Bayesian modeling and prediction for movies

Setup

Load packages

```
library(ggplot2)
library(dplyr)
```

```
## Warning: Installed Rcpp (0.12.10) different from Rcpp used to build dplyr (0.12.11
).
## Please reinstall dplyr to avoid random crashes or undefined behavior.
```

```
library(statsr)
library(BAS)

getwd()
```

```
## [1] "C:/Users/HP/Documents"
```

Load data

```
#load("movies.Rdata")
load("movies.Rdata")
movies<-data.frame(movies)
```

Part 1: Data

The dataset consists of variables pertaining to audience and critic review scores from Rotten Tomatoes and IMDB web sites for 651 movies released to theaters during 1970-2014.

Though Rotten Tomatoes and IMDB presumably uses randomly sampled data, no solid information was available regarding the specific sampling method used. This may impede the ability of the conclusions to be generalized to the population of all movies.

A possible selection bias may arise with regard to the audience if only the theater visiting audience is included and DVD-watching and online movie watching participants are excluded. Recall bias may affect the responses if there were a considerable time gap between the movie watched and the data collection.

Moreover, a sample size of 651 may be considered modest but perhaps not adequate to unearth subtle patterns underlying in the populations which could be elicited through big data analytics and data mining with a bigger sample. This may nevertheless be adequate for classical cross-sectional analyses.

This is an observational study, more specifically, a cross-sectional survey and therefore no causal relationships can be inferred or assumed from the conclusions drawn because temporality of associations cannot be elicited as well as confounding and bias may affect the results achieved.

```
str(movies)
```

```
## 'data.frame':    651 obs. of  32 variables:
##  $ title           : chr  "Filly Brown" "The Dish" "Waiting for Guffman" "The Age
of Innocence" ...
##  $ title_type      : Factor w/ 3 levels "Documentary",..: 2 2 2 2 2 1 2 2 1 2 ...
##  $ genre           : Factor w/ 11 levels "Action & Adventure",..: 6 6 4 6 7 5 6 6
5 6 ...
##  $ runtime         : num  80 101 84 139 90 78 142 93 88 119 ...
##  $ mpaa_rating     : Factor w/ 6 levels "G","NC-17","PG",..: 5 4 5 3 5 6 4 5 6 6 .
..
##  $ studio          : Factor w/ 211 levels "20th Century Fox",..: 91 202 167 34 13
163 147 118 88 84 ...
##  $ thtr_rel_year   : num  2013 2001 1996 1993 2004 ...
##  $ thtr_rel_month  : num  4 3 8 10 9 1 1 11 9 3 ...
##  $ thtr_rel_day    : num  19 14 21 1 10 15 1 8 7 2 ...
##  $ dvd_rel_year    : num  2013 2001 2001 2001 2005 ...
##  $ dvd_rel_month   : num  7 8 8 11 4 4 2 3 1 8 ...
##  $ dvd_rel_day     : num  30 28 21 6 19 20 18 2 21 14 ...
##  $ imdb_rating     : num  5.5 7.3 7.6 7.2 5.1 7.8 7.2 5.5 7.5 6.6 ...
##  $ imdb_num_votes  : int  899 12285 22381 35096 2386 333 5016 2272 880 12496 ...
##  $ critics_rating  : Factor w/ 3 levels "Certified Fresh",..: 3 1 1 1 3 2 3 3 2 1
...
##  $ critics_score   : num  45 96 91 80 33 91 57 17 90 83 ...
##  $ audience_rating : Factor w/ 2 levels "Spilled","Upright": 2 2 2 2 1 2 2 1 2 2 .
..
##  $ audience_score  : num  73 81 91 76 27 86 76 47 89 66 ...
##  $ best_pic_nom    : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ best_pic_win    : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ best_actor_win  : Factor w/ 2 levels "no","yes": 1 1 1 2 1 1 1 2 1 1 ...
##  $ best_actress_win: Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ best_dir_win    : Factor w/ 2 levels "no","yes": 1 1 1 2 1 1 1 1 1 1 ...
##  $ top200_box      : Factor w/ 2 levels "no","yes": 1 1 1 1 1 1 1 1 1 1 ...
##  $ director        : chr  "Michael D. Olmos" "Rob Sitch" "Christopher Guest" "Mart
in Scorsese" ...
##  $ actor1          : chr  "Gina Rodriguez" "Sam Neill" "Christopher Guest" "Daniel
Day-Lewis" ...
##  $ actor2          : chr  "Jenni Rivera" "Kevin Harrington" "Catherine O'Hara" "Mi
chelle Pfeiffer" ...
##  $ actor3          : chr  "Lou Diamond Phillips" "Patrick Warburton" "Parker Posey
" "Winona Ryder" ...
```

```
##  $ actor4          : chr  "Emilio Rivera" "Tom Long" "Eugene Levy" "Richard E. Gra
nt" ...
##  $ actor5          : chr  "Joseph Julian Soria" "Genevieve Mooy" "Bob Balaban" "Al
ec McCowen" ...
##  $ imdb_url        : chr  "http://www.imdb.com/title/tt1869425/" "http://www.imdb.
com/title/tt0205873/" "http://www.imdb.com/title/tt0118111/" "http://www.imdb.com/tit
le/tt0106226/" ...
##  $ rt_url          : chr  "//www.rottentomatoes.com/m/filly_brown_2012/" "//www.ro
ttentomatoes.com/m/dish/" "//www.rottentomatoes.com/m/waiting_for_guffman/" "//www.ro
ttentomatoes.com/m/age_of_innocence/" ...
```

```
head(movies)
```

```
##                   title   title_type       genre runtime mpaa_rating
## 1          Filly Brown Feature Film       Drama      80           R
## 2             The Dish Feature Film       Drama     101       PG-13
## 3  Waiting for Guffman Feature Film      Comedy      84           R
## 4 The Age of Innocence Feature Film       Drama     139          PG
## 5          Malevolence Feature Film      Horror      90           R
## 6          Old Partner  Documentary Documentary      78     Unrated
##                     studio thtr_rel_year thtr_rel_month thtr_rel_day
## 1      Indomina Media Inc.          2013              4           19
## 2    Warner Bros. Pictures          2001              3           14
## 3   Sony Pictures Classics          1996              8           21
## 4        Columbia Pictures          1993             10            1
## 5 Anchor Bay Entertainment          2004              9           10
## 6       Shcalo Media Group          2009              1           15
##   dvd_rel_year dvd_rel_month dvd_rel_day imdb_rating imdb_num_votes
## 1         2013             7          30         5.5            899
## 2         2001             8          28         7.3          12285
## 3         2001             8          21         7.6          22381
## 4         2001            11           6         7.2          35096
## 5         2005             4          19         5.1           2386
## 6         2010             4          20         7.8            333
##    critics_rating critics_score audience_rating audience_score
## 1          Rotten            45         Upright             73
## 2 Certified Fresh            96         Upright             81
## 3 Certified Fresh            91         Upright             91
## 4 Certified Fresh            80         Upright             76
## 5          Rotten            33         Spilled             27
## 6           Fresh            91         Upright             86
##   best_pic_nom best_pic_win best_actor_win best_actress_win best_dir_win
## 1           no           no             no               no           no
## 2           no           no             no               no           no
## 3           no           no             no               no           no
## 4           no           no            yes               no          yes
## 5           no           no             no               no           no
## 6           no           no             no               no           no
##   top200_box          director            actor1             actor2
```

```
## 1         no  Michael D. Olmos    Gina Rodriguez       Jenni Rivera
## 2         no         Rob Sitch         Sam Neill   Kevin Harrington
## 3         no Christopher Guest Christopher Guest   Catherine O'Hara
## 4         no   Martin Scorsese  Daniel Day-Lewis  Michelle Pfeiffer
## 5         no       Stevan Mena     Samantha Dark R. Brandon Johnson
## 6         no   Chung-ryoul Lee     Choi Won-kyun       Lee Sam-soon
##                 actor3           actor4              actor5
## 1 Lou Diamond Phillips    Emilio Rivera Joseph Julian Soria
## 2    Patrick Warburton         Tom Long      Genevieve Mooy
## 3         Parker Posey      Eugene Levy         Bob Balaban
## 4         Winona Ryder Richard E. Grant        Alec McCowen
## 5      Brandon Johnson    Heather Magee      Richard Glover
## 6                  Moo             <NA>               <NA>
##                              imdb_url
## 1 http://www.imdb.com/title/tt1869425/
## 2 http://www.imdb.com/title/tt0205873/
## 3 http://www.imdb.com/title/tt0118111/
## 4 http://www.imdb.com/title/tt0106226/
## 5 http://www.imdb.com/title/tt0388230/
## 6 http://www.imdb.com/title/tt1334549/
##                                          rt_url
## 1     //www.rottentomatoes.com/m/filly_brown_2012/
## 2                 //www.rottentomatoes.com/m/dish/
## 3  //www.rottentomatoes.com/m/waiting_for_guffman/
## 4     //www.rottentomatoes.com/m/age_of_innocence/
## 5 //www.rottentomatoes.com/m/10004684-malevolence/
## 6          //www.rottentomatoes.com/m/old-partner/
```

```
summary(movies)
```

```
##     title                    title_type                  genre
##  Length:651         Documentary : 55   Drama             :305
##  Class :character   Feature Film:591   Comedy            : 87
##  Mode  :character   TV Movie    :  5   Action & Adventure: 65
##                                        Mystery & Suspense: 59
##                                        Documentary       : 52
##                                        Horror            : 23
##                                        (Other)           : 60
##     runtime       mpaa_rating                                studio
##  Min.   : 39.0   G      : 19   Paramount Pictures                : 37
##  1st Qu.: 92.0   NC-17  :  2   Warner Bros. Pictures             : 30
##  Median :103.0   PG     :118   Sony Pictures Home Entertainment: 27
##  Mean   :105.8   PG-13  :133   Universal Pictures                : 23
##  3rd Qu.:115.8   R      :329   Warner Home Video                 : 19
##  Max.   :267.0   Unrated: 50   (Other)                           :507
##  NA's   :1                     NA's                              :  8
##  thtr_rel_year  thtr_rel_month   thtr_rel_day    dvd_rel_year
##  Min.   :1970   Min.   : 1.00   Min.   : 1.00   Min.   :1991
##  1st Qu.:1990   1st Qu.: 4.00   1st Qu.: 7.00   1st Qu.:2001
```

```
##  Median :2000   Median : 7.00   Median :15.00   Median :2004
##  Mean   :1998   Mean   : 6.74   Mean   :14.42   Mean   :2004
##  3rd Qu.:2007   3rd Qu.:10.00   3rd Qu.:21.00   3rd Qu.:2008
##  Max.   :2014   Max.   :12.00   Max.   :31.00   Max.   :2015
##                                                 NA's   :8
##  dvd_rel_month     dvd_rel_day     imdb_rating    imdb_num_votes
##  Min.   : 1.000   Min.   : 1.00   Min.   :1.900   Min.   :   180
##  1st Qu.: 3.000   1st Qu.: 7.00   1st Qu.:5.900   1st Qu.:  4546
##  Median : 6.000   Median :15.00   Median :6.600   Median : 15116
##  Mean   : 6.333   Mean   :15.01   Mean   :6.493   Mean   : 57533
##  3rd Qu.: 9.000   3rd Qu.:23.00   3rd Qu.:7.300   3rd Qu.: 58301
##  Max.   :12.000   Max.   :31.00   Max.   :9.000   Max.   :893008
##  NA's   :8        NA's   :8
##         critics_rating critics_score    audience_rating audience_score
##  Certified Fresh:135    Min.   :  1.00   Spilled:275     Min.   :11.00
##  Fresh          :209    1st Qu.: 33.00   Upright:376     1st Qu.:46.00
##  Rotten         :307    Median : 61.00                   Median :65.00
##                         Mean   : 57.69                   Mean   :62.36
##                         3rd Qu.: 83.00                   3rd Qu.:80.00
##                         Max.   :100.00                   Max.   :97.00
##
##  best_pic_nom best_pic_win best_actor_win best_actress_win best_dir_win
##  no :629      no :644      no :558        no :579           no :608
##  yes: 22      yes:  7      yes: 93        yes: 72           yes: 43
##
##
##
##
##
##  top200_box   director            actor1             actor2
##  no :636    Length:651         Length:651         Length:651
##  yes: 15    Class :character   Class :character   Class :character
##             Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##
##     actor3             actor4             actor5
##  Length:651         Length:651         Length:651
##  Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character
##
##
##
##
##    imdb_url           rt_url
##  Length:651         Length:651
##  Class :character   Class :character
##  Mode  :character   Mode  :character
##
```

```
##
##
##
```

