Fantastic movies
Where to find them?

Alexis Angel, Lin Zuo, Amy Xiong

Definition of popularity/success of a movie

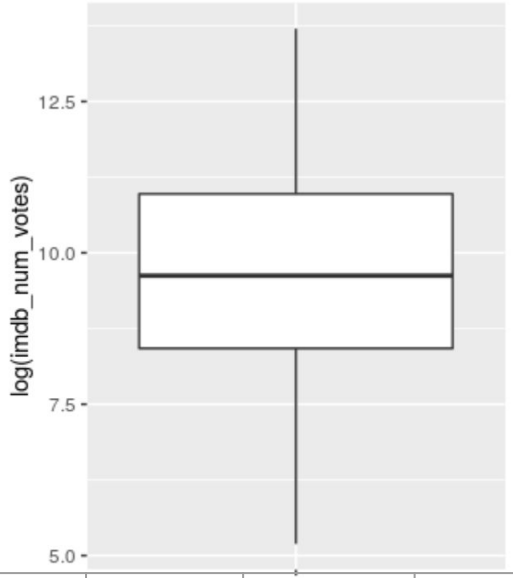
Three criteria:

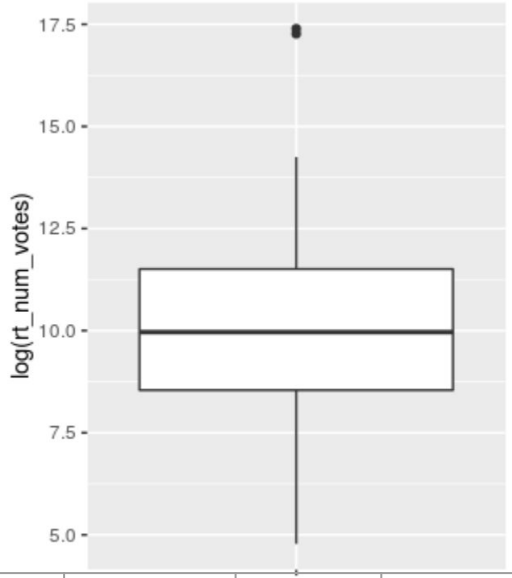
1. Public rating
2. Oscar awards
3. Profitability

Criterion # 1: Public Rating

Insight 1:

Which public rating is the most reliable/influential?





	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
IMDB	180	4546	15120	57530	58300	893000
RT	119	5133	21260	203300	99570	35790000

Criterion # 1: Public Rating

Insight 2: What are meaningful predictors of public rating?

-Linear model with stepwise selection

```
# Linear model for predicting movies' rating
  # removed title, director, actor1-5, urls, audience_rating, imdb_rating, critics_rating,
critics_score, thtr_rel_year, dvd_rel_year
movies_modified = na.omit(movies) %>% select(title_type, genre, runtime, mpaa_rating, studio,
thtr_rel_month, thtr_rel_day, dvd_rel_month, dvd_rel_day, imdb_num_votes,  audience_score, best_pic_nom,
best_pic_win, best_actor_win, best_actress_win, top200_box)

step(lm((audience_score)~., data=movies_modified), direction = "backward")

summary(lm(formula = (audience_score) ~ genre + mpaa_rating + dvd_rel_month +
    imdb_num_votes + best_pic_nom, data = movies_modified))$r.squared
```

The r-squared value of this final model is 0.345, meaning this model explains about 34.5% of the variability of public rating.

Since bribing the critics could be an option...

```
model = lm(critics_score ~ audience_score, data=movies)
summary(model)$r.squared
```

The r-squared value is 0.496

We calculate a second model where critics score and rating are candidate predictors:

```
summary(lm(formula = (audience_score) ~ genre + dvd_rel_day + imdb_num_votes +
    critics_score, data = movies_modified))$r.squared
```

-Meaningful predictors change

-The r-squared value increases from 0.345 to 0.563

Criteria #2: Oscar Awards

Insight 1: We are 95% confident that movies that won Oscars are generally 11.7 to 18.6 minutes longer than movies that didn't win Oscars.

```
data:  movies$runtime by movies$award
t = -8.6738, df = 266.09, p-value = 4.212e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -18.56375 -11.69510
sample estimates:
 mean in group no mean in group yes
        101.8413           116.9708
```

