Not-So-Bored Games: An Analysis of Board Game Ratings

Abigail Rieck

STATS 401: Applied Statistical Methods II

April 22, 2022

Background

This project sought to better understand what factors make a board game more or less popular, as measured by the average community rating from the gaming community website BoardGameGeek. Data for the project were taken from a tidytuesday post drawn from this website. The dataset included information on a game's category/genre, which describes the general mechanic or theme of each game. Following the example set by tidytuesday user Alyssa Goldberg, the dataset was filtered to exclude games with more than one listed category and then restricted to only the top 5 most popular categories (abstract strategy, card game, dice, economic, and party game) for ease of analysis. Additionally, only games published after 1950 were included in order to exclude classic but irrelevant games such as chess, which would be unlikely to give any interesting insights. The initial model for this dataset sought to predict the average game rating (out of 10 points) through variables measuring a game's maximum expected playtime (in minutes), minimum suggested age to play, title length (in characters), the number of users who reported owning the game, the game's age (in years, as of 2022), and its category.

Analysis

Initial Data Exploration

The mean game rating across the dataset was 6.20 points (out of a possible 10) with a minimum of 2.46 and a maximum of 8.87 points. Initial explorations of the variables of interest revealed a heavy right skew in the maximum playtime variable, which was corrected with a log transformation in subsequent models (Figures 1.1 and 1.2). A similar transformation was performed with the number of game owners and title length, although title length and minimum age both proved to be unsuccessful predictors of rating and were later dropped from the analysis.

The age of the game was originally included in the dataset as the year in which it was published. However, this variable showed a significant left skew that would have made it a poor linear predictor (Figure 2). Additionally, publishing year does not seem as immediately intuitive as the age of a game in years. To correct this, a new age variable was created by subtracting the year each game was published from 2022, the year in which the dataset was created and the year in which the most recent games in the set were published.

Game Category and Interaction With Age

Although average ratings for games did not initially seem to vary significantly across categories (Figure 3), there did seem to be some slight differences worth exploring. This was confirmed in the initial model with p-values for the estimated difference in intercepts (compared to abstract strategy games, the reference group) being very low for all groups, suggesting statistically significant differences between categories and supporting its inclusion in the model.

Various interactions between category and other variables were explored, with the age of the game ultimately seeming to be the strongest candidate for an interaction term. Graphing age vs. average rating by game category (Figure 4) did reveal potential for an interaction term; though the density of the graph makes it difficult to be certain, it seems possible that the different categories of games have different slopes for the relationship between age and average rating. Therefore, Model 1 was created using the log of maximum playtime, the log of the number of site users owning the game, and the interaction between category and age to predict a game's average rating. The summary output for this model is given in Table 1. In addition to the possible interaction between age and category, Figure 4 showed potential for a quadratic relationship between age and average rating. Therefore, Model 2 was created using the same variables as Model 1 but with a quadratic transformation for game age and no interaction term.

Model Comparison and Assumptions

Initial comparison of the adjusted R^2 values for the models pointed toward Model 2 as the better model; it had an adjusted R^2 value of 0.309 compared to 0.261 for Model 1. The more important factor in deciding between the models, however, was the diagnostic plots for each. Two of the assumptions necessary for OLS were met for both models by the background of the dataset. That is, the data for each game are presumed to be independent from one another, and it is reasonable to assume that the games in the dataset are a random sample from the population of all possible board games. However, the residuals vs fitted values plot for Model 1 (Figure 5) showed a violation of the assumptions of linearity and constant variance of the errors; the residual values in the plot were closely clumped together and unequally distributed across values.

By contrast, diagnostic plots showed that all necessary assumptions were met for Model 2 to be valid using OLS. Though there were still some slight imperfections, the residuals vs fitted values plot (Figure 6.1) showed a fairly random scatter of points, suggesting that the true errors have a constant variance and that this model represents an overall linear pattern. Additionally, despite some deviations, the QQ plot for the model (Figure 6.2) suggests that it is reasonable to assume that the true errors for the model are normally distributed. On this basis, Model 2 was selected as the final model.

Discussion of Final Model

Under Model 2, the final conditional mean function created to describe the variability in the average rating of board games is:

$$E(Y_{rating}|X) = \beta_0 + \beta_1 log(X_{playtime}) + \beta_2 log(X_{owners}) + \beta_3 X_{age} + \beta_4 X^2_{age} + \beta_5 Z_{cards} + \beta_6 Z_{dice} + \beta_7 Z_{economic} + \beta_8 Z_{party}$$

The summary output for this final model is given in Table 2. In terms of model fit, the R^2 value given for the model suggests that 31.19% of the variation in average game ratings can be explained by its linear relationship with the chosen predictors. The given RMSE value of 0.64 cannot be meaningfully interpreted, given the transformations included in the model. There is no evidence of overfitting or multicollinearity in the model; with over 2,000 games in the dataset, the sample size is sufficient for the number of predictors and none of the VIFs (Table 3) are large enough to provide evidence of multicollinearity.

