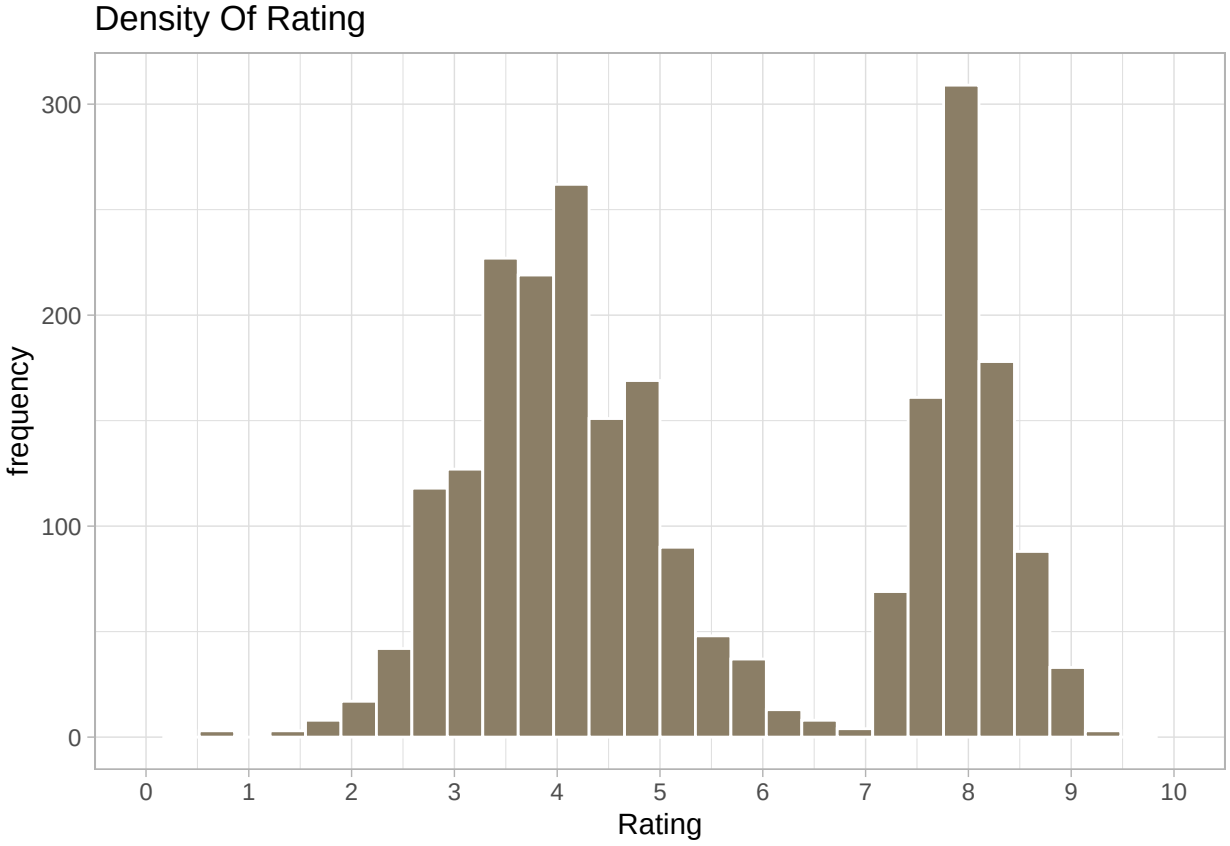
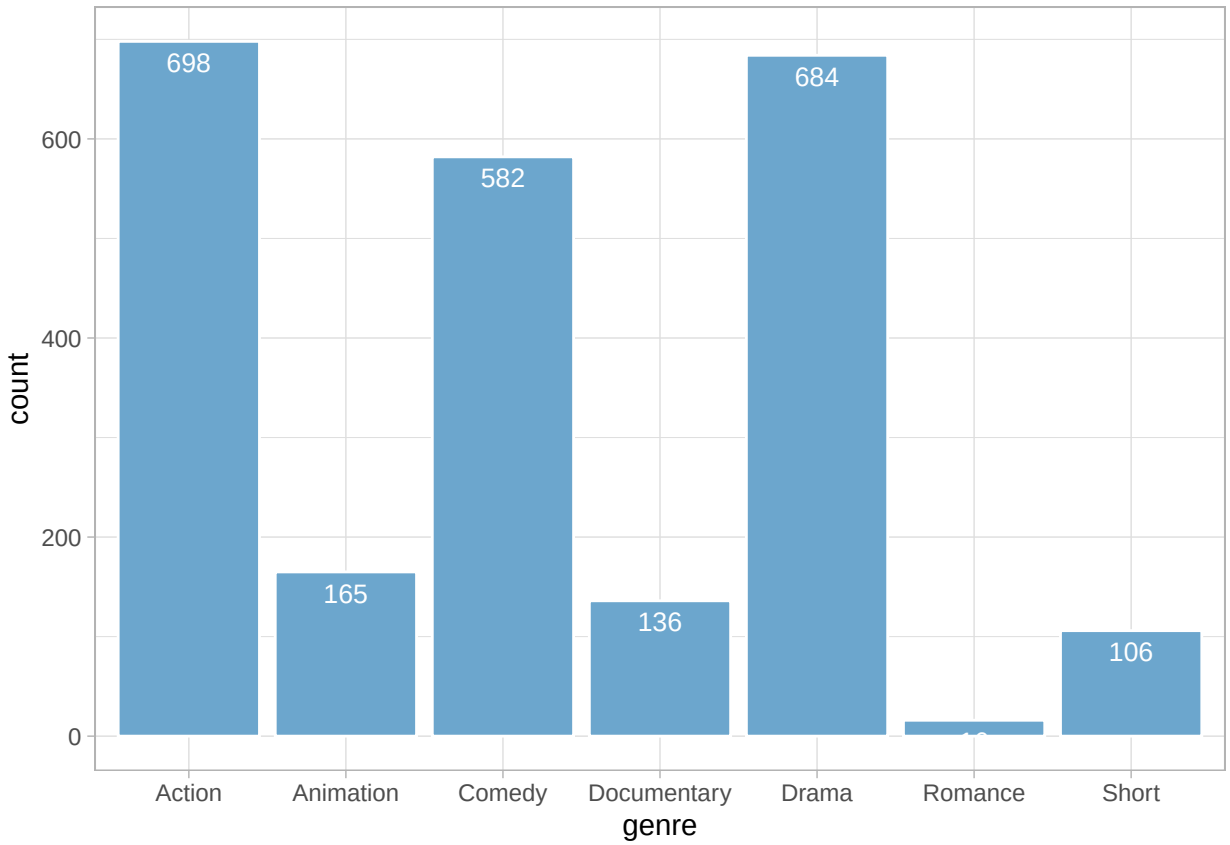
Group_07_Analysis