Criteria #2: Oscar Awards

Insight 2: We are 95% confident that movies with oscar awards are 0.07 to 0.4 higher in IMDB rating than movies without oscar awards.

```
data:  oscar_ratings$imdb_rating by oscar_ratings$award
t = -2.7392, df = 361.43, p-value = 0.006465
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.41260731 -0.06774357
sample estimates:
 mean in group no mean in group yes
        6.430000          6.670175
```

Criteria #2: Oscar Awards

Insight 3: The profitability is unrelated to the possibility to win an Oscar.

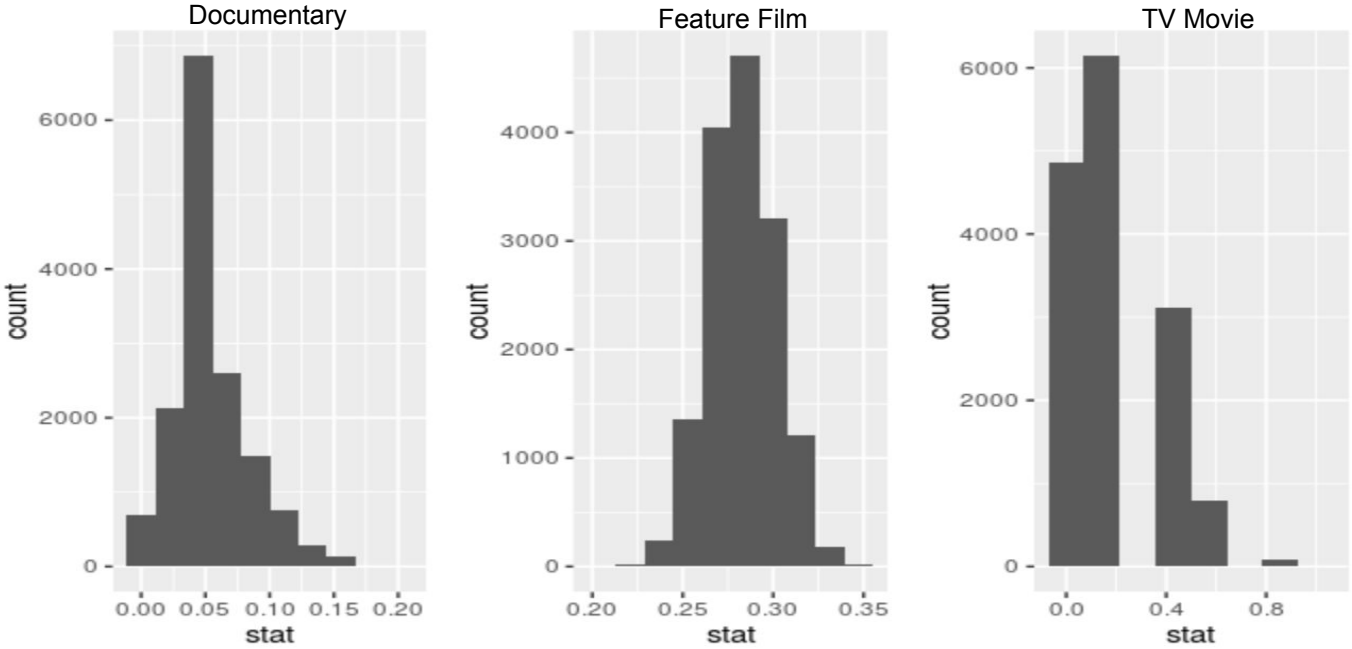
```
     profitable unprofitable Sum
no          132          127 259
yes          68           52 120
Sum         200          179 379

        2-sample test for equality of proportions without continuity correction

data:  c(oscar_roi_summ$x[1], oscar_roi_summ$x[2]) out of c(oscar_roi_summ$n[1],
oscar_roi_summ$n[2])
X-squared = 1.0696, df = 1, p-value = 0.301
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.04395542  0.14294984
sample estimates:
   prop 1    prop 2
0.3400000 0.2905028
```

