

Board Game Popularity

Crazy Eights - Kate, Bella, Daniel, Shreyans

2023-12-01

Introduction

In an increasingly digital world, board games have facilitated a renewed interest in face-to-face communication, serving as a common human leisure activity that combines entertainment, social interaction, and intellectual challenge. In a 2018 study, Gonzalo-Iglesia and colleagues examined how Spanish students in Communication and Biochemistry studies engaged with commercial board games and discovered that engaging with board games promoted learning, communication, decision-making, and teamwork (Gonzalo-Iglesia et al., 2018).

Having understood the benefits of board games as a learning tool and beyond, the motivation for our research question originates from the rapidly increasing sales of board games, as enthusiasts often find themselves navigating through an abundance of unique gaming choices. This led us to wonder: If a person walks through an aisle of board games, what catches his or her attention first? More importantly, what makes a board game successful and enjoyable? How do various factors influence a game's reception among players?

In this analysis, we will focus on a singular broad research question: What qualities constitute a good board game? We'll be looking at elements like game themes, categories, mechanics, popularity, and game structure where a higher average ranking denotes a better board game (i.e. Monopoly, Uno, Puzzles).

Before our exploratory data analysis, we crafted some hypotheses based on the context of board games. (1) Games that are owned more and have lower ranks will have a positive relationship with average user rating because popularity typically increases rating. (2) Board games from the fighting theme may have a positive relationship with average user rating because of dedicated fan bases. (3) While different players might prefer more different levels of difficulty, we hypothesize that difficult games will have a negative relationship with average user rating, on average, because it takes away from the fun and leisureness of board games.

The goal is to provide insights that can help both gamers and game creators navigate the world of board games more effectively.

Data description

This board games dataset involves data collected from BoardGameGeek(BGG), an online board game forum and database that stores information and ratings for over 125,600 different board games. The data was originally collected by web-scraping BGG’s website using BGG’s XML API, but it was sourced from Kaggle for the purpose of this analysis.

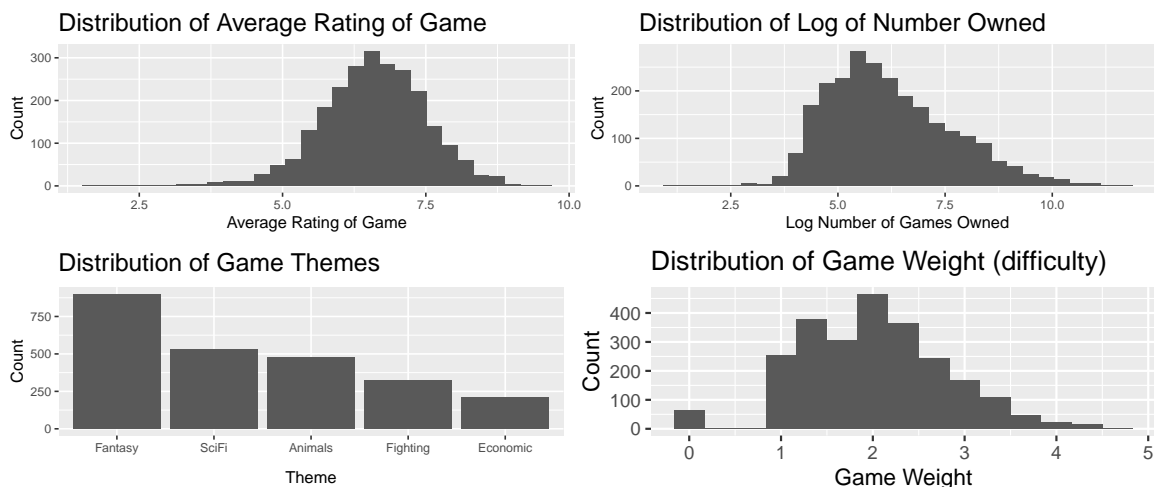
The initial board games data was split among several different files, but through data cleaning, merging, and variable selection, we reduced the size of the data significantly [see Section 3: EDA]. The final games dataset we use in our analysis has 1878 observations and 28 columns where 2 columns are identifiers (BBGid and name). There are characteristics that describe the structure of the game such as number of players, recommended time to play, and recommended age that are determined by the game designers. There are also more subjective characteristics that reflect the public opinion of each board game such as number of games owned/wanted, average rating, difficulty level, and average rank. Lastly, we also have a set of categorical variables that highlight what theme and subcategory each board game belongs to which gives more insight into the type of game and intended audience.

