I

PREDICTIVE ANALYSIS ON ANIME RATING

PROJECT BY ALI AHMAD

CONTENTS

CASE STUDY

- This dataset comprises the scrapped information about anime releases (Movie/Web series/etc.) from anime-planet (founded in 2001) which is the first anime & manga recommendation database. It comprises the anime & manga release logistics records from the year 2005 to 15th June 2020.
- It Consists 16 variables with 7029 rows.

DIFFERENT VARIABLES

IV

- title
- mediaType
- eps
- duration
- ongoing
- sznOfRelease
- description
- studios
- tags
- contentWarn
- watched
- watching
- wantWatch
- dropped
- rating
- votes

V

Objective

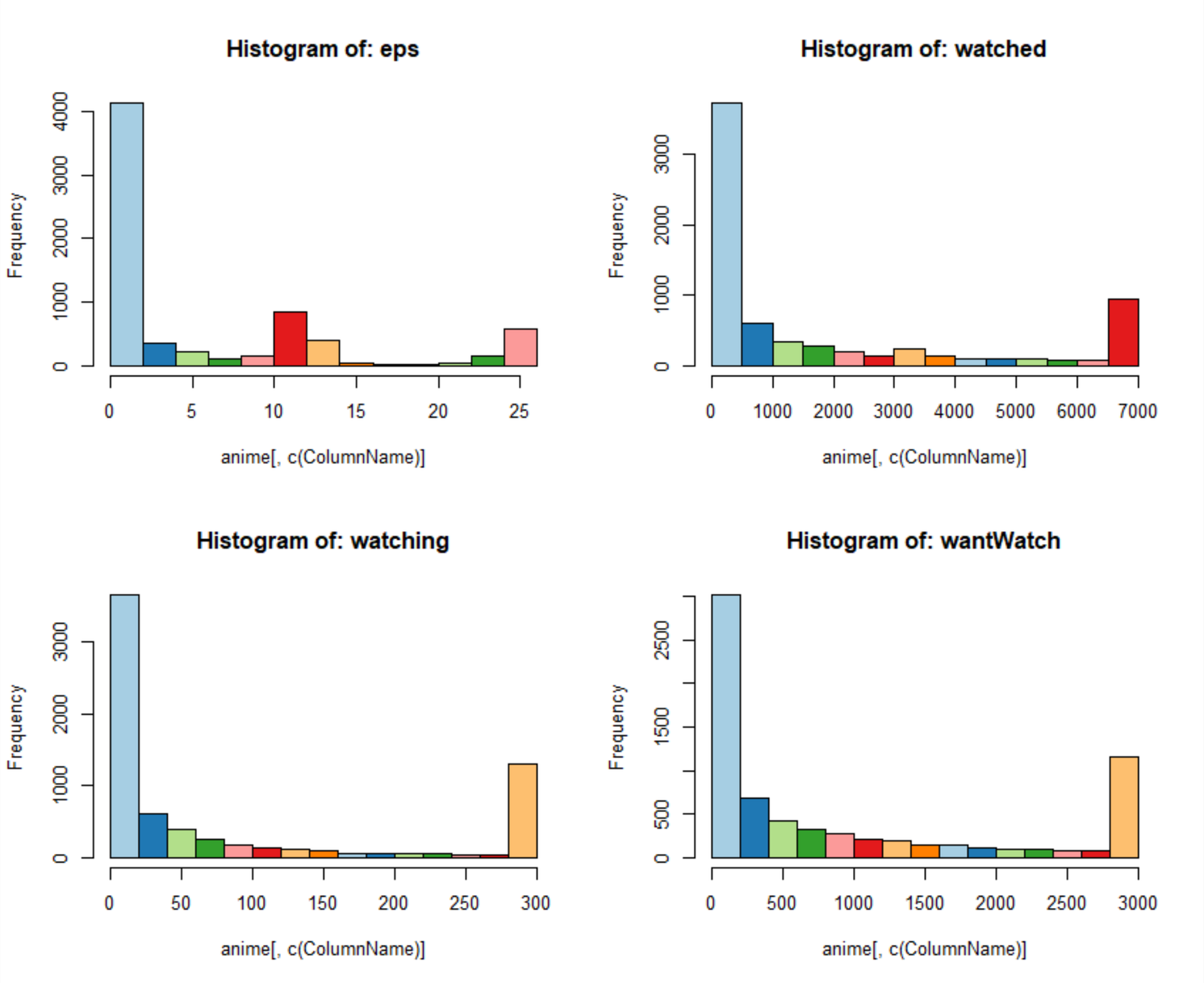
The objective here is to perform a Linear regression analysis to arrive at a model that can be used to predict the Ratings received by the enlisted anime releases like Movie/Web series such that in future, the anime production studios can develop their strategies which can improve the ratings.