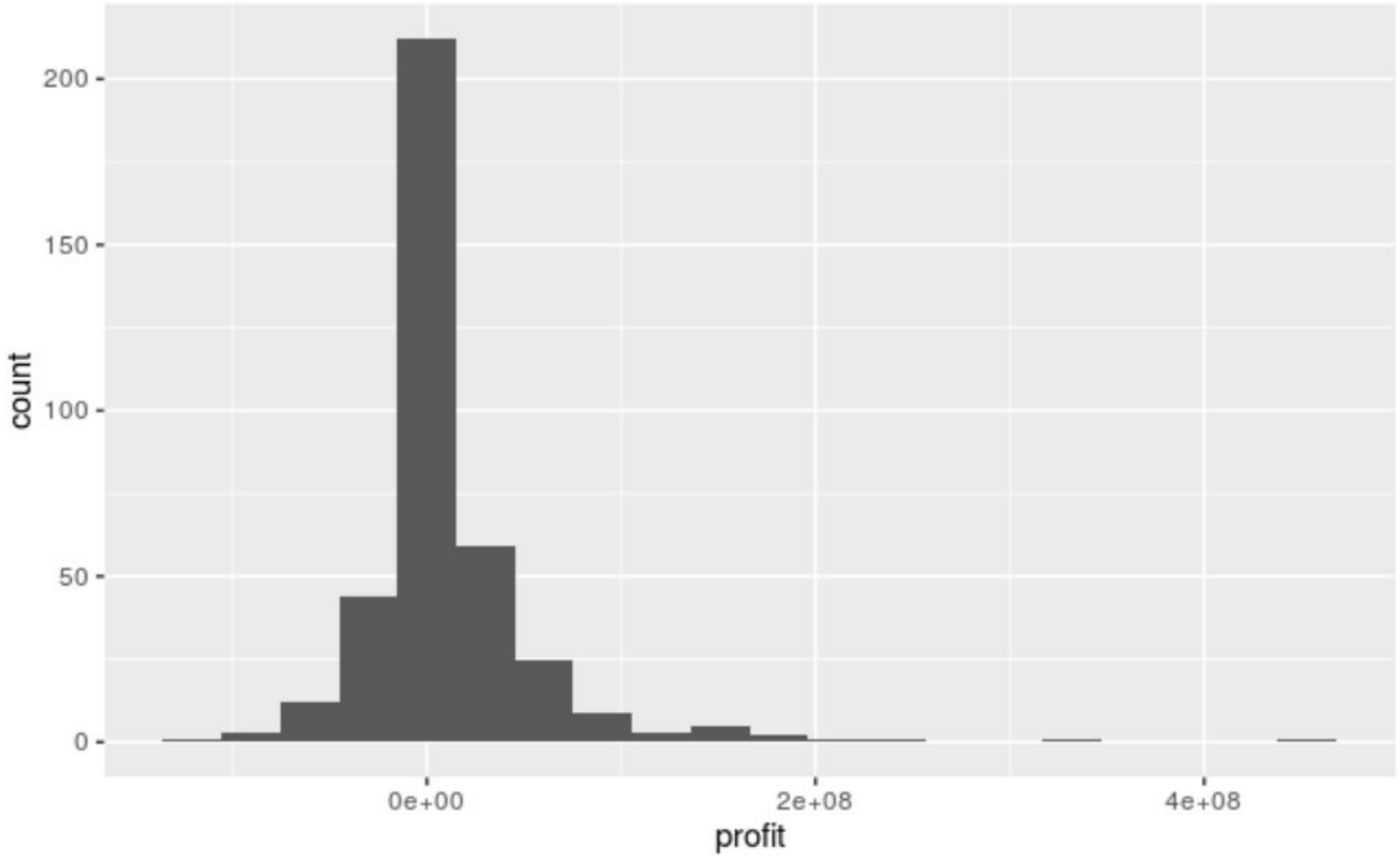
Criteria #2: Oscar Awards

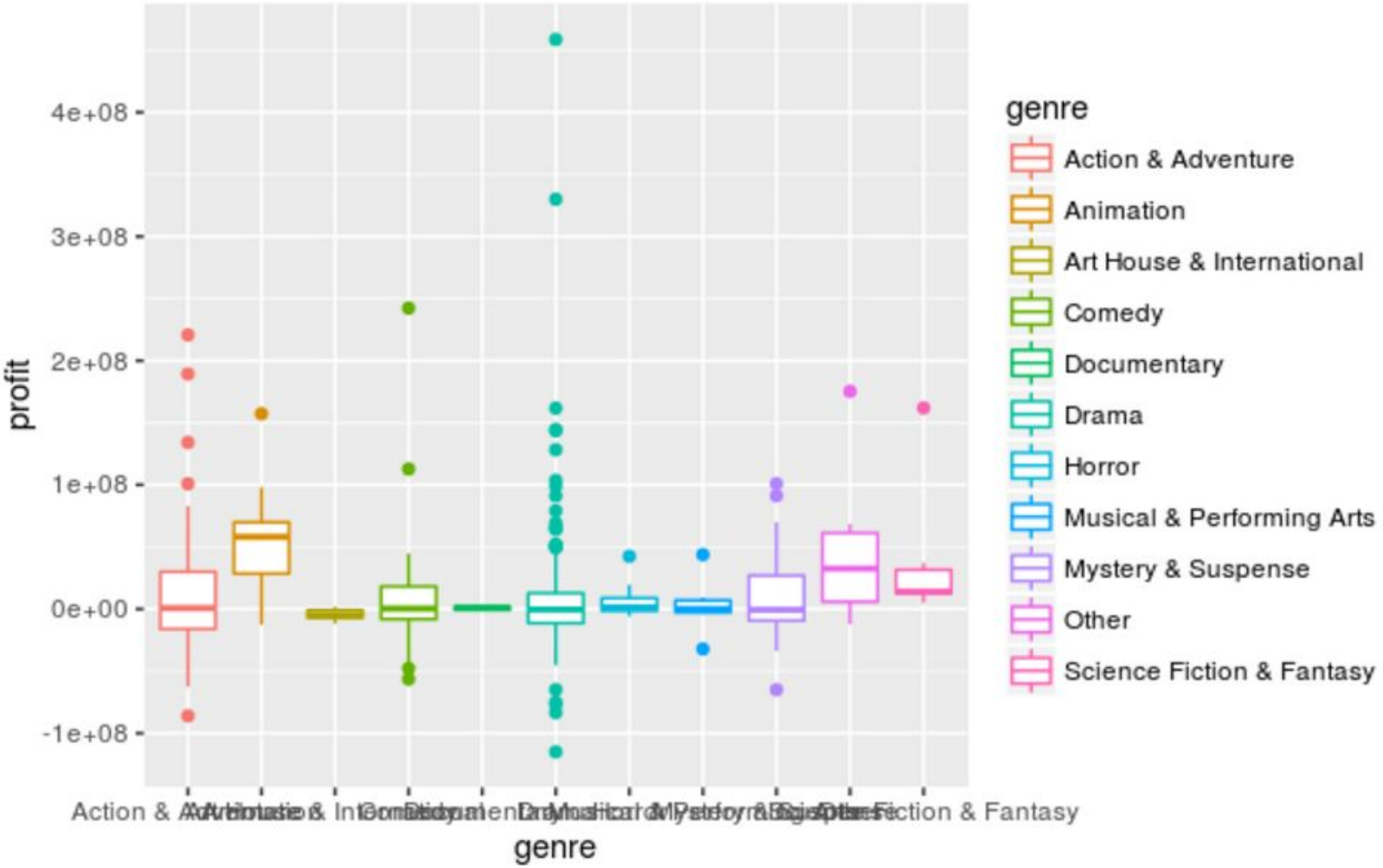
Insight 4: Feature films are qualitatively more likely to win Oscar awards.