Initial exploratory data analysis

We performed a lot of data wrangling in order to prepare our data for exploratory data analysis and modeling. We started with three separate datasets (games, themes, subcategory) where the “games” dataset was our target dataset we wanted to tidy/merge to based on “BBGid”. Our methodology is detailed below:

1. To remove the number of dimensions and reduce future model bias from underrepresented categories and themes, we decided to only include board games in the top five populated themes and categories from each dataset. In order to make categorical distinctions more clear, we also removed all board games that belonged to multiple categories. Some underrepresented categories include Word Games and Territory Building and some removed themes include Medical and Sports. While it would have been nice to include these, we wanted to focus on a subset of data (see Limitations).
2. We performed an inner-join to merge all three datasets. Columns were renamed to remove spacing and special characters.
3. We noticed that many observations for MaxPlayers and MfgPlaytime were zero, but it did not seem reasonable for these parameters to be 0 players and 0 minutes, respectively, in the context of board games. We concluded that the 0’s may have been used as the default if no data was provided. Therefore, we performed imputation to set the 0 values to the mean of the column (of all non-zero values).
4. Since our themes and categories were represented as indicator variables, we created a new factor column for each that classified each board game to aid in graphing.

Distribution AvgRating (response), log of NumOwned (quantitative variable), Theme (categorical variable), GameWeight (quantitative variable)



The distribution of Theme shows that the most popular game type is Fantasy by a significant margin. There at least 200 more Fantasy games than the next most popular game theme of Science Fiction.

Summary Statistics for Histogram of Average Rating

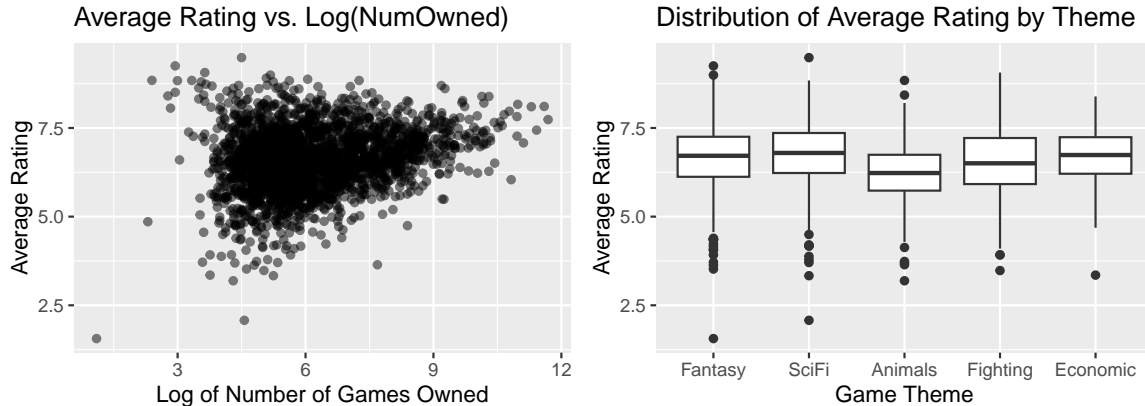
mean	sd	q0	q25	median	q75	q100	range
6.591	0.896	1.562	6.018	6.609	7.196	9.483	7.922

The distribution of AvgRating is approximately normal and unimodal. Thus, the best measure of the center is the mean, which is 6.639 points, and the standard deviation of 0.844 indicates that there is little variability in the data. There are no outliers shown in the histogram.