Part 2: Data manipulation

Following variables were synthesized as given in the instructions for the project.

feature_film: “yes” if title_type is “Feature Film”, “no” otherwise.

drama: “yes” if genre is “Drama”, “no” otherwise.

mpaa_rating_R: “yes” if mpaa_rating is “R”, “no” otherwise.

oscar_season: “yes” if movie is released in November, October, or December (based on thtr_rel_month), “no” otherwise.

summer_season: “yes” if movie is released in May, June, July, or August (based on thtr_rel_month), “no” otherwise.

```
movies <- mutate(movies, feature_film = factor(ifelse(title_type == 'Feature Film', '
yes', 'no')))
movies <- mutate(movies, drama = factor(ifelse(genre =='Drama', 'yes', 'no')))
movies <- mutate(movies, mpaa_rating_R = factor(ifelse(mpaa_rating == 'R', 'yes', 'no
')))
movies <- mutate(movies, oscar_season = factor(ifelse(thtr_rel_month >= 10, 'yes', 'n
o')))
movies <- mutate(movies, summer_season = factor(ifelse(thtr_rel_month %in% c(5,6,7,8)
, 'yes', 'no')))

head(movies[c("feature_film","drama", "mpaa_rating_R", "oscar_season", "summer_season
")])
```

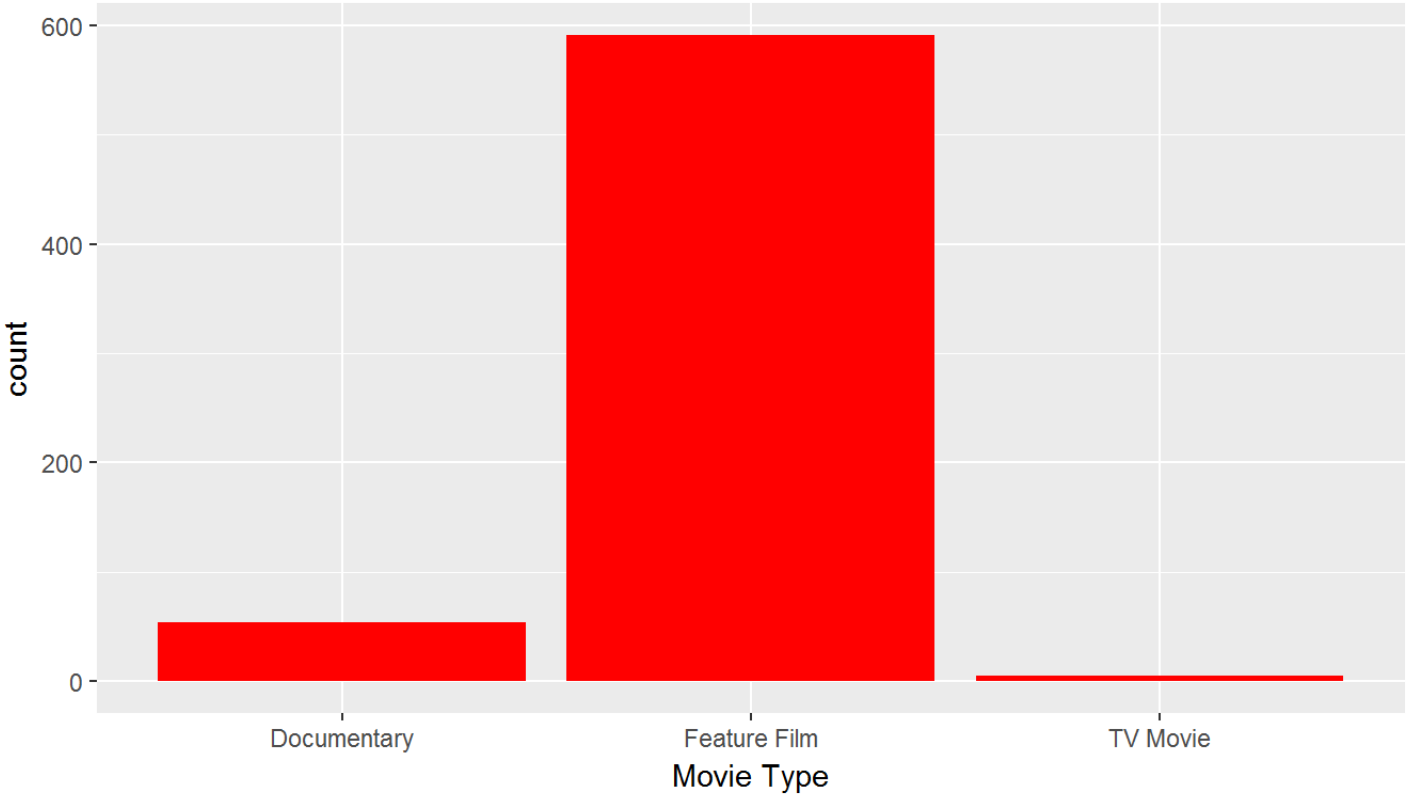
```
##   feature_film drama mpaa_rating_R oscar_season summer_season
## 1          yes   yes           yes           no            no
## 2          yes   yes            no           no            no
## 3          yes    no           yes           no           yes
## 4          yes   yes            no          yes            no
## 5          yes    no           yes           no            no
## 6           no    no            no           no            no
```

Calling the summary command showed that there is a single missing data point in the runtime variable.

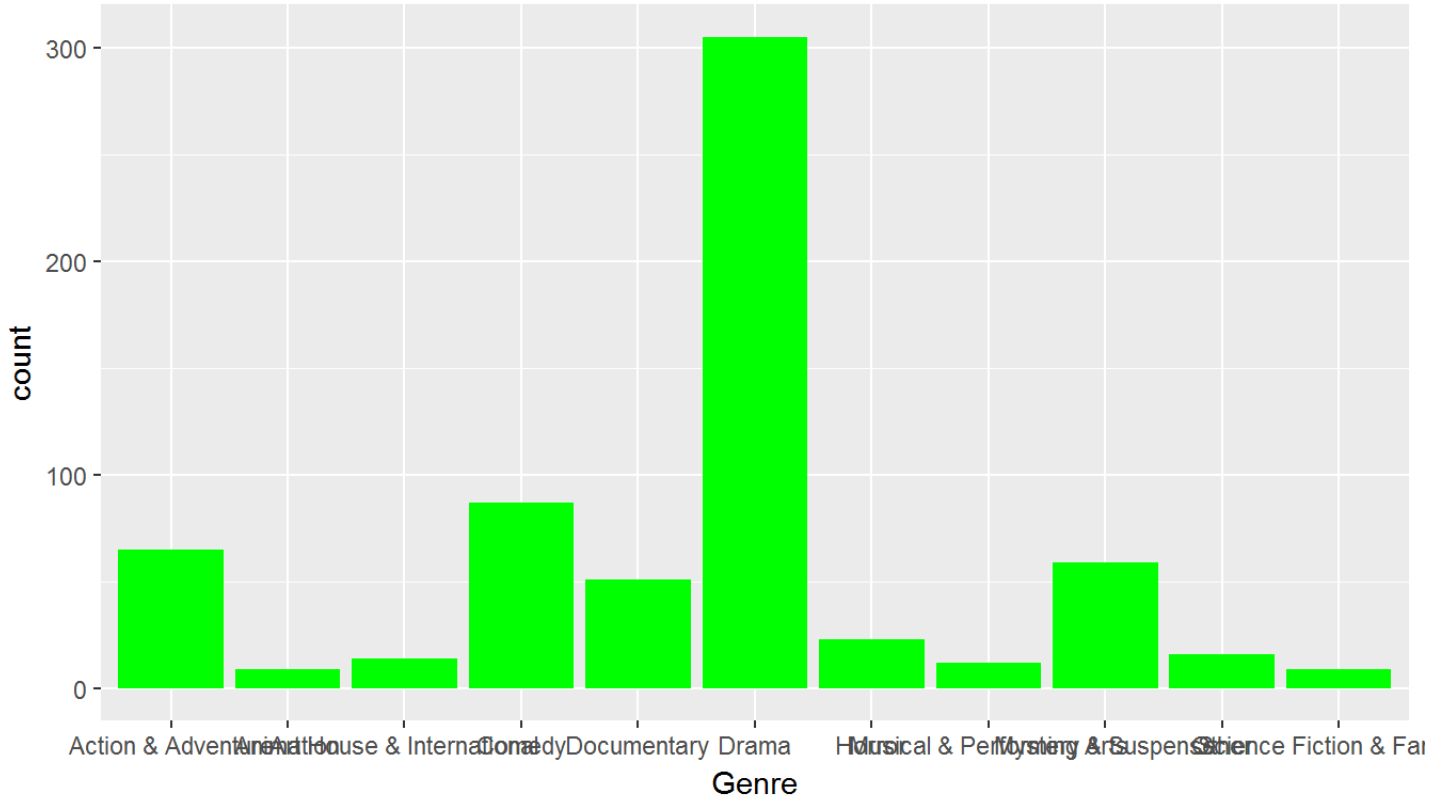
```
# Deleting the single missing data point in the runtime variable.
movies <- filter(movies, !is.na(runtime))
```

Part 3: Exploratory data analysis

```
#histograms and barplots
ggplot(data=movies, aes(x=title_type)) +
      geom_bar(fill="red") +
      xlab("Movie Type")
```

