Regression Analysis of IMDB 5000 Movies Datasets

Purpose: By doing a regresson analysis, we want to know: 1) Among the 27 variables given, which of them are critical in telling the IMDB rating of a movie. 2) Is there any correlation between genre & IMDB raging,face number in poster & IMDB rating,director name & IMDB rating and duration & IMDB rating. 3) Predict the IMDB Score using our model

m<- read.csv('movie\_metadata copy.csv')

## Step 1: Data Collection

This data set was found from Kaggle. The author scraped 5000+ movies from IMDB website using a Python library called "scrapy" and obtain all needed 28 variables for 5043 movies and 4906 posters (998MB), spanning across 100 years in 66 countries. There are 2399 unique director names, and thousands of actors/actresses. Below are the 28 variables: "movie\_title" "color" "num\_critic\_for\_reviews" "movie\_facebook\_likes" "duration" "director\_name" "director\_facebook\_likes" "actor\_3\_name" "actor\_3\_facebook\_likes" "actor\_2\_name" "actor\_2\_facebook\_likes" "actor\_1\_name" "actor\_1\_facebook\_likes" "gross" "genres" "num\_voted\_users" "cast\_total\_facebook\_likes" "facenumber\_in\_poster" "plot\_keywords" "movie\_imdb\_link" "num\_user\_for\_reviews" "language" "country" "content\_rating" "budget" "title\_year" "imdb\_score" "aspect\_ratio"

This dataset is a proof of concept. It can be used for experimental and learning purpose.For comprehensive movie analysis and accurate movie ratings prediction, 28 attributes from 5000 movies might not be enough. A decent dataset could contain hundreds of attributes from 50K or more movies, and requires tons of feature engineering.

## Step 2 : Data cleaning and exploration

Assign the first word of genres as the genre of each movie:(genres been split into words in Excel):

# remove columns X-X.8  
which(colnames(m)=='genres')

## [1] 10

which(colnames(m)=='X.8')

## [1] 19

m<-m[,-c(11:19)]

Only keep movie data for USA, bacause the "budget" variable was not all converted to US dollars, which might cause a problem in later analysis. If we want to convert all budgets into US dollarts, we have to take in to consideration for inflation as well. This might make the problem more complicated. Therefore, for pratice purpose, we decided to only study data for movies of USA.

movie.usa<-m[which(m[,'country']=='USA'),]

Double check:

movie.usa$country

## [1] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [18] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [35] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [52] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [69] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [86] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [103] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [120] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [137] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [154] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [171] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [188] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [205] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [222] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [239] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [256] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [273] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [290] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [307] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [324] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [341] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [358] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [375] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [392] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [409] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [426] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [443] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [460] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [477] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [494] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [511] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [528] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [545] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [562] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [579] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [596] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [613] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [630] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [647] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [664] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [681] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [698] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [715] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [732] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [749] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [766] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [783] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [800] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [817] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [834] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [851] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [868] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [885] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [902] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [919] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [936] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [953] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [970] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [987] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1004] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1021] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1038] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1055] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1072] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1089] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1106] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1123] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1140] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1157] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1174] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1191] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1208] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1225] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1242] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1259] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1276] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1293] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1310] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1327] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1344] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1361] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1378] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1395] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1412] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1429] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1446] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1463] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1480] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1497] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1514] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1531] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1548] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1565] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1582] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1599] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1616] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1633] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1650] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1667] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1684] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1701] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1718] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1735] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1752] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1769] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1786] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1803] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1820] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1837] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1854] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1871] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1888] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1905] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1922] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1939] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1956] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1973] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [1990] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2007] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2024] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2041] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2058] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2075] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2092] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2109] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2126] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2143] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2160] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2177] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2194] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2211] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2228] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2245] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2262] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2279] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2296] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2313] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2330] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2347] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2364] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2381] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2398] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2415] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2432] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2449] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2466] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2483] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2500] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2517] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2534] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2551] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2568] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2585] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2602] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2619] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2636] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2653] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2670] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2687] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2704] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2721] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2738] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2755] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2772] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2789] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2806] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2823] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2840] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2857] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2874] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2891] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2908] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2925] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2942] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2959] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2976] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [2993] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3010] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3027] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3044] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3061] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3078] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3095] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3112] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3129] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3146] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3163] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3180] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3197] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3214] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3231] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3248] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3265] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3282] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3299] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3316] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3333] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3350] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3367] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3384] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3401] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3418] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3435] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3452] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3469] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3486] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3503] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3520] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3537] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3554] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3571] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3588] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3605] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3622] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3639] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3656] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3673] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3690] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3707] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3724] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3741] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3758] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3775] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## [3792] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA  
## 66 Levels: Afghanistan Argentina Aruba Australia Bahamas ... West Germany

Remove 'language' since after removing all countries except for USA, there is only 4 languages aside from English, not meaningful for our prediction.

summary(movie.usa$language)

## Aboriginal Arabic Aramaic Bosnian Cantonese   
## 10 0 0 1 1 1   
## Chinese Czech Danish Dari Dutch Dzongkha   
## 0 0 0 1 0 0   
## English Filipino French German Greek Hebrew   
## 3779 1 0 0 0 1   
## Hindi Hungarian Icelandic Indonesian Italian Japanese   
## 1 0 0 0 0 1   
## Kannada Kazakh Korean Mandarin Maya Mongolian   
## 0 0 0 0 1 0   
## None Norwegian Panjabi Persian Polish Portuguese   
## 1 0 0 0 0 0   
## Romanian Russian Slovenian Spanish Swahili Swedish   
## 0 0 0 7 0 0   
## Tamil Telugu Thai Urdu Vietnamese Zulu   
## 0 0 0 0 1 0

movie.usa<-movie.usa[, -which(names(movie.usa)=='language')]

Remove 'movie\_imdb\_link' column since it's not useful for our analysis and store the rest od the data as 'movie'.

movie.df= data.frame(movie.usa)  
mm<-movie.df[, -which(names(movie.df)=='movie\_imdb\_link')]

str(mm)

## 'data.frame': 3807 obs. of 26 variables:  
## $ color : Factor w/ 3 levels ""," Black and White",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ director\_name : Factor w/ 2399 levels "","\xcc\xe4mile Gaudreault",..: 926 799 379 106 2030 1652 1225 2394 284 799 ...  
## $ num\_critic\_for\_reviews : int 723 302 813 462 392 324 635 673 434 313 ...  
## $ duration : int 178 169 164 132 156 100 141 183 169 151 ...  
## $ director\_facebook\_likes : int 0 563 22000 475 0 15 0 0 0 563 ...  
## $ actor\_3\_facebook\_likes : int 855 1000 23000 530 4000 284 19000 2000 903 1000 ...  
## $ actor\_2\_name : Factor w/ 3033 levels "","50 Cent","A. Michael Baldwin",..: 1408 2218 534 2549 1228 801 2440 1704 1911 2218 ...  
## $ actor\_1\_facebook\_likes : int 1000 40000 27000 640 24000 799 26000 15000 18000 40000 ...  
## $ gross : int 760505847 309404152 448130642 73058679 336530303 200807262 458991599 330249062 200069408 423032628 ...  
## $ genres : Factor w/ 21 levels "Action","Adventure",..: 1 1 1 1 1 2 1 1 1 1 ...  
## $ actor\_1\_name : Factor w/ 2098 levels "","\xcc\xd2lafur Darri \xcc\xd2lafsson",..: 303 982 1968 441 786 221 337 740 1104 982 ...  
## $ movie\_title : Factor w/ 4917 levels "[Rec] 2\xe5\xca",..: 397 2731 3707 1960 3289 3459 398 460 3416 2732 ...  
## $ num\_voted\_users : int 886204 471220 1144337 212204 383056 294810 462669 371639 240396 522040 ...  
## $ cast\_total\_facebook\_likes: int 4834 48350 106759 1873 46055 2036 92000 24450 29991 48486 ...  
## $ actor\_3\_name : Factor w/ 3522 levels "","\xcc\xd2scar Jaenada",..: 3442 1393 1769 2714 1969 2162 3018 57 1134 1393 ...  
## $ facenumber\_in\_poster : int 0 0 0 1 0 1 4 0 0 2 ...  
## $ plot\_keywords : Factor w/ 4761 levels "","10 year old|dog|florida|girl|supermarket",..: 1320 4283 3484 651 4745 29 1142 1564 3312 2188 ...  
## $ num\_user\_for\_reviews : int 3054 1238 2701 738 1902 387 1117 3018 2367 1832 ...  
## $ country : Factor w/ 66 levels "","Afghanistan",..: 65 65 65 65 65 65 65 65 65 65 ...  
## $ content\_rating : Factor w/ 19 levels "","Approved",..: 10 10 10 10 10 9 10 10 10 10 ...  
## $ budget : num 2.37e+08 3.00e+08 2.50e+08 2.64e+08 2.58e+08 ...  
## $ title\_year : int 2009 2007 2012 2012 2007 2010 2015 2016 2006 2006 ...  
## $ actor\_2\_facebook\_likes : int 936 5000 23000 632 11000 553 21000 4000 10000 5000 ...  
## $ imdb\_score : num 7.9 7.1 8.5 6.6 6.2 7.8 7.5 6.9 6.1 7.3 ...  
## $ aspect\_ratio : num 1.78 2.35 2.35 2.35 2.35 1.85 2.35 2.35 2.35 2.35 ...  
## $ movie\_facebook\_likes : int 33000 0 164000 24000 0 29000 118000 197000 0 5000 ...

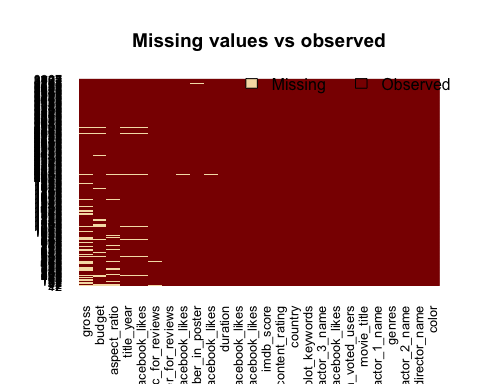
Check for missing values:

library(Amelia)

## Loading required package: Rcpp

## ##   
## ## Amelia II: Multiple Imputation  
## ## (Version 1.7.4, built: 2015-12-05)  
## ## Copyright (C) 2005-2017 James Honaker, Gary King and Matthew Blackwell  
## ## Refer to http://gking.harvard.edu/amelia/ for more information  
## ##

missmap(mm, main = "Missing values vs observed")



sapply(mm,function(x) sum(is.na(x))) # number of missing values for each variable

## color director\_name   
## 0 0   
## num\_critic\_for\_reviews duration   
## 39 6   
## director\_facebook\_likes actor\_3\_facebook\_likes   
## 74 13   
## actor\_2\_name actor\_1\_facebook\_likes   
## 0 4   
## gross genres   
## 572 0   
## actor\_1\_name movie\_title   
## 0 0   
## num\_voted\_users cast\_total\_facebook\_likes   
## 0 0   
## actor\_3\_name facenumber\_in\_poster   
## 0 12   
## plot\_keywords num\_user\_for\_reviews   
## 0 13   
## country content\_rating   
## 0 0   
## budget title\_year   
## 298 74   
## actor\_2\_facebook\_likes imdb\_score   
## 7 0   
## aspect\_ratio movie\_facebook\_likes   
## 222 0

We noticed that there are many missing values for budget,aspect ratio and gross.

Omit missing values:

movie<-na.omit(mm)  
sapply(movie,function(x) sum(is.na(x))) # double check for missing values

## color director\_name   
## 0 0   
## num\_critic\_for\_reviews duration   
## 0 0   
## director\_facebook\_likes actor\_3\_facebook\_likes   
## 0 0   
## actor\_2\_name actor\_1\_facebook\_likes   
## 0 0   
## gross genres   
## 0 0   
## actor\_1\_name movie\_title   
## 0 0   
## num\_voted\_users cast\_total\_facebook\_likes   
## 0 0   
## actor\_3\_name facenumber\_in\_poster   
## 0 0   
## plot\_keywords num\_user\_for\_reviews   
## 0 0   
## country content\_rating   
## 0 0   
## budget title\_year   
## 0 0   
## actor\_2\_facebook\_likes imdb\_score   
## 0 0   
## aspect\_ratio movie\_facebook\_likes   
## 0 0

library(psych)  
library(car)

##   
## Attaching package: 'car'

## The following object is masked from 'package:psych':  
##   
## logit

library(RColorBrewer)   
library(corrplot)  
library(ggplot2)

##   
## Attaching package: 'ggplot2'

## The following objects are masked from 'package:psych':  
##   
## %+%, alpha

Explore title\_year predictor:

range(movie$title\_year) # check movie title year

## [1] 1920 2016

sum(with(movie,title\_year=='2009')) # 145

## [1] 145

sum(with(movie,title\_year=='2014')) # 121

## [1] 121

Visualization of title Year vs. Score:

#library(scatter)  
#scatterplot(x=movie$title\_year,y=movie$imdb\_score)

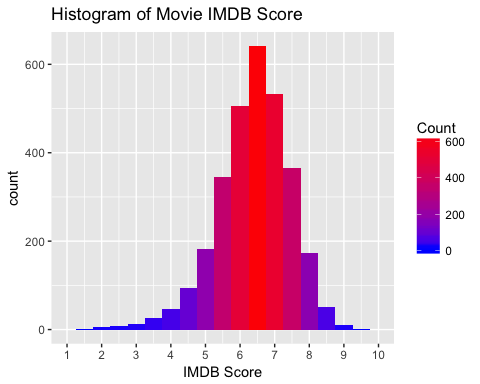
There are many outliers for title year. The mojority of data points are around the year of 2000 and later,which make sense that this is less movies in the early years. Also, an intering notice is that movies from early years tend to have higher scores.

Visualization of IMDB Score:

max(movie$imdb\_score) # 9.4

## [1] 9.3

ggplot(movie, aes(x = imdb\_score)) +  
 geom\_histogram(aes(fill = ..count..), binwidth =0.5) +  
 scale\_x\_continuous(name = "IMDB Score",  
 breaks = seq(0,10),  
 limits=c(1, 10)) +  
 ggtitle("Histogram of Movie IMDB Score") +  
 scale\_fill\_gradient("Count", low = "blue", high = "red")



sum(with(movie,imdb\_score>=8))

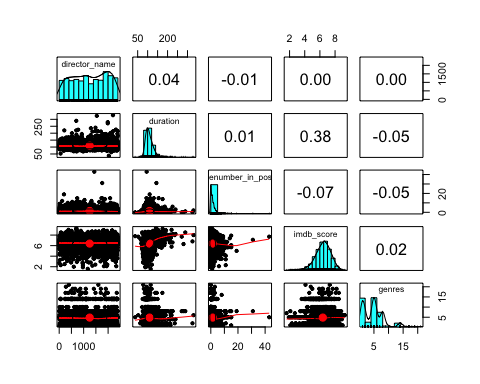
## [1] 148

# 148 movies with IMDB score greater or equal to 8.

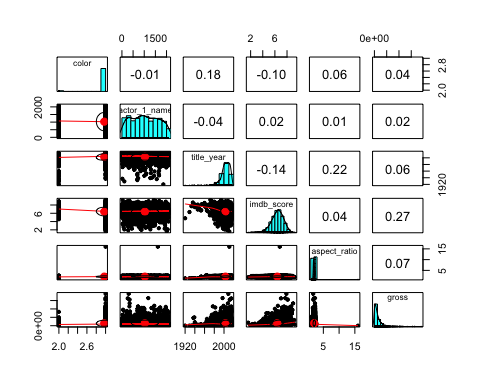
IMDB score looks normal.The highest score is 9.4 out of scale 10. And we can consider movies with a score greater or equal to 8 a great movie from many perspectives.

Exploring correlation :

pairs.panels(movie[c('director\_name','duration','facenumber\_in\_poster','imdb\_score','genres')])

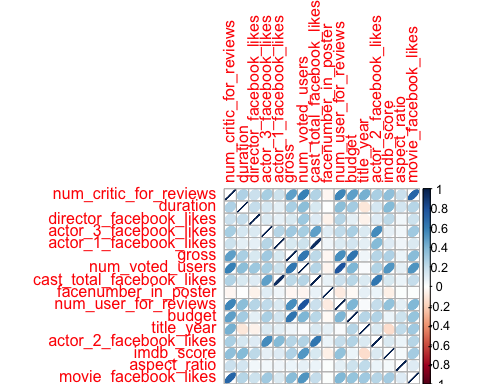
 from the plot, only duration and IMBD score has a high correlation. face number in posters has a negative correaltion with IMBD score. genre has little correlatin with score Interesting, director name has no correlation with IMDB score

pairs.panels(movie[c('color','actor\_1\_name','title\_year','imdb\_score','aspect\_ratio','gross')])

 Color and title year has highly positive correlation. Color and aspect ratia,gross has smaller positive correlations. Actor 1 namem has very small positive correlation with gross, meaning who plays the movies does not have impact on the gross. Title year and aspect ratio and color are highly positively correlated. IMDB score has very small positive correlation with actor 1 name ,which means who was the actor 1 does not make the movie has a higher score. Interestingly, IMDB score has a negative correlation with title year,which means the old movies seems to have a higher score. the result agrees with out pbservation from the scatter plot. IMDB and aspect ratio has small positive correlation. IMDB has a strong positive correlation with gross.

Corplot for all numerical variables:

nums<- sapply(movie,is.numeric) # select numeric columns  
movie.num<- movie[,nums]  
corrplot(cor(movie.num),method='ellipse')

 Note: corrplot cannot use data.frame, use cor() to change it to matrix.

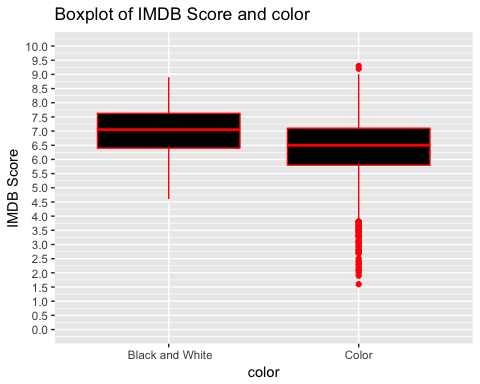
From the correlation plot, we can tell that: Face number in poster has negative correlation with all other predictors. Cast total facebook likes and actor 1 facebook likes has a stronger positive correlation. budget and gross have strong correaltion which is not surprising. Interestingly, IMDB scores has strong positive corrlation with number of critics for review, which means the more the critics review, the higher the score.Duration and number of voted users also have strong positive correlation with IMDB scores.

Find the pairs of correlations

corr.test(movie.num,y=NULL,use='pairwise',method='pearson',adjust='holm',alpha=0.05) # x must be numeric

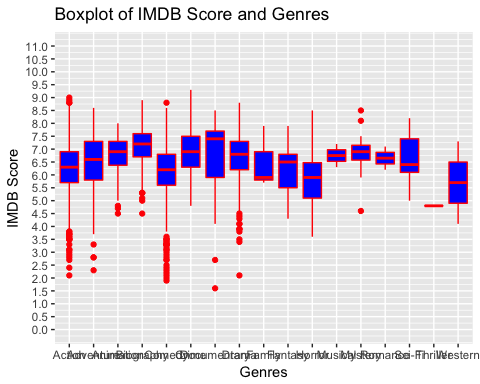
## Call:corr.test(x = movie.num, y = NULL, use = "pairwise", method = "pearson",   
## adjust = "holm", alpha = 0.05)  
## Correlation matrix   
## num\_critic\_for\_reviews duration  
## num\_critic\_for\_reviews 1.00 0.26  
## duration 0.26 1.00  
## director\_facebook\_likes 0.19 0.21  
## actor\_3\_facebook\_likes 0.28 0.14  
## actor\_1\_facebook\_likes 0.17 0.09  
## gross 0.48 0.28  
## num\_voted\_users 0.60 0.37  
## cast\_total\_facebook\_likes 0.25 0.13  
## facenumber\_in\_poster -0.03 0.01  
## num\_user\_for\_reviews 0.57 0.36  
## budget 0.49 0.30  
## title\_year 0.42 -0.11  
## actor\_2\_facebook\_likes 0.28 0.15  
## imdb\_score 0.36 0.38  
## aspect\_ratio 0.18 0.16  
## movie\_facebook\_likes 0.71 0.25  
## director\_facebook\_likes actor\_3\_facebook\_likes  
## num\_critic\_for\_reviews 0.19 0.28  
## duration 0.21 0.14  
## director\_facebook\_likes 1.00 0.12  
## actor\_3\_facebook\_likes 0.12 1.00  
## actor\_1\_facebook\_likes 0.09 0.25  
## gross 0.14 0.30  
## num\_voted\_users 0.32 0.28  
## cast\_total\_facebook\_likes 0.12 0.48  
## facenumber\_in\_poster -0.05 0.10  
## num\_user\_for\_reviews 0.24 0.22  
## budget 0.09 0.27  
## title\_year -0.06 0.13  
## actor\_2\_facebook\_likes 0.12 0.55  
## imdb\_score 0.22 0.09  
## aspect\_ratio 0.05 0.05  
## movie\_facebook\_likes 0.17 0.31  
## actor\_1\_facebook\_likes gross num\_voted\_users  
## num\_critic\_for\_reviews 0.17 0.48 0.60  
## duration 0.09 0.28 0.37  
## director\_facebook\_likes 0.09 0.14 0.32  
## actor\_3\_facebook\_likes 0.25 0.30 0.28  
## actor\_1\_facebook\_likes 1.00 0.13 0.17  
## gross 0.13 1.00 0.64  
## num\_voted\_users 0.17 0.64 1.00  
## cast\_total\_facebook\_likes 0.95 0.22 0.25  
## facenumber\_in\_poster 0.05 -0.04 -0.04  
## num\_user\_for\_reviews 0.12 0.55 0.78  
## budget 0.15 0.64 0.40  
## title\_year 0.09 0.06 0.03  
## actor\_2\_facebook\_likes 0.38 0.25 0.25  
## imdb\_score 0.12 0.27 0.51  
## aspect\_ratio 0.05 0.07 0.09  
## movie\_facebook\_likes 0.12 0.38 0.52  
## cast\_total\_facebook\_likes facenumber\_in\_poster  
## num\_critic\_for\_reviews 0.25 -0.03  
## duration 0.13 0.01  
## director\_facebook\_likes 0.12 -0.05  
## actor\_3\_facebook\_likes 0.48 0.10  
## actor\_1\_facebook\_likes 0.95 0.05  
## gross 0.22 -0.04  
## num\_voted\_users 0.25 -0.04  
## cast\_total\_facebook\_likes 1.00 0.07  
## facenumber\_in\_poster 0.07 1.00  
## num\_user\_for\_reviews 0.18 -0.09  
## budget 0.23 -0.03  
## title\_year 0.13 0.08  
## actor\_2\_facebook\_likes 0.63 0.07  
## imdb\_score 0.14 -0.07  
## aspect\_ratio 0.07 0.01  
## movie\_facebook\_likes 0.21 0.01  
## num\_user\_for\_reviews budget title\_year  
## num\_critic\_for\_reviews 0.57 0.49 0.42  
## duration 0.36 0.30 -0.11  
## director\_facebook\_likes 0.24 0.09 -0.06  
## actor\_3\_facebook\_likes 0.22 0.27 0.13  
## actor\_1\_facebook\_likes 0.12 0.15 0.09  
## gross 0.55 0.64 0.06  
## num\_voted\_users 0.78 0.40 0.03  
## cast\_total\_facebook\_likes 0.18 0.23 0.13  
## facenumber\_in\_poster -0.09 -0.03 0.08  
## num\_user\_for\_reviews 1.00 0.40 0.03  
## budget 0.40 1.00 0.25  
## title\_year 0.03 0.25 1.00  
## actor\_2\_facebook\_likes 0.20 0.25 0.13  
## imdb\_score 0.35 0.07 -0.14  
## aspect\_ratio 0.10 0.18 0.22  
## movie\_facebook\_likes 0.39 0.33 0.31  
## actor\_2\_facebook\_likes imdb\_score aspect\_ratio  
## num\_critic\_for\_reviews 0.28 0.36 0.18  
## duration 0.15 0.38 0.16  
## director\_facebook\_likes 0.12 0.22 0.05  
## actor\_3\_facebook\_likes 0.55 0.09 0.05  
## actor\_1\_facebook\_likes 0.38 0.12 0.05  
## gross 0.25 0.27 0.07  
## num\_voted\_users 0.25 0.51 0.09  
## cast\_total\_facebook\_likes 0.63 0.14 0.07  
## facenumber\_in\_poster 0.07 -0.07 0.01  
## num\_user\_for\_reviews 0.20 0.35 0.10  
## budget 0.25 0.07 0.18  
## title\_year 0.13 -0.14 0.22  
## actor\_2\_facebook\_likes 1.00 0.13 0.07  
## imdb\_score 0.13 1.00 0.04  
## aspect\_ratio 0.07 0.04 1.00  
## movie\_facebook\_likes 0.25 0.29 0.11  
## movie\_facebook\_likes  
## num\_critic\_for\_reviews 0.71  
## duration 0.25  
## director\_facebook\_likes 0.17  
## actor\_3\_facebook\_likes 0.31  
## actor\_1\_facebook\_likes 0.12  
## gross 0.38  
## num\_voted\_users 0.52  
## cast\_total\_facebook\_likes 0.21  
## facenumber\_in\_poster 0.01  
## num\_user\_for\_reviews 0.39  
## budget 0.33  
## title\_year 0.31  
## actor\_2\_facebook\_likes 0.25  
## imdb\_score 0.29  
## aspect\_ratio 0.11  
## movie\_facebook\_likes 1.00  
## Sample Size   
## [1] 3005  
## Probability values (Entries above the diagonal are adjusted for multiple tests.)   
## num\_critic\_for\_reviews duration  
## num\_critic\_for\_reviews 0.00 0.00  
## duration 0.00 0.00  
## director\_facebook\_likes 0.00 0.00  
## actor\_3\_facebook\_likes 0.00 0.00  
## actor\_1\_facebook\_likes 0.00 0.00  
## gross 0.00 0.00  
## num\_voted\_users 0.00 0.00  
## cast\_total\_facebook\_likes 0.00 0.00  
## facenumber\_in\_poster 0.09 0.66  
## num\_user\_for\_reviews 0.00 0.00  
## budget 0.00 0.00  
## title\_year 0.00 0.00  
## actor\_2\_facebook\_likes 0.00 0.00  
## imdb\_score 0.00 0.00  
## aspect\_ratio 0.00 0.00  
## movie\_facebook\_likes 0.00 0.00  
## director\_facebook\_likes actor\_3\_facebook\_likes  
## num\_critic\_for\_reviews 0.00 0.00  
## duration 0.00 0.00  
## director\_facebook\_likes 0.00 0.00  
## actor\_3\_facebook\_likes 0.00 0.00  
## actor\_1\_facebook\_likes 0.00 0.00  
## gross 0.00 0.00  
## num\_voted\_users 0.00 0.00  
## cast\_total\_facebook\_likes 0.00 0.00  
## facenumber\_in\_poster 0.00 0.00  
## num\_user\_for\_reviews 0.00 0.00  
## budget 0.00 0.00  
## title\_year 0.00 0.00  
## actor\_2\_facebook\_likes 0.00 0.00  
## imdb\_score 0.00 0.00  
## aspect\_ratio 0.01 0.01  
## movie\_facebook\_likes 0.00 0.00  
## actor\_1\_facebook\_likes gross num\_voted\_users  
## num\_critic\_for\_reviews 0.00 0.00 0.00  
## duration 0.00 0.00 0.00  
## director\_facebook\_likes 0.00 0.00 0.00  
## actor\_3\_facebook\_likes 0.00 0.00 0.00  
## actor\_1\_facebook\_likes 0.00 0.00 0.00  
## gross 0.00 0.00 0.00  
## num\_voted\_users 0.00 0.00 0.00  
## cast\_total\_facebook\_likes 0.00 0.00 0.00  
## facenumber\_in\_poster 0.01 0.05 0.02  
## num\_user\_for\_reviews 0.00 0.00 0.00  
## budget 0.00 0.00 0.00  
## title\_year 0.00 0.00 0.10  
## actor\_2\_facebook\_likes 0.00 0.00 0.00  
## imdb\_score 0.00 0.00 0.00  
## aspect\_ratio 0.00 0.00 0.00  
## movie\_facebook\_likes 0.00 0.00 0.00  
## cast\_total\_facebook\_likes facenumber\_in\_poster  
## num\_critic\_for\_reviews 0 0.65  
## duration 0 1.00  
## director\_facebook\_likes 0 0.06  
## actor\_3\_facebook\_likes 0 0.00  
## actor\_1\_facebook\_likes 0 0.07  
## gross 0 0.37  
## num\_voted\_users 0 0.17  
## cast\_total\_facebook\_likes 0 0.00  
## facenumber\_in\_poster 0 0.00  
## num\_user\_for\_reviews 0 0.00  
## budget 0 0.14  
## title\_year 0 0.00  
## actor\_2\_facebook\_likes 0 0.00  
## imdb\_score 0 0.00  
## aspect\_ratio 0 0.55  
## movie\_facebook\_likes 0 0.50  
## num\_user\_for\_reviews budget title\_year  
## num\_critic\_for\_reviews 0.00 0.00 0.00  
## duration 0.00 0.00 0.00  
## director\_facebook\_likes 0.00 0.00 0.04  
## actor\_3\_facebook\_likes 0.00 0.00 0.00  
## actor\_1\_facebook\_likes 0.00 0.00 0.00  
## gross 0.00 0.00 0.04  
## num\_voted\_users 0.00 0.00 0.65  
## cast\_total\_facebook\_likes 0.00 0.00 0.00  
## facenumber\_in\_poster 0.00 0.65 0.00  
## num\_user\_for\_reviews 0.00 0.00 0.65  
## budget 0.00 0.00 0.00  
## title\_year 0.12 0.00 0.00  
## actor\_2\_facebook\_likes 0.00 0.00 0.00  
## imdb\_score 0.00 0.00 0.00  
## aspect\_ratio 0.00 0.00 0.00  
## movie\_facebook\_likes 0.00 0.00 0.00  
## actor\_2\_facebook\_likes imdb\_score aspect\_ratio  
## num\_critic\_for\_reviews 0.00 0.00 0.00  
## duration 0.00 0.00 0.00  
## director\_facebook\_likes 0.00 0.00 0.10  
## actor\_3\_facebook\_likes 0.00 0.00 0.07  
## actor\_1\_facebook\_likes 0.00 0.00 0.05  
## gross 0.00 0.00 0.00  
## num\_voted\_users 0.00 0.00 0.00  
## cast\_total\_facebook\_likes 0.00 0.00 0.00  
## facenumber\_in\_poster 0.01 0.00 1.00  
## num\_user\_for\_reviews 0.00 0.00 0.00  
## budget 0.00 0.00 0.00  
## title\_year 0.00 0.00 0.00  
## actor\_2\_facebook\_likes 0.00 0.00 0.00  
## imdb\_score 0.00 0.00 0.34  
## aspect\_ratio 0.00 0.04 0.00  
## movie\_facebook\_likes 0.00 0.00 0.00  
## movie\_facebook\_likes  
## num\_critic\_for\_reviews 0  
## duration 0  
## director\_facebook\_likes 0  
## actor\_3\_facebook\_likes 0  
## actor\_1\_facebook\_likes 0  
## gross 0  
## num\_voted\_users 0  
## cast\_total\_facebook\_likes 0  
## facenumber\_in\_poster 1  
## num\_user\_for\_reviews 0  
## budget 0  
## title\_year 0  
## actor\_2\_facebook\_likes 0  
## imdb\_score 0  
## aspect\_ratio 0  
## movie\_facebook\_likes 0  
##   
## To see confidence intervals of the correlations, print with the short=FALSE option

# Boxplots for significant categorical predictors  
fill <- "Black"  
line <- "Red"  
ggplot(movie, aes(x = color, y =imdb\_score)) +  
 geom\_boxplot(fill = fill, colour = line) +  
 scale\_y\_continuous(name = "IMDB Score",  
 breaks = seq(0, 10, 0.5),  
 limits=c(0, 10)) +  
 scale\_x\_discrete(name = "color") +  
 ggtitle("Boxplot of IMDB Score and color")

 Black and white movies seems to have a hither meadian rate, and overall a little higher scores. Colors movies have many outliers.

