# Contex

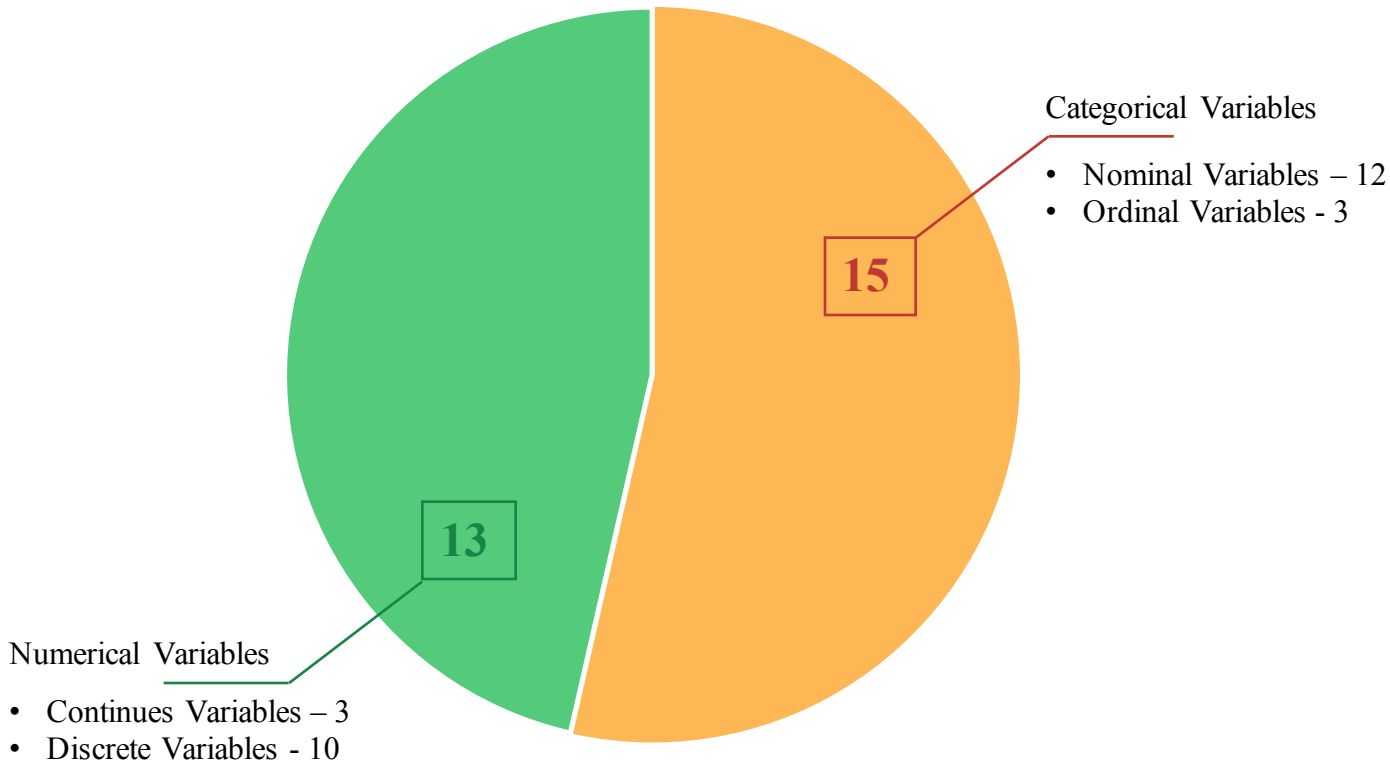
- The Data Set
- Univariate Analysis
- Bivariate Analysis
- Multivariate Analysis
- Reflections

