Valeri Vankov

Exploring the Role of Genre and Rating Among the Highest Grossing Movies of All Time

Introduction

How do we choose what movie to watch next? Among the two most useful descriptors are genre and MPA rating. These categories inform consumers what the movie might be about and help them decide if they want to watch it. At the same time, studios and streaming companies can use genre and rating to market a movie to a particular audience, or consider what types of movies they want to produce and distribute in the future. As a result, I am interested in exploring how genre and rating are related to each other as well as to other movie metrics like year of release, runtime, critical reception, and box office performance. For this purpose, I chose to analyze the top 1000 highest grossing movies of all time, as of November 21, 2023, since I am especially curious to observe whether genre and rating had any influence on their success.

Design and Primary Questions

I will use three types of multivariate analysis techniques (ordination, MANOVA, and cluster analysis) to examine the relationship between genre and rating and explore how these two metrics impact the rest of the variables in the dataset. First, I will use NMDS (non-metric multidimensional scaling) and CCA (canonical correspondence analysis) to think about the relative positions of genres and ratings and how additional variables like year, runtime, reception, and commercial success impact them. After that, I will create interaction plots, run a two-way MANOVA, perform multivariate and univariate contrasts, and run MRPP (multi-response permutation procedure) tests to discuss the interaction between genre and rating, and analyze their individual as well as combined effects on the other variables. Finally, I will conduct hierarchical cluster analysis and k-means clustering on the top 50 highest grossing movies to see if movies from the same group have, in fact, the same genre or rating. Ultimately, the goal of the project will be to compare the results of the three approaches and gain insight about the importance of a movie's genre and rating.

Data

The data was retrieved from a publicly available list on IMDb. Most of the information I needed for the analysis is readily available on IMDb but the gross values for all the films were manually added by the creator, who used Box Office Mojo data, which is also provided by IMDb. Thus, there should be no inaccuracies in the data. The only issue is that not all movies have a metascore or a rating listed. I have excluded movies with missing information at the beginning of my analysis so that they do not interfere with the methods used. I have also set the movie titles as row names so that the data frame itself contains the rest of the variables. Below I have provided more detailed descriptions about each of the variables and have indicated whether they are continuous or categorical:

Name	Description	Type
Year	Year of release	Continuous
Rating	MPA rating (e.g., PG, PG-13, R)	Categorical
Runtime	Duration, measured in minutes	Continuous
Genre	Contains between 1 and 3 genre categories (e.g., action and comedy)	Categorical
IMDb Score	User rating, given as a number between 1 and 10 rounded to first digit after the decimal point	Continuous
Metascore	An aggregate score based on reviews from Metacritic, given as an integer between 0 and 100	Continuous
Votes	Number of user ratings	Continuous
Gross	Worldwide lifetime gross, measured in USD	Continuous

Descriptive Plots and Summary Statistics

Let's first look at the distribution of genres:

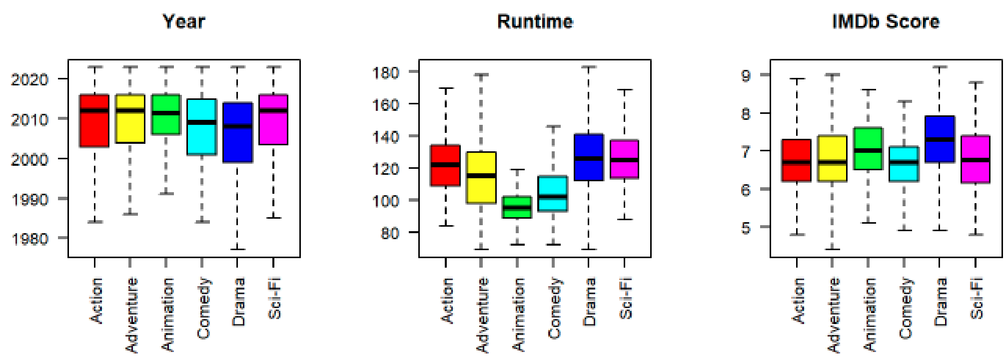
```
  Adventure      Action      Comedy       Drama   Animation      Sci-Fi    Thriller     Fantasy 
        514         469         355         270         158         152         144         120 
      Crime      Family     Romance     Mystery      Horror   Biography     History       Music 
        108          88          87          66          46          37          18          16 
        War     Musical       Sport     Western Documentary 
         10           8           7           5           3 
```

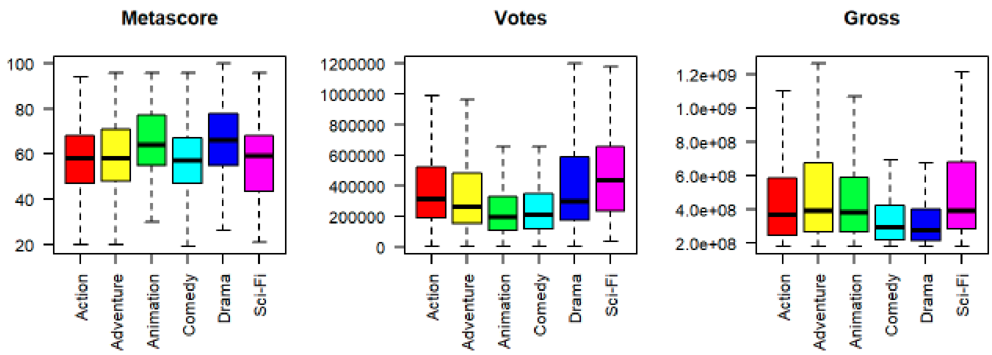
There are quite a few genres with varying frequencies. For ordination, I will only consider the most frequent genres up to horror and for MANOVA I will reduce the number of genre categories even further. Let's also calculate the number of times each rating appears in the dataset:

```
   PG-13         PG          R          G  Not Rated        13+   Approved     Passed      TV-Y7 
     456        236        209         44         11          1          1          1          1 
```

Again, the ratings are not evenly distributed. For ordination, I will analyze all PG-13, PG, R, G, or unrated movies (the last category indicates movies that were not submitted to the MPA for a rating) while for MANOVA I will only focus on PG-13, PG, and R movies.

