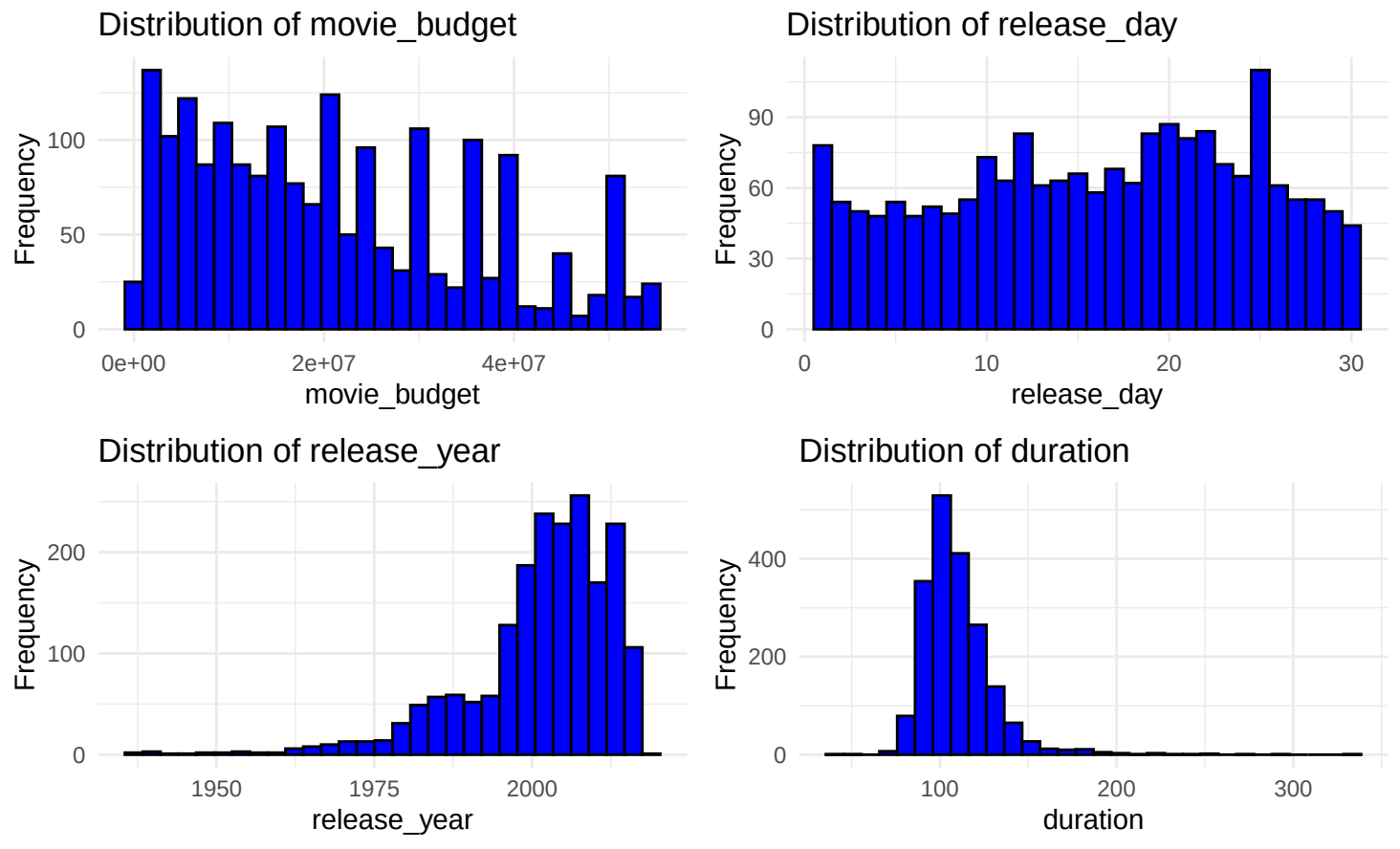
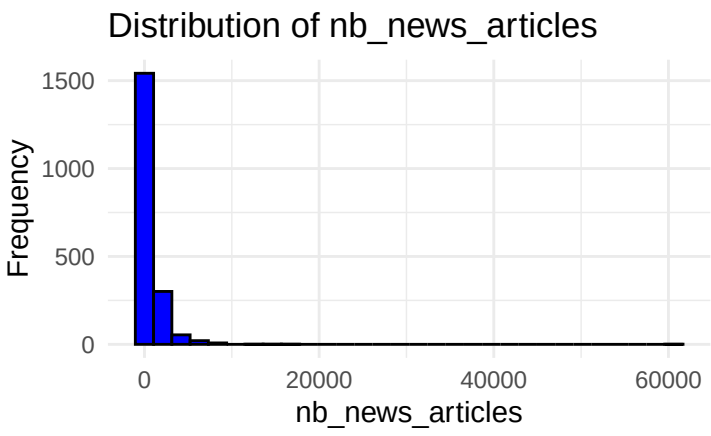
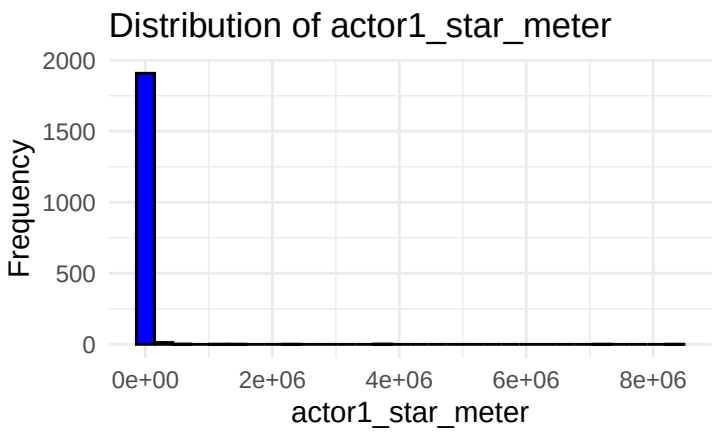
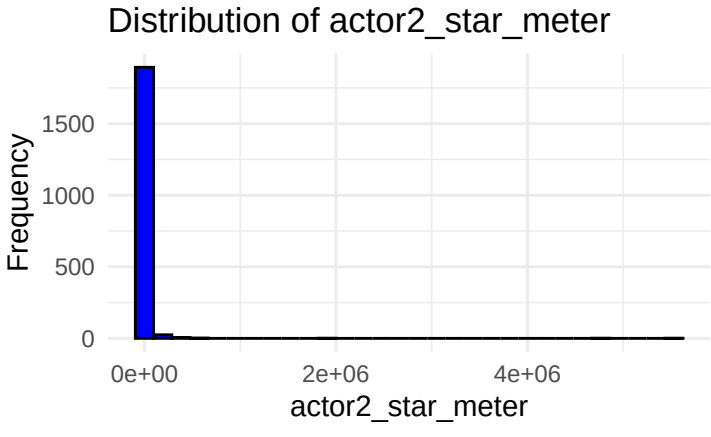
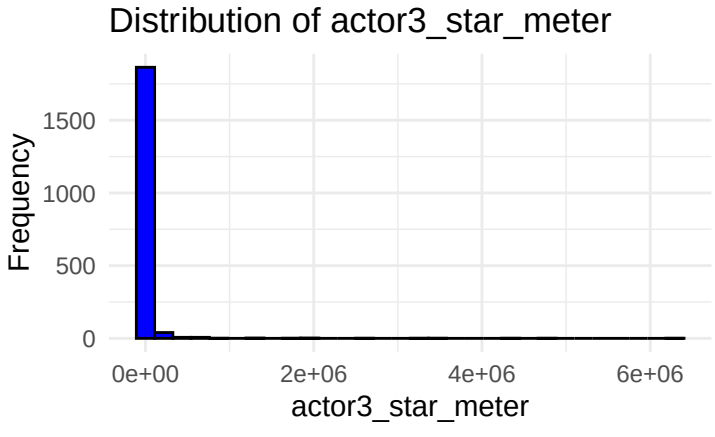
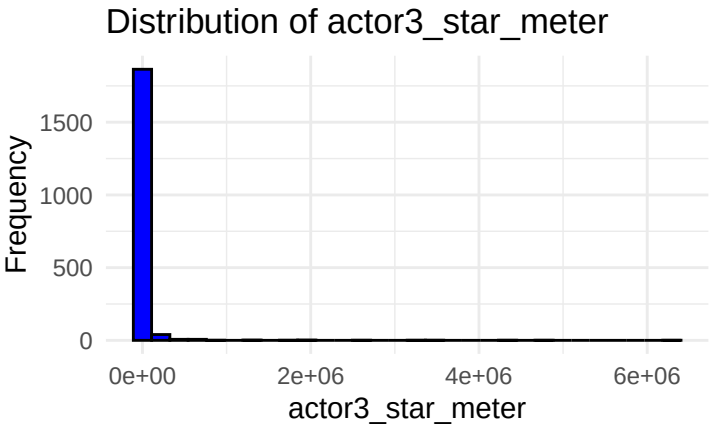