# The Data Set

**5042**
**Observations**

**28**
**Variables**

**1 profit**
**Variable Created**



# Univariate Analysis

## Main Feature

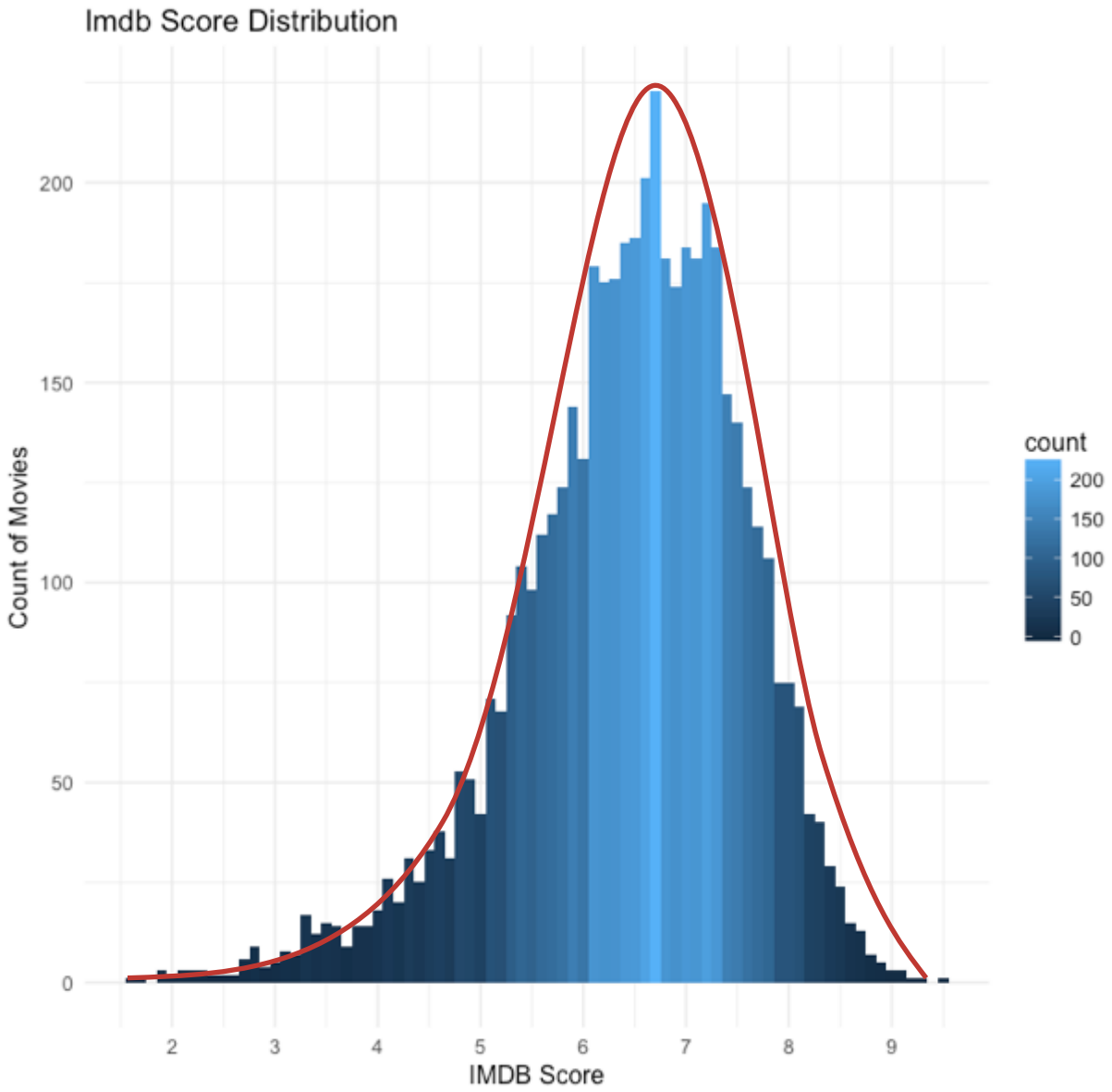
IMDB SCORE

## Supporting Features

Movie Director
Content Rating
Title Year
Profit Distribution
Movie Facebook Likes

