1 Introduction [10 points]

Group members: David Zheng and Amanda Li

Team name: cheese

Division of labor:

- David: Off-the-shelf Software (Surprise SVD, SVD++, and NMF) and GridSearch implementation
- Amanda: Data cleaning, basic visualizations, unbiased SVD (from Homework 5) and SVD with global bias/offset terms

2 Basic Visualizations [20 points]

Note that in our initial data cleaning, we dropped entries for movies with the same title as a previous entry but a different ID, so that each movie had one unique ID only. We then reindexed all of the movies so that the IDs were continuous (no missing IDs) and updated the data/train/test.txt files with the new IDs. All of our visualizations are based on the data that was cleaned as described above, and code for the data cleaning is included at the end of this section.

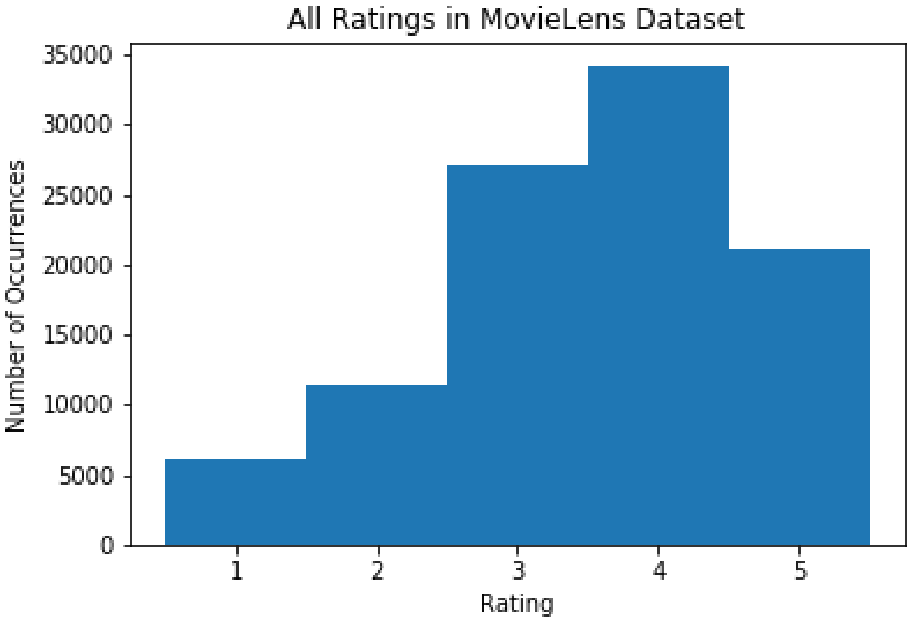


Figure 1: All Ratings in MovieLens Dataset

We observe that the histogram is slightly skewed to the left, with the majority of ratings being 3 or above. This is not unexpected, as many users only bother to rate movies that they enjoyed, and users generally tend to watch movies that they believe they will enjoy.

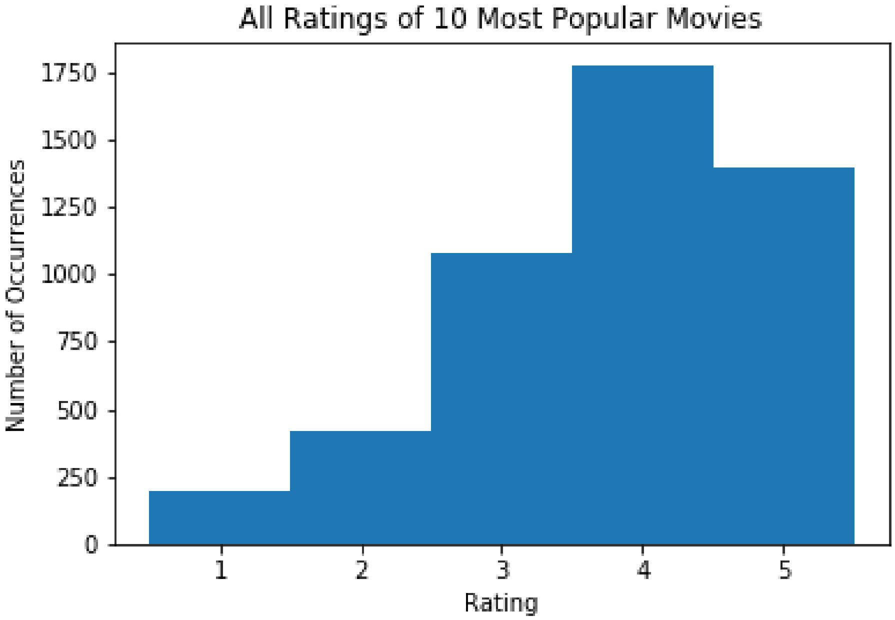


Figure 2: All Ratings of 10 Most Popular Movies

We found that the 10 most popular movies were ['Star Wars', 'Contact', 'Fargo', 'Return of the Jedi', 'Liar Liar', 'The English Patient', 'Scream', 'Toy Story', 'Air Force One', 'Independence Day']. We observe that the histogram is skewed to the left, with the vast majority of ratings being 3 or above. This indicates that the movies with the most ratings also tend to be highly rated on average, which is expected given that users are more likely to recommend movies that they enjoyed/rated highly to other users. Moreover, if a movie has high ratings, then users are more likely to watch it than a lower-rated movie, leading to a higher popularity. This distribution is also very similar in shape to the distribution of ratings for all movies; however, there is a slightly higher proportion of 5 ratings in this histogram than in the histogram with all movies. This is expected given that these are the 10 most popular movies; we would expect their ratings to be generally higher than ratings for the average movie because people keep watching them, which is an indication of their quality.

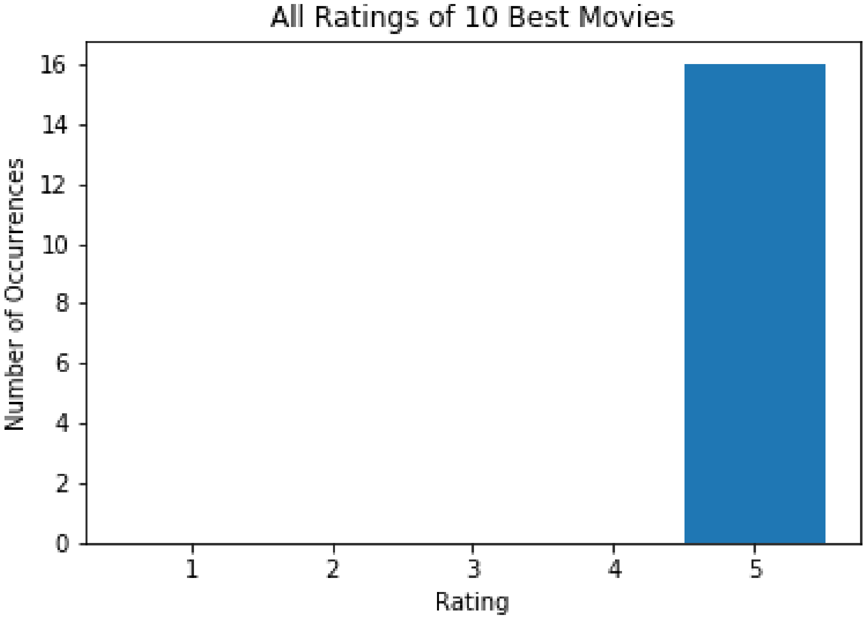


Figure 3: All Ratings of 10 Best Movies

We found the 10 best movies to be [”Someone Else’s America”, ’Prefontaine’, ’Aiqing wansui’, ’Star Kid’, ’Entertaining Angels: The Dorothy Day Story’, ’They Made Me a Criminal’, ’Marlene Dietrich: Shadow and Light’, ’A Great Day in Harlem’, ’The Saint of Fort Washington’, ’Santa With Muscles’]. We observe that every single rating of the 10 movies with the highest average ratings is 5. Moreover, the overall number of ratings for these movies is very low, indicating that these movies are generally not popular. This is expected, since there are some movies with very few, but very high, ratings that can achieve an average rating of 5. Indeed, we observed that out of these top 10 best-rated movies, none have more than 3 total ratings. Most movies with more ratings (more popular movies) tend to have a wider range of ratings, and so while they are generally fairly highly rated, they are unable to achieve an average rating of 5.

Thus, this explains the difference between the ratings of the most popular and the best movies. The best movies have all ratings of 5, but there are very few ratings total, while the most popular movies have a wider range of ratings but in total have many hundreds of times more ratings of 5 than the best movies do.