Distribution of movie_budget
Frequency
0
50
100
0e+00
2e+07
4e+07
movie_budget
Distribution of release_day
Frequency
0
30
60
90
0
10
20
30
release_day
Distribution of release_year
Frequency
0
100
200
1950
1975
2000
release_year
Distribution of duration
Frequency
0
200
400
100
200
300
duration

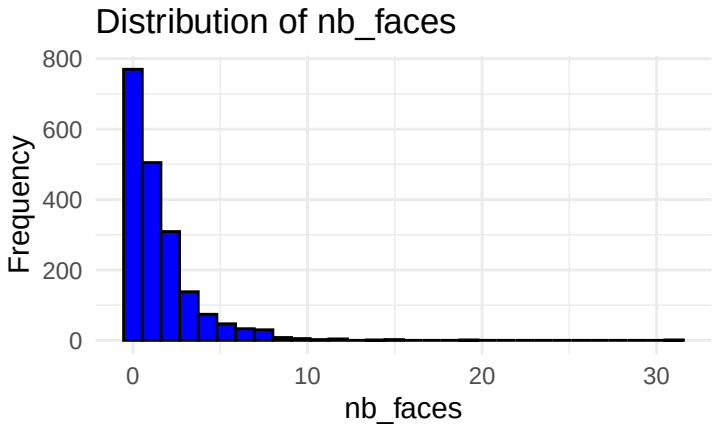

Distribution of nb_news_articles
Frequency
1500
1000
500
0
0
20000
40000
60000
nb_news_articles

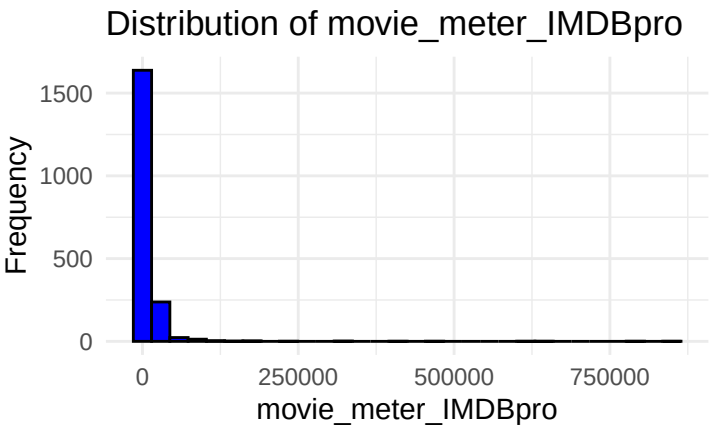

Distribution of actor1_star_meter
Frequency
2000
1500
1000
500
0
0e+00
2e+06
4e+06
6e+06
8e+06
actor1_star_meter


Distribution of actor2_star_meter
Frequency
1500
1000
500
0
0e+00
2e+06
4e+06
actor2_star_meter

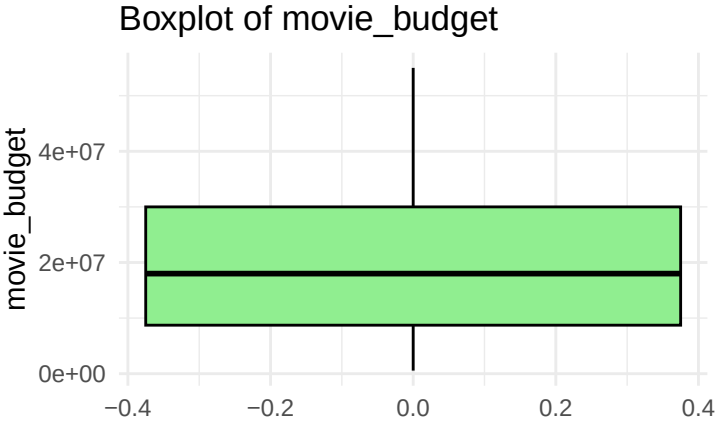

Distribution of actor3_star_meter
Frequency
1500
1000
500
0
0e+00
2e+06
4e+06
6e+06
actor3_star_meter

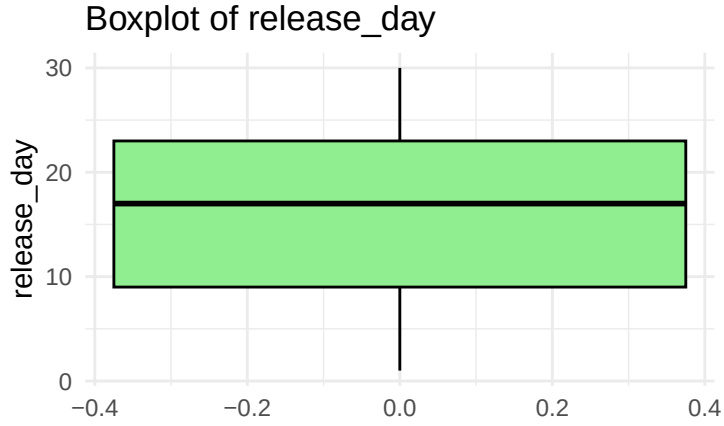

Distribution of actor3_star_meter
Frequency
1500
1000
500
0
0e+00
2e+06
4e+06
6e+06
actor3_star_meter