## Removed Features

Duration
Movie Color
Language
Critic Reviews
Number User for Reviews

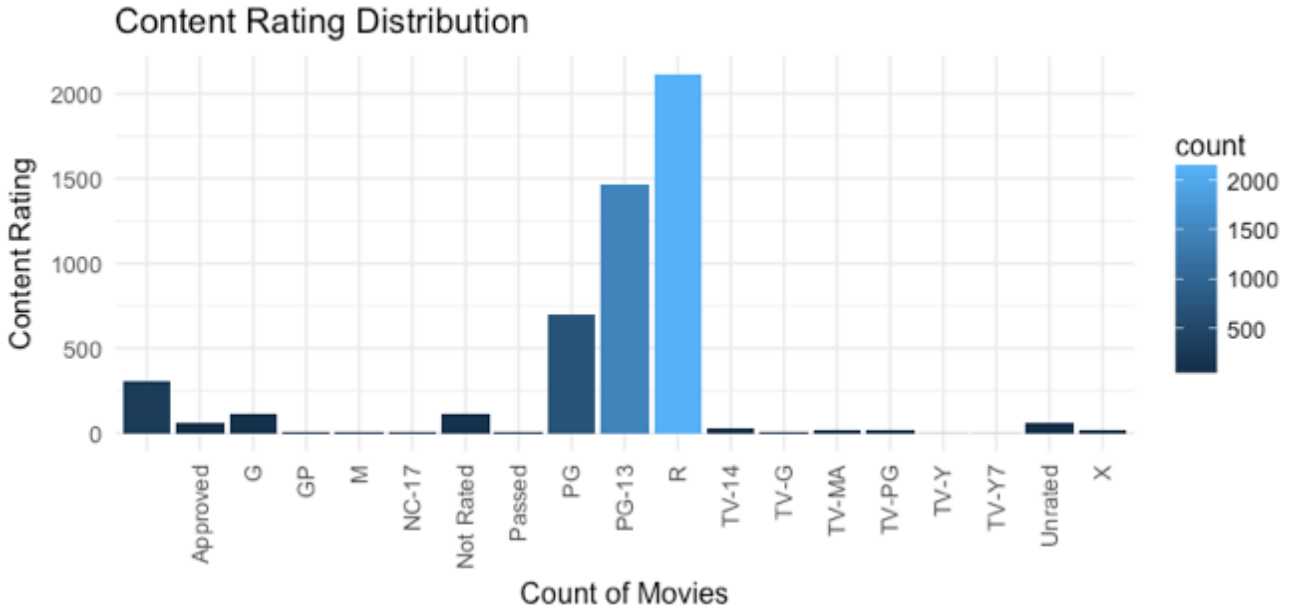


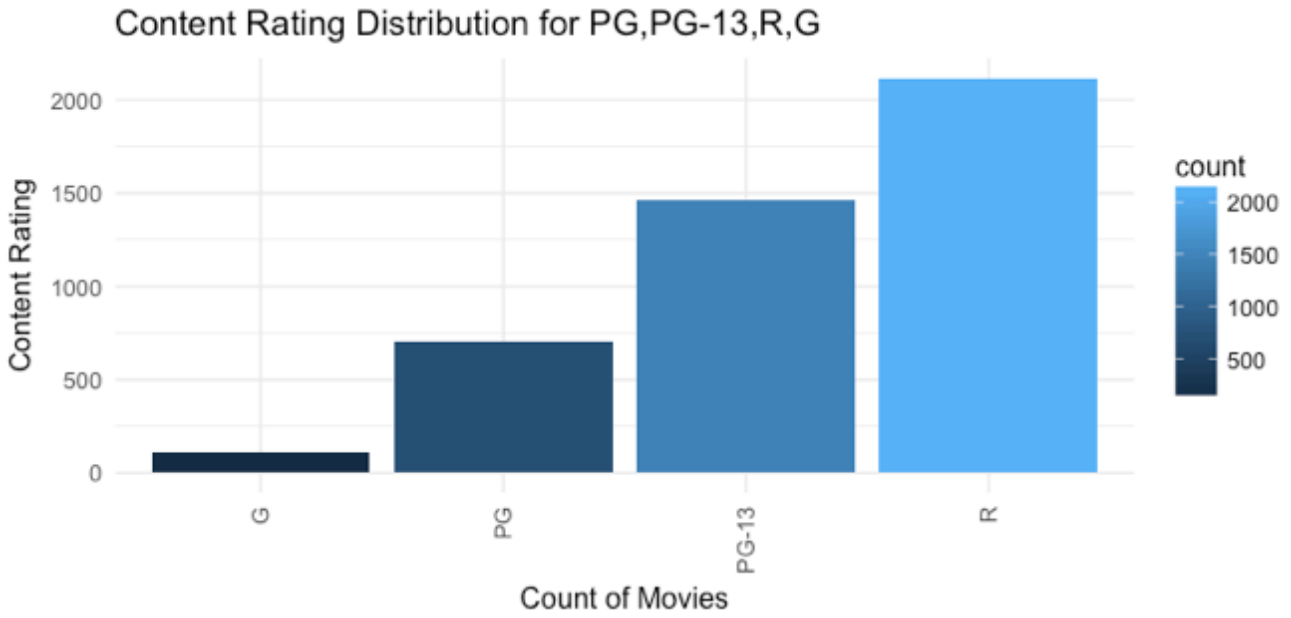
# Univariate Analysis

## Supporting Features

Movie Director
> **Content Rating**
Title Year
Profit Distribution
Movie Facebook Likes

Content_rating distribution - It seem like focusing only to "PG", "PG-13", "R", "G" makes sense to understand their effect on imdb score.



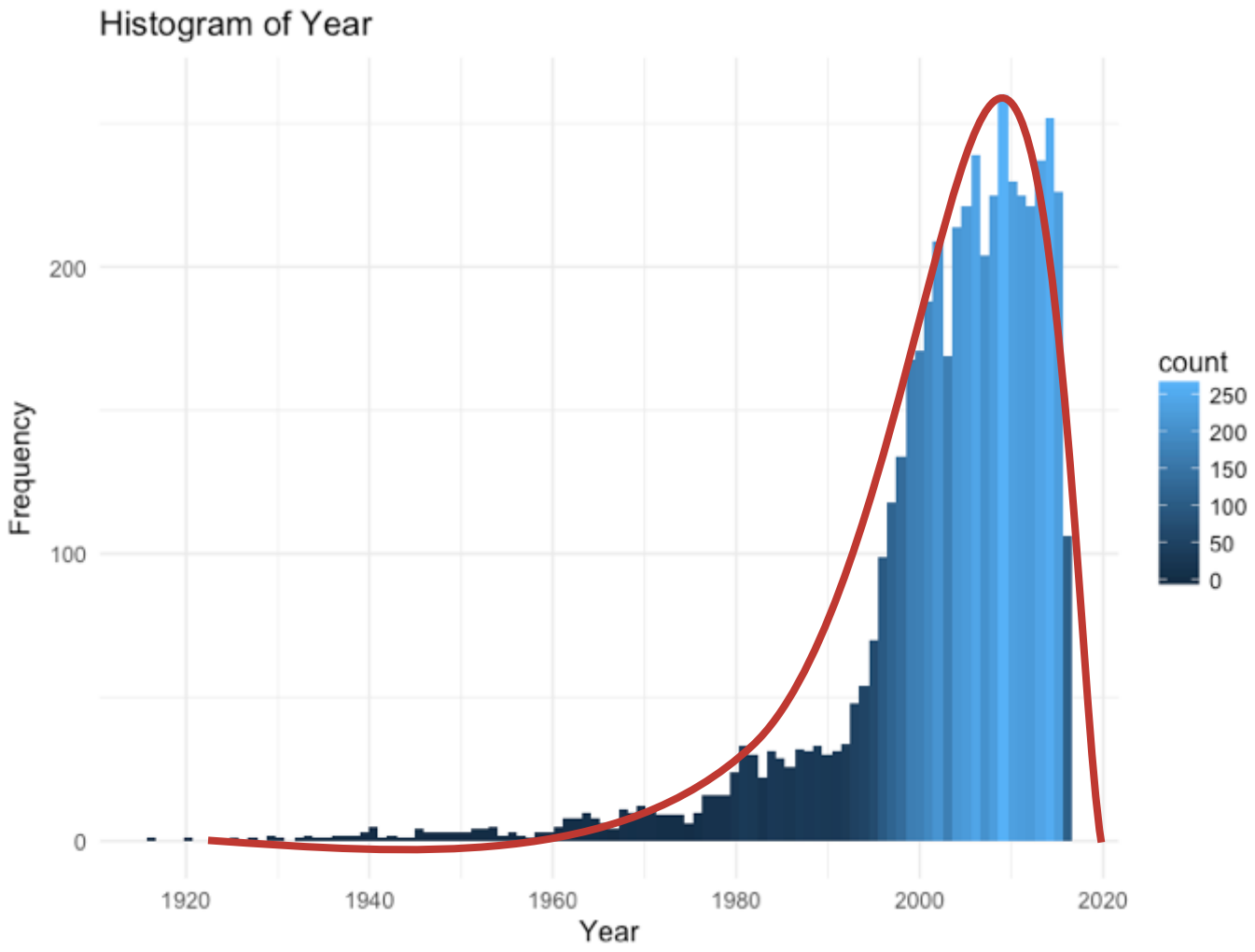
## Supporting Features

Movie Director
Content Rating
**> Title Year**
Profit Distribution
Movie Facebook Likes

It is obvious starting from 1990's movie sctor produce more movies than ever.

Movie Statistics

| | |
|---|---|
| Min. | : 1.00 |
| 1st Qu. | : 3.00 |
| Median | : 10.00 |
| Mean | : 54.82 |
| 3rd Qu. | : 58.00 |
| Max. | : 260.00 |