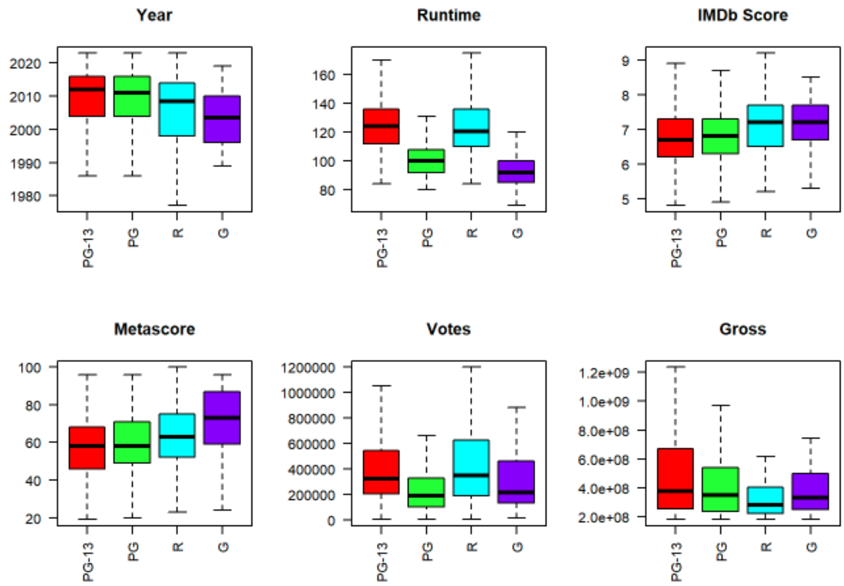
Below I have created a series of boxplots for all the genres containing at least 150 movies:





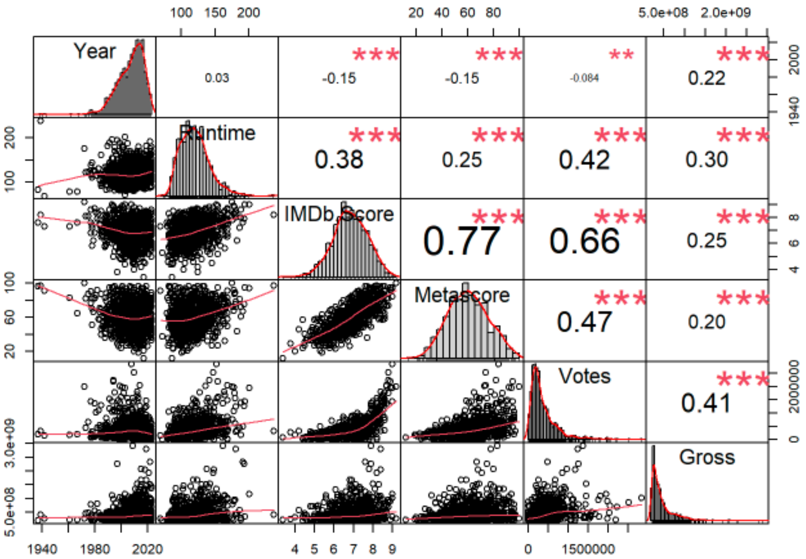
It appears that animation and comedy have noticeably shorter runtimes while dramas have higher runtime, IMDb score, and metascore. Also, sci-fi movies seem to have a higher number of votes compared to animation and comedy. Also, adventure and sci-fi movies appear to have the highest gross while dramas and comedies are the least profitable genres.

Similarly, I created boxplots for all the movies with ratings PG-13, PG, R, and G:



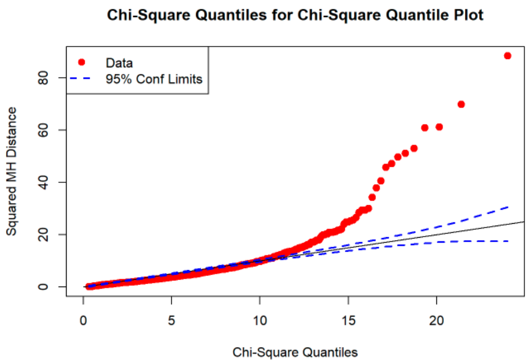
PG-13-rated and R-rated movies have longer runtimes and more votes on average than PG-rated and G-rated movies. R-rated and G-rated movies seem to have slightly higher IMDb scores and metascores. Finally, PG-13-rated movies tend to have the highest gross while R-rated movies have the lowest.

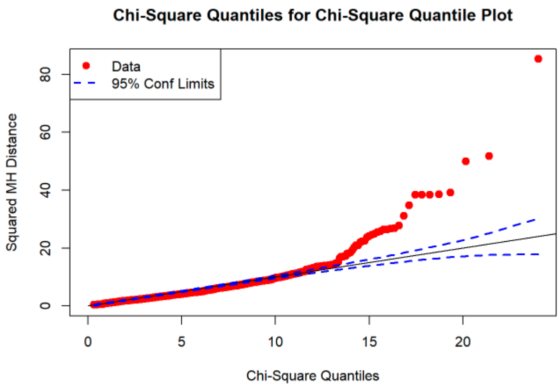
Apart from boxplots showing the relationships between genre, rating, and the continuous variables, we can create a matrix correlation plot to observe relations between the continuous variables as well:



The correlations between IMDb scores, metascores, and votes are relatively high and positive but the correlations between the rest of the variables are somewhat lower. This means that there may be relatively strong linear relationships between IMDb scores, metascores, and votes and weaker for the other variables. Additionally, all variables are somewhat normally distributed, except for `Votes` and `Gross` which are both highly skewed to the right.