Turning now to individual predictors, the model estimates that for a 10% increase in the maximum amount of time needed to play a game, the average rating of the game is expected to increase by 0.021 points, holding all other predictors constant. On a rating scale out of only 10 possible points, this could be considered a practically significant effect. With an associated p-value of 2e-16, this is also statistically significant, as were all other predictors included in the model. The estimated coefficient for age unfortunately does not have a simple interpretation given the quadratic transformation, but it does seem to have been important in predicting the average rating for a game. Finally, the different categories of games all had similar estimates. Party games, for example, were estimated to have an intercept that is on average. -0.4973 points lower than that of abstract strategy games Although the intercept estimates themselves for each category are practically meaningless, since it is impossible for all of the quantitative predictors to equal zero, it does seem from these estimates that the reference group of abstract strategy games are generally more popular with the highest intercept out of all groups.

Conclusion

This analysis showed that it is possible to predict the average community rating of a board game using the time it takes to play it, the number of people who own it, its age, and the genre it belongs to. Although this model did not seem to fit the data particularly well, it did shed some interesting insights on the factors that make a good board game. Estimates show that games that take longer to play and are already owned by more people tend to be rated more highly, and that it is possible that abstract strategy games are rated more highly than other genres. Although the relationship between a game's popularity and its age is more difficult to describe, age does appear to affect the ratings of games. Future analysis including games belonging to multiple genres could further explain the variation within board game ratings, and may be of interest to anyone trying to design a highly-rated board game.