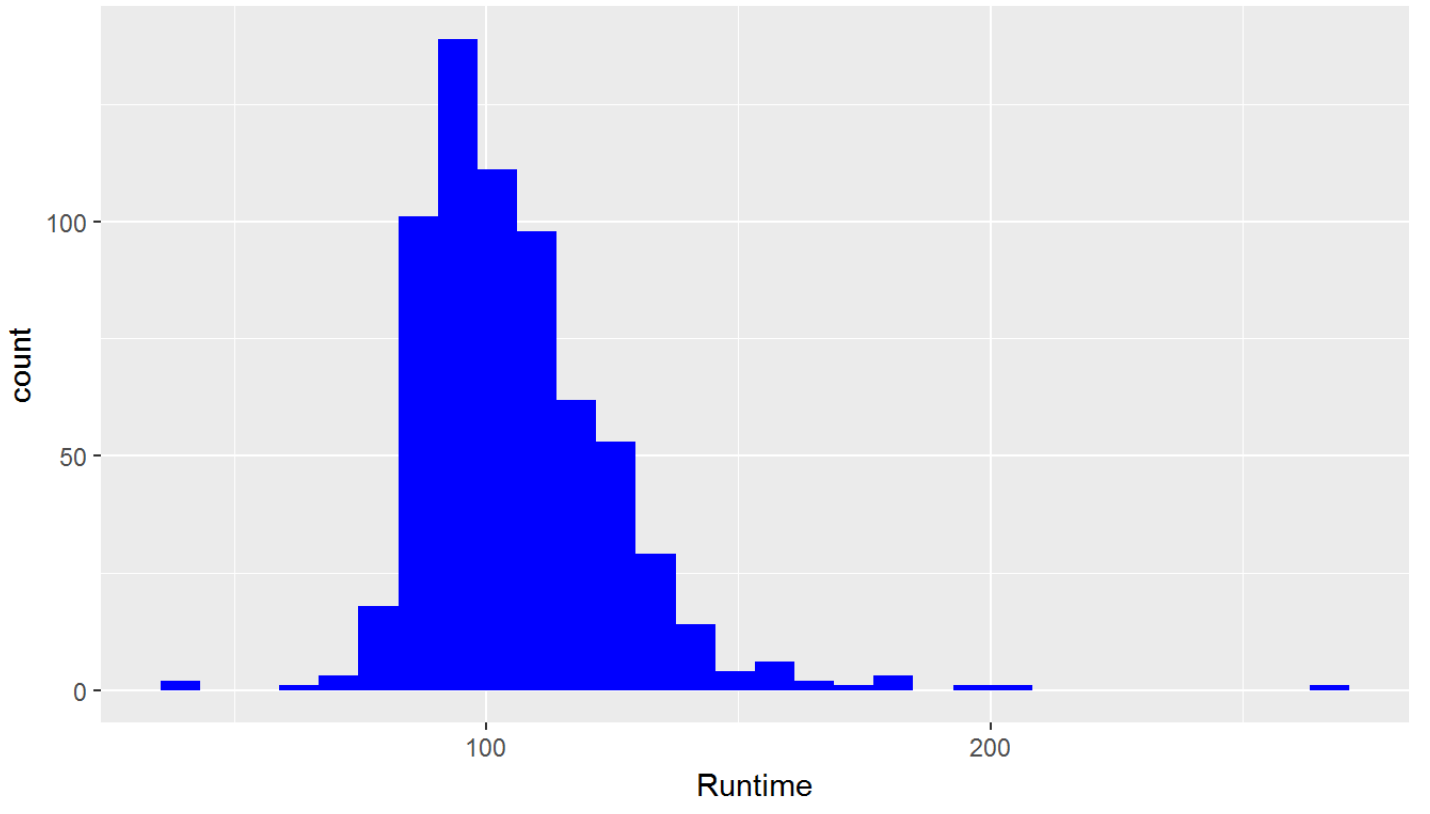



```
ggplot(data=movies, aes(x=genre)) +
      geom_bar(fill="green") +
      xlab("Genre")
```

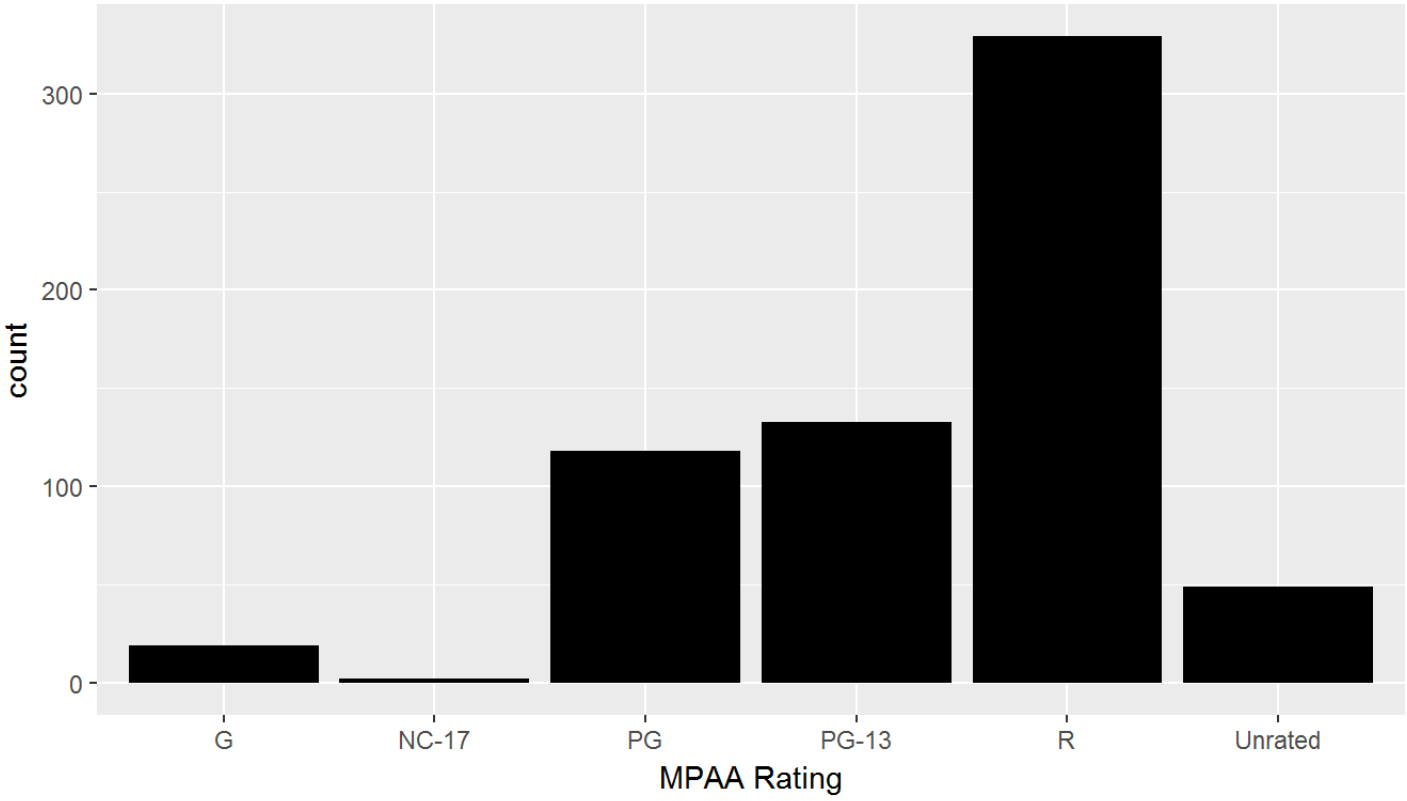



```
ggplot(data=movies, aes(x=runtime)) +
      geom_histogram(fill="blue") +
      xlab("Runtime")
```

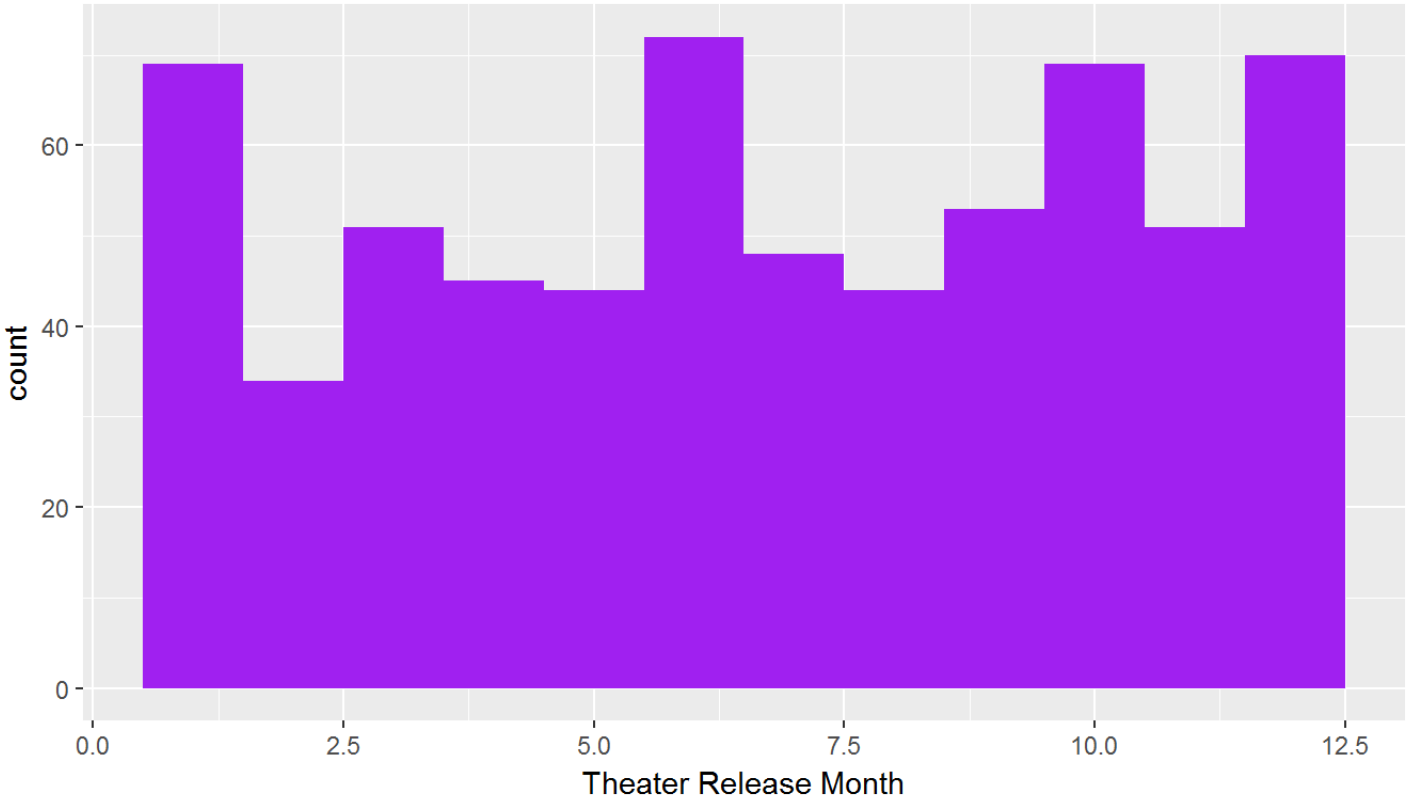
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



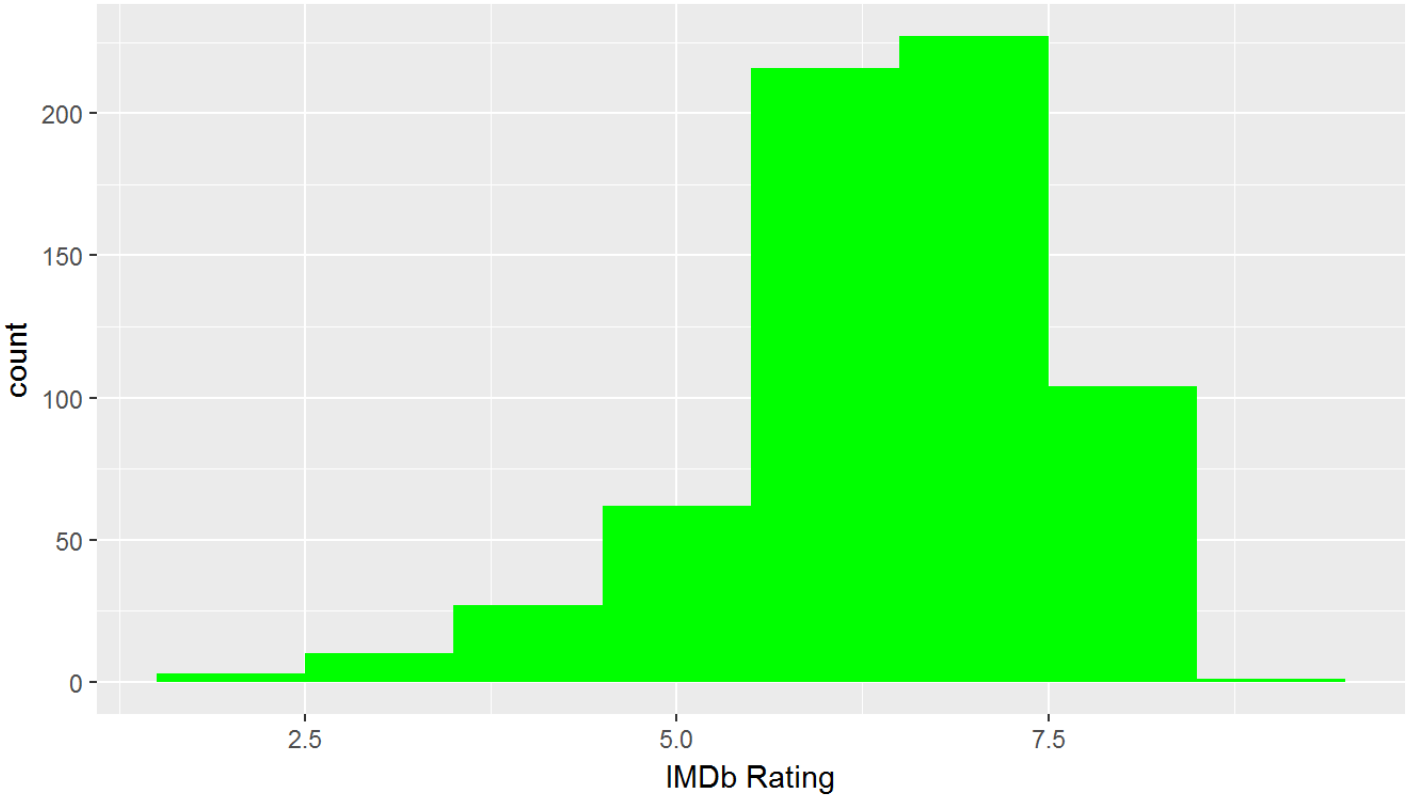
```
ggplot(data=movies, aes(x=mpaa_rating)) +
      geom_bar(fill="black") +
      xlab("MPAA Rating")
```