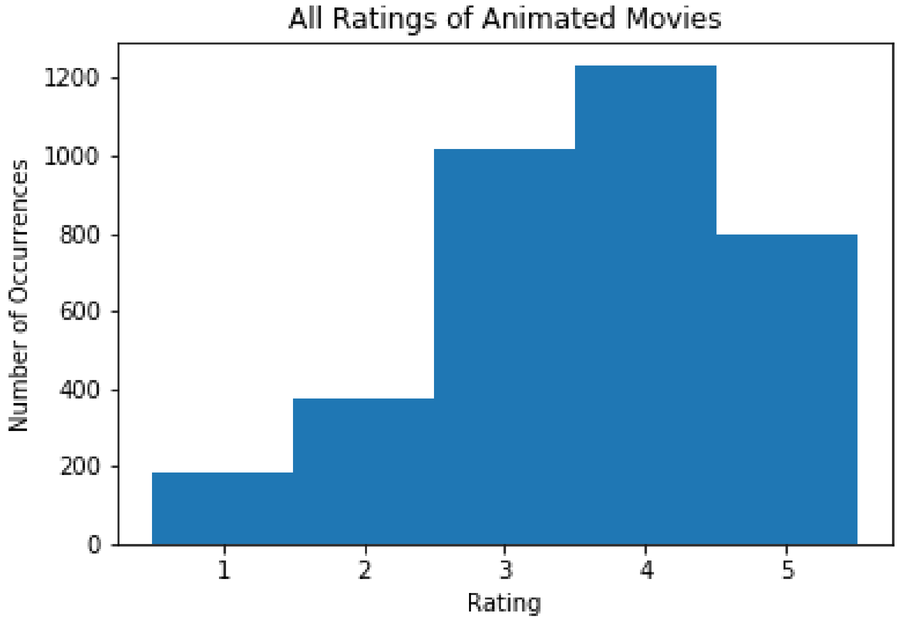


Figure 4: All Ratings of Animated Movies

We observe that the histogram is skewed to the left, with the vast majority of ratings being 3 or above. This indicates that animated movies are generally well-received, which makes sense given that animated movies are typically targeted towards kids and are meant to be lighthearted and unoffensive. In addition, the shape of the histogram is very similar to the shape of the histogram with all movie ratings. This shows that users, most likely children, who rate animated movies tend to give ratings in the same distribution as users do as a whole. This may be an indication that age does not play much of a factor in the rating of a movie, though we would need addition data on the users to make more concrete conclusions.

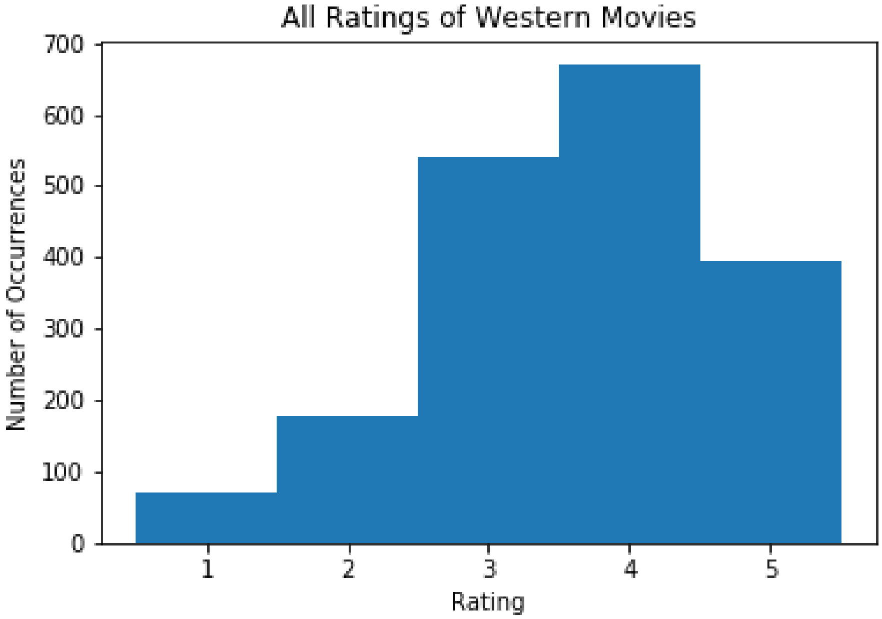


Figure 5: All Ratings of Western Movies

We observe that the histogram is skewed to the left, with the majority of ratings being 3 or above. This indicates that Western movies are generally well-received, which might be expected given that users who watch Westerns tend to be those who like Westerns in the first place, while users who do not enjoy Westerns would generally avoid them. We see the same general trend in the histogram as with animated movies, suggesting that seemingly very different genres can have similar distributions of ratings.

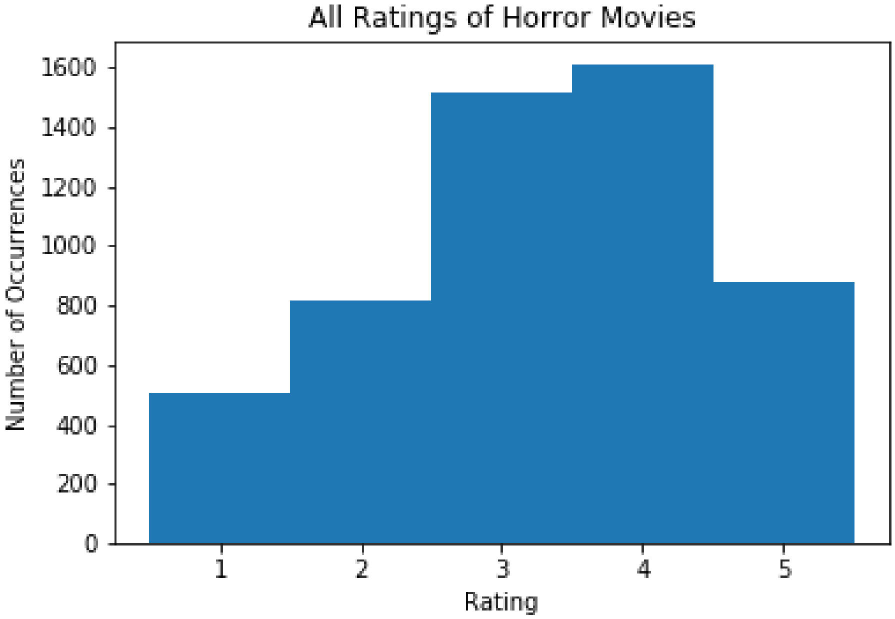


Figure 6: All Ratings of Horror Movies

We observe that the majority of ratings are 3 or 4, with about 1.5 times more ratings of 1 and 2 than ratings of 5. This histogram indicates that opinions regarding horror movies are fairly mixed, with many users leaving very good ratings and many users leaving very poor ratings. This is expected, since many times horror movies are ”hit or miss,” or elicit very strong reactions.

Comparing between the three genres, the histograms for animated movies and Westerns look fairly similar in shape (skewed to the left, with 4 being the most popular rating), while the histogram for horror movies has many more low ratings in comparison. As we discussed above, this is unsurprising given that horror movies may cause strong negative reactions, while animated movies and Westerns are less polarizing and generally liked by more people.

3 Matrix Factorization Methods [40 points]

Below are the different types of SVD that we implemented:

1. Normal (unbiased) SVD from Homework 5: this method seeks to minimize the regularized square error given by

$$\arg\min_{U,V} \frac{\lambda}{2}\left(||U||_F^2 + ||V||_F^2\right) + \frac{1}{2}\sum_{i,j}\left(y_{ij} - u_i^T v_j\right)^2,$$

where $Y \in \mathbb{R}^{m\times n}$ is a matrix such that y_{ij} represents user i's rating of movie j; and $U \in \mathbb{R}^{k\times m}$ and $V \in \mathbb{R}^{k\times n}$ are latent factors corresponding to users and movies, respectively, such that $Y \approx U^T V$.

We implemented this method using SGD, where we used the usual update rule

$$u_i = u_i - \eta\partial_{u_i}$$

$$v_j = v_j - \eta\partial_{v_j}$$

and the gradients are given by

$$\partial_{u_i} = \lambda u_i - \sum_j v_j(y_{ij} - u_i^T v_j)$$

$$\partial_{v_j} = \lambda v_j - \sum_i u_i(y_{ij} - u_i^T v_j)$$

To tune our model, we tested various values of the regularization constant λ and found the value that minimized test error to be $\lambda = 0.1$:

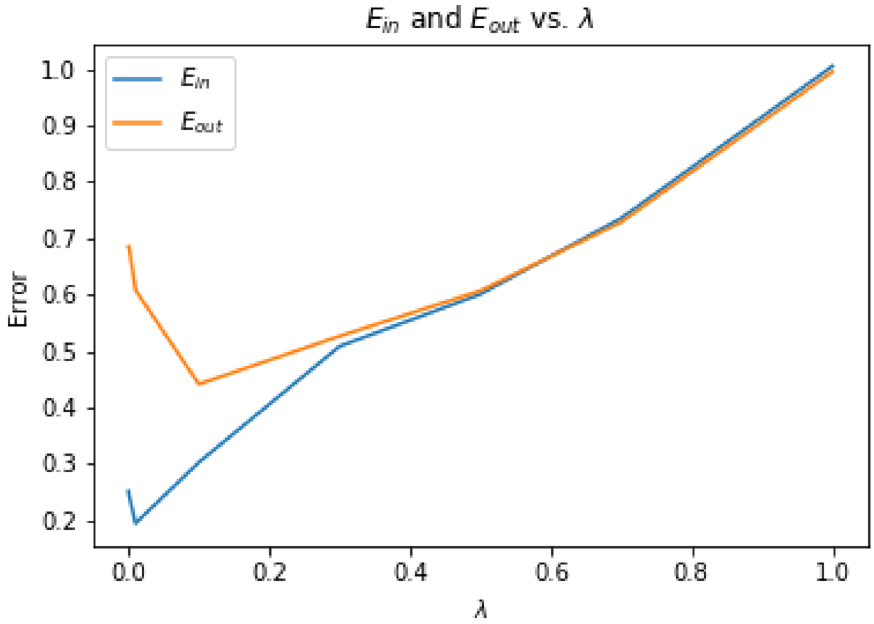
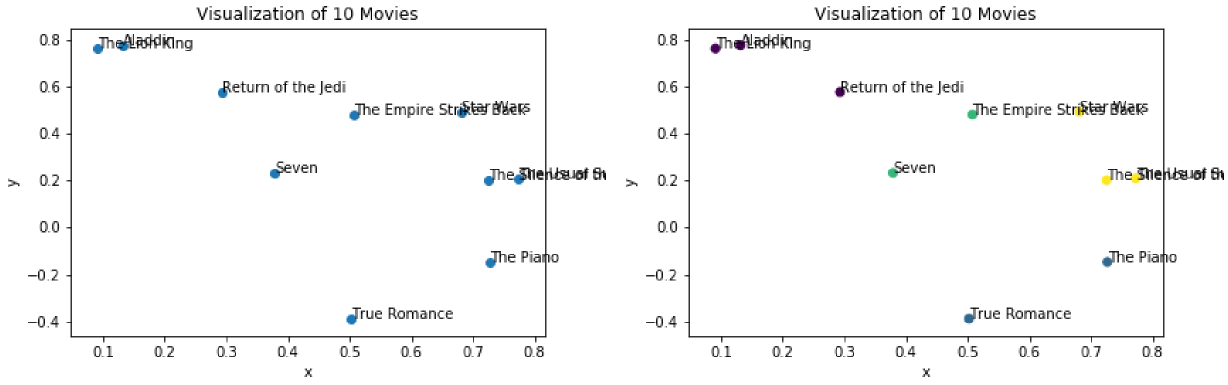


Figure 7: Error vs λ for Unbiased SVD

Additionally, we preserved the stopping condition and step size used in Homework 5. The test error we achieved using these hyperparameters was approximately 0.44.

Below, we include the 6 visualizations created using the results of this method. In addition to just plotting the movies, we also used the k-means algorithm to cluster them:

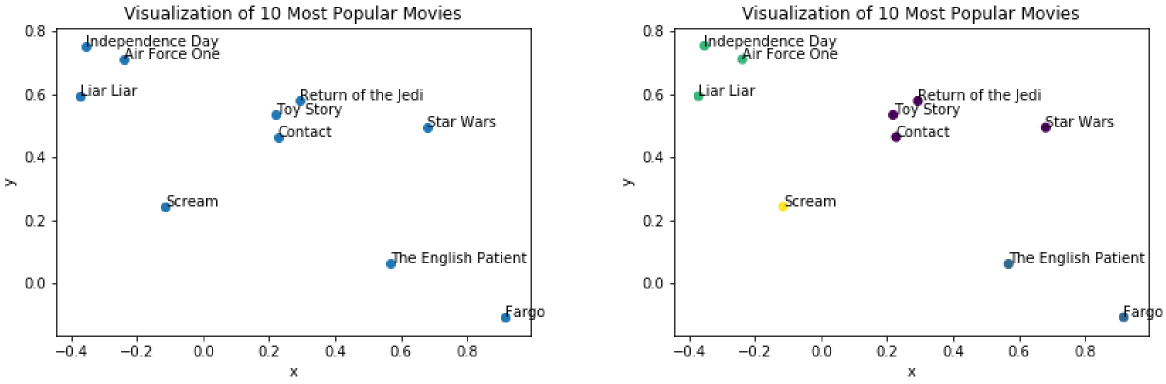


(a) unclustered

(b) clustered via k-means

Figure 8: Visualization of 10 Movies, Unbiased SVD

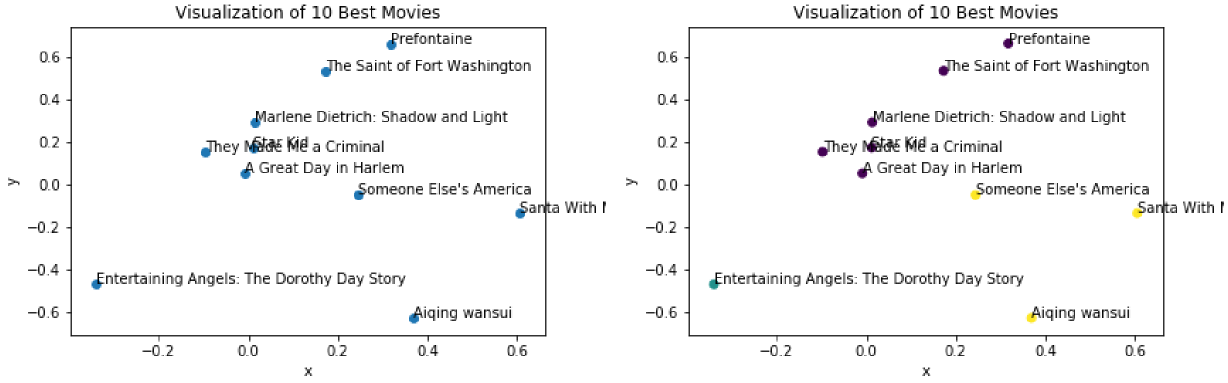
The 10 movies we chose for this first plot were ['Seven', 'The Usual Suspects', 'The Lion King', 'Aladdin', 'The Silence of the Lambs', 'True Romance', 'The Piano', 'Return of the Jedi', 'The Empire Strikes Back', 'Star Wars'].



(a) unclustered

(b) clustered via k-means

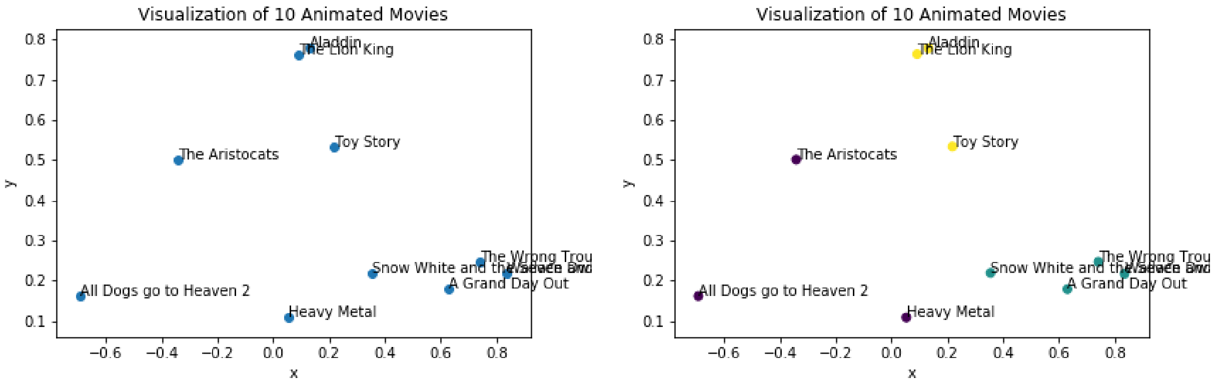
Figure 9: Visualization of 10 Most Popular Movies, Unbiased SVD



(a) unclustered

(b) clustered via k-means

Figure 10: Visualization of 10 Best Movies, Unbiased SVD

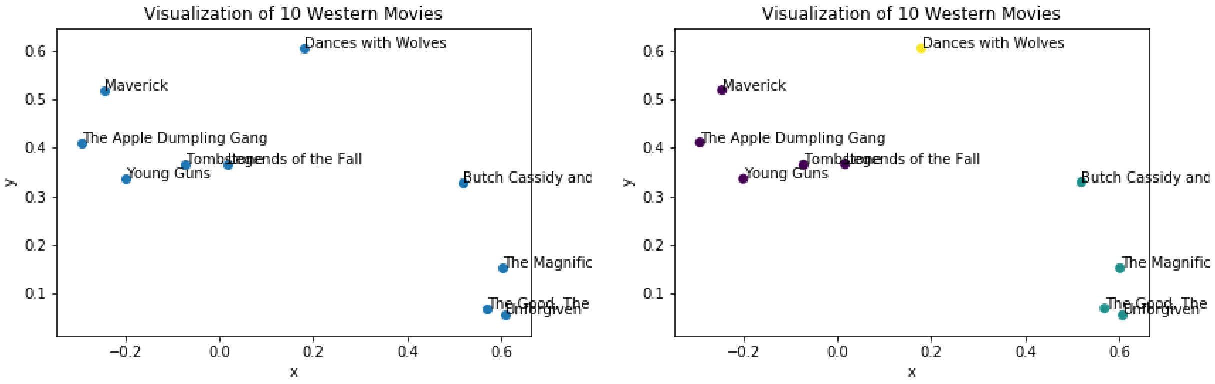


(a) unclustered

(b) clustered via k-means

Figure 11: Visualization of 10 Animated Movies, Unbiased SVD

As can be seen in the plot, the 10 animated movies we visualized were ['Toy Story', 'The Lion King', 'Aladdin', 'Snow White and the Seven Dwarfs', 'Heavy Metal', 'The Aristocats', 'All Dogs go to Heaven 2', 'Wallace and Gromit', 'The Wrong Trousers', 'A Grand Day Out'].

