FML Assignment 1

Akshaya Mamidipalli

2024-02-04

#loading the required packages

library(ggplot2)

## Warning: package 'ggplot2' was built under R version 4.3.2

library(tinytex)

#downloaded the dataset from the kaggle (<https://www.kaggle.com/datasets/arthurchongg/imdb-top-1000-movies>)

#loading the dataset

IMDB\_dataset <- read.csv("C:/Users/mamid/Downloads/IMDB\_1000.csv")

#summary of the dataset

summary(IMDB\_dataset)

## title director release\_year runtime   
## Length:1000 Length:1000 Length:1000 Length:1000   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## genre rating metascore gross   
## Length:1000 Min. :7.600 Min. : 0.00 Length:1000   
## Class :character 1st Qu.:7.800 1st Qu.: 64.75 Class :character   
## Mode :character Median :7.900 Median : 77.00 Mode :character   
## Mean :7.969 Mean : 66.65   
## 3rd Qu.:8.100 3rd Qu.: 86.00   
## Max. :9.300 Max. :100.00

#mean of the rating variable

mean(IMDB\_dataset$rating)

## [1] 7.9687

#descriptive statistics of the variables

IMDB\_quantitative <- IMDB\_dataset[, c('release\_year','runtime','rating','metascore','gross')]  
  
IMDB\_categorical <- IMDB\_dataset[, c('title' , 'director', 'genre')]

#printing the descriptive statistics

summary(IMDB\_quantitative)

## release\_year runtime rating metascore   
## Length:1000 Length:1000 Min. :7.600 Min. : 0.00   
## Class :character Class :character 1st Qu.:7.800 1st Qu.: 64.75   
## Mode :character Mode :character Median :7.900 Median : 77.00   
## Mean :7.969 Mean : 66.65   
## 3rd Qu.:8.100 3rd Qu.: 86.00   
## Max. :9.300 Max. :100.00   
## gross   
## Length:1000   
## Class :character   
## Mode :character   
##   
##   
##

summary(IMDB\_categorical)

## title director genre   
## Length:1000 Length:1000 Length:1000   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character

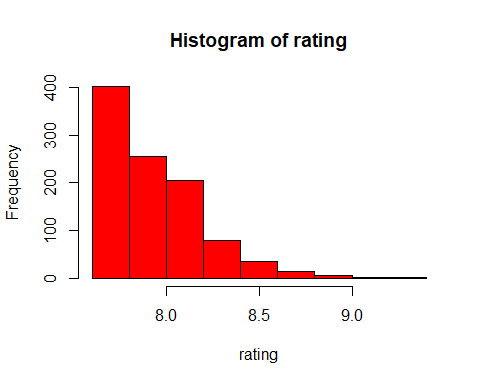
#log transformation of quantitative variable

log(IMDB\_dataset$metascore)

