# MGSC661 Midterm

2023-10-20

## Read in CSV

```
movie = read.csv("IMDB_data_fall_2023.csv")
attach(movie)
```

```
summary(movie)
```

```
##  movie_title           movie_id      imdb_link           imdb_score
##  Length:1930        Min.   :    2   Length:1930        Min.   :1.900
##  Class :character   1st Qu.: 2528   Class :character   1st Qu.:5.900
##  Mode  :character   Median : 5802   Mode  :character   Median :6.600
##                     Mean   : 7067                      Mean   :6.512
##                     3rd Qu.:10604                      3rd Qu.:7.300
##                     Max.   :21838                      Max.   :9.300
##   movie_budget       release_day    release_month       release_year
##  Min.   :  560000   Min.   : 1.00   Length:1930        Min.   :1936
##  1st Qu.: 8725000   1st Qu.: 9.00   Class :character   1st Qu.:1997
##  Median :18000000   Median :17.00   Mode  :character   Median :2004
##  Mean   :20973774   Mean   :15.95                      Mean   :2001
##  3rd Qu.:30000000   3rd Qu.:23.00                      3rd Qu.:2010
##  Max.   :55000000   Max.   :30.00                      Max.   :2018
##     duration       language           country          maturity_rating
##  Min.   : 37.0   Length:1930        Length:1930        Length:1930
##  1st Qu.: 96.0   Class :character   Class :character   Class :character
##  Median :106.0   Mode  :character   Mode  :character   Mode  :character
##  Mean   :109.7
##  3rd Qu.:118.0
##  Max.   :330.0
##   aspect_ratio   distributor        nb_news_articles    director
##  Min.   :1.180   Length:1930        Min.   :    0.0   Length:1930
##  1st Qu.:1.850   Class :character   1st Qu.:   78.0   Class :character
##  Median :2.350   Mode  :character   Median :  286.0   Mode  :character
##  Mean   :2.096                      Mean   :  770.6
##  3rd Qu.:2.350                      3rd Qu.:  845.5
##  Max.   :2.760                      Max.   :60620.0
##     actor1          actor1_star_meter    actor2          actor2_star_meter
##  Length:1930        Min.   :      9   Length:1930        Min.   :      3
##  Class :character   1st Qu.:    505   Class :character   1st Qu.:   1895
##  Mode  :character   Median :   1888   Mode  :character   Median :   3986
##                     Mean   :  21190                      Mean   :  17114
##                     3rd Qu.:   4665                      3rd Qu.:   7667
##                     Max.   :8342201                      Max.   :5529461
```

```
##     actor3          actor3_star_meter colour_film          genres
##  Length:1930        Min.   :      8   Length:1930        Length:1930
##  Class :character   1st Qu.:   3075   Class :character   Class :character
##  Mode  :character   Median :   5856   Mode  :character   Mode  :character
##                     Mean   :  35469
##                     3rd Qu.:  12250
##                     Max.   :6292982
##     nb_faces     plot_keywords          action         adventure
##  Min.   : 0.00   Length:1930        Min.   :0.0000   Min.   :0.0000
##  1st Qu.: 0.00   Class :character   1st Qu.:0.0000   1st Qu.:0.0000
##  Median : 1.00   Mode  :character   Median :0.0000   Median :0.0000
##  Mean   : 1.44                      Mean   :0.2005   Mean   :0.1264
##  3rd Qu.: 2.00                      3rd Qu.:0.0000   3rd Qu.:0.0000
##  Max.   :31.00                      Max.   :1.0000   Max.   :1.0000
##      scifi           thriller         musical           romance
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.00000   1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.00000   Median :0.0000
##  Mean   :0.1083   Mean   :0.2979   Mean   :0.07047   Mean   :0.2451
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:0.00000   3rd Qu.:0.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000   Max.   :1.0000
##     western            sport             horror          drama
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.000   Min.   :0.0000
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.000   1st Qu.:0.0000
##  Median :0.00000   Median :0.00000   Median :0.000   Median :1.0000
##  Mean   :0.01762   Mean   :0.04819   Mean   :0.113   Mean   :0.5492
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.000   3rd Qu.:1.0000
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.000   Max.   :1.0000
##       war            animation           crime        movie_meter_IMDBpro
##  Min.   :0.00000   Min.   :0.00000   Min.   :0.0000   Min.   :    71
##  1st Qu.:0.00000   1st Qu.:0.00000   1st Qu.:0.0000   1st Qu.:  2836
##  Median :0.00000   Median :0.00000   Median :0.0000   Median :  5406
##  Mean   :0.03627   Mean   :0.01036   Mean   :0.2161   Mean   : 11612
##  3rd Qu.:0.00000   3rd Qu.:0.00000   3rd Qu.:0.0000   3rd Qu.: 10198
##  Max.   :1.00000   Max.   :1.00000   Max.   :1.0000   Max.   :849550
##  cinematographer    production_company
##  Length:1930        Length:1930
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
##
##
```