The distribution of NumOwned was highly right skewed with a lot of points near 0, so we decided to take the log transformation. The distribution of the log of NumOwned is approximately normal and unimodal. Thus, the best measure of the center is the mean, which is 6.497 points. The standard deviation of 1.479 indicates that there is a moderate amount of variability in the data. There are no outliers shown in the histogram. [see Appendix for summary stats]

The distribution of GameWeight is slightly right-skewed, so the best measure of the center is the median, which is a difficulty of 2. The best measure of the spread of 50% of the data is the IQR of 1, suggesting there is some moderate variability in the data. There is an interesting shape to the distribution because there are not any observations between 0 and 1, while there are many observations between 1 and 5. [see Appendix for summary stats]

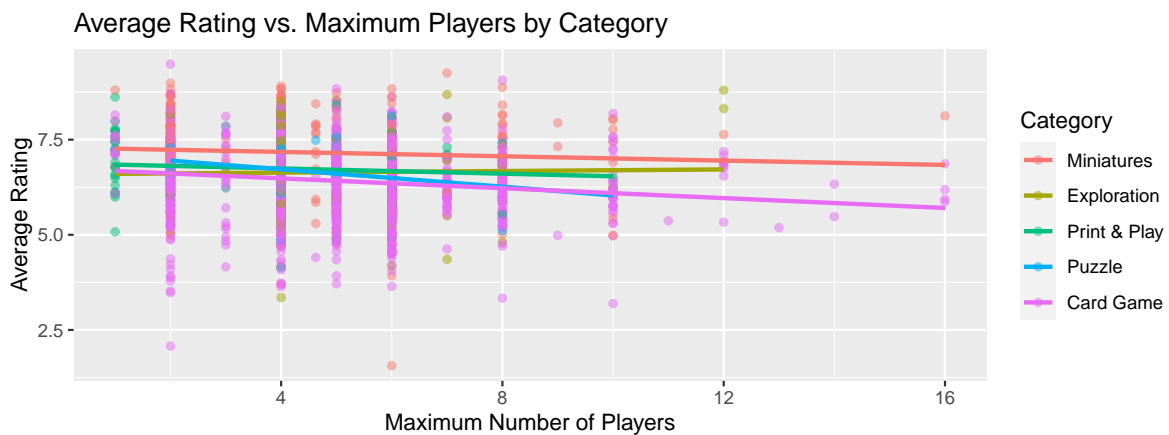
Plots of Response Variable and Predictor Variables



When plotting the number of games owned (continuous predictor) with the average user rating (response), we used a log transformation again. We can observe a moderate, positive, linear relationship between the log of number of games owned and user rating. However, there is a cluster on the top left, which suggests that the effect of the number of games owned on user rating is stronger when the number is very high.

When plotting the board game themes (categorical predictor) against the average user rating (response), we can observe that the median of average user rating for all themes fall between 6 and 7 with Fantasy, Scifi, and Economic themes above the 6.5 mark. Ranges and IQRs are pretty consistent as ratings typically fall between 5-9 with 50% of ratings in each theme within 0.75 points off the median. There are a few board game outliers with significantly lower ratings in each theme.

Visualizations of Potential Interaction Effects



In the plot, we wanted to visualize how the category would affect the maximum number of players because if the category was too difficult to play with multiple people, then the average

rating would decrease. Therefore, we want to quantify effect of category on the relationship between the maximum number of players and average rating. The differing slopes for Category suggest that there could be a significant interaction effect between MaxPlayers and Category.

Correlation Matrix

We can use the correlation matrix [see Appendix] to help motivate what variables we add to the model and check for multicollinearity. Based on the matrix above, AvgRating seems to have a relationship with Rank and GameWeight. In addition, there are moderate correlations between Rank x NumOwned, GameWeight x Playtime, and NumOwned x NumWant x NumWish.

Methodology

We are using a multiple linear regression model because we have multiple predictors and the response variable of our model, AvgRating, is a continuous value.

Creating a Baseline Model (Model 1): When we are considering the baseline for the rating of a certain board game, we are looking at classic hallmarks of success such as popularity i.e. number of sales and the rank that board game critics may give a game. Based on the information provided by BGG, how the values of Rank were determined is unclear, however we believe there is a relationship between AvgRating and Rank, so we included it in the model.

term	estimate	std.error	statistic	p.value
(Intercept)	6.585	0.016	415.983	0
Rank	-0.013	0.000	-35.151	0
logNumOwned	-0.257	0.016	-16.467	0

Breaking Down the Recipe 1: Just as we logged the NumOwned variable in our EDA, we decided to apply a log transformation to create logNumOwned. Following this, we also removed the original NumOwned predictor from the model.