## [1] 4.406719 4.605170 4.430817 4.553877 4.574711 4.543295 4.499810 4.454347  
## [9] 4.553877 4.304065 4.521789 4.204693 4.406719 4.499810 4.465908 -Inf  
## [17] -Inf 4.304065 4.521789 4.290459 4.430817 4.406719 -Inf -Inf  
## [25] 4.110874 4.317488 4.174387 4.510860 4.454347 4.499810 4.564348 4.369448  
## [33] 4.077537 4.584967 4.488636 4.442651 -Inf 4.465908 4.442651 4.204693  
## [41] 4.488636 4.564348 4.189655 4.488636 4.158883 4.477337 4.343805 4.605170  
## [49] 4.127134 4.574711 4.442651 4.043051 4.406719 4.543295 4.605170 4.382027  
## [57] 4.564348 4.595120 -Inf 4.442651 4.465908 4.356709 4.394449 4.543295  
## [65] 4.430817 4.077537 4.356709 4.219508 4.189655 4.343805 4.418841 4.477337  
## [73] -Inf 4.394449 4.394449 4.553877 4.204693 4.488636 4.574711 4.454347  
## [81] 4.317488 4.543295 4.499810 4.330733 -Inf 4.499810 -Inf -Inf  
## [89] -Inf -Inf 4.356709 4.330733 4.234107 4.356709 4.174387 4.430817  
## [97] 4.262680 4.248495 4.488636 4.343805 4.605170 4.219508 4.430817 4.394449  
## [105] 4.317488 4.343805 4.564348 4.488636 4.477337 4.060443 4.234107 4.382027  
## [113] 4.605170 4.521789 4.330733 4.584967 4.477337 4.605170 4.595120 4.584967  
## [121] 4.418841 4.553877 -Inf 4.543295 -Inf 4.521789 4.094345 -Inf  
## [129] 4.553877 -Inf -Inf -Inf -Inf -Inf -Inf -Inf  
## [137] -Inf -Inf -Inf 4.158883 4.174387 4.317488 4.262680 4.532599  
## [145] 4.219508 4.143135 4.521789 4.043051 4.248495 4.543295 4.356709 4.290459  
## [153] 4.007333 4.406719 4.442651 4.499810 4.234107 4.234107 4.510860 4.158883  
## [161] 4.499810 4.276666 4.276666 4.477337 4.406719 4.574711 4.521789 4.127134  
## [169] 4.584967 4.454347 4.584967 4.510860 4.356709 4.276666 4.382027 4.304065  
## [177] 4.574711 4.174387 -Inf 4.317488 4.584967 -Inf -Inf 4.488636  
## [185] 4.584967 4.584967 4.532599 4.454347 -Inf -Inf -Inf -Inf  
## [193] -Inf 4.454347 4.605170 4.343805 -Inf -Inf 4.442651 -Inf  
## [201] 4.584967 -Inf -Inf -Inf 4.521789 -Inf -Inf -Inf  
## [209] -Inf 4.465908 4.477337 4.248495 4.143135 4.430817 4.430817 4.499810  
## [217] 4.369448 4.394449 4.127134 4.343805 4.317488 4.262680 4.262680 4.317488  
## [225] 4.564348 4.442651 4.394449 4.543295 4.442651 4.248495 4.510860 4.454347  
## [233] 4.418841 4.521789 4.477337 4.369448 4.290459 4.454347 4.369448 4.406719  
## [241] 4.454347 4.343805 4.304065 4.564348 4.532599 4.143135 4.521789 4.317488  
## [249] 4.189655 4.262680 4.553877 4.276666 4.454347 4.488636 4.465908 -Inf  
## [257] -Inf 4.454347 4.499810 4.521789 4.442651 4.343805 4.510860 4.442651  
## [265] 4.356709 4.418841 4.499810 4.394449 4.477337 4.465908 4.454347 4.430817  
## [273] 4.369448 4.418841 -Inf 4.454347 -Inf 4.543295 4.317488 4.574711  
## [281] 4.343805 4.532599 4.564348 4.510860 -Inf 4.477337 -Inf 4.605170  
## [289] 4.406719 -Inf 4.465908 4.605170 -Inf 4.477337 -Inf 4.430817  
## [297] 4.543295 4.369448 4.564348 4.418841 4.290459 4.477337 -Inf -Inf  
## [305] -Inf 4.262680 4.488636 4.276666 -Inf -Inf -Inf -Inf  
## [313] -Inf -Inf -Inf 4.553877 -Inf -Inf 4.574711 -Inf  
## [321] -Inf -Inf -Inf -Inf 4.219508 -Inf -Inf -Inf  
## [329] -Inf -Inf -Inf -Inf 4.290459 -Inf -Inf -Inf  
## [337] 4.499810 -Inf -Inf -Inf -Inf -Inf -Inf 4.304065  
## [345] 4.394449 4.330733 4.174387 4.174387 4.369448 4.343805 4.234107 4.543295  
## [353] 4.382027 4.477337 4.499810 4.510860 4.356709 4.418841 4.382027 4.330733  
## [361] 4.262680 4.304065 