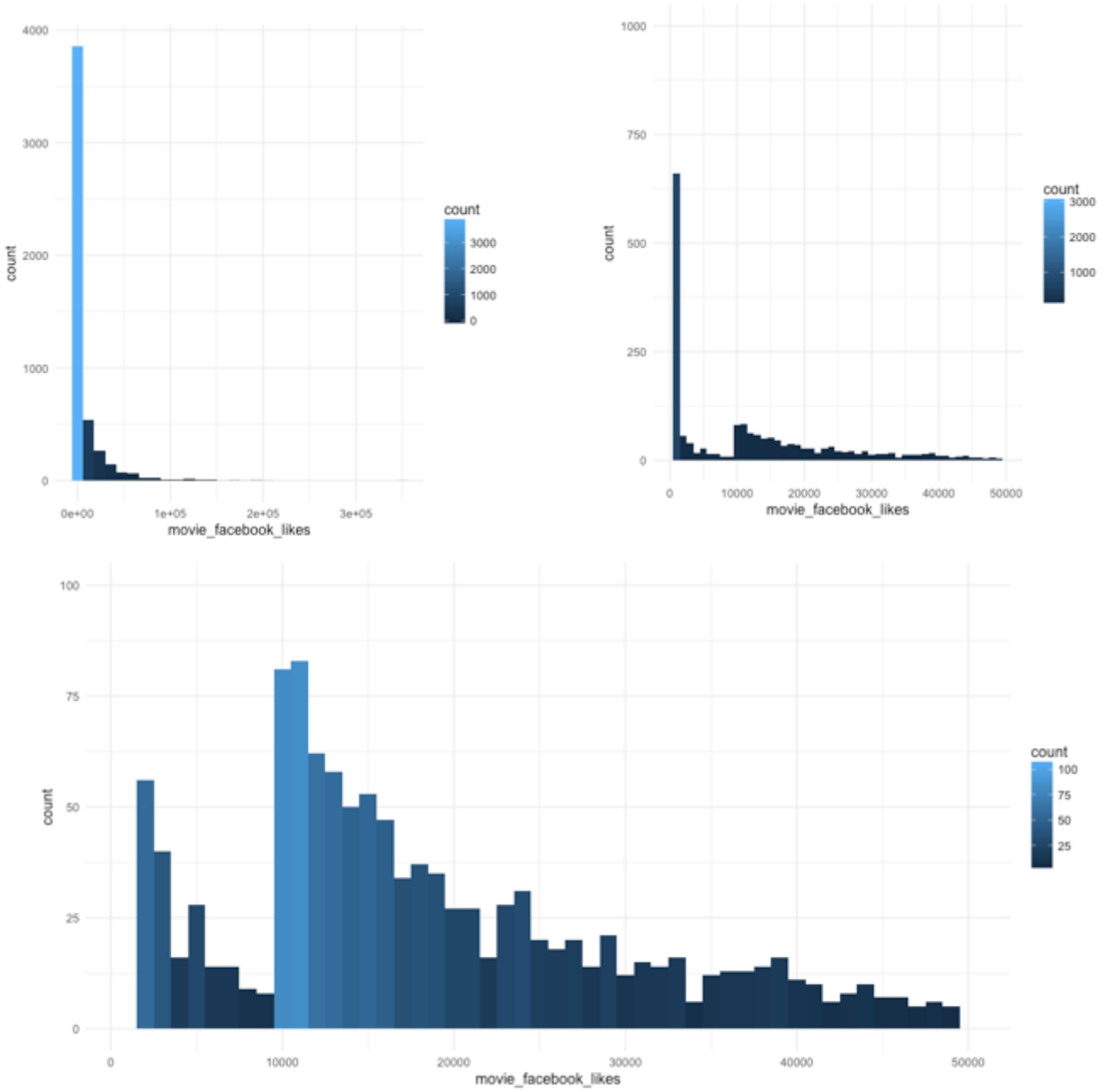


# Univariate Analysis

## Supporting Features

Movie Director
Content Rating
Title Year
Profit Distribution
**> Movie Facebook Likes**

Movie_facebook_likes distribution - It seem like focusing only to certain area gives me more understanding of data so I limit movie_facebook_likes 50000 and count with 1000 occurrences.



## Analysis Conducted

### Vs. Imdb Score

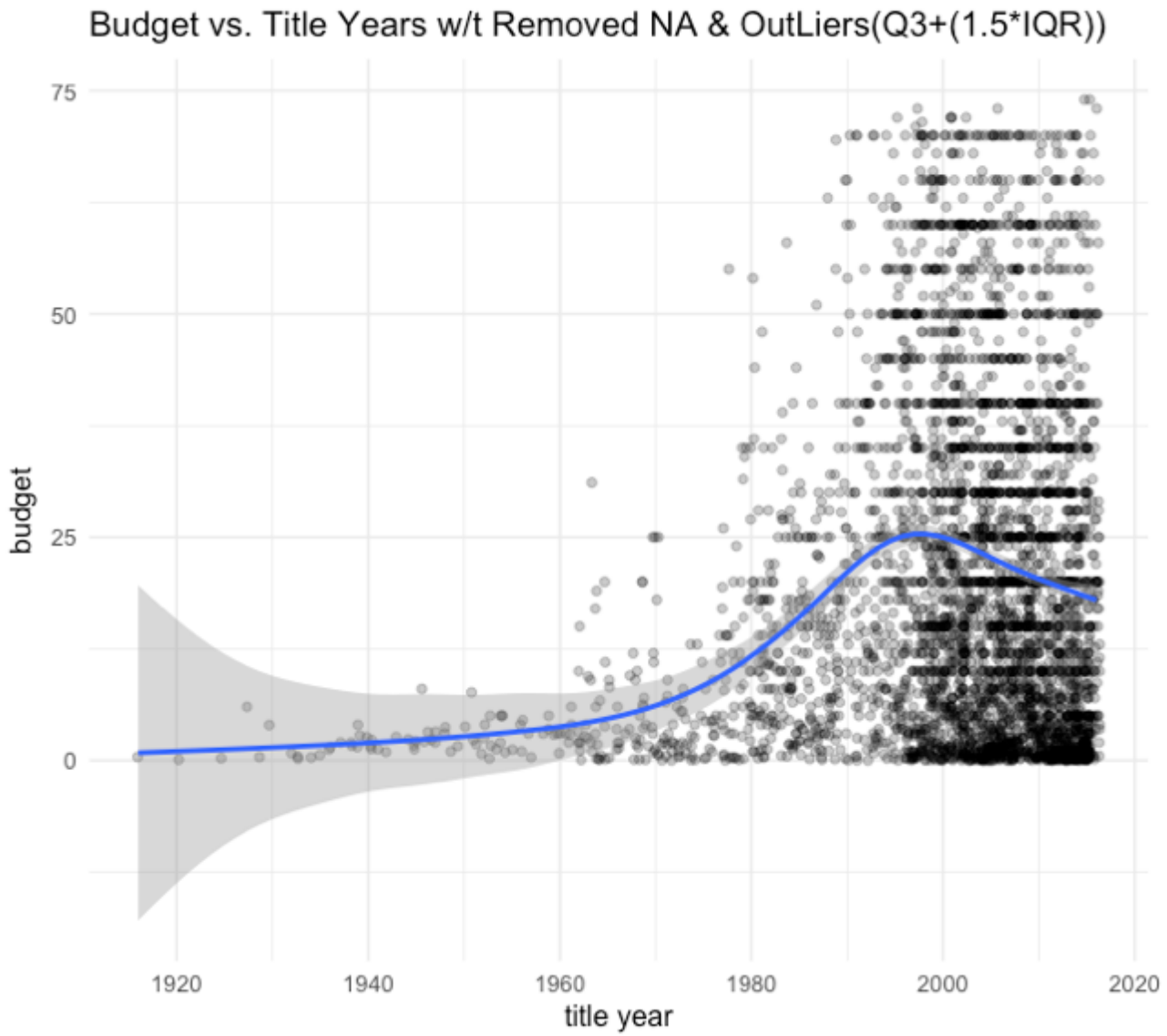
- Imdb Score vs. Title Year of Movies
- Facebook Likes vs. imdb score
- Number of voters vs. imdb score
- Content Rating vs. Average Imdb Score
- Profit vs. imdb Score
- Budget vs Imdb Score