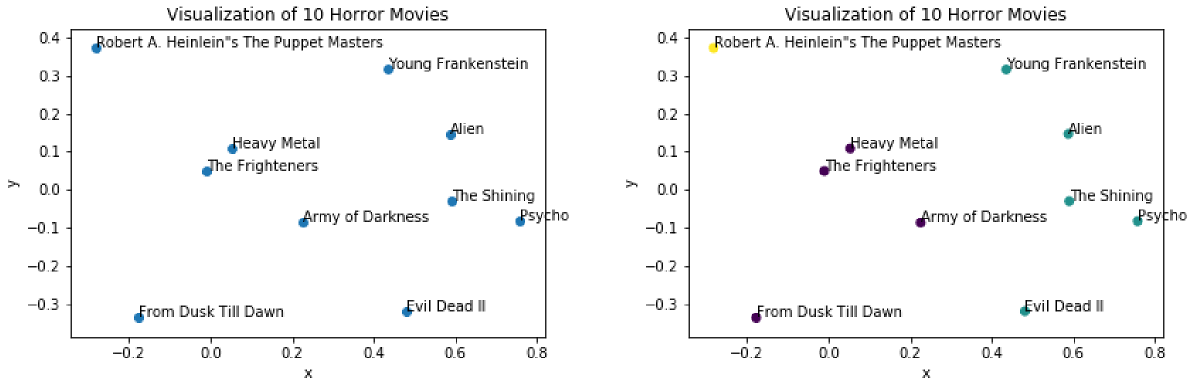


(a) unclustered

(b) clustered via k-means

Figure 12: Visualization of 10 Western Movies, Unbiased SVD

The 10 westerns that we visualized were ['Legends of the Fall', 'Maverick', 'Dances with Wolves', 'The Good, The Bad, and The Ugly', 'Unforgiven', 'Young Guns', 'The Apple Dumpling Gang', 'Butch Cassidy and the Sundance Kid', 'Tombstone', 'The Magnificent Seven'].



(a) unclustered

(b) clustered via k-means

Figure 13: Visualization of 10 Horror Movies, Unbiased SVD

The 10 horror movies that we chose were ['From Dusk Till Dawn', 'Robert A. Heinlein''s The Puppet Masters', 'Heavy Metal', 'The Frighteners', 'Alien', 'Army of Darkness', 'Psycho', 'The Shining', 'Evil Dead II', 'Young Frankenstein'].

2. SVD with global bias term from Slide 8 of Miniproject guide: this method seeks to minimize the regularized square error given by

$$\arg\min_{U,V,a,b} \frac{\lambda}{2}\left(||U||_F^2 + ||V||_F^2 + ||a||^2 + ||b||^2\right) + \frac{1}{2}\sum_{i,j}\left((y_{ij}-\mu) - (u_i^T v_j + a_i + b_j)\right)^2,$$

where a is a bias vector representing user-specific deviation from global bias, b is a bias vector representing movie-specific deviation from global bias, and the global bias μ is modeled as the average of all observations in Y. We can then approximate $y_{ij} \approx u_i^T v_j + a_i + b_j$.

We implemented this method using SGD, where we used the usual update rule

$$u_i = u_i - \eta\partial_{u_i}$$

$$v_j = v_j - \eta\partial_{v_j}$$

$$a_i = a_i - \eta\partial_{a_i}$$

$$b_j = b_j - \eta\partial_{b_j}$$

and the gradients are given by

$$\partial_{u_i} = \lambda u_i - \sum_j v_j((y_{ij}-\mu) - (u_i^T v_j + a_i + b_j))$$

$$\partial_{v_j} = \lambda v_j - \sum_i u_i((y_{ij}-\mu) - (u_i^T v_j + a_i + b_j))$$

$$\partial_{a_i} = \lambda a_i - \sum_j ((y_{ij} - \mu) - (u_i^T v_j + a_i + b_j))$$

$$\partial_{b_j} = \lambda b_j - \sum_i ((y_{ij} - \mu) - (u_i^T v_j + a_i + b_j))$$

Like with the first method, we determined the best regularization constant to be $\lambda = 0.1$:

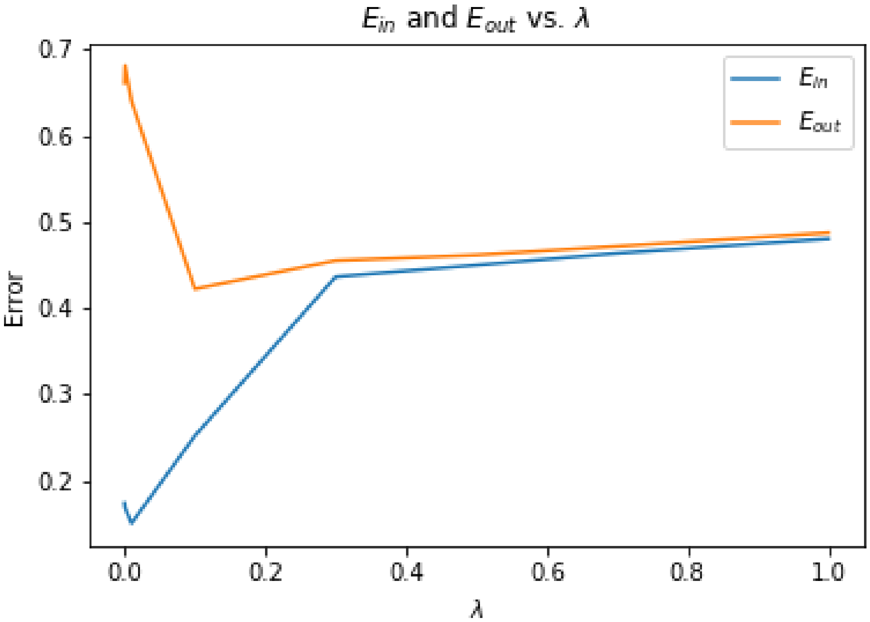
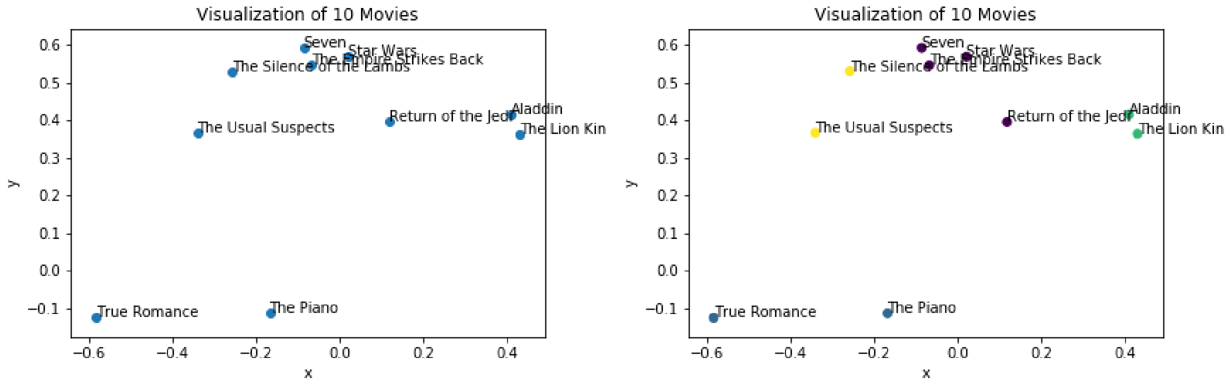


Figure 14: Error vs λ for Biased SVD

With the same stopping condition and step size as above, we achieved a test error of approximately 0.42, lower than the test error of the first method. This suggests that the bias-incorporated method is better at generalizing to new data than the method without bias, which may be explained by the fact that including the bias terms is meant to keep U and V more focused on variability between users and between movies, respectively. For example, if a particular user tends to give low ratings on all movies that they rate, the entry in the a bias vector corresponding to that user should be updated to reflect that user's offset from the average user. Thus, these results indicate that accounting for the user- and movie-specific deviations from global bias has non-trivial impacts on the performance of the final model.