Group 7

```
    rating
Min.   :0.700
1st Qu.:3.700
Median :4.700
Mean   :5.414
3rd Qu.:7.800
Max.   :9.200
```

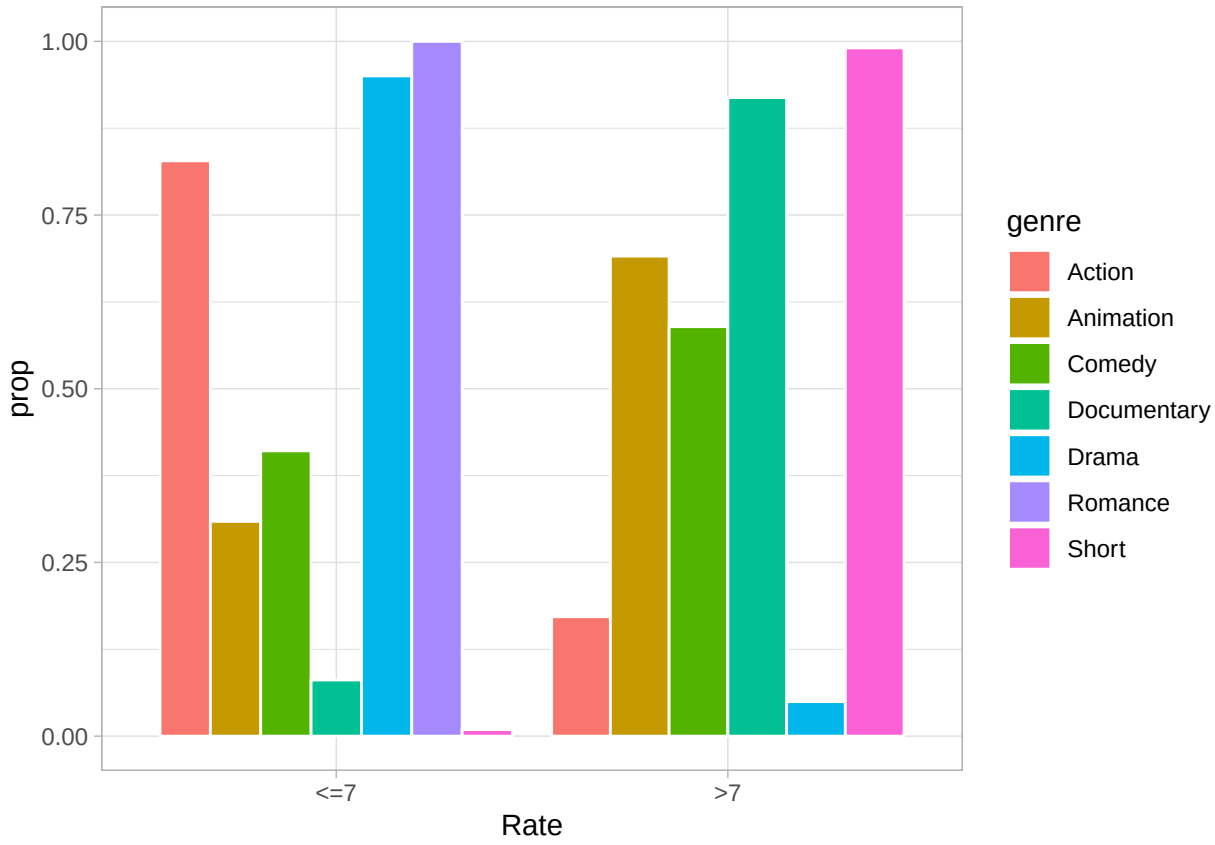


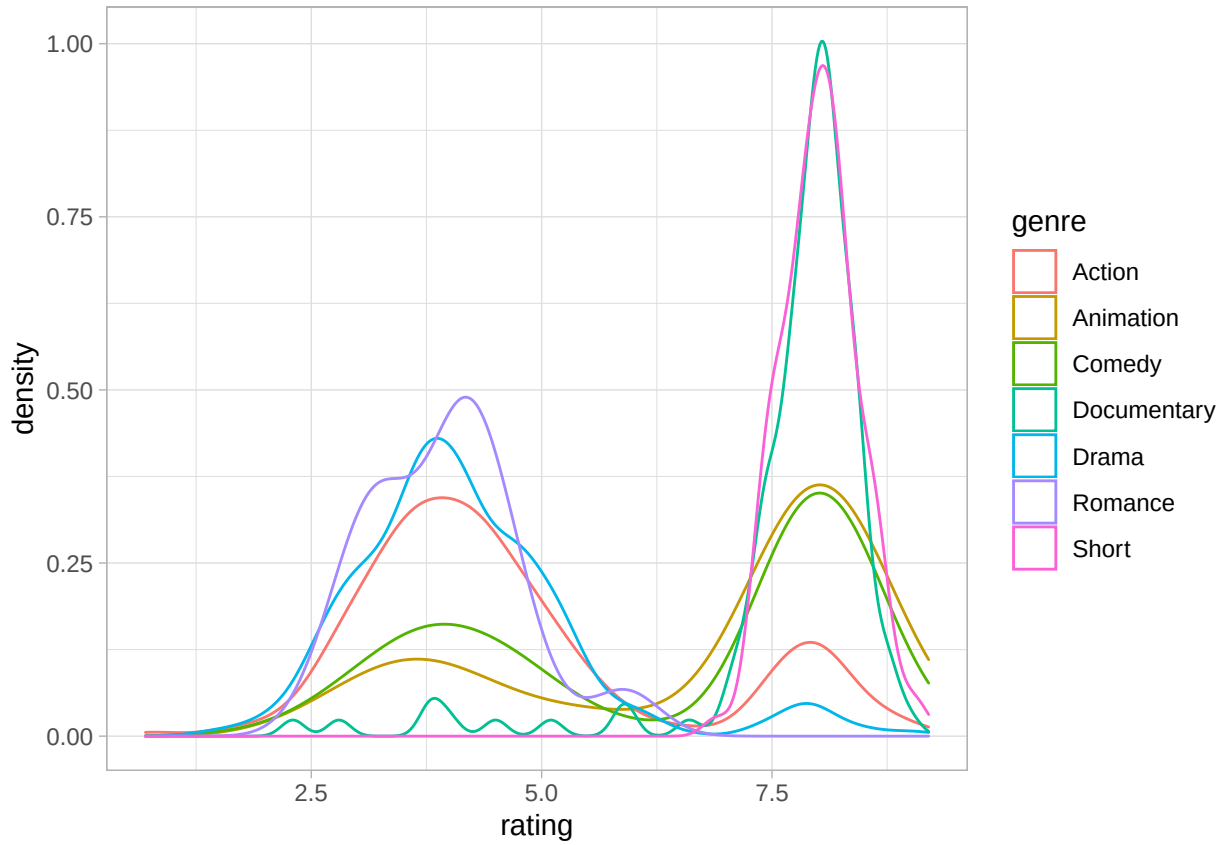

```
Rate      Action   Animation      Comedy Documentary       Drama   Romance
 <=7 37.4% (578)  3.3%  (51) 15.5% (239)  0.7%  (11) 42.0% (650) 1.0% (16)