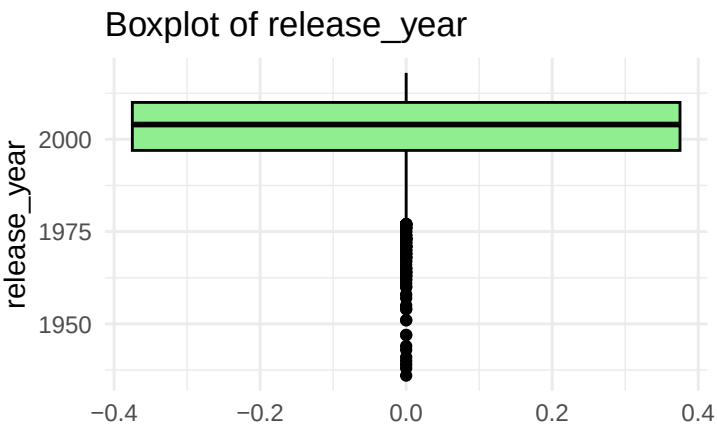

Distribution of nb_faces
Frequency
800
600
400
200
0
0
10
20
30
nb_faces

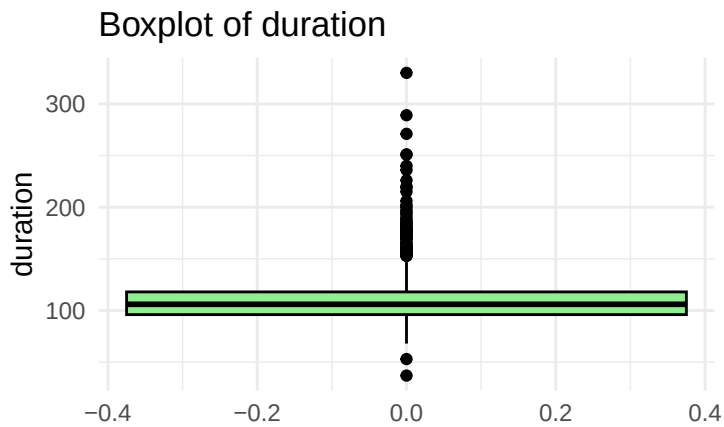

Distribution of movie_meter_IMDBpro
Frequency
1500
1000
500
0
0
250000
500000
750000
movie_meter_IMDBpro

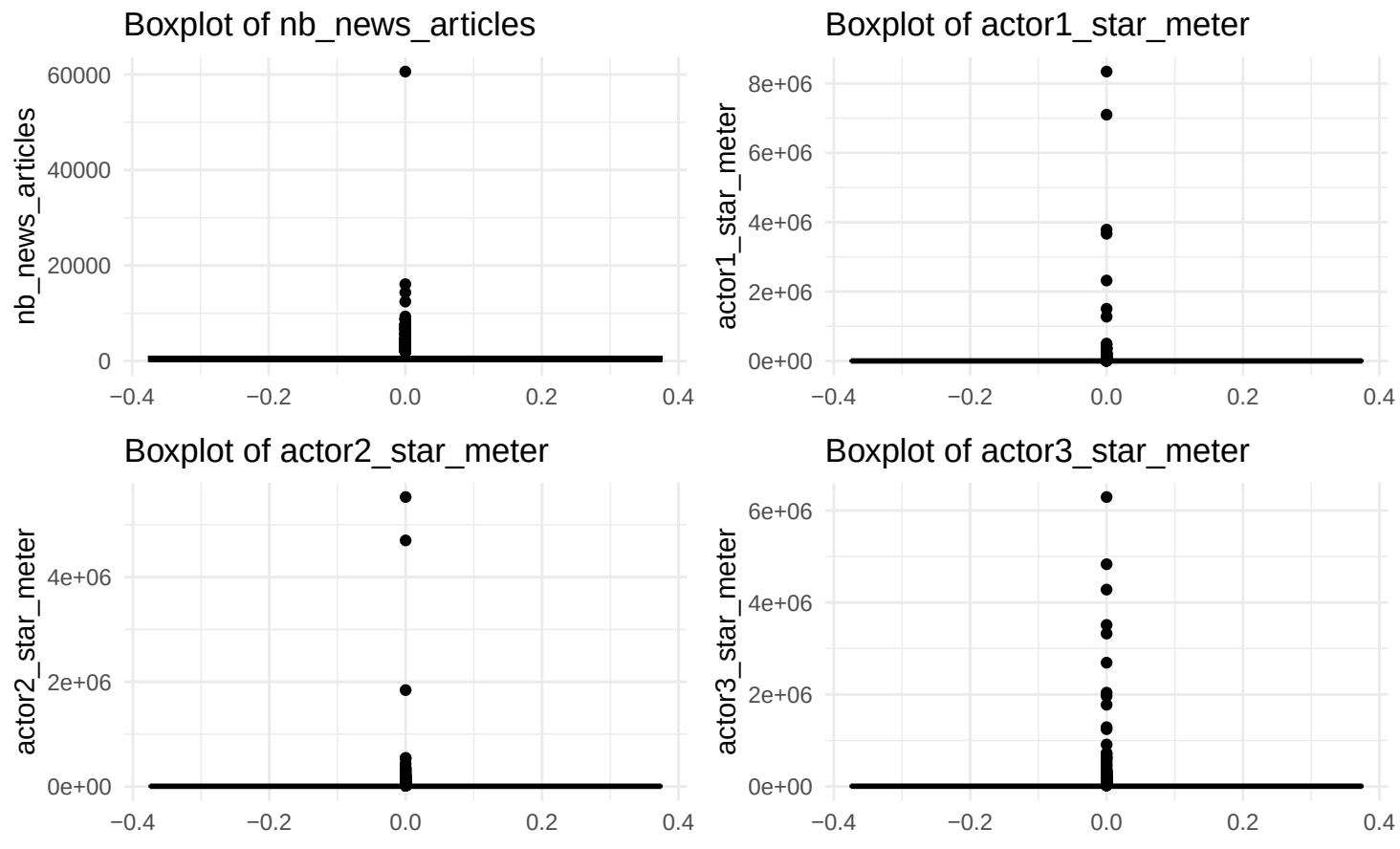
	Predictor	Skewness	Kurtosis
movie_budget	movie_budget	0.5379464	-0.7262042
release_day	release_day	-0.1666258	-1.0893363
release_year	release_year	-1.6267935	3.8669948
duration	duration	2.6770820	15.5853124
aspect_ratio	aspect_ratio	-0.3718203	-1.0763686
nb_news_articles	nb_news_articles	18.6088704	558.9580715
actor1_star_meter	actor1_star_meter	23.2678536	594.7305886
actor2_star_meter	actor2_star_meter	27.5774777	810.6270569
actor3_star_meter	actor3_star_meter	16.2897374	306.3312994
nb_faces	nb_faces	3.7344758	29.8917997
movie_meter_IMDBpro	movie_meter_IMDBpro	14.5328325	249.5463930