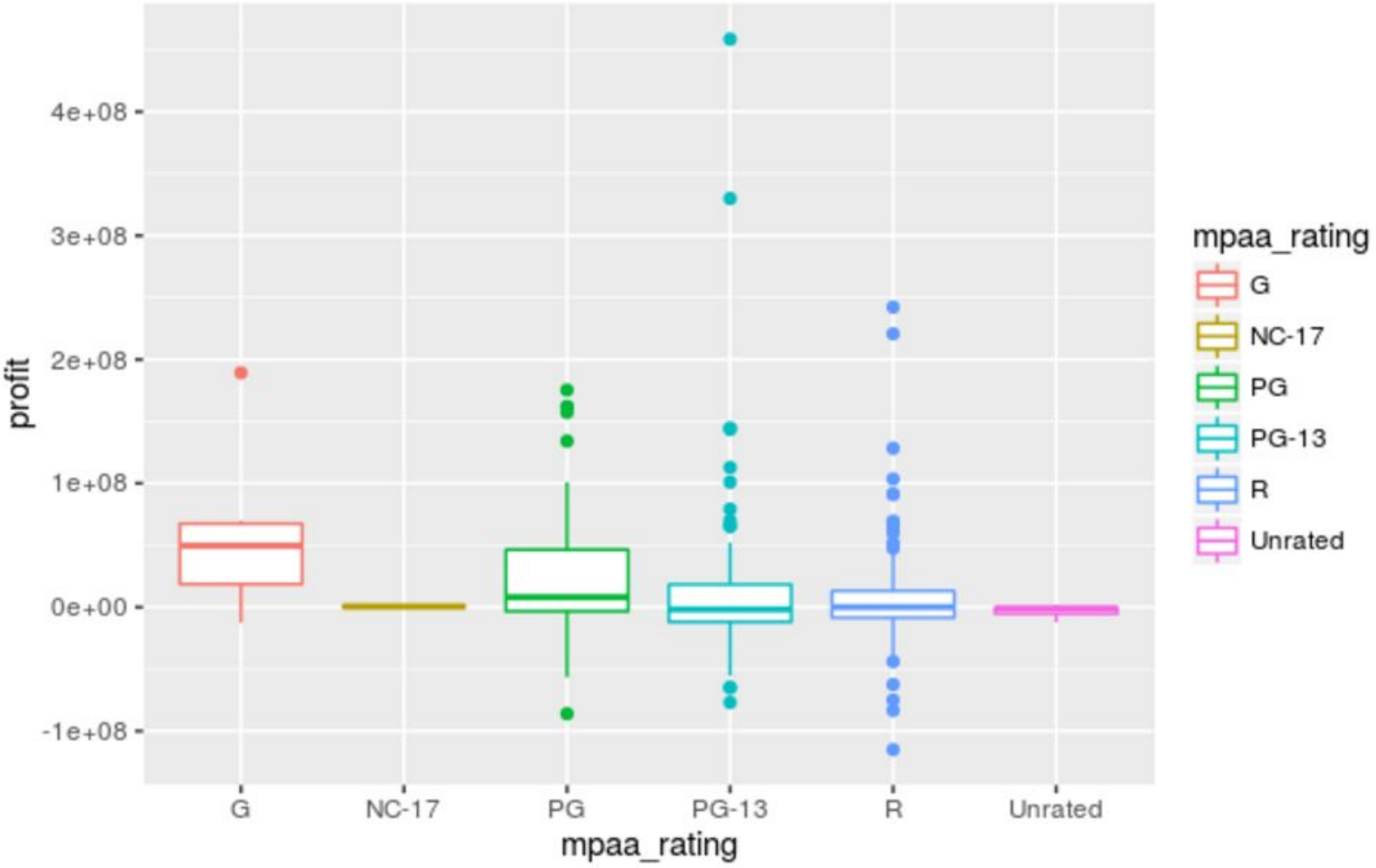


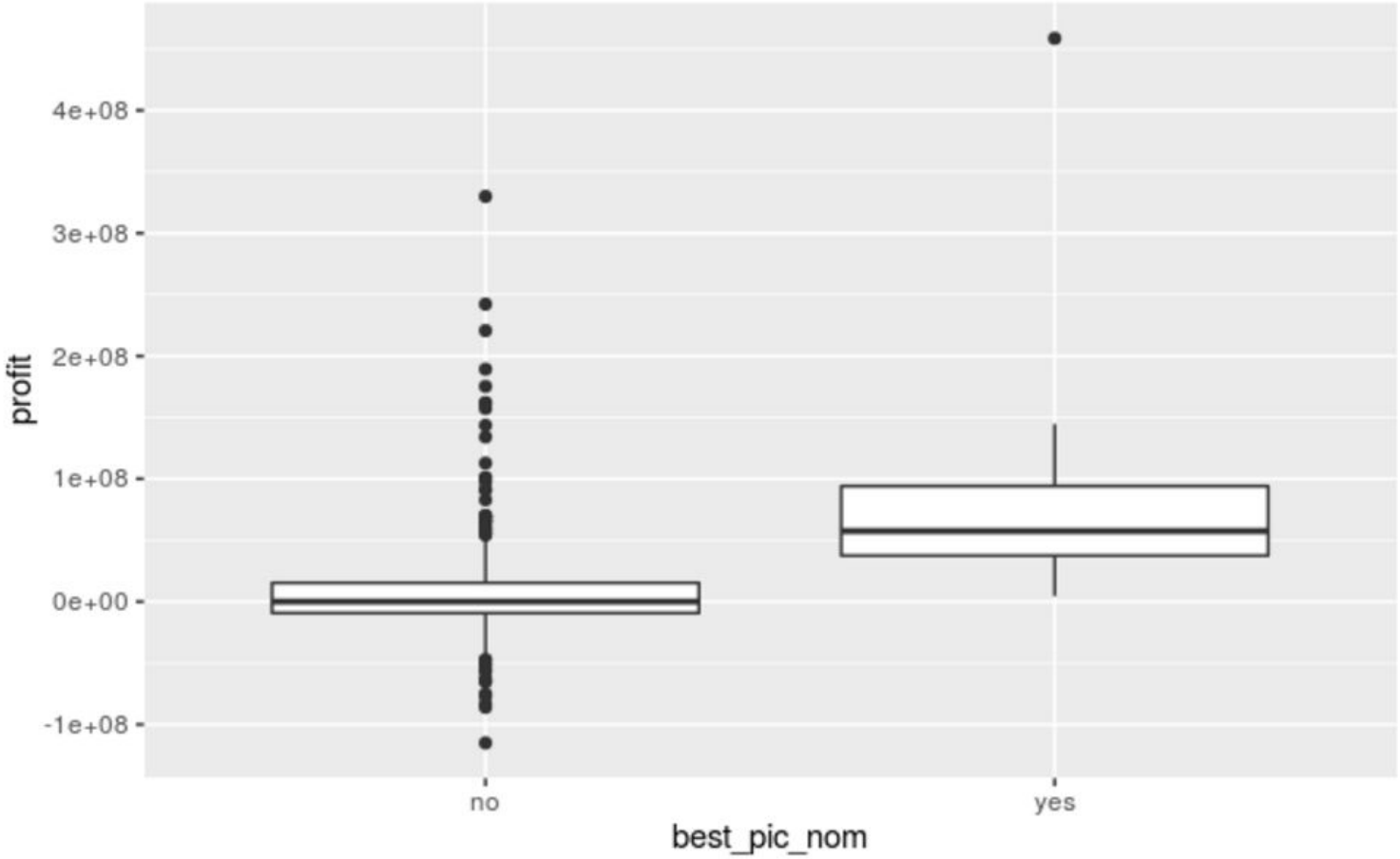