  >7 14.3% (120) 13.6% (114) 40.8% (343) 14.9% (125)  4.0%  (34) 0.0%  (0)
      Short
 0.1%   (1)
12.5% (105)
```



.

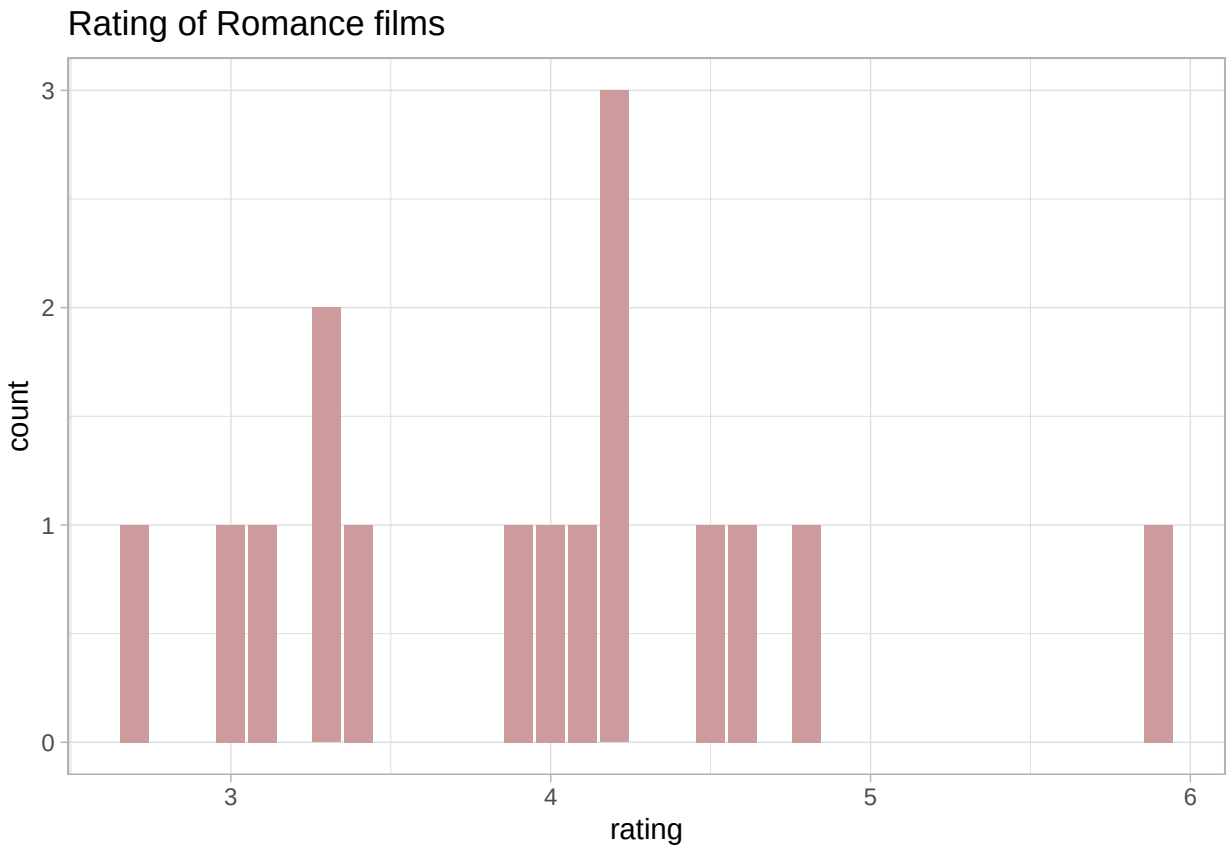
```
Action    Animation       Comedy Documentary        Drama      Romance
   698          165          582          136          684           16
 Short
   106
```