Boxplot for genre:

fill <- "Blue"  
line <- "Red"  
ggplot(movie, aes(x = genres, y =imdb\_score)) +  
 geom\_boxplot(fill = fill, colour = line) +  
 scale\_y\_continuous(name = "IMDB Score",  
 breaks = seq(0, 11, 0.5),  
 limits=c(0, 11)) +  
 scale\_x\_discrete(name = "Genres") +  
 ggtitle("Boxplot of IMDB Score and Genres")

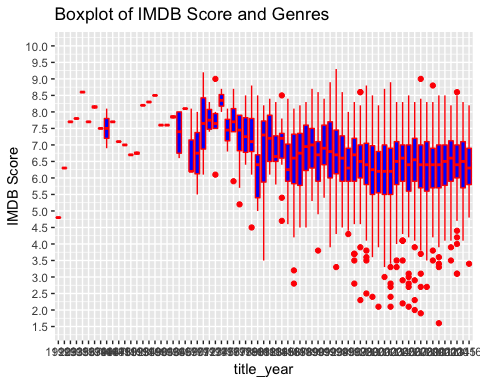
 From the boxplot of genres, "Documentation" has the highest median score.And Trill movies has the lowest median. But it is also because there is 1 observation for thrill movies in our data set.

summary(movie$genres)

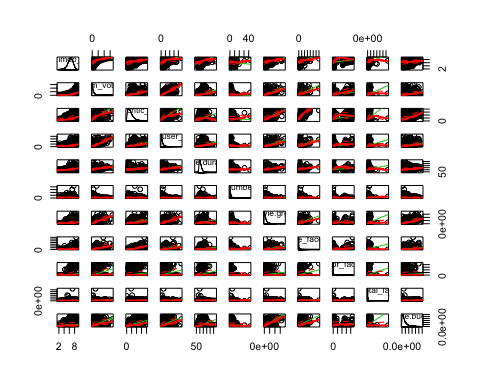
## Action Adventure Animation Biography Comedy Crime   
## 751 291 36 137 853 204   
## Documentary Drama Family Fantasy Film-Noir Game-Show   
## 25 506 3 31 0 0   
## History Horror Music Musical Mystery Romance   
## 0 138 0 2 16 2   
## Sci-Fi Thriller Western   
## 7 1 2

# Boxplots for "title year':

library(ggplot2)  
fill <- "Blue"  
line <- "Red"  
ggplot(movie, aes(x = as.factor(title\_year), y =imdb\_score)) +  
 geom\_boxplot(fill = fill, colour = line) +  
 scale\_y\_continuous(name = "IMDB Score",  
 breaks = seq(1.5, 10, 0.5),  
 limits=c(1.5, 10)) +  
 scale\_x\_discrete(name = "title\_year") +  
 ggtitle("Boxplot of IMDB Score and Genres")

 The median of imdb score of all years seem different. So let's try to treat title\_year as categorical.

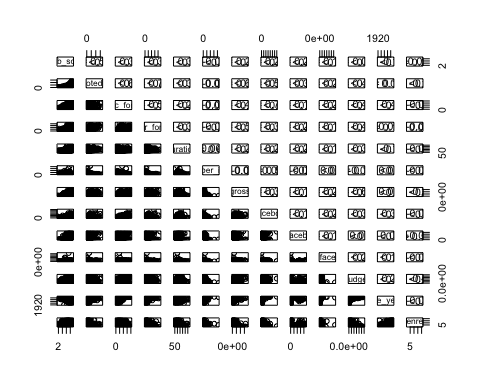
# Scatter plot matrix for correlation significant numerical variables  
scatterplotMatrix(~movie$imdb\_score+movie$num\_voted\_users+movie$num\_critic\_for\_reviews+movie$num\_user\_for\_reviews+movie$duration+movie$facenumber\_in\_poster+movie$gross+movie$movie\_facebook\_likes+movie$director\_facebook\_likes+movie$cast\_total\_facebook\_likes+movie$budget)



## Step 3: fitting regression model

movie.sig<-movie[,c('imdb\_score','num\_voted\_users','num\_critic\_for\_reviews','num\_user\_for\_reviews','duration','facenumber\_in\_poster','gross','movie\_facebook\_likes','director\_facebook\_likes','cast\_total\_facebook\_likes','budget','title\_year','genres')]

panel.cor <- function(x, y, digits = 2, cex.cor, ...)  
{  
 usr <- par("usr"); on.exit(par(usr))  
 par(usr = c(0, 1, 0, 1))  
 # correlation coefficient  
 r <- cor(x, y)  
 txt <- format(c(r, 0.123456789), digits = digits)[1]  
 txt <- paste("r= ", txt, sep = "")  
 text(0.5, 0.6, txt)  
  
 # p-value calculation  
 p <- cor.test(x, y)$p.value  
 txt2 <- format(c(p, 0.123456789), digits = digits)[1]  
 txt2 <- paste("p= ", txt2, sep = "")  
 if(p<0.01) txt2 <- paste("p= ", "<0.01", sep = "")  
 text(0.5, 0.4, txt2)  
}  
  
pairs(movie.sig, upper.panel = panel.cor)



Step function to check AIC criteria:

null=lm(movie.sig$imdb\_score~1) # set null model  
summary(null)

##   
## Call:  
## lm(formula = movie.sig$imdb\_score ~ 1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.7873 -0.5873 0.1127 0.7127 2.9127   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.3873 0.0192 332.6 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.053 on 3004 degrees of freedom

1. Full model is linear additive model

full1=lm(movie.sig$imdb\_score~movie.sig$num\_voted\_users+movie.sig$num\_critic\_for\_reviews+movie.sig$num\_user\_for\_reviews+movie.sig$duration+movie.sig$facenumber\_in\_poster+movie.sig$gross+movie.sig$movie\_facebook\_likes+movie.sig$director\_facebook\_likes+movie.sig$cast\_total\_facebook\_likes+movie.sig$budget+movie.sig$title\_year+factor(movie.sig$genres))  
summary(full1)

##   
## Call:  
## lm(formula = movie.sig$imdb\_score ~ movie.sig$num\_voted\_users +   
## movie.sig$num\_critic\_for\_reviews + movie.sig$num\_user\_for\_reviews +   
## movie.sig$duration + movie.sig$facenumber\_in\_poster + movie.sig$gross +   
## movie.sig$movie\_facebook\_likes + movie.sig$director\_facebook\_likes +   
## movie.sig$cast\_total\_facebook\_likes + movie.sig$budget +   
## movie.sig$title\_year + factor(movie.sig$genres))  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.9157 -0.3693 0.0835 0.4993 2.0350   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 5.413e+01 3.604e+00 15.019 < 2e-16  
## movie.sig$num\_voted\_users 3.158e-06 1.757e-07 17.969 < 2e-16  
## movie.sig$num\_critic\_for\_reviews 3.333e-03 2.119e-04 15.727 < 2e-16  
## movie.sig$num\_user\_for\_reviews -4.887e-04 5.976e-05 -8.177 4.26e-16  
## movie.sig$duration 8.491e-03 7.848e-04 10.820 < 2e-16  
## movie.sig$facenumber\_in\_poster -1.750e-02 6.947e-03 -2.519 0.01182  
## movie.sig$gross 2.247e-10 3.096e-10 0.726 0.46808  
## movie.sig$movie\_facebook\_likes -4.007e-06 9.702e-07 -4.131 3.72e-05  
## movie.sig$director\_facebook\_likes 2.832e-07 4.562e-06 0.062 0.95051  
## movie.sig$cast\_total\_facebook\_likes 1.110e-06 7.323e-07 1.516 0.12975  
## movie.sig$budget -4.486e-09 5.125e-10 -8.753 < 2e-16  
## movie.sig$title\_year -2.467e-02 1.797e-03 -13.727 < 2e-16  
## factor(movie.sig$genres)Adventure 3.458e-01 5.448e-02 6.347 2.53e-10  
## factor(movie.sig$genres)Animation 6.621e-01 1.345e-01 4.924 8.93e-07  
## factor(movie.sig$genres)Biography 6.557e-01 7.661e-02 8.558 < 2e-16  
## factor(movie.sig$genres)Comedy 1.532e-01 4.361e-02 3.513 0.00045  
## factor(movie.sig$genres)Crime 4.551e-01 6.464e-02 7.040 2.37e-12  
## factor(movie.sig$genres)Documentary 9.270e-01 1.608e-01 5.765 8.98e-09  
## factor(movie.sig$genres)Drama 5.326e-01 4.904e-02 10.861 < 2e-16  
## factor(movie.sig$genres)Family 2.201e-01 4.521e-01 0.487 0.62639  
## factor(movie.sig$genres)Fantasy -1.629e-01 1.448e-01 -1.125 0.26068  
## factor(movie.sig$genres)Horror -3.858e-01 7.777e-02 -4.961 7.41e-07  
## factor(movie.sig$genres)Musical -4.133e-01 5.573e-01 -0.742 0.45839  
## factor(movie.sig$genres)Mystery 1.968e-01 1.979e-01 0.995 0.32005  
## factor(movie.sig$genres)Romance 5.466e-01 5.506e-01 0.993 0.32095  
## factor(movie.sig$genres)Sci-Fi 2.551e-01 2.960e-01 0.862 0.38870  
## factor(movie.sig$genres)Thriller -4.301e-01 7.786e-01 -0.552 0.58077  
## factor(movie.sig$genres)Western -1.037e-01 5.521e-01 -0.188 0.85101  
##   
## (Intercept) \*\*\*  
## movie.sig$num\_voted\_users \*\*\*  
## movie.sig$num\_critic\_for\_reviews \*\*\*  
## movie.sig$num\_user\_for\_reviews \*\*\*  
## movie.sig$duration \*\*\*  
## movie.sig$facenumber\_in\_poster \*   
## movie.sig$gross   
## movie.sig$movie\_facebook\_likes \*\*\*  
## movie.sig$director\_facebook\_likes   
## movie.sig$cast\_total\_facebook\_likes   
## movie.sig$budget \*\*\*  
## movie.sig$title\_year \*\*\*  
## factor(movie.sig$genres)Adventure \*\*\*  
## factor(movie.sig$genres)Animation \*\*\*  
## factor(movie.sig$genres)Biography \*\*\*  
## factor(movie.sig$genres)Comedy \*\*\*  
## factor(movie.sig$genres)Crime \*\*\*  
## factor(movie.sig$genres)Documentary \*\*\*  
## factor(movie.sig$genres)Drama \*\*\*  
## factor(movie.sig$genres)Family   
## factor(movie.sig$genres)Fantasy   
## factor(movie.sig$genres)Horror \*\*\*  
## factor(movie.sig$genres)Musical   
## factor(movie.sig$genres)Mystery   
## factor(movie.sig$genres)Romance   
## factor(movie.sig$genres)Sci-Fi   
## factor(movie.sig$genres)Thriller   
## factor(movie.sig$genres)Western   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7768 on 2977 degrees of freedom  
## Multiple R-squared: 0.4604, Adjusted R-squared: 0.4555   
## F-statistic: 94.07 on 27 and 2977 DF, p-value: < 2.2e-16

step(null,scope = list(lower=null,upper=full1),direction = 'forward')

## Start: AIC=309.81  
## movie.sig$imdb\_score ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$num\_voted\_users 1 871.90 2457.2 -600.74  
## + movie.sig$duration 1 491.13 2838.0 -167.82  
## + movie.sig$num\_critic\_for\_reviews 1 428.38 2900.8 -102.10  
## + movie.sig$num\_user\_for\_reviews 1 407.62 2921.5 -80.68  
## + factor(movie.sig$genres) 16 331.02 2998.1 27.10  
## + movie.sig$movie\_facebook\_likes 1 282.82 3046.3 45.02  
## + movie.sig$gross 1 242.62 3086.5 84.42  
## + movie.sig$director\_facebook\_likes 1 166.17 3163.0 157.95  
## + movie.sig$title\_year 1 69.27 3259.9 248.63  
## + movie.sig$cast\_total\_facebook\_likes 1 64.28 3264.8 253.22  
## + movie.sig$budget 1 16.26 3312.9 297.09  
## + movie.sig$facenumber\_in\_poster 1 15.14 3314.0 298.11  
## <none> 3329.1 309.81  
##   
## Step: AIC=-600.74  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users  
##   
## Df Sum of Sq RSS AIC  
## + factor(movie.sig$genres) 16 311.531 2145.7 -976.12  
## + movie.sig$duration 1 147.786 2309.4 -785.13  
## + movie.sig$title\_year 1 84.649 2372.6 -704.08  
## + movie.sig$budget 1 73.211 2384.0 -689.63  
## + movie.sig$num\_user\_for\_reviews 1 21.297 2435.9 -624.90  
## + movie.sig$gross 1 16.929 2440.3 -619.51  
## + movie.sig$num\_critic\_for\_reviews 1 14.632 2442.6 -616.69  
## + movie.sig$director\_facebook\_likes 1 13.657 2443.6 -615.49  
## + movie.sig$facenumber\_in\_poster 1 6.789 2450.4 -607.05  
## + movie.sig$movie\_facebook\_likes 1 2.627 2454.6 -601.95  
## <none> 2457.2 -600.74  
## + movie.sig$cast\_total\_facebook\_likes 1 0.524 2456.7 -599.38  
##   
## Step: AIC=-976.12  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres)  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$title\_year 1 79.623 2066.1 -1087.75  
## + movie.sig$duration 1 74.584 2071.1 -1080.44  
## + movie.sig$budget 1 28.689 2117.0 -1014.57  
## + movie.sig$num\_critic\_for\_reviews 1 23.116 2122.6 -1006.67  
## + movie.sig$num\_user\_for\_reviews 1 12.251 2133.4 -991.33  
## + movie.sig$director\_facebook\_likes 1 3.707 2142.0 -979.32  
## + movie.sig$facenumber\_in\_poster 1 3.274 2142.4 -978.71  
## + movie.sig$movie\_facebook\_likes 1 1.686 2144.0 -976.49  
## <none> 2145.7 -976.12  
## + movie.sig$gross 1 1.391 2144.3 -976.07  
## + movie.sig$cast\_total\_facebook\_likes 1 0.362 2145.3 -974.63  
##   
## Step: AIC=-1087.75  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$num\_critic\_for\_reviews 1 125.091 1941.0 -1273.4  
## + movie.sig$duration 1 55.857 2010.2 -1168.1  
## + movie.sig$movie\_facebook\_likes 1 21.746 2044.3 -1117.5  
## + movie.sig$num\_user\_for\_reviews 1 11.741 2054.3 -1102.9  
## + movie.sig$budget 1 9.196 2056.9 -1099.2  
## + movie.sig$cast\_total\_facebook\_likes 1 2.923 2063.2 -1090.0  
## + movie.sig$director\_facebook\_likes 1 1.740 2064.3 -1088.3  
## <none> 2066.1 -1087.8  
## + movie.sig$facenumber\_in\_poster 1 1.084 2065.0 -1087.3  
## + movie.sig$gross 1 0.638 2065.4 -1086.7  
##   
## Step: AIC=-1273.43  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$budget 1 36.627 1904.4 -1328.7  
## + movie.sig$num\_user\_for\_reviews 1 35.326 1905.7 -1326.6  
## + movie.sig$duration 1 34.873 1906.1 -1325.9  
## + movie.sig$gross 1 7.359 1933.6 -1282.8  
## + movie.sig$movie\_facebook\_likes 1 1.397 1939.6 -1273.6  
## <none> 1941.0 -1273.4  
## + movie.sig$facenumber\_in\_poster 1 0.926 1940.1 -1272.9  
## + movie.sig$director\_facebook\_likes 1 0.644 1940.3 -1272.4  
## + movie.sig$cast\_total\_facebook\_likes 1 0.572 1940.4 -1272.3  
##   
## Step: AIC=-1328.68  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$duration 1 58.373 1846.0 -1420.2  
## + movie.sig$num\_user\_for\_reviews 1 27.052 1877.3 -1369.7  
## + movie.sig$movie\_facebook\_likes 1 2.576 1901.8 -1330.8  
## + movie.sig$cast\_total\_facebook\_likes 1 2.005 1902.3 -1329.8  
## <none> 1904.4 -1328.7  
## + movie.sig$facenumber\_in\_poster 1 1.071 1903.3 -1328.4  
## + movie.sig$director\_facebook\_likes 1 0.557 1903.8 -1327.6  
## + movie.sig$gross 1 0.074 1904.3 -1326.8  
##   
## Step: AIC=-1420.23  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$num\_user\_for\_reviews 1 33.825 1812.2 -1473.8  
## + movie.sig$movie\_facebook\_likes 1 4.702 1841.3 -1425.9  
## + movie.sig$facenumber\_in\_poster 1 2.488 1843.5 -1422.3  
## + movie.sig$cast\_total\_facebook\_likes 1 1.601 1844.4 -1420.8  
## <none> 1846.0 -1420.2  
## + movie.sig$gross 1 0.196 1845.8 -1418.5  
## + movie.sig$director\_facebook\_likes 1 0.043 1845.9 -1418.3  
##   
## Step: AIC=-1473.81  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration + movie.sig$num\_user\_for\_reviews  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$movie\_facebook\_likes 1 10.4792 1801.7 -1489.2  
## + movie.sig$facenumber\_in\_poster 1 3.7911 1808.4 -1478.1  
## <none> 1812.2 -1473.8  
## + movie.sig$cast\_total\_facebook\_likes 1 0.9926 1811.2 -1473.5  
## + movie.sig$gross 1 0.3569 1811.8 -1472.4  
## + movie.sig$director\_facebook\_likes 1 0.0128 1812.2 -1471.8  
##   
## Step: AIC=-1489.23  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration + movie.sig$num\_user\_for\_reviews +   
## movie.sig$movie\_facebook\_likes  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$facenumber\_in\_poster 1 3.5218 1798.2 -1493.1  
## <none> 1801.7 -1489.2  
## + movie.sig$cast\_total\_facebook\_likes 1 1.0918 1800.6 -1489.0  
## + movie.sig$gross 1 0.3413 1801.3 -1487.8  
## + movie.sig$director\_facebook\_likes 1 0.0167 1801.7 -1487.3  
##   
## Step: AIC=-1493.11  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration + movie.sig$num\_user\_for\_reviews +   
## movie.sig$movie\_facebook\_likes + movie.sig$facenumber\_in\_poster  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$cast\_total\_facebook\_likes 1 1.41883 1796.7 -1493.5  
## <none> 1798.2 -1493.1  
## + movie.sig$gross 1 0.33944 1797.8 -1491.7  
## + movie.sig$director\_facebook\_likes 1 0.00320 1798.2 -1491.1  
##   
## Step: AIC=-1493.48  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration + movie.sig$num\_user\_for\_reviews +   
## movie.sig$movie\_facebook\_likes + movie.sig$facenumber\_in\_poster +   
## movie.sig$cast\_total\_facebook\_likes  
##   
## Df Sum of Sq RSS AIC  
## <none> 1796.7 -1493.5  
## + movie.sig$gross 1 0.31546 1796.4 -1492.0  
## + movie.sig$director\_facebook\_likes 1 0.00000 1796.7 -1491.5

##   
## Call:  
## lm(formula = movie.sig$imdb\_score ~ movie.sig$num\_voted\_users +   
## factor(movie.sig$genres) + movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration + movie.sig$num\_user\_for\_reviews +   
## movie.sig$movie\_facebook\_likes + movie.sig$facenumber\_in\_poster +   
## movie.sig$cast\_total\_facebook\_likes)  
##   
## Coefficients:  
## (Intercept) movie.sig$num\_voted\_users   
## 5.446e+01 3.203e-06   
## factor(movie.sig$genres)Adventure factor(movie.sig$genres)Animation   
## 3.495e-01 6.687e-01   
## factor(movie.sig$genres)Biography factor(movie.sig$genres)Comedy   
## 6.564e-01 1.558e-01   
## factor(movie.sig$genres)Crime factor(movie.sig$genres)Documentary   
## 4.522e-01 9.302e-01   
## factor(movie.sig$genres)Drama factor(movie.sig$genres)Family   
## 5.326e-01 2.466e-01   
## factor(movie.sig$genres)Fantasy factor(movie.sig$genres)Horror   
## -1.616e-01 -3.839e-01   
## factor(movie.sig$genres)Musical factor(movie.sig$genres)Mystery   
## -4.044e-01 1.950e-01   
## factor(movie.sig$genres)Romance factor(movie.sig$genres)Sci-Fi   
## 5.455e-01 2.483e-01   
## factor(movie.sig$genres)Thriller factor(movie.sig$genres)Western   
## -4.271e-01 -9.845e-02   
## movie.sig$title\_year movie.sig$num\_critic\_for\_reviews   
## -2.483e-02 3.339e-03   
## movie.sig$budget movie.sig$duration   
## -4.311e-09 8.481e-03   
## movie.sig$num\_user\_for\_reviews movie.sig$movie\_facebook\_likes   
## -4.876e-04 -4.010e-06   
## movie.sig$facenumber\_in\_poster movie.sig$cast\_total\_facebook\_likes   
## -1.753e-02 1.121e-06

1. full model is polynomial regresison model with interaction terms:

full2=lm(movie.sig$imdb\_score~poly(movie.sig$num\_voted\_users,2)+poly(movie.sig$num\_critic\_for\_reviews,2)+poly(movie.sig$num\_user\_for\_reviews,2)+poly(movie.sig$duration,2)+movie.sig$facenumber\_in\_poster+poly(movie.sig$gross,2)+poly(movie.sig$movie\_facebook\_likes,2)+movie.sig$director\_facebook\_likes+movie.sig$cast\_total\_facebook\_likes+movie.sig$budget+movie.sig$title\_year+movie.sig$genres+movie.sig$facenumber\_in\_poster\*movie.sig$num\_critic\_for\_reviews+movie.sig$num\_user\_for\_reviews\*movie.sig$num\_voted\_users+movie.sig$num\_voted\_users\*movie.sig$gross+movie.sig$gross\*movie.sig$budget)  
summary(full2)

##   
## Call:  
## lm(formula = movie.sig$imdb\_score ~ poly(movie.sig$num\_voted\_users,   
## 2) + poly(movie.sig$num\_critic\_for\_reviews, 2) + poly(movie.sig$num\_user\_for\_reviews,   
## 2) + poly(movie.sig$duration, 2) + movie.sig$facenumber\_in\_poster +   
## poly(movie.sig$gross, 2) + poly(movie.sig$movie\_facebook\_likes,   
## 2) + movie.sig$director\_facebook\_likes + movie.sig$cast\_total\_facebook\_likes +   
## movie.sig$budget + movie.sig$title\_year + movie.sig$genres +   
## movie.sig$facenumber\_in\_poster \* movie.sig$num\_critic\_for\_reviews +   
## movie.sig$num\_user\_for\_reviews \* movie.sig$num\_voted\_users +   
## movie.sig$num\_voted\_users \* movie.sig$gross + movie.sig$gross \*   
## movie.sig$budget)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.3608 -0.3549 0.0642 0.4619 2.1792   
##   
## Coefficients: (4 not defined because of singularities)  
## Estimate  
## (Intercept) 5.948e+01  
## poly(movie.sig$num\_voted\_users, 2)1 2.305e+01  
## poly(movie.sig$num\_voted\_users, 2)2 -1.873e+01  
## poly(movie.sig$num\_critic\_for\_reviews, 2)1 1.393e+01  
## poly(movie.sig$num\_critic\_for\_reviews, 2)2 -9.490e+00  
## poly(movie.sig$num\_user\_for\_reviews, 2)1 -1.760e+01  
## poly(movie.sig$num\_user\_for\_reviews, 2)2 4.166e+00  
## poly(movie.sig$duration, 2)1 1.087e+01  
## poly(movie.sig$duration, 2)2 -3.883e+00  
## movie.sig$facenumber\_in\_poster -2.093e-02  
## poly(movie.sig$gross, 2)1 -1.454e+01  
## poly(movie.sig$gross, 2)2 -5.285e+00  
## poly(movie.sig$movie\_facebook\_likes, 2)1 2.580e+00  
## poly(movie.sig$movie\_facebook\_likes, 2)2 2.283e-01  
## movie.sig$director\_facebook\_likes 4.608e-06  
## movie.sig$cast\_total\_facebook\_likes 2.533e-07  
## movie.sig$budget -7.852e-09  
## movie.sig$title\_year -2.656e-02  
## movie.sig$genresAdventure 3.727e-01  
## movie.sig$genresAnimation 7.564e-01  
## movie.sig$genresBiography 6.264e-01  
## movie.sig$genresComedy 1.576e-01  
## movie.sig$genresCrime 4.558e-01  
## movie.sig$genresDocumentary 9.738e-01  
## movie.sig$genresDrama 5.230e-01  
## movie.sig$genresFamily 5.958e-01  
## movie.sig$genresFantasy -1.891e-01  
## movie.sig$genresHorror -3.533e-01  
## movie.sig$genresMusical -4.744e-01  
## movie.sig$genresMystery 1.947e-01  
## movie.sig$genresRomance 6.094e-01  
## movie.sig$genresSci-Fi 1.471e-01  
## movie.sig$genresThriller -3.085e-01  
## movie.sig$genresWestern -4.204e-02  
## movie.sig$num\_critic\_for\_reviews NA  
## movie.sig$num\_user\_for\_reviews NA  
## movie.sig$num\_voted\_users NA  
## movie.sig$gross NA  
## movie.sig$facenumber\_in\_poster:movie.sig$num\_critic\_for\_reviews -1.291e-06  
## movie.sig$num\_user\_for\_reviews:movie.sig$num\_voted\_users 7.966e-10  
## movie.sig$num\_voted\_users:movie.sig$gross 1.498e-15  
## movie.sig$budget:movie.sig$gross 2.946e-17  
## Std. Error  
## (Intercept) 3.617e+00  
## poly(movie.sig$num\_voted\_users, 2)1 3.426e+00  
## poly(movie.sig$num\_voted\_users, 2)2 2.200e+00  
## poly(movie.sig$num\_critic\_for\_reviews, 2)1 1.661e+00  
## poly(movie.sig$num\_critic\_for\_reviews, 2)2 1.004e+00  
## poly(movie.sig$num\_user\_for\_reviews, 2)1 2.325e+00  
## poly(movie.sig$num\_user\_for\_reviews, 2)2 1.593e+00  
## poly(movie.sig$duration, 2)1 9.246e-01  
## poly(movie.sig$duration, 2)2 7.809e-01  
## movie.sig$facenumber\_in\_poster 1.106e-02  
## poly(movie.sig$gross, 2)1 2.418e+00  
## poly(movie.sig$gross, 2)2 1.483e+00  
## poly(movie.sig$movie\_facebook\_likes, 2)1 1.322e+00  
## poly(movie.sig$movie\_facebook\_likes, 2)2 8.238e-01  
## movie.sig$director\_facebook\_likes 4.409e-06  
## movie.sig$cast\_total\_facebook\_likes 7.013e-07  
## movie.sig$budget 7.213e-10  
## movie.sig$title\_year 1.809e-03  
## movie.sig$genresAdventure 5.268e-02  
## movie.sig$genresAnimation 1.298e-01  
## movie.sig$genresBiography 7.351e-02  
## movie.sig$genresComedy 4.205e-02  
## movie.sig$genresCrime 6.236e-02  
## movie.sig$genresDocumentary 1.542e-01  
## movie.sig$genresDrama 4.726e-02  
## movie.sig$genresFamily 4.362e-01  
## movie.sig$genresFantasy 1.387e-01  
## movie.sig$genresHorror 7.597e-02  
## movie.sig$genresMusical 5.328e-01  
## movie.sig$genresMystery 1.891e-01  
## movie.sig$genresRomance 5.254e-01  
## movie.sig$genresSci-Fi 2.827e-01  
## movie.sig$genresThriller 7.433e-01  
## movie.sig$genresWestern 5.272e-01  
## movie.sig$num\_critic\_for\_reviews NA  
## movie.sig$num\_user\_for\_reviews NA  
## movie.sig$num\_voted\_users NA  
## movie.sig$gross NA  
## movie.sig$facenumber\_in\_poster:movie.sig$num\_critic\_for\_reviews 4.260e-05  
## movie.sig$num\_user\_for\_reviews:movie.sig$num\_voted\_users 2.817e-10  
## movie.sig$num\_voted\_users:movie.sig$gross 1.105e-15  
## movie.sig$budget:movie.sig$gross 4.104e-18  
## t value  
## (Intercept) 16.446  
## poly(movie.sig$num\_voted\_users, 2)1 6.727  
## poly(movie.sig$num\_voted\_users, 2)2 -8.514  
## poly(movie.sig$num\_critic\_for\_reviews, 2)1 8.388  
## poly(movie.sig$num\_critic\_for\_reviews, 2)2 -9.452  
## poly(movie.sig$num\_user\_for\_reviews, 2)1 -7.568  
## poly(movie.sig$num\_user\_for\_reviews, 2)2 2.615  
## poly(movie.sig$duration, 2)1 11.755  
## poly(movie.sig$duration, 2)2 -4.973  
## movie.sig$facenumber\_in\_poster -1.892  
## poly(movie.sig$gross, 2)1 -6.012  
## poly(movie.sig$gross, 2)2 -3.565  
## poly(movie.sig$movie\_facebook\_likes, 2)1 1.952  
## poly(movie.sig$movie\_facebook\_likes, 2)2 0.277  
## movie.sig$director\_facebook\_likes 1.045  
## movie.sig$cast\_total\_facebook\_likes 0.361  
## movie.sig$budget -10.886  
## movie.sig$title\_year -14.680  
## movie.sig$genresAdventure 7.075  
## movie.sig$genresAnimation 5.828  
## movie.sig$genresBiography 8.522  
## movie.sig$genresComedy 3.747  
## movie.sig$genresCrime 7.309  
## movie.sig$genresDocumentary 6.317  
## movie.sig$genresDrama 11.067  
## movie.sig$genresFamily 1.366  
## movie.sig$genresFantasy -1.364  
## movie.sig$genresHorror -4.650  
## movie.sig$genresMusical -0.890  
## movie.sig$genresMystery 1.029  
## movie.sig$genresRomance 1.160  
## movie.sig$genresSci-Fi 0.520  
## movie.sig$genresThriller -0.415  
## movie.sig$genresWestern -0.080  
## movie.sig$num\_critic\_for\_reviews NA  
## movie.sig$num\_user\_for\_reviews NA  
## movie.sig$num\_voted\_users NA  
## movie.sig$gross NA  
## movie.sig$facenumber\_in\_poster:movie.sig$num\_critic\_for\_reviews -0.03\0  
\\## movie.sig$num\_user\_for\_reviews:movie.sig$num\_voted\_users 2.