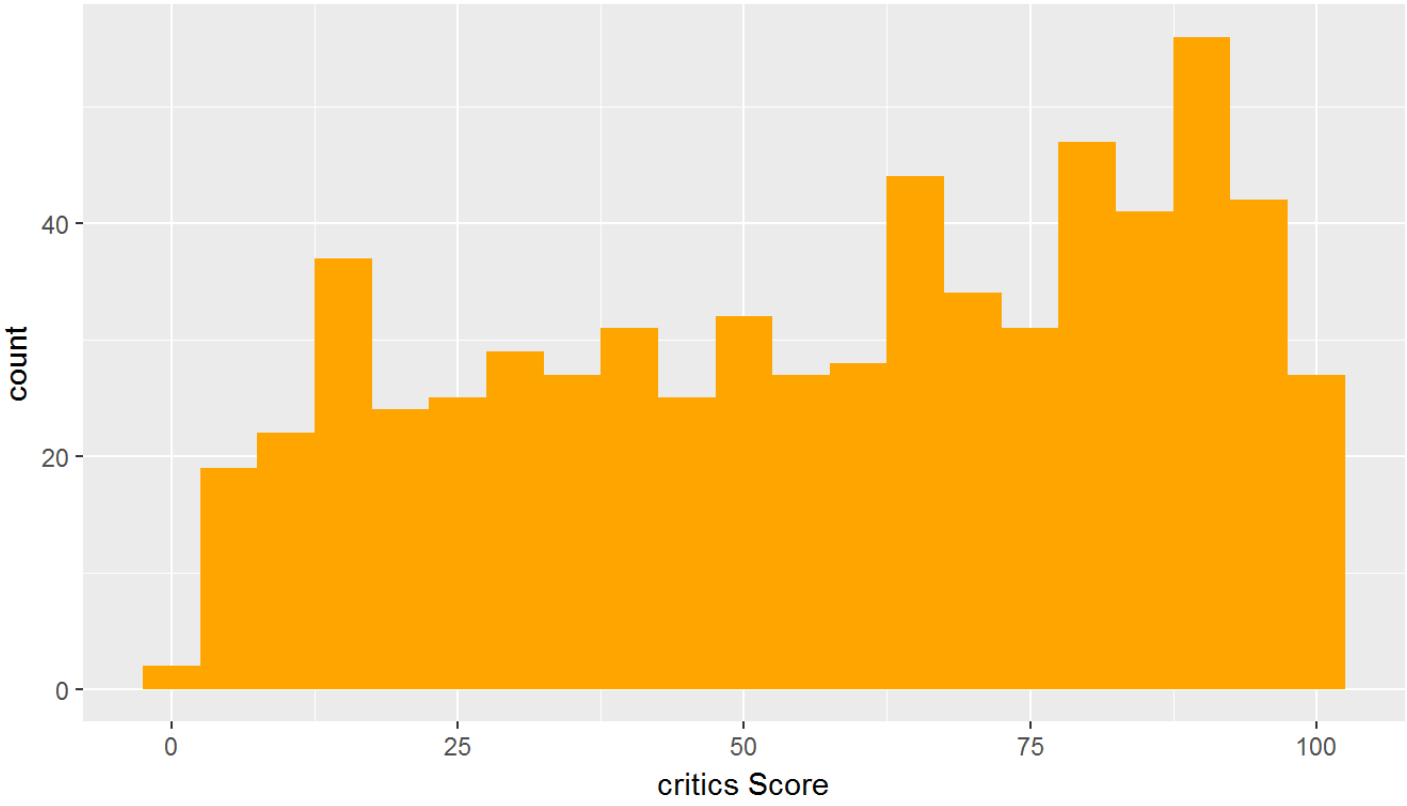

```
ggplot(data=movies, aes(x=thtr_rel_month)) +
      geom_histogram(binwidth=1, fill="purple") +
  xlab("Theater Release Month")
```



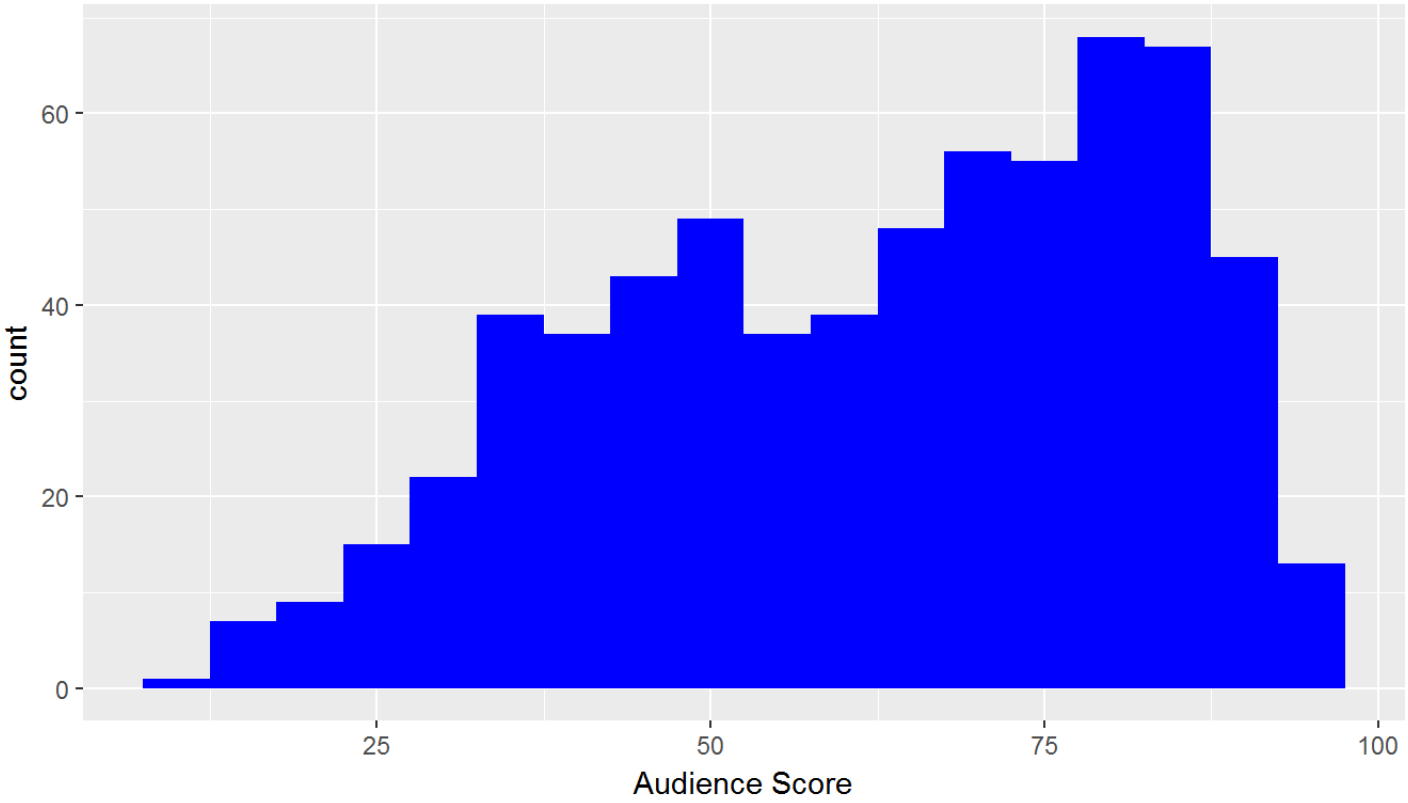

```
ggplot(data=movies, aes(x=imdb_rating)) +
      geom_histogram(binwidth=1, fill="green") +
  xlab("IMDb Rating")
```



```
ggplot(data=movies, aes(x=critics_score)) +
      geom_histogram(binwidth=5, fill="orange") +
  xlab("critics Score")
```




```
ggplot(data=movies, aes(x=audience_score)) +
      geom_histogram(binwidth=5, fill="blue") +
  xlab("Audience Score")
```

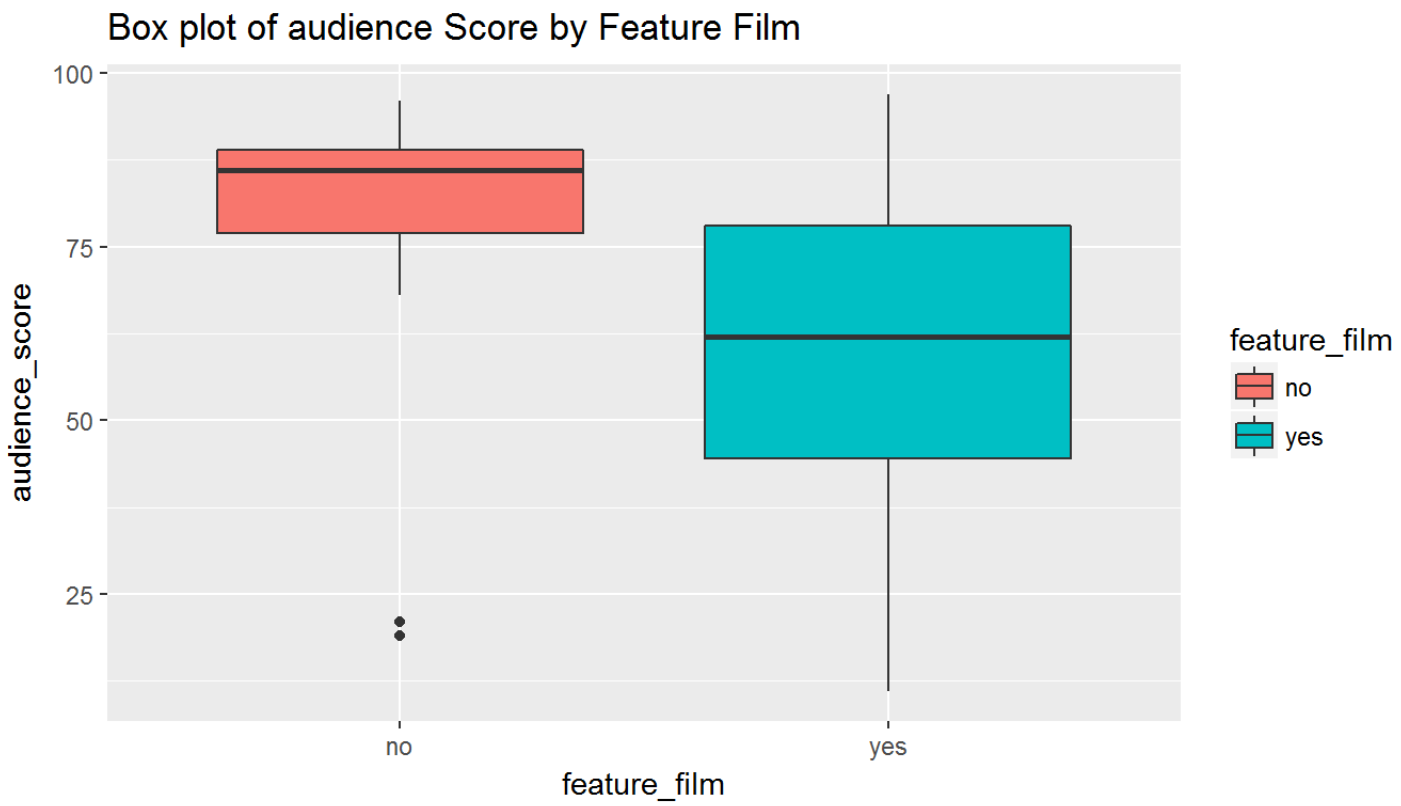


The runtime variable ranges from 39.0 to 267.0 minutes with a mean of 105.8 and a median of 103.0 minutes and the corresponding histogram indicates that the distribution is roughly normal.The audience score variable has a minimum of 11 and a maximum of 97 with a mean of 62.36 and a median of 65.00.