1.00
0.75
0.50
0.25
0.00
prop
<=7
>7
Rate
genre
Action
Animation
Comedy
Documentary
Drama
Romance
Short

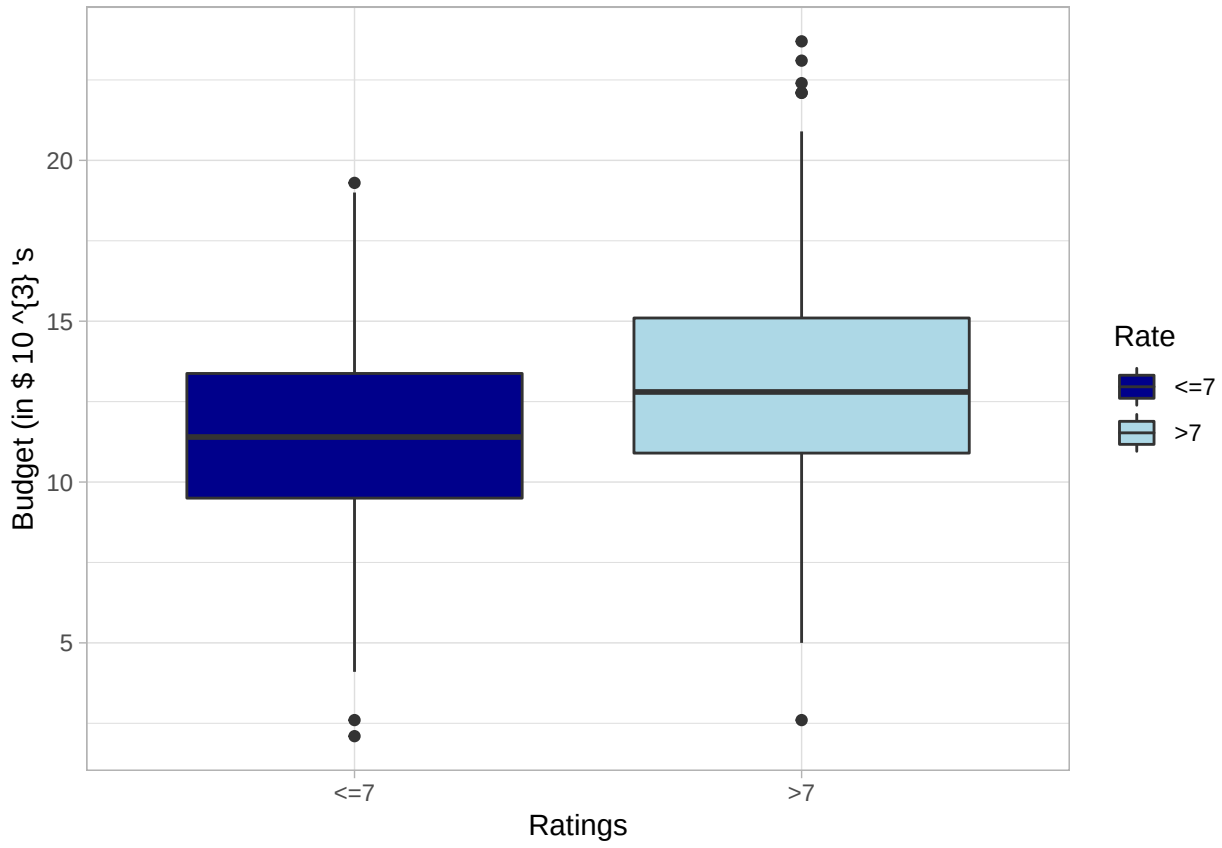


From this density gragh we can see that Short, Documentary,Comedy and Animation films are more likely to get rates over 7.5 compared to rates lower than 7 (5 to be exact)

We notice that the majority of films are within one of the 3 most dominant categories Action , Comedy and Drama . On the opposite Animation Documentary and Short films have relatively close frequencies . And Romance films are comparatively very rare among our list of films (only 16 films where Romantic among all 2387)