VI

UNIVARIATE ANALYSIS

Using Histogram: For Continuous columns

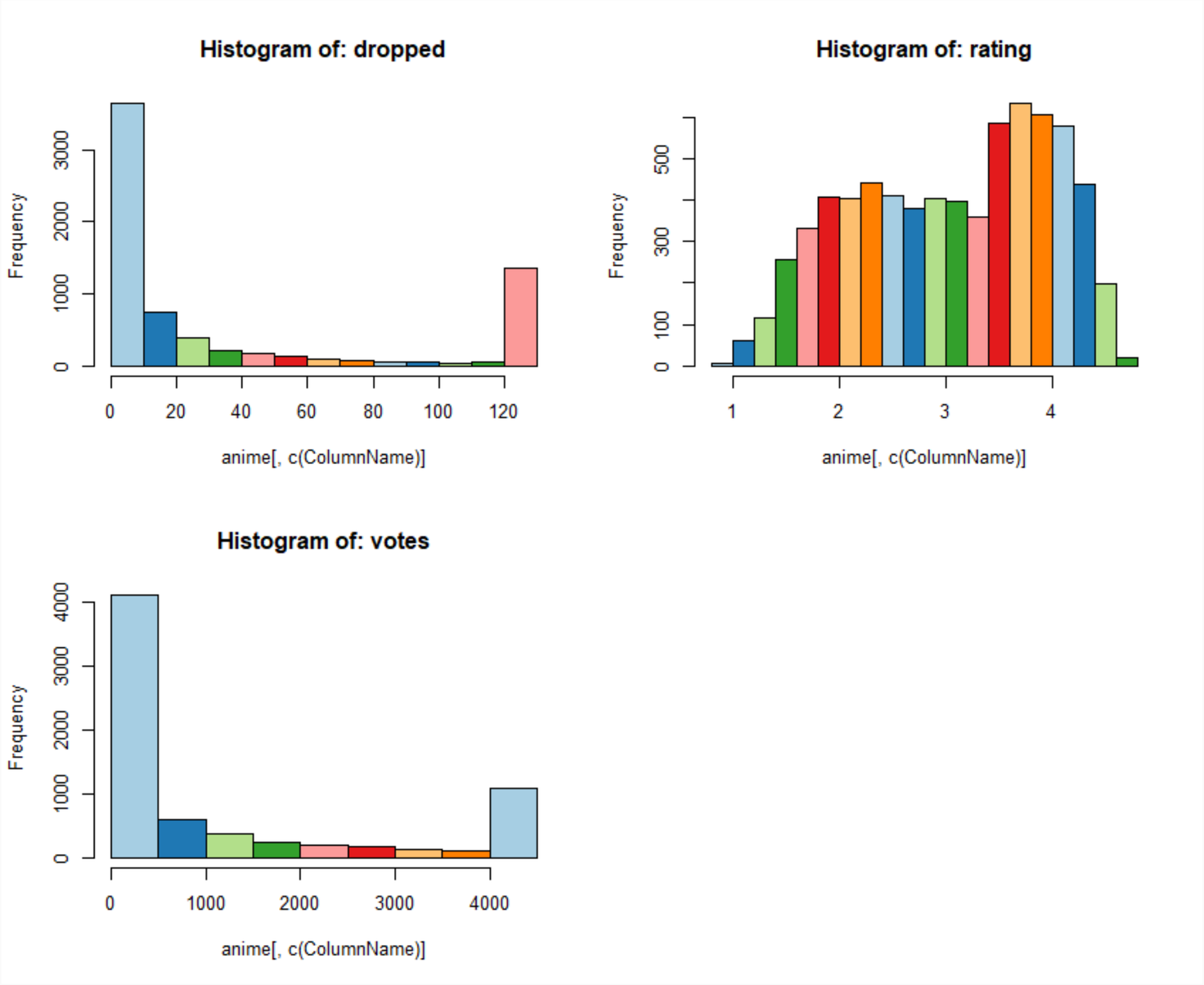
All the columns are right skewed. The columns are not normally distributed.