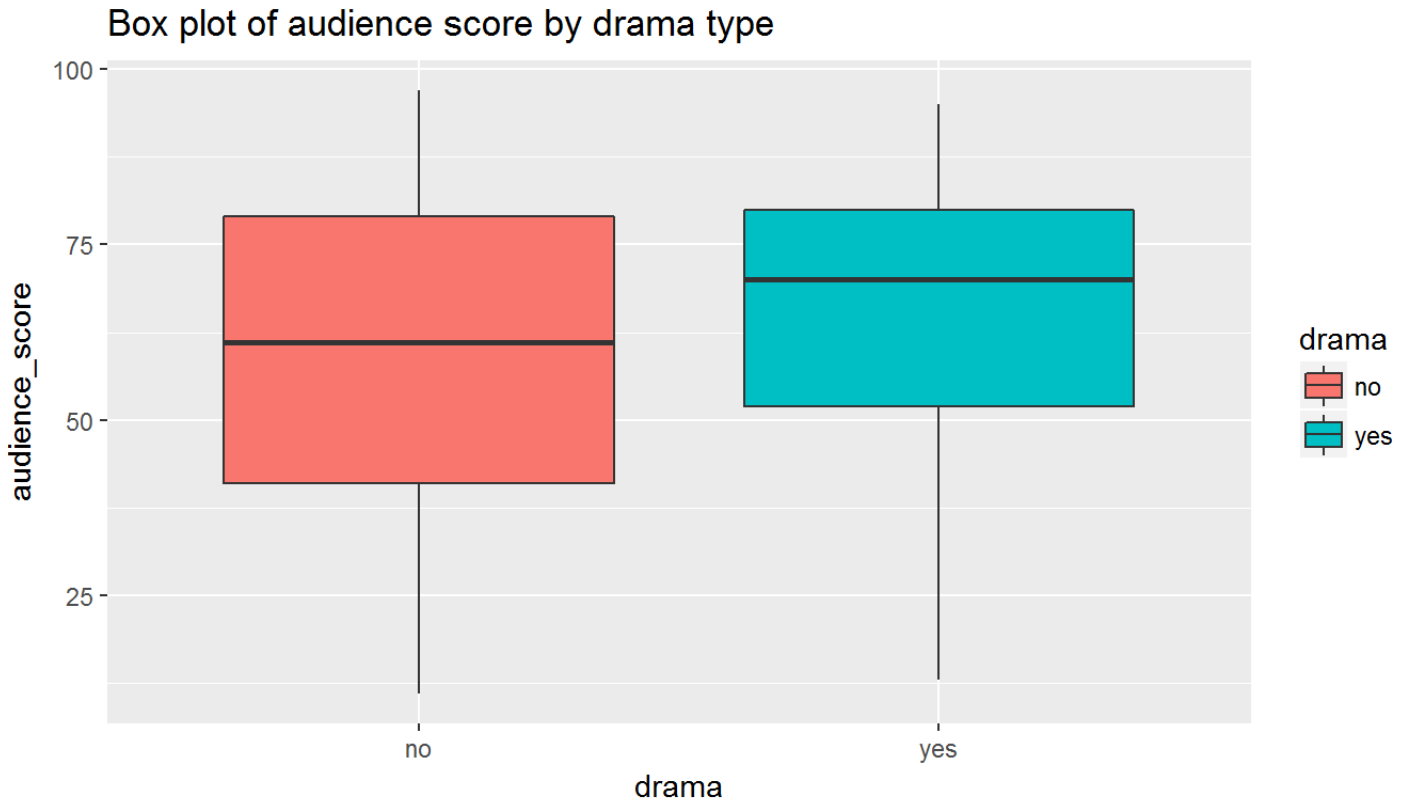
```
#distribution of the newly created dichotomous variables
summary(movies[,c("feature_film","drama", "mpaa_rating_R", "oscar_season", "summer_se
ason")])
```

```
##  feature_film drama     mpaa_rating_R oscar_season summer_season
##  no : 59      no :345   no :321       no :460      no :442
##  yes:591      yes:305   yes:329       yes:190      yes:208
```

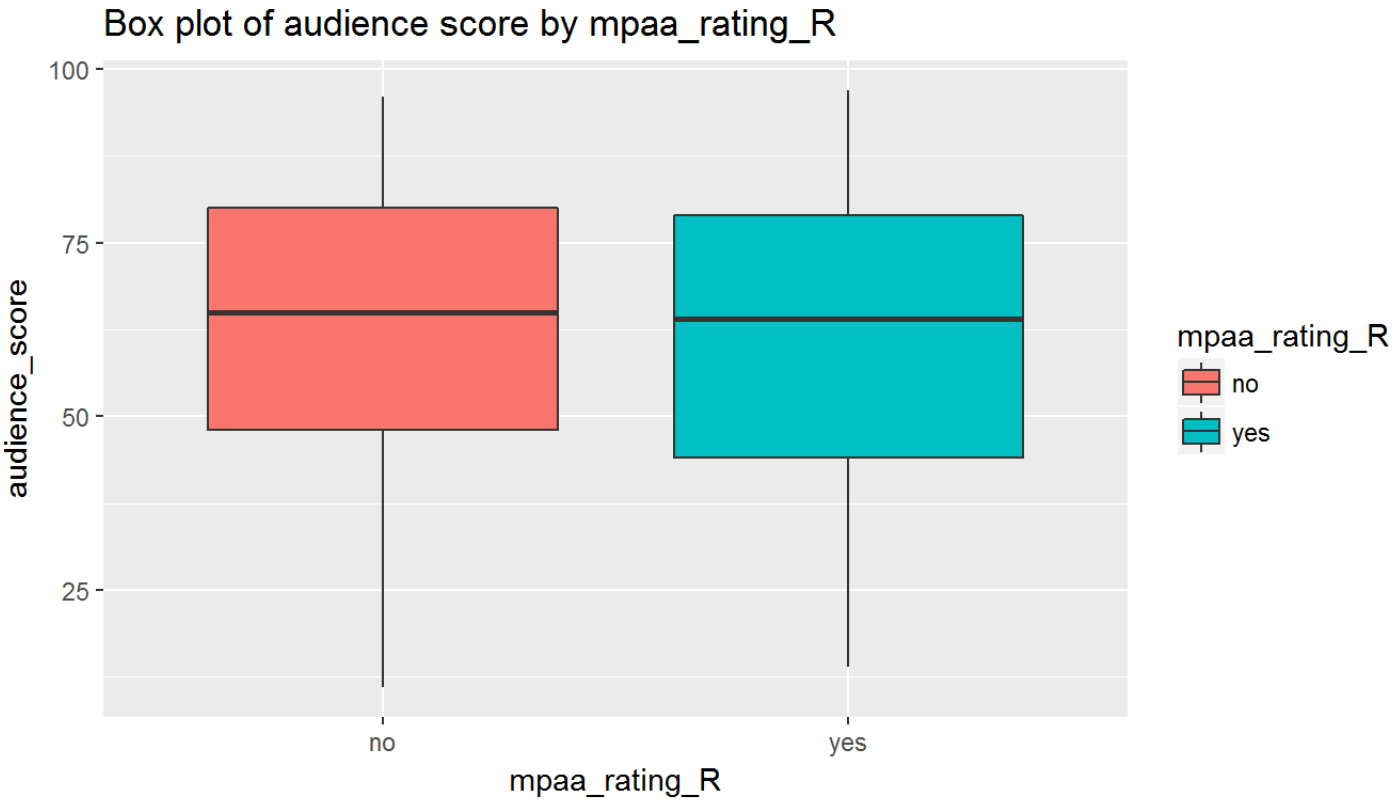
```
ggplot(movies, aes(factor(feature_film), audience_score, fill=feature_film)) +
    geom_boxplot() +
    ggtitle('Box plot of audience Score by Feature Film') +
    xlab('feature_film') +
    ylab('audience_score')
```



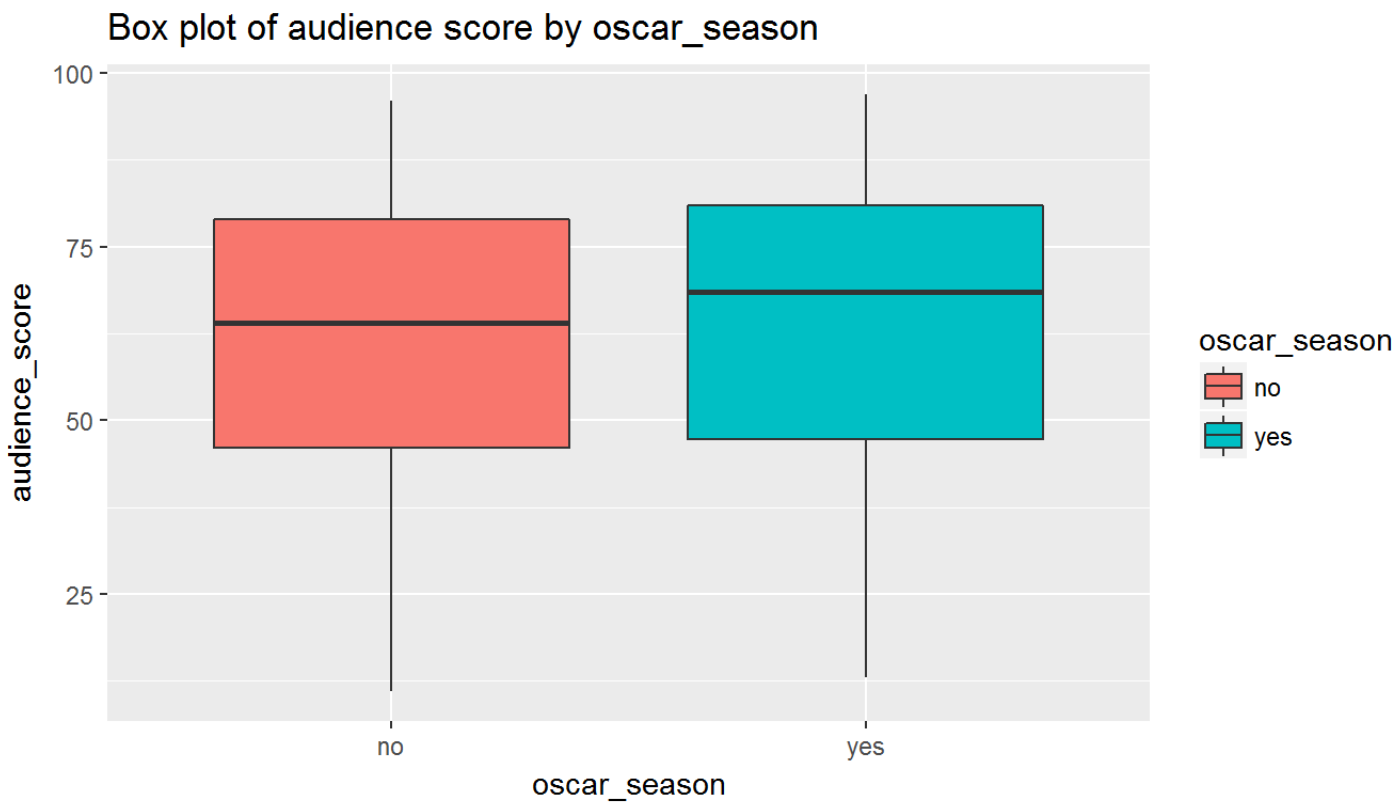

```
ggplot(movies, aes(factor(drama), audience_score, fill= drama)) +
    geom_boxplot() +
    ggtitle('Box plot of audience score by drama type') +
    xlab('drama') +
    ylab('audience_score')
```