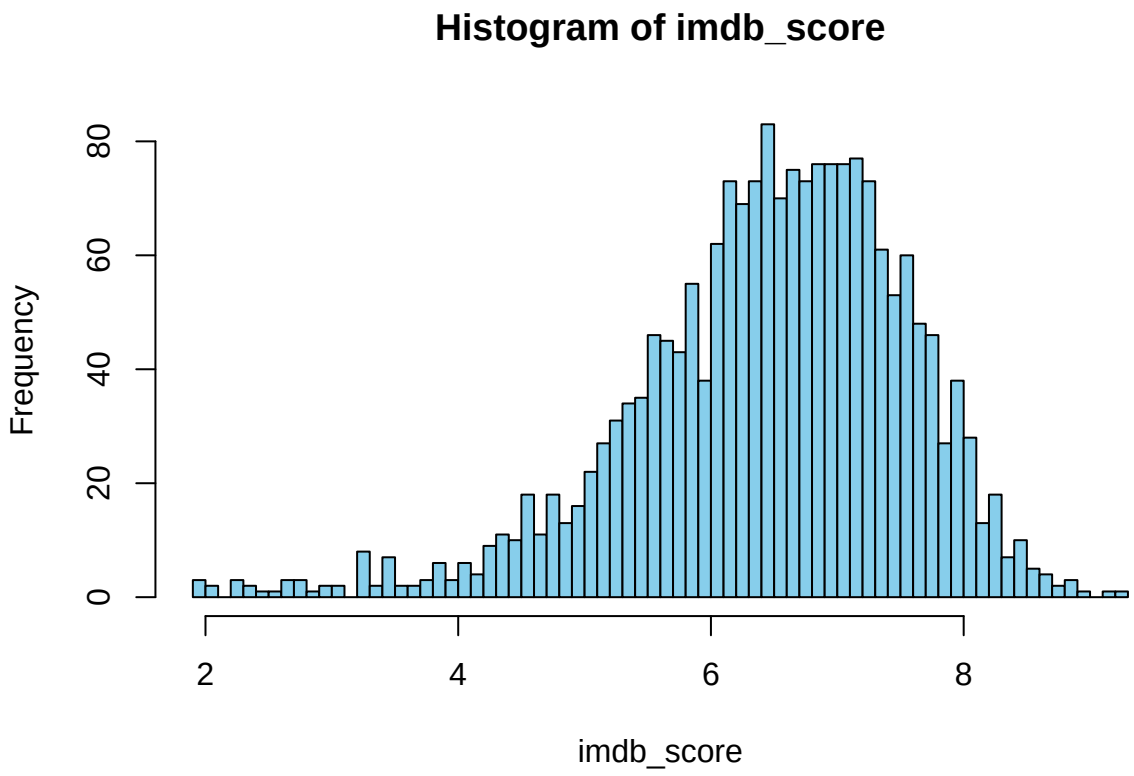
```
summary(movie$imdb_score)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.900   5.900   6.600   6.512   7.300   9.300
```

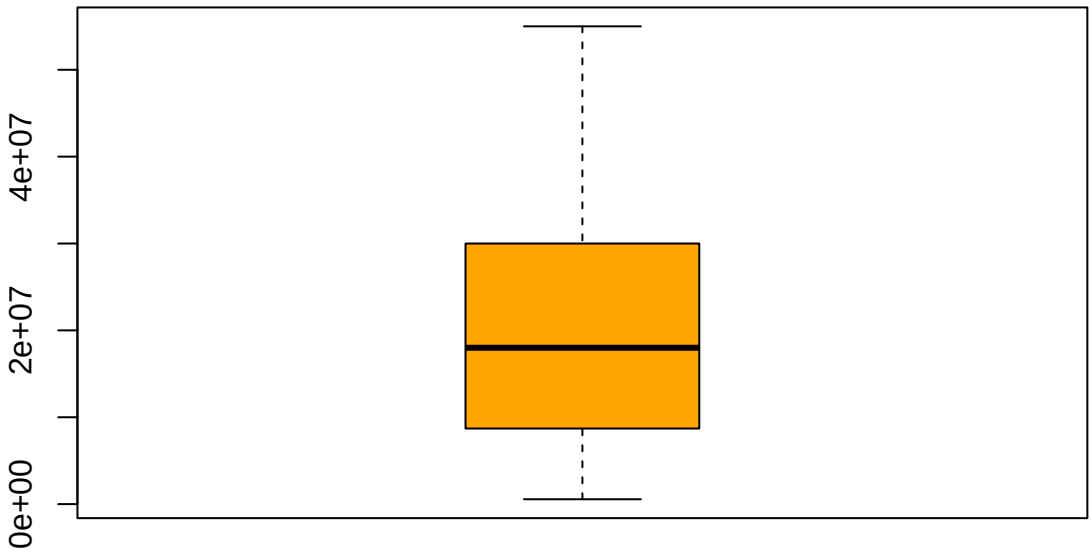
## Data Exploration - Plots

```r
library(ggplot2)
require(methods)
```

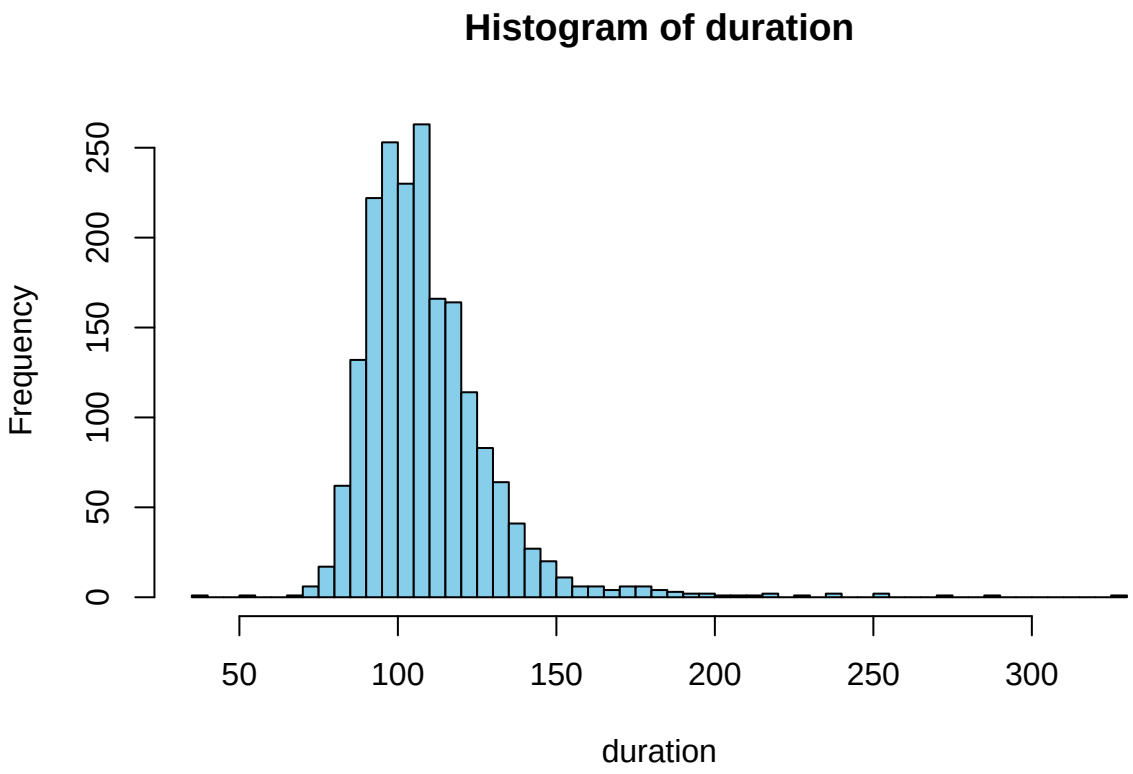
```r
hist(imdb_score, breaks = 100, col = "skyblue")
```



