Movie Ratings vs Domestic Box Office Revenue

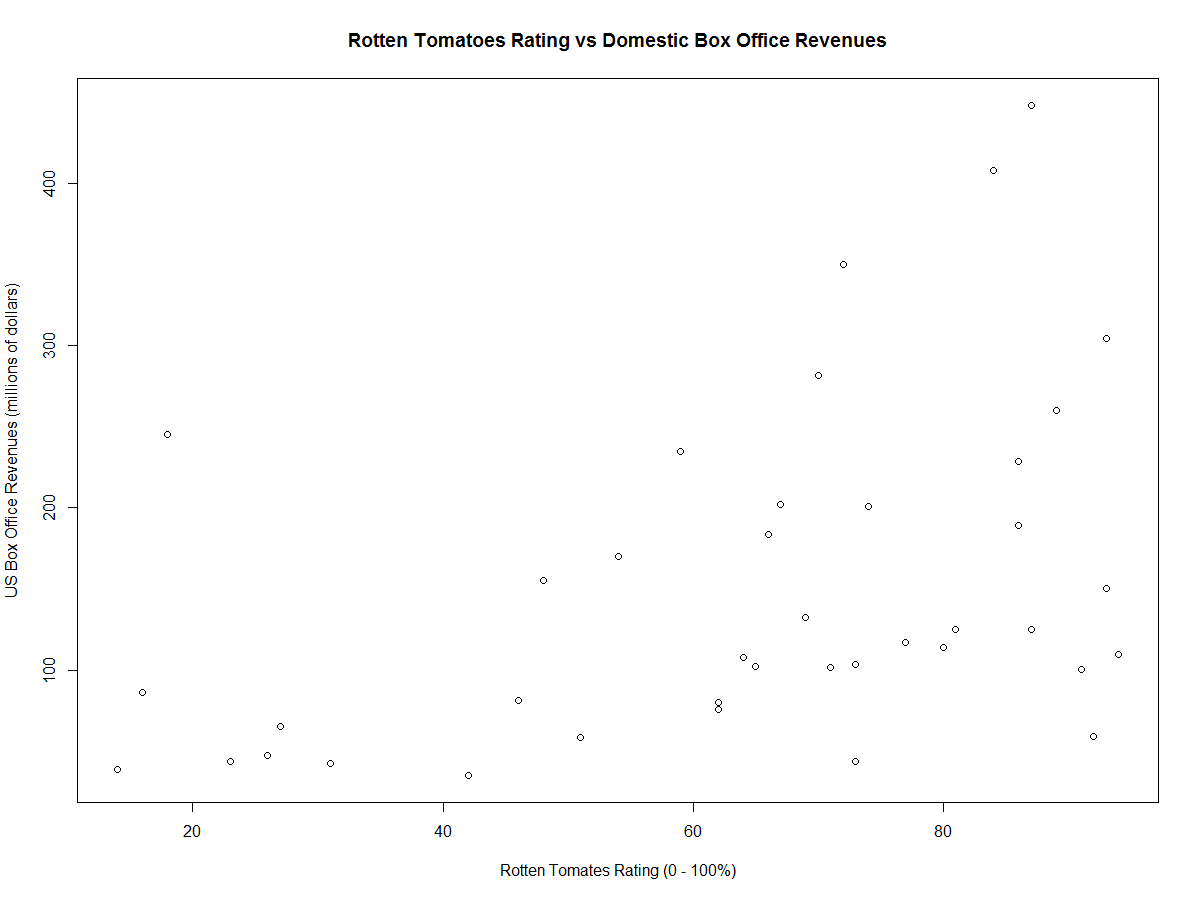
Going to the movies has been a leading American past time for decades. Hollywood generates billions of dollars annually from box office revenues alone and they reinvest a great deal of these billions into generating more cash flows. Movie producers analyze everything from who the director should be, which actors should be casted and what weekend the movie should be released, in order to maximize box office revenues. Studios routinely invest millions of dollars in finding the optimal combination. One complaint frequently heard on internet forums or group discussions is that movies don't necessarily need to be “good” to generate huge profits. Movies could have a mediocre plot and directing, but with a star cast and a hefty advertising budget, they could generate hundreds of millions in the box office (Movies like the recent Batman v. Superman might come to mind). Therefore, the focus of this linear regression will be to focus on the relationship between movie ratings and box office profits.

The data analyzed for this project are Box Office Revenues from 40 movies from 2012-2015 and the ratings these movies got on Rotten Tomatoes. For the movie selection, the top 50 movies in terms of domestic (US) box office revenues was examined for the years 2012, 2013, 2014 and 2015. From these lists, 10 movies were randomly selected from each (Using the sample function in R). Domestic Box Office Revenues were chosen instead of International since the Rotten Tomatoes ratings used were by critics based in the U.S. After these movies were chosen, a python script was written to scrape the website Rotten Tomatoes for the ratings of all these movies (BeautifulSoup was the package used to write the python web scrapper. At first, I tried to use the Rotten Tomatoes API, but got no response to my API request; so, alternative methods had to be used). The rating used was the Rotten Tomatoes Tomatometer Rating, which represents the percentage of professional critic reviews that are positive for a given film. General audience ratings are not factored in as they may be based more on whim.

Since this is an observational study, we cannot directly establish a causation relationship between movie ratings and box office revenues, but it would be reasonable to assume some type of relationship. Rotten Tomatoes’ Tomatometer Rating comes out before a film’s release and movie goers may use the rating to determine which movie they watch. Using this association, the linear regression equation can be formed:

A large β0 may indicate that a great deal of Domestic Box Office Revenue comes from other factors while a large β1 and small β0 may suggest that there is a large association between the Tomatometer Rating (or a movie’s rating in general: since the Tomatometer Rating is highly correlated with other movie rating metrics) and Box Office success.

The data set of the 40 movies was then graphed using R’s plot function. The Box Office revenues are listed in millions of dollars and the Tomatometer Rating is on a 0 – 100 scale.



As shown in the plot, there seems to be a weak, positive relationship between the Rotten Tomatoes Rating and the US Box Office Revenues. There also seem to be a few outliers, which can be expected of high performing Hollywood blockbusters.