Appendices

Abigail Rieck

Figure 1.1: Histogram of Maximum Playtime

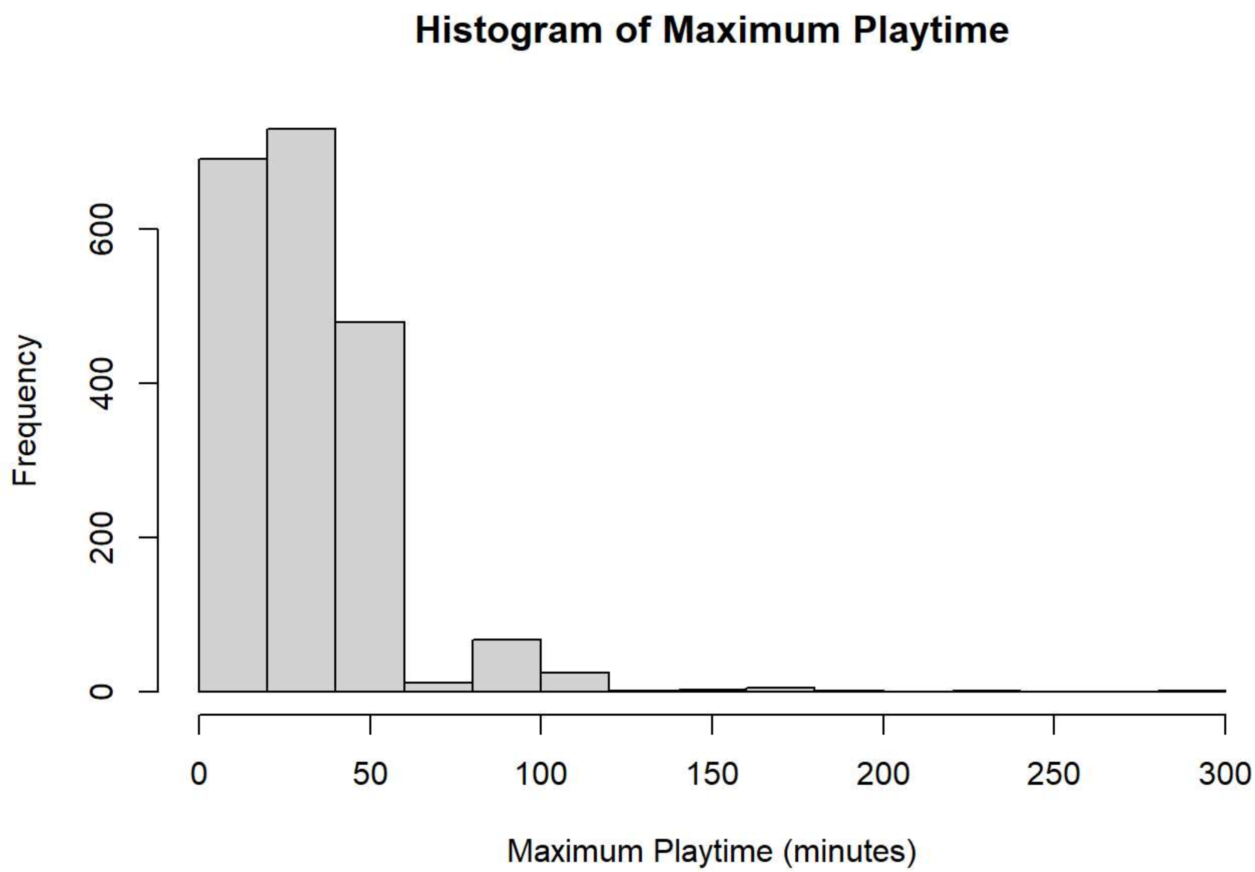

Figure 1.2: Histogram of Maximum Playtime with Log Transformation

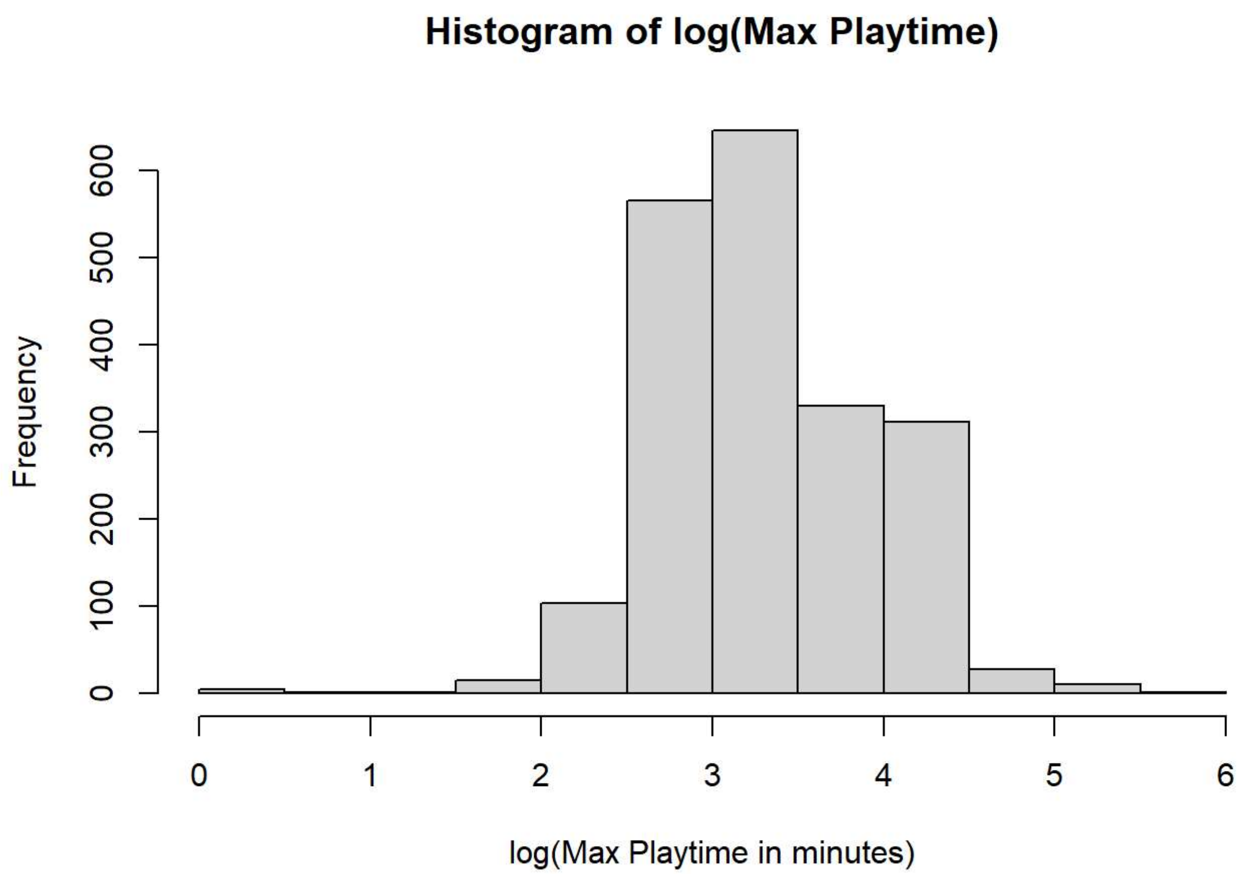

Figure 2: Histogram of Publishing Year

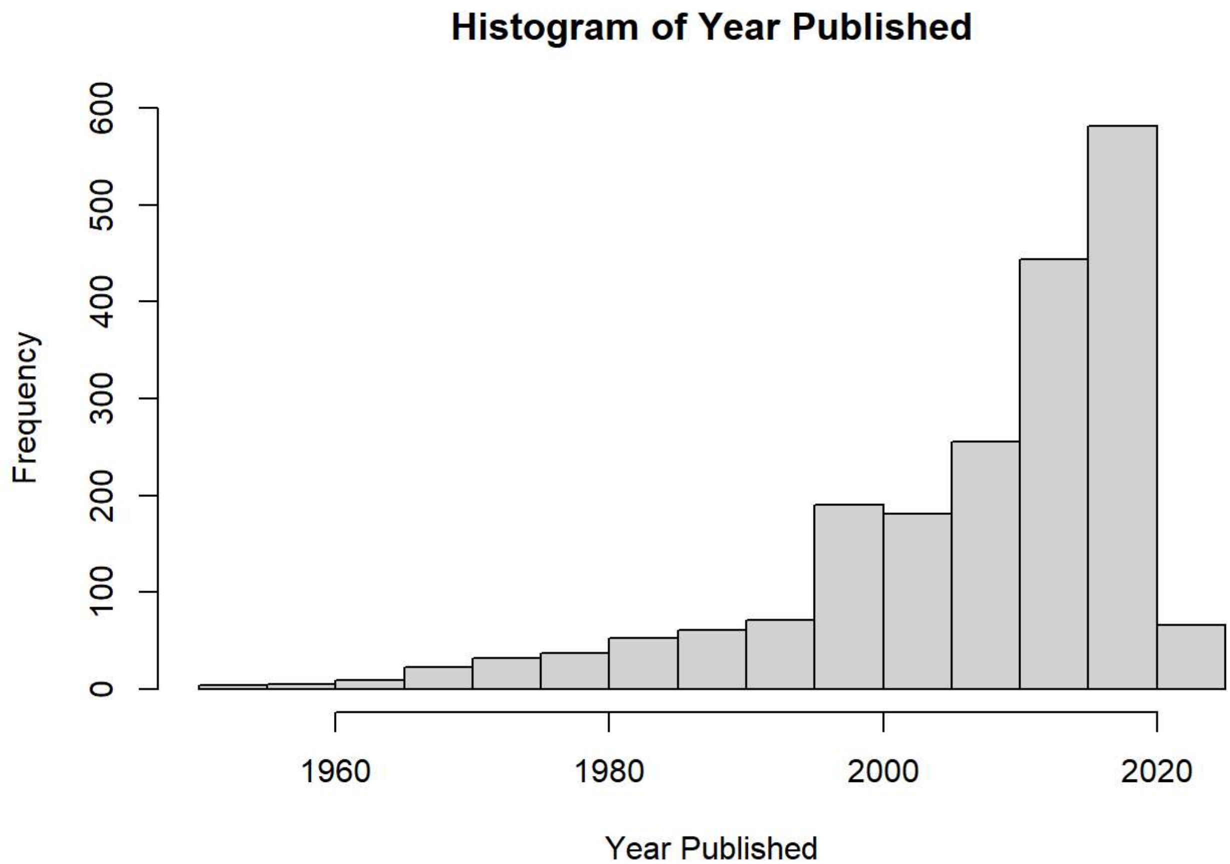

Figure 3: Boxplots of Average Rating by Category

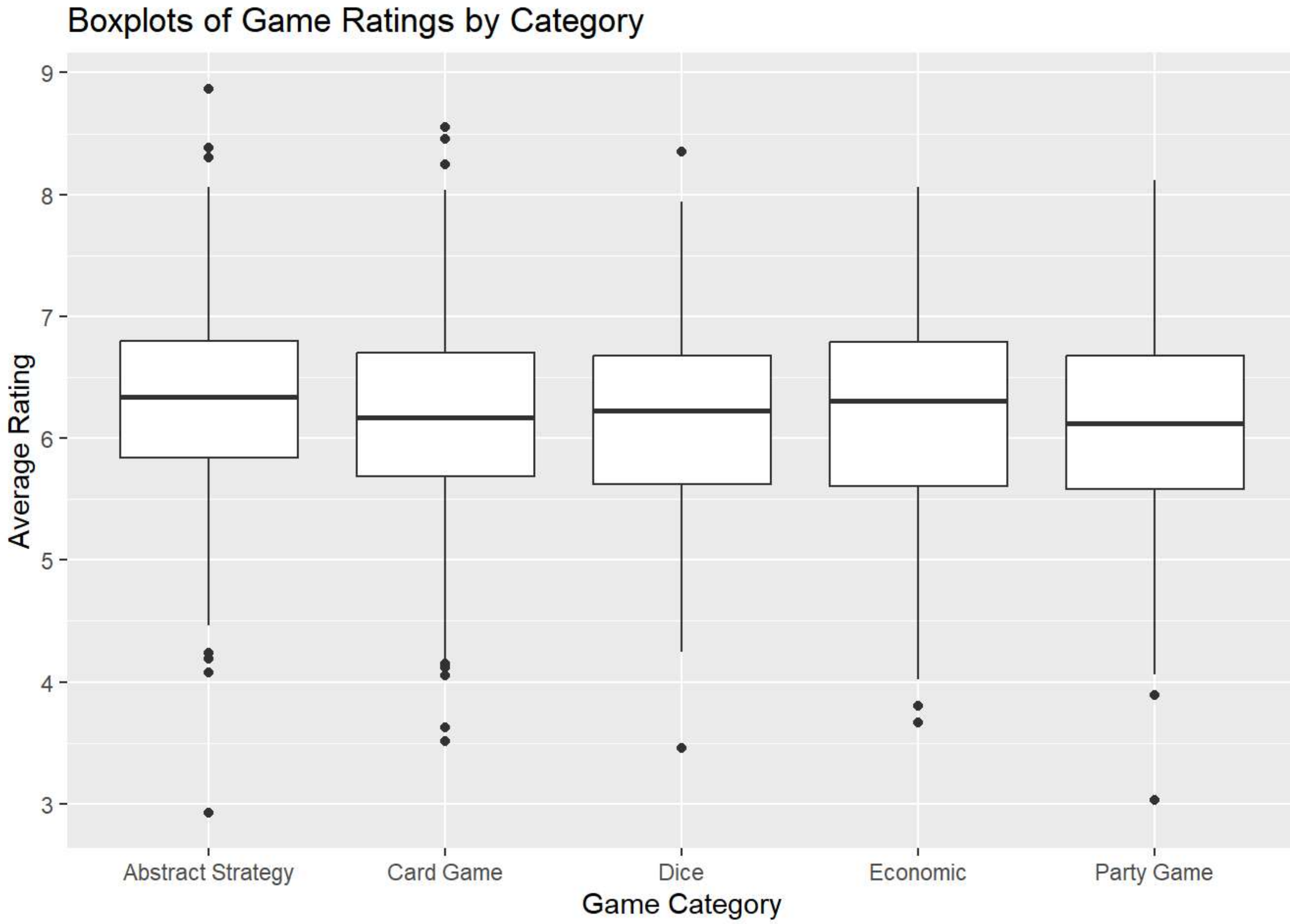

Figure 4: Interaction Between Category & Age