### Vs. Title Year

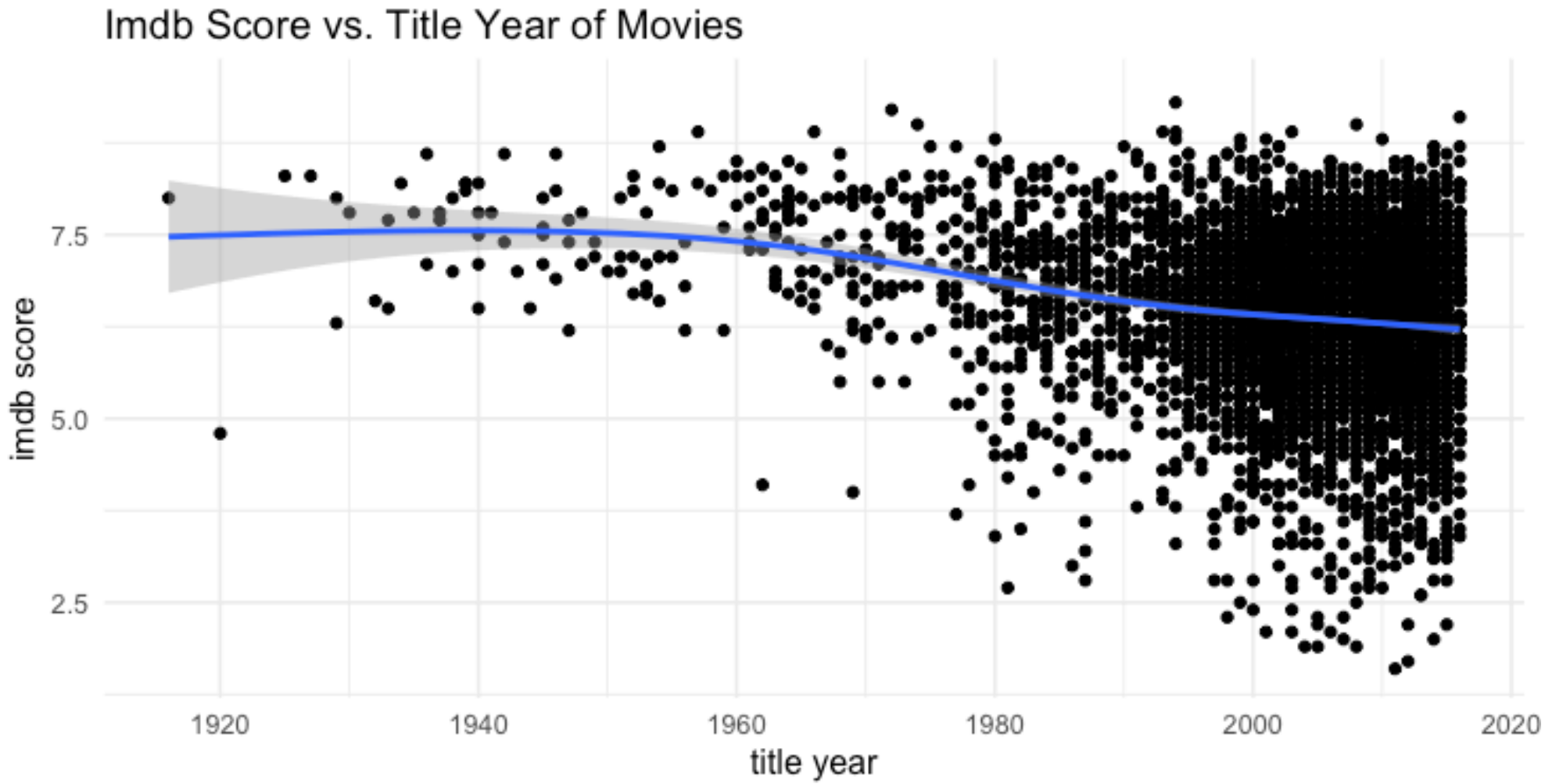
- Facebook Likes vs. Title Year
- Budget vs. Title Year of Movies
- Profit vs. Title Year

### Others

- Distribution of budget over years
- Number of Voted Users  vs. Title Year
- Budget vs profit