To reduce the skewness of Votes and Gross and make them more normally distributed, I will perform log transformations on these variables. This will also make the chi-square plot of all the continuous variables fit more closely to a multivariate normal distribution (the left plot is before transformation and the right is after transformation):



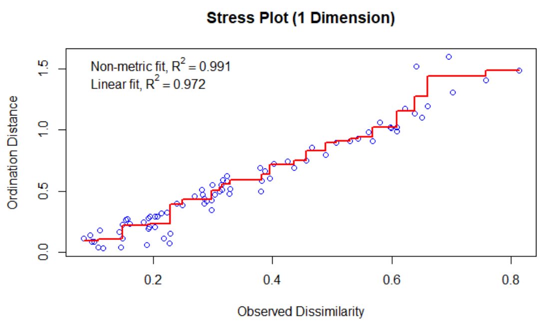


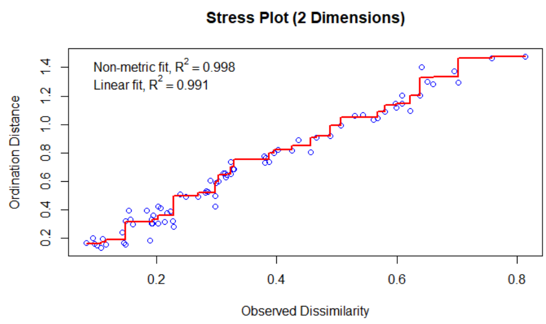
Multivariate Analysis

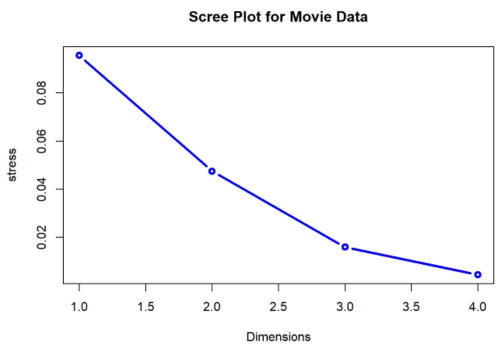
Ordination

As I stated above, I will only consider the 13 most frequent genres and 5 types of ratings. To perform the analysis, I have created a two-way table for genre and rating. Additionally, I have created a table that contains the average metric for each of the continuous variables per genre. I scaled the continuous variables prior to the analysis so that no variable outweighs the others when performing ordination.

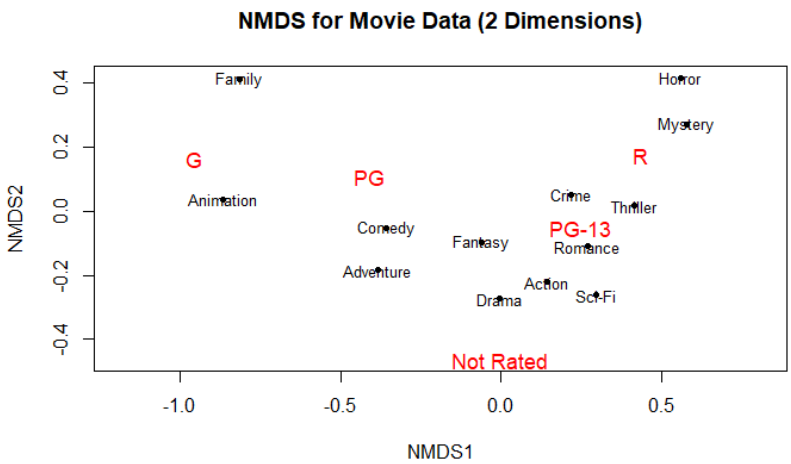
Let’s begin with NMDS using Bray-Curtis distance (which is set by default in R). The stress plots for one-dimensional NMDS and two-dimensionsional NMDS as well as the scree plot comparing the stress for the dimensional solutions ranging from 1 to 4 reveal that one or two dimensions are sufficient, but two dimensions are perhaps ideal for more precise results.







The two-dimensional plot of the NMDS results is shown below:

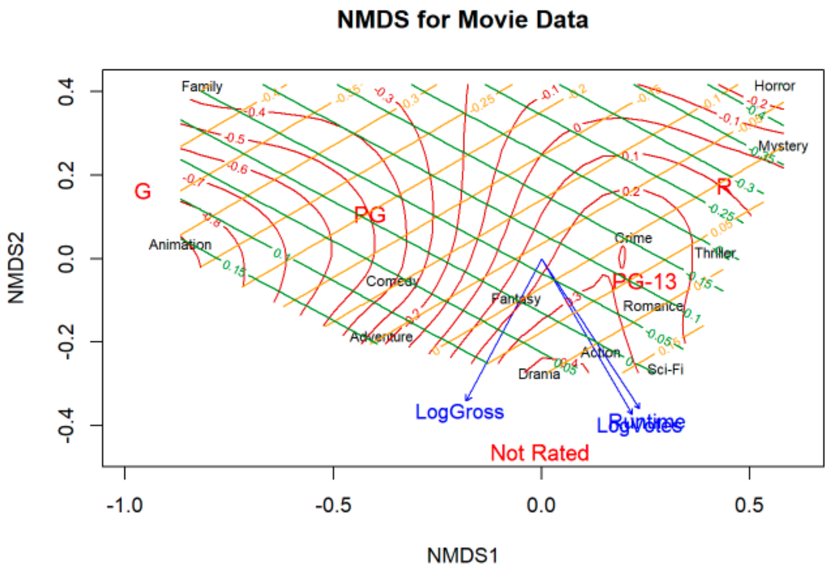


Family movies, and comedies are closest to the G and PG ratings while horror and mystery are closest to the R rating. Crime and thriller are close to the PG-13 rating as well as the R rating. Action, drama, romance, sci-fi, and fantasy movies are clearly predominantly rated PG-13. Adventure is split between PG-13 and PG ratings. The “Not Rated” category is closest to the more popular genres like adventure, drama, and action which contain the most unrated movies.

To explore the relationship between genre, rating, and the continuous variables, I will overlay additional variables. Before that, I will calculate their p-values to see which ones are significant:

```
***VECTORS

              NMDS1     NMDS2     r2   Pr(>r)
Year       -0.24083  0.97057 0.0384 0.80619
Runtime     0.54765 -0.83671 0.5947 0.02198 *
IMDb.Score  0.05011 -0.99874 0.2345 0.25475
Metascore  -0.36674 -0.93032 0.1402 0.46254
LogVotes    0.50401 -0.86370 0.6032 0.01099 *
LogGross   -0.47061 -0.88234 0.4781 0.02897 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
Permutation: free
Number of permutations: 1000
```

Only `Runtime`, `LogVotes`, and `LogGross` are significant (with p-values less than 0.05), so I will only overlay those variables. I will also create non-linear wireplots of the overlaid variables (red for `Runtime`, orange for `LogVotes`, and green for `LogGross`).