828  
## movie.sig$num\_voted\_users:movie.sig$gross 1.355

## movie.sig$budget:movie.sig$gross 7.180  
## Pr(>|t|)  
## (Intercept) < 2e-16  
## poly(movie.sig$num\_voted\_users, 2)1 2.07e-11  
## poly(movie.sig$num\_voted\_users, 2)2 < 2e-16  
## poly(movie.sig$num\_critic\_for\_reviews, 2)1 < 2e-16  
## poly(movie.sig$num\_critic\_for\_reviews, 2)2 < 2e-16  
## poly(movie.sig$num\_user\_for\_reviews, 2)1 5.01e-14  
## poly(movie.sig$num\_user\_for\_reviews, 2)2 0.008973  
## poly(movie.sig$duration, 2)1 < 2e-16  
## poly(movie.sig$duration, 2)2 6.98e-07  
## movie.sig$facenumber\_in\_poster 0.058589  
## poly(movie.sig$gross, 2)1 2.05e-09  
## poly(movie.sig$gross, 2)2 0.000370  
## poly(movie.sig$movie\_facebook\_likes, 2)1 0.051079  
## poly(movie.sig$movie\_facebook\_likes, 2)2 0.781673  
## movie.sig$director\_facebook\_likes 0.296005  
## movie.sig$cast\_total\_facebook\_likes 0.717999  
## movie.sig$budget < 2e-16  
## movie.sig$title\_year < 2e-16  
## movie.sig$genresAdventure 1.86e-12  
## movie.sig$genresAnimation 6.23e-09  
## movie.sig$genresBiography < 2e-16  
## movie.sig$genresComedy 0.000183  
## movie.sig$genresCrime 3.45e-13  
## movie.sig$genresDocumentary 3.06e-10  
## movie.sig$genresDrama < 2e-16  
## movie.sig$genresFamily 0.172120  
## movie.sig$genresFantasy 0.172721  
## movie.sig$genresHorror 3.46e-06  
## movie.sig$genresMusical 0.373334  
## movie.sig$genresMystery 0.303477  
## movie.sig$genresRomance 0.246252  
## movie.sig$genresSci-Fi 0.602769  
## movie.sig$genresThriller 0.678129  
## movie.sig$genresWestern 0.936448  
## movie.sig$num\_critic\_for\_reviews NA  
## movie.sig$num\_user\_for\_reviews NA  
## movie.sig$num\_voted\_users NA  
## movie.sig$gross NA  
## movie.sig$facenumber\_in\_poster:movie.sig$num\_critic\_for\_reviews 0.975820  
## movie.sig$num\_user\_for\_reviews:movie.sig$num\_voted\_users 0.004714  
## movie.sig$num\_voted\_users:movie.sig$gross 0.175492  
## movie.sig$budget:movie.sig$gross 8.80e-13  
##   
## (Intercept) \*\*\*  
## poly(movie.sig$num\_voted\_users, 2)1 \*\*\*  
## poly(movie.sig$num\_voted\_users, 2)2 \*\*\*  
## poly(movie.sig$num\_critic\_for\_reviews, 2)1 \*\*\*  
## poly(movie.sig$num\_critic\_for\_reviews, 2)2 \*\*\*  
## poly(movie.sig$num\_user\_for\_reviews, 2)1 \*\*\*  
## poly(movie.sig$num\_user\_for\_reviews, 2)2 \*\*   
## poly(movie.sig$duration, 2)1 \*\*\*  
## poly(movie.sig$duration, 2)2 \*\*\*  
## movie.sig$facenumber\_in\_poster .   
## poly(movie.sig$gross, 2)1 \*\*\*  
## poly(movie.sig$gross, 2)2 \*\*\*  
## poly(movie.sig$movie\_facebook\_likes, 2)1 .   
## poly(movie.sig$movie\_facebook\_likes, 2)2   
## movie.sig$director\_facebook\_likes   
## movie.sig$cast\_total\_facebook\_likes   
## movie.sig$budget \*\*\*  
## movie.sig$title\_year \*\*\*  
## movie.sig$genresAdventure \*\*\*  
## movie.sig$genresAnimation \*\*\*  
## movie.sig$genresBiography \*\*\*  
## movie.sig$genresComedy \*\*\*  
## movie.sig$genresCrime \*\*\*  
## movie.sig$genresDocumentary \*\*\*  
## movie.sig$genresDrama \*\*\*  
## movie.sig$genresFamily   
## movie.sig$genresFantasy   
## movie.sig$genresHorror \*\*\*  
## movie.sig$genresMusical   
## movie.sig$genresMystery   
## movie.sig$genresRomance   
## movie.sig$genresSci-Fi   
## movie.sig$genresThriller   
## movie.sig$genresWestern   
## movie.sig$num\_critic\_for\_reviews   
## movie.sig$num\_user\_for\_reviews   
## movie.sig$num\_voted\_users   
## movie.sig$gross   
## movie.sig$facenumber\_in\_poster:movie.sig$num\_critic\_for\_reviews   
## movie.sig$num\_user\_for\_reviews:movie.sig$num\_voted\_users \*\*   
## movie.sig$num\_voted\_users:movie.sig$gross   
## movie.sig$budget:movie.sig$gross \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.741 on 2967 degrees of freedom  
## Multiple R-squared: 0.5107, Adjusted R-squared: 0.5046   
## F-statistic: 83.69 on 37 and 2967 DF, p-value: < 2.2e-16

step(null,scope=list(lower=null,upper=full2),direction='forward')

## Start: AIC=309.81  
## movie.sig$imdb\_score ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + poly(movie.sig$num\_voted\_users, 2) 2 976.96 2352.2 -730.05  
## + movie.sig$num\_voted\_users 1 871.90 2457.2 -600.74  
## + poly(movie.sig$duration, 2) 2 536.11 2793.0 -213.83  
## + poly(movie.sig$num\_user\_for\_reviews, 2) 2 483.99 2845.1 -158.27  
## + poly(movie.sig$num\_critic\_for\_reviews, 2) 2 436.49 2892.6 -108.52  
## + movie.sig$num\_critic\_for\_reviews 1 428.38 2900.8 -102.10  
## + movie.sig$num\_user\_for\_reviews 1 407.62 2921.5 -80.68  
## + poly(movie.sig$movie\_facebook\_likes, 2) 2 317.80 3011.3 12.32  
## + movie.sig$genres 16 331.02 2998.1 27.10  
## + poly(movie.sig$gross, 2) 2 251.27 3077.9 77.99  
## + movie.sig$gross 1 242.62 3086.5 84.42  
## + movie.sig$director\_facebook\_likes 1 166.17 3163.0 157.95  
## + movie.sig$title\_year 1 69.27 3259.9 248.63  
## + movie.sig$cast\_total\_facebook\_likes 1 64.28 3264.8 253.22  
## + movie.sig$budget 1 16.26 3312.9 297.09  
## + movie.sig$facenumber\_in\_poster 1 15.14 3314.0 298.11  
## <none> 3329.1 309.81  
##   
## Step: AIC=-730.05  
## movie.sig$imdb\_score ~ poly(movie.sig$num\_voted\_users, 2)  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$genres 16 337.58 2014.6 -1163.60  
## + poly(movie.sig$duration, 2) 2 137.87 2214.3 -907.55  
## + movie.sig$budget 1 133.09 2219.1 -903.07  
## + movie.sig$title\_year 1 101.46 2250.7 -860.55  
## + poly(movie.sig$gross, 2) 2 58.78 2293.4 -802.09  
## + movie.sig$gross 1 54.53 2297.6 -798.53  
## + poly(movie.sig$num\_user\_for\_reviews, 2) 2 29.12 2323.1 -763.48  
## + movie.sig$num\_user\_for\_reviews 1 25.39 2326.8 -760.66  
## + movie.sig$director\_facebook\_likes 1 17.94 2334.2 -751.05  
## + movie.sig$facenumber\_in\_poster 1 6.62 2345.5 -736.52  
## + poly(movie.sig$num\_critic\_for\_reviews, 2) 2 5.36 2346.8 -732.90  
## <none> 2352.2 -730.05  
## + movie.sig$num\_critic\_for\_reviews 1 0.18 2352.0 -728.28  
## + movie.sig$cast\_total\_facebook\_likes 1 0.15 2352.0 -728.23  
## + poly(movie.sig$movie\_facebook\_likes, 2) 2 1.29 2350.9 -727.70  
##   
## Step: AIC=-1163.6  
## movie.sig$imdb\_score ~ poly(movie.sig$num\_voted\_users, 2) + movie.sig$genres  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$title\_year 1 97.775 1916.8 -1311.1  
## + movie.sig$budget 1 65.238 1949.3 -1260.5  
## + poly(movie.sig$duration, 2) 2 65.750 1948.8 -1259.3  
## + movie.sig$gross 1 19.722 1994.9 -1191.2  
## + poly(movie.sig$gross, 2) 2 20.698 1993.9 -1190.6  
## + poly(movie.sig$num\_user\_for\_reviews, 2) 2 20.024 1994.6 -1189.6  
## + movie.sig$num\_user\_for\_reviews 1 14.834 1999.8 -1183.8  
## + poly(movie.sig$num\_critic\_for\_reviews, 2) 2 9.375 2005.2 -1173.6  
## + movie.sig$director\_facebook\_likes 1 6.114 2008.5 -1170.7  
## + movie.sig$facenumber\_in\_poster 1 3.792 2010.8 -1167.3  
## <none> 2014.6 -1163.6  
## + movie.sig$cast\_total\_facebook\_likes 1 0.355 2014.2 -1162.1  
## + movie.sig$num\_critic\_for\_reviews 1 0.042 2014.5 -1161.7  
## + poly(movie.sig$movie\_facebook\_likes, 2) 2 0.813 2013.8 -1160.8  
##   
## Step: AIC=-1311.1  
## movie.sig$imdb\_score ~ poly(movie.sig$num\_voted\_users, 2) + movie.sig$genres +   
## movie.sig$title\_year  
##   
## Df Sum of Sq RSS AIC  
## + poly(movie.sig$num\_critic\_for\_reviews, 2) 2 73.976 1842.8 -1425.4  
## + poly(movie.sig$duration, 2) 2 49.885 1866.9 -1386.3  
## + movie.sig$num\_critic\_for\_reviews 1 43.723 1873.1 -1378.4  
## + movie.sig$budget 1 32.246 1884.6 -1360.1  
## + poly(movie.sig$num\_user\_for\_reviews, 2) 2 21.755 1895.0 -1341.4  
## + poly(movie.sig$gross, 2) 2 19.623 1897.2 -1338.0  
## + movie.sig$gross 1 17.879 1898.9 -1337.3  
## + poly(movie.sig$movie\_facebook\_likes, 2) 2 18.788 1898.0 -1336.7  
## + movie.sig$num\_user\_for\_reviews 1 14.396 1902.4 -1331.8  
## + movie.sig$director\_facebook\_likes 1 3.373 1913.4 -1314.4  
## <none> 1916.8 -1311.1  
## + movie.sig$facenumber\_in\_poster 1 1.216 1915.6 -1311.0  
## + movie.sig$cast\_total\_facebook\_likes 1 0.300 1916.5 -1309.6  
##   
## Step: AIC=-1425.37  
## movie.sig$imdb\_score ~ poly(movie.sig$num\_voted\_users, 2) + movie.sig$genres +   
## movie.sig$title\_year + poly(movie.sig$num\_critic\_for\_reviews,   
## 2)  
##   
## Df Sum of Sq RSS AIC  
## + poly(movie.sig$num\_user\_for\_reviews, 2) 2 54.189 1788.6 -1511.1  
## + movie.sig$budget 1 46.017 1796.8 -1499.4  
## + poly(movie.sig$duration, 2) 2 38.533 1804.3 -1484.9  
## + movie.sig$num\_user\_for\_reviews 1 33.751 1809.1 -1478.9  
## + poly(movie.sig$gross, 2) 2 20.602 1822.2 -1455.2  
## + movie.sig$gross 1 16.630 1826.2 -1450.6  
## + poly(movie.sig$movie\_facebook\_likes, 2) 2 8.227 1834.6 -1434.8  
## + movie.sig$director\_facebook\_likes 1 2.296 1840.5 -1427.1  
## <none> 1842.8 -1425.4  
## + movie.sig$facenumber\_in\_poster 1 0.831 1842.0 -1424.7  
## + movie.sig$cast\_total\_facebook\_likes 1 0.104 1842.7 -1423.5  
##   
## Step: AIC=-1511.06  
## movie.sig$imdb\_score ~ poly(movie.sig$num\_voted\_users, 2) + movie.sig$genres +   
## movie.sig$title\_year + poly(movie.sig$num\_critic\_for\_reviews,   
## 2) + poly(movie.sig$num\_user\_for\_reviews, 2)  
##   
## Df Sum of Sq RSS AIC  
## + poly(movie.sig$duration, 2) 2 51.219 1737.4 -1594.4  
## + movie.sig$budget 1 34.907 1753.7 -1568.3  
## + poly(movie.sig$gross, 2) 2 20.882 1767.8 -1542.3  
## + movie.sig$gross 1 16.727 1771.9 -1537.3  
## + poly(movie.sig$movie\_facebook\_likes, 2) 2 3.910 1784.7 -1513.6  
## + movie.sig$director\_facebook\_likes 1 2.540 1786.1 -1513.3  
## + movie.sig$facenumber\_in\_poster 1 1.970 1786.7 -1512.4  
## <none> 1788.6 -1511.1  
## + movie.sig$cast\_total\_facebook\_likes 1 0.022 1788.6 -1509.1  
##   
## Step: AIC=-1594.36  
## movie.sig$imdb\_score ~ poly(movie.sig$num\_voted\_users, 2) + movie.sig$genres +   
## movie.sig$title\_year + poly(movie.sig$num\_critic\_for\_reviews,   
## 2) + poly(movie.sig$num\_user\_for\_reviews, 2) + poly(movie.sig$duration,   
## 2)  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$budget 1 62.211 1675.2 -1701.9  
## + poly(movie.sig$gross, 2) 2 30.406 1707.0 -1643.4  
## + movie.sig$gross 1 23.936 1713.5 -1634.0  
## + movie.sig$facenumber\_in\_poster 1 4.139 1733.3 -1599.5  
## <none> 1737.4 -1594.4  
## + movie.sig$director\_facebook\_likes 1 0.946 1736.5 -1594.0  
## + poly(movie.sig$movie\_facebook\_likes, 2) 2 1.928 1735.5 -1593.7  
## + movie.sig$cast\_total\_facebook\_likes 1 0.064 1737.4 -1592.5  
##   
## Step: AIC=-1701.94  
## movie.sig$imdb\_score ~ poly(movie.sig$num\_voted\_users, 2) + movie.sig$genres +   
## movie.sig$title\_year + poly(movie.sig$num\_critic\_for\_reviews,   
## 2) + poly(movie.sig$num\_user\_for\_reviews, 2) + poly(movie.sig$duration,   
## 2) + movie.sig$budget  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$facenumber\_in\_poster 1 5.0599 1670.2 -1709.0  
## + poly(movie.sig$gross, 2) 2 4.5359 1670.7 -1706.1  
## + movie.sig$gross 1 1.8995 1673.3 -1703.3  
## <none> 1675.2 -1701.9  
## + movie.sig$director\_facebook\_likes 1 0.6239 1674.6 -1701.1  
## + movie.sig$cast\_total\_facebook\_likes 1 0.2471 1675.0 -1700.4  
## + poly(movie.sig$movie\_facebook\_likes, 2) 2 0.8695 1674.3 -1699.5  
##   
## Step: AIC=-1709.03  
## movie.sig$imdb\_score ~ poly(movie.sig$num\_voted\_users, 2) + movie.sig$genres +   
## movie.sig$title\_year + poly(movie.sig$num\_critic\_for\_reviews,   
## 2) + poly(movie.sig$num\_user\_for\_reviews, 2) + poly(movie.sig$duration,   
## 2) + movie.sig$budget + movie.sig$facenumber\_in\_poster  
##   
## Df Sum of Sq RSS AIC  
## + poly(movie.sig$gross, 2) 2 4.6247 1665.5 -1713.4  
## + movie.sig$gross 1 1.9720 1668.2 -1710.6  
## <none> 1670.2 -1709.0  
## + movie.sig$director\_facebook\_likes 1 0.4874 1669.7 -1707.9  
## + movie.sig$cast\_total\_facebook\_likes 1 0.4443 1669.7 -1707.8  
## + poly(movie.sig$movie\_facebook\_likes, 2) 2 0.8414 1669.3 -1706.5  
##   
## Step: AIC=-1713.36  
## movie.sig$imdb\_score ~ poly(movie.sig$num\_voted\_users, 2) + movie.sig$genres +   
## movie.sig$title\_year + poly(movie.sig$num\_critic\_for\_reviews,   
## 2) + poly(movie.sig$num\_user\_for\_reviews, 2) + poly(movie.sig$duration,   
## 2) + movie.sig$budget + movie.sig$facenumber\_in\_poster +   
## poly(movie.sig$gross, 2)  
##   
## Df Sum of Sq RSS AIC  
## <none> 1665.5 -1713.4  
## + movie.sig$director\_facebook\_likes 1 0.49076 1665.0 -1712.2  
## + movie.sig$cast\_total\_facebook\_likes 1 0.41310 1665.1 -1712.1  
## + poly(movie.sig$movie\_facebook\_likes, 2) 2 1.10614 1664.4 -1711.4

##   
## Call:  
## lm(formula = movie.sig$imdb\_score ~ poly(movie.sig$num\_voted\_users,   
## 2) + movie.sig$genres + movie.sig$title\_year + poly(movie.sig$num\_critic\_for\_reviews,   
## 2) + poly(movie.sig$num\_user\_for\_reviews, 2) + poly(movie.sig$duration,   
## 2) + movie.sig$budget + movie.sig$facenumber\_in\_poster +   
## poly(movie.sig$gross, 2))  
##   
## Coefficients:  
## (Intercept)   
## 5.851e+01   
## poly(movie.sig$num\_voted\_users, 2)1   
## 3.249e+01   
## poly(movie.sig$num\_voted\_users, 2)2   
## -1.320e+01   
## movie.sig$genresAdventure   
## 3.770e-01   
## movie.sig$genresAnimation   
## 7.306e-01   
## movie.sig$genresBiography   
## 6.559e-01   
## movie.sig$genresComedy   
## 1.875e-01   
## movie.sig$genresCrime   
## 4.845e-01   
## movie.sig$genresDocumentary   
## 1.037e+00   
## movie.sig$genresDrama   
## 5.524e-01   
## movie.sig$genresFamily   
## 2.093e-01   
## movie.sig$genresFantasy   
## -1.231e-01   
## movie.sig$genresHorror   
## -2.986e-01   
## movie.sig$genresMusical   
## -4.597e-01   
## movie.sig$genresMystery   
## 2.304e-01   
## movie.sig$genresRomance   
## 6.151e-01   
## movie.sig$genresSci-Fi   
## 1.706e-01   
## movie.sig$genresThriller   
## -2.631e-01   
## movie.sig$genresWestern   
## 5.056e-02   
## movie.sig$title\_year   
## -2.605e-02   
## poly(movie.sig$num\_critic\_for\_reviews, 2)1   
## 1.634e+01   
## poly(movie.sig$num\_critic\_for\_reviews, 2)2   
## -6.906e+00   
## poly(movie.sig$num\_user\_for\_reviews, 2)1   
## -1.209e+01   
## poly(movie.sig$num\_user\_for\_reviews, 2)2   
## 7.641e+00   
## poly(movie.sig$duration, 2)1   
## 1.072e+01   
## poly(movie.sig$duration, 2)2   
## -3.800e+00   
## movie.sig$budget   
## -4.048e-09   
## movie.sig$facenumber\_in\_poster   
## -2.026e-02   
## poly(movie.sig$gross, 2)1   
## -2.770e+00   
## poly(movie.sig$gross, 2)2   
## 1.851e+00

1. full3: additive model with interaction

full3=  
lm(movie.sig$imdb\_score ~movie.sig$num\_voted\_users+movie.sig$num\_critic\_for\_reviews+movie.sig$num\_user\_for\_reviews+movie.sig$duration+movie.sig$facenumber\_in\_poster+movie.sig$gross+movie.sig$movie\_facebook\_likes+movie.sig$director\_facebook\_likes+movie.sig$cast\_total\_facebook\_likes+movie.sig$budget+movie.sig$title\_year+factor(movie.sig$genres)+movie.sig$duration\*movie.sig$num\_voted\_users+movie.sig$num\_voted\_users\*movie.sig$num\_user\_for\_reviews+movie.sig$gross\*movie.sig$budget,data=movie.sig)  
summary(full3)

##   
## Call:  
## lm(formula = movie.sig$imdb\_score ~ movie.sig$num\_voted\_users +   
## movie.sig$num\_critic\_for\_reviews + movie.sig$num\_user\_for\_reviews +   
## movie.sig$duration + movie.sig$facenumber\_in\_poster + movie.sig$gross +   
## movie.sig$movie\_facebook\_likes + movie.sig$director\_facebook\_likes +   
## movie.sig$cast\_total\_facebook\_likes + movie.sig$budget +   
## movie.sig$title\_year + factor(movie.sig$genres) + movie.sig$duration \*   
## movie.sig$num\_voted\_users + movie.sig$num\_voted\_users \* movie.sig$num\_user\_for\_reviews +   
## movie.sig$gross \* movie.sig$budget, data = movie.sig)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.0519 -0.3700 0.0863 0.4828 2.0996   
##   
## Coefficients:  
## Estimate  
## (Intercept) 4.748e+01  
## movie.sig$num\_voted\_users 7.890e-06  
## movie.sig$num\_critic\_for\_reviews 2.427e-03  
## movie.sig$num\_user\_for\_reviews -3.039e-04  
## movie.sig$duration 1.277e-02  
## movie.sig$facenumber\_in\_poster -1.858e-02  
## movie.sig$gross -1.469e-09  
## movie.sig$movie\_facebook\_likes -2.370e-06  
## movie.sig$director\_facebook\_likes 3.969e-06  
## movie.sig$cast\_total\_facebook\_likes 7.641e-07  
## movie.sig$budget -5.900e-09  
## movie.sig$title\_year -2.154e-02  
## factor(movie.sig$genres)Adventure 3.308e-01  
## factor(movie.sig$genres)Animation 7.426e-01  
## factor(movie.sig$genres)Biography 6.551e-01  
## factor(movie.sig$genres)Comedy 1.515e-01  
## factor(movie.sig$genres)Crime 4.496e-01  
## factor(movie.sig$genres)Documentary 8.960e-01  
## factor(movie.sig$genres)Drama 4.965e-01  
## factor(movie.sig$genres)Family 3.329e-01  
## factor(movie.sig$genres)Fantasy -1.544e-01  
## factor(movie.sig$genres)Horror -3.577e-01  
## factor(movie.sig$genres)Musical -2.616e-01  
## factor(movie.sig$genres)Mystery 1.263e-01  
## factor(movie.sig$genres)Romance 5.476e-01  
## factor(movie.sig$genres)Sci-Fi 1.673e-01  
## factor(movie.sig$genres)Thriller -4.858e-01  
## factor(movie.sig$genres)Western -1.277e-01  
## movie.sig$num\_voted\_users:movie.sig$duration -3.052e-08  
## movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews -3.752e-10  
## movie.sig$gross:movie.sig$budget 1.411e-17  
## Std. Error  
## (Intercept) 3.592e+00  
## movie.sig$num\_voted\_users 4.790e-07  
## movie.sig$num\_critic\_for\_reviews 2.275e-04  
## movie.sig$num\_user\_for\_reviews 6.998e-05  
## movie.sig$duration 9.200e-04  
## movie.sig$facenumber\_in\_poster 6.806e-03  
## movie.sig$gross 4.191e-10  
## movie.sig$movie\_facebook\_likes 9.659e-07  
## movie.sig$director\_facebook\_likes 4.482e-06  
## movie.sig$cast\_total\_facebook\_likes 7.181e-07  
## movie.sig$budget 5.917e-10  
## movie.sig$title\_year 1.790e-03  
## factor(movie.sig$genres)Adventure 5.338e-02  
## factor(movie.sig$genres)Animation 1.319e-01  
## factor(movie.sig$genres)Biography 7.512e-02  
## factor(movie.sig$genres)Comedy 4.284e-02  
## factor(movie.sig$genres)Crime 6.353e-02  
## factor(movie.sig$genres)Documentary 1.579e-01  
## factor(movie.sig$genres)Drama 4.835e-02  
## factor(movie.sig$genres)Family 4.432e-01  
## factor(movie.sig$genres)Fantasy 1.419e-01  
## factor(movie.sig$genres)Horror 7.638e-02  
## factor(movie.sig$genres)Musical 5.459e-01  
## factor(movie.sig$genres)Mystery 1.939e-01  
## factor(movie.sig$genres)Romance 5.392e-01  
## factor(movie.sig$genres)Sci-Fi 2.900e-01  
## factor(movie.sig$genres)Thriller 7.627e-01  
## factor(movie.sig$genres)Western 5.408e-01  
## movie.sig$num\_voted\_users:movie.sig$duration 3.447e-09  
## movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews 9.851e-11  
## movie.sig$gross:movie.sig$budget 2.887e-18  
## t value Pr(>|t|)  
## (Intercept) 13.218 < 2e-16  
## movie.sig$num\_voted\_users 16.472 < 2e-16  
## movie.sig$num\_critic\_for\_reviews 10.669 < 2e-16  
## movie.sig$num\_user\_for\_reviews -4.343 1.46e-05  
## movie.sig$duration 13.882 < 2e-16  
## movie.sig$facenumber\_in\_poster -2.730 0.006371  
## movie.sig$gross -3.505 0.000463  
## movie.sig$movie\_facebook\_likes -2.454 0.014175  
## movie.sig$director\_facebook\_likes 0.885 0.376035  
## movie.sig$cast\_total\_facebook\_likes 1.064 0.287447  
## movie.sig$budget -9.971 < 2e-16  
## movie.sig$title\_year -12.032 < 2e-16  
## factor(movie.sig$genres)Adventure 6.196 6.60e-10  
## factor(movie.sig$genres)Animation 5.629 1.98e-08  
## factor(movie.sig$genres)Biography 8.720 < 2e-16  
## factor(movie.sig$genres)Comedy 3.537 0.000411  
## factor(movie.sig$genres)Crime 7.077 1.83e-12  
## factor(movie.sig$genres)Documentary 5.676 1.51e-08  
## factor(movie.sig$genres)Drama 10.269 < 2e-16  
## factor(movie.sig$genres)Family 0.751 0.452648  
## factor(movie.sig$genres)Fantasy -1.089 0.276414  
## factor(movie.sig$genres)Horror -4.683 2.95e-06  
## factor(movie.sig$genres)Musical -0.479 0.631791  
## factor(movie.sig$genres)Mystery 0.652 0.514773  
## factor(movie.sig$genres)Romance 1.016 0.309947  
## factor(movie.sig$genres)Sci-Fi 0.577 0.563982  
## factor(movie.sig$genres)Thriller -0.637 0.524230  
## factor(movie.sig$genres)Western -0.236 0.813336  
## movie.sig$num\_voted\_users:movie.sig$duration -8.852 < 2e-16  
## movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews -3.809 0.000143  
## movie.sig$gross:movie.sig$budget 4.886 1.08e-06  
##   
## (Intercept) \*\*\*  
## movie.sig$num\_voted\_users \*\*\*  
## movie.sig$num\_critic\_for\_reviews \*\*\*  
## movie.sig$num\_user\_for\_reviews \*\*\*  
## movie.sig$duration \*\*\*  
## movie.sig$facenumber\_in\_poster \*\*   
## movie.sig$gross \*\*\*  
## movie.sig$movie\_facebook\_likes \*   
## movie.sig$director\_facebook\_likes   
## movie.sig$cast\_total\_facebook\_likes   
## movie.sig$budget \*\*\*  
## movie.sig$title\_year \*\*\*  
## factor(movie.sig$genres)Adventure \*\*\*  
## factor(movie.sig$genres)Animation \*\*\*  
## factor(movie.sig$genres)Biography \*\*\*  
## factor(movie.sig$genres)Comedy \*\*\*  
## factor(movie.sig$genres)Crime \*\*\*  
## factor(movie.sig$genres)Documentary \*\*\*  
## factor(movie.sig$genres)Drama \*\*\*  
## factor(movie.sig$genres)Family   
## factor(movie.sig$genres)Fantasy   
## factor(movie.sig$genres)Horror \*\*\*  
## factor(movie.sig$genres)Musical   
## factor(movie.sig$genres)Mystery   
## factor(movie.sig$genres)Romance   
## factor(movie.sig$genres)Sci-Fi   
## factor(movie.sig$genres)Thriller   
## factor(movie.sig$genres)Western   
## movie.sig$num\_voted\_users:movie.sig$duration \*\*\*  
## movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews \*\*\*  
## movie.sig$gross:movie.sig$budget \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7607 on 2974 degrees of freedom  
## Multiple R-squared: 0.483, Adjusted R-squared: 0.4778   
## F-statistic: 92.63 on 30 and 2974 DF, p-value: < 2.2e-16

step(null,scope=list(lower=null,upper=full3),direction='forward')

## Start: AIC=309.81  
## movie.sig$imdb\_score ~ 1  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$num\_voted\_users 1 871.90 2457.2 -600.74  
## + movie.sig$duration 1 491.13 2838.0 -167.82  
## + movie.sig$num\_critic\_for\_reviews 1 428.38 2900.8 -102.10  
## + movie.sig$num\_user\_for\_reviews 1 407.62 2921.5 -80.68  
## + factor(movie.sig$genres) 16 331.02 2998.1 27.10  
## + movie.sig$movie\_facebook\_likes 1 282.82 3046.3 45.02  
## + movie.sig$gross 1 242.62 3086.5 84.42  
## + movie.sig$director\_facebook\_likes 1 166.17 3163.0 157.95  
## + movie.sig$title\_year 1 69.27 3259.9 248.63  
## + movie.sig$cast\_total\_facebook\_likes 1 64.28 3264.8 253.22  
## + movie.sig$budget 1 16.26 3312.9 297.09  
## + movie.sig$facenumber\_in\_poster 1 15.14 3314.0 298.11  
## <none> 3329.1 309.81  
##   
## Step: AIC=-600.74  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users  
##   
## Df Sum of Sq RSS AIC  
## + factor(movie.sig$genres) 16 311.531 2145.7 -976.12  
## + movie.sig$duration 1 147.786 2309.4 -785.13  
## + movie.sig$title\_year 1 84.649 2372.6 -704.08  
## + movie.sig$budget 1 73.211 2384.0 -689.63  
## + movie.sig$num\_user\_for\_reviews 1 21.297 2435.9 -624.90  
## + movie.sig$gross 1 16.929 2440.3 -619.51  
## + movie.sig$num\_critic\_for\_reviews 1 14.632 2442.6 -616.69  
## + movie.sig$director\_facebook\_likes 1 13.657 2443.6 -615.49  
## + movie.sig$facenumber\_in\_poster 1 6.789 2450.4 -607.05  
## + movie.sig$movie\_facebook\_likes 1 2.627 2454.6 -601.95  
## <none> 2457.2 -600.74  
## + movie.sig$cast\_total\_facebook\_likes 1 0.524 2456.7 -599.38  
##   
## Step: AIC=-976.12  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres)  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$title\_year 1 79.623 2066.1 -1087.75  
## + movie.sig$duration 1 74.584 2071.1 -1080.44  
## + movie.sig$budget 1 28.689 2117.0 -1014.57  
## + movie.sig$num\_critic\_for\_reviews 1 23.116 2122.6 -1006.67  
## + movie.sig$num\_user\_for\_reviews 1 12.251 2133.4 -991.33  
## + movie.sig$director\_facebook\_likes 1 3.707 2142.0 -979.32  
## + movie.sig$facenumber\_in\_poster 1 3.274 2142.4 -978.71  
## + movie.sig$movie\_facebook\_likes 1 1.686 2144.0 -976.49  
## <none> 2145.7 -976.12  
## + movie.sig$gross 1 1.391 2144.3 -976.07  
## + movie.sig$cast\_total\_facebook\_likes 1 0.362 2145.3 -974.63  
##   
## Step: AIC=-1087.75  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$num\_critic\_for\_reviews 1 125.091 1941.0 -1273.4  
## + movie.sig$duration 1 55.857 2010.2 -1168.1  
## + movie.sig$movie\_facebook\_likes 1 21.746 2044.3 -1117.5  
## + movie.sig$num\_user\_for\_reviews 1 11.741 2054.3 -1102.9  
## + movie.sig$budget 1 9.196 2056.9 -1099.2  
## + movie.sig$cast\_total\_facebook\_likes 1 2.923 2063.2 -1090.0  
## + movie.sig$director\_facebook\_likes 1 1.740 2064.3 -1088.3  
## <none> 2066.1 -1087.8  
## + movie.sig$facenumber\_in\_poster 1 1.084 2065.0 -1087.3  
## + movie.sig$gross 1 0.638 2065.4 -1086.7  
##   
## Step: AIC=-1273.43  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$budget 1 36.627 1904.4 -1328.7  
## + movie.sig$num\_user\_for\_reviews 1 35.326 1905.7 -1326.6  
## + movie.sig$duration 1 34.873 1906.1 -1325.9  
## + movie.sig$gross 1 7.359 1933.6 -1282.8  
## + movie.sig$movie\_facebook\_likes 1 1.397 1939.6 -1273.6  
## <none> 1941.0 -1273.4  
## + movie.sig$facenumber\_in\_poster 1 0.926 1940.1 -1272.9  
## + movie.sig$director\_facebook\_likes 1 0.644 1940.3 -1272.4  
## + movie.sig$cast\_total\_facebook\_likes 1 0.572 1940.4 -1272.3  
##   
## Step: AIC=-1328.68  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$duration 1 58.373 1846.0 -1420.2  
## + movie.sig$num\_user\_for\_reviews 1 27.052 1877.3 -1369.7  
## + movie.sig$movie\_facebook\_likes 1 2.576 1901.8 -1330.8  
## + movie.sig$cast\_total\_facebook\_likes 1 2.005 1902.3 -1329.8  
## <none> 1904.4 -1328.7  
## + movie.sig$facenumber\_in\_poster 1 1.071 1903.3 -1328.4  
## + movie.sig$director\_facebook\_likes 1 0.557 1903.8 -1327.6  
## + movie.sig$gross 1 0.074 1904.3 -1326.8  
##   
## Step: AIC=-1420.23  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$num\_voted\_users:movie.sig$duration 1 70.848 1775.1 -1535.8  
## + movie.sig$num\_user\_for\_reviews 1 33.825 1812.2 -1473.8  
## + movie.sig$movie\_facebook\_likes 1 4.702 1841.3 -1425.9  
## + movie.sig$facenumber\_in\_poster 1 2.488 1843.5 -1422.3  
## + movie.sig$cast\_total\_facebook\_likes 1 1.601 1844.4 -1420.8  
## <none> 1846.0 -1420.2  
## + movie.sig$gross 1 0.196 1845.8 -1418.5  
## + movie.sig$director\_facebook\_likes 1 0.043 1845.9 -1418.3  
##   
## Step: AIC=-1535.83  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration + movie.sig$num\_voted\_users:movie.sig$duration  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$num\_user\_for\_reviews 1 26.4426 1748.7 -1578.9  
## + movie.sig$facenumber\_in\_poster 1 2.9576 1772.2 -1538.8  
## + movie.sig$cast\_total\_facebook\_likes 1 1.1823 1774.0 -1535.8  
## <none> 1775.1 -1535.8  
## + movie.sig$movie\_facebook\_likes 1 0.9446 1774.2 -1535.4  
## + movie.sig$director\_facebook\_likes 1 0.3854 1774.8 -1534.5  
## + movie.sig$gross 1 0.0191 1775.1 -1533.9  
##   
## Step: AIC=-1578.93  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration + movie.sig$num\_user\_for\_reviews +   
## movie.sig$num\_voted\_users:movie.sig$duration  
##   
## Df Sum of Sq  
## + movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews 1 5.4845  
## + movie.sig$facenumber\_in\_poster 1 4.1664  
## + movie.sig$movie\_facebook\_likes 1 3.9301  
## <none>   
## + movie.sig$cast\_total\_facebook\_likes 1 0.7354  
## + movie.sig$director\_facebook\_likes 1 0.2660  
## + movie.sig$gross 1 0.0008  
## RSS AIC  
## + movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews 1743.2 -1586.4  
## + movie.sig$facenumber\_in\_poster 1744.5 -1584.1  
## + movie.sig$movie\_facebook\_likes 1744.8 -1583.7  
## <none> 1748.7 -1578.9  
## + movie.sig$cast\_total\_facebook\_likes 1748.0 -1578.2  
## + movie.sig$director\_facebook\_likes 1748.4 -1577.4  
## + movie.sig$gross 1748.7 -1576.9  
##   
## Step: AIC=-1586.37  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration + movie.sig$num\_user\_for\_reviews +   
## movie.sig$num\_voted\_users:movie.sig$duration + movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$facenumber\_in\_poster 1 4.0181 1739.2 -1591.3  
## + movie.sig$movie\_facebook\_likes 1 3.2754 1739.9 -1590.0  
## <none> 1743.2 -1586.4  
## + movie.sig$cast\_total\_facebook\_likes 1 0.6359 1742.6 -1585.5  
## + movie.sig$director\_facebook\_likes 1 0.3798 1742.8 -1585.0  
## + movie.sig$gross 1 0.0475 1743.2 -1584.5  
##   
## Step: AIC=-1591.31  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration + movie.sig$num\_user\_for\_reviews +   
## movie.sig$facenumber\_in\_poster + movie.sig$num\_voted\_users:movie.sig$duration +   
## movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews  
##   
## Df Sum of Sq RSS AIC  
## + movie.sig$movie\_facebook\_likes 1 3.11243 1736.1 -1594.7  
## <none> 1739.2 -1591.3  
## + movie.sig$cast\_total\_facebook\_likes 1 0.90996 1738.3 -1590.9  
## + movie.sig$director\_facebook\_likes 1 0.29041 1738.9 -1589.8  
## + movie.sig$gross 1 0.04757 1739.1 -1589.4  
##   
## Step: AIC=-1594.69  
## movie.sig$imdb\_score ~ movie.sig$num\_voted\_users + factor(movie.sig$genres) +   
## movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration + movie.sig$num\_user\_for\_reviews +   
## movie.sig$facenumber\_in\_poster + movie.sig$movie\_facebook\_likes +   
## movie.sig$num\_voted\_users:movie.sig$duration + movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews  
##   
## Df Sum of Sq RSS AIC  
## <none> 1736.1 -1594.7  
## + movie.sig$cast\_total\_facebook\_likes 1 0.97305 1735.1 -1594.4  
## + movie.sig$director\_facebook\_likes 1 0.27990 1735.8 -1593.2  
## + movie.sig$gross 1 0.03634 1736.0 -1592.8