Criteria #3: Profitability

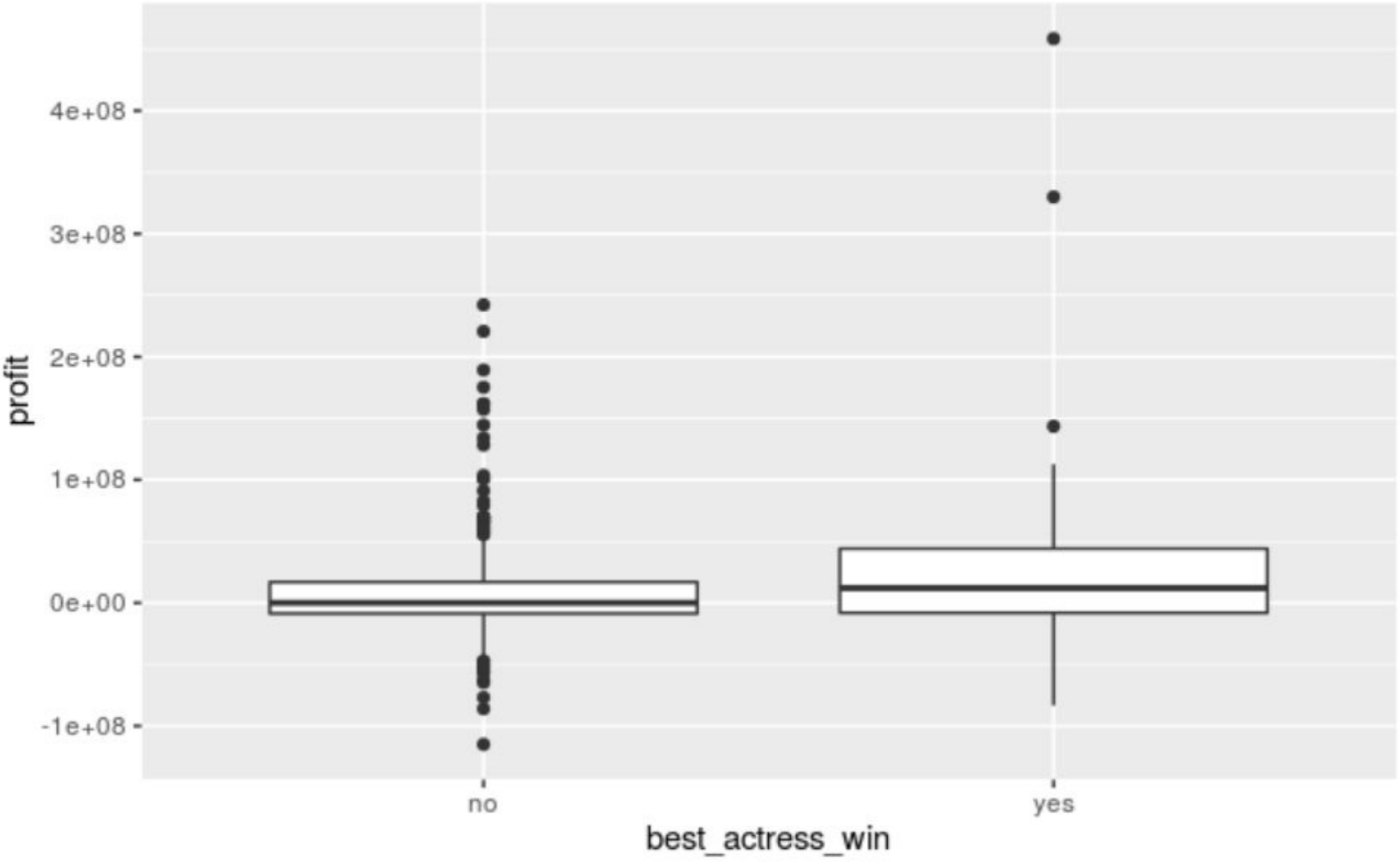
profitable <chr>	count <int>	median(profit) <dbl>
no	179	-9430929
yes	200	19093239