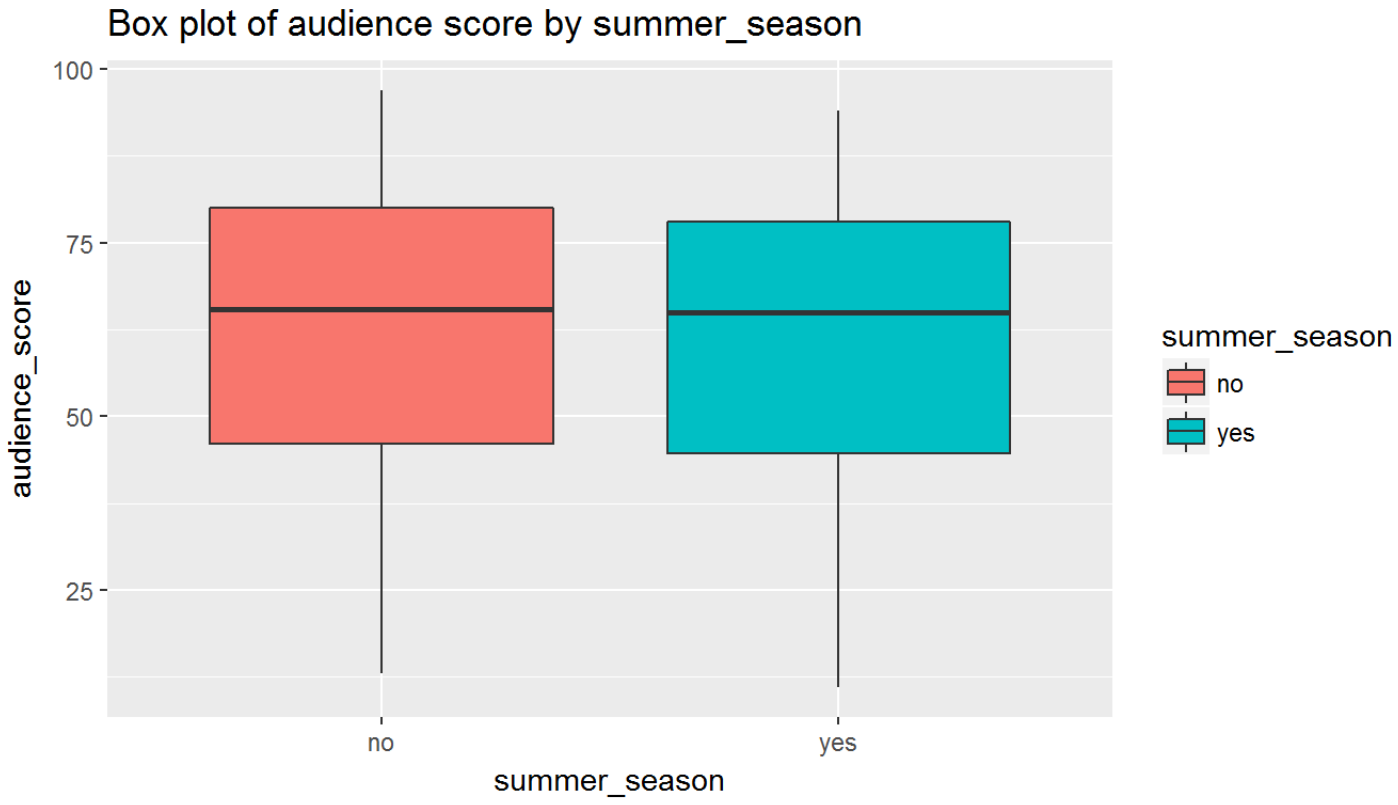
```
ggplot(movies, aes(factor(mpaa_rating_R), audience_score, fill= mpaa_rating_R)) +
    geom_boxplot() +
    ggtitle('Box plot of audience score by mpaa_rating_R') +
    xlab('mpaa_rating_R') +
    ylab('audience_score')
```




```
ggplot(movies, aes(factor(oscar_season), audience_score, fill= oscar_season)) +
    geom_boxplot() +
    ggtitle('Box plot of audience score by oscar_season') +
    xlab('oscar_season') +
    ylab('audience_score')
```



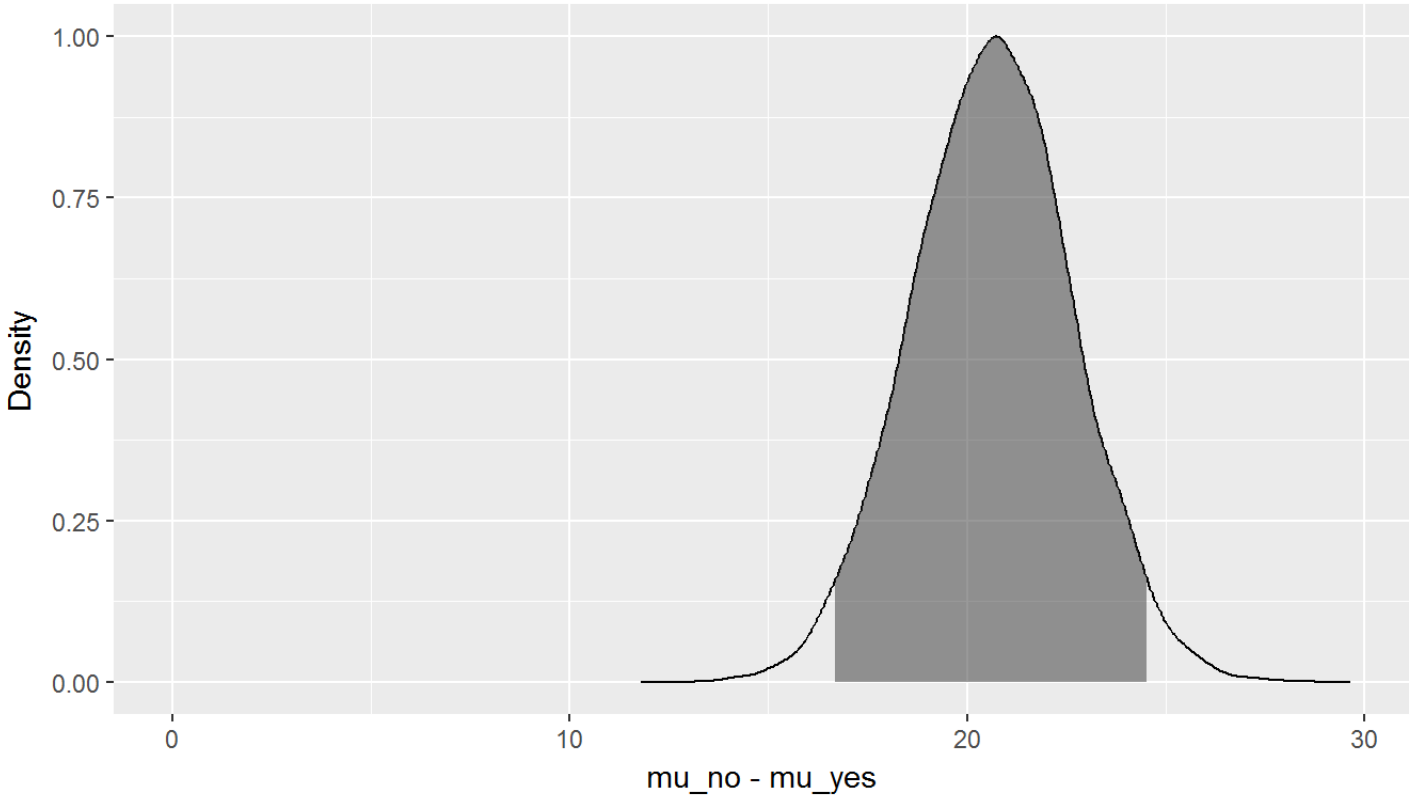

```
ggplot(movies, aes(factor(summer_season), audience_score, fill= summer_season)) +
    geom_boxplot() +
    ggtitle('Box plot of audience score by summer_season') +
    xlab('summer_season') +
    ylab('audience_score')
```



According to the boxplots, out of the 5 newly created binary variables, three variables, namely, feature_film, drama, and oscar_season, show a clearly visible difference in the distribution by the audience score.

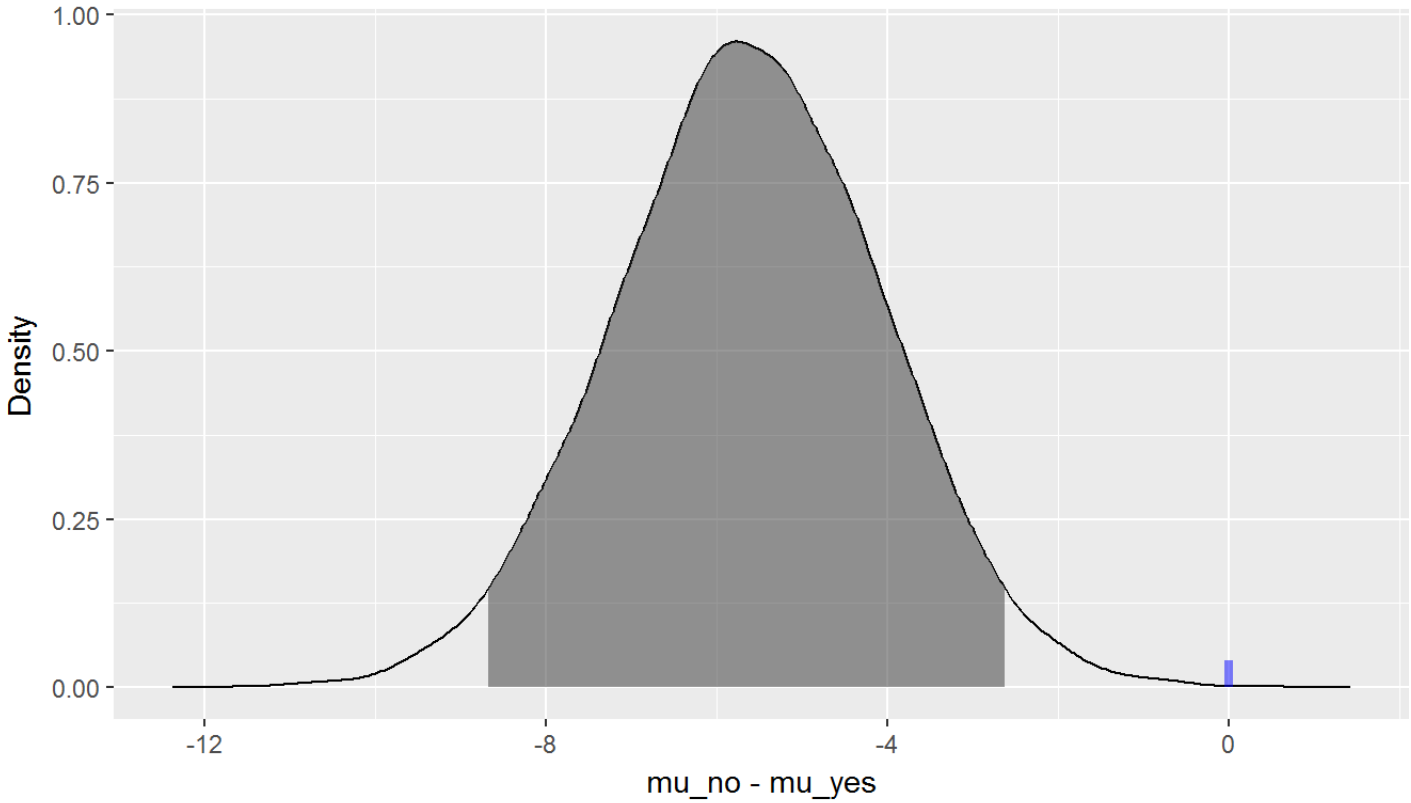
```
bayes_inference(y = audience_score, x = feature_film, data = movies, statistic = "mea
n", type = "ht", null = 0, alternative = "twosided")
```

```
## Response variable: numerical, Explanatory variable: categorical (2 levels)
## n_no = 59, y_bar_no = 81.2034, s_no = 13.6404
## n_yes = 591, y_bar_yes = 60.4653, s_yes = 19.824
## (Assuming intrinsic prior on parameters)
## Hypotheses:
## H1: mu_no  = mu_yes
## H2: mu_no != mu_yes
##
## Priors:
## P(H1) = 0.5
## P(H2) = 0.5
##
## Results:
## BF[H2:H1] = 2.794604e+13
## P(H1|data) = 0
## P(H2|data) = 1
```