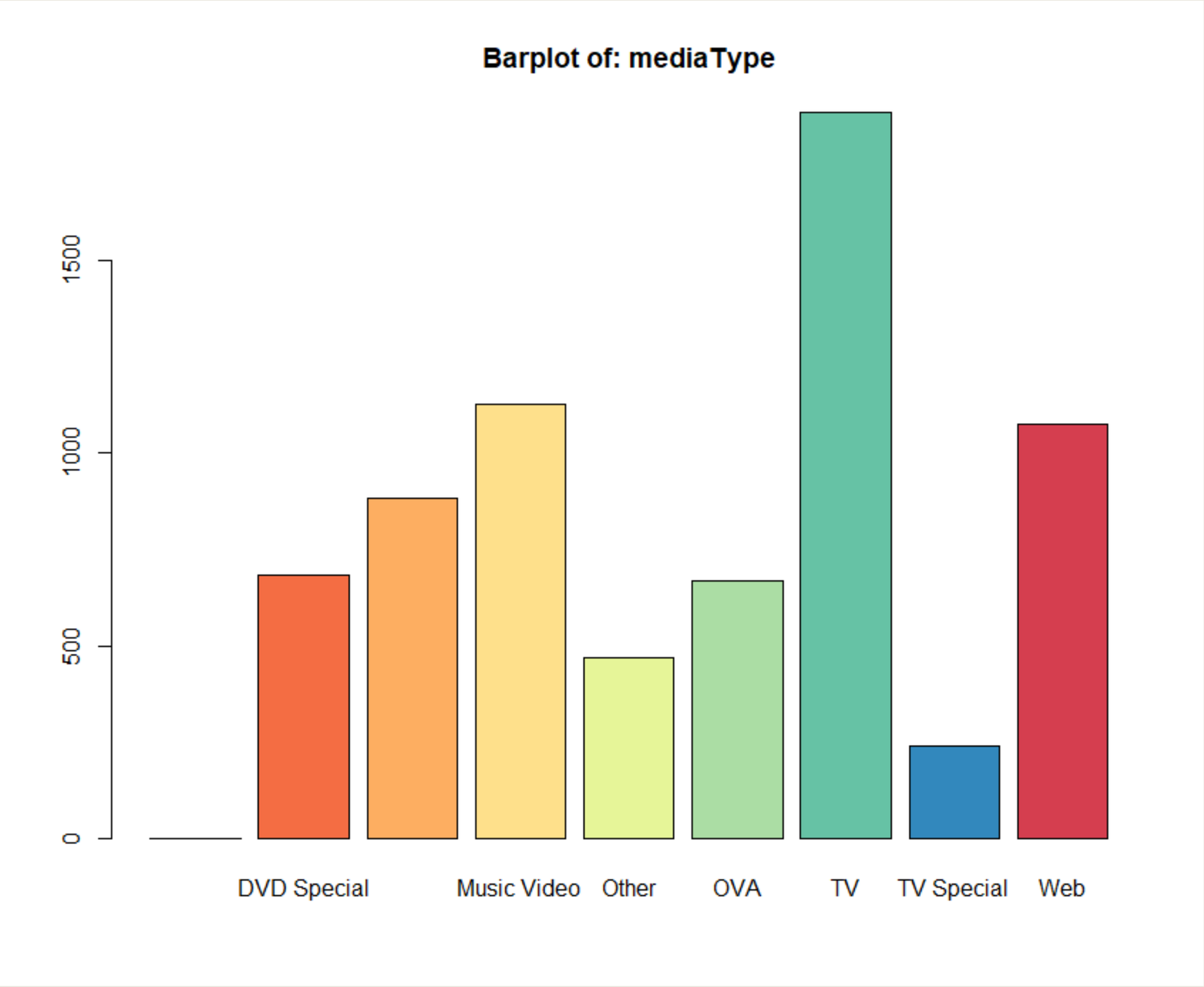


VII

UNIVARIATE ANALYSIS

The variables “dropped” and “votes” are also right skewed. Theses columns are not normally distributed.