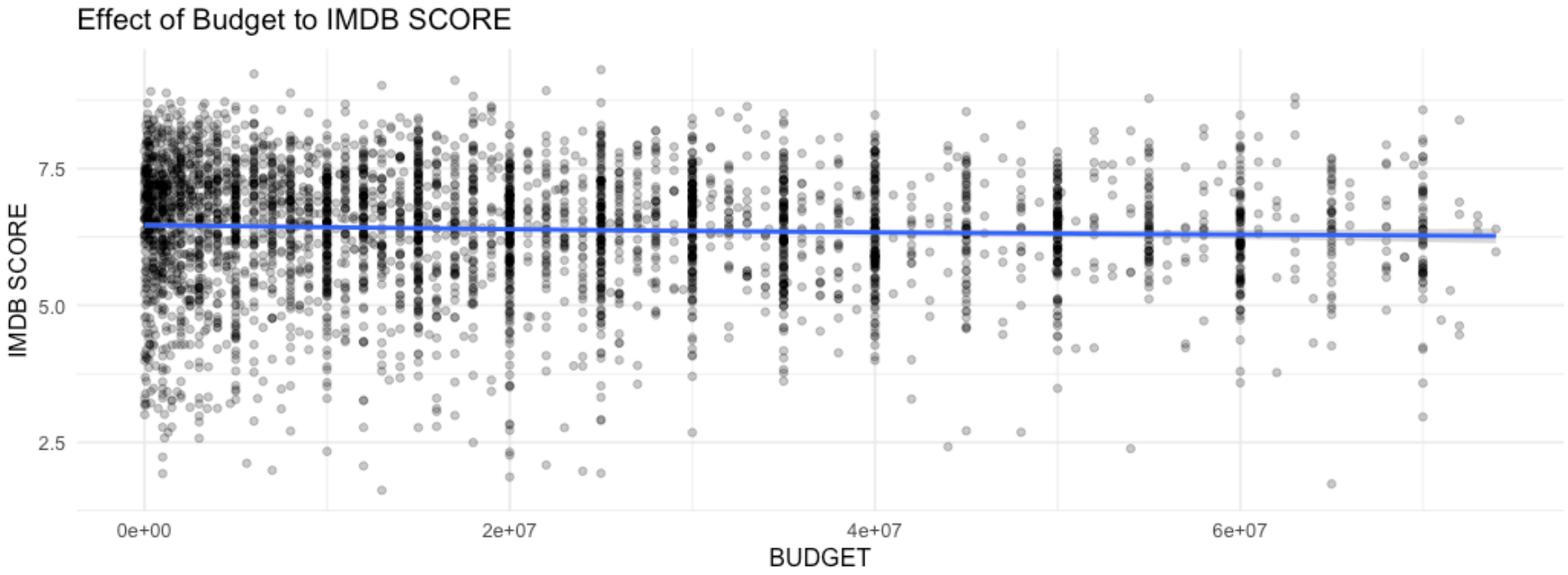


# Bivariate Analysis



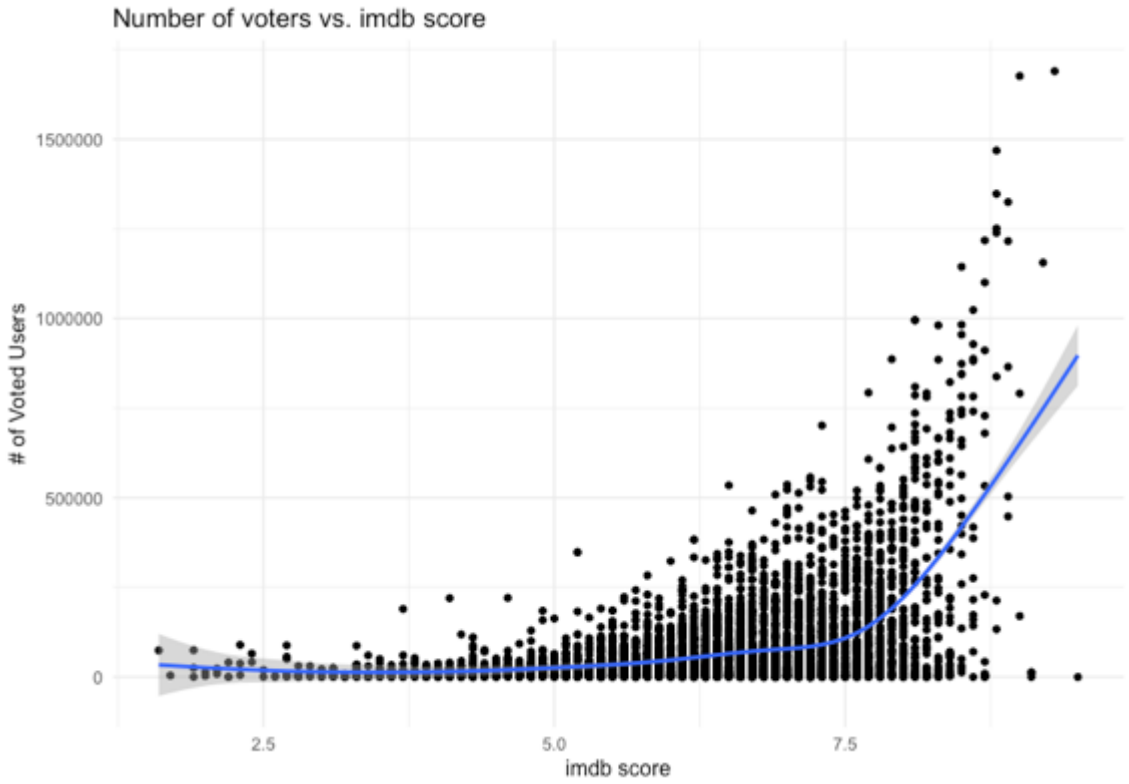

We clearly see that amount of movies produced increase over time but in contrary imdb score goes down. This might be the reason of low quality movie productions.