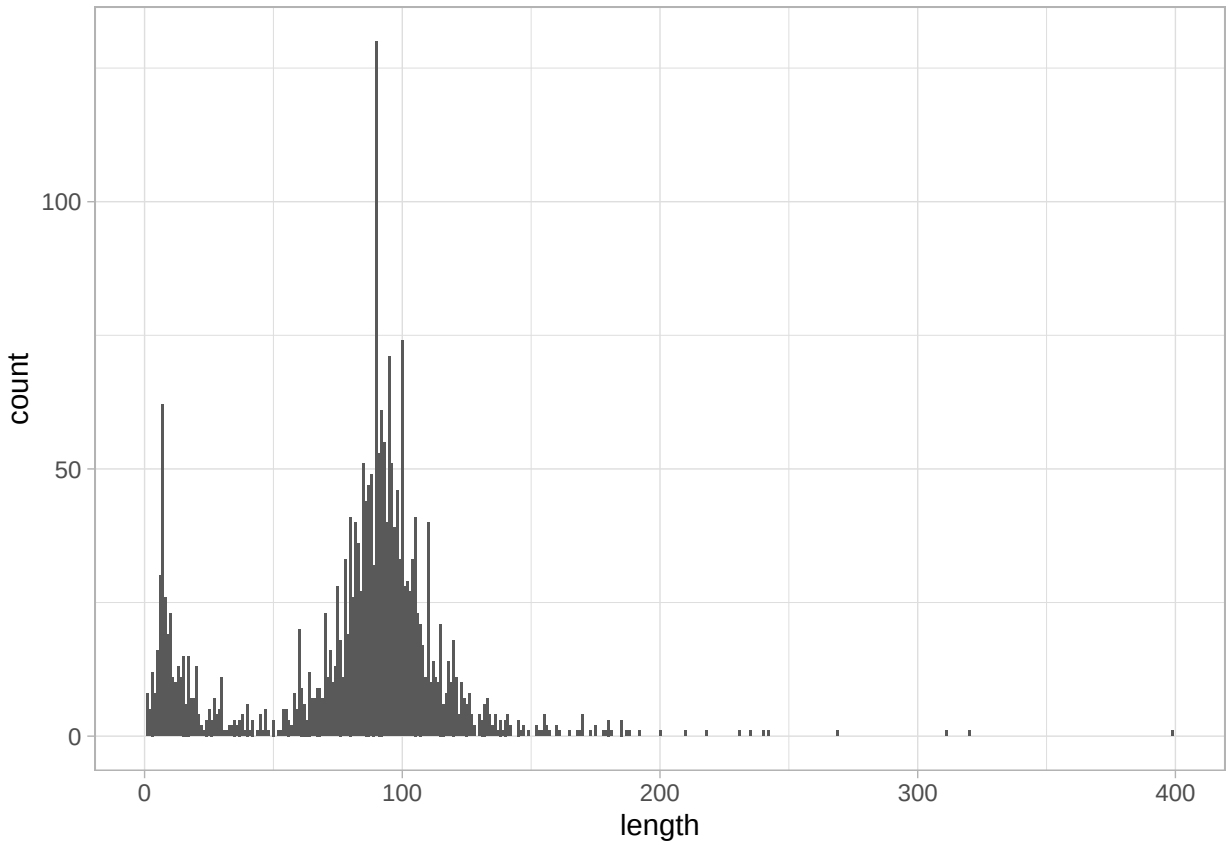


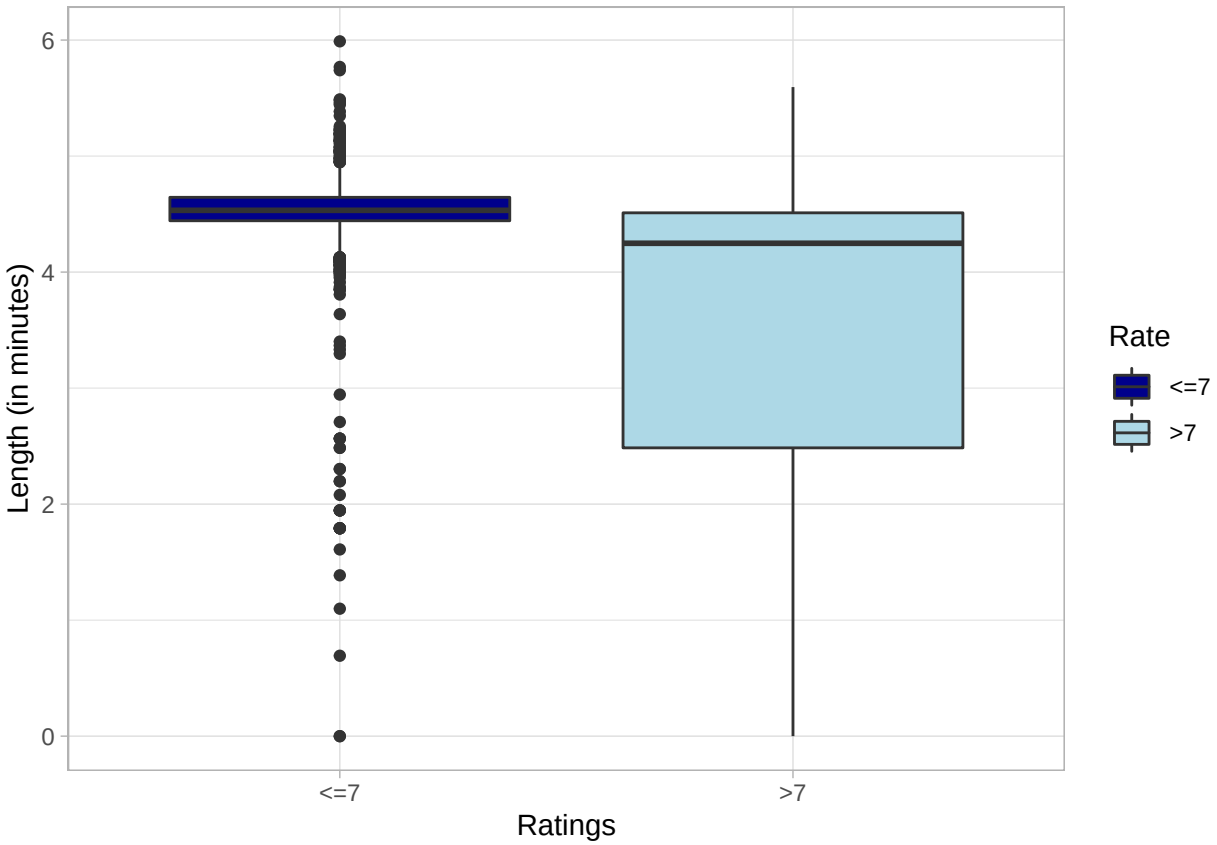
Given the barchart we notice that in our datasel all Romance films have ratings below 7. This might indicate that Romance is not a strong explanation a