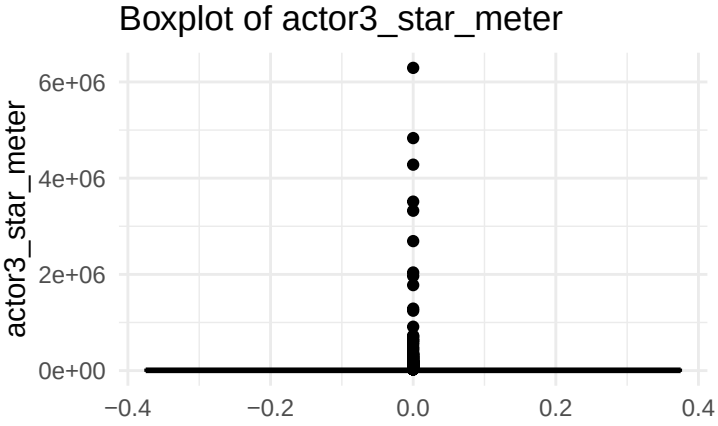

Boxplot of movie_budget
movie_budget
4e+07
2e+07
0e+00
−0.4
−0.2
0.0
0.2
0.4

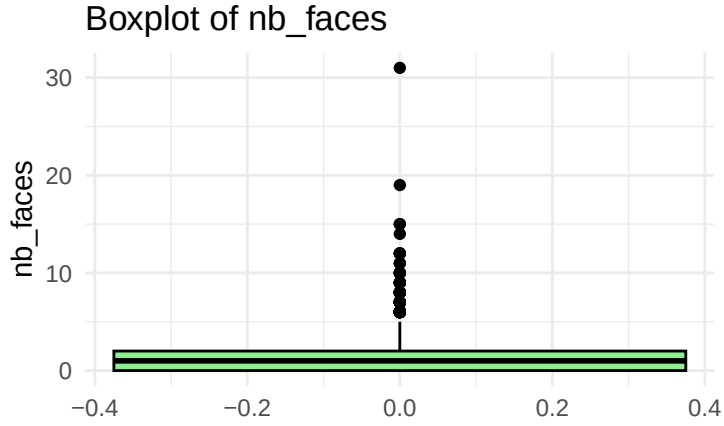

Boxplot of release_day
release_day
30
20
10
0
−0.4
−0.2
0.0
0.2
0.4

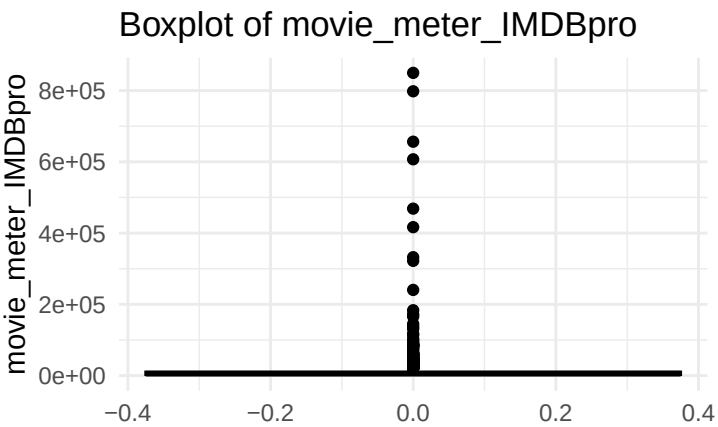

Boxplot of release_year
release_year
2000
1975
1950
−0.4
−0.2
0.0
0.2
0.4


Boxplot of duration
duration
300
200
100
−0.4
−0.2
0.0
0.2
0.4

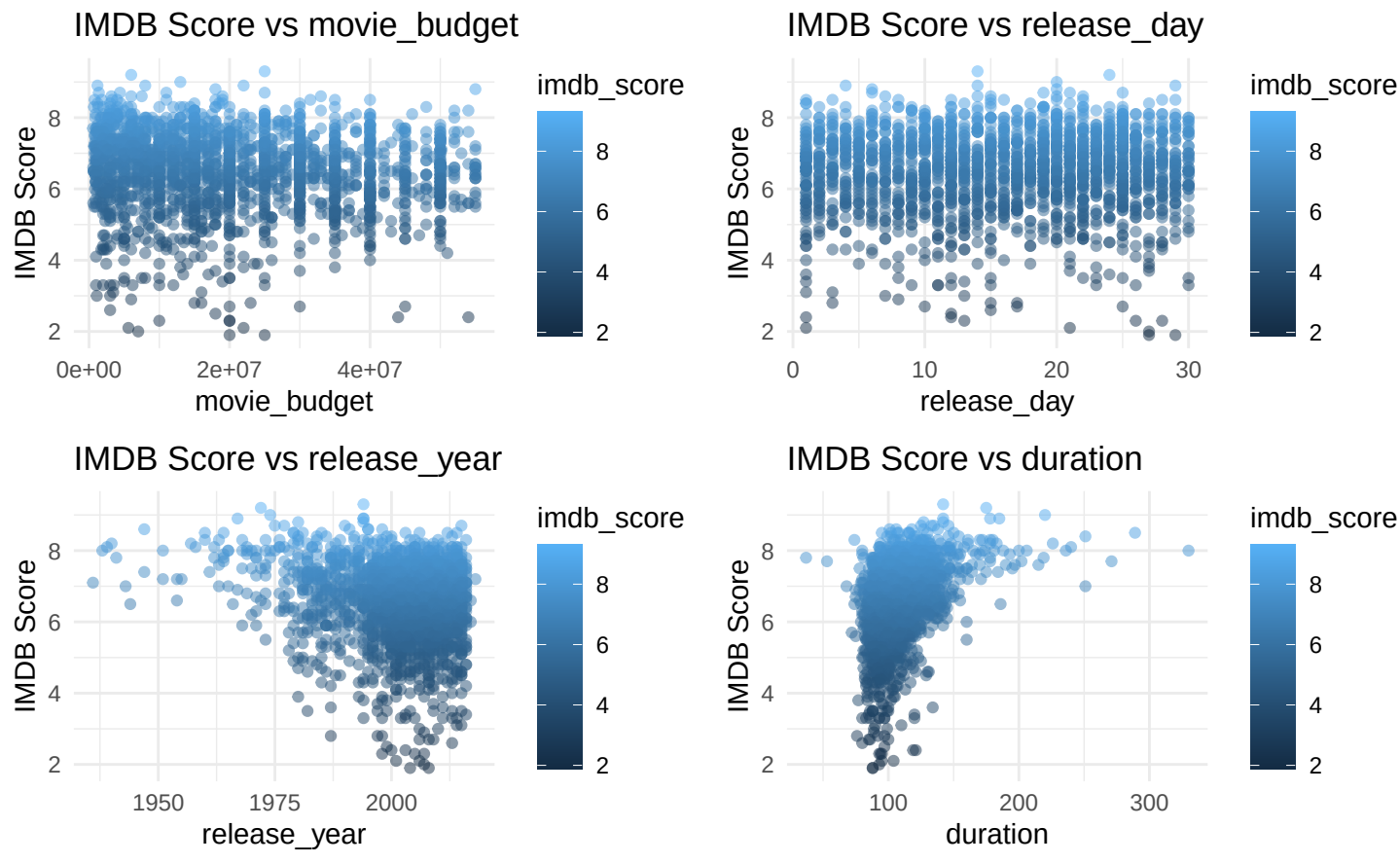

Boxplot of nb_news_articles
nb_news_articles
60000
40000
20000
0
−0.4
−0.2
0.0
0.2
0.4
Boxplot of actor1_star_meter
actor1_star_meter
8e+06
6e+06
4e+06
2e+06
0e+00
Boxplot of actor2_star_meter
actor2_star_meter
4e+06
2e+06
0e+00
Boxplot of actor3_star_meter
actor3_star_meter
6e+06
4e+06
2e+06
0e+00

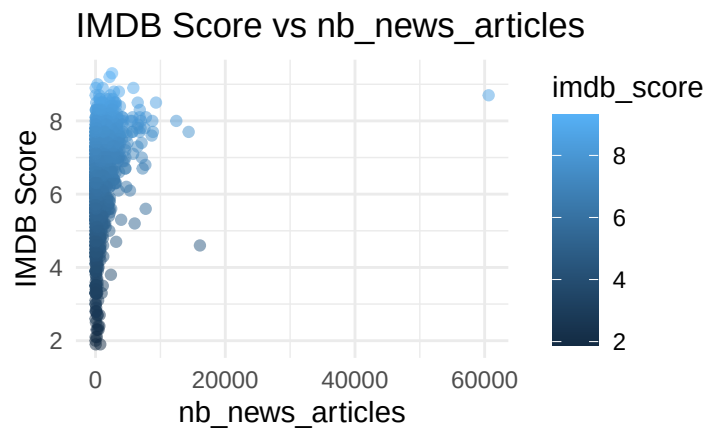

Boxplot of actor3_star_meter
actor3_star_meter
6e+06
4e+06
2e+06
0e+00
−0.4
−0.2
0.0
0.2
0.4

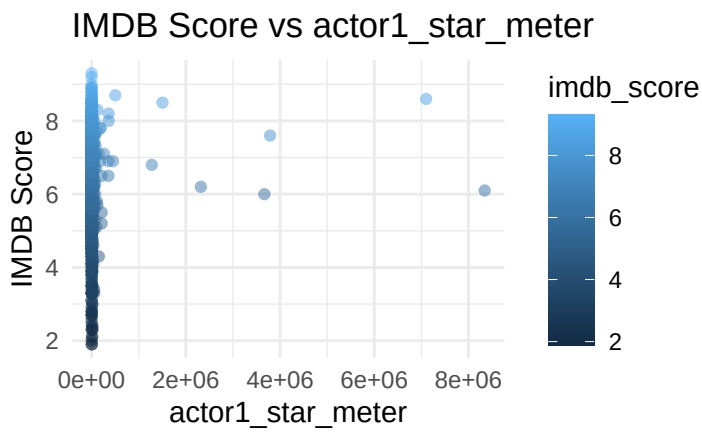

Boxplot of nb_faces
nb_faces
30
20
10
0
−0.4
−0.2
0.0
0.2
0.4

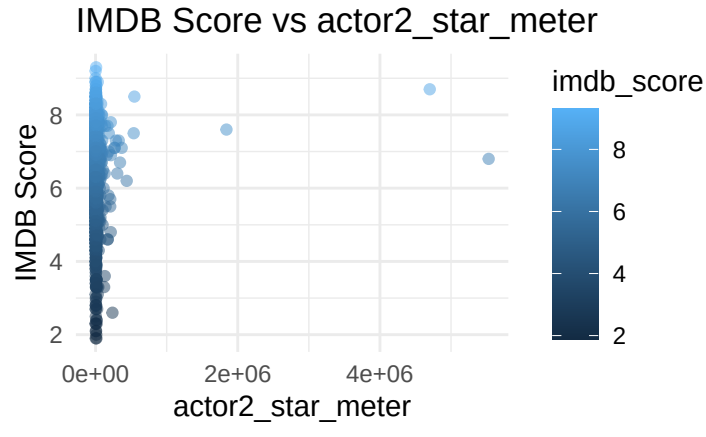

Boxplot of movie_meter_IMDBpro
movie_meter_IMDBpro
8e+05
6e+05
4e+05
2e+05
0e+00
−0.4
−0.2
0.0
0.2
0.4

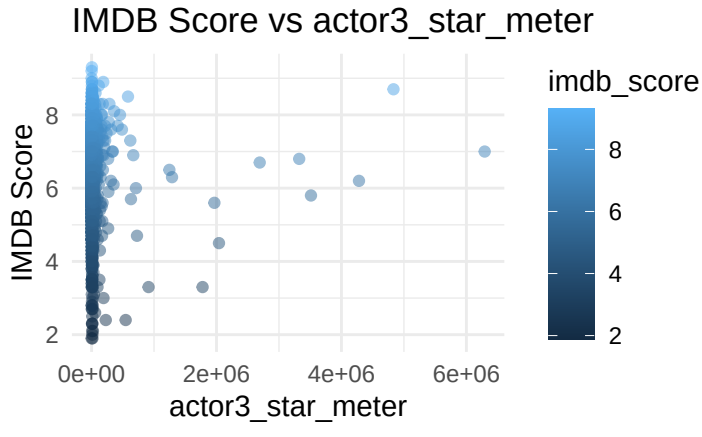
```
 movie_budget       release_day     release_year     duration      aspect_ratio