We also decided to step_mutate and scale the rank down by 100. This was motivated by the fact that the rank variable had a much larger range (9 to 21,926) than our avgRating response variable (1 to 10), making the coefficient too small to interpret without re-scaling. Mean-centering was also used so we can better interpret the intercept and reduce VIF.

Creating a Second Model (Model 2): Analysis of the correlation matrix helped us to figure out which predictors would have the greatest correlation with the average rating of a board game. We created a model built off of the baseline model that adds the top 5 Themes, top 5 Categories, GameWeight, and MaxPlayers.

The top 5 themes and categories were selected because they would have the greatest impact on the rating of a game. Analyzing every single theme and category available would create too many levels and minimize the impact of each level.

The difficulty of the game was added because we believed that the difficulty would have an impact on how the user enjoys the game. If the difficulty is too high, the game may not be enjoyable for the consumer and the average rating may be lowered. In addition, our correlation matrix confirmed this relationship. The maximum number of players was added because we believed that the rating of a game may be lowered if people felt that the maximum number of players for the game was too limiting.

Breaking Down the Recipe 2: Based on our EDA, we decided to add an interaction term between game Category and MaxPlayers because the difference in the slopes may be statistically significant.

term	estimate	std.error	statistic	p.value
(Intercept)	6.963	0.083	84.141	0.000
GameWeight	0.167	0.022	7.446	0.000
Rank	-0.011	0.000	-31.102	0.000
MaxPlayers	-0.021	0.017	-1.241	0.215
logNumOwned	-0.228	0.015	-15.046	0.000
MaxPlayers_x_CategoryExploration	0.080	0.037	2.130	0.033
MaxPlayers_x_CategoryPrint & Play	-0.024	0.039	-0.620	0.535
MaxPlayers_x_CategoryPuzzle	-0.035	0.071	-0.493	0.622
MaxPlayers_x_CategoryCard Game	0.011	0.020	0.560	0.575
Category_Exploration	-0.652	0.185	-3.521	0.000
Category_Print...Play	-0.144	0.168	-0.855	0.393
Category_Puzzle	-0.144	0.315	-0.457	0.648
Category_Card.Game	-0.387	0.096	-4.045	0.000
Theme_SciFi	-0.047	0.042	-1.127	0.260
Theme_Animals	-0.153	0.044	-3.444	0.001
Theme_Fighting	-0.027	0.048	-0.558	0.577
Theme_Economic	-0.076	0.058	-1.310	0.190

Creating a Third Model (Model 3): In this model, we got rid of the interaction effect between MaxPlayers and Category to check if the interaction effect is actually useful. We observed that the slopes, while different, may not be statistically significant. Additionally, when analyzing the output of model 2, the p-values for all but one of the coefficients in the interaction term between MaxPlayers and Category were very high.

term	estimate	std.error	statistic	p.value
(Intercept)	6.913	0.046	149.967	0.000
GameWeight	0.165	0.022	7.360	0.000
Rank	-0.011	0.000	-31.089	0.000
MaxPlayers	-0.011	0.008	-1.407	0.160
logNumOwned	-0.228	0.015	-15.102	0.000
Category_Exploration	-0.280	0.066	-4.243	0.000
Category_Print...Play	-0.219	0.093	-2.363	0.018
Category_Puzzle	-0.293	0.090	-3.248	0.001
Category_Card.Game	-0.342	0.046	-7.369	0.000
Theme_SciFi	-0.051	0.042	-1.239	0.216
Theme_Animals	-0.158	0.044	-3.578	0.000
Theme_Fighting	-0.031	0.048	-0.643	0.520
Theme_Economic	-0.083	0.058	-1.441	0.150