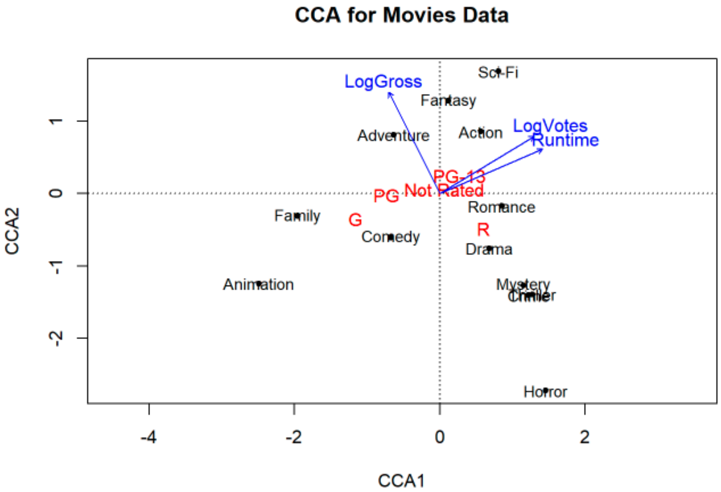
It can be easily observed that there is a nonlinear relationship between runtime and the NMDS dimensions while votes and gross are more linearly related to the dimensions. The contour lines as well as the directions of the overlaid variables seem to suggest that G-rated movies, like family and animation, tend to have shorter runtimes and lower number of votes (perhaps because these movies are often targeting children and parents). R-rated movies, such as horror and mystery, seem to have the lowest gross which is partly due to the fact that they can be viewed by a smaller or more niche audience. In contrast, PG-13 movies that belong to mainstream genres, like adventure, action, drama, and sci-fi, tend to not only have the highest runtimes but also the highest gross and number of votes.

For comparison, I also consider CCA (I have omitted CA (correspondence analysis) because it provides nearly identical information as CCA while CCA also provides insight into the relationship between genre, rating, and other variables). First, let’s look at the summary statistics of CCA:

```
Call:
cca(X = movies.count, Y = movies.env)

Partitioning of scaled Chi-square:
              Inertia Proportion
Total         0.43050    1.00000
Constrained   0.40397    0.93839
Unconstrained 0.02652    0.06161
```

```
Accumulated constrained eigenvalues
Importance of components:
                        CCA1   CCA2     CCA3     CCA4
Eigenvalue            0.3176 0.0811 0.003880 0.001348
Proportion Explained  0.7863 0.2007 0.009605 0.003337
Cumulative Proportion 0.7863 0.9871 0.996663 1.000000
```

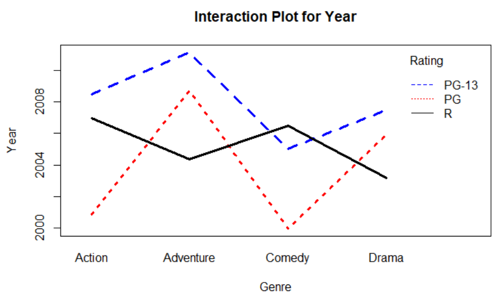
The total inertia equals 0.4305 while the constrained inertia (inertia explained by the continuous variables) constitutes 93.839% of the total inertia. Additionally, the first dimension accounts for 78.63% of the total inertia while the first and the second dimension explain 98.71% of the total inertia. Therefore, CCA also indicates that two dimensions are likely ideal for understanding the relationship between genres and ratings. Below I have included a plot of the results:

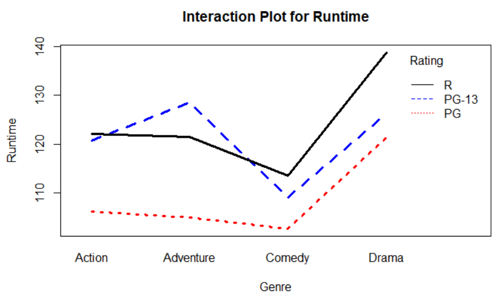


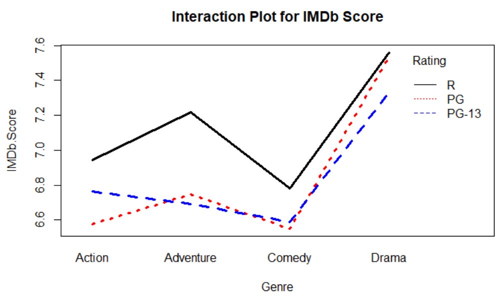
Once again, we see that the directions of the continuous variables indicate that action, adventure, fantasy, and sci-fi movies have higher gross, number of votes, and runtime, while family and animation tend to have lower runtime and number of votes, and horror, thriller, crime, and mystery movies have lower gross.

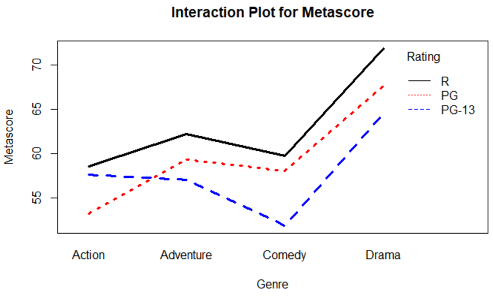
MANOVA

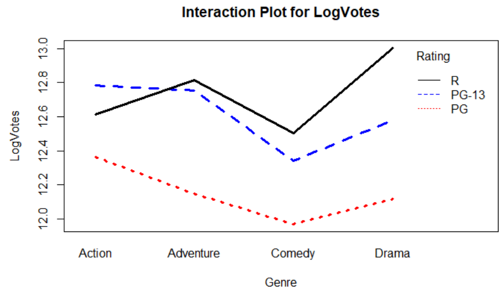
To perform MANOVA, I will only focus on three ratings: PG, PG-13, and R. Additionally, I will replace each movie's collection of genres with the most popular genre it contains among all the genres in the dataset (for example, a movie that is both "Action" and "Sci-Fi" will be assigned "Action"). When this task is performed, most of the movies are either adventure, action, comedy, or drama, so only these four genres are retained. Now, I will create interaction plots for genre and rating against each of the continuous variables:

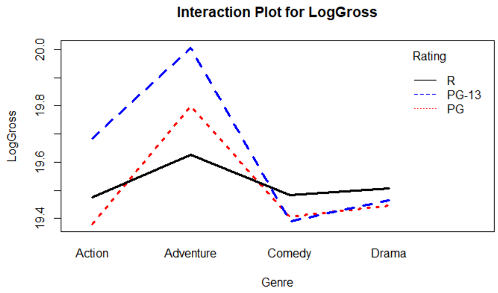












All interaction plots suggest there might be an interaction between genre and rating. It seems that among the top grossing movies of all time the more recent ones are PG-13 or R-rated in the action and comedy genres while older movies are typically PG in the action and comedy genre and R in the adventure and drama genre. Additionally, R-rated and PG-13-rated movies tend to have higher runtime and log votes than PG movies. In the interaction plots for IMDb score metascore, R-rated and PG-rated movies, except for the action genre as well as the comedy score in the IMDb score plot, tend to have higher values than PG-13-rated movies. However, PG-13-rated and PG-rated movies seem to have higher log gross than R-rated movies, at least in the action and adventure genres and a similar one in the comedy and drama genres. In general, it appears that adventures and dramas tend to have the highest values in all categories, except year, with dramas tending to have the highest IMDb scores and metascores while adventures tending to have the highest log gross.

Let's now perform a two-way MANOVA for genre and rating:

```
Type III MANOVA Tests:

Sum of squares and products for error:
                  Year    Runtime IMDb.Score   Metascore  LogVotes  LogGross
Year        75169.9897  -5138.560  -579.0198   -6411.920 -672.3227  674.3057
Runtime     -5138.5598 297767.867  5540.2540   76945.594 4426.6794 2440.7901
IMDb.Score   -579.0198   5540.254   600.7466    8128.223  394.2128  127.0733
Metascore   -6411.9203  76945.594  8128.2234  192506.522 5082.8867 1871.2764
LogVotes     -672.3227   4426.679   394.2128    5082.887  552.7671  175.0569
LogGross      674.3057   2440.790   127.0733    1871.276  175.0569  213.9798

------------------------------------------

Term: (Intercept)

Sum of squares and products for the hypothesis:
                  Year    Runtime IMDb.Score   Metascore   LogVotes    LogGross
Year        1357170786 79809759.4 4696992.63  40702722.2 8457720.01 13232352.41
Runtime       79809759  4693291.2  276211.26   2393563.5  497364.52   778141.46
IMDb.Score     4696993   276211.3   16255.68    140866.9   29271.08    45795.46
Metascore     40702722  2393563.5  140866.86   1220709.7  253654.32   396849.66
LogVotes       8457720   497364.5   29271.08    253654.3   52707.46    82462.38
LogGross      13232352   778141.5   45795.46    396849.7   82462.38   129014.82

Multivariate Tests: (Intercept)
                 Df test stat approx F num Df den Df     Pr(>F)
Pillai            1      1.00  2654994      6    847 < 2.22e-16 ***
Wilks             1      0.00  2654994      6    847 < 2.22e-16 ***
Hotelling-Lawley  1  18807.51  2654994      6    847 < 2.22e-16 ***
Roy               1  18807.51  2654994      6    847 < 2.22e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

------------------------------------------
```

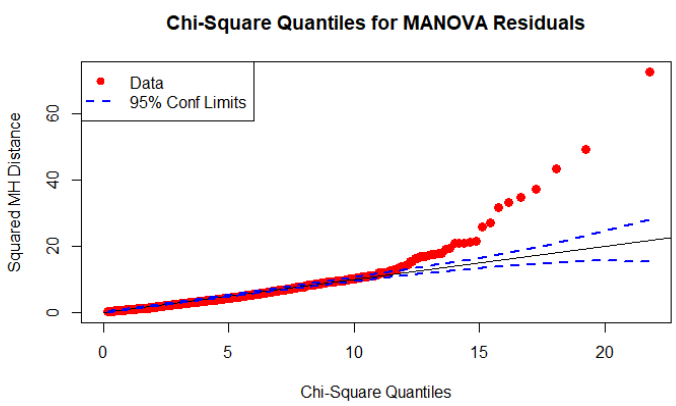
```
Term: Genre

Sum of squares and products for the hypothesis:
                 Year    Runtime  IMDb.Score   Metascore   LogVotes    LogGross
Year       1241.59824  2111.4071  42.3860289  517.294705  73.160819 118.5961241
Runtime    2111.40711 17517.0353 695.9499074 9458.310669 276.900083 109.3753396
IMDb.Score   42.38603   695.9499  30.6415299  431.814363   8.093405   0.1547925
Metascore   517.29470  9458.3107 431.8143626 6214.789696  94.835202  -1.6738274
LogVotes     73.16082   276.9001   8.0934054   94.835202   7.204273   5.7387095
LogGross    118.59612   109.3753   0.1547925   -1.673827   5.738709  11.9865869

Multivariate Tests: Genre
                 Df test stat  approx F num Df   den Df     Pr(>F)
Pillai            3 0.1711908  8.563141     18 2547.000 < 2.22e-16 ***
Wilks             3 0.8367993  8.655387     18 2396.163 < 2.22e-16 ***
Hotelling-Lawley  3 0.1855803  8.718836     18 2537.000 < 2.22e-16 ***
Roy               3 0.1065123 15.071496      6  849.000 < 2.22e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

------------------------------------------

Term: Rating

Sum of squares and products for the hypothesis:
                  Year     Runtime IMDb.Score   Metascore   LogVotes    LogGross
Year        1338.13774  1446.46742 -65.742913 -1509.53332 50.6927485  47.8772907
Runtime     1446.46742 10183.13429 143.602526  1561.37208 387.2309576  29.6123257
IMDb.Score   -65.74291   143.60253   8.576193   153.68705  5.7886335  -2.9036310
Metascore -1509.53332  1561.37208 153.687047  2885.76687 65.9642510 -62.2117352
LogVotes     50.69275   387.23096   5.788633    65.96425 14.7415267   0.9598212
LogGross     47.87729    29.61233  -2.903631   -62.21174  0.9598212   1.7698766

Multivariate Tests: Rating
                 Df test stat  approx F num Df den Df     Pr(>F)
Pillai            2 0.1139478  8.538804     12   1696 6.1449e-16 ***
Wilks             2 0.8889068  8.561687     12   1694 5.4733e-16 ***
Hotelling-Lawley  2 0.1217659  8.584499     12   1692 4.8769e-16 ***
Roy               2 0.0831392 11.750340      6    848 1.1103e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

------------------------------------------
```

According to all multivariate tests, the main effects of `Genre` and `Rating` as well as the interaction term are all statistically significant. The univariate tests suggest that `Genre`, `Rating`, and their interaction all have a significant effect on `Year`, `Runtime` and `LogVotes`. We do not have evidence to conclude that the interaction term is statistically significant for `IMDb Score` and `Metascore` or that the main effect of `Rating` is significant for `LogGross`.