Min.   :  560000   Min.   : 1.00   Min.   :1936   Min.   : 37.0   Min.   :1.180
1st Qu.: 8725000   1st Qu.: 9.00   1st Qu.:1997   1st Qu.: 96.0   1st Qu.:1.850
Median :18000000   Median :17.00   Median :2004   Median :106.0   Median :2.350
Mean   :20973774   Mean   :15.95   Mean   :2001   Mean   :109.7   Mean   :2.096
3rd Qu.:30000000   3rd Qu.:23.00   3rd Qu.:2010   3rd Qu.:118.0   3rd Qu.:2.350
Max.   :55000000   Max.   :30.00   Max.   :2018   Max.   :330.0   Max.   :2.760
nb_news_articles  actor1_star_meter actor2_star_meter actor3_star_meter
Min.   :    0.0   Min.   :      9   Min.   :      3   Min.   :      8
1st Qu.:   78.0   1st Qu.:    505   1st Qu.:   1895   1st Qu.:   3075
Median :  286.0   Median :   1888   Median :   3986   Median :   5856
Mean   :  770.6   Mean   :  21190   Mean   :  17114   Mean   :  35469
3rd Qu.:  845.5   3rd Qu.:   4665   3rd Qu.:   7667   3rd Qu.:  12250
Max.   :60620.0   Max.   :8342201   Max.   :5529461   Max.   :6292982
   nb_faces     movie_meter_IMDBpro
Min.   : 0.00   Min.   :    71
1st Qu.: 0.00   1st Qu.:  2836
Median : 1.00   Median :  5406
Mean   : 1.44   Mean   : 11612
3rd Qu.: 2.00   3rd Qu.: 10198
Max.   :31.00   Max.   :849550
```