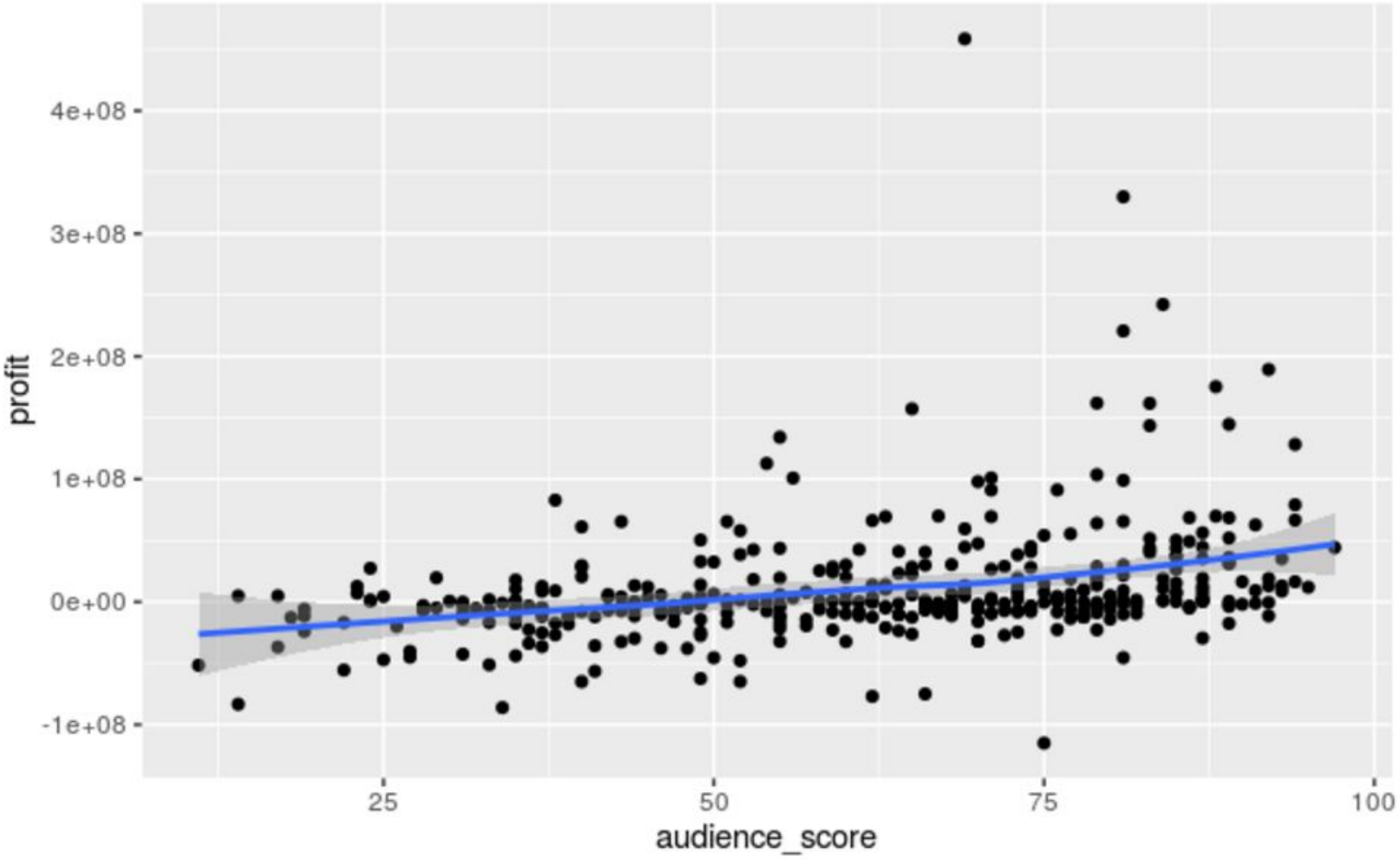

200
150
100
50
0
count
0e+00
2e+08
4e+08
profit


profit
4e+08
3e+08
2e+08
1e+08
0e+00
-1e+08
genre
Action & Adventure
Animation
Art House & International
Comedy
Documentary
Drama
Horror
Musical & Performing Arts
Mystery & Suspense
Other
Science Fiction & Fantasy


4e+08
3e+08
2e+08
1e+08
0e+00
-1e+08
profit
G
NC-17
PG
PG-13
R
Unrated
mpaa_rating
mpaa_rating
G
NC-17
PG
PG-13
R
Unrated


4e+08
3e+08
2e+08
1e+08
0e+00
-1e+08
profit
no
yes
best_pic_nom


4e+08
3e+08
2e+08
1e+08
0e+00
-1e+08
profit
no
yes
best_actress_win


4e+08
3e+08
2e+08
1e+08
0e+00
-1e+08
profit
25
50
75
100
audience_score

Now, remind me why Passengers is a PG13 116 minute long sci-fi movie with a budget of $120 million released on the 21st of December with Jennifer Lawrence who won the Oscar for best actress?