##   
## Call:  
## lm(formula = movie.sig$imdb\_score ~ movie.sig$num\_voted\_users +   
## factor(movie.sig$genres) + movie.sig$title\_year + movie.sig$num\_critic\_for\_reviews +   
## movie.sig$budget + movie.sig$duration + movie.sig$num\_user\_for\_reviews +   
## movie.sig$facenumber\_in\_poster + movie.sig$movie\_facebook\_likes +   
## movie.sig$num\_voted\_users:movie.sig$duration + movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews)  
##   
## Coefficients:  
## (Intercept)   
## 4.817e+01   
## movie.sig$num\_voted\_users   
## 7.152e-06   
## factor(movie.sig$genres)Adventure   
## 3.300e-01   
## factor(movie.sig$genres)Animation   
## 7.097e-01   
## factor(movie.sig$genres)Biography   
## 6.794e-01   
## factor(movie.sig$genres)Comedy   
## 1.675e-01   
## factor(movie.sig$genres)Crime   
## 4.784e-01   
## factor(movie.sig$genres)Documentary   
## 9.449e-01   
## factor(movie.sig$genres)Drama   
## 5.252e-01   
## factor(movie.sig$genres)Family   
## 2.260e-01   
## factor(movie.sig$genres)Fantasy   
## -1.422e-01   
## factor(movie.sig$genres)Horror   
## -3.440e-01   
## factor(movie.sig$genres)Musical   
## -3.165e-01   
## factor(movie.sig$genres)Mystery   
## 1.499e-01   
## factor(movie.sig$genres)Romance   
## 5.682e-01   
## factor(movie.sig$genres)Sci-Fi   
## 1.953e-01   
## factor(movie.sig$genres)Thriller   
## -4.097e-01   
## factor(movie.sig$genres)Western   
## -4.521e-02   
## movie.sig$title\_year   
## -2.189e-02   
## movie.sig$num\_critic\_for\_reviews   
## 2.566e-03   
## movie.sig$budget   
## -4.370e-09   
## movie.sig$duration   
## 1.206e-02   
## movie.sig$num\_user\_for\_reviews   
## -3.210e-04   
## movie.sig$facenumber\_in\_poster   
## -1.750e-02   
## movie.sig$movie\_facebook\_likes   
## -2.239e-06   
## movie.sig$num\_voted\_users:movie.sig$duration   
## -2.661e-08   
## movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews   
## -2.729e-10

lm.fit10<-lm(imdb\_score ~ num\_voted\_users+factor(genres)+title\_year+num\_critic\_for\_reviews+budget+duration+num\_voted\_users:duration,data=movie.sig)  
summary(lm.fit10)

##   
## Call:  
## lm(formula = imdb\_score ~ num\_voted\_users + factor(genres) +   
## title\_year + num\_critic\_for\_reviews + budget + duration +   
## num\_voted\_users:duration, data = movie.sig)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -5.2315 -0.3600 0.0840 0.4917 2.1054   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.805e+01 3.537e+00 13.584 < 2e-16 \*\*\*  
## num\_voted\_users 6.966e-06 4.418e-07 15.767 < 2e-16 \*\*\*  
## factor(genres)Adventure 3.524e-01 5.358e-02 6.578 5.63e-11 \*\*\*  
## factor(genres)Animation 7.748e-01 1.329e-01 5.828 6.21e-09 \*\*\*  
## factor(genres)Biography 7.244e-01 7.554e-02 9.589 < 2e-16 \*\*\*  
## factor(genres)Comedy 1.654e-01 4.274e-02 3.869 0.000111 \*\*\*  
## factor(genres)Crime 5.014e-01 6.397e-02 7.838 6.33e-15 \*\*\*  
## factor(genres)Documentary 9.329e-01 1.594e-01 5.852 5.37e-09 \*\*\*  
## factor(genres)Drama 5.123e-01 4.864e-02 10.532 < 2e-16 \*\*\*  
## factor(genres)Family 2.522e-01 4.470e-01 0.564 0.572591   
## factor(genres)Fantasy -1.510e-01 1.437e-01 -1.051 0.293246   
## factor(genres)Horror -3.709e-01 7.646e-02 -4.851 1.29e-06 \*\*\*  
## factor(genres)Musical -3.720e-01 5.522e-01 -0.674 0.500586   
## factor(genres)Mystery 1.508e-01 1.961e-01 0.769 0.441944   
## factor(genres)Romance 5.959e-01 5.468e-01 1.090 0.275917   
## factor(genres)Sci-Fi 1.695e-01 2.936e-01 0.577 0.563791   
## factor(genres)Thriller -4.447e-01 7.732e-01 -0.575 0.565232   
## factor(genres)Western -2.710e-02 5.470e-01 -0.050 0.960492   
## title\_year -2.185e-02 1.761e-03 -12.410 < 2e-16 \*\*\*  
## num\_critic\_for\_reviews 2.235e-03 1.807e-04 12.366 < 2e-16 \*\*\*  
## budget -4.489e-09 4.427e-10 -10.141 < 2e-16 \*\*\*  
## duration 1.240e-02 8.855e-04 14.001 < 2e-16 \*\*\*  
## num\_voted\_users:duration -3.334e-08 3.056e-09 -10.909 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.7715 on 2982 degrees of freedom  
## Multiple R-squared: 0.4668, Adjusted R-squared: 0.4629   
## F-statistic: 118.7 on 22 and 2982 DF, p-value: < 2.2e-16

For convinience to interpret the result, I will start with Full3(additive mode with interactiin terms). After checking residual, then decide should we add higher order terms.

Split data into Test and Train:

indx = sample(1:nrow(movie.sig), as.integer(0.9\*nrow(movie.sig)))  
indx # ramdomize rows, save 90% of data into index

## [1] 16 1371 734 1481 1991 606 905 1208 2534 2540 2444 1951 1131  
## [14] 1381 1505 97 2635 2356 616 1761 337 1416 640 2618 10 2730  
## [27] 2784 1610 1795 1955 1207 2842 281 1583 2383 1697 1323 1880 2462  
## [40] 1892 2858 123 940 1580 797 1555 169 1210 2188 906 1372 712  
## [53] 2786 2368 2182 1379 2287 2569 1665 1031 657 2120 1302 2439 1041  
## [66] 1445 2883 2610 1924 520 2057 910 1201 2888 1980 1988 2140 1945  
## [79] 679 2241 161 2296 1614 1675 1613 2411 771 645 1563 85 820  
## [92] 2103 1509 2851 831 491 671 667 279 1014 1596 2301 1064 2207  
## [105] 332 953 397 1887 1597 1756 2558 2854 2694 2055 2523 33 1224  
## [118] 2397 1044 1256 564 2997 1855 1944 1931 420 2430 1799 2964 2941  
## [131] 1168 1038 1452 250 1060 599 1990 2457 2865 69 828 1312 1801  
## [144] 1879 1280 2248 1300 1315 2393 1543 1554 2846 1571 1172 892 1261  
## [157] 1883 1893 523 881 1969 2148 354 656 917 290 829 1284 1485  
## [170] 2620 1278 1851 2360 434 2548 2039 1194 2272 2904 1557 1961 2533  
## [183] 1729 2831 68 280 2765 669 130 2450 2179 2175 385 1490 1033  
## [196] 192 872 938 151 1329 1158 2331 562 2913 896 785 911 2852  
## [209] 1822 2358 1768 695 1318 238 1544 2256 2715 2691 2340 2758 2626  
## [222] 211 2164 2838 899 91 1248 1078 2461 1739 1625 1092 2190 538  
## [235] 105 458 197 2682 2901 446 2073 2642 1458 1018 1383 1006 2734  
## [248] 1262 2884 2100 2986 1874 40 2806 2592 1662 2887 347 1128 692  
## [261] 2517 333 2906 628 1741 170 1578 1218 2174 2963 120 795 2422  
## [274] 793 1110 1783 483 1657 2077 1807 2468 251 1258 2979 1073 1664  
## [287] 2086 2714 1671 2757 2597 1647 1386 629 2705 1856 2695 38 2647  
## [300] 2547 1217 898 1688 913 1547 2347 2001 2649 265 2409 566 2367  
## [313] 2508 76 649 2652 2079 2940 172 2108 1639 1255 126 216 1655  
## [326] 1025 2954 2737 2309 2545 610 970 1134 442 1330 11 2921 1407  
## [339] 1645 1975 760 2664 683 2588 2669 1607 1591 955 615 2240 993  
## [352] 508 289 1263 2424 1930 2137 1093 536 2139 1776 1496 1989 2595  
## [365] 1605 1200 1376 1328 2022 2320 407 2848 409 2455 2739 2202 1285  
## [378] 596 1611 314 1854 807 1334 2318 877 1886 1118 1165 2633 175  
## [391] 1532 2251 591 857 1633 1476 2764 928 2441 1701 253 2908 2850  
## [404] 2489 1057 1292 2890 2308 1754 57 433 2333 2974 472 2781 1793  
## [417] 20 2823 2085 2425 236 190 825 2306 1586 2227 2898 1722 1773  
## [430] 376 1673 891 2602 2662 1431 2776 1908 1502 1221 2604 2570 425  
## [443] 991 101 2601 1800 1787 635 2816 1873 2744 142 533 2895 2135  
## [456] 763 1850 2491 2795 1621 2609 1661 808 2040 2634 2224 2452 2791  
## [469] 2168 128 2246 2478 715 503 246 2185 1891 2464 62 2313 2741  
## [482] 2443 1842 1876 2034 2065 2000 1985 1234 2162 2155 894 950 1847  
## [495] 2473 726 2885 1236 897 437 2316 1658 196 2414 114 2549 2405  
## [508] 2805 1437 2221 2014 109 2021 2156 1816 1833 842 697 2574 2242  
## [521] 1395 943 1191 1693 843 2357 2735 276 1771 1870 2923 2625 1983  
## [534] 944 2150 2211 412 2275 2472 264 1451 2233 2957 484 2133 298  
## [547] 445 1726 1738 2058 145 1909 576 1550 729 2359 1477 1927 598  
## [560] 2341 792 2917 2994 2408 1462 1109 1450 1553 2161 2226 371 2243  
## [573] 2650 4 2546 2519 687 2440 1299 2157 1635 1443 1434 1852 245  
## [586] 2173 849 1144 705 714 302 1572 1949 1368 1877 149 50 515  
## [599] 1065 2914 1453 2797 2683 2219 475 98 2799 2130 2319 459 2366  
## [612] 1260 1629 1839 772 930 1098 968 531 1392 1398 2144 2766 1736  
## [625] 2878 1275 2989 2276 2260 2555 2098 202 605 1499 839 959 565  
## [638] 1257 189 1259 2944 1958 1581 2934 329 1047 1867 2896 107 544  
## [651] 757 1430 2953 215 2789 1036 1418 2196 745 1732 223 682 1579  
## [664] 82 769 1199 2036 384 2839 2438 2328 218 1155 2655 1947 159  
## [677] 1291 256 626 1125 1304 2431 88 1923 2873 1938 1705 1023 2083  
## [690] 287 1684 2134 1341 588 2238 924 2772 1750 1620 1281 556 847  
## [703] 2428 2717 2727 780 293 1813 2713 1525 2295 2684 1340 2515 885  
## [716] 1559 2600 477 2573 1506 2092 1686 2719 2069 962 1755 2576 660  
## [729] 972 2172 39 2967 787 2761 1063 2063 1068 83 427 2955 1987  
## [742] 2093 701 220 2641 2681 89 827 2566 152 1105 1995 2010 147  
## [755] 52 646 110 2234 631 1358 2503 1364 956 1901 2981 327 865  
## [768] 555 912 174 717 205 1998 2775 1935 961 733 545 951 1567  
## [781] 1501 932 2759 979 36 73 2167 794 1045 2327 349 627 1008  
## [794] 2706 1135 179 93 260 2252 2589 2973 2738 1770 504 2899 2426  
## [807] 1212 2751 967 2794 2205 1024 2526 1632 1231 27 234 1775 5  
## [820] 2194 2046 2288 2459 2314 1659 2920 1193 826 1757 135 1492 644  
## [833] 2101 749 878 720 1187 2191 44 2381 1646 2038 3001 835 751  
## [846] 2442 2856 2075 1124 1016 2418 622 800 855 2264 1956 1824 675  
## [859] 2770 2774 1223 1748 494 1860 1460 2429 678 1112 2567 2245 1948  
## [872] 3005 574 485 59 452 2492 316 2218 470 1289 2911 207 336  
## [885] 2269 987 869 597 2124 2229 2419 2787 2925 2235 206 2290 75  
## [898] 2812 1086 633 1150 2583 405 732 1471 1487 790 1088 2363 53  
## [911] 2572 2342 2375 1902 1997 274 2253 1356 252 1040 370 2346 1209  
## [924] 2959 1744 2902 2067 850 419 537 1637 1594 2565 509 853 1374  
## [937] 1802 2918 719 2292 941 1089 1032 81 1237 619 365 462 2273  
## [950] 1878 761 495 153 529 1071 2131 1837 2564 600 24 1920 2893  
## [963] 58 2469 2237 2283 2257 1560 18 2619 2274 2998 2436 1039 506  
## [976] 474 2722 2661 1640 28 278 2778 1138 2960 2158 1439 2206 886  
## [989] 1615 1733 1180 2249 471 727 367 2630 2119 307 1708 707 2813  
## [1002] 754 781 463 272 1624 415 980 1573 2490 401 294 1316 2187  
## [1015] 823 673 2966 1232 2169 1992 186 1 204 417 1570 2922 1163  
## [1028] 41 1986 903 2048 1964 2646 482 1809 2866 2894 150 2496 2826  
## [1041] 1825 396 2059 1650 1786 1423 2352 342 2110 1781 2843 2720 322  
## [1054] 428 1287 704 1104 2076 607 315 72 2020 923 2365 2106 862  
## [1067] 2690 2423 1797 2933 1143 1042 449 184 286 521 2391 2447 1526  
## [1080] 496 513 1698 444 2696 532 2853 456 1592 13 1814 1730 1725  
## [1093] 1002 2796 1359 1380 2138 155 2995 164 604 1332 1709 1903 2560  
## [1106] 300 461 1351 411 1922 2819 1727 2198 90 292 2165 1270 2855  
## [1119] 1403 1482 541 112 156 239 2869 1230 178 762 2402 1737 299  
## [1132] 1719 2394 841 2474 560 2840 2830 1062 1498 1778 746 181 2321  
## [1145] 942 1537 308 47 1844 435 1360 2886 219 1779 1694 688 1077  
## [1158] 1702 1582 2504 1598 1339 424 154 2177 1264 1132 29 764 8  
## [1171] 528 1140 432 374 530 2976 1551 709 1751 2551 1053 399 466  
## [1184] 2956 2128 784 585 476 1401 2704 1051 1010 2697 2033 2307 2870  
## [1197] 2281 2154 1052 716 1700 2836 776 2213 1385 2451 1656 1521 1767  
## [1210] 2259 1835 2371 1085 325 2015 2814 2005 2586 2113 817 652 2122  
## [1223] 2881 248 1456 1798 2919 1454 2266 1838 165 2824 1164 1679 1348  
## [1236] 2700 1440 1517 2756 684 815 1845 2239 360 2875 736 1475 113  
## [1249] 2127 388 497 2835 2577 2948 275 971 2291 1119 2828 655 338  
## [1262] 2821 2084 261 1906 799 2736 266 103 124 880 927 2672 1106  
## [1275] 479 2936 1303 1346 2126 1966 2562 2030 1190 438 1254 1746 1156  
## [1288] 3004 1826 989 391 1428 879 1473 685 1484 2762 1568 2676 2111  
## [1301] 382 2687 1382 108 1145 1011 387 2088 1996 2404 1848 2623 2666  
## [1314] 1069 430 1513 460 636 1859 637 2822 2553 163 1116 1562 1435  
## [1327] 874 2184 1363 2768 2332 373 2311 2262 858 584 1228 392 1714  
## [1340] 1448 1342 51 1777 803 1411 1491 902 1917 1600 1384 1333 2087  
## [1353] 310 228 2112 2511 630 1916 2410 1638 647 295 2334 2603 602  
## [1366] 1784 1667 1399 1139 2520 1561 2529 1971 2250 2708 12 1538 577  
## [1379] 1785 2598 2018 1083 2605 973 2484 1782 832 824 747 890 552  
## [1392] 2415 2070 2798 1871 614 2382 2121 2372 572 2608 999 1169 2688  
## [1405] 915 1812 168 2531 945 2215 255 918 2386 1626 339 889 2044  
## [1418] 119 1619 372 1241 1283 2743 1803 1962 2740 752 856 1059 364  
## [1431] 1114 106 836 2903 1240 2539 1055 2493 2972 621 1912 2023 2731  
## [1444] 2178 416 2012 2377 1097 759 441 517 2361 2689 1507 2721 1616  
## [1457] 623 519 1184 871 2143 2658 1202 1100 1206 2480 217 2384 1157  
## [1470] 1792 1466 2398 1141 1182 801 1827 25 1566 638 2028 590 2527  
## [1483] 1804 2114 2091 19 2312 1429 2837 1464 2500 1175 2535 620 1717  
## [1496] 1282 1022 2817 2116 353 2475 2524 954 1815 63 1178 122 2448  
## [1509] 1067 1888 268 1752 929 2117 2400 595 1627 2506 676 2874 1527  
## [1522] 2579 1247 1529 235 742 2149 2072 402 2258 1027 2009 241 1099  
## [1535] 2692 1226 1122 17 1310 2556 2946 775 812 1818 2068 996 1541  
## [1548] 609 546 2889 2993 1642 379 2325 798 66 2892 867 514 158  
## [1561] 2747 2265 1347 1427 22 226 2544 221 330 1519 1365 2977 2910  
## [1574] 618 478 1508 413 2035 2052 2349 1723 1311 344 2300 2659 1928  
## [1587] 1214 2718 1389 2755 2528 2197 1469 2716 875 79 2818 1307 45  
## [1600] 1968 690 870 426 351 782 1829 2412 567 3002 180 1636 80  
## [1613] 244 1142 1294 779 984 744 510 2943 340 1420 262 457 1147  
## [1626] 2685 1618 1718 1049 2952 1884 132 569 1602 919 229 1865 1159  
## [1639] 403 505 767 1050 1005 2355 1574 1849 334 1367 1993 1290 1161  
## [1652] 2793 2581 1066 6 111 2742 947 2485 1183 1699 1480 1823 2323  
## [1665] 26 1967 2109 952 1743 2042 2016 1820 2037 115 2125 2090 960  
## [1678] 2557 2289 1021 481 559 1162 1913 755 1362 1296 2627 480 32  
## [1691] 1523 2970 698 317 1465 2643 1609 319 1715 177 1712 2749 14  
## [1704] 2483 2868 2142 1846 2712 285 1853 893 2255 2699 2271 1478 2811  
## [1717] 1449 2170 1146 1497 2247 1457 1152 166 21 1177 854 258 1905  
## [1730] 198 1872 1769 406 786 1522 1528 706 1393 2479 724 2458 1663  
## [1743] 2673 1421 2582 2494 1153 2982 948 303 375 2159 1447 571 2988  
## [1756] 2779 689 700 926 2622 267 2905 2099 1079 1102 1012 1486 2470  
## [1769] 1868 2024 547 455 1677 301 651 575 802 2326 1279 2209 1911  
## [1782] 2002 1107 998 1576 1612 2512 1512 2984 2521 2329 2417 2017 699  
## [1795] 46 2935 1817 1863 1375 212 2006 1043 2180 1836 2054 1710 3000  
## [1808] 1074 2203 1160 1222 1774 1020 568 2390 2004 254 691 2152 2370  
## [1821] 2062 2507 1408 1366 2877 345 2612 2261 2374 2808 2230 2437 1641  
## [1834] 2047 1470 900 535 2388 2181 1425 2362 2653 2849 1780 1734 1229  
## [1847] 2983 778 1305 1048 1461 2665 1203 978 1390 1272 2621 2857 990  
## [1860] 1108 247 2501 1219 2559 450 199 1634 2433 356 2163 2019 282  
## [1873] 34 2790 2082 2631 818 304 1267 1488 2081 2027 2495 1706 1890  
## [1886] 2723 173 1628 1678 2606 65 1387 583 1037 708 1238 2166 2310  
## [1899] 2859 1674 1549 2645 2115 489 171 2614 1740 958 468 2780 1669  
## [1912] 2991 866 1361 404 2118 1649 1933 1349 1345 936 2097 60 2223  
## [1925] 1479 625 713 2336 1821 194 176 429 2990 499 548 2632 2160  
## [1938] 931 1707 1984 389 821 343 723 1495 1898 2760 658 3 2456  
## [1951] 1120 861 320 191 563 2104 2299 1436 1711 2200 1056 230 1297  
## [1964] 2710 2373 1001 838 1094 2971 1409 56 1075 518 167 1493 2949  
## [1977] 1373 348 1338 1716 1013 1321 1271 586 1388 1530 2043 1914 1545  
## [1990] 1314 2807 451 703 284 1394 1728 2420 1009 1273 756 2011 2876  
## [2003] 868 672 2845 2975 2678 1511 15 1896 214 1438 2013 2825 1546  
## [2016] 2930 1666 486 1939 1900 834 283 2670 852 2563 1331 2466 1205  
## [2029] 1400 2656 1806 2337 2725 2406 2385 1565 2502 2268 67 2486 1417  
## [2042] 2407 694 1154 2141 2674 1791 1179 1919 2815 1766 1520 1654 601  
## [2055] 1841 2844 2771 1026 99 270 1504 1943 2969 2907 328 2639 1585  
## [2068] 1630 933 937 2613 2753 753 9 2915 1185 1575 2578 2344 2924  
## [2081] 2465 2804 1745 816 1327 668 551 2624 2834 1881 1894 1322 963  
## [2094] 144 904 1869 1973 2961 1216 2931 2413 240 2189 140 554 813  
## [2107] 1676 2403 2416 2728 102 2937 100 378 2151 2611 30 1253 1772  
## [2120] 2860 860 117 1690 748 1414 1760 1514 2928 1034 309 136 87  
## [2133] 731 1090 465 1149 200 2354 377 139 603 558 796 2746 2376  
## [2146] 1858 2864 2965 1503 1081 581 783 2671 1054 809 467 2568 524  
## [2159] 557 2399 43 2267 845 1325 1137 291 2767 473 901 2593 1308  
## [2172] 2537 593 1019 440 561 2785 540 352 527 680 1584 2244 2396  
## [2185] 914 1148 840 2702 662 2636 1683 1535 725 94 469 2638 2176  
## [2198] 985 2350 2996 2145 995 2752 1556 306 1524 1834 1915 2607 681  
## [2211] 1463 1174 664 1404 1672 888 23 1660 1293 2217 1548 1651 553  
## [2224] 2820 277 611 1595 1882 710 312 464 2675 1355 516 1981 331  
## [2237] 2322 1213 1720 1344 42 2186 2777 974 1265 887 418 2317 2053  
## [2250] 2467 1337 2446 976 2183 2628 2580 1875 1861 2552 393 2060 617  
## [2263] 1788 1643 2711 2724 740 2882 2435 758 1406 2707 542 582 702  
## [2276] 659 188 613 64 711 1703 686 2051 1648 587 227 2096 386  
## [2289] 2945 2298 2369 2668 1269 2031 1843 1790 773 2029 1235 92 183  
## [2302] 2929 788 488 766 1335 361 2304 735 129 1533 394 1593 2763  
## [2315] 2516 2343 1885 2232 2153 1058 341 2543 964 86 2123 925 1130  
## [2328] 2594 1692 1121 2102 2032 1910 1426 1286 2732 2891 2514 2575 2912  
## [2341] 1864 1724 908 2792 907 1899 1441 2962 1295 1243 422 31 400  
## [2354] 436 837 1601 920 490 1204 1309 1405 84 448 1695 1941 2395  
## [2367] 612 1539 665 2293 2680 2132 1215 2201 1413 2554 1758 326 1831  
## [2380] 2733 1474 642 1326 2210 1668 957 677 1167 369 634 988 570  
## [2393] 2064 257 1242 2216 1298 1103 1000 2900 501 2927 2987 398 946  
## [2406] 2078 1245 1007 1577 2657 1970 666 848 2499 1239 2660 549 983  
## [2419] 768 2648 61 1805 368 2525 1432 414 1516 1952 2074 2095 355  
## [2432] 224 1369 1061 2107 2748 2550 2947 1979 1840 1188 1357 1455 992  
## [2445] 592 195 1599 1319 366 2726 1251 2867 1534 1276 1685 2208 1982  
## [2458] 2351 1433 543 1558 2497 148 2951 1531 721 1934 969 1004 1749  
## [2471] 2220 410 2968 738 1518 1317 2481 324 408 2214 1170 2449 242  
## [2484] 1589 231 2729 1343 2654 1354 2872 1978 1696 578 1166 909 1301  
## [2497] 741 201 1608 1940 805 313 2926 335 765 1977 1277 2909 965  
## [2510] 1765 37 431 2518 1759 579 395 2841 500 2980 2584 1494 921  
## [2523] 1076 209 362 311 2637 2315 55 522 1198 2335 297 1113 1976  
## [2536] 1932 2950 2477 447 2651 883 2542 1189 573 1015 810 1974 193  
## [2549] 1288 1082 160 1084 1353 1459 2434 95 650 358 2663 1391 2  
## [2562] 185 2498 1588 2509 2338 323 125 1687 2025 1830 133 2809 1921  
## [2575] 1080 1937 511 2348 1606 138 1536 1552 2421 1266 1794 2629 78  
## [2588] 1246 2596 2007 1244 2345 141 213 661 2863 2302 1604 1819 492  
## [2601] 1828 2195 2303 1422 1689 1617 232 2802 830 116 3003 1133 2463  
## [2614] 1127 1424 2679 1442 271 718 2783 2401 525 822 1483 1197 2199  
## [2627] 1467 1176 743 2380 1994 2324 423 2279 269 2389 643 2387 876  
## [2640] 1192 2080 833 674 1747 873 608 2094 1965 2754 203 381 2538  
## [2653] 2282 2045 383 2003 2146 2453 654 1196 1415 2066 1101 1897 1796  
## [2666] 162 49 922 811 730 1742 7 357 1704 1810 493 1603 2879  
## [2679] 1087 1721 118 789 580 2709 2773 550 2476 2041 844 1370 1762  
## [2692] 986 70 263 502 2089 774 2212 864 421 2769 1115 884 2254

movie\_train = movie.sig[indx,]  
movie\_test = movie.sig[-indx,]

# lm.fit 1: linear model with interaction term dropping insig predictors.

# insig terms: director facebooklike','movie fb like' and 'cast total fb likes' from summary(full3)

# Note: nothing to do with step function we choose for full3.

lm.fit1<-lm(movie\_train$imdb\_score~movie\_train$num\_voted\_users+movie\_train$num\_critic\_for\_reviews+movie\_train$num\_user\_for\_reviews+movie\_train$duration+movie\_train$facenumber\_in\_poster+movie\_train$gross+movie\_train$budget+movie\_train$title\_year+factor(movie\_train$genres)+movie\_train$duration\*movie\_train$num\_voted\_users+movie\_train$num\_voted\_users\*movie\_train$num\_user\_for\_reviews+movie\_train$gross\*movie\_train$budget)

summary(lm.fit1)

Call:

lm(formula = movie\_train$imdb\_score ~ movie\_train$num\_voted\_users +

movie\_train$num\_critic\_for\_reviews + movie\_train$num\_user\_for\_reviews +

movie\_train$duration + movie\_train$facenumber\_in\_poster +

movie\_train$gross + movie\_train$budget + movie\_train$title\_year +

factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users +

movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews +

movie\_train$gross \* movie\_train$budget)

Residuals:

Min 1Q Median 3Q Max

-4.1147 -0.3741 0.0888 0.4871 2.1317

Coefficients:

Estimate Std. Error

(Intercept) 4.560e+01 3.808e+00

movie\_train$num\_voted\_users 8.634e-06 5.295e-07

movie\_train$num\_critic\_for\_reviews 2.112e-03 2.089e-04

movie\_train$num\_user\_for\_reviews -3.048e-04 7.757e-05

movie\_train$duration 1.466e-02 1.067e-03

movie\_train$facenumber\_in\_poster -2.045e-02 7.348e-03

movie\_train$gross -1.647e-09 4.448e-10

movie\_train$budget -6.108e-09 6.324e-10

movie\_train$title\_year -2.068e-02 1.896e-03

factor(movie\_train$genres)Adventure 3.217e-01 5.659e-02

factor(movie\_train$genres)Animation 7.176e-01 1.420e-01

factor(movie\_train$genres)Biography 6.369e-01 7.902e-02

factor(movie\_train$genres)Comedy 1.569e-01 4.533e-02

factor(movie\_train$genres)Crime 4.599e-01 6.644e-02

factor(movie\_train$genres)Documentary 1.274e+00 1.678e-01

factor(movie\_train$genres)Drama 4.859e-01 5.117e-02

factor(movie\_train$genres)Family 3.362e-01 4.422e-01

factor(movie\_train$genres)Fantasy -1.805e-01 1.465e-01

factor(movie\_train$genres)Horror -3.590e-01 8.119e-02

factor(movie\_train$genres)Musical 3.105e-01 7.627e-01

factor(movie\_train$genres)Mystery 1.919e-01 2.064e-01

factor(movie\_train$genres)Romance 8.404e-01 7.608e-01

factor(movie\_train$genres)Sci-Fi 2.792e-01 3.821e-01

factor(movie\_train$genres)Western 9.478e-01 7.606e-01

movie\_train$num\_voted\_users:movie\_train$duration -3.535e-08 3.909e-09

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews -4.491e-10 1.258e-10

movie\_train$gross:movie\_train$budget 1.550e-17 3.080e-18

t value Pr(>|t|)

(Intercept) 11.977 < 2e-16 \*\*\*

movie\_train$num\_voted\_users 16.306 < 2e-16 \*\*\*

movie\_train$num\_critic\_for\_reviews 10.109 < 2e-16 \*\*\*

movie\_train$num\_user\_for\_reviews -3.929 8.73e-05 \*\*\*

movie\_train$duration 13.744 < 2e-16 \*\*\*

movie\_train$facenumber\_in\_poster -2.783 0.005428 \*\*

movie\_train$gross -3.704 0.000217 \*\*\*

movie\_train$budget -9.658 < 2e-16 \*\*\*

movie\_train$title\_year -10.903 < 2e-16 \*\*\*

factor(movie\_train$genres)Adventure 5.685 1.45e-08 \*\*\*

factor(movie\_train$genres)Animation 5.055 4.60e-07 \*\*\*

factor(movie\_train$genres)Biography 8.061 1.14e-15 \*\*\*

factor(movie\_train$genres)Comedy 3.461 0.000547 \*\*\*

factor(movie\_train$genres)Crime 6.923 5.54e-12 \*\*\*

factor(movie\_train$genres)Documentary 7.594 4.27e-14 \*\*\*

factor(movie\_train$genres)Drama 9.496 < 2e-16 \*\*\*

factor(movie\_train$genres)Family 0.760 0.447116

factor(movie\_train$genres)Fantasy -1.233 0.217782

factor(movie\_train$genres)Horror -4.422 1.02e-05 \*\*\*

factor(movie\_train$genres)Musical 0.407 0.683996

factor(movie\_train$genres)Mystery 0.930 0.352623

factor(movie\_train$genres)Romance 1.105 0.269455

factor(movie\_train$genres)Sci-Fi 0.731 0.465043

factor(movie\_train$genres)Western 1.246 0.212842

movie\_train$num\_voted\_users:movie\_train$duration -9.043 < 2e-16 \*\*\*

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews -3.572 0.000361 \*\*\*

movie\_train$gross:movie\_train$budget 5.031 5.20e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7591 on 2647 degrees of freedom

Multiple R-squared: 0.4834, Adjusted R-squared: 0.4783

F-statistic: 95.26 on 26 and 2647 DF, p-value: < 2.2e-16

Do Lack of fit test to see if removing the predictors improve model performance:

Do Lack of fit test to see if removing the predictors improve model performance:

```{r}

#lm.full: full linear model with interaction terms on train dataset.

lm.full<-lm(movie\_train$imdb\_score~movie\_train$num\_voted\_users+movie\_train$num\_critic\_for\_reviews+movie\_train$num\_user\_for\_reviews+movie\_train$duration+movie\_train$facenumber\_in\_poster+movie\_train$gross+movie\_train$movie\_facebook\_likes+movie\_train$director\_facebook\_likes+movie\_train$cast\_total\_facebook\_likes+movie\_train$budget+movie\_train$title\_year+factor(movie\_train$genres)+movie\_train$duration\*movie\_train$num\_voted\_users+movie\_train$num\_voted\_users\*movie\_train$num\_user\_for\_reviews+movie\_train$gross\*movie\_train$budget)

summary(lm.full)

Call:

lm(formula = movie\_train$imdb\_score ~ movie\_train$num\_voted\_users +

movie\_train$num\_critic\_for\_reviews + movie\_train$num\_user\_for\_reviews +

movie\_train$duration + movie\_train$facenumber\_in\_poster +

movie\_train$gross + movie\_train$movie\_facebook\_likes + movie\_train$director\_facebook\_likes +

movie\_train$cast\_total\_facebook\_likes + movie\_train$budget +

movie\_train$title\_year + factor(movie\_train$genres) + movie\_train$duration \*

movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \*

movie\_train$num\_user\_for\_reviews + movie\_train$gross \* movie\_train$budget)

Residuals:

Min 1Q Median 3Q Max

-4.1019 -0.3741 0.0885 0.4895 2.1395

Coefficients:

Estimate Std. Error

(Intercept) 4.559e+01 3.821e+00

movie\_train$num\_voted\_users 8.574e-06 5.307e-07

movie\_train$num\_critic\_for\_reviews 2.287e-03 2.484e-04

movie\_train$num\_user\_for\_reviews -3.272e-04 7.929e-05

movie\_train$duration 1.466e-02 1.070e-03

movie\_train$facenumber\_in\_poster -2.074e-02 7.387e-03

movie\_train$gross -1.632e-09 4.448e-10

movie\_train$movie\_facebook\_likes -1.518e-06 1.105e-06

movie\_train$director\_facebook\_likes 1.685e-06 4.839e-06

movie\_train$cast\_total\_facebook\_likes 8.757e-07 7.410e-07

movie\_train$budget -6.211e-09 6.349e-10

movie\_train$title\_year -2.068e-02 1.903e-03

factor(movie\_train$genres)Adventure 3.250e-01 5.665e-02

factor(movie\_train$genres)Animation 7.116e-01 1.420e-01

factor(movie\_train$genres)Biography 6.290e-01 7.941e-02

factor(movie\_train$genres)Comedy 1.552e-01 4.534e-02

factor(movie\_train$genres)Crime 4.541e-01 6.652e-02

factor(movie\_train$genres)Documentary 1.281e+00 1.678e-01

factor(movie\_train$genres)Drama 4.848e-01 5.119e-02

factor(movie\_train$genres)Family 3.485e-01 4.427e-01

factor(movie\_train$genres)Fantasy -1.782e-01 1.465e-01

factor(movie\_train$genres)Horror -3.663e-01 8.136e-02

factor(movie\_train$genres)Musical 3.270e-01 7.627e-01

factor(movie\_train$genres)Mystery 1.861e-01 2.067e-01

factor(movie\_train$genres)Romance 8.519e-01 7.608e-01

factor(movie\_train$genres)Sci-Fi 2.869e-01 3.821e-01

factor(movie\_train$genres)Western 9.181e-01 7.642e-01

movie\_train$num\_voted\_users:movie\_train$duration -3.497e-08 3.921e-09

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews -4.250e-10 1.264e-10

movie\_train$gross:movie\_train$budget 1.558e-17 3.088e-18

t value Pr(>|t|)

(Intercept) 11.931 < 2e-16 \*\*\*

movie\_train$num\_voted\_users 16.155 < 2e-16 \*\*\*

movie\_train$num\_critic\_for\_reviews 9.205 < 2e-16 \*\*\*

movie\_train$num\_user\_for\_reviews -4.127 3.79e-05 \*\*\*

movie\_train$duration 13.709 < 2e-16 \*\*\*

movie\_train$facenumber\_in\_poster -2.808 0.005023 \*\*

movie\_train$gross -3.669 0.000249 \*\*\*

movie\_train$movie\_facebook\_likes -1.374 0.169453

movie\_train$director\_facebook\_likes 0.348 0.727767

movie\_train$cast\_total\_facebook\_likes 1.182 0.237372

movie\_train$budget -9.783 < 2e-16 \*\*\*

movie\_train$title\_year -10.866 < 2e-16 \*\*\*

factor(movie\_train$genres)Adventure 5.736 1.08e-08 \*\*\*

factor(movie\_train$genres)Animation 5.011 5.76e-07 \*\*\*

factor(movie\_train$genres)Biography 7.922 3.42e-15 \*\*\*

factor(movie\_train$genres)Comedy 3.423 0.000630 \*\*\*

factor(movie\_train$genres)Crime 6.826 1.08e-11 \*\*\*

factor(movie\_train$genres)Documentary 7.635 3.14e-14 \*\*\*

factor(movie\_train$genres)Drama 9.471 < 2e-16 \*\*\*

factor(movie\_train$genres)Family 0.787 0.431219

factor(movie\_train$genres)Fantasy -1.217 0.223723

factor(movie\_train$genres)Horror -4.502 7.04e-06 \*\*\*

factor(movie\_train$genres)Musical 0.429 0.668111

factor(movie\_train$genres)Mystery 0.900 0.368101

factor(movie\_train$genres)Romance 1.120 0.262933

factor(movie\_train$genres)Sci-Fi 0.751 0.452860

factor(movie\_train$genres)Western 1.201 0.229704

movie\_train$num\_voted\_users:movie\_train$duration -8.918 < 2e-16 \*\*\*

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews -3.362 0.000785 \*\*\*

movie\_train$gross:movie\_train$budget 5.045 4.83e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.759 on 2644 degrees of freedom

Multiple R-squared: 0.484, Adjusted R-squared: 0.4784

F-statistic: 85.53 on 29 and 2644 DF, p-value: < 2.2e-16

anova(lm.full,lm.fit1) # H0: reduced model fits===lack of fit=0

Analysis of Variance Table

Model 1: movie\_train$imdb\_score ~ movie\_train$num\_voted\_users + movie\_train$num\_critic\_for\_reviews +

movie\_train$num\_user\_for\_reviews + movie\_train$duration +

movie\_train$facenumber\_in\_poster + movie\_train$gross + movie\_train$movie\_facebook\_likes +

movie\_train$director\_facebook\_likes + movie\_train$cast\_total\_facebook\_likes +

movie\_train$budget + movie\_train$title\_year + factor(movie\_train$genres) +

movie\_train$duration \* movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \*

movie\_train$num\_user\_for\_reviews + movie\_train$gross \* movie\_train$budget

Model 2: movie\_train$imdb\_score ~ movie\_train$num\_voted\_users + movie\_train$num\_critic\_for\_reviews +

movie\_train$num\_user\_for\_reviews + movie\_train$duration +

movie\_train$facenumber\_in\_poster + movie\_train$gross + movie\_train$budget +

movie\_train$title\_year + factor(movie\_train$genres) + movie\_train$duration \*

movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \*

movie\_train$num\_user\_for\_reviews + movie\_train$gross \* movie\_train$budget

Res.Df RSS Df Sum of Sq F Pr(>F)

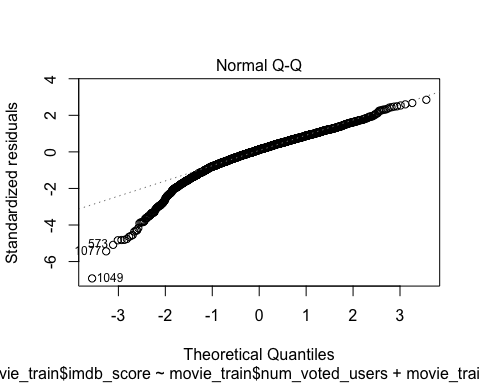
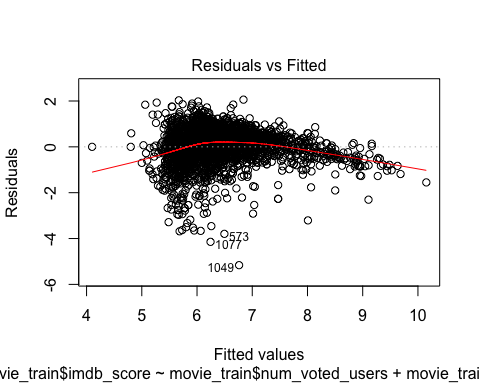
1 2644 1523.3

2 2647 1525.3 -3 -1.9874 1.1498 0.3276

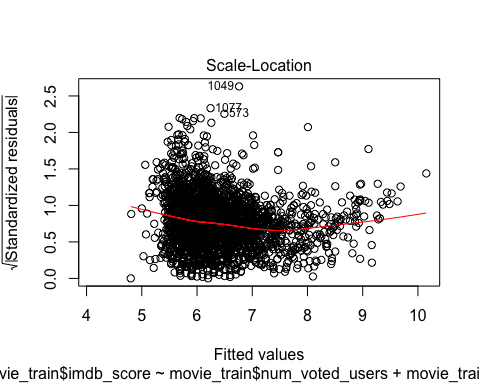
Diagnostics:

plot(lm.fit1)

## Warning: not plotting observations with leverage one:  
## 408,1125,1679

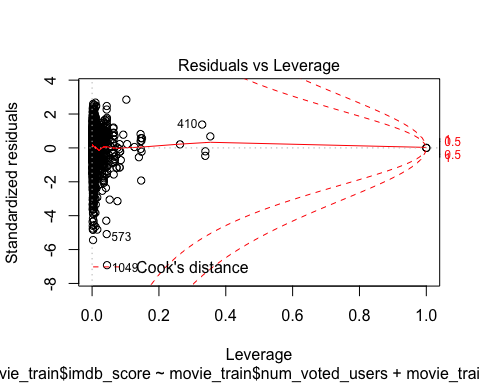


## Warning: not plotting observations with leverage one:  
## 122, 2563



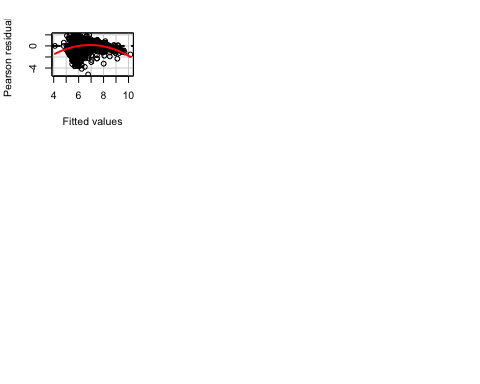
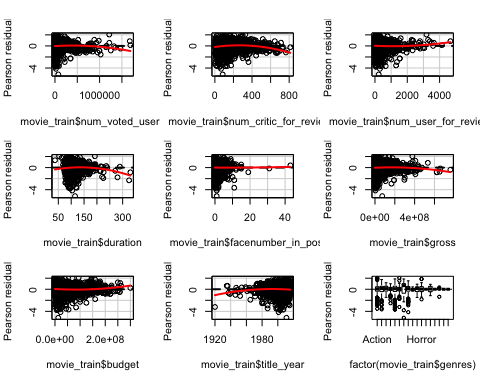
## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced



# residual vs fitted indicates might be higher order term. Normal plot not good.

library(car)  
residualPlots(lm.fit1)



## Test stat Pr(>|t|)

Test stat Pr(>|t|)

movie\_train$num\_voted\_users -8.262 0.000

movie\_train$num\_critic\_for\_reviews -7.891 0.000

movie\_train$num\_user\_for\_reviews 3.602 0.000

movie\_train$duration -2.977 0.003

movie\_train$facenumber\_in\_poster 0.478 0.633

movie\_train$gross -3.859 0.000

movie\_train$budget 4.742 0.000

movie\_train$title\_year -3.893 0.000

factor(movie\_train$genres) NA NA

Tukey test -14.024 0.000

Fit model with higer order terms:

# lm.fit2: model based on lm.fit1 adding higer order for all variables except for 'face number in poster' and 'title-year'.

lm.fit2<-lm(movie\_train$imdb\_score~poly(movie\_train$num\_voted\_users,2)+poly(movie\_train$num\_critic\_for\_reviews,2)+poly(movie\_train$num\_user\_for\_reviews,2)+poly(movie\_train$duration,2)+movie\_train$facenumber\_in\_poster+poly(movie\_train$gross,2)+poly(movie\_train$budget,2)+movie\_train$title\_year+factor(movie\_train$genres)+movie\_train$duration\*movie\_train$num\_voted\_users+movie\_train$num\_voted\_users\*movie\_train$num\_user\_for\_reviews+movie\_train$gross\*movie\_train$budget)

summary(lm.fit2)

Call:

lm(formula = movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users,

2) + poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

poly(movie\_train$gross, 2) + poly(movie\_train$budget, 2) +

movie\_train$title\_year + factor(movie\_train$genres) + movie\_train$duration \*

movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \*

movie\_train$num\_user\_for\_reviews + movie\_train$gross \* movie\_train$budget)

Residuals:

Min 1Q Median 3Q Max

-3.9101 -0.3534 0.0686 0.4626 2.1872

Coefficients: (5 not defined because of singularities)

Estimate Std. Error

(Intercept) 5.191e+01 3.815e+00

poly(movie\_train$num\_voted\_users, 2)1 4.092e+01 5.128e+00

poly(movie\_train$num\_voted\_users, 2)2 -1.645e+01 2.197e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)1 1.290e+01 1.326e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -7.365e+00 8.388e-01

poly(movie\_train$num\_user\_for\_reviews, 2)1 -1.836e+01 2.394e+00

poly(movie\_train$num\_user\_for\_reviews, 2)2 2.421e+00 1.578e+00

poly(movie\_train$duration, 2)1 1.391e+01 1.127e+00

poly(movie\_train$duration, 2)2 -2.466e+00 8.048e-01

movie\_train$facenumber\_in\_poster -2.298e-02 7.123e-03

poly(movie\_train$gross, 2)1 -6.234e+00 2.091e+00

poly(movie\_train$gross, 2)2 -1.703e+00 1.204e+00

poly(movie\_train$budget, 2)1 -1.346e+01 1.943e+00

poly(movie\_train$budget, 2)2 5.738e+00 1.084e+00

movie\_train$title\_year -2.278e-02 1.910e-03

factor(movie\_train$genres)Adventure 3.517e-01 5.560e-02

factor(movie\_train$genres)Animation 7.280e-01 1.386e-01

factor(movie\_train$genres)Biography 6.269e-01 7.659e-02

factor(movie\_train$genres)Comedy 1.373e-01 4.429e-02

factor(movie\_train$genres)Crime 4.575e-01 6.460e-02

factor(movie\_train$genres)Documentary 1.309e+00 1.634e-01

factor(movie\_train$genres)Drama 4.885e-01 4.980e-02

factor(movie\_train$genres)Family 3.536e-01 4.337e-01

factor(movie\_train$genres)Fantasy -2.195e-01 1.421e-01

factor(movie\_train$genres)Horror -3.916e-01 8.049e-02

factor(movie\_train$genres)Musical -2.307e-02 7.386e-01

factor(movie\_train$genres)Mystery 1.637e-01 2.000e-01

factor(movie\_train$genres)Romance 8.961e-01 7.362e-01

factor(movie\_train$genres)Sci-Fi 1.123e-01 3.701e-01

factor(movie\_train$genres)Western 9.097e-01 7.358e-01

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$num\_user\_for\_reviews NA NA

movie\_train$gross NA NA

movie\_train$budget NA NA

movie\_train$duration:movie\_train$num\_voted\_users -1.869e-08 4.142e-09

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 1.184e-09 3.635e-10

movie\_train$gross:movie\_train$budget 1.198e-17 5.662e-18

t value Pr(>|t|)

(Intercept) 13.606 < 2e-16 \*\*\*

poly(movie\_train$num\_voted\_users, 2)1 7.980 2.16e-15 \*\*\*

poly(movie\_train$num\_voted\_users, 2)2 -7.487 9.56e-14 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)1 9.728 < 2e-16 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -8.780 < 2e-16 \*\*\*

poly(movie\_train$num\_user\_for\_reviews, 2)1 -7.671 2.39e-14 \*\*\*

poly(movie\_train$num\_user\_for\_reviews, 2)2 1.534 0.12508

poly(movie\_train$duration, 2)1 12.342 < 2e-16 \*\*\*

poly(movie\_train$duration, 2)2 -3.065 0.00220 \*\*

movie\_train$facenumber\_in\_poster -3.226 0.00127 \*\*

poly(movie\_train$gross, 2)1 -2.982 0.00289 \*\*

poly(movie\_train$gross, 2)2 -1.415 0.15730

poly(movie\_train$budget, 2)1 -6.926 5.41e-12 \*\*\*

poly(movie\_train$budget, 2)2 5.292 1.31e-07 \*\*\*

movie\_train$title\_year -11.926 < 2e-16 \*\*\*

factor(movie\_train$genres)Adventure 6.326 2.96e-10 \*\*\*

factor(movie\_train$genres)Animation 5.253 1.61e-07 \*\*\*

factor(movie\_train$genres)Biography 8.186 4.18e-16 \*\*\*

factor(movie\_train$genres)Comedy 3.101 0.00195 \*\*

factor(movie\_train$genres)Crime 7.082 1.81e-12 \*\*\*

factor(movie\_train$genres)Documentary 8.011 1.70e-15 \*\*\*

factor(movie\_train$genres)Drama 9.808 < 2e-16 \*\*\*

factor(movie\_train$genres)Family 0.815 0.41505

factor(movie\_train$genres)Fantasy -1.544 0.12266

factor(movie\_train$genres)Horror -4.865 1.21e-06 \*\*\*

factor(movie\_train$genres)Musical -0.031 0.97509

factor(movie\_train$genres)Mystery 0.819 0.41308

factor(movie\_train$genres)Romance 1.217 0.22361

factor(movie\_train$genres)Sci-Fi 0.304 0.76151

factor(movie\_train$genres)Western 1.236 0.21643

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$num\_user\_for\_reviews NA NA

movie\_train$gross NA NA

movie\_train$budget NA NA

movie\_train$duration:movie\_train$num\_voted\_users -4.513 6.67e-06 \*\*\*

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 3.257 0.00114 \*\*

movie\_train$gross:movie\_train$budget 2.116 0.03441 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7341 on 2632 degrees of freedom

Multiple R-squared: 0.5175, Adjusted R-squared: 0.5116

F-statistic: 88.21 on 32 and 2632 DF, p-value: < 2.2e-16

# lm.fit3: based on lm.fit2 dropping second order term for 'gross' and budget\*gross

lm.fit3<-lm(movie\_train$imdb\_score~poly(movie\_train$num\_voted\_users,2)+poly(movie\_train$num\_critic\_for\_reviews,2)+poly(movie\_train$num\_user\_for\_reviews,2)+poly(movie\_train$duration,2)+movie\_train$facenumber\_in\_poster+movie\_train$gross+poly(movie\_train$budget,2)+movie\_train$title\_year+factor(movie\_train$genres)+movie\_train$duration\*movie\_train$num\_voted\_users+movie\_train$num\_voted\_users\*movie\_train$num\_user\_for\_reviews)  
summary(lm.fit3)

Call:

lm(formula = movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users,

2) + poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year +

factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users +

movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews)

Residuals:

Min 1Q Median 3Q Max

-3.9076 -0.3552 0.0688 0.4625 2.1776

Coefficients: (3 not defined because of singularities)

Estimate Std. Error

(Intercept) 5.111e+01 3.788e+00

poly(movie\_train$num\_voted\_users, 2)1 3.936e+01 5.016e+00

poly(movie\_train$num\_voted\_users, 2)2 -1.704e+01 2.181e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)1 1.277e+01 1.325e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -7.141e+00 8.217e-01

poly(movie\_train$num\_user\_for\_reviews, 2)1 -1.851e+01 2.393e+00

poly(movie\_train$num\_user\_for\_reviews, 2)2 2.427e+00 1.578e+00

poly(movie\_train$duration, 2)1 1.372e+01 1.122e+00

poly(movie\_train$duration, 2)2 -2.483e+00 8.049e-01

movie\_train$facenumber\_in\_poster -2.261e-02 7.124e-03

movie\_train$gross -7.074e-10 3.240e-10

poly(movie\_train$budget, 2)1 -1.022e+01 1.167e+00

poly(movie\_train$budget, 2)2 7.221e+00 8.067e-01

movie\_train$title\_year -2.234e-02 1.895e-03

factor(movie\_train$genres)Adventure 3.559e-01 5.554e-02

factor(movie\_train$genres)Animation 7.284e-01 1.386e-01

factor(movie\_train$genres)Biography 6.334e-01 7.655e-02

factor(movie\_train$genres)Comedy 1.408e-01 4.423e-02

factor(movie\_train$genres)Crime 4.674e-01 6.446e-02

factor(movie\_train$genres)Documentary 1.312e+00 1.634e-01

factor(movie\_train$genres)Drama 4.938e-01 4.976e-02

factor(movie\_train$genres)Family 2.018e-01 4.273e-01

factor(movie\_train$genres)Fantasy -2.120e-01 1.421e-01

factor(movie\_train$genres)Horror -3.848e-01 8.040e-02

factor(movie\_train$genres)Musical -1.028e-01 7.377e-01

factor(movie\_train$genres)Mystery 1.728e-01 2.000e-01

factor(movie\_train$genres)Romance 8.931e-01 7.365e-01

factor(movie\_train$genres)Sci-Fi 1.134e-01 3.701e-01

factor(movie\_train$genres)Western 9.152e-01 7.361e-01

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$num\_user\_for\_reviews NA NA

movie\_train$duration:movie\_train$num\_voted\_users -1.770e-08 4.096e-09

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 1.246e-09 3.616e-10

t value Pr(>|t|)

(Intercept) 13.492 < 2e-16 \*\*\*

poly(movie\_train$num\_voted\_users, 2)1 7.846 6.21e-15 \*\*\*

poly(movie\_train$num\_voted\_users, 2)2 -7.813 7.98e-15 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)1 9.639 < 2e-16 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -8.690 < 2e-16 \*\*\*

poly(movie\_train$num\_user\_for\_reviews, 2)1 -7.737 1.44e-14 \*\*\*

poly(movie\_train$num\_user\_for\_reviews, 2)2 1.537 0.124298

poly(movie\_train$duration, 2)1 12.229 < 2e-16 \*\*\*

poly(movie\_train$duration, 2)2 -3.085 0.002058 \*\*

movie\_train$facenumber\_in\_poster -3.174 0.001523 \*\*

movie\_train$gross -2.184 0.029077 \*

poly(movie\_train$budget, 2)1 -8.755 < 2e-16 \*\*\*

poly(movie\_train$budget, 2)2 8.950 < 2e-16 \*\*\*

movie\_train$title\_year -11.790 < 2e-16 \*\*\*

factor(movie\_train$genres)Adventure 6.408 1.74e-10 \*\*\*

factor(movie\_train$genres)Animation 5.257 1.58e-07 \*\*\*

factor(movie\_train$genres)Biography 8.273 < 2e-16 \*\*\*

factor(movie\_train$genres)Comedy 3.184 0.001471 \*\*

factor(movie\_train$genres)Crime 7.251 5.40e-13 \*\*\*

factor(movie\_train$genres)Documentary 8.028 1.48e-15 \*\*\*

factor(movie\_train$genres)Drama 9.925 < 2e-16 \*\*\*

factor(movie\_train$genres)Family 0.472 0.636816

factor(movie\_train$genres)Fantasy -1.492 0.135833

factor(movie\_train$genres)Horror -4.787 1.79e-06 \*\*\*

factor(movie\_train$genres)Musical -0.139 0.889184

factor(movie\_train$genres)Mystery 0.864 0.387742

factor(movie\_train$genres)Romance 1.213 0.225378

factor(movie\_train$genres)Sci-Fi 0.306 0.759288

factor(movie\_train$genres)Western 1.243 0.213871

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$num\_user\_for\_reviews NA NA

movie\_train$duration:movie\_train$num\_voted\_users -4.322 1.60e-05 \*\*\*

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 3.447 0.000576 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7344 on 2634 degrees of freedom

Multiple R-squared: 0.5167, Adjusted R-squared: 0.5112

F-statistic: 93.85 on 30 and 2634 DF, p-value: < 2.2e-16

anova(lm.fit2,lm.fit3)

Analysis of Variance Table

Model 1: movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) +

poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

poly(movie\_train$gross, 2) + poly(movie\_train$budget, 2) +

movie\_train$title\_year + factor(movie\_train$genres) + movie\_train$duration \*

movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \*

movie\_train$num\_user\_for\_reviews + movie\_train$gross \* movie\_train$budget

Model 2: movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) +

poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year +

factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users +

movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews

Res.Df RSS Df Sum of Sq F Pr(>F)

1 2632 1418.2

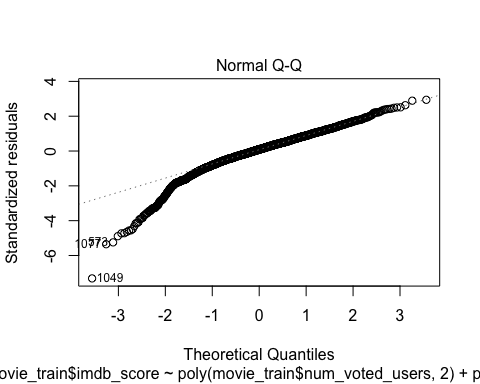
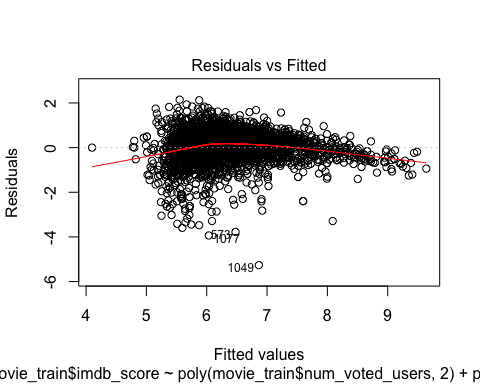
2 2634 1420.6 -2 -2.4178 2.2436 0.1063

P-value for lack of fit test is : 0.1063. Meaning lm.fit3 is better than lm.fit2. R^2 for lm.fit3: 0.5075, 50.75% of variation could be explained by this model.

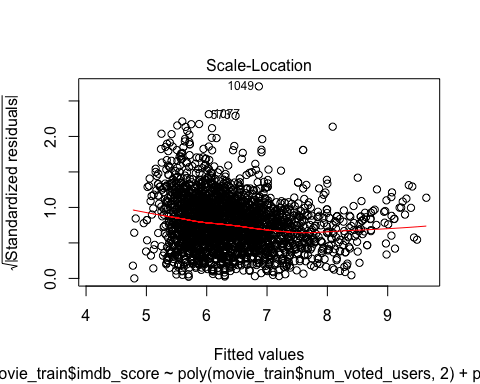
Diagnostics for lm.fit3:

plot(lm.fit3)

## Warning: not plotting observations with leverage one:  
## 408, 1679

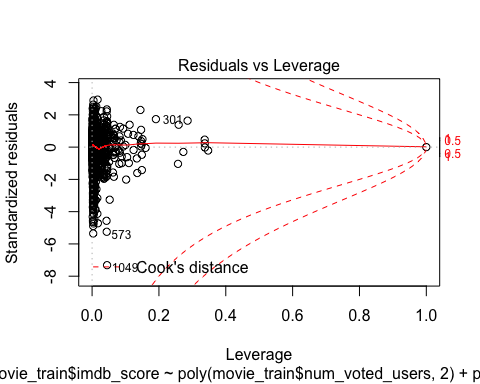


## Warning: not plotting observations with leverage one:  
## 122, 843, 2563

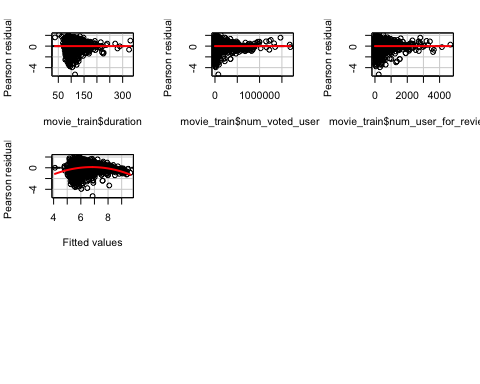
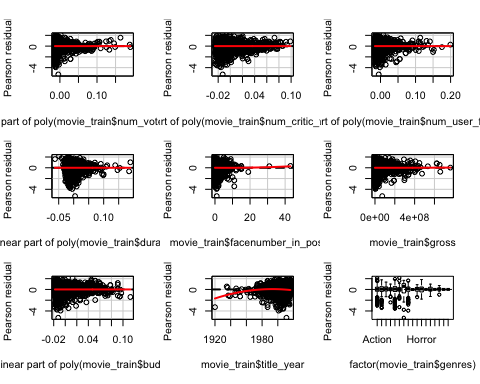


## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced



library(car)  
residualPlots(lm.fit3)



##

Test stat Pr(>|t|)

poly(movie\_train$num\_voted\_users, 2) NA NA

poly(movie\_train$num\_critic\_for\_reviews, 2) NA NA

poly(movie\_train$num\_user\_for\_reviews, 2) NA NA

poly(movie\_train$duration, 2) NA NA

movie\_train$facenumber\_in\_poster 0.564 0.573

movie\_train$gross 0.089 0.929

poly(movie\_train$budget, 2) NA NA

movie\_train$title\_year -6.247 0.000

factor(movie\_train$genres) NA NA

movie\_train$duration -0.432 0.665

movie\_train$num\_voted\_users -0.774 0.439

movie\_train$num\_user\_for\_reviews -0.581 0.562

Tukey test -12.602 0.000

The plot is way better than lm.fit2. All the residuals vs predictors are strainght lines except for title year. So, let't try to add second order for title year.

# lm.fit4: based on lm.fit3 addting second order for title year.  
lm.fit4<-lm(movie\_train$imdb\_score~poly(movie\_train$num\_voted\_users,2)+poly(movie\_train$num\_critic\_for\_reviews,2)+movie\_train$num\_user\_for\_reviews+poly(movie\_train$duration,2)+movie\_train$facenumber\_in\_poster+movie\_train$gross+poly(movie\_train$budget,2)+poly(movie\_train$title\_year,2)+factor(movie\_train$genres)+movie\_train$duration\*movie\_train$num\_voted\_users+movie\_train$num\_voted\_users\*movie\_train$num\_user\_for\_reviews)  
summary(lm.fit4)

Call:

lm(formula = movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users,

2) + poly(movie\_train$num\_critic\_for\_reviews, 2) + movie\_train$num\_user\_for\_reviews +

poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + poly(movie\_train$title\_year,

2) + factor(movie\_train$genres) + movie\_train$duration \*

movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \*

movie\_train$num\_user\_for\_reviews)

Residuals:

Min 1Q Median 3Q Max

-3.9264 -0.3543 0.0557 0.4566 2.1730

Coefficients: (2 not defined because of singularities)

Estimate Std. Error

(Intercept) 6.679e+00 6.701e-02

poly(movie\_train$num\_voted\_users, 2)1 3.349e+01 4.593e+00

poly(movie\_train$num\_voted\_users, 2)2 -1.878e+01 1.987e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)1 1.664e+01 1.475e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -7.765e+00 8.219e-01

movie\_train$num\_user\_for\_reviews -1.159e-03 1.044e-04

poly(movie\_train$duration, 2)1 1.358e+01 1.118e+00

poly(movie\_train$duration, 2)2 -2.192e+00 8.000e-01

movie\_train$facenumber\_in\_poster -1.776e-02 7.107e-03

movie\_train$gross -5.902e-10 3.204e-10

poly(movie\_train$budget, 2)1 -1.054e+01 1.159e+00

poly(movie\_train$budget, 2)2 7.608e+00 8.042e-01

poly(movie\_train$title\_year, 2)1 -1.309e+01 1.003e+00

poly(movie\_train$title\_year, 2)2 -5.135e+00 8.539e-01

factor(movie\_train$genres)Adventure 3.712e-01 5.497e-02

factor(movie\_train$genres)Animation 7.905e-01 1.379e-01

factor(movie\_train$genres)Biography 6.365e-01 7.601e-02

factor(movie\_train$genres)Comedy 1.446e-01 4.388e-02

factor(movie\_train$genres)Crime 4.706e-01 6.397e-02

factor(movie\_train$genres)Documentary 1.375e+00 1.627e-01

factor(movie\_train$genres)Drama 5.046e-01 4.940e-02

factor(movie\_train$genres)Family 1.591e-01 4.247e-01

factor(movie\_train$genres)Fantasy -2.701e-01 1.413e-01

factor(movie\_train$genres)Horror -4.130e-01 7.972e-02

factor(movie\_train$genres)Musical -1.662e-01 7.332e-01

factor(movie\_train$genres)Mystery 1.747e-01 1.987e-01

factor(movie\_train$genres)Romance 9.884e-01 7.321e-01

factor(movie\_train$genres)Sci-Fi 6.816e-02 3.677e-01

factor(movie\_train$genres)Western 8.681e-01 7.316e-01

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$duration:movie\_train$num\_voted\_users -1.737e-08 4.063e-09

movie\_train$num\_user\_for\_reviews:movie\_train$num\_voted\_users 1.834e-09 2.449e-10

t value Pr(>|t|)

(Intercept) 99.660 < 2e-16 \*\*\*

poly(movie\_train$num\_voted\_users, 2)1 7.292 4.03e-13 \*\*\*

poly(movie\_train$num\_voted\_users, 2)2 -9.453 < 2e-16 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)1 11.285 < 2e-16 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -9.447 < 2e-16 \*\*\*

movie\_train$num\_user\_for\_reviews -11.093 < 2e-16 \*\*\*

poly(movie\_train$duration, 2)1 12.152 < 2e-16 \*\*\*

poly(movie\_train$duration, 2)2 -2.740 0.006179 \*\*

movie\_train$facenumber\_in\_poster -2.499 0.012501 \*

movie\_train$gross -1.842 0.065598 .

poly(movie\_train$budget, 2)1 -9.098 < 2e-16 \*\*\*

poly(movie\_train$budget, 2)2 9.461 < 2e-16 \*\*\*

poly(movie\_train$title\_year, 2)1 -13.056 < 2e-16 \*\*\*

poly(movie\_train$title\_year, 2)2 -6.013 2.07e-09 \*\*\*

factor(movie\_train$genres)Adventure 6.753 1.77e-11 \*\*\*

factor(movie\_train$genres)Animation 5.731 1.11e-08 \*\*\*

factor(movie\_train$genres)Biography 8.375 < 2e-16 \*\*\*

factor(movie\_train$genres)Comedy 3.295 0.000998 \*\*\*

factor(movie\_train$genres)Crime 7.356 2.52e-13 \*\*\*

factor(movie\_train$genres)Documentary 8.452 < 2e-16 \*\*\*

factor(movie\_train$genres)Drama 10.215 < 2e-16 \*\*\*

factor(movie\_train$genres)Family 0.375 0.708030

factor(movie\_train$genres)Fantasy -1.911 0.056104 .

factor(movie\_train$genres)Horror -5.180 2.38e-07 \*\*\*

factor(movie\_train$genres)Musical -0.227 0.820641

factor(movie\_train$genres)Mystery 0.880 0.379159

factor(movie\_train$genres)Romance 1.350 0.177099

factor(movie\_train$genres)Sci-Fi 0.185 0.852950

factor(movie\_train$genres)Western 1.187 0.235469

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$duration:movie\_train$num\_voted\_users -4.275 1.98e-05 \*\*\*

movie\_train$num\_user\_for\_reviews:movie\_train$num\_voted\_users 7.490 9.36e-14 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7299 on 2643 degrees of freedom

Multiple R-squared: 0.5231, Adjusted R-squared: 0.5177

F-statistic: 96.65 on 30 and 2643 DF, p-value: < 2.2e-16

anova(lm.fit3,lm.fit4)

Analysis of Variance Table