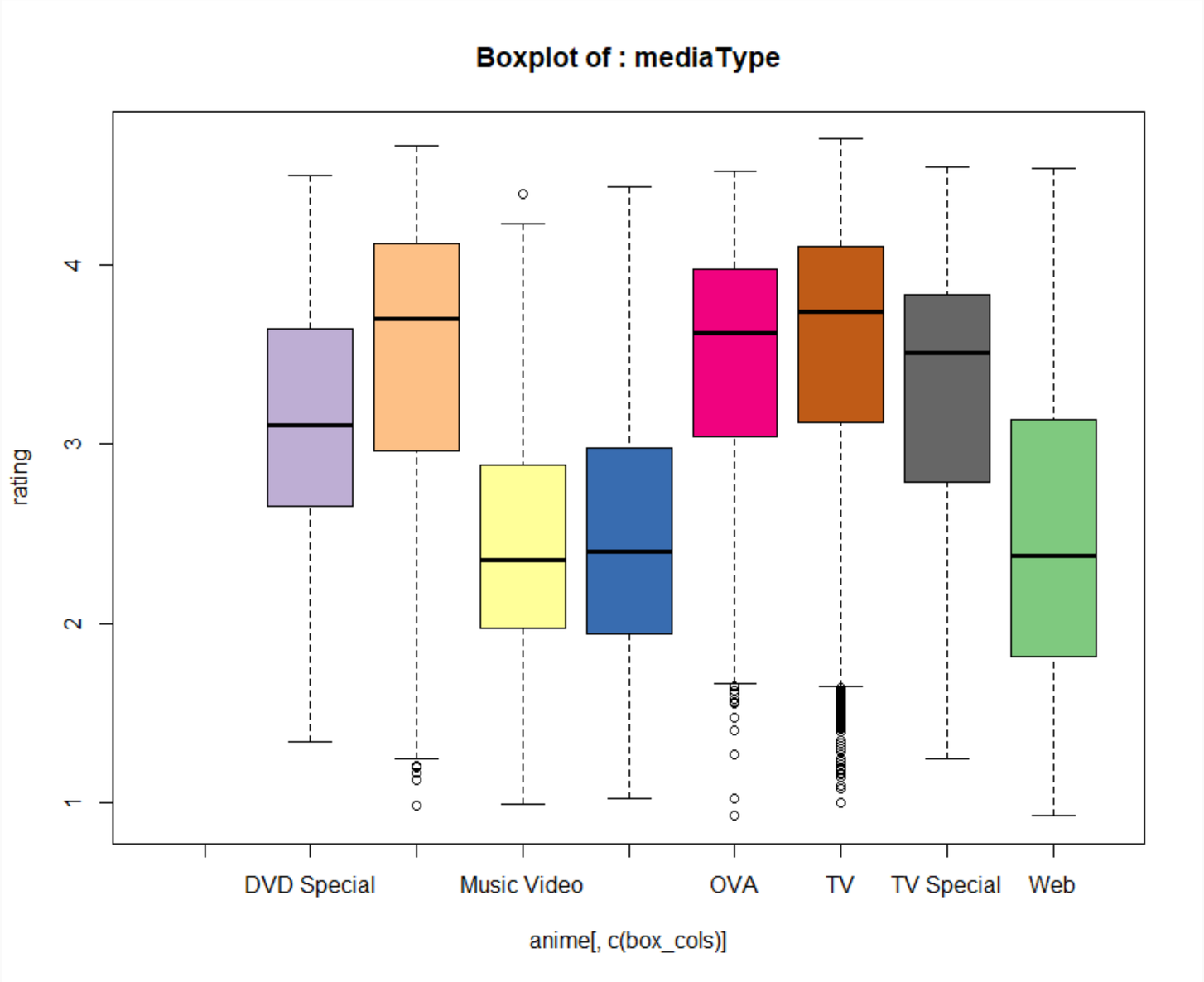
UNIVARIATE ANALYSIS

Using Barplot: Categorical

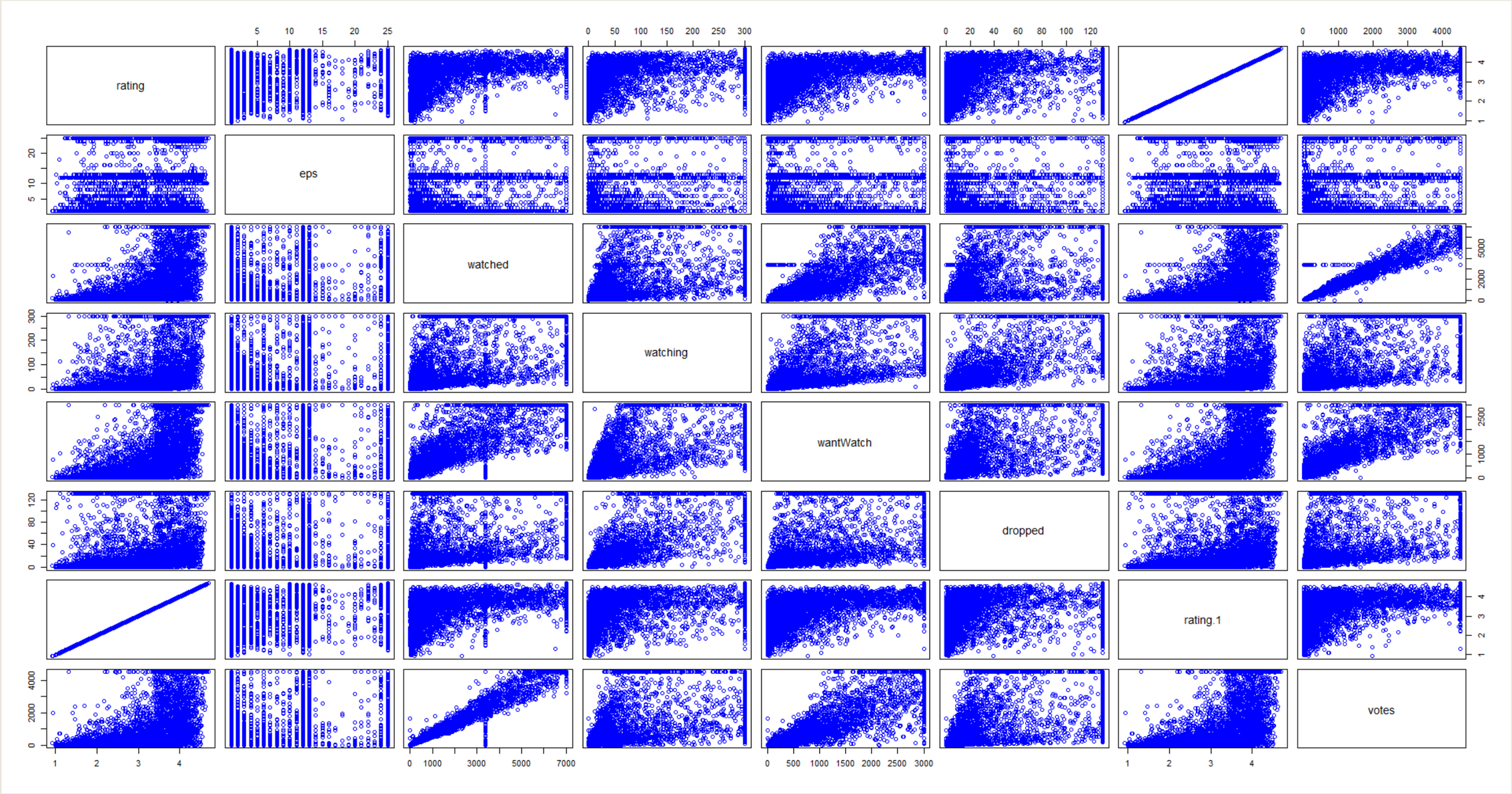
TV type has higher rating as compared to other types like OVA and TV Special which have significantly low Viewer rating