Considering that only the main effect of `Rating` did not appear to be significant for `LogGross`, I decided to perform univariate contrasts on `LogGross` and compare each pair of ratings as well as each rating to the other two. It turns out that there is a significant difference between the log gross of PG-13 movies and R-rated movies, and also between PG-13 movies and the other two ratings (PG and R). This result makes sense given that R-rated movies only allow older audiences while PG-rated movies, in contrast, might not attract older audiences, depending on the maturity of the content.

I also performed multivariate contrasts on all continuous variables (year, runtime, IMDb score, metascore, log votes, log gross) against each pair of genres and ratings. The results showed that all pairs of genres are significantly different in terms of the dependent variables except action and comedy. When it comes to ratings, only R and PG-13 ratings have a significant difference in terms of the dependent variables. For genres, a possible explanation is that many blockbuster actions incorporate elements of comedy, and vice versa. For ratings, the simplest explanation would be that PG-13-rated movies tend to be more mainstream while R-rated movies tend to be more niche, and, as we saw above, PG-13-rated movies tend to perform better especially in terms of popularity metrics, like gross.

The diagnostic plot for the MANOVA residuals is shown below:



Overall, the residuals fit a multivariate normal distribution relatively well. Transformations of the variables do not necessarily improve the fit, so the data is left as it is.

Finally, I will run MRPP tests on the genre and rating variables to compare the multivariate means between groups:

```
Call:
mrpp(dat = movies_trans2[, c(2, 4, 6:9)], grouping = movies_trans2$Genre)

Dissimilarity index: euclidean
Weights for groups:  n

Class means and counts:

      Action Adventure Comedy Drama
delta 32.69  35.79     30.47  37.83
n     156    466       121    121

Chance corrected within-group agreement A: 0.03319
Based on observed delta 34.77 and expected delta 35.96

Significance of delta: 0.001
Permutation: free
Number of permutations: 999
```

```
Call:
mrpp(dat = movies_trans2[, c(2, 4, 6:9)], grouping = movies_trans2$Rating)

Dissimilarity index: euclidean
Weights for groups:  n

Class means and counts:

      PG    PG-13 R
delta 32.32 33.26 37.73
n     235   444   185

Chance corrected within-group agreement A: 0.05553
Based on observed delta 33.96 and expected delta 35.96

Significance of delta: 0.001
Permutation: free
Number of permutations: 999
```

The significance of delta (0.001) suggests that the multivariate means across genre groups are statistically significantly different. Likewise, the multivariate means for movies across the groups of ratings are also significantly different.

Cluster Analysis

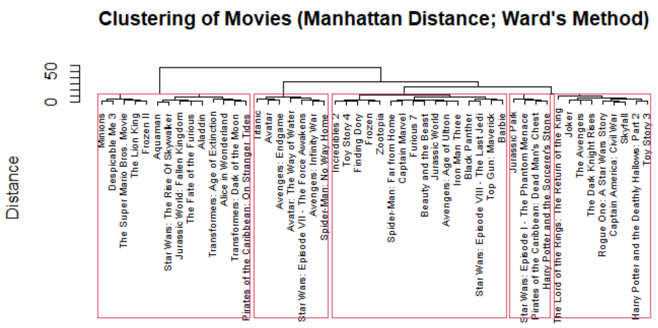
Finally, I will apply cluster analysis to my data. More specifically, I will focus on the 50 highest grossing movies of all time. For this approach, I am only clustering the movies using the continuous variables but I will also explore if the same genres and ratings are grouped together. I will standardize the variables again so that they are on the same scale.

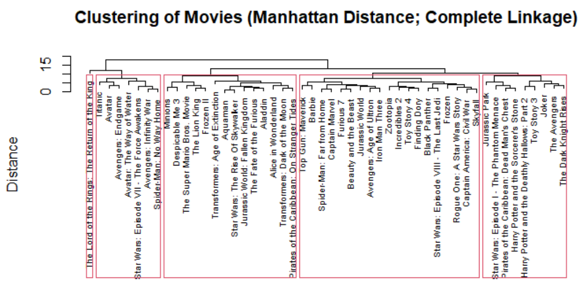
Before that, let's look at the distribution of genres and ratings:

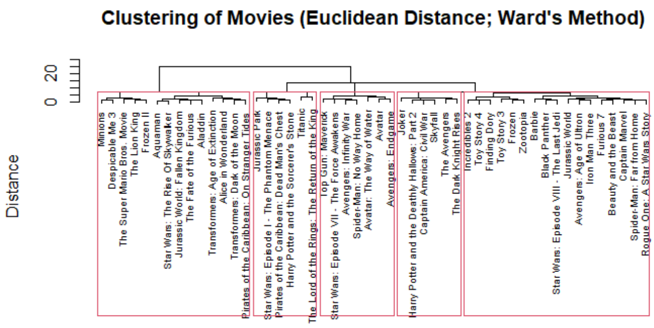
```
Adventure    Action   Fantasy    Sci-Fi    Comedy Animation     Drama    Family  Thriller
       42        32        14        14        12        11         7         5         5
    Crime   Romance
        3         1

PG-13    PG     G     R
   33    14     2     1
```

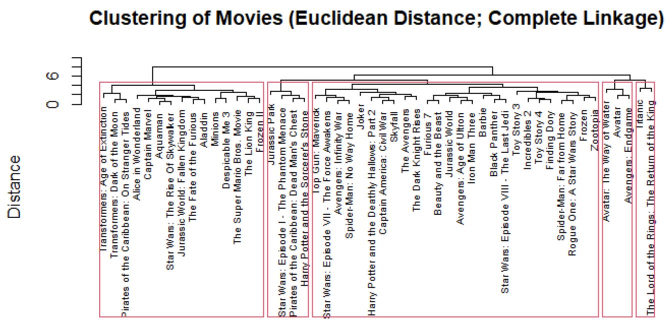
It is clear most of the movies are rated PG-13 and part of the adventure or action genres. Nevertheless, we can make inferences about whether films from less frequent genres (e.g., fantasy, sci-fi, comedy, animation) and ratings (e.g., PG) are grouped together. Now I proceed with the methods.