## BUDGET VS. IMDB SCORE

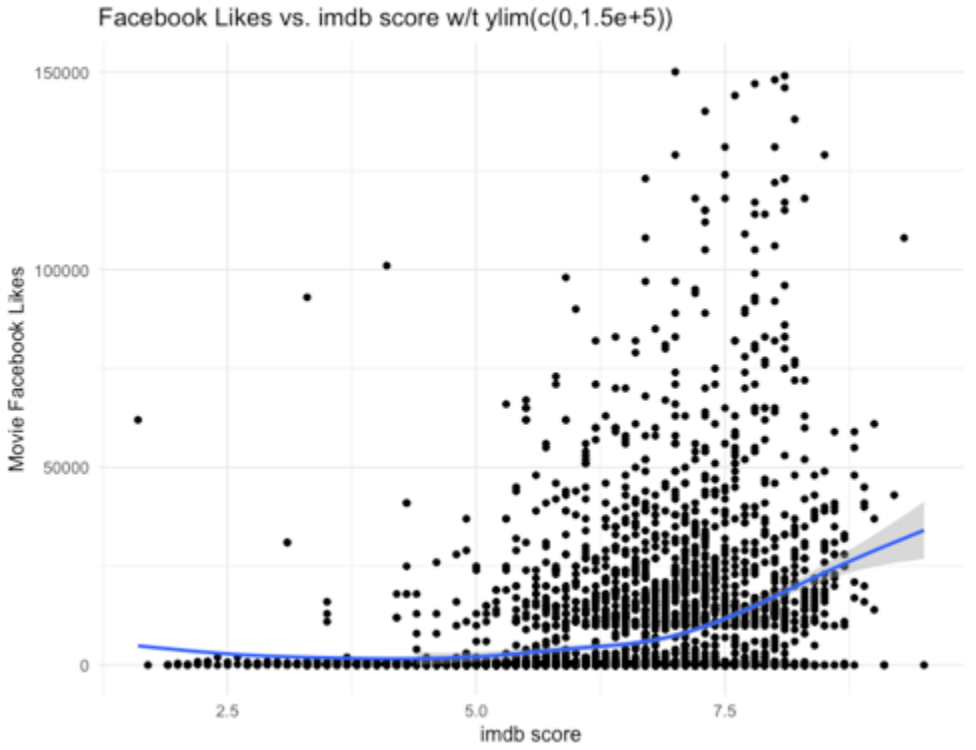



The effect of budget to imdb score appeared to be uniform. So we can easily say that budget have no real effect on imdb score.

# Bivariate Analysis



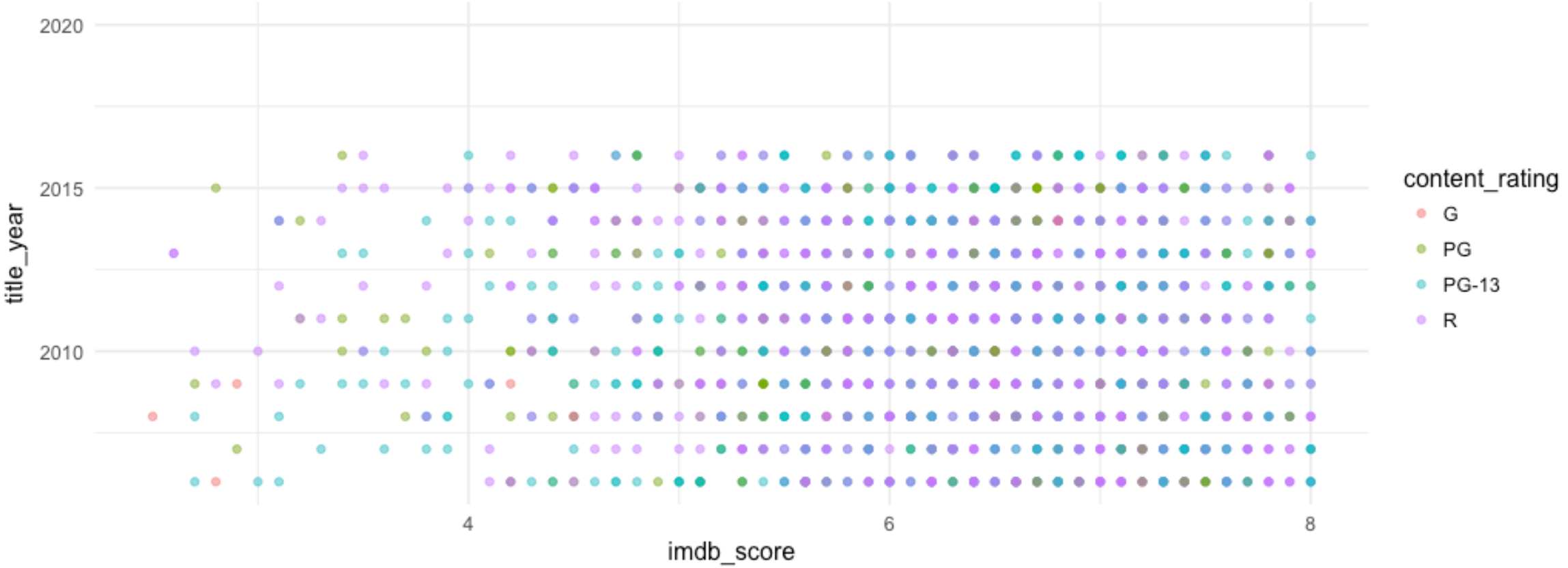

Number of voted users increase exponentially. When the imdb score goes up.




Number of facebook likes increase exponentially. When the imdb score goes up.

