**Motivation:**

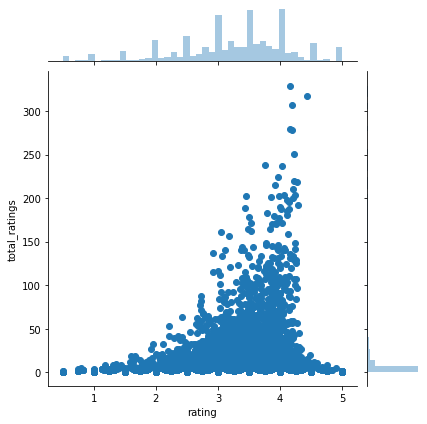
**Problem Definition:**

**Solution Explanation:**

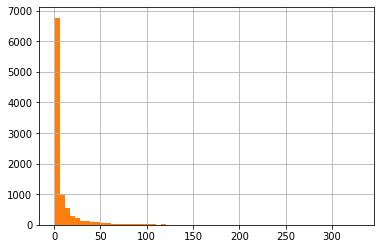
The plots show us a relationship between the number of ratings and the ratings for the movies. It shows that there is a positive relationship between the number of ratings and the average rating of a movie. Also, the graph indicates that the more ratings a movie has, the higher its average rating becomes. This causes a problem, because for the movies with few ratings but high ratings, it could allow them to be recommended even though they are not similar. For that reason, we must create a threshold when comparing the movies. By viewing the histogram of Fig. 1 we can see that a majority of the ratings per movie are under 100. Thus, only search for movies that have at least 100 reviews and then sort them ascendingly by correlation.

**Data Description:**

In this assignment we used the Movie Lens Dataset. The dataset was built by a research group at the University of Minnesota. It contains 100,000 ratings applied to 9,000 movies by 600 users. The data set will change over time and is not appropriate for reporting research results.

Fig.1

From looking at Fig. 1 we can see that most of the films are rated between 2.5 to 4.

Fig. 2

From looking at Fig. 2 we can see that the that most of the movies have a small amount of ratings. The movies with the most ratings are the most famous films.

**Result Analysis:**

We used a dataset to find the number of ratings and the average rating for each movie. We then used the ratings to find a correlation between the movies. We used a correlation function to indicate how the two variables fluctuate together. The movies with a high correlation are the movies that have the most similarity with one another. The correlation numbers lie between –1 and 1. -1 Indicates a strong negative correlation, zero indicates no correlation (not similar at all) and 1 indicates a positive correlation. The most similar movies to Forrest Gump was Good Will Hunting; with the highest correlation of 0.48, Aladdin and American History X. The most similar movies to Pulp Fiction was Fight Club, with the highest correlation of 0.54, Kill Bill 1 and Trainspotting.