First, I will use hierarchical cluster analysis. Since I am dealing with continuous variables, the types of distance metrics that could be used include Euclidean, Manhattan, Minkowski, squared Euclidean, and maximum distance. I will only focus on Euclidean and Manhattan distance, given that they are similar to and yet more commonly used than Minkowski (which is a generalization of these two methods) and squared Euclidean due to their relative simplicity and interpretability. Additionally, maximum distance prioritizes variables that have the most variability which is not my goal as I want to consider all variables evenly. Apart from the two distance metrics, I will also use two agglomeration methods, Ward's Method and complete linkage. For each pairing of a distance metric with an agglomeration method I will create five clusters. The four dendrograms are shown below:



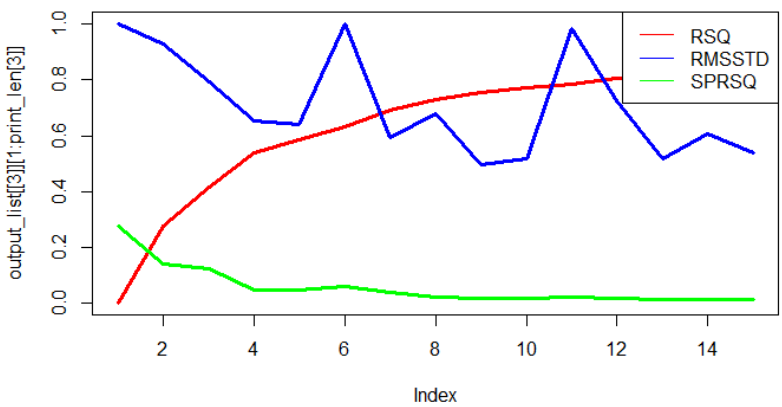




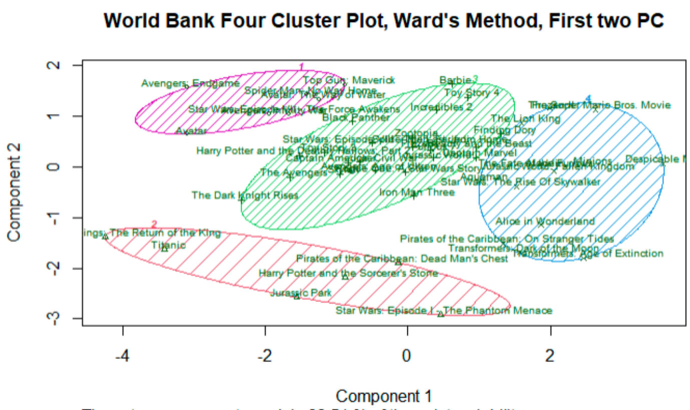


I will proceed with an analysis of the clustering method using Euclidean distance and Ward's method because the distribution of films within clusters is most even when this approach is used. In terms of how the movies are grouped, it appears that movies that target older audiences (like sci-fi or action) tend to be in the same cluster and, similarly, those that are typically aimed at younger demographics (especially animation) are clustered together. Movies that are in the same genre or that are part of the same franchise tend to be in the same group as well, although that is not always the case. Notably the three highest grossing movies (*Avatar*, *Avengers: Endgame*, *Avatar: The Way of Water*) are always clustered together and even form their own group in the case of the combination between Euclidean distance and complete linkage. This is probably because even after transformation and standardization these three movies have much higher gross values than the rest in the list.

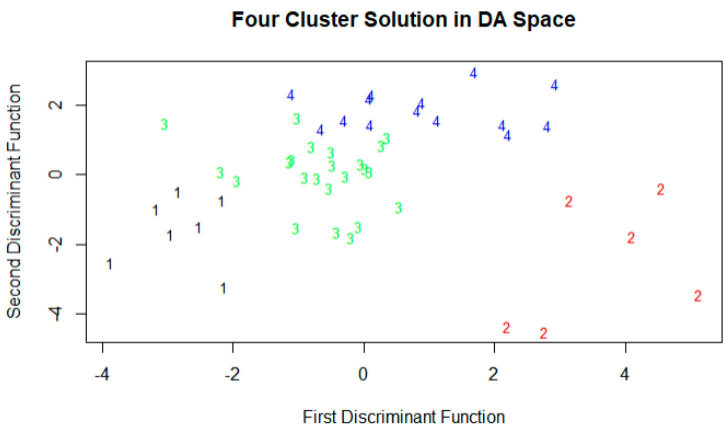
The plots for root-mean-square standard deviation, R-Squared, and semi-partial R-Squared statistics suggest that 4 or 5 clusters should be sufficient.



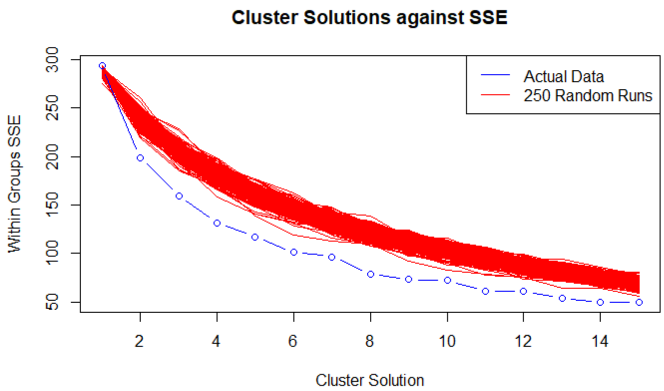
Additionally, it seems that the movies can be clustered really well in 4 groups in the discriminant analysis space and in the PCA space, as shown below.