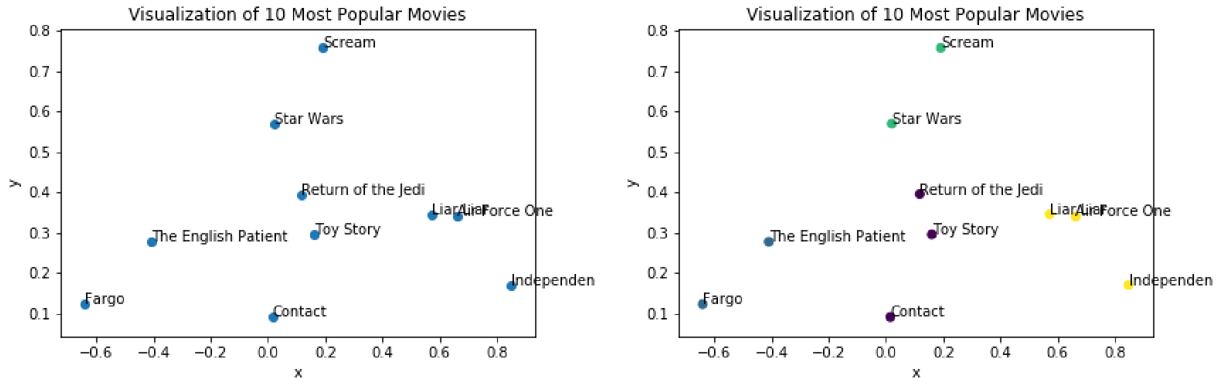
Below, we include the 6 visualizations using the results of this method, along with the results of k-means clustering:



(a) unclustered

(b) clustered via k-means

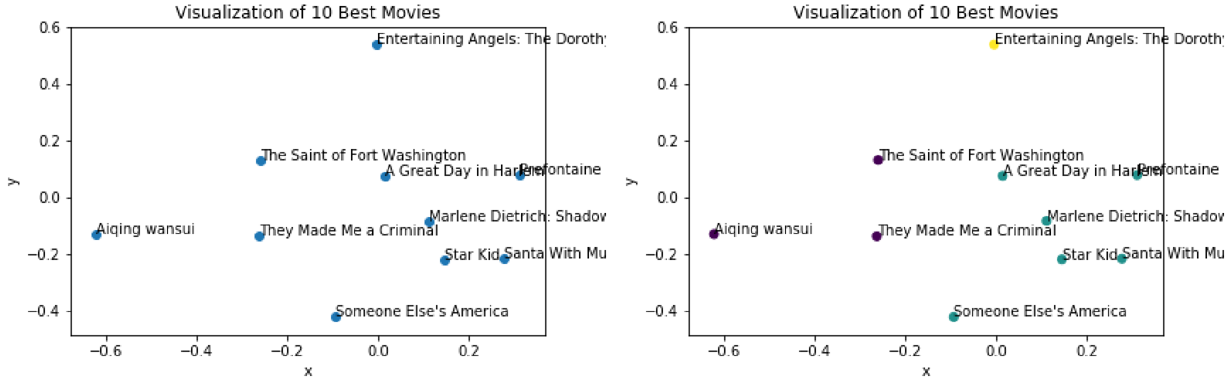
Figure 15: Visualization of 10 Movies, Biased SVD



(a) unclustered

(b) clustered via k-means

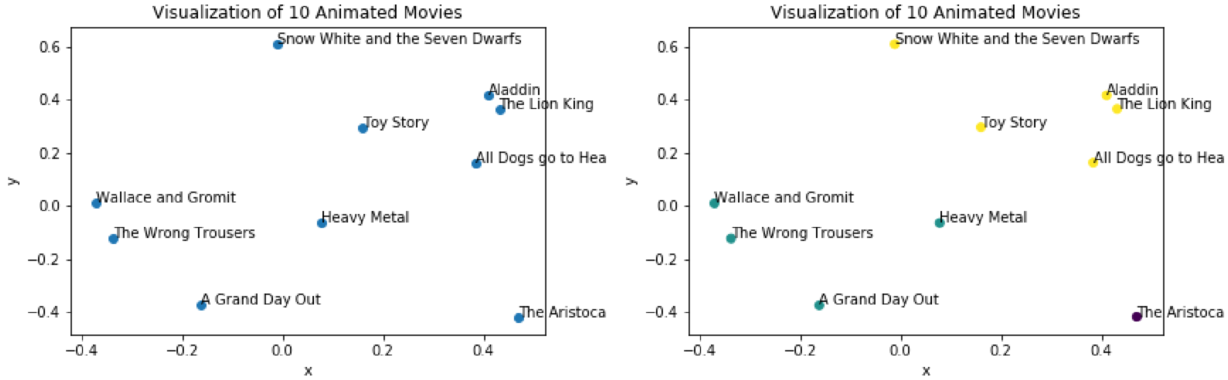
Figure 16: Visualization of 10 Most Popular Movies, Biased SVD



(a) unclustered

(b) clustered via k-means

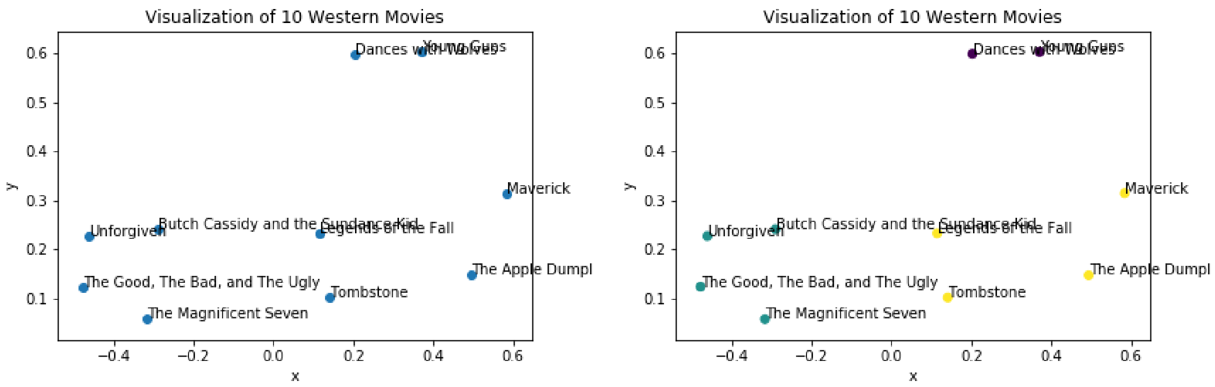
Figure 17: Visualization of 10 Best Movies, Biased SVD



(a) unclustered

(b) clustered via k-means

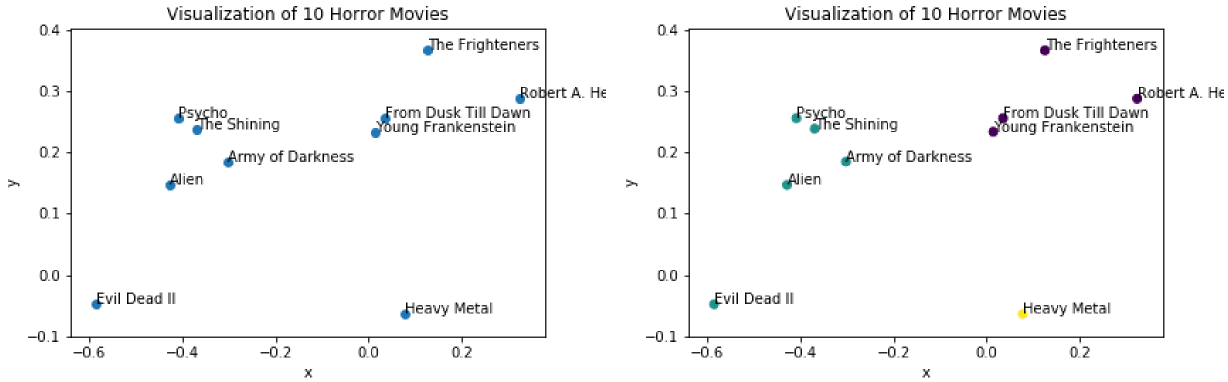
Figure 18: Visualization of 10 Animated Movies, Biased SVD



(a) unclustered

(b) clustered via k-means

Figure 19: Visualization of 10 Western Movies, Biased SVD



(a) unclustered

(b) clustered via k-means

Figure 20: Visualization of 10 Horror Movies, Biased SVD

3. SVD using off-the-shelf matrix factorization in Surprise. This package is called SVD++ and is similar to regular SVD except that it also takes into account implicit ratings. It tries to minimize the same error as SVD but has an extra component in the prediction process. SVD++ uses the following algorithm to predict y_{ij} with the following formula instead of just the u, v, and the bias terms:

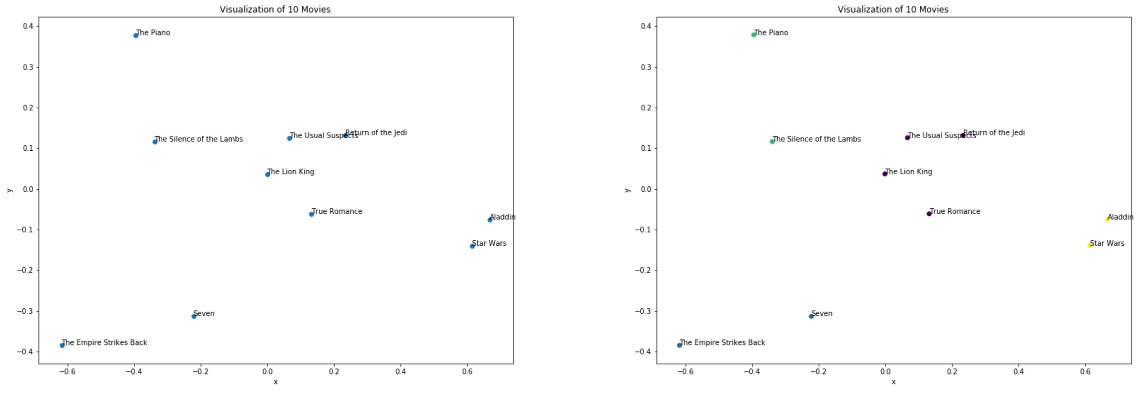
$$y_{ij} = \mu + a_y + b_j + u_i^T \left(v_j + |J_i|^{-\frac{1}{2}} \sum_{k \subset I_j} c_k \right)$$

In this algorithm, c_k is a new set of item factors that capture the implicit ratings of each movie taken from the matrix that contains the rating for each movie by a specific user. This is determined by how the user rated items regardless of the actual rating of the movie; for example, the fact that a user rated an item in the first place contains implicit information, since the chances that the user liked a movie that they rated are higher than for a random movie they did not rate.