IX

BIVARIATE ANALYSIS



BIVARIATE ANALYSIS : Scatterplot



CORRELATION TEST

Correlation is a statistical technique that predicts whether and how strongly pairs of variables are related. It gives the relationship between continuous and continuous columns whether they are statistically significant or not.
It ranges from -1.0 to +1.0.

In this dataset, the variables that are found significant from Correlation test:

- Eps
- Watched
- Watching
- Wantwatch
- Dropped
- Votes

The variable “Eps” is not correlated with the target variable Rating, so we reject this variable for further analysis.

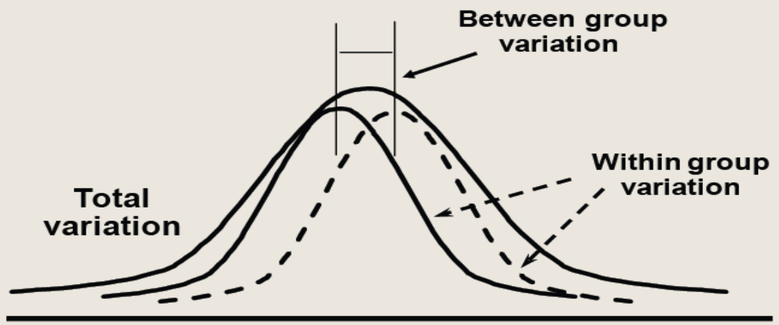
XII

ANOVA (Analysis of Variance)