4.276666 4.276666 4.356709 4.430817 4.077537 4.418841  
## [369] 4.276666 4.532599 4.454347 4.418841 4.454347 4.158883 4.304065 4.158883  
## [377] 4.060443 4.553877 4.276666 4.488636 4.369448 4.442651 4.204693 4.564348  
## [385] 4.204693 -Inf 4.418841 4.343805 4.189655 4.477337 4.110874 4.204693  
## [393] 4.234107 4.317488 4.077537 4.430817 4.356709 4.356709 4.060443 4.304065  
## [401] 4.418841 4.532599 4.532599 4.521789 4.454347 4.574711 4.532599 4.356709  
## [409] 4.532599 3.850148 4.234107 4.060443 4.574711 -Inf 4.553877 4.369448  
## [417] 4.369448 4.317488 4.317488 4.454347 -Inf 4.553877 4.465908 4.595120  
## [425] 4.406719 -Inf 4.499810 4.605170 4.442651 -Inf 3.496508 4.605170  
## [433] 4.488636 4.477337 -Inf 4.430817 4.510860 4.521789 4.442651 4.304065  
## [441] -Inf -Inf 4.521789 -Inf -Inf 4.219508 4.499810 -Inf  
## [449] -Inf 4.290459 4.574711 4.454347 4.477337 -Inf 4.442651 -Inf  
## [457] 4.262680 4.564348 -Inf -Inf -Inf 4.382027 -Inf -Inf  
## [465] 4.158883 -Inf -Inf -Inf -Inf -Inf -Inf 4.317488  
## [473] 4.356709 4.418841 4.290459 3.988984 4.394449 4.418841 4.406719 4.369448  
## [481] 4.454347 4.406719 4.406719 4.077537 4.442651 4.304065 4.262680 4.204693  
## [489] 4.317488 4.330733 4.499810 4.060443 4.219508 3.891820 4.304065 4.418841  
## [497] 4.521789 4.430817 4.290459 4.369448 4.204693 4.430817 4.094345 4.430817  
## [505] 4.343805 4.394449 4.543295 4.189655 4.430817 4.477337 4.543295 4.605170  
## [513] 4.521789 4.189655 -Inf 3.850148 4.430817 4.330733 4.406719 4.543295  
## [521] 4.174387 4.465908 4.382027 4.574711 -Inf 4.543295 4.406719 4.584967  
## [529] 4.330733 4.234107 4.262680 -Inf 4.543295 4.343805 4.369448 4.127134  
## [537] 4.234107 4.605170 4.442651 -Inf 4.488636 4.521789 4.532599 4.343805  
## [545] 4.510860 4.143135 4.465908 4.477337 4.406719 4.521789 4.454347 4.454347  
## [553] 4.465908 4.564348 4.430817 -Inf 4.356709 4.605170 4.553877 4.454347  
## [561] 3.988984 4.454347 -Inf -Inf -Inf 4.574711 3.912023 4.234107  
## [569] 4.477337 -Inf 4.488636 -Inf -Inf 4.276666 4.394449 4.262680  
## [577] 4.430817 4.174387 4.477337 4.382027 4.543295 -Inf -Inf 4.574711  
## [585] 4.574711 -Inf -Inf -Inf 4.595120 -Inf 4.406719 4.521789  
## [593] 4.430817 -Inf 4.430817 -Inf -Inf -Inf 4.276666 4.394449  
## [601] 3.912023 4.382027 4.543295 4.382027 4.430817 4.189655 3.931826 4.442651  
## [609] 3.970292 4.330733 4.356709 4.406719 4.394449 4.394449 4.543295 4.418841  
## [617] 4.007333 4.219508 4.262680 4.543295 4.510860 4.262680 4.007333 3.850148  
## [625] 4.553877 4.406719 3.871201 4.110874 4.060443 4.174387 4.382027 4.219508  
## [633] 4.330733 4.465908 4.304065 4.330733 4.442651 4.330733 4.248495 4.317488  
## [641] 4.369448 3.931826 4.290459 4.043051 4.499810 4.204693 4.234107 4.248495  
## [649] 4.394449 4.406719 4.564348 4.276666 4.189655 4.158883 4.317488 4.007333  
## [657] 4.304065 4.189655 4.369448 4.330733 4.204693 4.477337 4.276666 4.276666  
## [665] 4.276666 4.465908 4.356709 4.304065 4.304065 4.204693 4.330733 4.510860  
## [673] 4.369448 4.174387 4.290459 4.158883 4.477337 4.330733 4.442651 4.343805  
## [681] 4.454347 4.369448 4.394449 4.430817 4.143135 4.248495 4.488636 4.418841  
## [689] 3.891820 4.262680 4.553877 4.343805 4.574711 4.382027 4.510860 4.382027  
## [697] 4.262680 4.532599 4.143135 4.499810 -Inf 4.382027 4.465908 4.276666  
## [705] 4.488636 4.418841 4.317488 4.356709 4.158883 4.418841 