Cross Validation and Model Selection: Based on the size of our training set ($n = 1828$), we performed 12-fold cross-validation on each model using the training data. Across the 12 folds, the average RMSE for the three models were 0.674, 0.643, and 0.642, respectively [see Appendix]. The slightly lower RMSE value indicates that the third model is a better fit for the data than the first two. The average R^2 for the three models were 0.443, 0.492, and 0.494, respectively [see Appendix]. This indicates that the third model explains more variation in the data, so the third model is the best predictor of AvgRating. We also calculated the adjusted R^2 , AIC, and BIC values for each model on the training data to confirm our results. While the adjusted R^2 value is slightly lower for the third model than the second model, the AIC and BIC values are lower for the third model. Moreover, when analyzing the VIF of model 2, we observed multiple predictors to be highly correlated with each other ($VIF > 10$) [see Appendix]. Removing the interaction term removed the multicollinearity between the predictors. Although one of the p-values for the interaction term between MaxPlayers and Category Exploration is statistically significant compared to the baseline Category Miniatures, all other interaction terms were not statistically significant. Since Model 3 is more parsimonious with lower VIFs than Model 2, and the difference in adjusted R^2 is negligible, we chose the third model to predict average board game rating.

Results

Model Assumptions

After checking model conditions (linearity, constant variance, normality, independence, and multicollinearity), we decided that while our data was not perfect, it was reasonable to proceed with mathematical inference [see Appendix for supporting plots and Discussion for limitations].

Understanding the Model

$$\begin{aligned}\widehat{AvgRating} = & 6.913 + 0.165 * GameWeight \\ & - 0.011 * Rank - 0.011 * MaxPlayers \\ & - 0.228 * logNumOwned - 0.280 * Category_Exploration \\ & - 0.219 * Category_Print...Play - 0.293 * Category_Puzzle \\ & - 0.342 * Category_Card.Game - 0.051 * Theme_SciFi \\ & - 0.158 * Theme_Animals - 0.031 * Theme_Fighting \\ & - 0.083 * Theme_Economic\end{aligned}$$

The intercept of 6.913 in the model represents the predicted average rating when a game has a GameWeight of 2.021, Rank of 10,047, MaxPlayers of 4.626, and logNumOwned of 6.21 and the board game is fantasy themed (baseline theme) in the Miniatures category (baseline category).

The coefficients of the categorical predictors shift the intercept by their value when all other predictors are held constant. For example, when all of the predictor variables are held constant, the average rating of a board game is expected to decrease by 0.293 points when a board game's category changes from a Miniatures game (baseline category) to a Puzzle game. The fact that every categorical coefficient is negative suggests that a board game that belongs to Miniatures (baseline category) typically has the highest average rating. The p-values of the coefficients in every category are lower than the significance level of 0.05, providing sufficient evidence that they are all statistically significant relative to the baseline of Miniatures.

However, the only theme with a p-value lower than 0.05 is Animals, meaning that only the Animals Theme is statistically significant relative to the baseline of Fantasy. This means that when a board game's theme changes from Fantasy to Animals, the average rating is expected to decrease by 0.158 points, holding all else constant.

The coefficients of the quantitative predictors represent the expected change in the average rating of a board game for each unit increase in the predictor variables when all other predictors are held constant. For example, for every 1 point increase in the GameWeight of a board game, the expected average rating is expected to increase by 0.165 points, holding all else constant. Similarly, for every 1 unit increase in the log of the number of games owned, the expected average rating is expected to decrease by 0.228 points, holding all else constant. Rank is also negatively correlated with average rating due to a negative coefficient. The p-values for GameWeight, Rank, and logNumOwned are all approximately 0, which is less than the 0.05 threshold, so they are statistically significant predictors, while MaxPlayers is not.

Model Performance

RMSE and Adjusted R^2 for Model 3 (Chosen Model) on Testing Data

.metric	.estimator	.estimate
rmse	standard	0.681

.metric	.estimator	.estimate
rsq	standard	0.43

Let's analyze performance metrics for Model 3 on our testing data. The R^2 is 0.430, suggesting that approximately 43.0% or nearly half of the variation in average board game ratings can be explained by the Category, Theme, logNumOwned, GameWeight, Rank, and MaxPlayers. When we break down our RMSE, the RMSE value of 0.681 means that the average distance between the actual and predicted values of the average rating is 0.681 points. Since average board game ratings have a range from 1-10, the average error of our chosen model is reasonable. The RMSE is not significantly higher on the testing data than on the training data, meaning the chosen model is not overfit (0.681 vs 0.637).