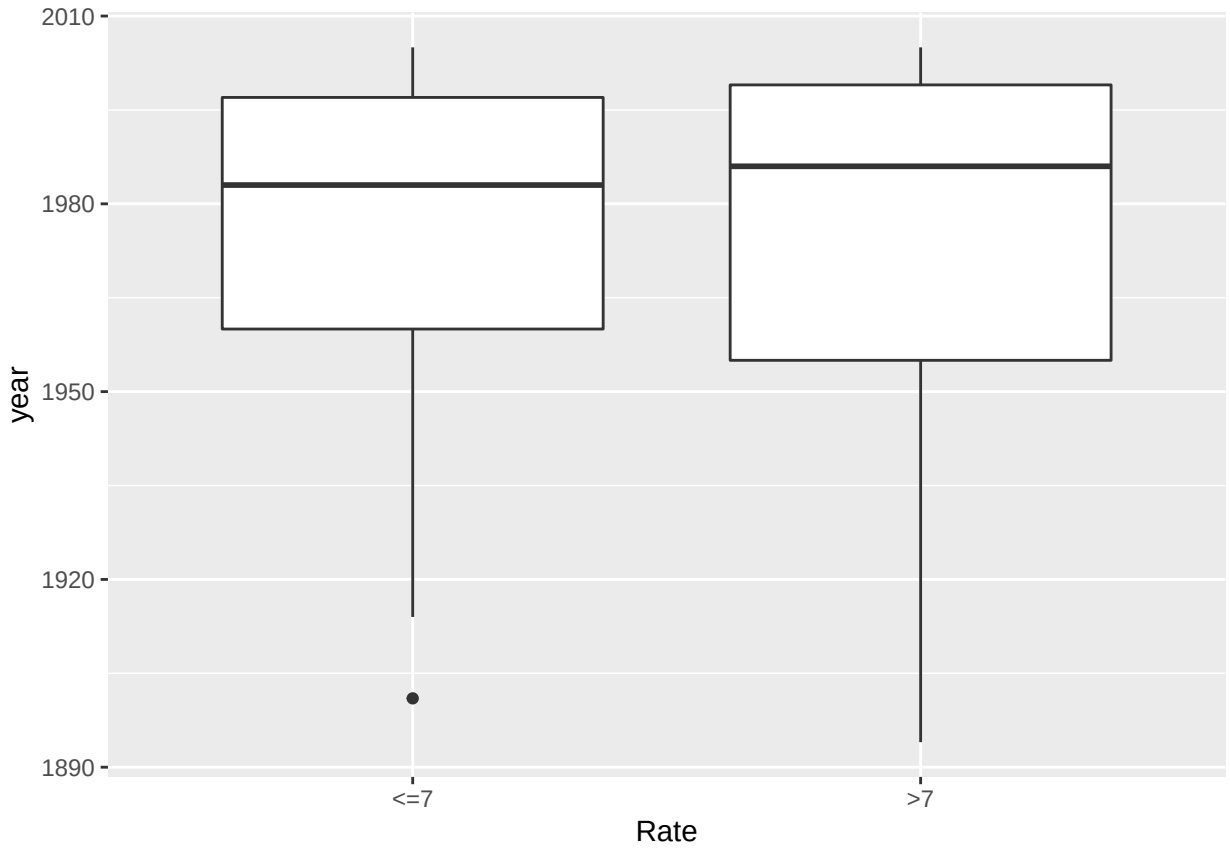


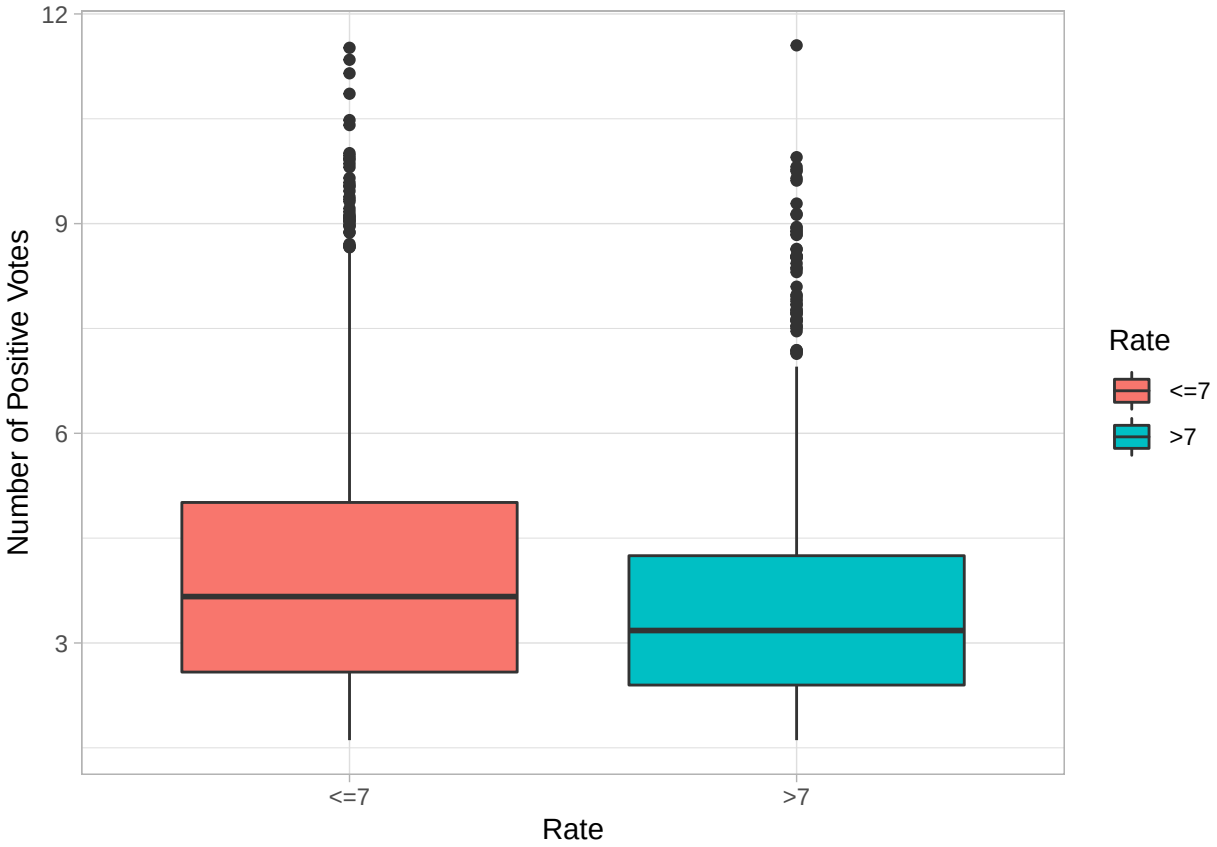

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
   1.00   72.00   90.00   81.41  100.00  399.00      92
```

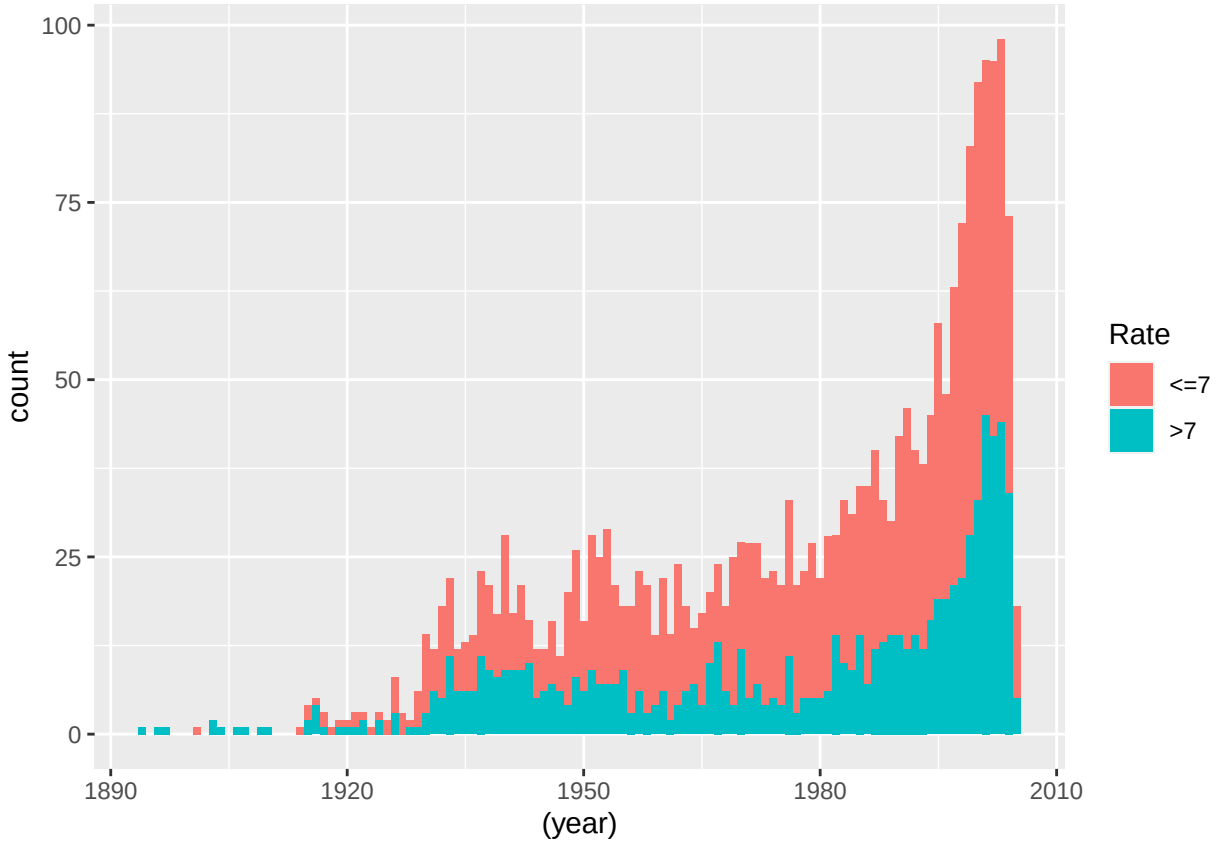
```
[1] 87
```

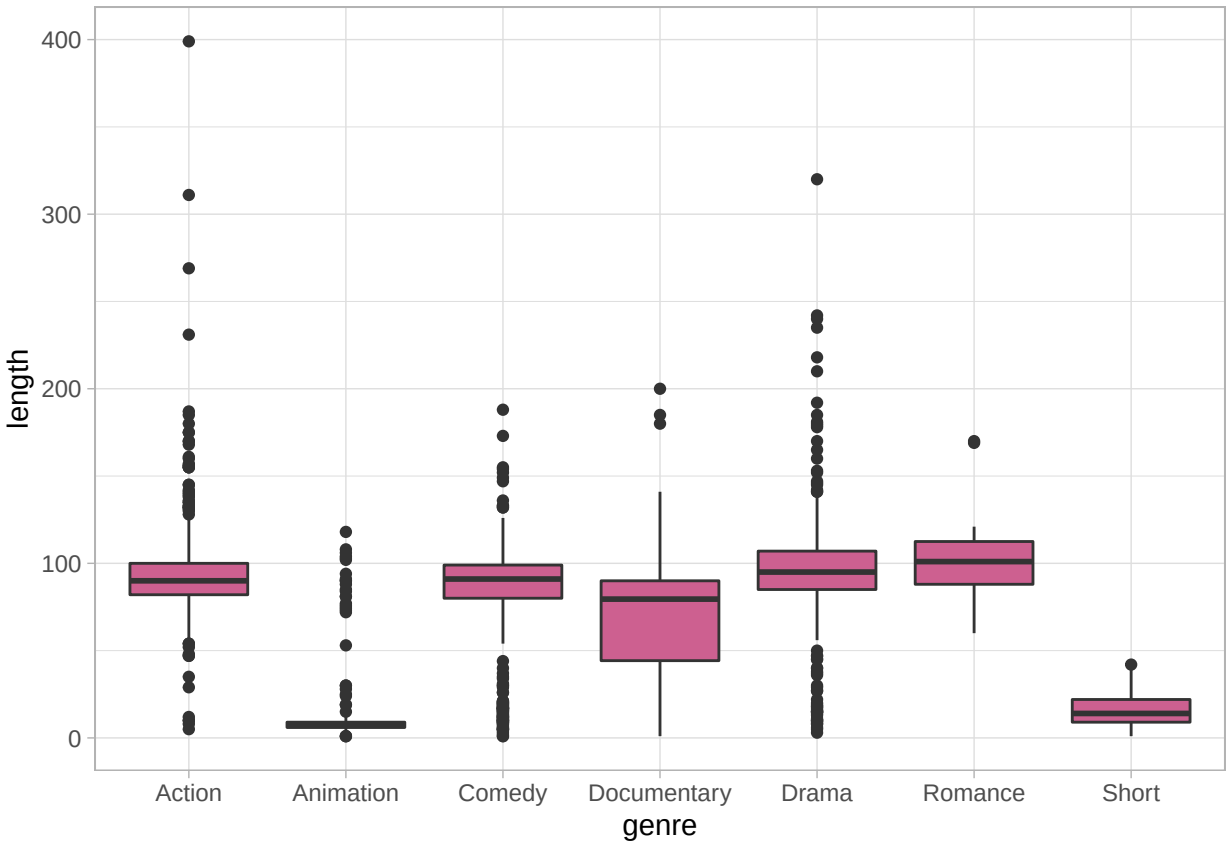

count
100
50
0
0
100
200
300
400
length

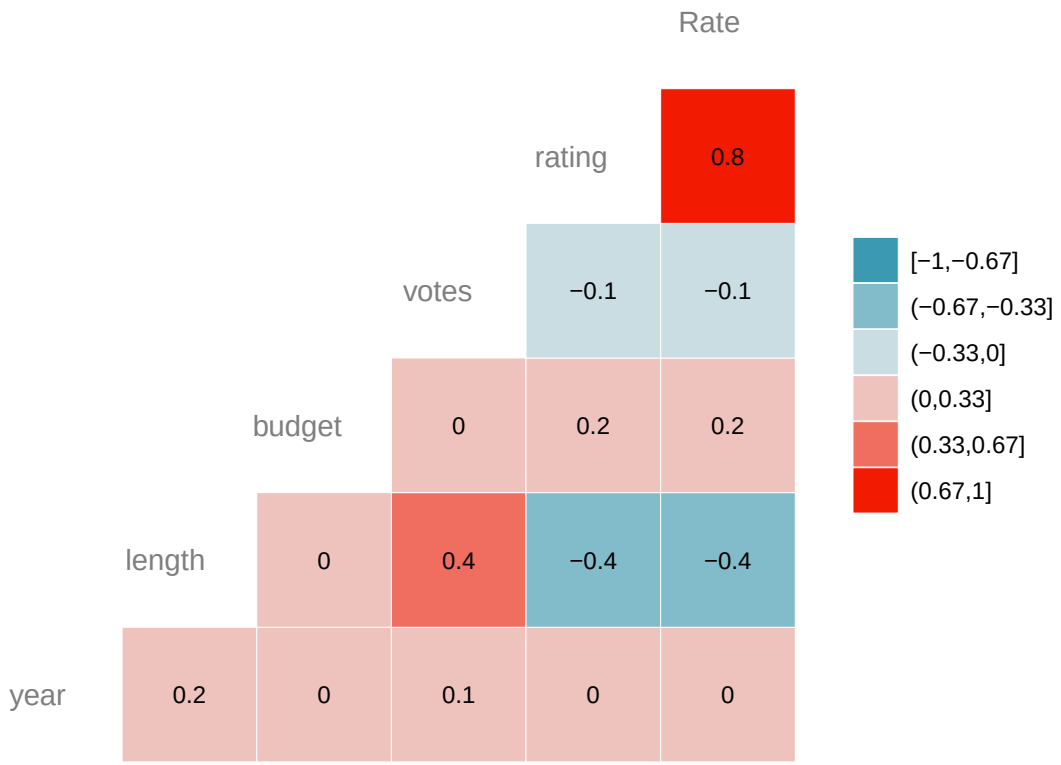

6
4
2
0
Length (in minutes)
<=7
>7
Ratings
Rate
<=7
>7


2010
1980
1950
1920
1890
year
<=7
>7
Rate


12
9
6
3
Number of Positive Votes
<=7
>7
Rate
Rate
<=7
>7


100
75
50
25
0
count
1890
1920
1950
1980
2010
(year)
Rate
<=7
>7


400
300
200
100
0
length
Action
Animation
Comedy
Documentary
Drama
Romance
Short
genre



We see in general films' duration are roughly in same range except Animation and Short films. Both have comperatively smaller lenth

```
Call:
glm(formula = Rate ~ genre + length, family = binomial(link = "logit"),
    data = film)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.9520  -0.5677  -0.2469   0.3637   4.1533

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)        1.832605   0.269055   6.811 9.67e-12 ***
genreAnimation    -0.257000   0.299480  -0.858 0.390809
genreComedy        1.949419   0.143447  13.590  < 2e-16 ***
genreDocumentary   3.989061   0.373338  10.685  < 2e-16 ***
genreDrama        -1.576207   0.222284  -7.091 1.33e-12 ***
genreRomance     -13.669351 348.890679  -0.039 0.968747
genreShort         3.499947   1.028815   3.402 0.000669 ***
length            -0.038875   0.002934 -13.250  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2982.5  on 2294  degrees of freedom
```