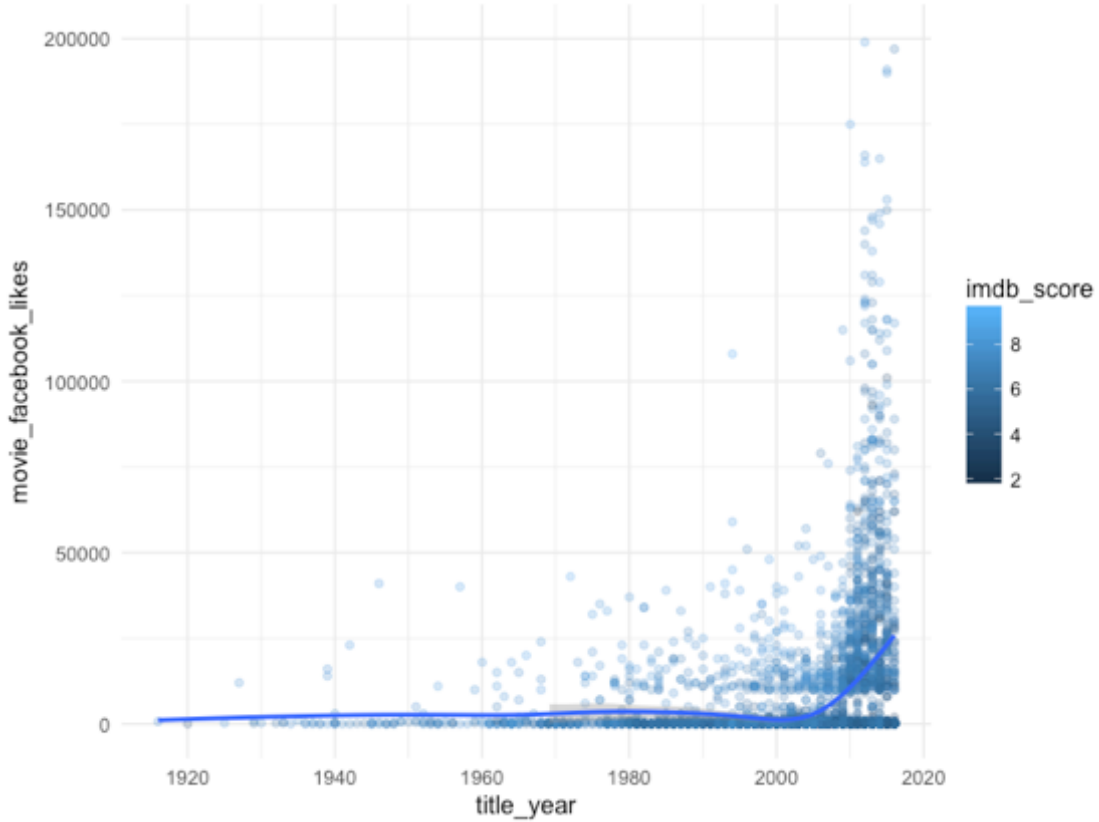
# Multivariate Analysis

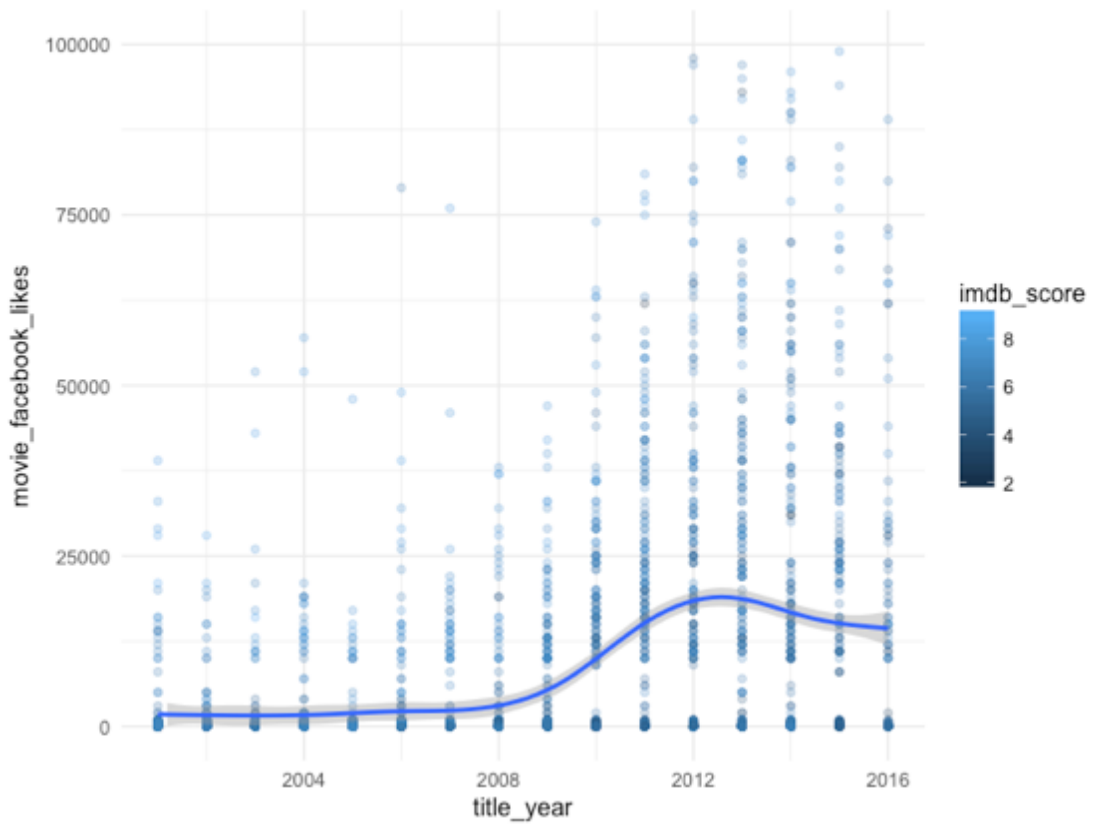
## TITLE YEAR VS. IMDB SCORE VS. CONTENT RATING




One of the surprising case appeared when I analyze content rating. In the univariate analysis "R" and "pg-13" content has a huge dominance over general audience. By this fact in mind I assume that the people tend to like and rate high imdb scores for this kind of movies. But when I check it in multivariate it is clearly seen that content rating have nearly uniform distribution over imdb score. I concentrate on the booming years of movie production.

## TITLE YEAR VS. IMDB SCORE VS. FACEBOOK LIKES




There is a clear trend starting on 2000's that shows when the imdb_score increases movie facebook likes increases. Also this effect strengthens after 2000's. Keeping in mind that the early adapters of facebook is millennials it is expected that they mostly vote for movies that are created after 2000's. Assuming this is not hard since the ability to create life like animations is increased significantly.

# Reflections

- The imdb movie dataset which is downloaded by kaggle.com consists of 28 variables, with almost 5043 observations. This data set has a distribution of 15 categorical and 13 Numerical variables.

- I explored the imdb score of movies across many variables and they to find the patterns that effect the imdb score.

- Budget seem to have a uniform effect on imdb score.

- In addition to this profit seems to have a near exponential effect on imdb score which makes the real identifier gross. in the further analysis with other individual variable facebook likes and directors make an possible effect on imdb score.

Some limitations of this research include the source of the data. The source data is limited and didn't not cover huge number of observations. More over the there is a lack of data (NA) in many of the individual variables. Since this kind of data is eliminated while doing analysis, it might have a positive or negative effect on to the patterns presented. A dataset with more observations and fully filled data, would be better to make predictions of on effects of individual variables on imdb score. To investigate this data further, I would examine how profit is effected by the individual variables that exist in this data set.