We used grid search to tune the hyperparameters for this method and the following method. When we used grid search, we found that the most optimal values were to a number of epochs of 20, a learning rate of 0.01, and a regularization constant of 0.1 to create the least error when using the root mean square error calculator, which produced an error of around 0.94. Since the error functions in our own implementations have an extra factor of $\frac{1}{2}$, we can divide this error by 2 to get about 0.47, which is comparable to, albeit slightly higher than, the errors we achieved with our own implementations.

This method did not perform as well empirically as the previous methods, some notable movies that we intuitively expected to be clustered together were not; this algorithm separated the 3 Star Wars movies and was not able to classify 'Aladdin' and 'The Lion King' together as children's animations. However, we did notice that the central group had some characteristics in common since 'True Romance' and 'The Usual Suspects' are both categorized as crime movies and 'Return of the Jedi' and 'True Romance' are both classified as romances. In addition to that, we see that 'The Usual Suspects' and 'The Silence of the Lamb', both of which are thrillers, are relatively close.

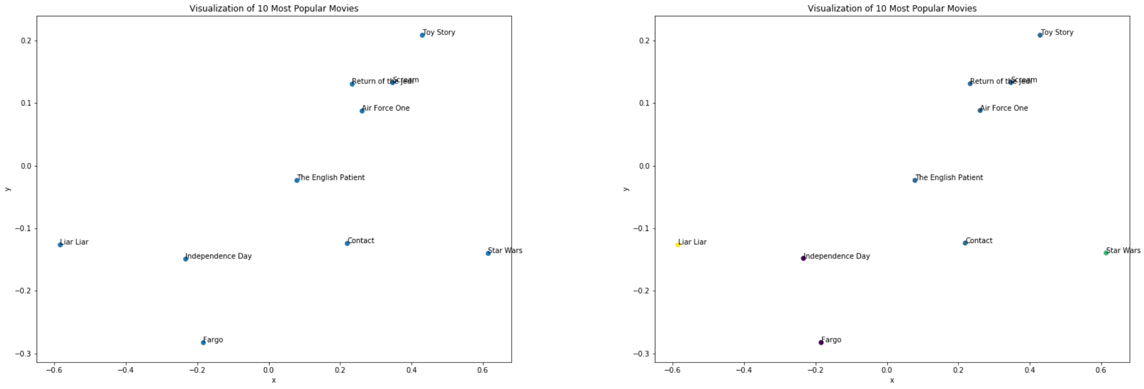
Below, we include the 6 visualizations using the results of this method, along with the results of k-means clustering:



(a) unclustered

(b) clustered via k-means

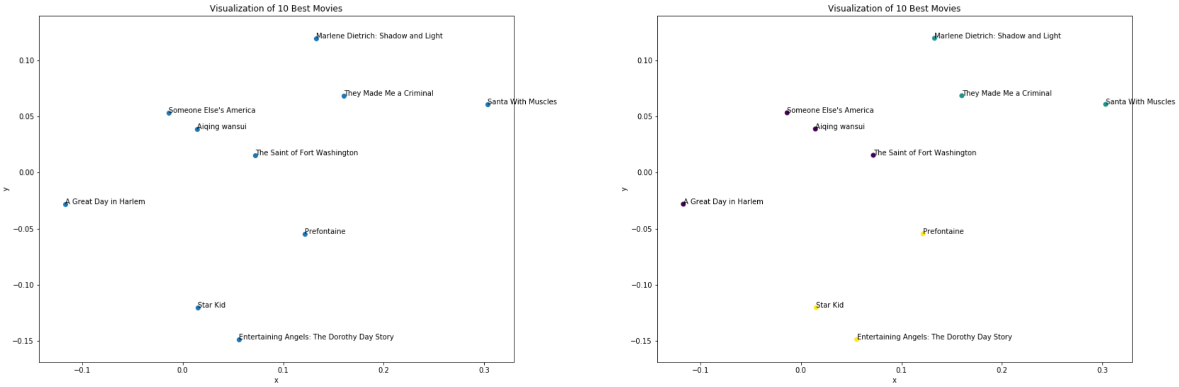
Figure 21: Visualization of 10 Movies, SVD++



(a) unclustered

(b) clustered via k-means

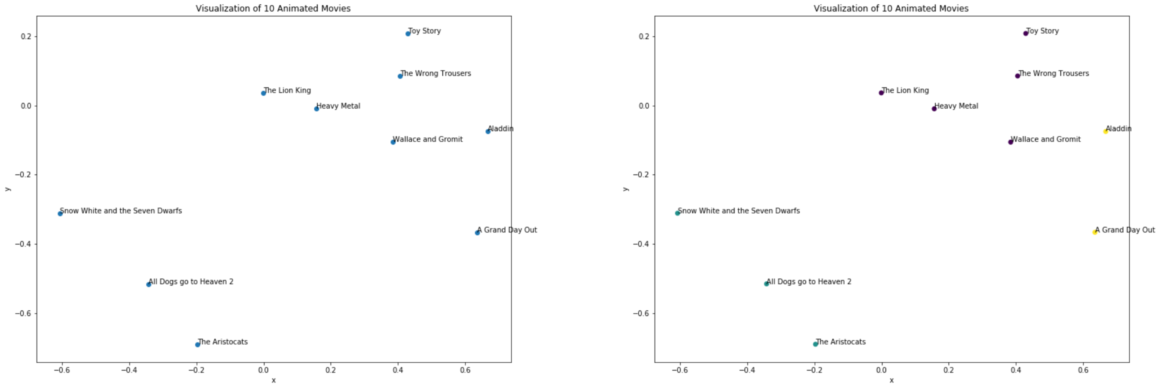
Figure 22: Visualization of 10 Most Popular Movies, SVD++



(a) unclustered

(b) clustered via k-means

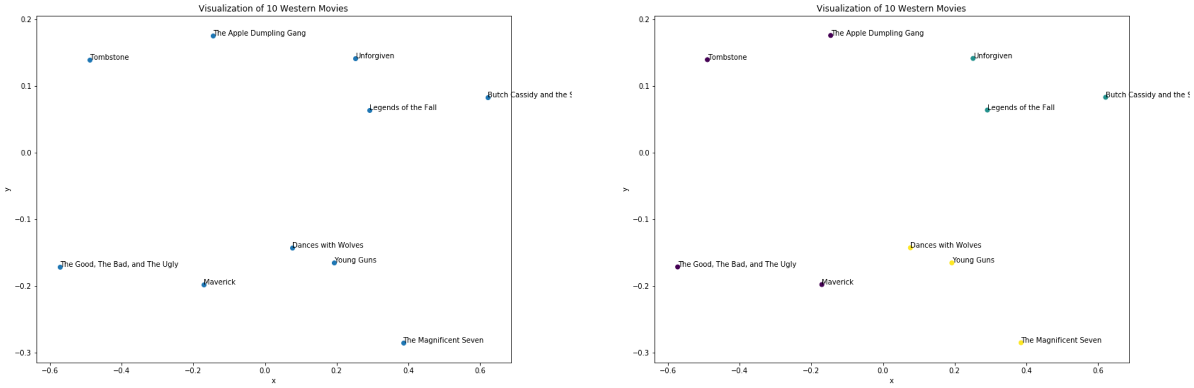
Figure 23: Visualization of 10 Best Movies, SVD++



(a) unclustered

(b) clustered via k-means

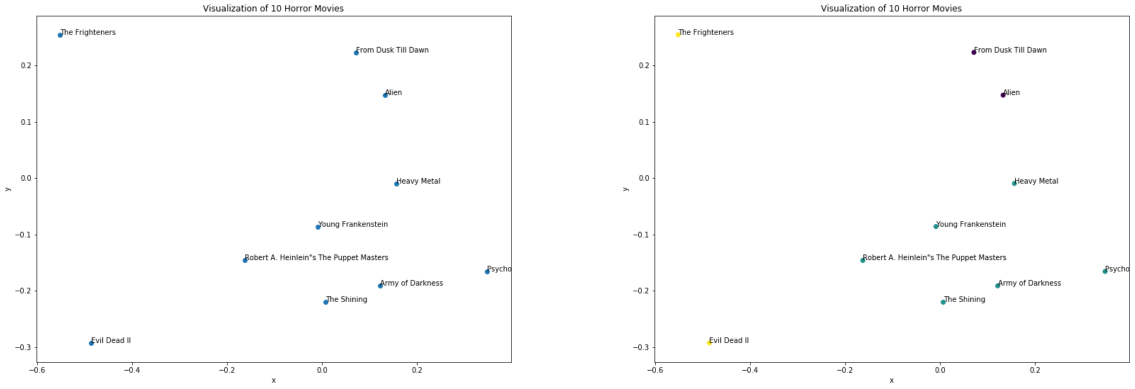
Figure 24: Visualization of 10 Animated Movies, SVD++