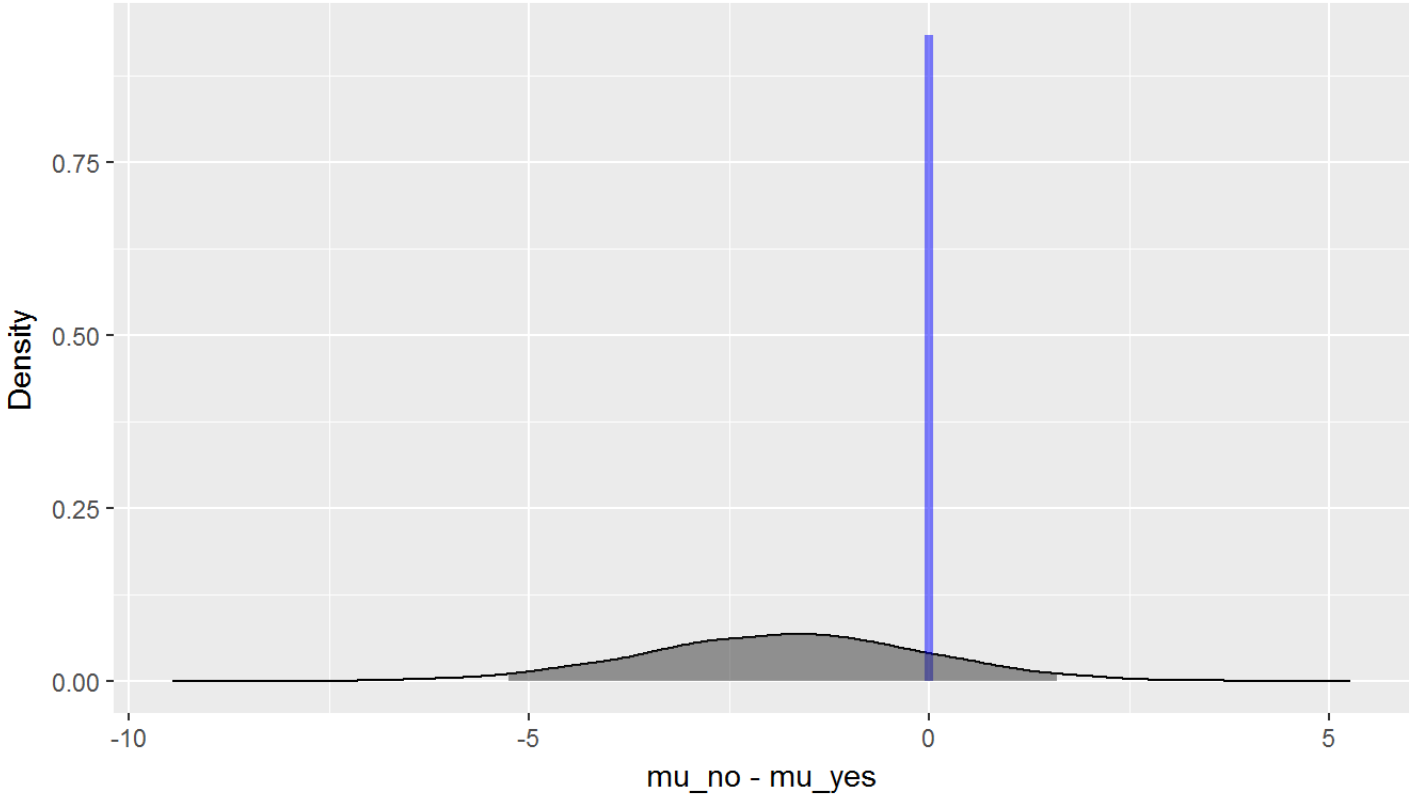

```
bayes_inference(y = audience_score, x = drama, data = movies, statistic = "mean", typ
e = "ht", null = 0, alternative = "twosided")
```

```
## Response variable: numerical, Explanatory variable: categorical (2 levels)
## n_no = 345, y_bar_no = 59.6957, s_no = 21.2981
## n_yes = 305, y_bar_yes = 65.3475, s_yes = 18.5418
## (Assuming intrinsic prior on parameters)
## Hypotheses:
## H1: mu_no  = mu_yes
## H2: mu_no != mu_yes
##
## Priors:
## P(H1) = 0.5
## P(H2) = 0.5
##
## Results:
## BF[H2:H1] = 24.1609
## P(H1|data) = 0.0397
## P(H2|data) = 0.9603
```




```
bayes_inference(y = audience_score, x = oscar_season, data = movies, statistic = "mea
n", type = "ht", null = 0, alternative = "twosided")
```

```
## Response variable: numerical, Explanatory variable: categorical (2 levels)
## n_no = 460, y_bar_no = 61.813, s_no = 20.1196
## n_yes = 190, y_bar_yes = 63.6421, s_yes = 20.5062
## (Assuming intrinsic prior on parameters)
## Hypotheses:
## H1: mu_no  = mu_yes
## H2: mu_no != mu_yes
##
## Priors:
## P(H1) = 0.5
## P(H2) = 0.5
##
## Results:
## BF[H1:H2] = 13.7738
## P(H1|data) = 0.9323
## P(H2|data) = 0.0677
```



In terms of Bayes Factors, after Jeffrey's criteria, there seem to be a clear difference in the audience score by the three selected binary variables above.

Part 4: Modeling

The audience_score was the dependent variable chosen as the proxy of the popularity of a movie, and a linear regression model and bayesian model averaging were used for the modelling. Following 12 variables were selected as predictors; run_time, thtr_rel_month, dvd_rel_month, imdb_rating,

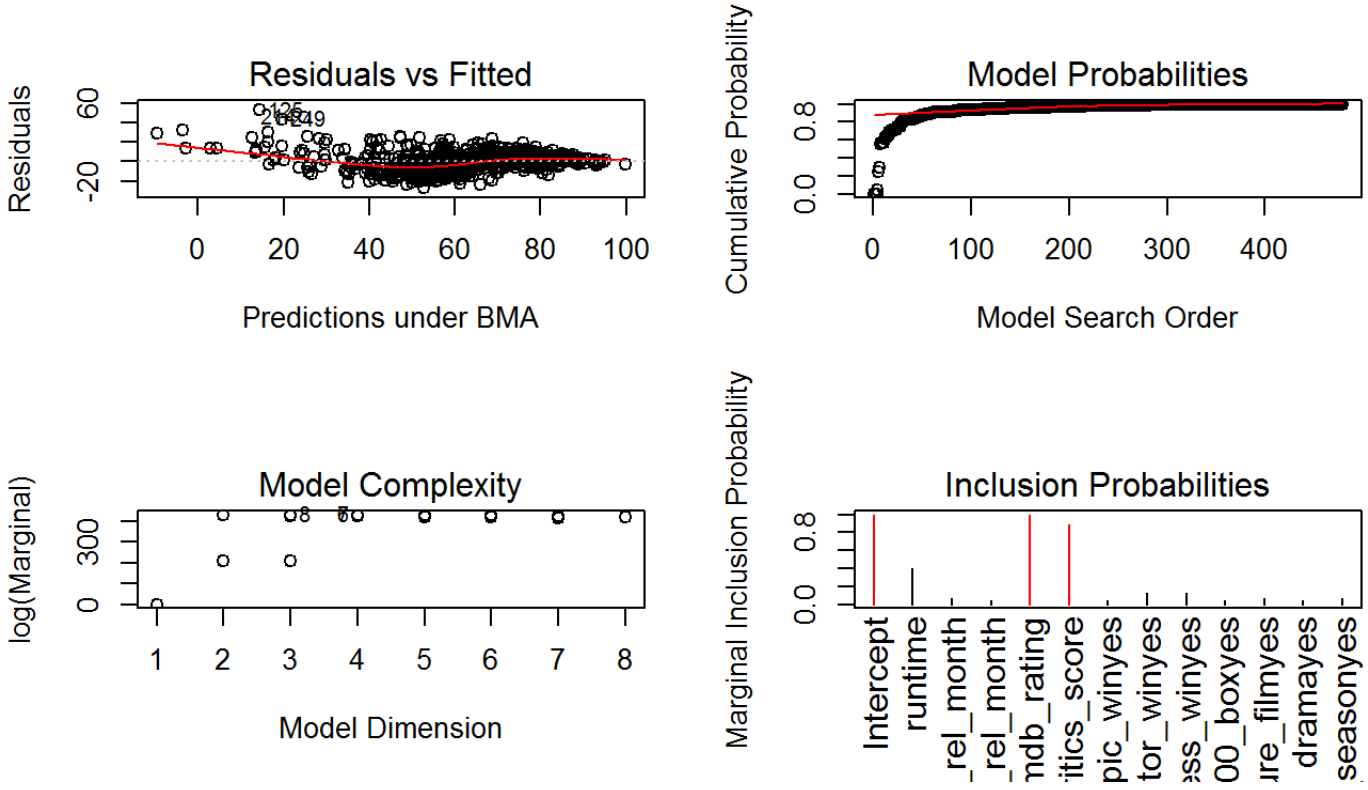
critics_score,best_pic_win, best_actor_win, best_actress_win, top200_box, feature_film, drama, and oscar_season.

```
#extracting the variables
movies <- select(movies, runtime, thtr_rel_month, dvd_rel_month, imdb_rating, critics
_score, audience_score, best_pic_win, best_actor_win, best_actress_win, top200_box, f
eature_film, drama, oscar_season)

#model fitting
model <- bas.lm(audience_score ~ ., data=movies, method='MCMC', prior='ZS-null', mode
lprior=uniform())
```