Model 1: movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) +

poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year +

factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users +

movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews

Model 2: movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) +

poly(movie\_train$num\_critic\_for\_reviews, 2) + movie\_train$num\_user\_for\_reviews +

poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + poly(movie\_train$title\_year,

2) + factor(movie\_train$genres) + movie\_train$duration \*

movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \*

movie\_train$num\_user\_for\_reviews

Res.Df RSS Df Sum of Sq F Pr(>F)

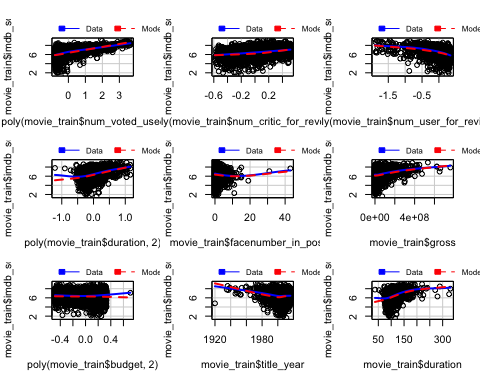
1 2643 1425.8

2 2643 1408.0 0 17.8

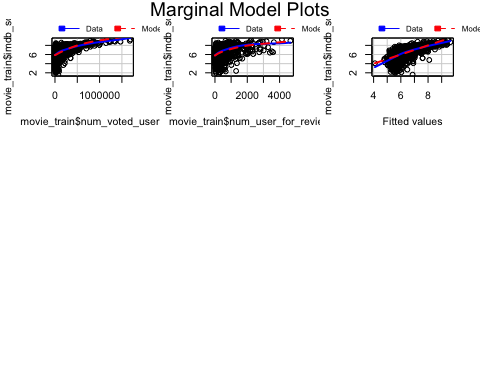
Marginal Model plot:

marginalModelPlots(lm.fit3)

## Warning in mmps(...): Splines and/or polynomials replaced by a fitted  
## linear combination

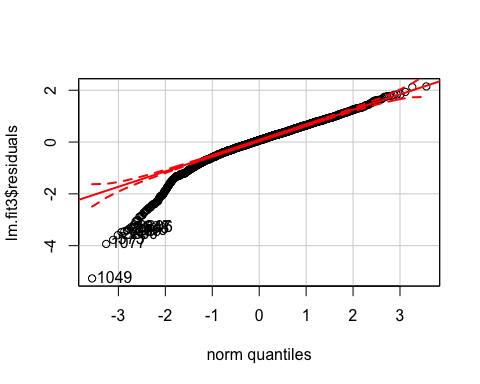


## Warning in mmps(...): Interactions and/or factors skipped

 The plots of the response versus the individual predictors display the conditional distribution of the response given each predictor, ignoring the other predictors. From our plots, our model is really good.since the marginal relationship between the response and the predictor are overlapping.

Check for residual ourliers:

library(car)  
qqPlot(lm.fit3$residuals,id.n = 10)



248 2269 1365 2056 2243 467 703 714 2629 1196

1 2 3 4 5 6 7 8 9 10

outlierTest(lm.fit3) # H0: residual is not an outlier

Show in New WindowClear OutputExpand/Collapse Output

[1] 10

[1] 19

Show in New WindowClear OutputExpand/Collapse Output

[1] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[22] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[43] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[64] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[85] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[106] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[127] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[148] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[169] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[190] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[211] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[232] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[253] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[274] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[295] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[316] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[337] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[358] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[379] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[400] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[421] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[442] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[463] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[484] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[505] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[526] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[547] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[568] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[589] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[610] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[631] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[652] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[673] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[694] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[715] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[736] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[757] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[778] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[799] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[820] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[841] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[862] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[883] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[904] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[925] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[946] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[967] USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA USA

[988] USA USA USA USA USA USA USA USA USA USA USA USA USA

[ reached getOption("max.print") -- omitted 2807 entries ]

66 Levels: Afghanistan Argentina Aruba Australia Bahamas Belgium Brazil ... West Germany

Show in New WindowClear OutputExpand/Collapse Output

Aboriginal Arabic Aramaic Bosnian Cantonese Chinese Czech

10 0 0 1 1 1 0 0

Danish Dari Dutch Dzongkha English Filipino French German

0 1 0 0 3779 1 0 0

Greek Hebrew Hindi Hungarian Icelandic Indonesian Italian Japanese

0 1 1 0 0 0 0 1

Kannada Kazakh Korean Mandarin Maya Mongolian None Norwegian

0 0 0 0 1 0 1 0

Panjabi Persian Polish Portuguese Romanian Russian Slovenian Spanish

0 0 0 0 0 0 0 7

Swahili Swedish Tamil Telugu Thai Urdu Vietnamese Zulu

0 0 0 0 0 0 1 0

Show in New WindowClear OutputExpand/Collapse Output

'data.frame': 3807 obs. of 26 variables:

$ color : Factor w/ 3 levels ""," Black and White",..: 3 3 3 3 3 3 3 3 3 3 ...

$ director\_name : Factor w/ 2399 levels "","\xcc\xe4mile Gaudreault",..: 926 799 379 106 2030 1652 1225 2394 284 799 ...

$ num\_critic\_for\_reviews : int 723 302 813 462 392 324 635 673 434 313 ...

$ duration : int 178 169 164 132 156 100 141 183 169 151 ...

$ director\_facebook\_likes : int 0 563 22000 475 0 15 0 0 0 563 ...

$ actor\_3\_facebook\_likes : int 855 1000 23000 530 4000 284 19000 2000 903 1000 ...

$ actor\_2\_name : Factor w/ 3033 levels "","50 Cent","A. Michael Baldwin",..: 1408 2218 534 2549 1228 801 2440 1704 1911 2218 ...

$ actor\_1\_facebook\_likes : int 1000 40000 27000 640 24000 799 26000 15000 18000 40000 ...

$ gross : int 760505847 309404152 448130642 73058679 336530303 200807262 458991599 330249062 200069408 423032628 ...

$ genres : Factor w/ 21 levels "Action","Adventure",..: 1 1 1 1 1 2 1 1 1 1 ...

$ actor\_1\_name : Factor w/ 2098 levels "","\xcc\xd2lafur Darri \xcc\xd2lafsson",..: 303 982 1968 441 786 221 337 740 1104 982 ...

$ movie\_title : Factor w/ 4917 levels "[Rec] 2\xe5\xca",..: 397 2731 3707 1960 3289 3459 398 460 3416 2732 ...

$ num\_voted\_users : int 886204 471220 1144337 212204 383056 294810 462669 371639 240396 522040 ...

$ cast\_total\_facebook\_likes: int 4834 48350 106759 1873 46055 2036 92000 24450 29991 48486 ...

$ actor\_3\_name : Factor w/ 3522 levels "","\xcc\xd2scar Jaenada",..: 3442 1393 1769 2714 1969 2162 3018 57 1134 1393 ...

$ facenumber\_in\_poster : int 0 0 0 1 0 1 4 0 0 2 ...

$ plot\_keywords : Factor w/ 4761 levels "","10 year old|dog|florida|girl|supermarket",..: 1320 4283 3484 651 4745 29 1142 1564 3312 2188 ...

$ num\_user\_for\_reviews : int 3054 1238 2701 738 1902 387 1117 3018 2367 1832 ...

$ country : Factor w/ 66 levels "","Afghanistan",..: 65 65 65 65 65 65 65 65 65 65 ...

$ content\_rating : Factor w/ 19 levels "","Approved",..: 10 10 10 10 10 9 10 10 10 10 ...

$ budget : num 2.37e+08 3.00e+08 2.50e+08 2.64e+08 2.58e+08 ...

$ title\_year : int 2009 2007 2012 2012 2007 2010 2015 2016 2006 2006 ...

$ actor\_2\_facebook\_likes : int 936 5000 23000 632 11000 553 21000 4000 10000 5000 ...

$ imdb\_score : num 7.9 7.1 8.5 6.6 6.2 7.8 7.5 6.9 6.1 7.3 ...

$ aspect\_ratio : num 1.78 2.35 2.35 2.35 2.35 1.85 2.35 2.35 2.35 2.35 ...

$ movie\_facebook\_likes : int 33000 0 164000 24000 0 29000 118000 197000 0 5000 ...

Show in New WindowClear OutputExpand/Collapse Output

color director\_name num\_critic\_for\_reviews

0 0 39

duration director\_facebook\_likes actor\_3\_facebook\_likes

6 74 13

actor\_2\_name actor\_1\_facebook\_likes gross

0 4 572

genres actor\_1\_name movie\_title

0 0 0

num\_voted\_users cast\_total\_facebook\_likes actor\_3\_name

0 0 0

facenumber\_in\_poster plot\_keywords num\_user\_for\_reviews

12 0 13

country content\_rating budget

0 0 298

title\_year actor\_2\_facebook\_likes imdb\_score

74 7 0

aspect\_ratio movie\_facebook\_likes

222 0

R Console

Show in New WindowClear OutputExpand/Collapse Output

color director\_name num\_critic\_for\_reviews

0 0 0

duration director\_facebook\_likes actor\_3\_facebook\_likes

0 0 0

actor\_2\_name actor\_1\_facebook\_likes gross

0 0 0

genres actor\_1\_name movie\_title

0 0 0

num\_voted\_users cast\_total\_facebook\_likes actor\_3\_name

0 0 0

facenumber\_in\_poster plot\_keywords num\_user\_for\_reviews

0 0 0

country content\_rating budget

0 0 0

title\_year actor\_2\_facebook\_likes imdb\_score

0 0 0

aspect\_ratio movie\_facebook\_likes

0 0

Show in New WindowClear OutputExpand/Collapse Output

[1] 1920 2016

[1] 145

[1] 121

Show in New WindowClear OutputExpand/Collapse Output

Show Traceback

Error in library(scatter) : there is no package called ‘scatter’

Show in New WindowClear OutputExpand/Collapse Output

[1] 9.3

R Console

Show in New WindowClear OutputExpand/Collapse Output

[1] 148

Show in New WindowClear OutputExpand/Collapse Output

Show in New WindowClear OutputExpand/Collapse Output

Show in New WindowClear OutputExpand/Collapse Output

Show in New WindowClear OutputExpand/Collapse Output

Call:corr.test(x = movie.num, y = NULL, use = "pairwise", method = "pearson",

adjust = "holm", alpha = 0.05)

Correlation matrix

num\_critic\_for\_reviews duration director\_facebook\_likes

num\_critic\_for\_reviews 1.00 0.26 0.19

duration 0.26 1.00 0.21

director\_facebook\_likes 0.19 0.21 1.00

actor\_3\_facebook\_likes 0.28 0.14 0.12

actor\_1\_facebook\_likes 0.17 0.09 0.09

gross 0.48 0.28 0.14

num\_voted\_users 0.60 0.37 0.32

cast\_total\_facebook\_likes 0.25 0.13 0.12

facenumber\_in\_poster -0.03 0.01 -0.05

num\_user\_for\_reviews 0.57 0.36 0.24

budget 0.49 0.30 0.09

title\_year 0.42 -0.11 -0.06

actor\_2\_facebook\_likes 0.28 0.15 0.12

imdb\_score 0.36 0.38 0.22

aspect\_ratio 0.18 0.16 0.05

movie\_facebook\_likes 0.71 0.25 0.17

actor\_3\_facebook\_likes actor\_1\_facebook\_likes gross

num\_critic\_for\_reviews 0.28 0.17 0.48

duration 0.14 0.09 0.28

director\_facebook\_likes 0.12 0.09 0.14

actor\_3\_facebook\_likes 1.00 0.25 0.30

actor\_1\_facebook\_likes 0.25 1.00 0.13

gross 0.30 0.13 1.00

num\_voted\_users 0.28 0.17 0.64

cast\_total\_facebook\_likes 0.48 0.95 0.22

facenumber\_in\_poster 0.10 0.05 -0.04

num\_user\_for\_reviews 0.22 0.12 0.55

budget 0.27 0.15 0.64

title\_year 0.13 0.09 0.06

actor\_2\_facebook\_likes 0.55 0.38 0.25

imdb\_score 0.09 0.12 0.27

aspect\_ratio 0.05 0.05 0.07

movie\_facebook\_likes 0.31 0.12 0.38

num\_voted\_users cast\_total\_facebook\_likes facenumber\_in\_poster

num\_critic\_for\_reviews 0.60 0.25 -0.03

duration 0.37 0.13 0.01

director\_facebook\_likes 0.32 0.12 -0.05

actor\_3\_facebook\_likes 0.28 0.48 0.10

actor\_1\_facebook\_likes 0.17 0.95 0.05

gross 0.64 0.22 -0.04

num\_voted\_users 1.00 0.25 -0.04

cast\_total\_facebook\_likes 0.25 1.00 0.07

facenumber\_in\_poster -0.04 0.07 1.00

num\_user\_for\_reviews 0.78 0.18 -0.09

budget 0.40 0.23 -0.03

title\_year 0.03 0.13 0.08

actor\_2\_facebook\_likes 0.25 0.63 0.07

imdb\_score 0.51 0.14 -0.07

aspect\_ratio 0.09 0.07 0.01

movie\_facebook\_likes 0.52 0.21 0.01

num\_user\_for\_reviews budget title\_year actor\_2\_facebook\_likes

num\_critic\_for\_reviews 0.57 0.49 0.42 0.28

duration 0.36 0.30 -0.11 0.15

director\_facebook\_likes 0.24 0.09 -0.06 0.12

actor\_3\_facebook\_likes 0.22 0.27 0.13 0.55

actor\_1\_facebook\_likes 0.12 0.15 0.09 0.38

gross 0.55 0.64 0.06 0.25

num\_voted\_users 0.78 0.40 0.03 0.25

cast\_total\_facebook\_likes 0.18 0.23 0.13 0.63

facenumber\_in\_poster -0.09 -0.03 0.08 0.07

num\_user\_for\_reviews 1.00 0.40 0.03 0.20

budget 0.40 1.00 0.25 0.25

title\_year 0.03 0.25 1.00 0.13

actor\_2\_facebook\_likes 0.20 0.25 0.13 1.00

imdb\_score 0.35 0.07 -0.14 0.13

aspect\_ratio 0.10 0.18 0.22 0.07

movie\_facebook\_likes 0.39 0.33 0.31 0.25

imdb\_score aspect\_ratio movie\_facebook\_likes

num\_critic\_for\_reviews 0.36 0.18 0.71

duration 0.38 0.16 0.25

director\_facebook\_likes 0.22 0.05 0.17

actor\_3\_facebook\_likes 0.09 0.05 0.31

actor\_1\_facebook\_likes 0.12 0.05 0.12

gross 0.27 0.07 0.38

num\_voted\_users 0.51 0.09 0.52

cast\_total\_facebook\_likes 0.14 0.07 0.21

facenumber\_in\_poster -0.07 0.01 0.01

num\_user\_for\_reviews 0.35 0.10 0.39

budget 0.07 0.18 0.33

title\_year -0.14 0.22 0.31

actor\_2\_facebook\_likes 0.13 0.07 0.25

imdb\_score 1.00 0.04 0.29

aspect\_ratio 0.04 1.00 0.11

movie\_facebook\_likes 0.29 0.11 1.00

Sample Size

[1] 3005

Probability values (Entries above the diagonal are adjusted for multiple tests.)

num\_critic\_for\_reviews duration director\_facebook\_likes

num\_critic\_for\_reviews 0.00 0.00 0.00

duration 0.00 0.00 0.00

director\_facebook\_likes 0.00 0.00 0.00

actor\_3\_facebook\_likes 0.00 0.00 0.00

actor\_1\_facebook\_likes 0.00 0.00 0.00

gross 0.00 0.00 0.00

num\_voted\_users 0.00 0.00 0.00

cast\_total\_facebook\_likes 0.00 0.00 0.00

facenumber\_in\_poster 0.09 0.66 0.00

num\_user\_for\_reviews 0.00 0.00 0.00

budget 0.00 0.00 0.00

title\_year 0.00 0.00 0.00

actor\_2\_facebook\_likes 0.00 0.00 0.00

imdb\_score 0.00 0.00 0.00

aspect\_ratio 0.00 0.00 0.01

movie\_facebook\_likes 0.00 0.00 0.00

actor\_3\_facebook\_likes actor\_1\_facebook\_likes gross

num\_critic\_for\_reviews 0.00 0.00 0.00

duration 0.00 0.00 0.00

director\_facebook\_likes 0.00 0.00 0.00

actor\_3\_facebook\_likes 0.00 0.00 0.00

actor\_1\_facebook\_likes 0.00 0.00 0.00

gross 0.00 0.00 0.00

num\_voted\_users 0.00 0.00 0.00

cast\_total\_facebook\_likes 0.00 0.00 0.00

facenumber\_in\_poster 0.00 0.01 0.05

num\_user\_for\_reviews 0.00 0.00 0.00

budget 0.00 0.00 0.00

title\_year 0.00 0.00 0.00

actor\_2\_facebook\_likes 0.00 0.00 0.00

imdb\_score 0.00 0.00 0.00

aspect\_ratio 0.01 0.00 0.00

movie\_facebook\_likes 0.00 0.00 0.00

num\_voted\_users cast\_total\_facebook\_likes facenumber\_in\_poster

num\_critic\_for\_reviews 0.00 0 0.65

duration 0.00 0 1.00

director\_facebook\_likes 0.00 0 0.06

actor\_3\_facebook\_likes 0.00 0 0.00

actor\_1\_facebook\_likes 0.00 0 0.07

gross 0.00 0 0.37

num\_voted\_users 0.00 0 0.17

cast\_total\_facebook\_likes 0.00 0 0.00

facenumber\_in\_poster 0.02 0 0.00

num\_user\_for\_reviews 0.00 0 0.00

budget 0.00 0 0.14

title\_year 0.10 0 0.00

actor\_2\_facebook\_likes 0.00 0 0.00

imdb\_score 0.00 0 0.00

aspect\_ratio 0.00 0 0.55

movie\_facebook\_likes 0.00 0 0.50

num\_user\_for\_reviews budget title\_year actor\_2\_facebook\_likes

num\_critic\_for\_reviews 0.00 0.00 0.00 0.00

duration 0.00 0.00 0.00 0.00

director\_facebook\_likes 0.00 0.00 0.04 0.00

actor\_3\_facebook\_likes 0.00 0.00 0.00 0.00

actor\_1\_facebook\_likes 0.00 0.00 0.00 0.00

gross 0.00 0.00 0.04 0.00

num\_voted\_users 0.00 0.00 0.65 0.00

cast\_total\_facebook\_likes 0.00 0.00 0.00 0.00

facenumber\_in\_poster 0.00 0.65 0.00 0.01

num\_user\_for\_reviews 0.00 0.00 0.65 0.00

budget 0.00 0.00 0.00 0.00

title\_year 0.12 0.00 0.00 0.00

actor\_2\_facebook\_likes 0.00 0.00 0.00 0.00

imdb\_score 0.00 0.00 0.00 0.00

aspect\_ratio 0.00 0.00 0.00 0.00

movie\_facebook\_likes 0.00 0.00 0.00 0.00

imdb\_score aspect\_ratio movie\_facebook\_likes

num\_critic\_for\_reviews 0.00 0.00 0

duration 0.00 0.00 0

director\_facebook\_likes 0.00 0.10 0

actor\_3\_facebook\_likes 0.00 0.07 0

actor\_1\_facebook\_likes 0.00 0.05 0

gross 0.00 0.00 0

num\_voted\_users 0.00 0.00 0

cast\_total\_facebook\_likes 0.00 0.00 0

facenumber\_in\_poster 0.00 1.00 1

num\_user\_for\_reviews 0.00 0.00 0

budget 0.00 0.00 0

title\_year 0.00 0.00 0

actor\_2\_facebook\_likes 0.00 0.00 0

imdb\_score 0.00 0.34 0

aspect\_ratio 0.04 0.00 0

movie\_facebook\_likes 0.00 0.00 0

To see confidence intervals of the correlations, print with the short=FALSE option

Modify Chunk OptionsRun All Chunks AboveRun Current Chunk

Show in New WindowClear OutputExpand/Collapse Output

Call:

lm(formula = imdb\_score ~ num\_voted\_users + factor(genres) +

title\_year + num\_critic\_for\_reviews + budget + duration +

num\_voted\_users:duration, data = movie.sig)

Residuals:

Min 1Q Median 3Q Max

-5.2315 -0.3600 0.0840 0.4917 2.1054

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 4.805e+01 3.537e+00 13.584 < 2e-16 \*\*\*

num\_voted\_users 6.966e-06 4.418e-07 15.767 < 2e-16 \*\*\*

factor(genres)Adventure 3.524e-01 5.358e-02 6.578 5.63e-11 \*\*\*

factor(genres)Animation 7.748e-01 1.329e-01 5.828 6.21e-09 \*\*\*

factor(genres)Biography 7.244e-01 7.554e-02 9.589 < 2e-16 \*\*\*

factor(genres)Comedy 1.654e-01 4.274e-02 3.869 0.000111 \*\*\*

factor(genres)Crime 5.014e-01 6.397e-02 7.838 6.33e-15 \*\*\*

factor(genres)Documentary 9.329e-01 1.594e-01 5.852 5.37e-09 \*\*\*

factor(genres)Drama 5.123e-01 4.864e-02 10.532 < 2e-16 \*\*\*

factor(genres)Family 2.522e-01 4.470e-01 0.564 0.572591

factor(genres)Fantasy -1.510e-01 1.437e-01 -1.051 0.293246

factor(genres)Horror -3.709e-01 7.646e-02 -4.851 1.29e-06 \*\*\*

factor(genres)Musical -3.720e-01 5.522e-01 -0.674 0.500586

factor(genres)Mystery 1.508e-01 1.961e-01 0.769 0.441944

factor(genres)Romance 5.959e-01 5.468e-01 1.090 0.275917

factor(genres)Sci-Fi 1.695e-01 2.936e-01 0.577 0.563791

factor(genres)Thriller -4.447e-01 7.732e-01 -0.575 0.565232

factor(genres)Western -2.710e-02 5.470e-01 -0.050 0.960492

title\_year -2.185e-02 1.761e-03 -12.410 < 2e-16 \*\*\*

num\_critic\_for\_reviews 2.235e-03 1.807e-04 12.366 < 2e-16 \*\*\*

budget -4.489e-09 4.427e-10 -10.141 < 2e-16 \*\*\*

duration 1.240e-02 8.855e-04 14.001 < 2e-16 \*\*\*

num\_voted\_users:duration -3.334e-08 3.056e-09 -10.909 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7715 on 2982 degrees of freedom

Multiple R-squared: 0.4668, Adjusted R-squared: 0.4629

F-statistic: 118.7 on 22 and 2982 DF, p-value: < 2.2e-16

Show in New WindowClear OutputExpand/Collapse Output

Call:

lm(formula = movie.sig$imdb\_score ~ movie.sig$num\_voted\_users +

movie.sig$num\_critic\_for\_reviews + movie.sig$num\_user\_for\_reviews +

movie.sig$duration + movie.sig$facenumber\_in\_poster + movie.sig$gross +

movie.sig$movie\_facebook\_likes + movie.sig$director\_facebook\_likes +

movie.sig$cast\_total\_facebook\_likes + movie.sig$budget +

movie.sig$title\_year + factor(movie.sig$genres) + movie.sig$duration \*

movie.sig$num\_voted\_users + movie.sig$num\_voted\_users \* movie.sig$num\_user\_for\_reviews +

movie.sig$gross \* movie.sig$budget, data = movie.sig)

Residuals:

Min 1Q Median 3Q Max

-5.0519 -0.3700 0.0863 0.4828 2.0996

Coefficients:

Estimate Std. Error t value

(Intercept) 4.748e+01 3.592e+00 13.218

movie.sig$num\_voted\_users 7.890e-06 4.790e-07 16.472

movie.sig$num\_critic\_for\_reviews 2.427e-03 2.275e-04 10.669

movie.sig$num\_user\_for\_reviews -3.039e-04 6.998e-05 -4.343

movie.sig$duration 1.277e-02 9.200e-04 13.882

movie.sig$facenumber\_in\_poster -1.858e-02 6.806e-03 -2.730

movie.sig$gross -1.469e-09 4.191e-10 -3.505

movie.sig$movie\_facebook\_likes -2.370e-06 9.659e-07 -2.454

movie.sig$director\_facebook\_likes 3.969e-06 4.482e-06 0.885

movie.sig$cast\_total\_facebook\_likes 7.641e-07 7.181e-07 1.064

movie.sig$budget -5.900e-09 5.917e-10 -9.971

movie.sig$title\_year -2.154e-02 1.790e-03 -12.032

factor(movie.sig$genres)Adventure 3.308e-01 5.338e-02 6.196

factor(movie.sig$genres)Animation 7.426e-01 1.319e-01 5.629

factor(movie.sig$genres)Biography 6.551e-01 7.512e-02 8.720

factor(movie.sig$genres)Comedy 1.515e-01 4.284e-02 3.537

factor(movie.sig$genres)Crime 4.496e-01 6.353e-02 7.077

factor(movie.sig$genres)Documentary 8.960e-01 1.579e-01 5.676

factor(movie.sig$genres)Drama 4.965e-01 4.835e-02 10.269

factor(movie.sig$genres)Family 3.329e-01 4.432e-01 0.751

factor(movie.sig$genres)Fantasy -1.544e-01 1.419e-01 -1.089

factor(movie.sig$genres)Horror -3.577e-01 7.638e-02 -4.683

factor(movie.sig$genres)Musical -2.616e-01 5.459e-01 -0.479

factor(movie.sig$genres)Mystery 1.263e-01 1.939e-01 0.652

factor(movie.sig$genres)Romance 5.476e-01 5.392e-01 1.016

factor(movie.sig$genres)Sci-Fi 1.673e-01 2.900e-01 0.577

factor(movie.sig$genres)Thriller -4.858e-01 7.627e-01 -0.637

factor(movie.sig$genres)Western -1.277e-01 5.408e-01 -0.236

movie.sig$num\_voted\_users:movie.sig$duration -3.052e-08 3.447e-09 -8.852

movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews -3.752e-10 9.851e-11 -3.809

movie.sig$gross:movie.sig$budget 1.411e-17 2.887e-18 4.886

Pr(>|t|)

(Intercept) < 2e-16 \*\*\*

movie.sig$num\_voted\_users < 2e-16 \*\*\*

movie.sig$num\_critic\_for\_reviews < 2e-16 \*\*\*

movie.sig$num\_user\_for\_reviews 1.46e-05 \*\*\*

movie.sig$duration < 2e-16 \*\*\*

movie.sig$facenumber\_in\_poster 0.006371 \*\*

movie.sig$gross 0.000463 \*\*\*

movie.sig$movie\_facebook\_likes 0.014175 \*

movie.sig$director\_facebook\_likes 0.376035

movie.sig$cast\_total\_facebook\_likes 0.287447

movie.sig$budget < 2e-16 \*\*\*

movie.sig$title\_year < 2e-16 \*\*\*

factor(movie.sig$genres)Adventure 6.60e-10 \*\*\*

factor(movie.sig$genres)Animation 1.98e-08 \*\*\*

factor(movie.sig$genres)Biography < 2e-16 \*\*\*

factor(movie.sig$genres)Comedy 0.000411 \*\*\*

factor(movie.sig$genres)Crime 1.83e-12 \*\*\*

factor(movie.sig$genres)Documentary 1.51e-08 \*\*\*

factor(movie.sig$genres)Drama < 2e-16 \*\*\*

factor(movie.sig$genres)Family 0.452648

factor(movie.sig$genres)Fantasy 0.276414

factor(movie.sig$genres)Horror 2.95e-06 \*\*\*

factor(movie.sig$genres)Musical 0.631791

factor(movie.sig$genres)Mystery 0.514773

factor(movie.sig$genres)Romance 0.309947

factor(movie.sig$genres)Sci-Fi 0.563982

factor(movie.sig$genres)Thriller 0.524230

factor(movie.sig$genres)Western 0.813336

movie.sig$num\_voted\_users:movie.sig$duration < 2e-16 \*\*\*

movie.sig$num\_voted\_users:movie.sig$num\_user\_for\_reviews 0.000143 \*\*\*

movie.sig$gross:movie.sig$budget 1.08e-06 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7607 on 2974 degrees of freedom

Multiple R-squared: 0.483, Adjusted R-squared: 0.4778

F-statistic: 92.63 on 30 and 2974 DF, p-value: < 2.2e-16

Modify Chunk OptionsRun All Chunks AboveRun Current Chunk

Show in New WindowClear OutputExpand/Collapse Output

Call:

lm(formula = movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users,

2) + poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

poly(movie\_train$gross, 2) + poly(movie\_train$budget, 2) +

movie\_train$title\_year + factor(movie\_train$genres) + movie\_train$duration \*

movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \*

movie\_train$num\_user\_for\_reviews + movie\_train$gross \* movie\_train$budget)

Residuals:

Min 1Q Median 3Q Max

-3.9101 -0.3534 0.0686 0.4626 2.1872

Coefficients: (5 not defined because of singularities)

Estimate Std. Error

(Intercept) 5.191e+01 3.815e+00

poly(movie\_train$num\_voted\_users, 2)1 4.092e+01 5.128e+00

poly(movie\_train$num\_voted\_users, 2)2 -1.645e+01 2.197e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)1 1.290e+01 1.326e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -7.365e+00 8.388e-01

poly(movie\_train$num\_user\_for\_reviews, 2)1 -1.836e+01 2.394e+00

poly(movie\_train$num\_user\_for\_reviews, 2)2 2.421e+00 1.578e+00

poly(movie\_train$duration, 2)1 1.391e+01 1.127e+00

poly(movie\_train$duration, 2)2 -2.466e+00 8.048e-01

movie\_train$facenumber\_in\_poster -2.298e-02 7.123e-03

poly(movie\_train$gross, 2)1 -6.234e+00 2.091e+00

poly(movie\_train$gross, 2)2 -1.703e+00 1.204e+00

poly(movie\_train$budget, 2)1 -1.346e+01 1.943e+00

poly(movie\_train$budget, 2)2 5.738e+00 1.084e+00

movie\_train$title\_year -2.278e-02 1.910e-03

factor(movie\_train$genres)Adventure 3.517e-01 5.560e-02

factor(movie\_train$genres)Animation 7.280e-01 1.386e-01

factor(movie\_train$genres)Biography 6.269e-01 7.659e-02

factor(movie\_train$genres)Comedy 1.373e-01 4.429e-02

factor(movie\_train$genres)Crime 4.575e-01 6.460e-02

factor(movie\_train$genres)Documentary 1.309e+00 1.634e-01

factor(movie\_train$genres)Drama 4.885e-01 4.980e-02

factor(movie\_train$genres)Family 3.536e-01 4.337e-01

factor(movie\_train$genres)Fantasy -2.195e-01 1.421e-01

factor(movie\_train$genres)Horror -3.916e-01 8.049e-02

factor(movie\_train$genres)Musical -2.307e-02 7.386e-01

factor(movie\_train$genres)Mystery 1.637e-01 2.000e-01

factor(movie\_train$genres)Romance 8.961e-01 7.362e-01

factor(movie\_train$genres)Sci-Fi 1.123e-01 3.701e-01

factor(movie\_train$genres)Western 9.097e-01 7.358e-01

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$num\_user\_for\_reviews NA NA

movie\_train$gross NA NA

movie\_train$budget NA NA

movie\_train$duration:movie\_train$num\_voted\_users -1.869e-08 4.142e-09

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 1.184e-09 3.635e-10

movie\_train$gross:movie\_train$budget 1.198e-17 5.662e-18

t value Pr(>|t|)

(Intercept) 13.606 < 2e-16 \*\*\*

poly(movie\_train$num\_voted\_users, 2)1 7.980 2.16e-15 \*\*\*