Discussion

Implications

Our model displayed negative coefficients for non-baseline categories and themes, which indicate that games falling outside the fantasy-themed and miniatures category might have lower average ratings. Additionally, higher difficulty tends to improve average ratings while the log of the number of games owned and average rank tends to decrease average ratings.

The negative correlation between logNumOwned and average rating in our model was unexpected, as we initially hypothesized that the increased popularity from a greater number of games owned would be correlated with higher ratings. This could be due to the fact that as the number of games owned increases, the number of raters follows accordingly, and the number of people that dislike the game may increase at a faster rate.

In addition, while we hypothesized that games that are fighting-themed would have higher ratings, our model showed that games that are fantasy-themed tend to have higher average ratings. This tells us that the gaming community tends towards less violent games.

Our analysis provides valuable insight for both board game enthusiasts and board game manufacturers. Game enthusiasts can use this information to guide what game to play since higher ratings are tied to overall enjoyment. Similarly, these findings can help game manufacturers enhance their game development strategies, improve their business models, and increase sales. Given our model, game enthusiasts and game manufacturers might consider purchasing and

producing games that are more of a challenge (higher GameWeight) and focus on elements of miniatures and fantasy themes. The gaming community can also look towards board games with lower rankings and are owned less to get a better idea of a recommended game.

Limitations/Validity

One limitation of our model is that not all of the themes and categories are used, as only 5 are chosen for both, reducing the total number of observations that the model considers. Therefore, the predictions of the model are only generalizable to the themes and categories that are utilized in the model. Additionally, our selected model has an R^2 value of 0.494 on the training data and 0.430 on the testing data. Thus, our model does not explain most of the variability in the data. Making accurate inferences about the average rating of a board game would be difficult using our model. Lastly, when testing inference conditions on our selected model, the plots for linearity and constant variance were not ideal. In the residual vs. fitted plot, there appears to be a very slight fan shape, suggesting that constant variance may be violated. In the plot of the residuals vs. logNumOwned, the residuals looked slightly parabolic and in the plot of the residuals vs. Rank, the residuals looked slightly cubic. Since these conditions may not be satisfied, making inferences about the average ratings of board games may not be appropriate. In addition, since BGG ratings are voluntary, we are getting data from either lovers or haters, which may not be representative of the entire gaming population. However, we still believe the data provides a relatively good indication of a board game's rating.

Future Work/Improvements

Our model can be improved through the combination of different categories and themes which would create different interaction terms with the pre-existing predictors. This would alter the fit of the model to the data and may give better fits that would help the predictive performance and alter the decisions that game manufacturers and consumers make.

Another improvement that we could make would be to add more data from games that are more recent or to find other game features from different datasets to merge into the pre-existing data. This would provide us with more predictors to utilize in our model that may increase the amount of variability explained by our model. One example would be price, as the cost of a game tends to play a large factor in whether people enjoy a game because if a game is too expensive, people may rate it lower even if they enjoyed it.

Conclusion

In conclusion, while our model has a moderate fit and may not be generalizable to every board game, our findings provide general insight to consumers and game manufacturers on the optimal board game. So next time you're choosing a game, consider a difficult, fantasy-themed board game!