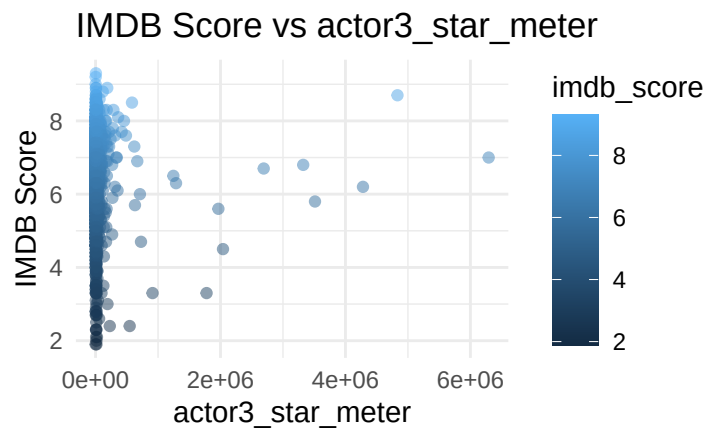

IMDB Score vs movie_budget
IMDB Score
movie_budget
0e+00
2e+07
4e+07
imdb_score
IMDB Score vs release_day
release_day
0
10
20
30
IMDB Score vs release_year
release_year
1950
1975
2000
IMDB Score vs duration
duration
100
200
300
2
4
6
8

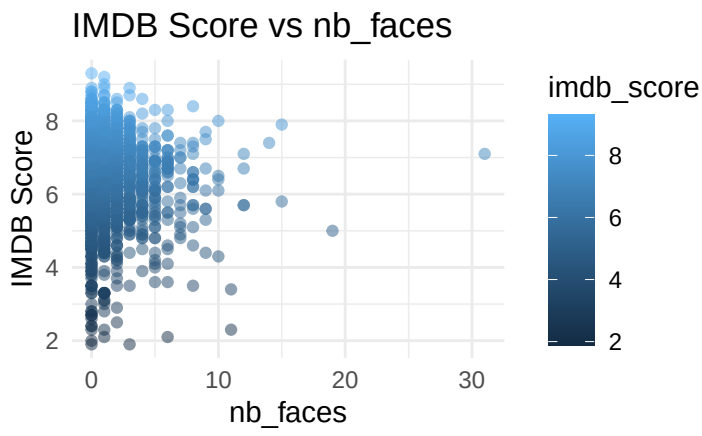

IMDB Score vs nb_news_articles
IMDB Score
nb_news_articles
imdb_score
0
20000
40000
60000
2
4
6
8

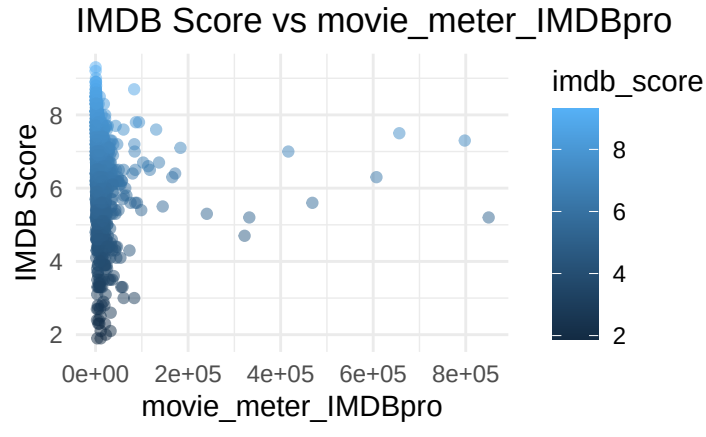

IMDB Score vs actor1_star_meter
IMDB Score
actor1_star_meter
imdb_score
0e+00
2e+06
4e+06
6e+06
8e+06
2
4
6
8


IMDB Score vs actor2_star_meter
IMDB Score
actor2_star_meter
imdb_score
0e+00
2e+06
4e+06
2
4
6
8


IMDB Score vs actor3_star_meter
IMDB Score
actor3_star_meter
imdb_score
0e+00
2e+06
4e+06
6e+06
2
4
6
8


IMDB Score vs actor3_star_meter
IMDB Score
8
6
4
2
0e+00
2e+06
4e+06
6e+06
actor3_star_meter
imdb_score
8
6
4
2


IMDB Score vs nb_faces
IMDB Score
8
6
4
2
0
10
20
30
nb_faces
imdb_score
8
6
4
2

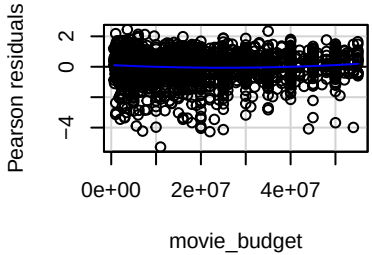

IMDB Score vs movie_meter_IMDBpro
IMDB Score
8
6
4
2
0e+00
2e+05
4e+05
6e+05
8e+05
movie_meter_IMDBpro
imdb_score
8
6
4
2

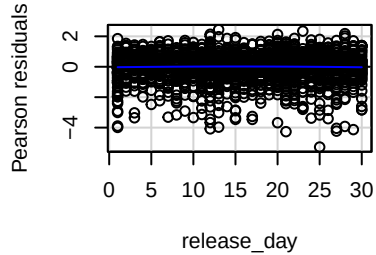
1	movie_budget	5.417691e-04	6.188867e-03
2	release_day	3.488141e-01	4.552876e-04
3	release_month	2.907435e-01	2.606009e-02
4	release_year	5.812340e-18	3.795539e-02
5	duration	2.124863e-79	1.686284e-01
6	language	7.963938e-01	1.646691e-02
7	country	1.725032e-01	3.767972e-02
8	maturity_rating	9.261149e-04	4.739582e-02
9	aspect_ratio	6.245747e-01	1.242413e-04
10	distributor	1.166891e-02	2.567400e-01
11	nb_news_articles	1.154764e-23	5.082841e-02
12	director	7.382646e-03	7.821907e-01
13	actor1	8.795616e-01	6.359676e-01
14	actor1_star_meter	2.039799e-01	8.368206e-04
15	actor2	6.181009e-01	7.677719e-01
16	actor2_star_meter	9.276008e-02	1.464975e-03
17	actor3	1.151085e-01	8.562379e-01
18	actor3_star_meter	8.581515e-01	1.657242e-05
19	colour_film	1.187884e-12	2.586263e-02
20	genres	9.474060e-02	4.331155e-01
21	nb_faces	8.394547e-05	7.992422e-03
22	plot_keywords	NaN	1.000000e+00
23	action	2.096054e-12	2.529932e-02
24	adventure	3.282087e-03	4.474328e-03
25	scifi	3.669494e-05	8.798989e-03
26	thriller	4.323968e-04	6.405711e-03
27	musical	3.198531e-01	5.132520e-04
28	romance	5.134628e-01	2.215080e-04
29	western	3.974098e-03	4.294571e-03
30	sport	1.566731e-02	3.025159e-03
31	horror	2.102335e-13	2.757971e-02
32	drama	7.482679e-53	1.143819e-01
33	war	1.756616e-06	1.178186e-02
34	animation	4.666386e-01	2.748906e-04
35	crime	6.930645e-03	3.775400e-03
36	movie_meter_IMDBpro	7.897205e-05	8.051840e-03
37	cinematographer	8.280564e-01	5.487727e-01
38	production_company	2.643196e-02	5.158081e-01