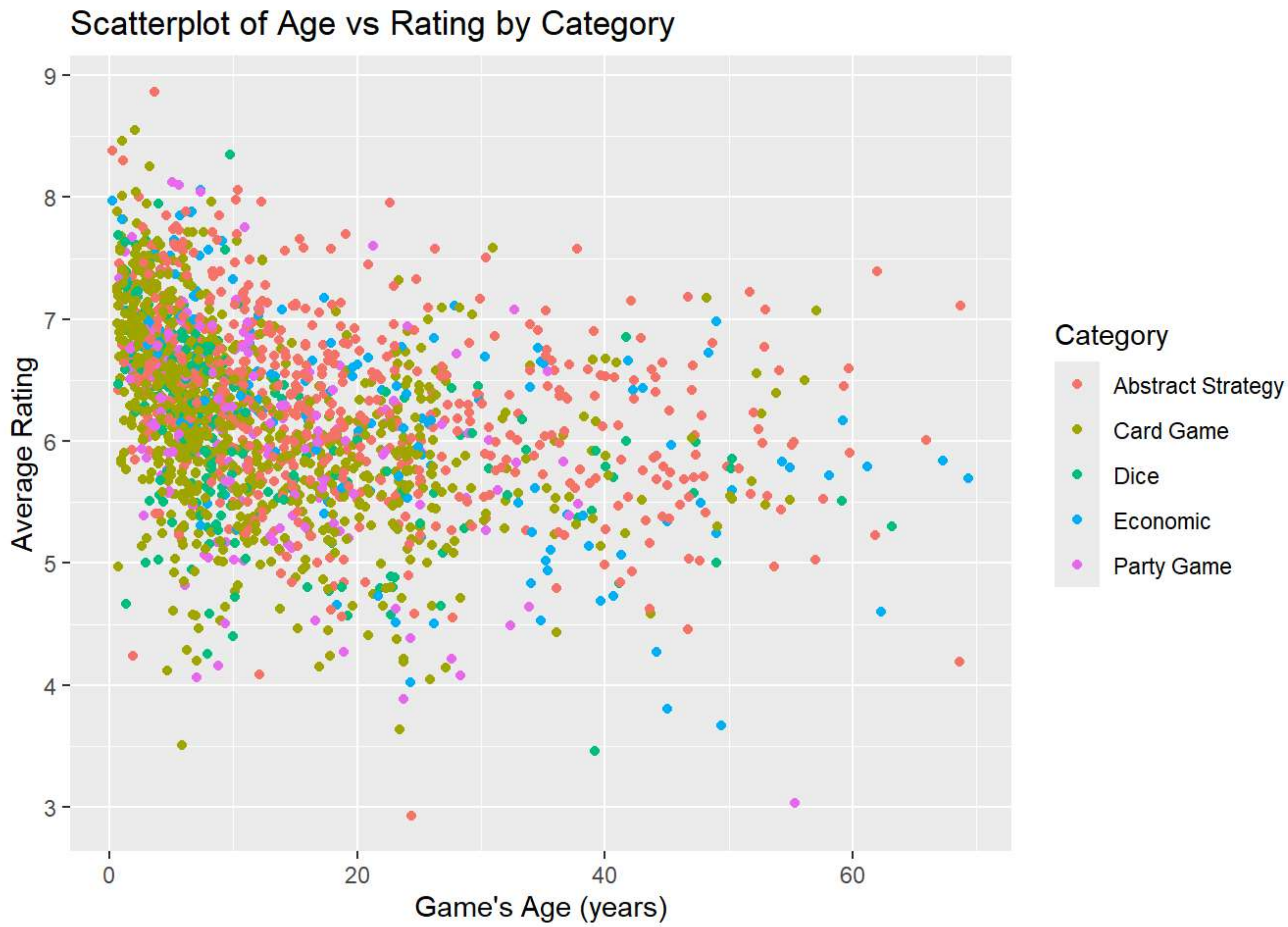

Table 1: Summary of Model 1

```
##
## Call:
## lm(formula = average ~ log(maxplaytime) + log(owned) + age *
##     category, data = games)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5117 -0.3891  0.0193  0.4042  3.0942
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)             5.134216   0.114299  44.919  < 2e-16 ***
## log(maxplaytime)        0.245121   0.027462   8.926  < 2e-16 ***
## log(owned)              0.134753   0.011919  11.306  < 2e-16 ***
## age                    -0.016649   0.001906  -8.737  < 2e-16 ***
## categoryCard Game      -0.203256   0.057334  -3.545 0.000401 ***
## categoryDice           -0.233242   0.078183  -2.983 0.002886 **
## categoryEconomic       -0.183623   0.107230  -1.712 0.086972 .
## categoryParty Game     -0.152913   0.096740  -1.581 0.114115
## age:categoryCard Game  -0.014464   0.002915  -4.963 7.54e-07 ***
## age:categoryDice       -0.012383   0.004211  -2.940 0.003314 **
## age:categoryEconomic   -0.011789   0.003932  -2.998 0.002751 **
## age:categoryParty Game -0.024127   0.005726  -4.213 2.63e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6655 on 2004 degrees of freedom
## Multiple R-squared:  0.2654, Adjusted R-squared:  0.2613
## F-statistic:  65.8 on 11 and 2004 DF,  p-value: < 2.2e-16
```