Appendix

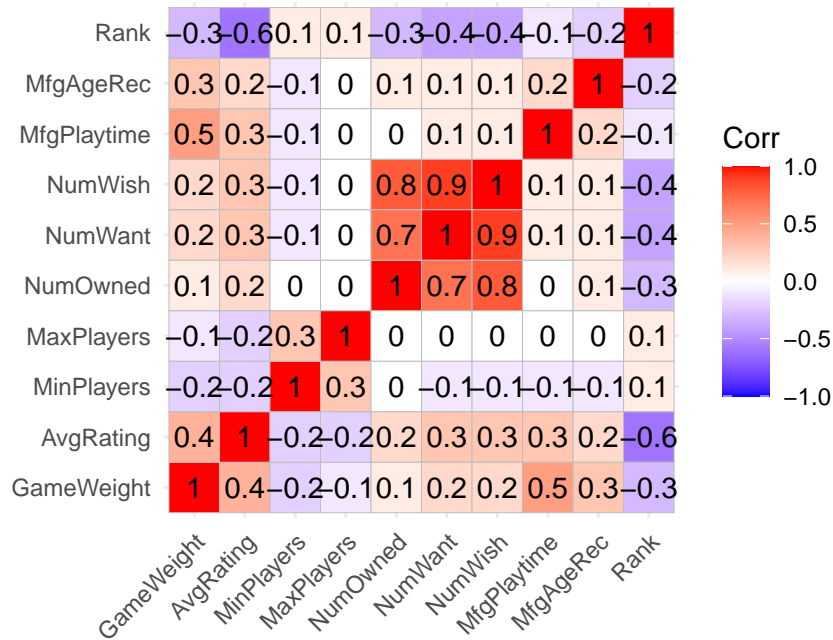
Summary Statistics for Histogram of Log of Number of Games Owned

mean	sd	q0	q25	median	q75	q100	range
6.206	1.452	1.099	5.13	5.965	7.109	11.689	10.59

Summary Statistics for Game Weight Histogram

mean	sd	q0	q25	median	q75	q100	range
2.021	0.789	0	1.5	2	2.5	4.667	4.667

Correlation Matrix



Average R^2 and RMSE across all 12 folds for Model 1

.metric	.estimator	mean	n	std_err
rmse	standard	0.674	12	0.027
rsq	standard	0.443	12	0.041

Adjusted R^2 , AIC, BIC across all 12 folds for Model 1

mean_adj_rsq	mean_aic	mean_bic
0.431	3450.781	3472.477

Average R^2 and RMSE across all 12 folds for Model 2

.metric	.estimator	mean	n	std_err
rmse	standard	0.643	12	0.025
rsq	standard	0.492	12	0.038

Adjusted R^2 , AIC, BIC across all 12 folds for Model 2

mean_adj_rsq	mean_aic	mean_bic
0.492	3273.414	3371.046

Average R^2 and RMSE across all 12 folds for Model 3

.metric	.estimator	mean	n	std_err
rmse	standard	0.642	12	0.025
rsq	standard	0.494	12	0.038

Adjusted R^2 , AIC, BIC across all 12 folds for Model 3

mean_adj_rsq	mean_aic	mean_bic
0.491	3271.627	3347.563

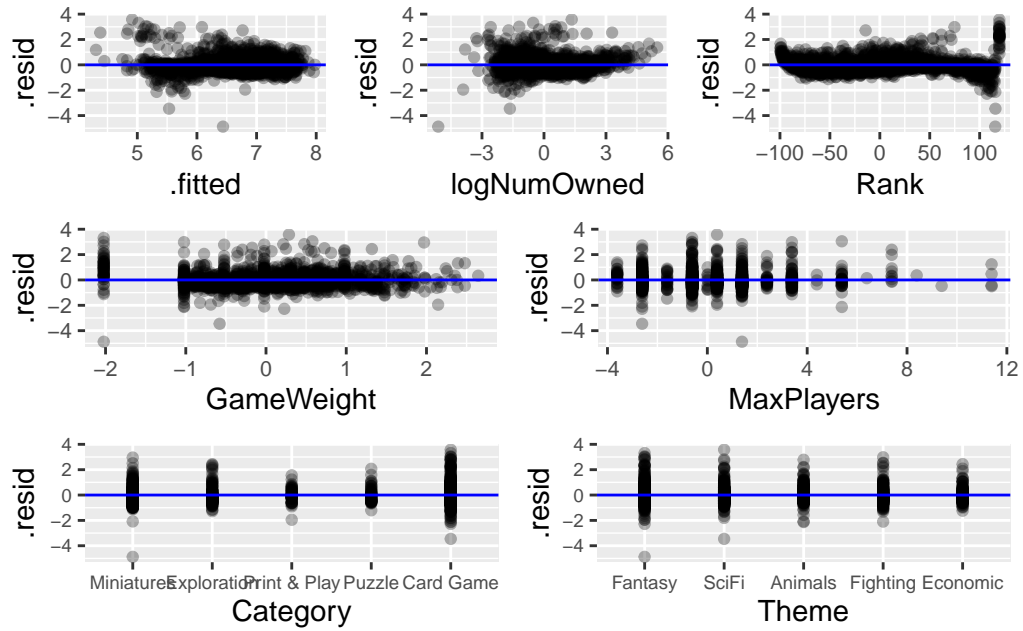
R^2 and RMSE for Model 3 on Training Data