(a) unclustered

(b) clustered via k-means

Figure 25: Visualization of 10 Western Movies, SVD++



(a) unclustered

(b) clustered via k-means

Figure 26: Visualization of 10 Horror Movies, SVD++

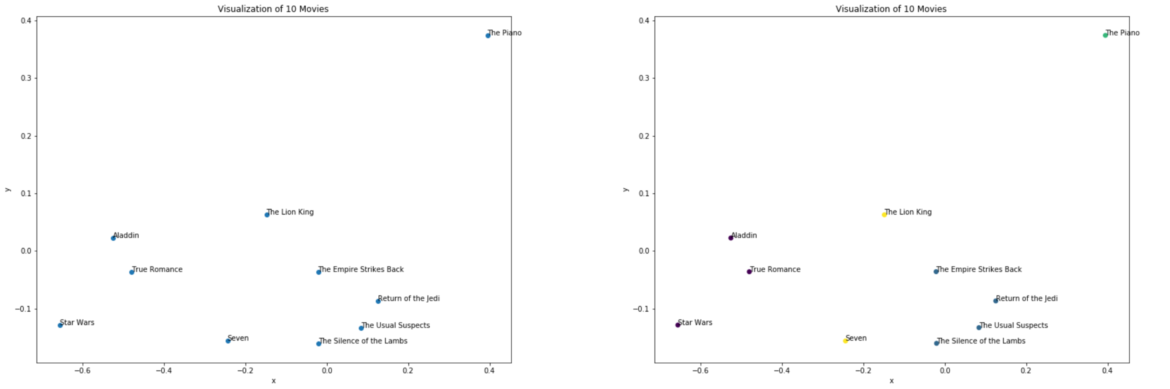
4. We also implemented the non-negative matrix factorization (NMF) as another off-the-shelf implementation from Surprise. This method slightly differs from SVD in how it updates the parameters. It calculates the expected rating the same way as SVD without bias, but it updates U and V differently, in the following format:

$$u_{if} = u_{if}\left(\frac{\sum_{j\subset J_i} v_{jf}r_{ij}}{\sum_{j\subset J_i} v_{jf}\mu_r + \lambda_i|J_i|u_{if}}\right)$$
$$v_{jf} = v_{jf}\left(\frac{\sum_{i\subset I_j} u_{if}r_{ij}}{\sum_{i\subset I_j} u_{if}\mu_r + \lambda_j|I_j|v_{jf}}\right)$$

In this algorithm, the goal is to learn the user and movie matrices such that they have all positive elements. This implementation is possible for this application because the ratings matrix Y has only positive elements (ratings range from 1-5). NMF uses regularized stochastic gradient descent to determine the values within the update formula as above. This includes keeping λ positive to ensure that all the components of the matrix of U and V are positive.

We can see that this model performed better than the SVD++ model. From this model we see that the separation of more movies was more intuitive. We observe that 2 of the Star Wars movies were mapped closer together. In addition, we see that 'The Usual Suspects' and 'The Silence of the Lambs' are close to each other as well as 'Seven', which makes sense given that all three are thrillers. We also notice that 'Aladdin' and 'The Lion King' are relative close. However, not all movies that were "close" together were clustered together by k-means, which suggests that more points are needed for the clusters to become more well-defined. In addition, the number of clusters could be adjusted. Some problems/unintuitive results that resulted from this method included 'Star Wars' not being close to the other Star Wars franchise movies and 'True Romance' and 'The Piano' not being near each other despite both being romance movies.

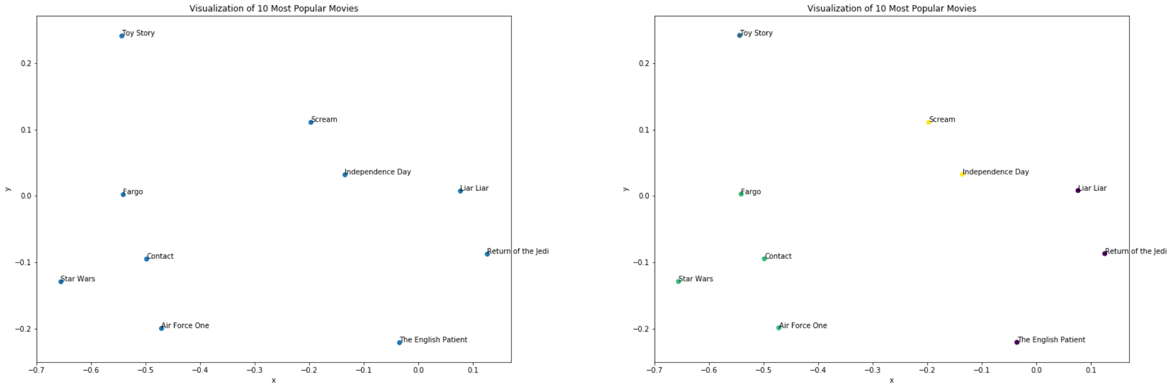
Below, we include the 6 visualizations using the results of this method, along with the results of k-means clustering:



(a) unclustered

(b) clustered via k-means

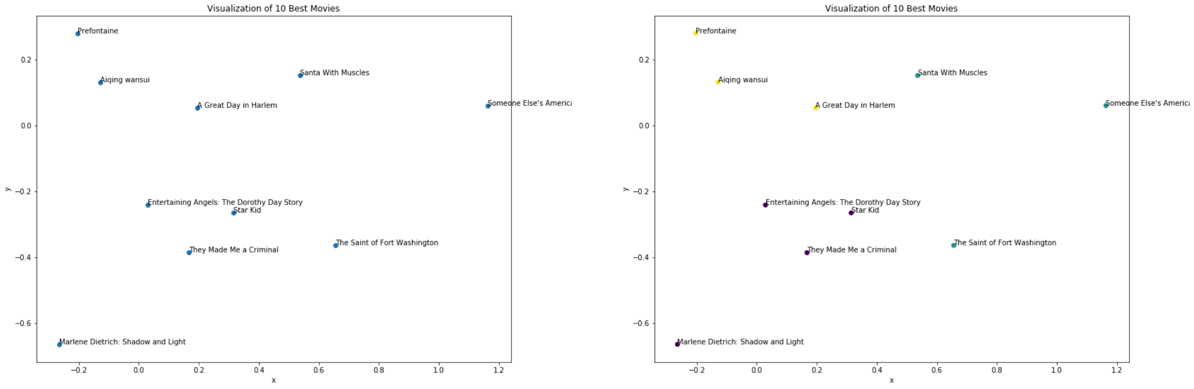
Figure 27: Visualization of 10 Movies, NMF



(a) unclustered

(b) clustered via k-means

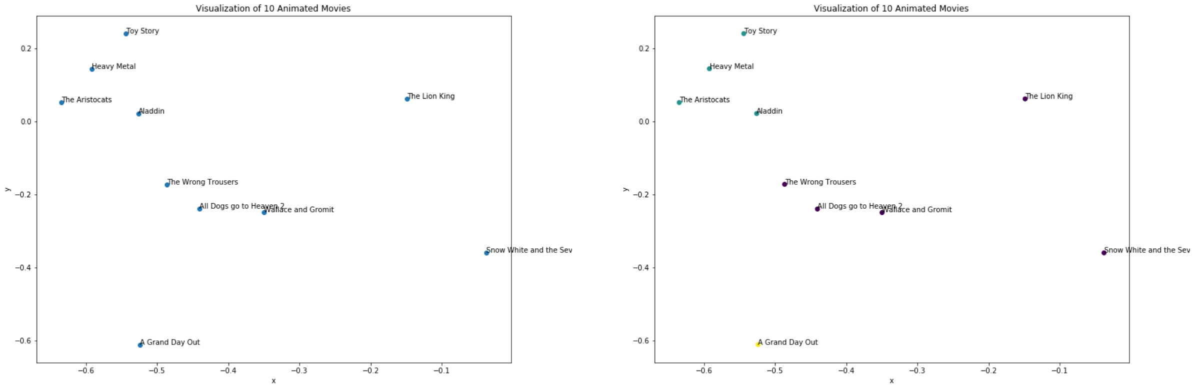
Figure 28: Visualization of 10 Most Popular Movies, NMF



(a) unclustered

(b) clustered via k-means

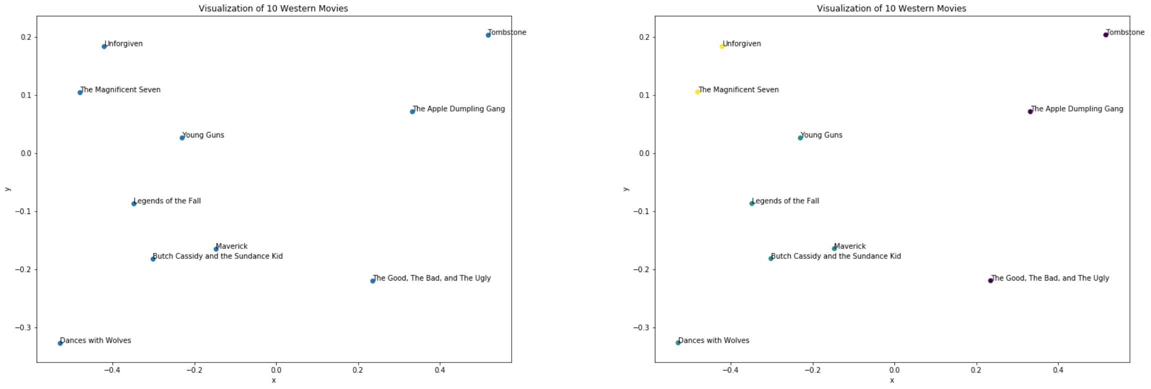
Figure 29: Visualization of 10 Best Movies, NMF