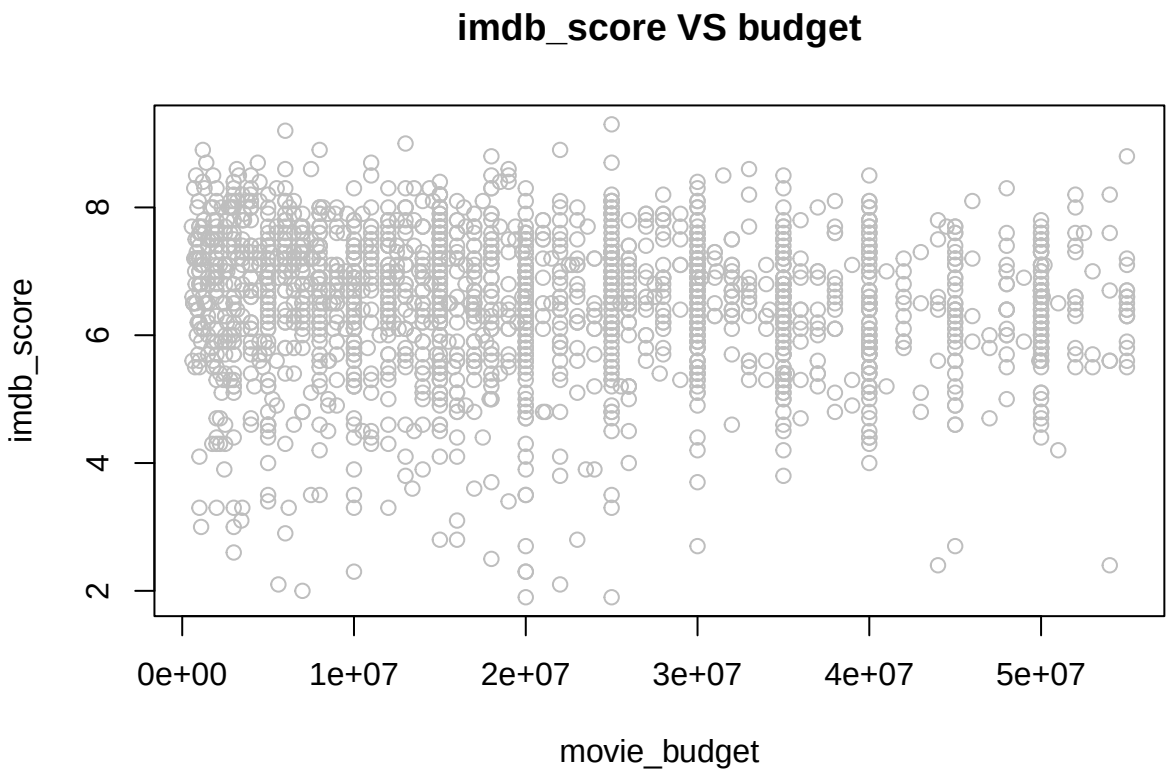

```r
boxplot(movie_budget, col = "orange", boxwex = 0.5)
title("Boxplot for movie_budget Variable")
```



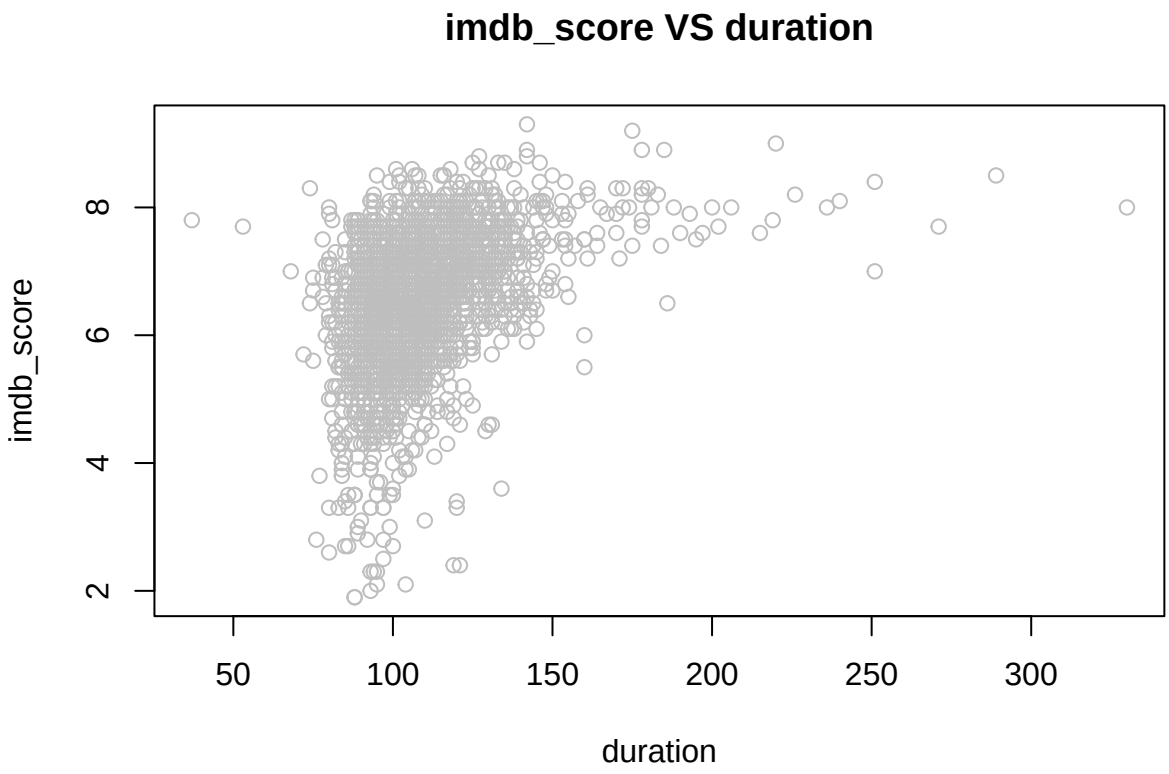

```
hist(duration, breaks = 100, col = "skyblue")
```



