2 1 1 1 1 1 1 1 2 1 2 1 1 1 2   
## 665 666 667 668 669 671 672 673 674 675 676 677 678 679 681   
## 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1   
## 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696   
## 1 1 1 1 1 1 1 1 1 1 1 2 1 1 1   
## 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711   
## 2 2 2 1 1 1 2 1 1 1 2 2 1 1 2   
## 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726   
## 1 2 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741   
## 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1   
## 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756   
## 1 1 1 1 1 1 1 1 2 1 1 2 1 1 1   
## 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771   
## 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1   
## 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786   
## 2 1 1 1 2 1 1 1 1 1 2 1 1 1 1   
## 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801   
## 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1   
## 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816   
## 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1   
## 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831   
## 2 1 1 1 1 2 1 1 1 1 1 1 1 1 1   
## 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846   
## 1 1 1 1 1 2 1 1 1 1 1 1 1 1 1   
## 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861   
## 1 1 1 1 2 1 1 1 1 1 1 1 1 1 1   
## 862 863 864 866 867 868 869 870 871 872 873 874 875 876 877   
## 1 2 1 1 2 2 1 1 1 1 1 2 1 1 1   
## 878 879 880 882 883 884 885 886 887 888 889 890 891 892 893   
## 1 1 2 2 2 1 1 1 2 2 1 1 1 1 1   
## 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908   
## 1 1 2 1 2 2 2 1 1 1 1 1 1 1 1   
## 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923   
## 2 2 1 1 1 1 2 1 1 1 1 2 1 1 2   
## 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938   
## 1 1 1 2 1 1 1 1 1 2 2 1 1 2 1   
## 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953   
## 1 1 1 2 1 1 1 1 1 1 1 2 1 1 1   
## 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968   
## 1 1 2 1 2 1 2 2 2 1 1 1 1 2 1   
## 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983   
## 1 1 1 2 1 1 1 1 1 1 1 2 2 1 1   
## 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998   
## 1 2 1 1 1 1 1 1 1 1 2 1 1 1 1   
## 999 1000 1001 1002 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014   
## 1 1 1 1 1 2 1 1 2 2 2 1 1 1 1   
## 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029   
## 1 1 2 2 1 1 2 1 1 1 1 1 1 1 1   
## 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044   
## 1 1 1 1 1 1 1 1 1 2 1 1 1 1 1   
## 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059   
## 2 1 1 1 1 1 1 1 2 1 2 1 1 1 1   
## 1060 1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074   
## 1 1 2 1 1 1 1 1 2 1 1 1 1 1 1   
## 1075 1076 1077 1078 1079 1080 1081 1082 1083 1084 1085 1086 1087 1088 1089   
## 2 1 1 1 1 1 1 2 1 1 1 2 1 1 1   
## 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100 1101 1102 1103 1104   
## 1 1 1 1 1 1 1 2 1 2 1 1 1 1 1   
## 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119   
## 1 1 2 2 1 1 2 2 2 1 1 1 2 1 1   
## 1120 1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134   
## 1 1 1 2 1 1 1 2 1 1 1 1 1 1 2   
## 1135 1136 1137 1138 1139 1140 1141 1142 1143 1144 1145 1146 1147 1148 1149   
## 1 2 1 1 1 1 1 1 1 2 2 2 2 2 2   
## 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160 1161 1162 1163 1164   
## 2 1 1 1 1 1 1 1 2 1 1 1 1 2 1   
## 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179   
## 1 1 1 2 2 1 2 1 1 1 1 1 1 1 1   
## 1180 1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194   
## 1 1 1 1 1 1 1 1 1 2 1 1 2 2 1   
## 1195 1196 1197 1198 