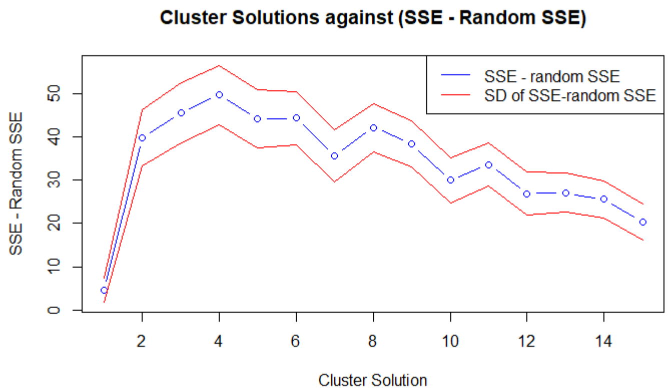


These two components explain 68.51 % of the point variability.

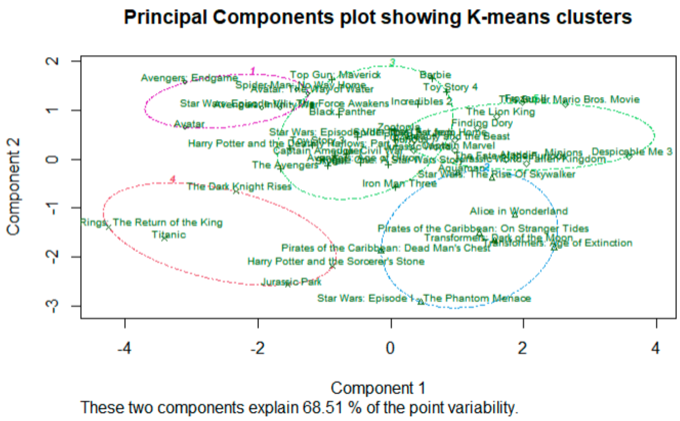


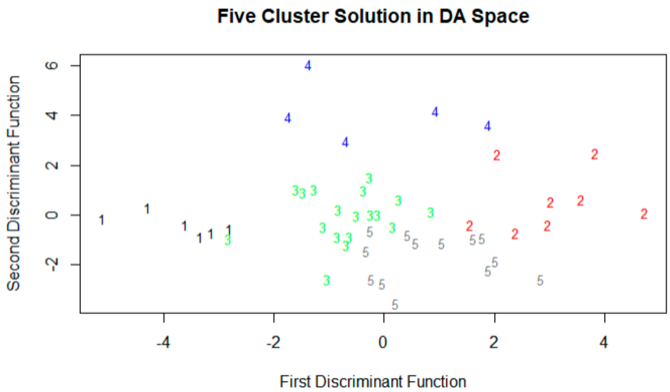
I will now run k-means clustering on the movie data and plot the within groups SSE (sum of squared error) for actual and randomized data against the number of cluster solutions. I will also plot the difference between actual SSE and mean SSE from 250 randomized datasets.





The two plots indicate that between 5 and 7 clusters should be enough for k-means clustering. It also seems that if this method is used, the movies can be clustered relatively well in 5 groups in the PCA space and the discriminant analysis space, as shown below.





Overall it seems that both clustering methods are good at partitioning the movies into groups, and many of the movies that are grouped together based on the first method are also grouped together according to the second one, and vice versa. Both methods create two bigger groups along with two or three smaller sized groups. In general the two approaches are quite accurate at clustering movies based on their characteristics, with hierarchical cluster analysis having a slight edge over k-means clustering. It is clear that along with the predominant genres (action and adventures) and ratings (PG-13), the movies from less popular genres or ratings are also typically clustered together. For example, one of the clusters produced by hierarchical cluster analysis using Euclidean distance and Ward's method contains primarily PG-rated movies and animations while another contains primarily fantasy and sci-fi movies. The other two contain action and adventure movies, and, as mentioned above, one of those groups contains most of the top grossing movies. Thus, cluster analysis is somewhat good at grouping movies by genre and rating, which in turn means that the highest grossing movies from the same genre and rating tend to also have similar year of release, runtime, IMDb score, metascore, number of votes, and gross.

Conclusions and Discussion

.As suggested by the interaction plots, there seems to be an interaction between genres and ratings. This was suggested by the ordination methods which showed that certain genres tend to be closer to particular ratings than others. For example, family and animation movies are usually G-rated genres while horrors and thrillers are mostly R-rated. These results make sense because most films concern topics that are usually appropriate for one or two ratings.

Using two-way MANOVA, it was discovered that genre and rating as well as their interaction have a significant effect on the combination of all continuous variables (year, runtime, IMDB score, metascore, votes, gross) and, for the most part, on each of the individual variables. Additionally, multivariate contrasts were used to verify that there is a significant difference between the multivariate means of most pairs of genres and ratings (only exceptions were action and comedy for genres and R and PG as well as PG and PG-13 for ratings). The univariate contrasts showed that there is a significant difference between the log gross means for movies with ratings PG-13 and R but not for PG and another rating. The MRPP tests confirmed that the multivariate means for genres and ratings are each significantly different. Like MANOVA and related methods, ordination also showed that runtime, log votes, and log gross, in particular, are influenced by genres and ratings, by showing the higher runtime, votes, and gross of PG and PG-13 rated movies in the popular genre categories, such as adventure, drama, action, sci-fi, comedy, fantasy, and romance, compared to the lower gross of horror and mystery movies and the lower runtime of animation and family movies.

While cluster analysis is not as explicit at relating genres and ratings to each other or showing how they are related to the continuous variables, it was observed that especially with hierarchical cluster analysis that combines Euclidean distance metric and Ward's method, similar genres and ratings do, in fact, tend to be placed in the same groups. These results indicate that even in the top 50 grossing movies, among the entire list of highest grossing movies, there is a tendency for movies that have the same genre and rating to have comparable continuous metrics.

Points for Further Analysis

I believe it would be interesting to see if similar tendencies exist beyond the highest grossing movies of all time and if the effects of genres and ratings on a movie's characteristics are more or less pronounced than they are for the most successful movies. Moreover, I think it would be useful to look at other categorical variables, like notable actors or directors, whose names might bring additional recognition to a film, and, therefore, improve its reception or performance.