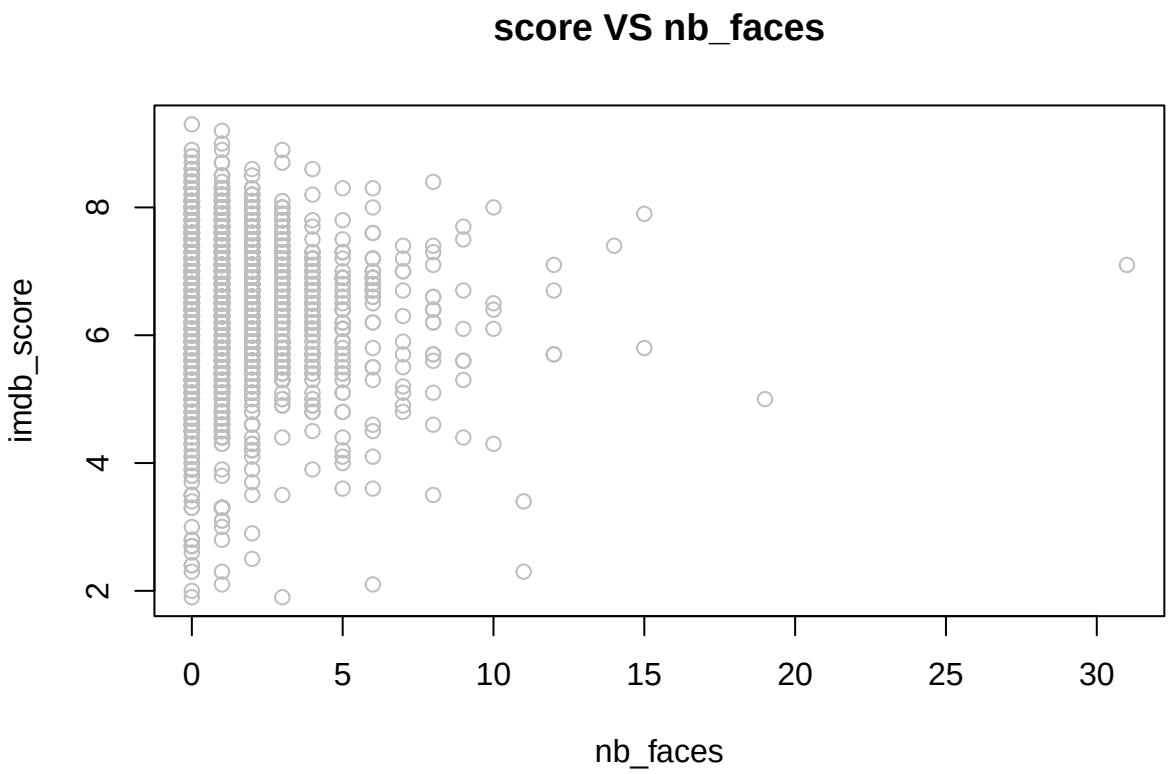

```
plot(movie$movie_budget, movie$imdb_score, col = "grey", main = "imdb_score VS budget", xlab = "movie_bu
```



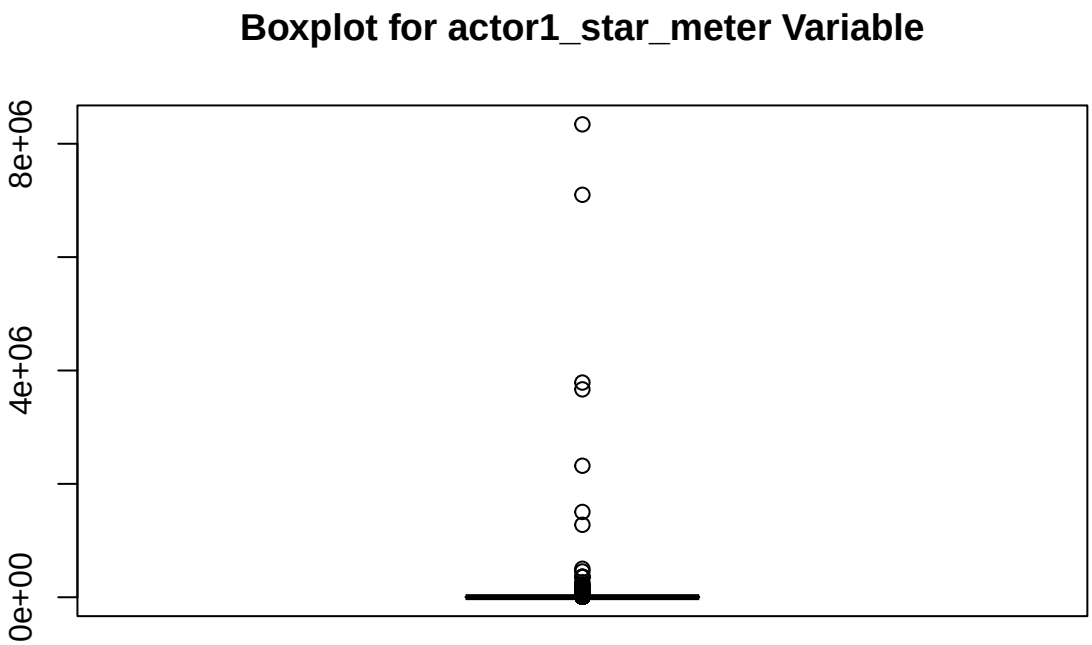

```
plot(movie$duration, movie$imdb_score, col = "grey", main = "imdb_score VS duration", xlab = "duration"
```




```
plot(movie$nb_faces, movie$imdb_score, col = "grey", main = "score VS nb_faces", xlab = "nb_faces", ylab
```