1199 1200 1201 1202 1203 1204 1205 1206 1207 1208 1209   
## 2 2 1 1 1 1 1 1 1 1 2 1 1 1 1   
## 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220 1221 1222 1223 1224   
## 1 1 1 1 2 1 1 1 1 1 2 2 1 2 2   
## 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239   
## 1 1 2 1 1 1 1 1 2 1 1 2 1 2 1   
## 1240 1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254   
## 1 1 1 1 1 1 1 1 1 1 1 1 2 1 1   
## 1255 1256 1258 1259 1260 1261 1262 1263 1264 1265 1266 1267 1268 1269 1270   
## 2 2 1 1 2 2 1 2 2 2 2 2 2 1 1   
## 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280 1281 1282 1283 1284 1285   
## 1 1 1 1 1 2 1 1 1 1 2 2 1 1 2   
## 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300   
## 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1   
## 1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315   
## 1 1 1 1 1 2 1 1 2 2 1 1 1 1 1   
## 1316 1317 1318 1319 1320 1321 1322 1323 1324 1325 1326 1327 1328 1329 1330   
## 2 2 1 2 2 1 1 2 1 2 2 2 2 2 2   
## 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340 1341 1342 1343 1344 1345   
## 1 2 1 2 2 1 1 2 1 1 2 2 2 2 2   
## 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360   
## 2 2 2 1 2 1 2 1 2 2 1 1 1 2 1   
## 1361 1362 1363 1364 1365 1366 1367 1368 1369 1370 1371 1372 1373 1374 1375   
## 1 1 1 1 1 1 1 2 2 1 1 1 1 1 2   
## 1376 1377 1378 1379 1380 1381 1382 1383 1384 1385 1386 1387 1388 1389 1390   
## 1 1 1 1 2 2 2 1 2 1 1 1 2 2 1   
## 1391 1392 1393 1394 1395 1396 1397 1398 1399 1400 1401 1402 1403 1404 1405   
## 2 2 1 1 2 2 2 2 1 1 2 2 1 1 2   
## 1406 1407 1408 1409 1410 1411 1412 1413 1414 1415 1416 1417 1418 1419 1420   
## 1 1 1 1 2 1 1 1 1 1 1 2 2 1 1   
## 1421 1422 1423 1424 1425 1426 1427 1428 1429 1430 1431 1432 1433 1434 1435   
## 2 1 1 1 1 1 2 1 1 2 1 2 1 1 1   
## 1436 1437 1438 1439 1440 1441 1442 1443 1444 1445 1446 1447 1448 1449 1450   
## 1 1 2 1 2 1 2 2 1 2 1 2 2 2 1   
## 1451 1452 1453 1454 1455 1456 1457 1458 1459 1460 1461 1462 1463 1464 1465   
## 1 1 1 1 1 1 1 1 1 1 1 1 1 2 1   
## 1466 1467 1468 1469 1470 1471 1472 1473 1474 1475 1476 1477 1478 1479 1480   
## 2 2 2 1 1 1 1 1 1 1 1 1 1 1 1   
## 1481 1482 1483 1484 1485 1486 1487 1488 1489 1490 1491 1492 1493 1494 1495   
## 2 1 1 1 1 2 1 2 1 1 1 1 1 1 2   
## 1496 1497 1498 1499 1500 1501 1502 1503 1504 1505 1506 1507 1508 1509 1510   
## 1 1 1 1 1 1 1 1 1 1 2 1 1 1 1   
## 1511 1512 1513 1514 1515 1516 1517 1518 1519 1520 1521 1522 1523 1524 1525   
## 1 2 1 2 1 2 2 2 1 1 1 1 1 2 1   
## 1526 1527 1528 1529 1530 1531 1532 1533 1534 1535 1536 1537 1538 1539 1540   
## 1 1 2 1 2 1 2 2 1 2 2 1 2 2 1   
## 1541 1542 1543 1544 1545 1546 1547 1548 1549 1550 1551 1552 1553 1554 1555   
## 2 2 2 1 1 2 1 1 1 1 2 1 1 2 1   
## 1556 1557 1558 1559 1560 1561 1562 1563 1564 1565 1566 1567 1568 1569 1570   
## 1 1 2 1 1 1 1 1 2 2 2 1 1 1 1   
## 1571 1572 1573 1574 1575 1576 1577 1578 1579 1580 1581 1582 1583 1584 1585   
## 2 2 1 1 2 1 2 2 1 1 1 1 2 2 1   
## 1586 1587 1588 1589 1590 1591 1592 1593 1594 1595 1596 1597 1598 1599 1600   
## 1 1 1 1 2 2 2 1 1 1 1 1 1 2 1   
## 1601 1602 1603 1604 1605 1608 1609 1610 1611 1612 1613 1614 1615 1616 1618   
## 1 2 2 1 1 1 1 1 1 1 1 1 1 2 1   
## 1619 1620 1621 1622 1623 1624 1626 1627 1628 1629 1630 1631 1632 1633 1634   
## 2 1 1 1 2 1 1 1 2 1 2 1 1 1 2   
## 1635 1636 1637 1638 1639 1640 1641 1642 1643 1644 1645 1646 1647 1648 1649   
## 2 2 2 1 2 2 1 2 2 1 2 1 1 2 1   
## 1651 1652 1653 1655 1656 1657 1659 1660 1661 1662 1663 1664 1665 1666 1667   
## 1 1 2 1 1 1 1 2 2 1 1 1 2 2 2   
## 1668 1669 1670 1671 1672 1673 1674 1675 1676 1677 1678 1679 1681 1682   
## 1 1 1 2 2 1 2 2 2 2 2 1 1 2

table(clusterGroups)

## clusterGroups  
## 1 2   
## 1294 370