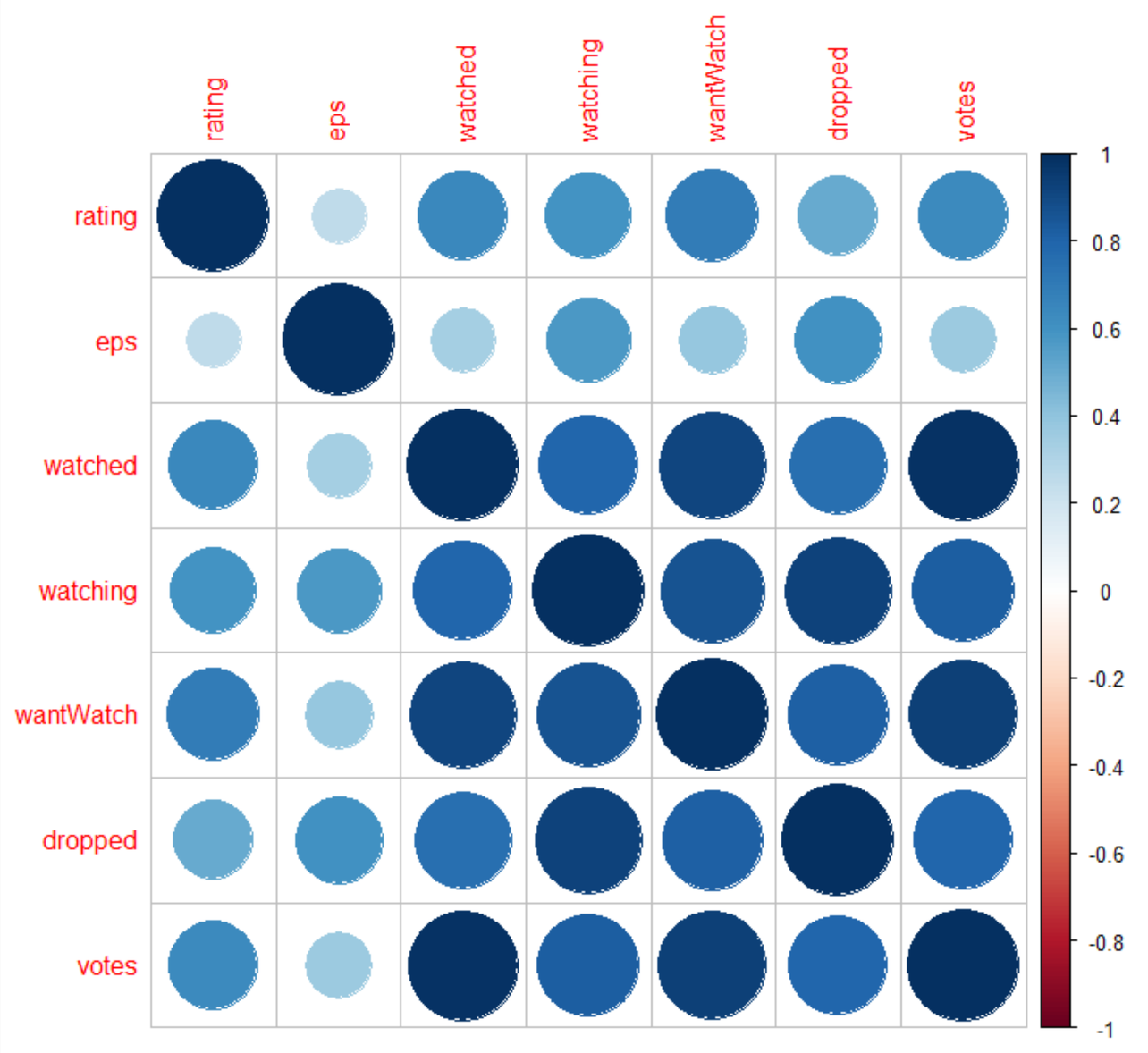
It gives the relationship between continuous and categorical columns whether they are statistically significant or not.

H0- Null hypothesis means the variables are not correlated:
1. If Small P-Value < 5% (0.05) means the variables are correlated. Null hypothesis H0 is rejected
2. If Large P-Value > 5% (0.05) means the variables are not correlated and the Null hypothesis H0 is accepted

In this dataset, the variables (p-value < 0.05) that are found significant from ANOVA test:
- ❖ mediaType