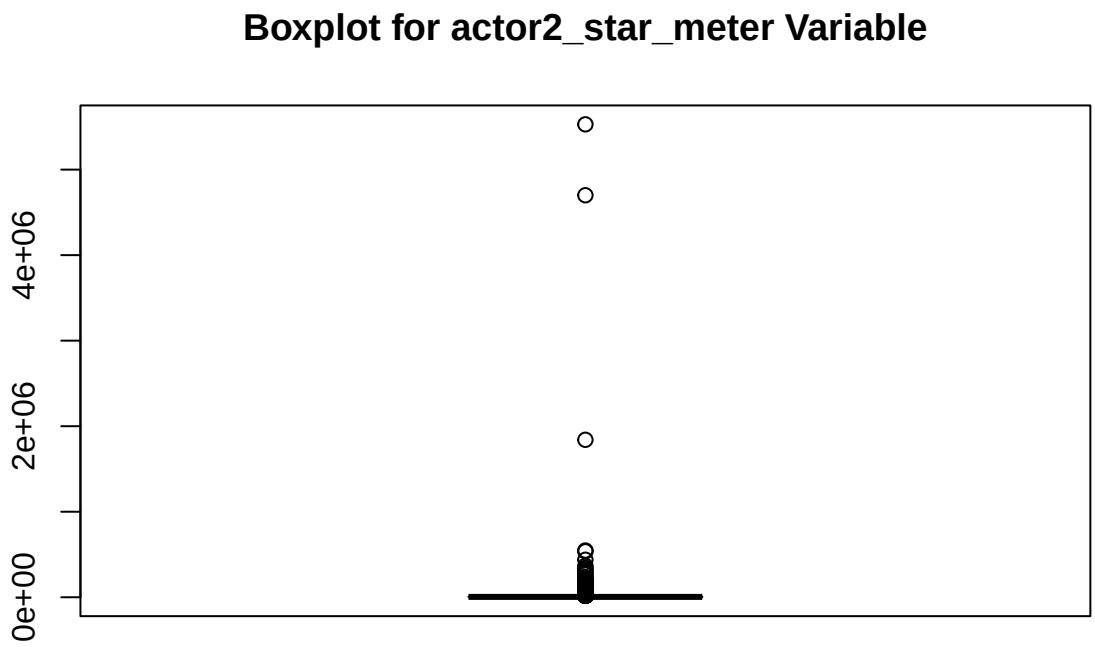



```
boxplot(actor1_star_meter, col = "orange", boxwex = 0.5)
title("Boxplot for actor1_star_meter Variable")
```

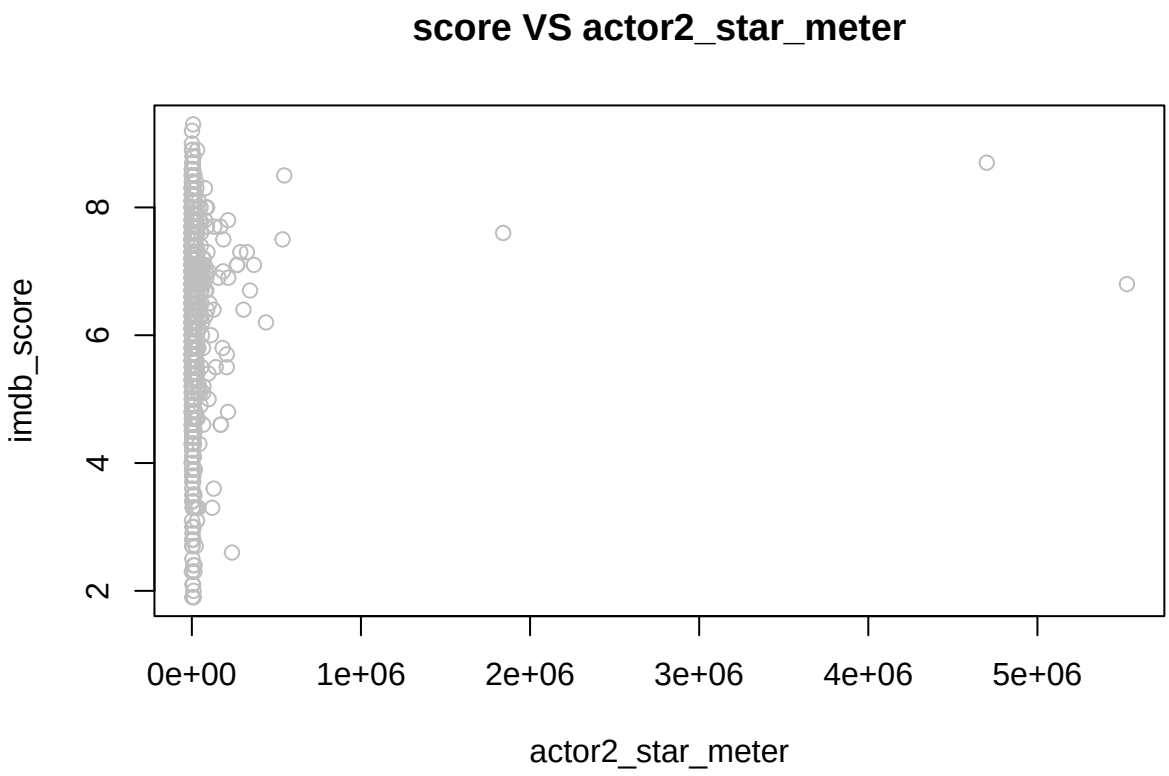



```
plot(movie$actor1_star_meter, movie$imdb_score, col = "grey", main = "score VS actor1_star_meter", xlab
```

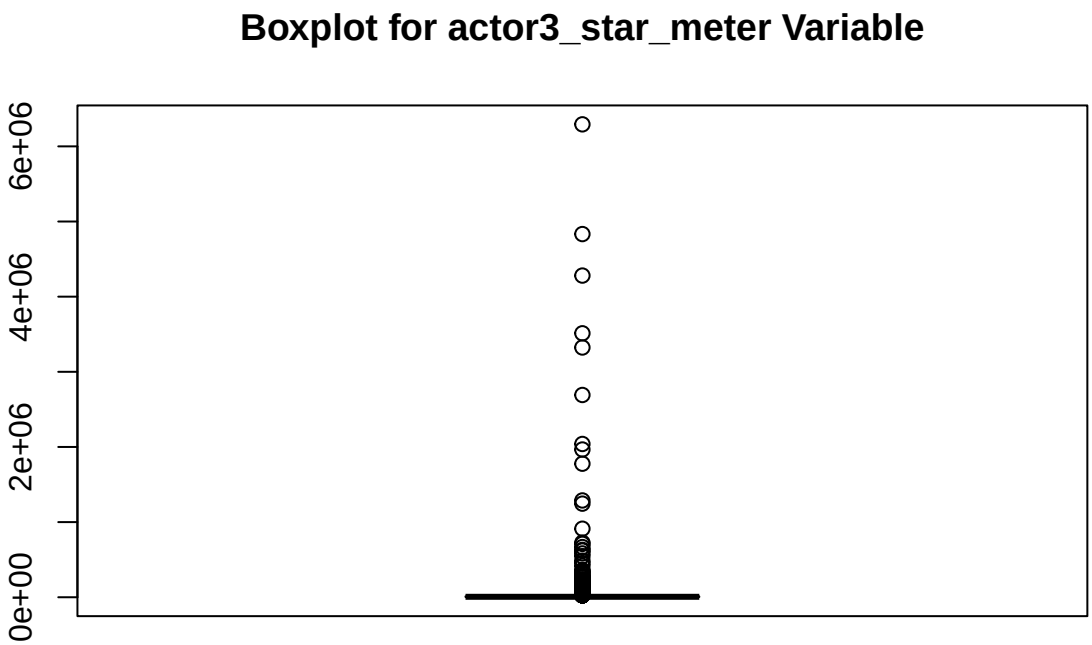

```
boxplot(actor2_star_meter, col = "orange", boxwex = 0.5)
title("Boxplot for actor2_star_meter Variable")
```