4.394449 4.127134  
## [713] 4.219508 4.510860 4.499810 4.532599 4.564348 4.330733 4.317488 -Inf  
## [721] 4.174387 -Inf 4.499810 -Inf 4.510860 4.430817 4.543295 4.499810  
## [729] 4.553877 4.454347 4.356709 4.477337 4.406719 4.394449 -Inf 4.110874  
## [737] 4.025352 4.304065 4.343805 4.406719 4.442651 4.204693 4.532599 4.290459  
## [745] -Inf 4.532599 -Inf 4.248495 -Inf 4.442651 -Inf -Inf  
## [753] -Inf 4.094345 4.454347 4.465908 4.127134 4.553877 4.094345 4.394449  
## [761] 4.290459 4.304065 4.394449 4.369448 4.304065 4.454347 4.127134 4.499810  
## [769] 4.290459 4.510860 4.276666 4.174387 4.465908 4.330733 3.850148 4.290459  
## [777] 4.174387 4.356709 4.330733 4.330733 4.356709 4.488636 4.143135 4.234107  
## [785] 4.330733 4.418841 4.382027 4.406719 4.394449 4.143135 4.110874 4.143135  
## [793] 4.418841 4.189655 4.262680 3.850148 4.465908 4.382027 4.110874 4.430817  
## [801] 4.219508 4.234107 4.290459 4.276666 4.343805 4.394449 4.430817 4.465908  
## [809] 4.369448 4.110874 4.234107 4.189655 4.290459 4.304065 4.543295 4.564348  
## [817] 4.330733 4.330733 4.204693 4.276666 3.784190 4.442651 4.127134 4.356709  
## [825] 4.406719 4.330733 4.007333 4.248495 4.276666 4.158883 4.356709 4.465908  
## [833] 4.394449 3.951244 4.394449 4.499810 4.262680 4.290459 3.970292 4.418841  
## [841] 4.248495 4.317488 4.454347 3.806662 3.988984 4.465908 4.454347 4.290459  
## [849] 4.454347 4.343805 4.454347 4.290459 4.304065 4.189655 4.394449 4.418841  
## [857] 4.290459 4.564348 4.060443 4.343805 -Inf 4.330733 -Inf 4.430817  
## [865] 4.290459 -Inf 4.369448 4.532599 4.510860 -Inf 4.369448 4.204693  
## [873] 4.304065 4.442651 4.248495 -Inf 4.262680 4.394449 -Inf 4.564348  
## [881] 4.454347 4.454347 4.343805 4.430817 4.418841 -Inf 4.521789 4.276666  
## [889] 4.442651 -Inf 4.276666 4.219508 4.454347 4.248495 4.510860 4.219508  
## [897] 4.248495 4.330733 4.499810 -Inf 4.317488 4.584967 4.317488 4.143135  
## [905] 4.382027 4.382027 -Inf 4.442651 -Inf -Inf 4.094345 4.110874  
## [913] 3.931826 4.189655 4.418841 4.356709 4.330733 4.356709 4.574711 4.488636  
## [921] 4.418841 4.317488 4.158883 4.442651 -Inf 4.442651 4.204693 4.174387  
## [929] 4.418841 4.158883 3.951244 4.330733 4.330733 3.951244 4.025352 4.189655  
## [937] 4.204693 4.394449 4.248495 4.189655 4.406719 4.219508 4.454347 3.828641  
## [945] 4.219508 4.465908 4.477337 3.401197 4.330733 4.219508 4.189655 4.382027  
## [953] 4.304065 4.382027 4.477337 3.970292 4.418841 4.304065 4.499810 4.262680  
## [961] 4.369448 4.219508 4.330733 4.394449 4.356709 4.276666 4.262680 4.290459  
## [969] 4.343805 4.189655 4.276666 4.248495 4.454347 4.174387 4.418841 4.219508  
## [977] 4.290459 4.369448 3.332205 4.276666 4.234107 4.499810 4.276666 4.060443  
## [985] 4.488636 4.564348 4.442651 4.276666 4.234107 4.248495 4.499810 4.304065  
## [993] 4.234107 4.219508 4.356709 4.330733 4.465908 4.343805 4.465908 -Inf

#histogram of a quantitative variable

hist(IMDB\_dataset$rating , main = "Histogram of rating", xlab = "rating", col = "red")



#scatterplot of the quantitative variables

plot(IMDB\_dataset$release\_year , IMDB\_dataset$gross , main = "Scatterplot of release\_year vs. gross", xlab = "release\_year", ylab = "gross", col = "blue")

## Warning in xy.coords(x, y, xlabel, ylabel, log): NAs introduced by coercion  
  
## Warning in xy.coords(x, y, xlabel, ylabel, log): NAs introduced by coercion