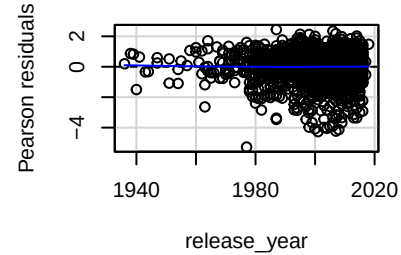
```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 8.574293, Df = 1, p = 0.0034094
```

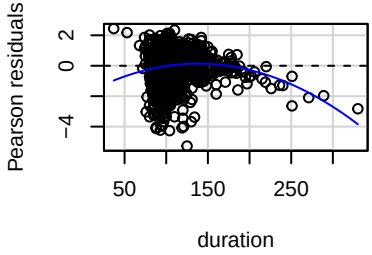
```
Call:
lm(formula = formula_numeric, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-5.2782 -0.4624  0.1466  0.6354  2.4424

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          2.101e+01  4.021e+00   5.225 1.93e-07 ***
movie_budget        -1.148e-08  1.585e-09  -7.244 6.28e-13 ***
release_day          1.685e-03  2.625e-03   0.642 0.521022
release_year        -8.353e-03  2.014e-03  -4.148 3.50e-05 ***
duration             2.050e-02  1.096e-03  18.706  < 2e-16 ***
aspect_ratio         8.328e-02  8.451e-02   0.985 0.324524
nb_news_articles     1.130e-04  1.185e-05   9.535  < 2e-16 ***
actor1_star_meter    9.566e-08  7.749e-08   1.235 0.217154
actor2_star_meter    1.513e-07  1.340e-07   1.129 0.258948
actor3_star_meter   -2.264e-08  8.765e-08  -0.258 0.796165
nb_faces            -4.055e-02  1.060e-02  -3.824 0.000136 ***
movie_meter_IMDBpro -1.732e-06  5.479e-07  -3.161 0.001596 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9555 on 1918 degrees of freedom
Multiple R-squared:  0.2497,    Adjusted R-squared:  0.2454
F-statistic: 58.04 on 11 and 1918 DF,  p-value: < 2.2e-16
```