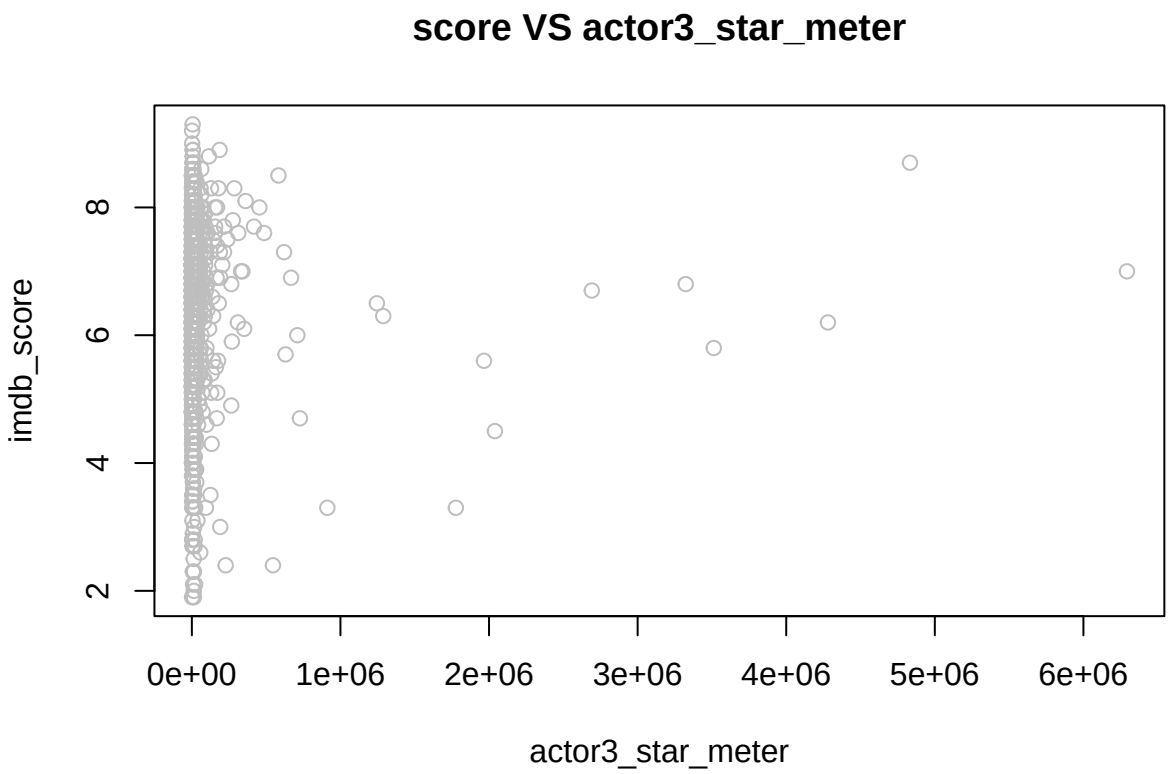

```
plot(movie$actor2_star_meter, movie$imdb_score, col = "grey", main = "score VS actor2_star_meter", xlab
```



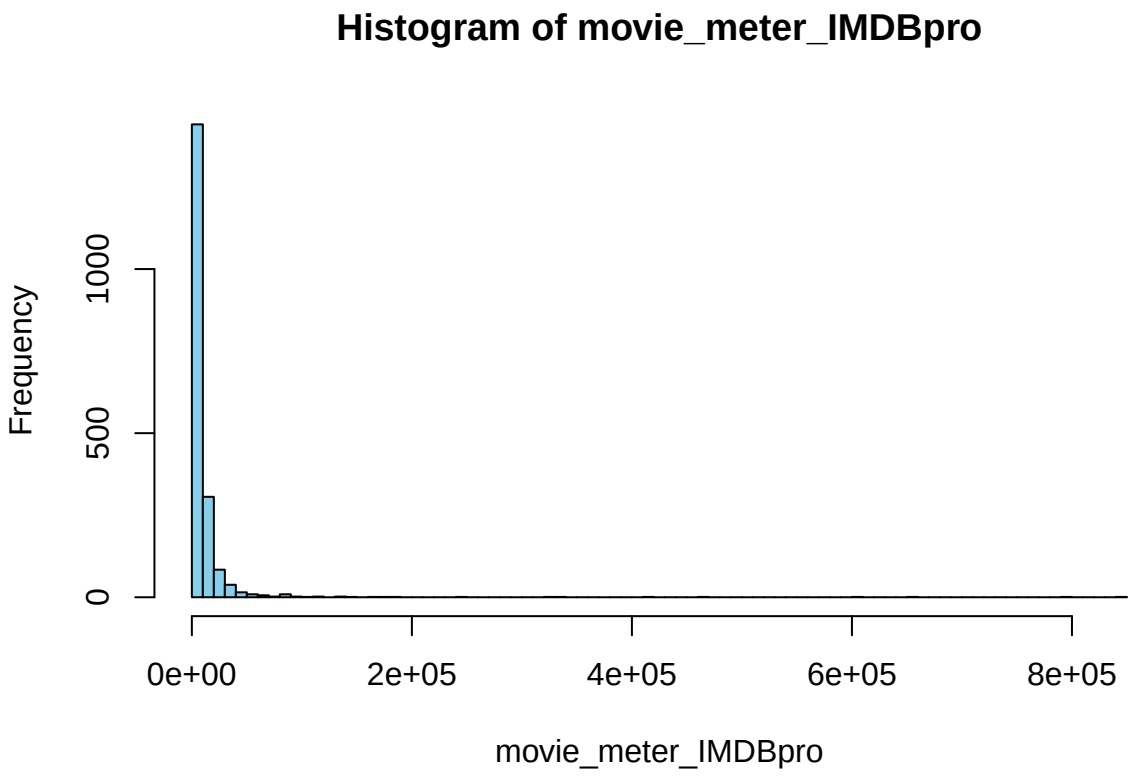

```
boxplot(actor3_star_meter, col = "orange", boxwex = 0.5)
title("Boxplot for actor3_star_meter Variable")
```