poly(movie\_train$num\_voted\_users, 2)2 -7.487 9.56e-14 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)1 9.728 < 2e-16 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -8.780 < 2e-16 \*\*\*

poly(movie\_train$num\_user\_for\_reviews, 2)1 -7.671 2.39e-14 \*\*\*

poly(movie\_train$num\_user\_for\_reviews, 2)2 1.534 0.12508

poly(movie\_train$duration, 2)1 12.342 < 2e-16 \*\*\*

poly(movie\_train$duration, 2)2 -3.065 0.00220 \*\*

movie\_train$facenumber\_in\_poster -3.226 0.00127 \*\*

poly(movie\_train$gross, 2)1 -2.982 0.00289 \*\*

poly(movie\_train$gross, 2)2 -1.415 0.15730

poly(movie\_train$budget, 2)1 -6.926 5.41e-12 \*\*\*

poly(movie\_train$budget, 2)2 5.292 1.31e-07 \*\*\*

movie\_train$title\_year -11.926 < 2e-16 \*\*\*

factor(movie\_train$genres)Adventure 6.326 2.96e-10 \*\*\*

factor(movie\_train$genres)Animation 5.253 1.61e-07 \*\*\*

factor(movie\_train$genres)Biography 8.186 4.18e-16 \*\*\*

factor(movie\_train$genres)Comedy 3.101 0.00195 \*\*

factor(movie\_train$genres)Crime 7.082 1.81e-12 \*\*\*

factor(movie\_train$genres)Documentary 8.011 1.70e-15 \*\*\*

factor(movie\_train$genres)Drama 9.808 < 2e-16 \*\*\*

factor(movie\_train$genres)Family 0.815 0.41505

factor(movie\_train$genres)Fantasy -1.544 0.12266

factor(movie\_train$genres)Horror -4.865 1.21e-06 \*\*\*

factor(movie\_train$genres)Musical -0.031 0.97509

factor(movie\_train$genres)Mystery 0.819 0.41308

factor(movie\_train$genres)Romance 1.217 0.22361

factor(movie\_train$genres)Sci-Fi 0.304 0.76151

factor(movie\_train$genres)Western 1.236 0.21643

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$num\_user\_for\_reviews NA NA

movie\_train$gross NA NA

movie\_train$budget NA NA

movie\_train$duration:movie\_train$num\_voted\_users -4.513 6.67e-06 \*\*\*

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 3.257 0.00114 \*\*

movie\_train$gross:movie\_train$budget 2.116 0.03441 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7341 on 2632 degrees of freedom

Multiple R-squared: 0.5175, Adjusted R-squared: 0.5116

F-statistic: 88.21 on 32 and 2632 DF, p-value: < 2.2e-16

Show in New WindowClear OutputExpand/Collapse Output

Call:

lm(formula = movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users,

2) + poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year +

factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users +

movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews)

Residuals:

Min 1Q Median 3Q Max

-3.9076 -0.3552 0.0688 0.4625 2.1776

Coefficients: (3 not defined because of singularities)

Estimate Std. Error

(Intercept) 5.111e+01 3.788e+00

poly(movie\_train$num\_voted\_users, 2)1 3.936e+01 5.016e+00

poly(movie\_train$num\_voted\_users, 2)2 -1.704e+01 2.181e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)1 1.277e+01 1.325e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -7.141e+00 8.217e-01

poly(movie\_train$num\_user\_for\_reviews, 2)1 -1.851e+01 2.393e+00

poly(movie\_train$num\_user\_for\_reviews, 2)2 2.427e+00 1.578e+00

poly(movie\_train$duration, 2)1 1.372e+01 1.122e+00

poly(movie\_train$duration, 2)2 -2.483e+00 8.049e-01

movie\_train$facenumber\_in\_poster -2.261e-02 7.124e-03

movie\_train$gross -7.074e-10 3.240e-10

poly(movie\_train$budget, 2)1 -1.022e+01 1.167e+00

poly(movie\_train$budget, 2)2 7.221e+00 8.067e-01

movie\_train$title\_year -2.234e-02 1.895e-03

factor(movie\_train$genres)Adventure 3.559e-01 5.554e-02

factor(movie\_train$genres)Animation 7.284e-01 1.386e-01

factor(movie\_train$genres)Biography 6.334e-01 7.655e-02

factor(movie\_train$genres)Comedy 1.408e-01 4.423e-02

factor(movie\_train$genres)Crime 4.674e-01 6.446e-02

factor(movie\_train$genres)Documentary 1.312e+00 1.634e-01

factor(movie\_train$genres)Drama 4.938e-01 4.976e-02

factor(movie\_train$genres)Family 2.018e-01 4.273e-01

factor(movie\_train$genres)Fantasy -2.120e-01 1.421e-01

factor(movie\_train$genres)Horror -3.848e-01 8.040e-02

factor(movie\_train$genres)Musical -1.028e-01 7.377e-01

factor(movie\_train$genres)Mystery 1.728e-01 2.000e-01

factor(movie\_train$genres)Romance 8.931e-01 7.365e-01

factor(movie\_train$genres)Sci-Fi 1.134e-01 3.701e-01

factor(movie\_train$genres)Western 9.152e-01 7.361e-01

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$num\_user\_for\_reviews NA NA

movie\_train$duration:movie\_train$num\_voted\_users -1.770e-08 4.096e-09

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 1.246e-09 3.616e-10

t value Pr(>|t|)

(Intercept) 13.492 < 2e-16 \*\*\*

poly(movie\_train$num\_voted\_users, 2)1 7.846 6.21e-15 \*\*\*

poly(movie\_train$num\_voted\_users, 2)2 -7.813 7.98e-15 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)1 9.639 < 2e-16 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -8.690 < 2e-16 \*\*\*

poly(movie\_train$num\_user\_for\_reviews, 2)1 -7.737 1.44e-14 \*\*\*

poly(movie\_train$num\_user\_for\_reviews, 2)2 1.537 0.124298

poly(movie\_train$duration, 2)1 12.229 < 2e-16 \*\*\*

poly(movie\_train$duration, 2)2 -3.085 0.002058 \*\*

movie\_train$facenumber\_in\_poster -3.174 0.001523 \*\*

movie\_train$gross -2.184 0.029077 \*

poly(movie\_train$budget, 2)1 -8.755 < 2e-16 \*\*\*

poly(movie\_train$budget, 2)2 8.950 < 2e-16 \*\*\*

movie\_train$title\_year -11.790 < 2e-16 \*\*\*

factor(movie\_train$genres)Adventure 6.408 1.74e-10 \*\*\*

factor(movie\_train$genres)Animation 5.257 1.58e-07 \*\*\*

factor(movie\_train$genres)Biography 8.273 < 2e-16 \*\*\*

factor(movie\_train$genres)Comedy 3.184 0.001471 \*\*

factor(movie\_train$genres)Crime 7.251 5.40e-13 \*\*\*

factor(movie\_train$genres)Documentary 8.028 1.48e-15 \*\*\*

factor(movie\_train$genres)Drama 9.925 < 2e-16 \*\*\*

factor(movie\_train$genres)Family 0.472 0.636816

factor(movie\_train$genres)Fantasy -1.492 0.135833

factor(movie\_train$genres)Horror -4.787 1.79e-06 \*\*\*

factor(movie\_train$genres)Musical -0.139 0.889184

factor(movie\_train$genres)Mystery 0.864 0.387742

factor(movie\_train$genres)Romance 1.213 0.225378

factor(movie\_train$genres)Sci-Fi 0.306 0.759288

factor(movie\_train$genres)Western 1.243 0.213871

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$num\_user\_for\_reviews NA NA

movie\_train$duration:movie\_train$num\_voted\_users -4.322 1.60e-05 \*\*\*

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 3.447 0.000576 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7344 on 2634 degrees of freedom

Multiple R-squared: 0.5167, Adjusted R-squared: 0.5112

F-statistic: 93.85 on 30 and 2634 DF, p-value: < 2.2e-16

Show in New WindowClear OutputExpand/Collapse Output

Analysis of Variance Table

Model 1: movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) +

poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

poly(movie\_train$gross, 2) + poly(movie\_train$budget, 2) +

movie\_train$title\_year + factor(movie\_train$genres) + movie\_train$duration \*

movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \*

movie\_train$num\_user\_for\_reviews + movie\_train$gross \* movie\_train$budget

Model 2: movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) +

poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year +

factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users +

movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews

Res.Df RSS Df Sum of Sq F Pr(>F)

1 2632 1418.2

2 2634 1420.6 -2 -2.4178 2.2436 0.1063

Show in New WindowClear OutputExpand/Collapse Output

not plotting observations with leverage one:

408, 1679not plotting observations with leverage one:

408, 1679NaNs producedNaNs produced

R Console

not plotting observations with leverage one:

408, 1679not plotting observations with leverage one:

408, 1679NaNs producedNaNs produced

Show in New WindowClear OutputExpand/Collapse Output

Show in New WindowClear OutputExpand/Collapse Output

Call:

lm(formula = movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users,

2) + poly(movie\_train$num\_critic\_for\_reviews, 2) + movie\_train$num\_user\_for\_reviews +

poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + poly(movie\_train$title\_year,

2) + factor(movie\_train$genres) + movie\_train$duration \*

movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \*

movie\_train$num\_user\_for\_reviews)

Residuals:

Min 1Q Median 3Q Max

-3.9264 -0.3543 0.0557 0.4566 2.1730

Coefficients: (2 not defined because of singularities)

Estimate Std. Error

(Intercept) 6.679e+00 6.701e-02

poly(movie\_train$num\_voted\_users, 2)1 3.349e+01 4.593e+00

poly(movie\_train$num\_voted\_users, 2)2 -1.878e+01 1.987e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)1 1.664e+01 1.475e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -7.765e+00 8.219e-01

movie\_train$num\_user\_for\_reviews -1.159e-03 1.044e-04

poly(movie\_train$duration, 2)1 1.358e+01 1.118e+00

poly(movie\_train$duration, 2)2 -2.192e+00 8.000e-01

movie\_train$facenumber\_in\_poster -1.776e-02 7.107e-03

movie\_train$gross -5.902e-10 3.204e-10

poly(movie\_train$budget, 2)1 -1.054e+01 1.159e+00

poly(movie\_train$budget, 2)2 7.608e+00 8.042e-01

poly(movie\_train$title\_year, 2)1 -1.309e+01 1.003e+00

poly(movie\_train$title\_year, 2)2 -5.135e+00 8.539e-01

factor(movie\_train$genres)Adventure 3.712e-01 5.497e-02

factor(movie\_train$genres)Animation 7.905e-01 1.379e-01

factor(movie\_train$genres)Biography 6.365e-01 7.601e-02

factor(movie\_train$genres)Comedy 1.446e-01 4.388e-02

factor(movie\_train$genres)Crime 4.706e-01 6.397e-02

factor(movie\_train$genres)Documentary 1.375e+00 1.627e-01

factor(movie\_train$genres)Drama 5.046e-01 4.940e-02

factor(movie\_train$genres)Family 1.591e-01 4.247e-01

factor(movie\_train$genres)Fantasy -2.701e-01 1.413e-01

factor(movie\_train$genres)Horror -4.130e-01 7.972e-02

factor(movie\_train$genres)Musical -1.662e-01 7.332e-01

factor(movie\_train$genres)Mystery 1.747e-01 1.987e-01

factor(movie\_train$genres)Romance 9.884e-01 7.321e-01

factor(movie\_train$genres)Sci-Fi 6.816e-02 3.677e-01

factor(movie\_train$genres)Western 8.681e-01 7.316e-01

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$duration:movie\_train$num\_voted\_users -1.737e-08 4.063e-09

movie\_train$num\_user\_for\_reviews:movie\_train$num\_voted\_users 1.834e-09 2.449e-10

t value Pr(>|t|)

(Intercept) 99.660 < 2e-16 \*\*\*

poly(movie\_train$num\_voted\_users, 2)1 7.292 4.03e-13 \*\*\*

poly(movie\_train$num\_voted\_users, 2)2 -9.453 < 2e-16 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)1 11.285 < 2e-16 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -9.447 < 2e-16 \*\*\*

movie\_train$num\_user\_for\_reviews -11.093 < 2e-16 \*\*\*

poly(movie\_train$duration, 2)1 12.152 < 2e-16 \*\*\*

poly(movie\_train$duration, 2)2 -2.740 0.006179 \*\*

movie\_train$facenumber\_in\_poster -2.499 0.012501 \*

movie\_train$gross -1.842 0.065598 .

poly(movie\_train$budget, 2)1 -9.098 < 2e-16 \*\*\*

poly(movie\_train$budget, 2)2 9.461 < 2e-16 \*\*\*

poly(movie\_train$title\_year, 2)1 -13.056 < 2e-16 \*\*\*

poly(movie\_train$title\_year, 2)2 -6.013 2.07e-09 \*\*\*

factor(movie\_train$genres)Adventure 6.753 1.77e-11 \*\*\*

factor(movie\_train$genres)Animation 5.731 1.11e-08 \*\*\*

factor(movie\_train$genres)Biography 8.375 < 2e-16 \*\*\*

factor(movie\_train$genres)Comedy 3.295 0.000998 \*\*\*

factor(movie\_train$genres)Crime 7.356 2.52e-13 \*\*\*

factor(movie\_train$genres)Documentary 8.452 < 2e-16 \*\*\*

factor(movie\_train$genres)Drama 10.215 < 2e-16 \*\*\*

factor(movie\_train$genres)Family 0.375 0.708030

factor(movie\_train$genres)Fantasy -1.911 0.056104 .

factor(movie\_train$genres)Horror -5.180 2.38e-07 \*\*\*

factor(movie\_train$genres)Musical -0.227 0.820641

factor(movie\_train$genres)Mystery 0.880 0.379159

factor(movie\_train$genres)Romance 1.350 0.177099

factor(movie\_train$genres)Sci-Fi 0.185 0.852950

factor(movie\_train$genres)Western 1.187 0.235469

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$duration:movie\_train$num\_voted\_users -4.275 1.98e-05 \*\*\*

movie\_train$num\_user\_for\_reviews:movie\_train$num\_voted\_users 7.490 9.36e-14 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7299 on 2643 degrees of freedom

Multiple R-squared: 0.5231, Adjusted R-squared: 0.5177

F-statistic: 96.65 on 30 and 2643 DF, p-value: < 2.2e-16

Show in New WindowClear OutputExpand/Collapse Output

Analysis of Variance Table

Model 1: movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) +

poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year +

factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users +

movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews

Model 2: movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) +

poly(movie\_train$num\_critic\_for\_reviews, 2) + movie\_train$num\_user\_for\_reviews +

poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + poly(movie\_train$title\_year,

2) + factor(movie\_train$genres) + movie\_train$duration \*

movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \*

movie\_train$num\_user\_for\_reviews

Res.Df RSS Df Sum of Sq F Pr(>F)

1 2643 1425.8

2 2643 1408.0 0 17.8

Show in New WindowClear OutputExpand/Collapse Output

Splines and/or polynomials replaced by a fitted linear combination

R Console

pseudoinverse used at 1

neighborhood radius 1

reciprocal condition number 0

pseudoinverse used at 1

neighborhood radius 1

reciprocal condition number 0

pseudoinverse used at 1

neighborhood radius 1

reciprocal condition number 0

pseudoinverse used at 1

neighborhood radius 1

reciprocal condition number 0

pseudoinverse used at 1

neighborhood radius 1

reciprocal condition number 0

pseudoinverse used at 1

neighborhood radius 1

reciprocal condition number 0

pseudoinverse used at 1

neighborhood radius 1

reciprocal condition number 0

pseudoinverse used at 1

neighborhood radius 1

reciprocal condition number 0

pseudoinverse used at 1

neighborhood radius 1

reciprocal condition number 0

pseudoinverse used at 1

neighborhood radius 1

reciprocal condition number 0

Show in New WindowClear OutputExpand/Collapse Output

248 2269 1365 2056 2243 467 703 714 2629 1196

1 2 3 4 5 6 7 8 9 10

R Console

248 2269 1365 2056 2243 467 703 714 2629 1196

1 2 3 4 5 6 7 8 9 10

Show in New WindowClear OutputExpand/Collapse Output

rstudent unadjusted p-value Bonferonni p

248 -5.353742 9.3508e-08 0.00025163

2269 -5.287391 1.3415e-07 0.00036099

1365 -4.923691 9.0119e-07 0.00242510

2056 -4.777208 1.8739e-06 0.00504280

467 -4.661852 3.2883e-06 0.00884880

2243 -4.637936 3.6891e-06 0.00992740

703 -4.561085 5.3194e-06 0.01431400

1196 -4.488397 7.4813e-06 0.02013200

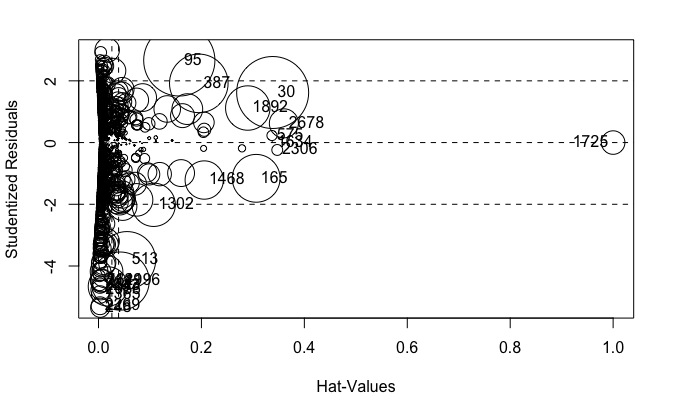
714 -4.479549 7.7957e-06 0.02097800

2629 -4.474565 7.9784e-06 0.02147000

All of the 10 residuals have significant p-values, therefore, we can drop them.

Before we drop, let's do some digsnostics to double check which to drop.

library(car)  
influencePlot(lm.fit3, id.n=10)



## StudRes Hat CookD  
30 1.622138959 0.338406745 4.203471e-02

95 2.662022785 0.157003065 4.114943e-02

165 -1.158716315 0.306363071 1.852903e-02

248 -5.353742201 0.003326096 2.958400e-03

387 1.918383293 0.194886925 2.781055e-02

418 NaN 1.000000000 NaN

467 -4.661852478 0.018231931 1.251473e-02

513 -3.811062717 0.055473567 2.652243e-02

575 0.228123357 0.337102118 8.272924e-04

703 -4.561084814 0.002700268 1.747224e-03

714 -4.479548636 0.006321509 3.960900e-03

1137 NaN 1.000000000 NaN

1196 -4.488396735 0.042017486 2.741528e-02

1302 -2.016447371 0.107168678 1.523429e-02

1365 -4.923691422 0.003286630 2.476490e-03

1468 -1.214142572 0.205494827 1.191286e-02

1634 0.009693384 0.337619819 1.497217e-06

1693 NaN 1.000000000 NaN

1725 0.000000000 1.000000000 4.480570e-03

1892 1.119323787 0.290045774 1.599398e-02

2056 -4.777207533 0.004441782 3.156045e-03

2243 -4.637936105 0.002819597 1.886164e-03

2269 -5.287390686 0.002984647 2.589097e-03

2306 -0.239672582 0.347347252 9.557014e-04

2629 -4.474564548 0.006833358 4.274362e-03

2678 0.623015593 0.360198498 6.830372e-03

From the influcence plot, we decided to drop observations: 1725,165,30,95,387,1892,2678,1468,634,57

# lm.fit5: model based on lm.fit3 removing 10 outliers.  
movie\_train<-movie\_train[-c(1725,165,30,95,387,1892,2678,1468,634,57),]  
  
lm.fit5<-lm(movie\_train$imdb\_score~poly(movie\_train$num\_voted\_users,2)+poly(movie\_train$num\_critic\_for\_reviews,2)+poly(movie\_train$num\_user\_for\_reviews,2)+poly(movie\_train$duration,2)+movie\_train$facenumber\_in\_poster+movie\_train$gross+poly(movie\_train$budget,2)+movie\_train$title\_year+factor(movie\_train$genres)+movie\_train$duration\*movie\_train$num\_voted\_users+movie\_train$num\_voted\_users\*movie\_train$num\_user\_for\_reviews)  
summary(lm.fit5)

##

Call:

lm(formula = movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users,

2) + poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year +

factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users +

movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews)

Residuals:

Min 1Q Median 3Q Max

-3.9076 -0.3552 0.0688 0.4625 2.1776

Coefficients: (3 not defined because of singularities)

Estimate Std. Error

(Intercept) 5.111e+01 3.788e+00

poly(movie\_train$num\_voted\_users, 2)1 3.936e+01 5.016e+00

poly(movie\_train$num\_voted\_users, 2)2 -1.704e+01 2.181e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)1 1.277e+01 1.325e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -7.141e+00 8.217e-01

poly(movie\_train$num\_user\_for\_reviews, 2)1 -1.851e+01 2.393e+00

poly(movie\_train$num\_user\_for\_reviews, 2)2 2.427e+00 1.578e+00

poly(movie\_train$duration, 2)1 1.372e+01 1.122e+00

poly(movie\_train$duration, 2)2 -2.483e+00 8.049e-01

movie\_train$facenumber\_in\_poster -2.261e-02 7.124e-03

movie\_train$gross -7.074e-10 3.240e-10

poly(movie\_train$budget, 2)1 -1.022e+01 1.167e+00

poly(movie\_train$budget, 2)2 7.221e+00 8.067e-01

movie\_train$title\_year -2.234e-02 1.895e-03

factor(movie\_train$genres)Adventure 3.559e-01 5.554e-02

factor(movie\_train$genres)Animation 7.284e-01 1.386e-01

factor(movie\_train$genres)Biography 6.334e-01 7.655e-02

factor(movie\_train$genres)Comedy 1.408e-01 4.423e-02

factor(movie\_train$genres)Crime 4.674e-01 6.446e-02

factor(movie\_train$genres)Documentary 1.312e+00 1.634e-01

factor(movie\_train$genres)Drama 4.938e-01 4.976e-02

factor(movie\_train$genres)Family 2.018e-01 4.273e-01

factor(movie\_train$genres)Fantasy -2.120e-01 1.421e-01

factor(movie\_train$genres)Horror -3.848e-01 8.040e-02

factor(movie\_train$genres)Musical -1.028e-01 7.377e-01

factor(movie\_train$genres)Mystery 1.728e-01 2.000e-01

factor(movie\_train$genres)Romance 8.931e-01 7.365e-01

factor(movie\_train$genres)Sci-Fi 1.134e-01 3.701e-01

factor(movie\_train$genres)Western 9.152e-01 7.361e-01

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$num\_user\_for\_reviews NA NA

movie\_train$duration:movie\_train$num\_voted\_users -1.770e-08 4.096e-09

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 1.246e-09 3.616e-10

t value Pr(>|t|)

(Intercept) 13.492 < 2e-16 \*\*\*

poly(movie\_train$num\_voted\_users, 2)1 7.846 6.21e-15 \*\*\*

poly(movie\_train$num\_voted\_users, 2)2 -7.813 7.98e-15 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)1 9.639 < 2e-16 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -8.690 < 2e-16 \*\*\*

poly(movie\_train$num\_user\_for\_reviews, 2)1 -7.737 1.44e-14 \*\*\*

poly(movie\_train$num\_user\_for\_reviews, 2)2 1.537 0.124298

poly(movie\_train$duration, 2)1 12.229 < 2e-16 \*\*\*

poly(movie\_train$duration, 2)2 -3.085 0.002058 \*\*

movie\_train$facenumber\_in\_poster -3.174 0.001523 \*\*

movie\_train$gross -2.184 0.029077 \*

poly(movie\_train$budget, 2)1 -8.755 < 2e-16 \*\*\*

poly(movie\_train$budget, 2)2 8.950 < 2e-16 \*\*\*

movie\_train$title\_year -11.790 < 2e-16 \*\*\*

factor(movie\_train$genres)Adventure 6.408 1.74e-10 \*\*\*

factor(movie\_train$genres)Animation 5.257 1.58e-07 \*\*\*

factor(movie\_train$genres)Biography 8.273 < 2e-16 \*\*\*

factor(movie\_train$genres)Comedy 3.184 0.001471 \*\*

factor(movie\_train$genres)Crime 7.251 5.40e-13 \*\*\*

factor(movie\_train$genres)Documentary 8.028 1.48e-15 \*\*\*

factor(movie\_train$genres)Drama 9.925 < 2e-16 \*\*\*

factor(movie\_train$genres)Family 0.472 0.636816

factor(movie\_train$genres)Fantasy -1.492 0.135833

factor(movie\_train$genres)Horror -4.787 1.79e-06 \*\*\*

factor(movie\_train$genres)Musical -0.139 0.889184

factor(movie\_train$genres)Mystery 0.864 0.387742

factor(movie\_train$genres)Romance 1.213 0.225378

factor(movie\_train$genres)Sci-Fi 0.306 0.759288

factor(movie\_train$genres)Western 1.243 0.213871

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$num\_user\_for\_reviews NA NA

movie\_train$duration:movie\_train$num\_voted\_users -4.322 1.60e-05 \*\*\*

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 3.447 0.000576 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7344 on 2634 degrees of freedom

Multiple R-squared: 0.5167, Adjusted R-squared: 0.5112

F-statistic: 93.85 on 30 and 2634 DF, p-value: < 2.2e-16

compareCoefs(lm.fit3, lm.fit5)

##

Call:

1: lm(formula = movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) + poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews, 2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster + movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year + factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews)

2: lm(formula = movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) + poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews, 2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year + factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users + movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews)

Est. 1 SE 1

(Intercept) 5.12e+01 3.79e+00

poly(movie\_train$num\_voted\_users, 2)1 3.98e+01 5.00e+00

poly(movie\_train$num\_voted\_users, 2)2 -1.71e+01 2.14e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)1 1.28e+01 1.33e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -7.11e+00 8.22e-01

poly(movie\_train$num\_user\_for\_reviews, 2)1 -1.85e+01 2.36e+00

poly(movie\_train$num\_user\_for\_reviews, 2)2 2.57e+00 1.56e+00

poly(movie\_train$duration, 2)1 1.38e+01 1.12e+00

poly(movie\_train$duration, 2)2 -2.50e+00 8.04e-01

movie\_train$facenumber\_in\_poster -2.24e-02 7.12e-03

movie\_train$gross -6.93e-10 3.23e-10

poly(movie\_train$budget, 2)1 -1.03e+01 1.17e+00

poly(movie\_train$budget, 2)2 7.13e+00 8.07e-01

movie\_train$title\_year -2.24e-02 1.89e-03

factor(movie\_train$genres)Adventure 3.60e-01 5.53e-02

factor(movie\_train$genres)Animation 7.27e-01 1.39e-01

factor(movie\_train$genres)Biography 6.33e-01 7.65e-02

factor(movie\_train$genres)Comedy 1.39e-01 4.41e-02

factor(movie\_train$genres)Crime 4.66e-01 6.44e-02

factor(movie\_train$genres)Documentary 1.31e+00 1.63e-01

factor(movie\_train$genres)Drama 4.96e-01 4.97e-02

factor(movie\_train$genres)Family 1.99e-01 4.27e-01

factor(movie\_train$genres)Fantasy -2.12e-01 1.42e-01

factor(movie\_train$genres)Horror -3.85e-01 8.04e-02

factor(movie\_train$genres)Musical -1.06e-01 7.38e-01

factor(movie\_train$genres)Mystery 1.73e-01 2.00e-01

factor(movie\_train$genres)Romance 8.92e-01 7.37e-01

factor(movie\_train$genres)Sci-Fi 1.13e-01 3.70e-01

factor(movie\_train$genres)Western 9.13e-01 7.36e-01

movie\_train$duration

movie\_train$num\_voted\_users

movie\_train$num\_user\_for\_reviews

movie\_train$duration:movie\_train$num\_voted\_users -1.79e-08 4.09e-09

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 1.23e-09 3.53e-10

Est. 2 SE 2

(Intercept) 5.11e+01 3.79e+00

poly(movie\_train$num\_voted\_users, 2)1 3.94e+01 5.02e+00

poly(movie\_train$num\_voted\_users, 2)2 -1.70e+01 2.18e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)1 1.28e+01 1.33e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -7.14e+00 8.22e-01

poly(movie\_train$num\_user\_for\_reviews, 2)1 -1.85e+01 2.39e+00

poly(movie\_train$num\_user\_for\_reviews, 2)2 2.43e+00 1.58e+00

poly(movie\_train$duration, 2)1 1.37e+01 1.12e+00

poly(movie\_train$duration, 2)2 -2.48e+00 8.05e-01

movie\_train$facenumber\_in\_poster -2.26e-02 7.12e-03

movie\_train$gross -7.07e-10 3.24e-10

poly(movie\_train$budget, 2)1 -1.02e+01 1.17e+00

poly(movie\_train$budget, 2)2 7.22e+00 8.07e-01

movie\_train$title\_year -2.23e-02 1.90e-03

factor(movie\_train$genres)Adventure 3.56e-01 5.55e-02

factor(movie\_train$genres)Animation 7.28e-01 1.39e-01

factor(movie\_train$genres)Biography 6.33e-01 7.66e-02

factor(movie\_train$genres)Comedy 1.41e-01 4.42e-02

factor(movie\_train$genres)Crime 4.67e-01 6.45e-02

factor(movie\_train$genres)Documentary 1.31e+00 1.63e-01

factor(movie\_train$genres)Drama 4.94e-01 4.98e-02

factor(movie\_train$genres)Family 2.02e-01 4.27e-01

factor(movie\_train$genres)Fantasy -2.12e-01 1.42e-01

factor(movie\_train$genres)Horror -3.85e-01 8.04e-02

factor(movie\_train$genres)Musical -1.03e-01 7.38e-01

factor(movie\_train$genres)Mystery 1.73e-01 2.00e-01

factor(movie\_train$genres)Romance 8.93e-01 7.37e-01