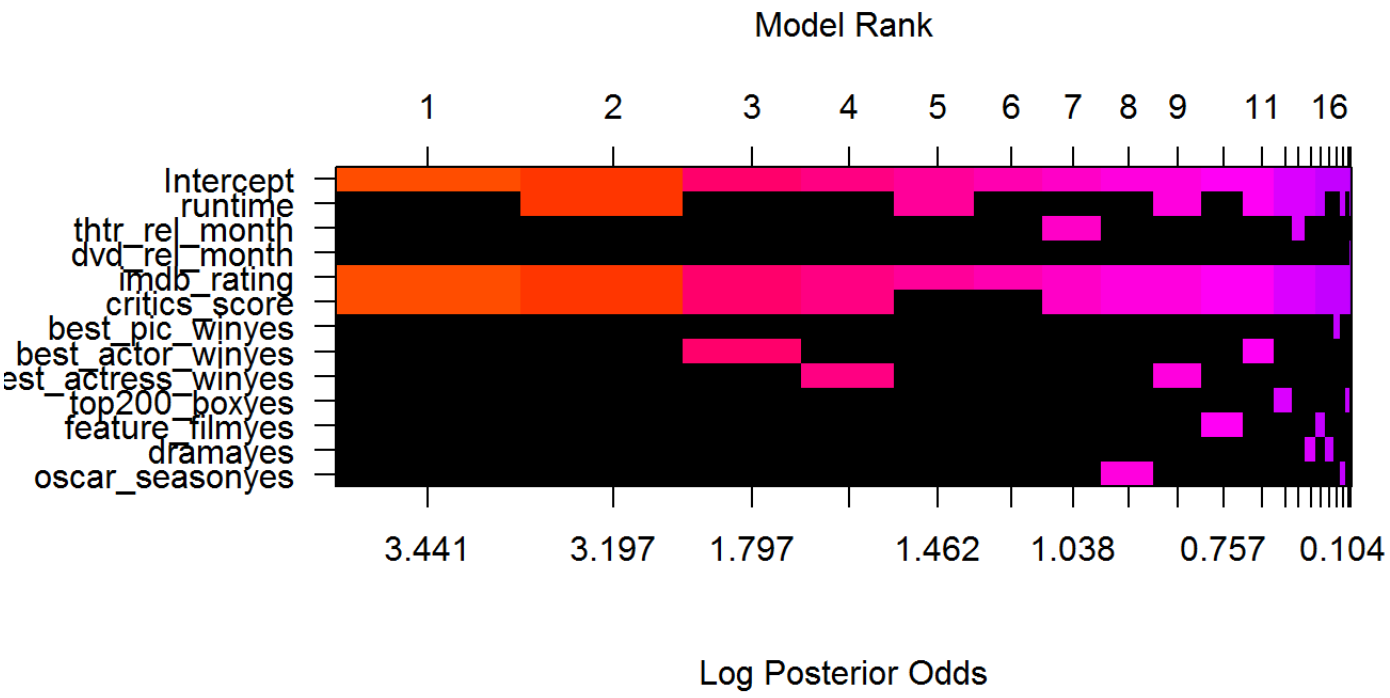
```
## Warning in bas.lm(audience_score ~ ., data = movies, method = "MCMC", prior
## = "ZS-null", : dropping 8 rows due to missing data
```

```
#Diagnostic plots
par(mfrow=c(2,2))
plot(model, which=c(1, 2, 3, 4), ask=FALSE)
```



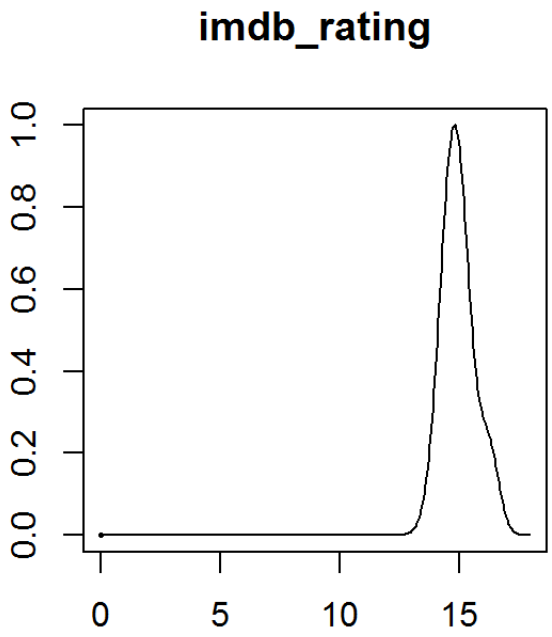
The first plot of residuals versus fitted values shows the random error variance assumption is not met throughout the predicted values. In fact, the random scatter is acceptable in the range 30-90 but outside that range it is not well met. The second plot shows that the convergnce of the posterior probability occured with about 350 model combinations sampling. The fourth plot suggests that imdb_rating and critics_score are important predictors with higher inclusion probabilities.

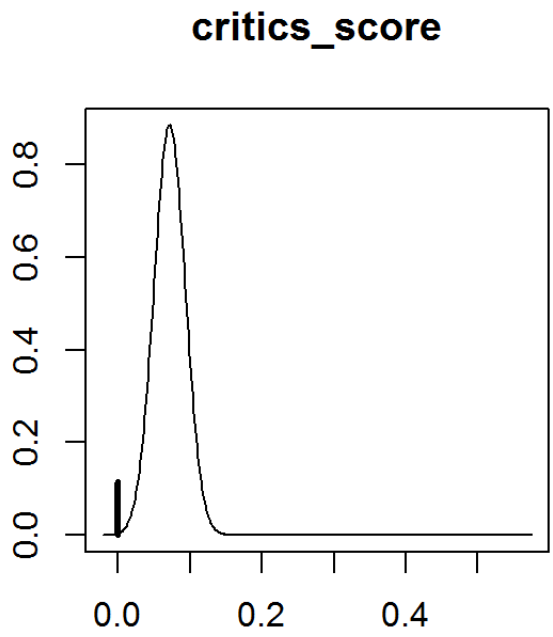
```
image(model, rotate = FALSE)
```



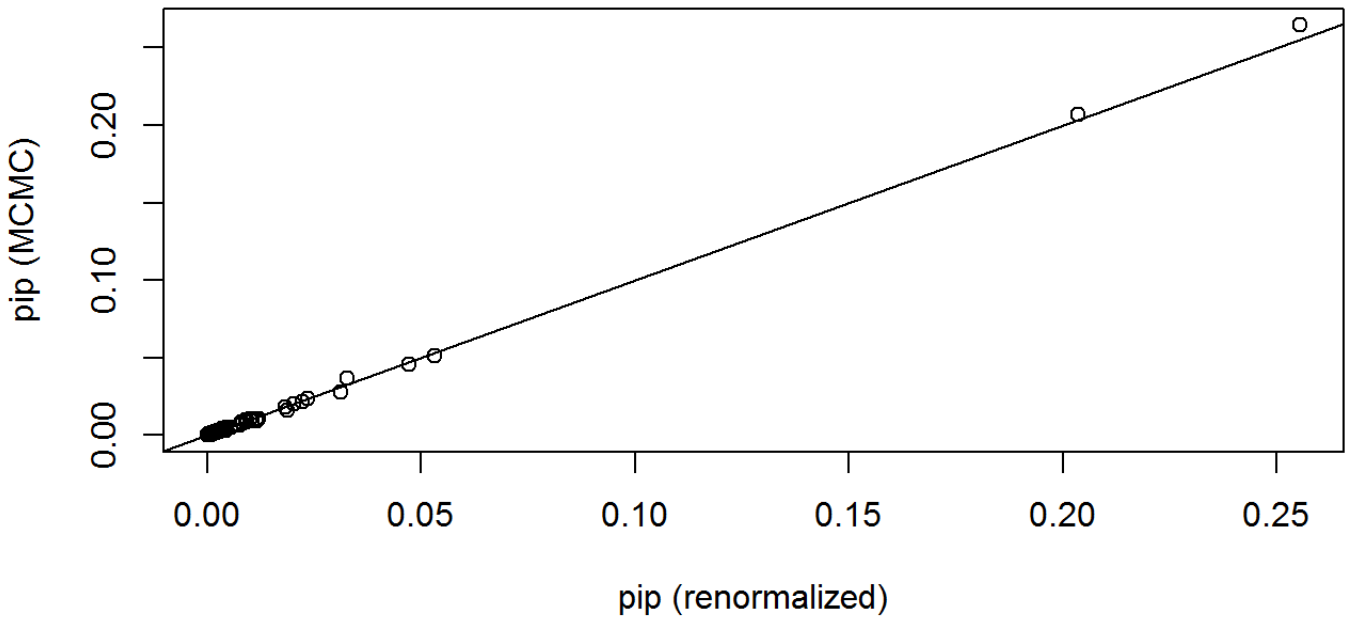
The best model with the highest posterior probability contains only two predictors, namely, imdb_rating and critics_score. In fact, both these variables are almost always included in the best 15 models.

```
par(mfrow=c(1,2))
plot(coefficients(model), subset=c(5, 6), ask=FALSE)
```






```
#checking the normality of posterior probabilities
diagnostics(model, type="model")
```



The above plot indicates that normality is satisfactory and the MCMC iterations were suffificient.

```
#predictions on the dataset and making credible intervals
BMA_model = predict(model, estimator="BMA", se.fit=TRUE)
BMA_confint_fit = confint(BMA_model, parm="mean")
BMA_confint_pred = confint(BMA_model, parm="pred")
head(cbind(BMA_confint_fit, BMA_confint_pred))
```

```
##          2.5%    97.5%     mean     2.5%     97.5%     pred
## [1,] 45.76137 48.98393 47.29617 27.10530  66.52969 47.29617
## [2,] 75.20927 78.76845 77.06186 57.66762  97.38306 77.06186
## [3,] 79.71160 83.61221 81.53372 61.23741 100.73663 81.53372
## [4,] 71.12974 75.41777 73.37822 53.35927  93.01701 73.37822
## [5,] 38.84557 41.75027 40.27196 19.96140  60.21597 40.27196
## [6,] 82.71204 87.21853 84.81863 64.84948 104.79403 84.81863
```

Part 5: Prediction

The best movie of the year 2016 that won the Oscar was chosen for prediction. Relevant data were gained from the IMDB and Rotten Tomatoes web sites.

```
#making a dataframe with moonlight movie's data
moonlight <- data.frame(runtime= 111,
                         thtr_rel_month = 11,
                        dvd_rel_month = 2,
                         imdb_rating= 7.5,
                         critics_score= 98,
                         audience_score= 80,
                         best_pic_win= "yes",
                         best_actor_win= "no",
                         best_actress_win= "no",
                         top200_box= "yes",
                         feature_film= "yes",
                         drama= "yes",
                         oscar_season= "yes"
                         )

# predicting of audience_score using bayesian model averaging.
BMA_moonlight <- predict(model, newdata= moonlight, estimator="BMA", se.fit=TRUE)

# prediction intervals
moonlight_pred <- qt(0.95, df=BMA_moonlight$se.bma.pred[1]) *
                     mean(BMA_moonlight$se.bma.pred)

# Show prediction results.
outcome <- data.frame(t="moonlight",
                  p=sprintf("%2.1f", BMA_moonlight$Ybma),
                  i=sprintf("%2.1f - %2.1f", BMA_moonlight$Ybma - moonlight_pred,
                            BMA_moonlight$Ybma + moonlight_pred),
                  r= 80)
colnames(outcome) <- c("Movie", "Predicted", "95% Interval",
                   "Actual")
print(outcome)
```

```
##       Movie Predicted 95% Interval Actual
## 1 moonlight      79.9  61.5 - 98.4     80
```

And lo behold! :)

Part 6: Conclusion

In conclusion, a parsimonious bayesian regression model has been produced that made pretty impressive predictions on a new movie. However, in view of the model assumptions that were not met according to the diagnostic plots, there is room for further analyses and improving the model.