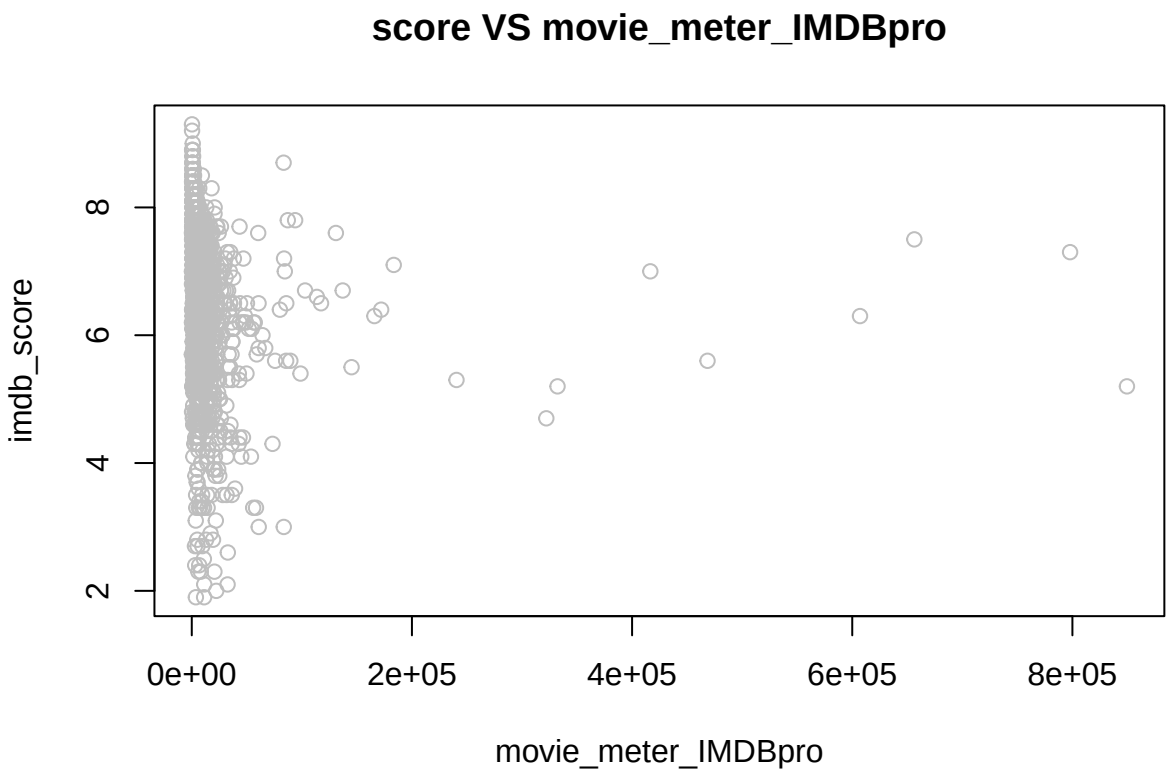

```
plot(movie$actor3_star_meter, movie$imdb_score, col = "grey", main = "score VS actor3_star_meter", xlab
```




```r
hist(movie_meter_IMDBpro, breaks = 100, col = "skyblue")
```




```
plot(movie$movie_meter_IMDBpro, movie$imdb_score, col = "grey", main = "score VS movie_meter_IMDBpro", 
```



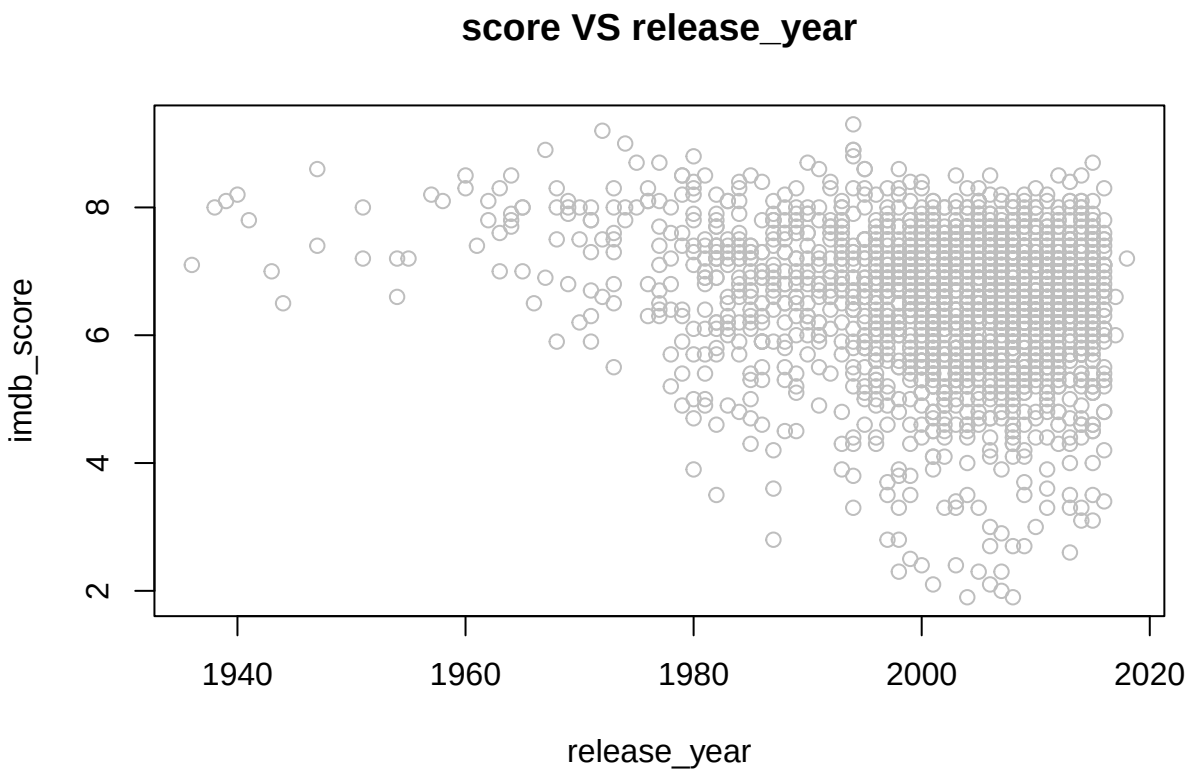

```
summary(movie$movie_meter_IMDBpro)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      71    2836    5406   11612   10198  849550
```