factor(movie\_train$genres)Sci-Fi 1.13e-01 3.70e-01

factor(movie\_train$genres)Western 9.15e-01 7.36e-01

movie\_train$duration

movie\_train$num\_voted\_users

movie\_train$num\_user\_for\_reviews

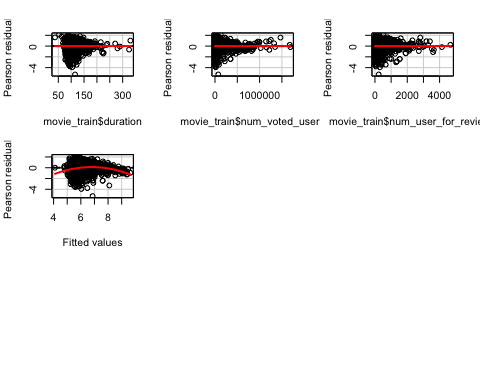
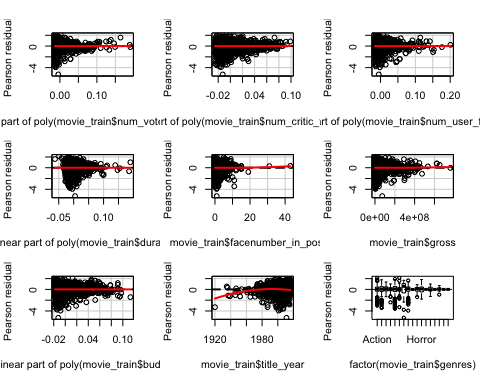
movie\_train$duration:movie\_train$num\_voted\_users -1.77e-08 4.10e-09

movie\_train$num\_voted\_users:movie\_train$num\_user\_for\_reviews 1.25e-09 3.62e-10

Removing outliers did not change the result too much.

Diagnostics for lm.fit5:

library(car)  
residualPlots(lm.fit5)



##

Test stat Pr(>|t|)

poly(movie\_train$num\_voted\_users, 2) NA NA

poly(movie\_train$num\_critic\_for\_reviews, 2) NA NA

poly(movie\_train$num\_user\_for\_reviews, 2) NA NA

poly(movie\_train$duration, 2) NA NA

movie\_train$facenumber\_in\_poster 0.564 0.573

movie\_train$gross 0.089 0.929

poly(movie\_train$budget, 2) NA NA

movie\_train$title\_year -6.247 0.000

factor(movie\_train$genres) NA NA

movie\_train$duration -0.432 0.665

movie\_train$num\_voted\_users -0.774 0.439

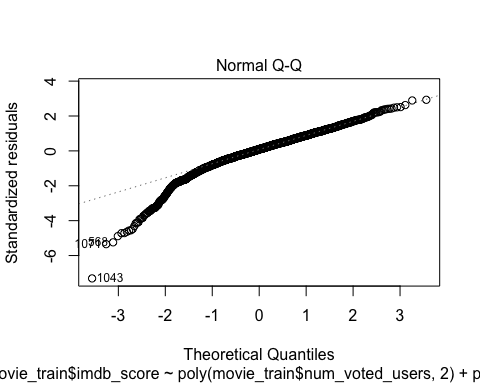
movie\_train$num\_user\_for\_reviews -0.581 0.562

Tukey test -12.602 0.000

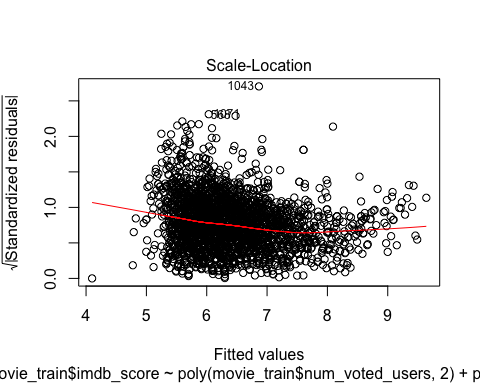
Looks good except for residuals vs fitted values show some curviture.

plot(lm.fit5)

## Warning: not plotting observations with leverage one:## 647, 837, 2554

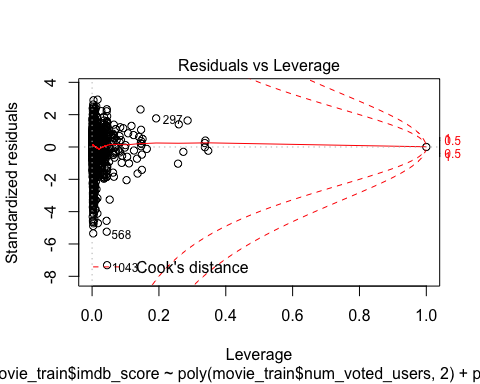


## Warning: not plotting observations with leverage one:  
## 647, 837, 2554



## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced



Now,let's look at model assumption for both lm.fit3 and lm.fit5:

# normality  
shapiro.test(lm.fit3$residuals)

Shapiro-Wilk normality test

data: lm.fit3$residuals

W = 0.94761, p-value < 2.2e-16

Shapiro-Wilk normality test

data: lm.fit5$residuals

W = 0.94742, p-value < 2.2e-16

Both models failed the normality assumption. I think this is due to the many outliers in the data set.

# equal variance : H0: variance is not constant  
ncvTest(lm.fit3)

Non-constant Variance Score Test

Variance formula: ~ fitted.values

Chisquare = 172.5443 Df = 1 p = 2.058198e-39

Non-constant Variance Score Test

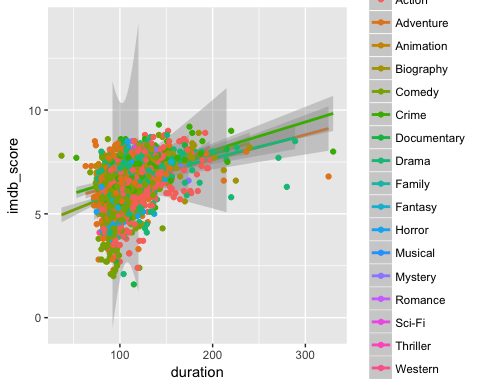
Variance formula: ~ fitted.values

Chisquare = 172.5443 Df = 1 p = 2.058198e-39

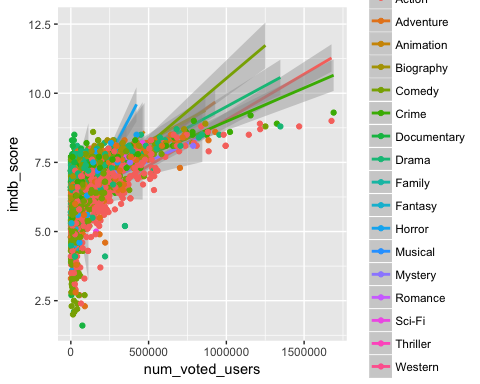
Both models passed the equal variance assumption.

This is just to explore more interesting facts Plots for data with fitted regression line:

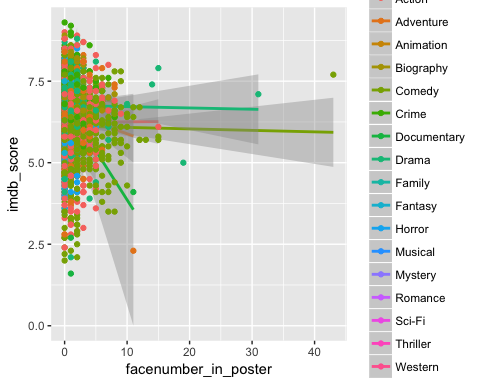
library(ggplot2)  
ggplot(data=movie\_train,aes(x=duration,y=imdb\_score,colour=factor(genres)))+stat\_smooth(method=lm,fullrange = FALSE)+geom\_point()



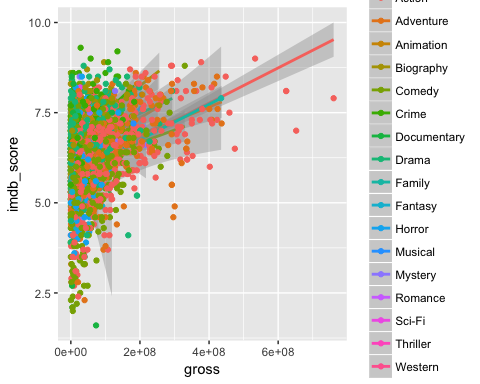
library(ggplot2)  
ggplot(data=movie\_train,aes(x=num\_voted\_users,y=imdb\_score,colour=factor(genres)))+stat\_smooth(method=lm,fullrange = FALSE)+geom\_point()



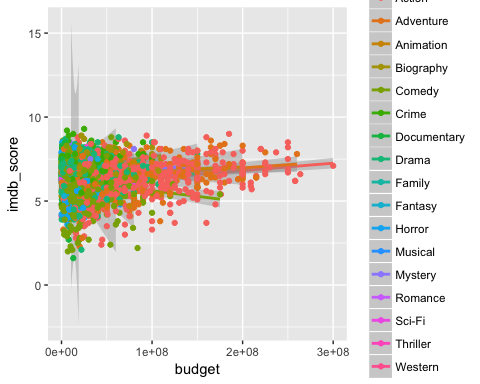
library(ggplot2)  
ggplot(data=movie\_train,aes(x=facenumber\_in\_poster,y=imdb\_score,colour=factor(genres)))+stat\_smooth(method=lm,fullrange = FALSE)+geom\_point()



library(ggplot2)  
ggplot(data=movie\_train,aes(x=gross,y=imdb\_score,colour=factor(genres)))+stat\_smooth(method=lm,fullrange = FALSE)+geom\_point()



library(ggplot2)  
ggplot(data=movie\_train,aes(x=budget,y=imdb\_score,colour=factor(genres)))+stat\_smooth(method=lm,fullrange = FALSE)+geom\_point()



## Step 4: Making predictions on the test dataset

Rewriting model lm.fit5 in another notation: # Note, if write in lm(trainx1+train$x2....), it will create the same number of values with the train data set when predict().

# lm.fit6 =lm.fit 5 using difference writing  
lm.fit6<-lm(imdb\_score~poly(num\_voted\_users,2)+poly(num\_critic\_for\_reviews,2)+poly(num\_user\_for\_reviews,2)+poly(duration,2)+facenumber\_in\_poster+gross+poly(budget,2)+title\_year+genres+duration\*num\_voted\_users+num\_voted\_users\*num\_user\_for\_reviews,data=data.frame(movie\_train))  
summary(lm.fit6)

Call:

lm(formula = imdb\_score ~ poly(num\_voted\_users, 2) + poly(num\_critic\_for\_reviews,

2) + poly(num\_user\_for\_reviews, 2) + poly(duration, 2) +

facenumber\_in\_poster + gross + poly(budget, 2) + title\_year +

genres + duration \* num\_voted\_users + num\_voted\_users \* num\_user\_for\_reviews,

data = data.frame(movie\_train))

Residuals:

Min 1Q Median 3Q Max

-3.9076 -0.3552 0.0688 0.4625 2.1776

Coefficients: (3 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept) 5.111e+01 3.788e+00 13.492 < 2e-16 \*\*\*

poly(num\_voted\_users, 2)1 3.936e+01 5.016e+00 7.846 6.21e-15 \*\*\*

poly(num\_voted\_users, 2)2 -1.704e+01 2.181e+00 -7.813 7.98e-15 \*\*\*

poly(num\_critic\_for\_reviews, 2)1 1.277e+01 1.325e+00 9.639 < 2e-16 \*\*\*

poly(num\_critic\_for\_reviews, 2)2 -7.141e+00 8.217e-01 -8.690 < 2e-16 \*\*\*

poly(num\_user\_for\_reviews, 2)1 -1.851e+01 2.393e+00 -7.737 1.44e-14 \*\*\*

poly(num\_user\_for\_reviews, 2)2 2.427e+00 1.578e+00 1.537 0.124298

poly(duration, 2)1 1.372e+01 1.122e+00 12.229 < 2e-16 \*\*\*

poly(duration, 2)2 -2.483e+00 8.049e-01 -3.085 0.002058 \*\*

facenumber\_in\_poster -2.261e-02 7.124e-03 -3.174 0.001523 \*\*

gross -7.074e-10 3.240e-10 -2.184 0.029077 \*

poly(budget, 2)1 -1.022e+01 1.167e+00 -8.755 < 2e-16 \*\*\*

poly(budget, 2)2 7.221e+00 8.067e-01 8.950 < 2e-16 \*\*\*

title\_year -2.234e-02 1.895e-03 -11.790 < 2e-16 \*\*\*

genresAdventure 3.559e-01 5.554e-02 6.408 1.74e-10 \*\*\*

genresAnimation 7.284e-01 1.386e-01 5.257 1.58e-07 \*\*\*

genresBiography 6.334e-01 7.655e-02 8.273 < 2e-16 \*\*\*

genresComedy 1.408e-01 4.423e-02 3.184 0.001471 \*\*

genresCrime 4.674e-01 6.446e-02 7.251 5.40e-13 \*\*\*

genresDocumentary 1.312e+00 1.634e-01 8.028 1.48e-15 \*\*\*

genresDrama 4.938e-01 4.976e-02 9.925 < 2e-16 \*\*\*

genresFamily 2.018e-01 4.273e-01 0.472 0.636816

genresFantasy -2.120e-01 1.421e-01 -1.492 0.135833

genresHorror -3.848e-01 8.040e-02 -4.787 1.79e-06 \*\*\*

genresMusical -1.028e-01 7.377e-01 -0.139 0.889184

genresMystery 1.728e-01 2.000e-01 0.864 0.387742

genresRomance 8.931e-01 7.365e-01 1.213 0.225378

genresSci-Fi 1.134e-01 3.701e-01 0.306 0.759288

genresWestern 9.152e-01 7.361e-01 1.243 0.213871

duration NA NA NA NA

num\_voted\_users NA NA NA NA

num\_user\_for\_reviews NA NA NA NA

duration:num\_voted\_users -1.770e-08 4.096e-09 -4.322 1.60e-05 \*\*\*

num\_voted\_users:num\_user\_for\_reviews 1.246e-09 3.616e-10 3.447 0.000576 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7344 on 2634 degrees of freedom

Multiple R-squared: 0.5167, Adjusted R-squared: 0.5112

F-statistic: 93.85 on 30 and 2634 DF, p-value: < 2.2e-16

lm.fit7<-lm(movie\_train$imdb\_score~poly(movie\_train$num\_voted\_users,2)+poly(movie\_train$num\_critic\_for\_reviews,2)+movie\_train$num\_user\_for\_reviews+poly(movie\_train$duration,2)+movie\_train$facenumber\_in\_poster+movie\_train$gross+poly(movie\_train$budget,2)+movie\_train$title\_year+factor(movie\_train$genres)+movie\_train$duration\*movie\_train$num\_voted\_users+movie\_train$num\_voted\_users\*movie\_train$num\_user\_for\_reviews)  
summary(lm.fit7)

Call:

lm(formula = movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users,

2) + poly(movie\_train$num\_critic\_for\_reviews, 2) + movie\_train$num\_user\_for\_reviews +

poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year +

factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users +

movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews)

Residuals:

Min 1Q Median 3Q Max

-3.9231 -0.3519 0.0683 0.4620 2.1810

Coefficients: (2 not defined because of singularities)

Estimate Std. Error

(Intercept) 5.082e+01 3.772e+00

poly(movie\_train$num\_voted\_users, 2)1 3.626e+01 4.594e+00

poly(movie\_train$num\_voted\_users, 2)2 -1.830e+01 2.021e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)1 1.244e+01 1.308e+00

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -6.983e+00 8.155e-01

movie\_train$num\_user\_for\_reviews -1.024e-03 1.037e-04

poly(movie\_train$duration, 2)1 1.368e+01 1.122e+00

poly(movie\_train$duration, 2)2 -2.445e+00 8.047e-01

movie\_train$facenumber\_in\_poster -2.222e-02 7.121e-03

movie\_train$gross -6.636e-10 3.228e-10

poly(movie\_train$budget, 2)1 -1.034e+01 1.164e+00

poly(movie\_train$budget, 2)2 7.266e+00 8.064e-01

movie\_train$title\_year -2.204e-02 1.886e-03

factor(movie\_train$genres)Adventure 3.587e-01 5.553e-02

factor(movie\_train$genres)Animation 7.349e-01 1.385e-01

factor(movie\_train$genres)Biography 6.365e-01 7.655e-02

factor(movie\_train$genres)Comedy 1.411e-01 4.424e-02

factor(movie\_train$genres)Crime 4.667e-01 6.448e-02

factor(movie\_train$genres)Documentary 1.314e+00 1.634e-01

factor(movie\_train$genres)Drama 4.931e-01 4.977e-02

factor(movie\_train$genres)Family 1.963e-01 4.274e-01

factor(movie\_train$genres)Fantasy -2.196e-01 1.420e-01

factor(movie\_train$genres)Horror -3.938e-01 8.021e-02

factor(movie\_train$genres)Musical -1.073e-01 7.378e-01

factor(movie\_train$genres)Mystery 1.636e-01 2.000e-01

factor(movie\_train$genres)Romance 8.950e-01 7.367e-01

factor(movie\_train$genres)Sci-Fi 9.746e-02 3.700e-01

factor(movie\_train$genres)Western 9.165e-01 7.363e-01

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$duration:movie\_train$num\_voted\_users -1.791e-08 4.095e-09

movie\_train$num\_user\_for\_reviews:movie\_train$num\_voted\_users 1.650e-09 2.488e-10

t value Pr(>|t|)

(Intercept) 13.472 < 2e-16 \*\*\*

poly(movie\_train$num\_voted\_users, 2)1 7.891 4.35e-15 \*\*\*

poly(movie\_train$num\_voted\_users, 2)2 -9.054 < 2e-16 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)1 9.513 < 2e-16 \*\*\*

poly(movie\_train$num\_critic\_for\_reviews, 2)2 -8.563 < 2e-16 \*\*\*

movie\_train$num\_user\_for\_reviews -9.875 < 2e-16 \*\*\*

poly(movie\_train$duration, 2)1 12.193 < 2e-16 \*\*\*

poly(movie\_train$duration, 2)2 -3.038 0.00240 \*\*

movie\_train$facenumber\_in\_poster -3.119 0.00183 \*\*

movie\_train$gross -2.056 0.03990 \*

poly(movie\_train$budget, 2)1 -8.886 < 2e-16 \*\*\*

poly(movie\_train$budget, 2)2 9.011 < 2e-16 \*\*\*

movie\_train$title\_year -11.691 < 2e-16 \*\*\*

factor(movie\_train$genres)Adventure 6.460 1.25e-10 \*\*\*

factor(movie\_train$genres)Animation 5.305 1.22e-07 \*\*\*

factor(movie\_train$genres)Biography 8.315 < 2e-16 \*\*\*

factor(movie\_train$genres)Comedy 3.189 0.00144 \*\*

factor(movie\_train$genres)Crime 7.238 5.95e-13 \*\*\*

factor(movie\_train$genres)Documentary 8.037 1.37e-15 \*\*\*

factor(movie\_train$genres)Drama 9.908 < 2e-16 \*\*\*

factor(movie\_train$genres)Family 0.459 0.64610

factor(movie\_train$genres)Fantasy -1.546 0.12222

factor(movie\_train$genres)Horror -4.909 9.70e-07 \*\*\*

factor(movie\_train$genres)Musical -0.145 0.88438

factor(movie\_train$genres)Mystery 0.818 0.41324

factor(movie\_train$genres)Romance 1.215 0.22450

factor(movie\_train$genres)Sci-Fi 0.263 0.79228

factor(movie\_train$genres)Western 1.245 0.21332

movie\_train$duration NA NA

movie\_train$num\_voted\_users NA NA

movie\_train$duration:movie\_train$num\_voted\_users -4.375 1.26e-05 \*\*\*

movie\_train$num\_user\_for\_reviews:movie\_train$num\_voted\_users 6.632 4.00e-11 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7346 on 2635 degrees of freedom

Multiple R-squared: 0.5162, Adjusted R-squared: 0.5109

F-statistic: 96.96 on 29 and 2635 DF, p-value: < 2.2e-16

anova(lm.fit5,lm.fit7)

Analysis of Variance Table

Model 1: movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) +

poly(movie\_train$num\_critic\_for\_reviews, 2) + poly(movie\_train$num\_user\_for\_reviews,

2) + poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year +

factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users +

movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews

Model 2: movie\_train$imdb\_score ~ poly(movie\_train$num\_voted\_users, 2) +

poly(movie\_train$num\_critic\_for\_reviews, 2) + movie\_train$num\_user\_for\_reviews +

poly(movie\_train$duration, 2) + movie\_train$facenumber\_in\_poster +

movie\_train$gross + poly(movie\_train$budget, 2) + movie\_train$title\_year +

factor(movie\_train$genres) + movie\_train$duration \* movie\_train$num\_voted\_users +

movie\_train$num\_voted\_users \* movie\_train$num\_user\_for\_reviews

Res.Df RSS Df Sum of Sq F Pr(>F)

1 2634 1420.6

2 2635 1421.9 -1 -1.2749 2.3638 0.1243

pr<-predict.lm(lm.fit6,newdata = data.frame(movie\_test),interval = 'confidence')

## Warning in predict.lm(lm.fit6, newdata = data.frame(movie\_test), interval =  
## "confidence"): prediction from a rank-deficient fit may be misleading

pr

## fit lwr upr

4 7.129879 6.601719 7.658040

18 6.804547 6.346576 7.262518

36 7.154564 6.972710 7.336418

41 6.828136 6.711522 6.944749

52 6.800389 6.623916 6.976861

54 7.073166 6.908433 7.237899

62 6.490910 6.181509 6.800311

72 5.489215 5.346246 5.632185

75 5.995054 5.837136 6.152971

97 7.832485 7.400888 8.264081

102 8.098861 7.917276 8.280446

109 5.689237 5.555231 5.823243

149 6.096397 5.962084 6.230709

158 5.717597 5.613821 5.821372

164 5.495273 5.380139 5.610406

165 5.488178 5.379152 5.597204

184 7.513695 7.399354 7.628036

190 6.718810 6.322116 7.115503

192 5.827696 5.711272 5.944119

194 6.096459 5.968611 6.224307

220 6.776458 6.661682 6.891234

223 6.248098 6.140688 6.355508

233 5.981958 5.873856 6.090060

241 7.675872 6.997657 8.354087

262 6.444692 6.355051 6.534334

292 6.581843 6.478196 6.685490

297 7.110811 6.568327 7.653295

304 5.868093 5.720774 6.015413

305 6.089019 5.977549 6.200489

313 5.705722 5.627517 5.783927

324 5.435453 5.310373 5.560533

340 7.678189 7.099021 8.257357

346 7.191456 7.032743 7.350169

362 8.600062 8.363232 8.836893

379 6.110887 6.003127 6.218647

395 6.043866 5.926990 6.160742

399 7.040110 6.936514 7.143706

404 6.831316 6.565035 7.097597

406 6.499418 6.414172 6.584664

410 6.373834 6.268414 6.479254

427 5.713178 5.623953 5.802403

435 8.076745 7.915924 8.237566

440 7.719358 7.385065 8.053650

453 8.104699 7.636570 8.572828

463 6.611594 6.472484 6.750704

475 5.683846 5.593161 5.774531

481 5.781978 5.513675 6.050282

494 5.762165 5.670297 5.854032

499 5.520375 5.429795 5.610955

502 5.697431 5.605593 5.789269

519 5.319947 5.200686 5.439207

525 6.895383 6.762886 7.027880

534 5.408915 5.324487 5.493343

536 6.263727 6.166085 6.361368

551 6.383371 6.121723 6.645020

553 6.004131 5.912677 6.095586

566 6.135612 6.006348 6.264876

570 6.667781 6.554252 6.781310

574 8.411075 8.095212 8.726939

576 6.813014 6.678954 6.947073

577 6.524214 6.412930 6.635498

599 5.764465 5.662462 5.866469

616 5.477677 5.400804 5.554550

617 5.883344 5.793426 5.973262

634 8.000097 7.700984 8.299210

642 6.809254 6.715172 6.903335

673 5.570513 5.465615 5.675411

680 6.726826 6.631322 6.822331

702 5.391216 5.309683 5.472749

706 5.807916 5.720578 5.895254

710 7.916905 7.709705 8.124104

723 6.565277 6.473309 6.657246

730 5.546177 5.423775 5.668579

741 5.901221 5.812799 5.989643

767 5.876873 5.806360 5.947386

789 7.093227 6.940793 7.245661

801 6.395657 6.297152 6.494162

805 6.098607 6.010300 6.186914

807 5.794258 5.716980 5.871537

816 5.986424 5.872011 6.100838

820 5.900768 5.796700 6.004836

842 5.912498 5.818877 6.006119

844 5.510298 5.418937 5.601658

849 7.017008 6.920682 7.113335

851 6.192140 6.112388 6.271893

872 6.066216 5.975162 6.157270

884 6.598269 5.771500 7.425037

897 6.049837 5.949201 6.150472

929 7.695470 6.249330 9.141611

933 7.234769 7.131099 7.338440

934 6.332482 6.172002 6.492962

941 7.478458 7.365635 7.591281

952 5.719089 5.646883 5.791295

1012 6.676421 6.555409 6.797432

1020 6.156160 6.031700 6.280619

1035 5.519650 5.434125 5.605174

1039 7.434400 7.233392 7.635407

1049 6.869870 6.780802 6.958937

1066 5.717017 5.630988 5.803047

1074 5.685882 5.589208 5.782557

1085 8.225810 8.060569 8.391051

1094 6.135485 6.054943 6.216027

1103 5.480237 5.414681 5.545794

1112 5.736857 5.633490 5.840224

1119 5.932489 5.842589 6.022389

1120 5.460822 5.351631 5.570012

1140 6.235391 6.156304 6.314479

1147 6.233684 6.147270 6.320098

1151 5.399389 5.311222 5.487555

1158 5.957840 5.883606 6.032074

1180 6.403172 6.309657 6.496687

1208 6.315142 6.239073 6.391211

1211 5.488830 5.393219 5.584441

1229 6.171675 6.099918 6.243432

1237 6.998408 6.881459 7.115357

1241 6.423667 6.036319 6.811015

1255 7.329921 7.208554 7.451289

1271 6.261405 6.180550 6.342260

1273 5.911551 5.838111 5.984990

1302 6.185764 6.043325 6.328203

1318 6.004711 5.910194 6.099229

1343 5.770357 5.696084 5.844630

1353 6.136402 6.054461 6.218344

1365 6.541353 6.090551 6.992154

1371 5.925128 5.774452 6.075804

1372 7.179208 7.079405 7.279012

1375 7.599312 7.406848 7.791777

1390 5.768766 5.697453 5.840080

1392 6.418355 6.335446 6.501264

1397 5.441832 5.370803 5.512861

1428 5.672205 5.588801 5.755609

1431 6.850461 6.753479 6.947443

1517 5.521986 5.441422 5.602550

1519 6.514907 6.419784 6.610030

1534 6.788030 6.703776 6.872284

1544 6.309473 6.231475 6.387471

1546 7.203227 7.081341 7.325113

1572 8.440940 7.614709 9.267172

1581 6.222448 6.146017 6.298880

1584 6.186199 6.072914 6.299485

1591 7.094998 6.987345 7.202650

1593 7.384152 7.241913 7.526390

1608 6.709391 6.631729 6.787052

1616 7.722129 7.516864 7.927395

1623 5.915616 5.851092 5.980140

1640 6.242537 6.177099 6.307975

1681 6.751281 6.655609 6.846954

1695 6.849389 6.727389 6.971389

1699 6.612137 6.512331 6.711943

1700 7.060396 6.898439 7.222352

1702 6.025601 5.942189 6.109014

1708 6.216294 6.138427 6.294161

1736 7.732113 7.574336 7.889889

1758 6.151356 5.991467 6.311245

1760 6.652922 6.537812 6.768033

1772 6.261768 6.197013 6.326523

1807 6.599329 6.485419 6.713238

1814 7.503488 7.278006 7.728971

1860 6.091739 5.972299 6.211179

1882 5.294651 5.198055 5.391247

1884 5.839828 5.748134 5.931523

1901 6.920514 6.810505 7.030522

1914 5.512259 5.400279 5.624239

1929 5.627333 5.554115 5.700552

1944 6.084466 6.025578 6.143354

1965 5.709568 5.634590 5.784545

1982 6.417559 6.339578 6.495541

1999 5.398261 5.300035 5.496487

2017 6.886101 6.797158 6.975044

2022 6.234111 6.077633 6.390590

2027 6.485194 6.392428 6.577960

2055 6.271453 6.127360 6.415547

2077 6.743979 6.651770 6.836187

2086 6.166397 6.086115 6.246678

2093 6.675307 6.476274 6.874341

2095 6.421349 6.342899 6.499800

2105 6.211968 6.093879 6.330057

2107 6.200788 6.089885 6.311692

2110 6.248472 6.161411 6.335532

2138 8.274940 8.087480 8.462401

2145 6.313794 6.235620 6.391968

2151 5.976658 5.872490 6.080826

2183 5.322050 5.205356 5.438743

2197 5.874918 5.716487 6.033348

2198 6.013236 5.924982 6.101490

2276 5.984150 5.914245 6.054055

2277 6.247489 6.095911 6.399066

2281 6.385939 6.312127 6.459752

2321 5.563688 5.479233 5.648142

2361 6.666432 6.517267 6.815596

2362 5.428520 5.291209 5.565830

2366 5.614390 5.516315 5.712465

2369 5.966485 5.910141 6.022830

2395 7.040163 6.881628 7.198697

2397 6.485774 6.390357 6.581191

2411 6.039443 5.964509 6.114376

2414 6.200808 6.118950 6.282666

2415 7.462884 7.235415 7.690353

2417 6.662253 6.540807 6.783698

2428 6.542666 6.425043 6.660289

2429 6.262545 6.172041 6.353048

2493 6.739390 6.557150 6.921630

2499 6.259760 6.195159 6.324360

2503 6.015693 5.870145 6.161241

2505 6.678693 6.528808 6.828579

2524 7.884262 7.768314 8.000209

2583 6.784890 6.683991 6.885789

2602 7.478941 7.315722 7.642159

2616 6.354617 6.256899 6.452335

2624 7.046294 6.951595 7.140993

2632 5.903946 5.800411 6.007481

2648 6.944136 6.823380 7.064893

2654 7.777866 7.459165 8.096567

2671 5.524389 5.386790 5.661987

2674 5.364658 5.278806 5.450510

2693 6.321836 6.207688 6.435984

2700 5.949545 5.865393 6.033697

2726 6.241788 6.155404 6.328173

2748 5.566248 5.498107 5.634390

2780 6.217779 6.101931 6.333628

2799 5.951655 5.848787 6.054523

2835 7.333314 7.020621 7.646006

2836 8.288568 7.996680 8.580456

2848 5.662567 5.561900 5.763234

2862 5.800709 5.684811 5.916607

2882 5.598427 5.524738 5.672115

2898 6.370293 6.296825 6.443762

2919 7.243303 7.102616 7.383991

2952 5.820413 5.687366 5.953460

2972 6.031895 5.908112 6.155678

2977 6.264199 5.536966 6.991432

2981 5.802273 5.723753 5.880794

2985 5.922611 5.850503 5.994719

3027 8.140564 7.889692 8.391437

3053 5.863304 5.799148 5.927461

3082 5.938765 5.870187 6.007344

3098 6.050786 5.911330 6.190241

3101 7.040554 6.947374 7.133735

3103 5.994181 5.883235 6.105127

3114 5.904737 5.757691 6.051783

3123 6.090794 5.817251 6.364338

3133 6.087172 5.973225 6.201119

3145 6.747028 6.649189 6.844867

3151 7.722884 7.611007 7.834762

3171 6.411048 6.321377 6.500720

3182 5.881579 5.804441 5.958717

3203 6.164654 6.080607 6.248702

3222 6.514979 6.369345 6.660612

3229 5.889601 5.791843 5.987358

3261 5.937535 5.873176 6.001893

3316 5.830113 5.732558 5.927667

3334 6.975751 6.864647 7.086855

3516 6.562407 6.422160 6.702654

3548 5.998012 5.927612 6.068412

3571 6.192320 6.113013 6.271627

3607 5.876612 5.802018 5.951207

3609 5.802287 5.661945 5.942629

3626 5.740117 5.649465 5.830769

3648 6.397714 6.277700 6.517728

3715 8.622920 8.437913 8.807926

3727 6.392880 6.311022 6.474739

3740 5.502112 5.366128 5.638096

3747 6.335726 6.228853 6.442598

3748 6.180521 6.080246 6.280797

3749 5.547474 5.409086 5.685863

3850 8.891945 8.641328 9.142562

3880 5.868224 5.720906 6.015541

3893 6.780227 6.686757 6.873697

3894 6.468088 6.367736 6.568440

3907 8.102296 7.912003 8.292588

3924 6.903891 6.595190 7.212592

3939 6.963215 6.862916 7.063514

4026 7.246280 7.070115 7.422444

4028 6.215870 6.133493 6.298246

4066 5.792091 5.644070 5.940112

4189 6.702242 6.616436 6.788048

4193 6.495390 6.416523 6.574257

4208 5.985247 5.890036 6.080458

4220 5.968388 5.858358 6.078417

4239 8.458368 8.225272 8.691463

4334 5.938129 5.859164 6.017093

4384 5.961219 5.896438 6.026000

4403 6.418801 6.337379 6.500222

4406 6.380558 6.281240 6.479877

4436 5.917929 5.820692 6.015167

4487 6.357209 6.280151 6.434266

4533 6.450410 6.166879 6.733941

4535 5.546226 5.401618 5.690834

4537 6.120125 5.962413 6.277838

4577 5.901948 5.825905 5.977991

4584 5.421954 5.336494 5.507414

4654 5.871819 5.788344 5.955294

4785 6.062158 5.985173 6.139142

4789 6.267175 6.189468 6.344883

4813 7.061382 5.601669 8.521095

4831 5.808854 5.721292 5.896415

4841 6.219130 6.120124 6.318136

4874 5.537256 4.807866 6.266646

4894 6.958121 6.751135 7.165106

5005 6.036437 4.591497 7.481378

5043 6.843895 6.534990 7.152801

Check Accuracy: Mean Absolute Error: how far, on average, prediction is from the true value.

MAE <- function(actual, predicted) {  
mean(abs(actual - predicted))  
}  
MAE(pr, movie\_test$imdb\_score)

[1] 0.5174271

Checking the impact significance of predictors on IMDB score.

# stantdardized regression coefficients  
library(QuantPsyc)

## Loading required package: boot

##   
## Attaching package: 'boot'

## The following object is masked from 'package:car':  
##   
## logit

## The following object is masked from 'package:psych':  
##   
## logit

## Loading required package: MASS

##   
## Attaching package: 'QuantPsyc'

## The following object is masked from 'package:base':  
##   
## norm

lm.beta(lm.fit6)

## Warning in var(if (is.vector(x) || is.factor(x)) x else as.double(x), na.rm = na.rm): Calling var(x) on a factor x is deprecated and will become an error.  
## Use something like 'all(duplicated(x)[-1L])' to test for a constant vector.

## Warning in b \* sx: longer object length is not a multiple of shorter object  
## length

Calling var(x) on a factor x is deprecated and will become an error.

Use something like 'all(duplicated(x)[-1L])' to test for a constant vector.longer object length is not a multiple of shorter object length poly(num\_voted\_users, 2)1 poly(num\_voted\_users, 2)2

7.258695e-01 -3.142231e-01

poly(num\_critic\_for\_reviews, 2)1 poly(num\_critic\_for\_reviews, 2)2

2.355688e-01 -1.317021e-01

poly(num\_user\_for\_reviews, 2)1 poly(num\_user\_for\_reviews, 2)2

-3.685777e+01 1.657757e+08

poly(duration, 2)1 poly(duration, 2)2

2.531245e-01 -2.335428e+01

facenumber\_in\_poster gross

-6.433104e-02 -1.394813e-08

poly(budget, 2)1 poly(budget, 2)2

-1.440924e+06 2.649445e+03

title\_year genresAdventure

-4.121020e-04 6.564319e-03

genresAnimation genresBiography

1.343445e-02 1.168137e-02

genresComedy genresCrime

2.803634e-01 3.193309e+07

genresDocumentary genresDrama

2.419307e-02 4.645184e+00

genresFamily genresFantasy

5.740956e-01 -4.179221e+00

genresHorror genresMusical

-5.427878e+04 -3.771834e+01

genresMystery genresRomance

3.186344e-03 1.647229e-02

genresSci-Fi genresWestern

2.091717e-03 1.687876e-02

duration:num\_voted\_users num\_voted\_users:num\_user\_for\_reviews

-3.524642e-08 8.514159e-02

Conclusion: The most important factor that affects movie rating is the duration. The longer the movie is, the higher the rating will be. num\_critic\_for\_reviews is also an important predictor. Budget is important, although there is no strong correlation between budget and movie rating. The number of faces in movie poster has a non-neglectable effect to the movie rating.