.metric	.estimator	.estimate
rmse	standard	0.637

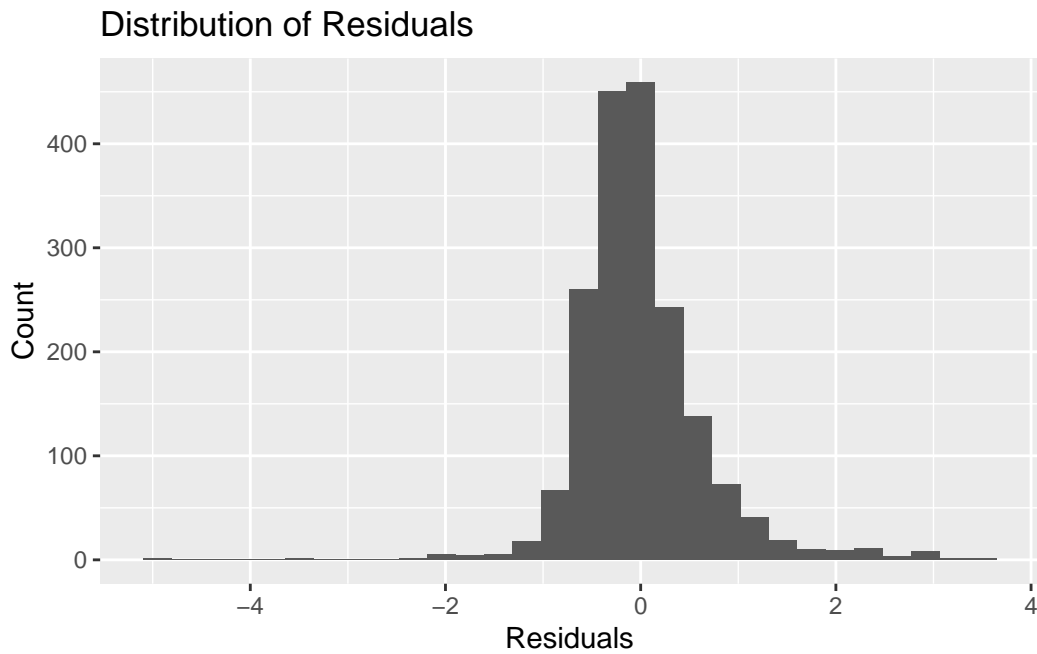
.metric	.estimator	.estimate
rsq	standard	0.494

Model Conditions

Linearity and Constant Variance



Normality



Independence

We have no reason to believe that the qualities of one board game affect the qualities of another board game. Data was collected from one time period, and we controlled for potential clusters such as category in our model. Hence, independence is satisfied.

Multicollinearity

VIF: We looked at the VIF of the 2nd and 3rd model to measure the multicollinearity between the predictors.

	x
GameWeight	1.416
Rank	2.270
MaxPlayers	5.058
logNumOwned	2.125
MaxPlayers_x_CategoryExploration	11.745
MaxPlayers_x_CategoryPrint & Play	3.767
MaxPlayers_x_CategoryPuzzle	15.223
MaxPlayers_x_CategoryCard Game	12.817
Category_Exploration	11.410
Category_Print...Play	3.939
Category_Puzzle	15.400
Category_Card.Game	8.787
Theme_SciFi	1.331
Theme_Animals	1.377
Theme_Fighting	1.203
Theme_Economic	1.178

	x
GameWeight	1.412
Rank	2.265
MaxPlayers	1.067
logNumOwned	2.118
Category_Exploration	1.448
Category_Print...Play	1.197
Category_Puzzle	1.268
Category_Card.Game	2.061
Theme_SciFi	1.324
Theme_Animals	1.369
Theme_Fighting	1.200
Theme_Economic	1.173