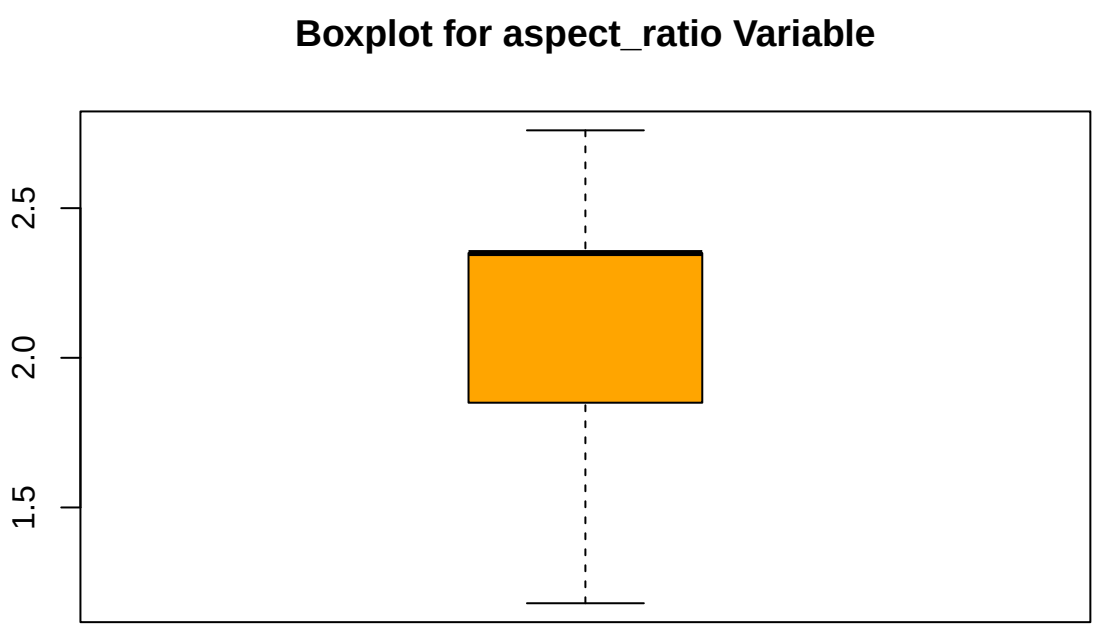
```
summary(movie$release_year)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1936    1997    2004    2001    2010    2018
```

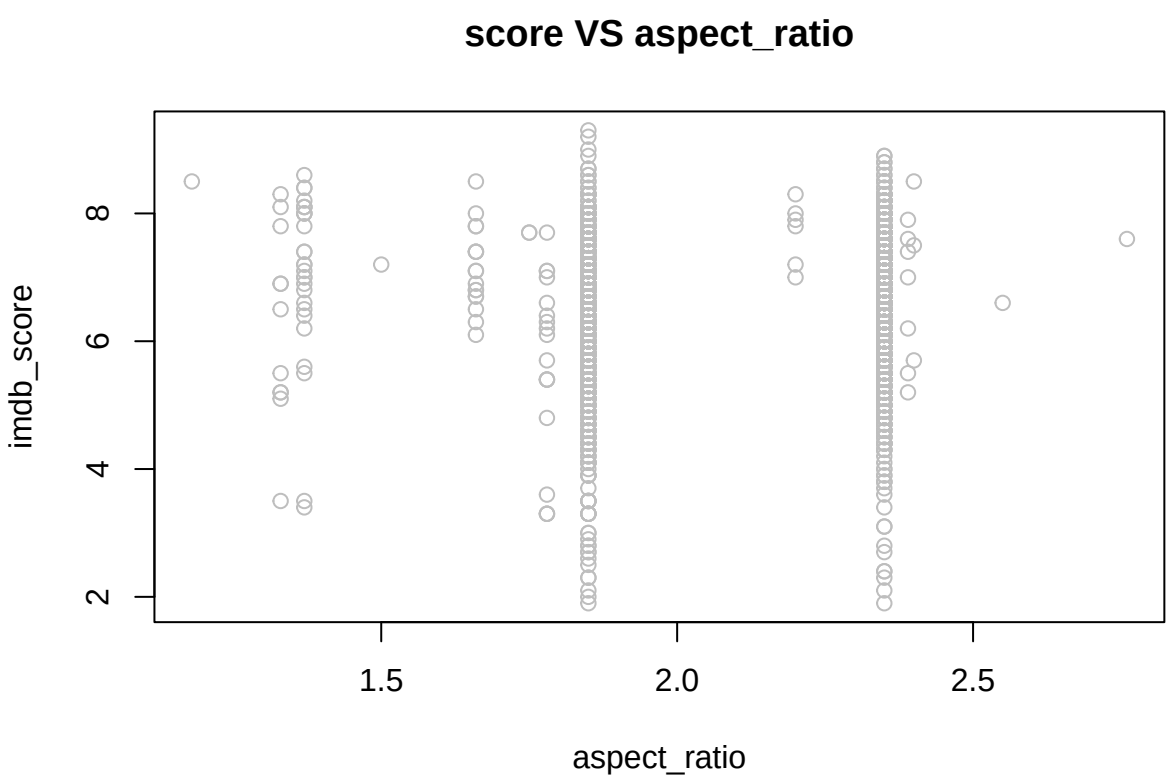
```
plot(movie$release_year, movie$imdb_score, col = "grey", main = "score VS release_year", xlab = "releas
```




```
boxplot(aspect_ratio, col = "orange", boxwex = 0.5)
title("Boxplot for aspect_ratio Variable")
```




```
plot(movie$aspect_ratio, movie$imdb_score, col = "grey", main = "score VS aspect_ratio", xlab = "aspect_
```


score VS aspect_ratio
imdb_score
2
4
6
8
1.5
2.0
2.5
aspect_ratio