HEATMAP OF CORRELATED VARIABLES



XIV

MODEL BUILDING WITH SIGNIFICANT VARIABLES

- Votes
- MediaType
- Dropped
- WantWatch
- Watching
- Watched

XV

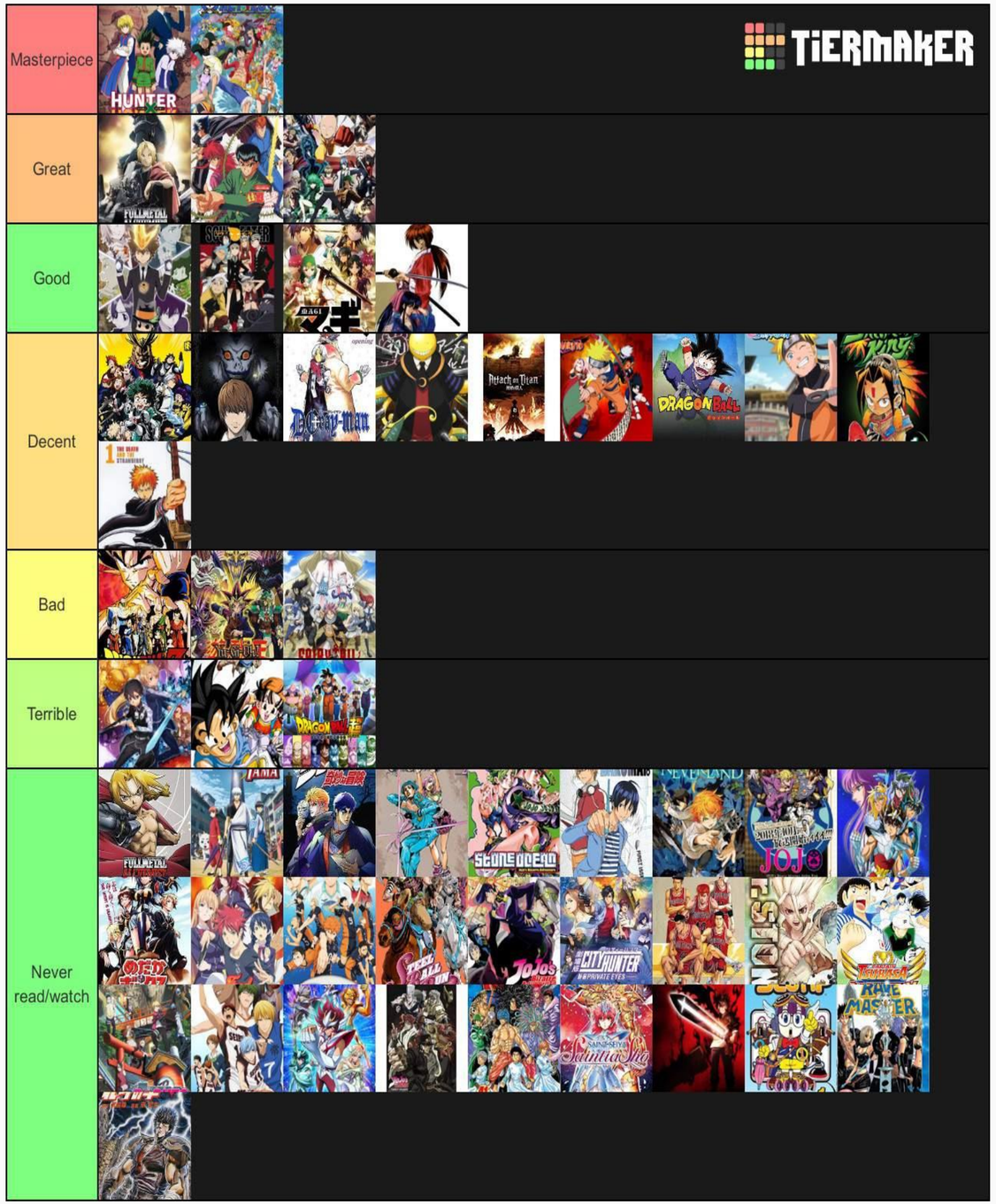
OBSERVATIONS

- Multiple R-Squared of Linear Regression : 0.5533
- Adjusted R-squared of Linear Regression : 0.5524
- Mean Absolute Percentage Error (MAPE) : 18.97585
- Mean Accuracy of Linear Regression : 81.02
- Median Absolute Percentage Error (MDAPE) : 13.58057
- Median Accuracy of Linear Regression : 86.41

XVI

BUSINESS RECOMMENDATIONS

- Production studios needs to promote their shows in TV Special format because this format can provide higher rating of viewers.
- The studios should broadcast through OVA, as the MediaType can increase the viewer's rating as compared to DVD Special.
- The studios should release more on Web platform as nowadays present generation like to watch on internet. This will elevate the rating of the anime.



THANK YOU