Figure 5: Residuals vs. Fitted Values Plot For Model 1

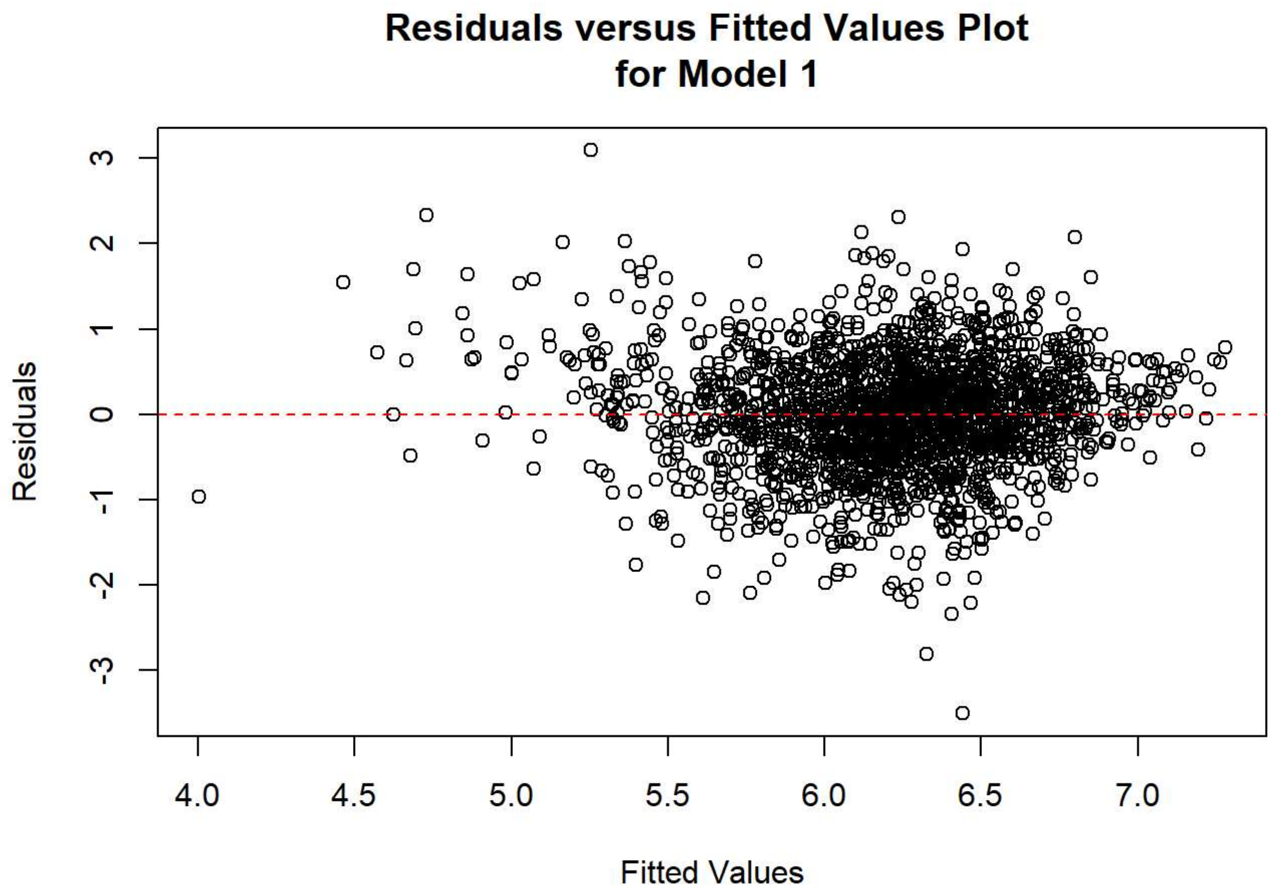

Figure 6.1: Residuals vs Fitted Values Plot for Model 2

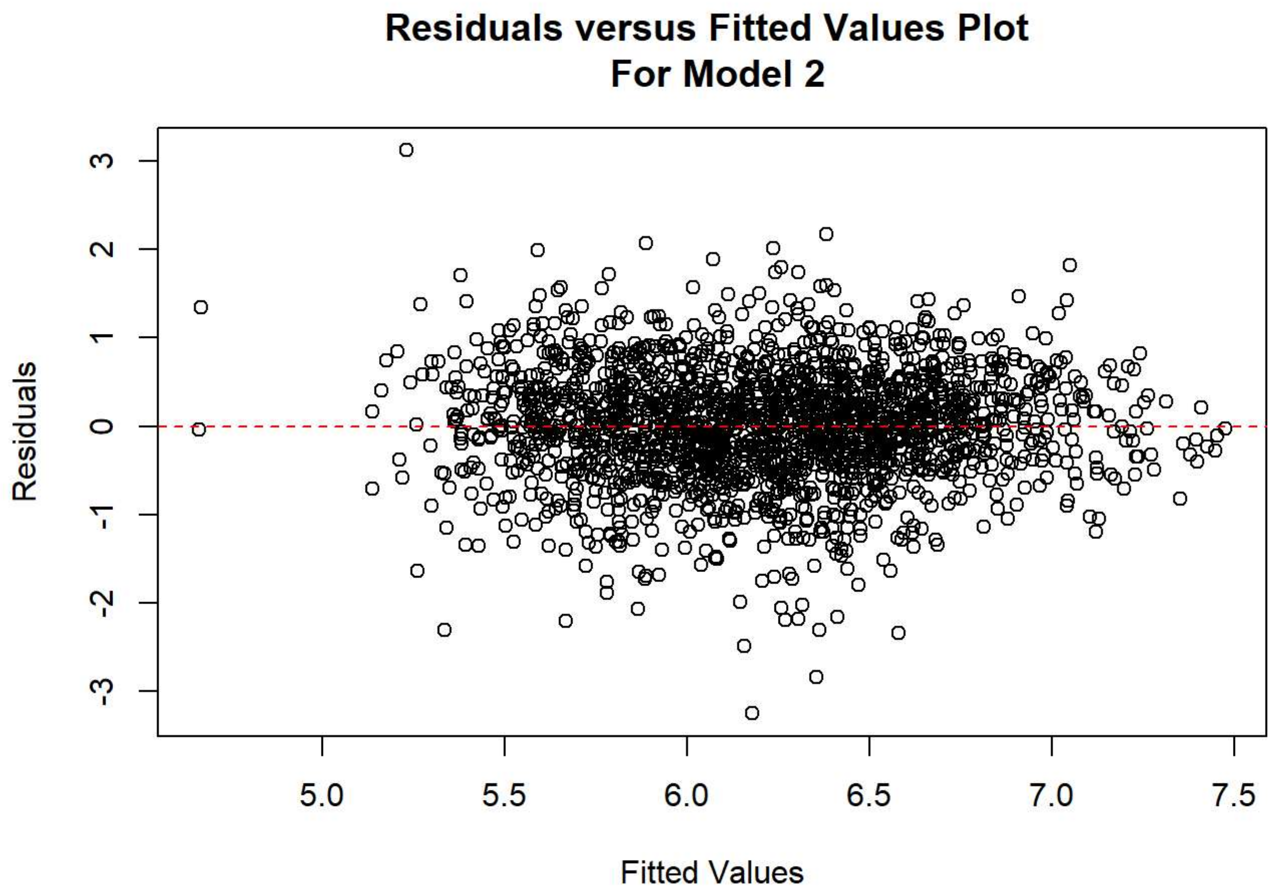

Figure 6.2: QQ Plot for Model 2

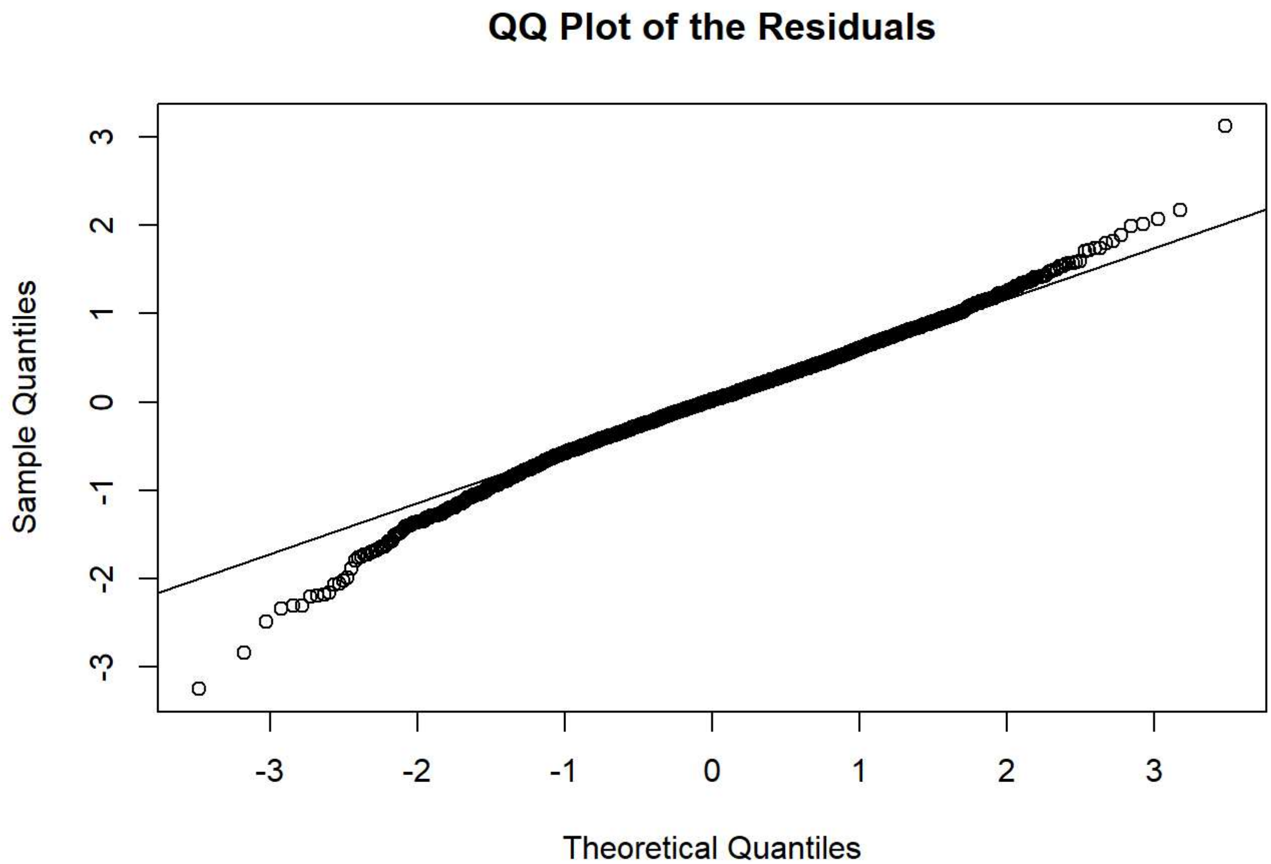

Table 2: Summary of Final Model

```
##
## Call:
## lm(formula = average ~ log(maxplaytime) + log(owned) + age +
##     I(age^2) + category, data = games)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2498 -0.3774  0.0222  0.4009  3.1186
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         5.663e+00  1.080e-01  52.437  < 2e-16 ***
## log(maxplaytime)    2.258e-01  2.645e-02   8.537  < 2e-16 ***
## log(owned)          1.353e-01  1.149e-02  11.772  < 2e-16 ***
## age                -6.754e-02  3.409e-03 -19.814  < 2e-16 ***
## I(age^2)            8.797e-04  6.633e-05  13.261  < 2e-16 ***
## categoryCard Game  -4.545e-01  3.603e-02 -12.616  < 2e-16 ***
## categoryDice       -4.920e-01  5.162e-02  -9.532  < 2e-16 ***
## categoryEconomic   -4.282e-01  6.589e-02  -6.499 1.02e-10 ***
## categoryParty Game -4.973e-01  5.708e-02  -8.712  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6436 on 2007 degrees of freedom
## Multiple R-squared:  0.3119, Adjusted R-squared:  0.3091
## F-statistic: 113.7 on 8 and 2007 DF,  p-value: < 2.2e-16
```

Table 3: VIFs for Final Model

```
##                      GVIF Df GVIF^(1/(2*Df))
## log(maxplaytime) 1.246878  1        1.116637
## log(owned)       1.065205  1        1.032088
## age              9.224084  1        3.037118
## I(age^2)         9.066037  1        3.010986
## category         1.411375  4        1.044012
```