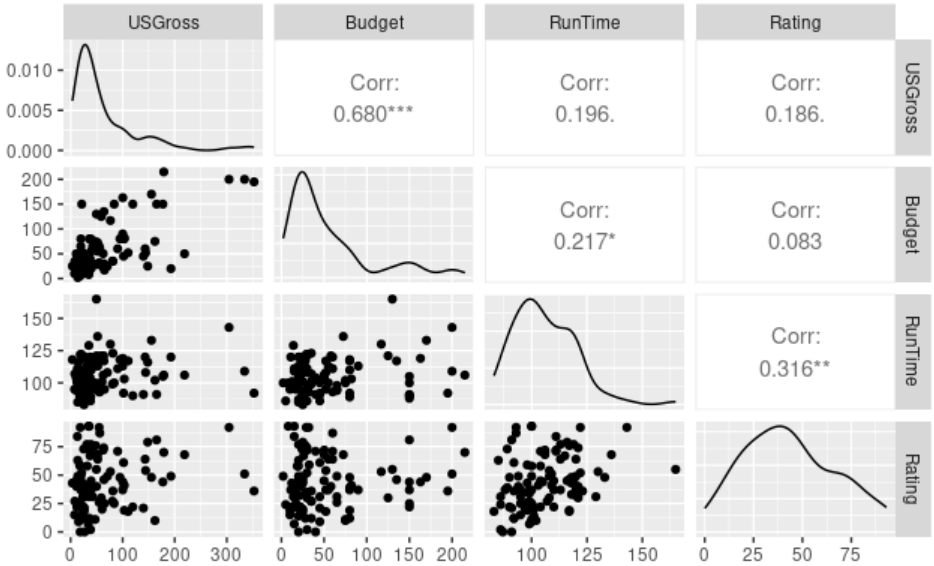
Purpose

The purpose of this analysis is to expand the previous model used to predict gross earnings based on film budget. In this model, Questionably Normal Productions is interested in using more explanatory variables including budget, length of the movie, and rating, to better estimate their potential returns.

Data

Data on 100 movies were provided by Questionably Normal Productions. This analysis focuses on film budget (median = 33.75 million USD, IQR = 45.88 million USD), length of the movie (median = 104.0 minutes, IQR = 21.2 minutes), rating (median = 41.0 percent, IQR = 36.5 percent), and gross earnings (median = 37.98 million USD, IQR = 56.63 million USD). Summary statistics for these variables are provided in Appendix A. Figure 1 shows the matrix plots of film budget, runtime, and rating vs gross earnings for this sample of movies, suggesting a moderate positive linear relationship between the film budget and gross earning (r = 0.68). However, the relationships between runtime and rating vs gross earnings show little relationship between them with correlation of 0.196 for run time and 0.186 for rating.

Figure 1: Matrix plots of film budget, runtime and rating vs gross earnings



Method

Multiple linear regression was used to examine if movie budget, run time, and rating predict gross earnings. USGross was the response variable, and Budget, RunTime, and Rating were the explanatory variables. To address issues with equal variance, normality, and linearity, the natural log transformation was applied to both gross earnings and budget. The population regression model was:

$$\log(\text{USGross}) = \beta_0 + \beta_1 * \log(\text{Budget}) + \beta_2 * \text{RunTime} + \beta_3 * \text{Rating} + \varepsilon$$

Results

The estimated regression line was:

log(USGross)-hat = $\beta_0 + \beta_1 * 0.59 + \beta_2 * 0.005 + \beta_3 * 0.005$

A hypothesis test of the slope shows that the multilinear regression model is useful in predicting gross earnings (p-value < 0.001, df = 3 and 96, F = 20.76). THe estimated slope indicates that for each 1% increase in movie budget, the median gross earnings are estimated to increase by 0.614% (95% CI: 0.451% to 0.779%). For each minute increase in runtime, the gross earning increases by 0.59% (95% CI:-0.0062% to 0.016%), while simultaneously holding run time and rating constant. And for each percent increase in rating, the gross earning increases by 0.005%, ,while simultaneously holding a log of film budget and run time constant.

Appendix A: Exploratory Data Analysis

Table 1A: Summary Statistics of USGross, Budget, RunTime, and Rating

Variable	Min	Q1	Median	Q3	Max
Gross Earnings	3.352	23.868	37.982	80.498	352.390
Budget	2.0	21.0	33.75	66.88	215.0
RunTime	83.0	95.0	104.0	116.2	165.0
Rating	0.0	25.0	41.0	61.5	93.0