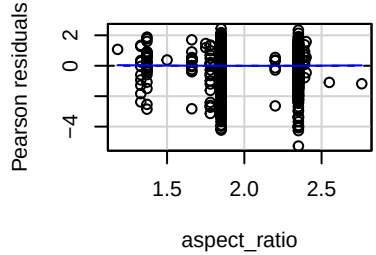
```
t test of coefficients:

                    Estimate  Std. Error t value  Pr(>|t|)
(Intercept)        2.1008e+01  4.6352e+00  4.5322 6.199e-06 ***
movie_budget      -1.1482e-08  1.5227e-09 -7.5409 7.161e-14 ***
release_day        1.6849e-03  2.6749e-03  0.6299 0.5288339
release_year      -8.3530e-03  2.3290e-03 -3.5866 0.0003434 ***
duration           2.0501e-02  1.5301e-03 13.3983 < 2.2e-16 ***
aspect_ratio       8.3280e-02  8.6971e-02  0.9576 0.3384018
nb_news_articles   1.1303e-04  4.9917e-05  2.2643 0.0236690 *
actor1_star_meter  9.5660e-08  6.3152e-08  1.5148 0.1300000
actor2_star_meter  1.5131e-07  1.6960e-07  0.8921 0.3724349
actor3_star_meter -2.2645e-08  7.9944e-08 -0.2833 0.7770111
nb_faces          -4.0548e-02  1.1473e-02 -3.5344 0.0004184 ***
movie_meter_IMDBpro -1.7320e-06  8.5195e-07 -2.0330 0.0421924 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

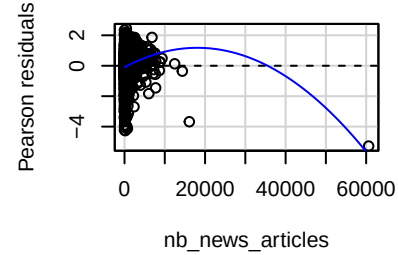

Pearson residuals
−4
0
2
0e+00
2e+07
4e+07
movie_budget

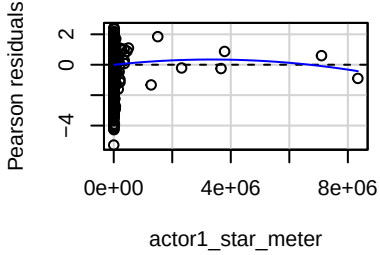

Pearson residuals
−4
0
2
0
5
10
15
20
25
30
release_day

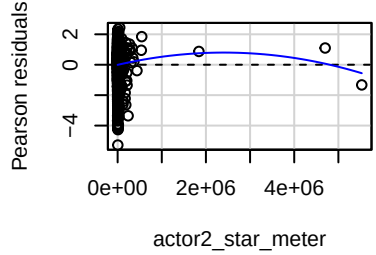

Pearson residuals
−4
0
2
1940
1980
2020
release_year

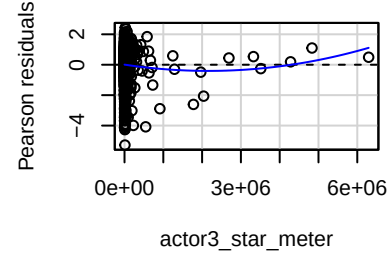

Pearson residuals
−4
0
2
50
150
250
duration

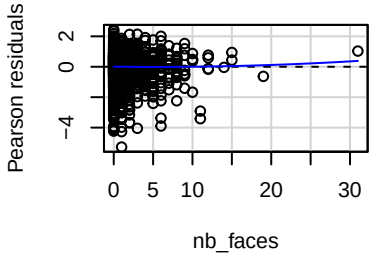

Pearson residuals
−4
0
2
1.5
2.0
2.5
aspect_ratio

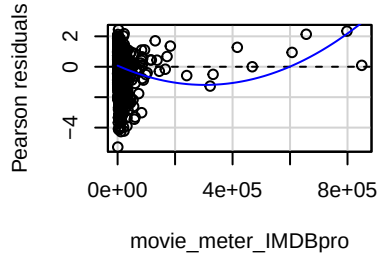

Pearson residuals
−4
0
2
0
20000
40000
60000
nb_news_articles

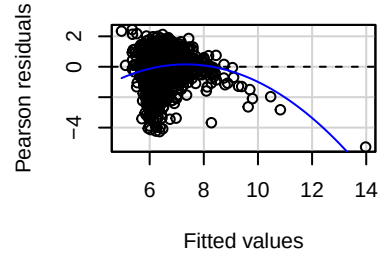

Pearson residuals
−4
0
2
0e+00
4e+06
8e+06
actor1_star_meter


Pearson residuals
−4
0
2
0e+00
2e+06
4e+06
actor2_star_meter

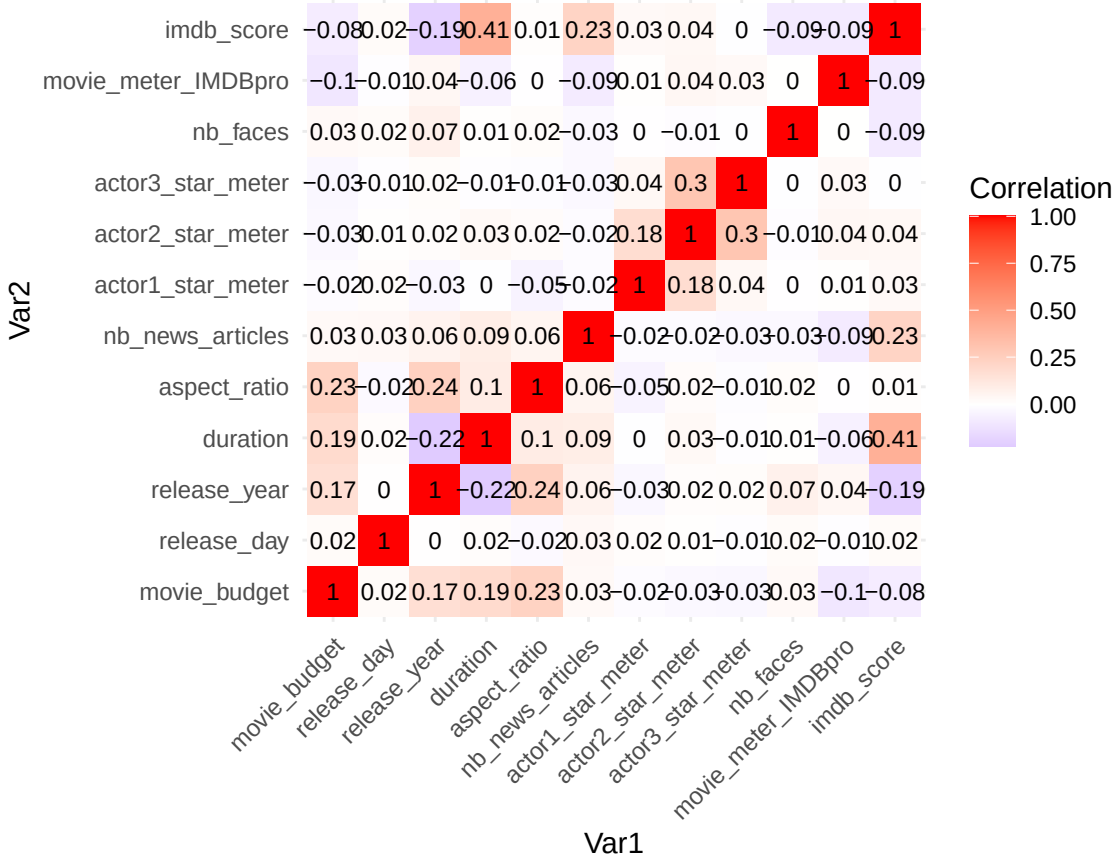

Pearson residuals
−4
0
2
0e+00
3e+06
6e+06
actor3_star_meter


Pearson residuals
−4
0
2
0
5
10
20
30
nb_faces


Pearson residuals
−4
0
2
0e+00
4e+05
8e+05
movie_meter_IMDBpro


Pearson residuals
−4
0
2
6
8
10
12
14
Fitted values

```
                    Test stat Pr(>|Test stat|)
movie_budget           2.7960          0.005226 **
release_day           -0.6533          0.513635
release_year           0.4243          0.671388
duration              -7.4554         1.349e-13 ***
aspect_ratio           0.1851          0.853141
nb_news_articles      -9.4878         < 2.2e-16 ***
actor1_star_meter     -0.9317          0.351608
actor2_star_meter     -1.4345          0.151589
actor3_star_meter      1.8199          0.068937 .
nb_faces               0.4940          0.621382
movie_meter_IMDBpro    6.7627         1.792e-11 ***
Tukey test           -11.8803         < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```


imdb_score −0.08 0.02 −0.19 0.41 0.01 0.23 0.03 0.04 0 −0.09 −0.09 1
movie_meter_IMDBpro −0.1 −0.01 0.04 −0.06 0 −0.09 0.01 0.04 0.03 0 1 −0.09
nb_faces 0.03 0.02 0.07 0.01 0.02 −0.03 0 −0.01 0 1 0 −0.09
actor3_star_meter −0.03 −0.01 0.02 −0.01 −0.01 −0.03 0.04 0.3 1 0 0.03 0
actor2_star_meter −0.03 0.01 0.02 0.03 0.02 −0.02 0.18 1 0.3 −0.01 0.04 0.04
actor1_star_meter −0.02 0.02 −0.03 0 −0.05 −0.02 1 0.18 0.04 0 0.01 0.03
nb_news_articles 0.03 0.03 0.06 0.09 0.06 1 −0.02 −0.02 −0.03 −0.03 −0.09 0.23
aspect_ratio 0.23 −0.02 0.24 0.1 1 0.06 −0.05 0.02 −0.01 0.02 0 0.01
duration 0.19 0.02 −0.22 1 0.1 0.09 0 0.03 −0.01 0.01 −0.06 0.41
release_year 0.17 0 1 −0.22 0.24 0.06 −0.03 0.02 0.02 0.07 0.04 −0.19
release_day 0.02 1 0 0.02 −0.02 0.03 0.02 0.01 −0.01 0.02 −0.01 0.02
movie_budget 1 0.02 0.17 0.19 0.23 0.03 −0.02 −0.03 −0.03 0.03 −0.1 −0.08
Var2
movie_budget
release_day
release_year
duration
aspect_ratio
nb_news_articles
actor1_star_meter
actor2_star_meter
actor3_star_meter
nb_faces
movie_meter_IMDBpro
imdb_score
Var1
Correlation
1.00
0.75
0.50
0.25
0.00