(a) unclustered

(b) clustered via k-means

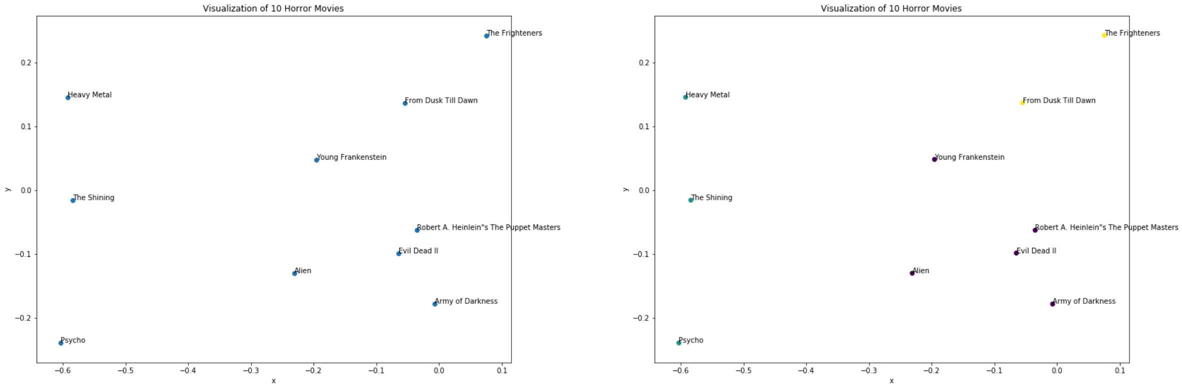
Figure 30: Visualization of 10 Animated Movies, NMF



(a) unclustered

(b) clustered via k-means

Figure 31: Visualization of 10 Western Movies, NMF



(a) unclustered

(b) clustered via k-means

Figure 32: Visualization of 10 Horror Movies, NMF

We will discuss our plots and other interesting visualizations in detail in the following section.

4 Matrix Factorization Visualizations [30 points]

Generally, we observe that movies of similar genres tend to be clustered closer together than movies of wildly different genres. For example, we used k-means clustering on the 10 movies in Visualization (a) for each different matrix factorization method and obtained the following plots:

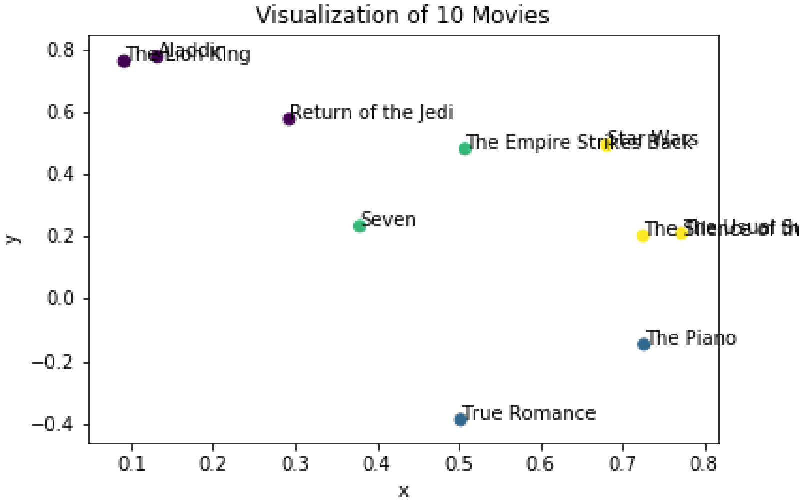


Figure 33: Visualization of 10 Movies, Clustered, SVD without Bias

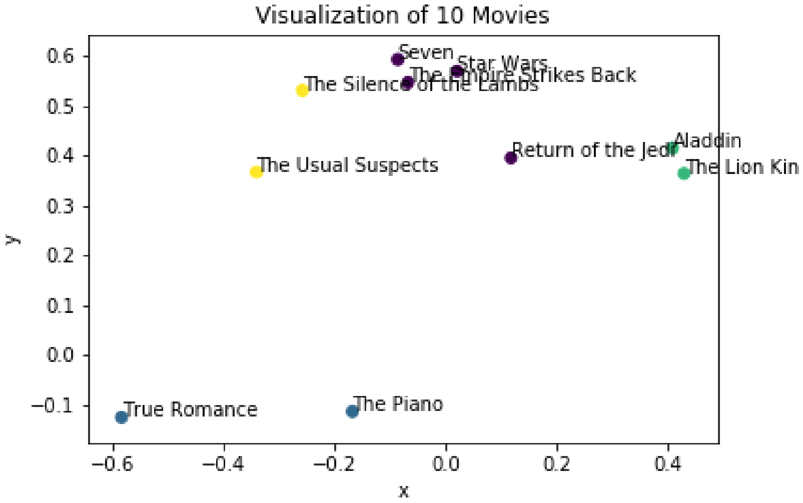


Figure 34: Visualization of 10 Movies, Clustered, SVD with Bias

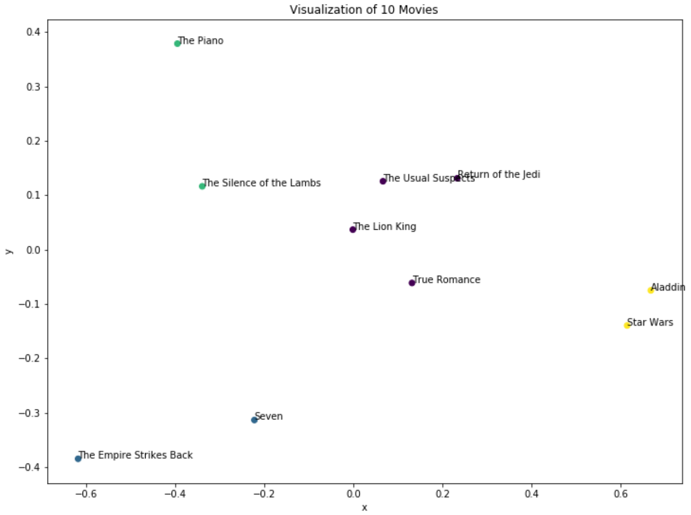


Figure 35: Visualization of 10 Movies, Clustered, SVD++

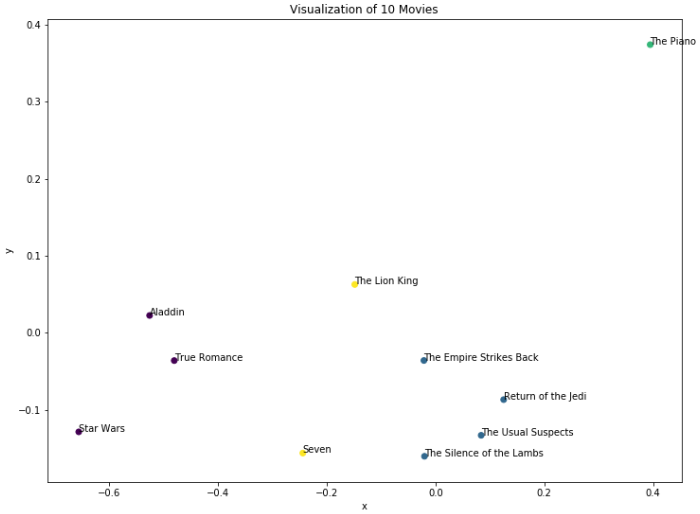


Figure 36: Visualization of 10 Movies, Clustered, NMF

We chose these movies because they encompass a variety of genres, including thrillers, romance movies, animated children's movies, and action/adventure movies. Additionally, there are certain movies that we expect to cluster together, so we can use these groupings (or lack thereof) to empirically evaluate the models.

From the clusters resulting from SVD without bias, we observe that while some of the groupings are intuitively expected ('The Usual Suspects' and 'The Silence of the Lambs' are both thrillers; 'The Piano' and 'True Romance' are both romance movies; and 'Aladdin' and 'The Lion King' are both animated children's movies), some groupings do not make as much sense (Star Wars movies are not all clustered together, and 'Seven' and 'The Usual Suspects' are both crime thrillers, yet are not grouped together).

From the clusters resulting from bias-incorporated SVD, we observe that the Star Wars movies are all clustered together, as are the animated children's movies ('The Lion King' and 'Aladdin'), and the romance movies ('The Piano' and 'True Romance'). Like the first visualization, there are some unexpected results ('Seven' is a crime thriller, yet it is not near 'The Usual Suspects', which is also a crime thriller). However, it does makes sense that the action movies (Star Wars movies) are relatively close to and have some overlap with the thrillers ('The Silence of the Lambs', 'Seven', 'The Usual Suspects'), since thrillers often have elements of action as well. This may be a reflection of the bias-incorporated method's ability to learn more nuanced information from the ratings that translates to genres with similar elements being grouped together, since users may tend to give more similar ratings to movies that are similar in genre.

From the clusters resulting from Surprise SVD++, we observed some interesting grouping. We notice that some of the romance movies ('True Romance' and 'Return of the Jedi'), the drama movies ('The Piano' and 'The Silence of the Lamb'), and the crime movies ('True Romance' and 'The Usual Suspect') are clustered together. However, we do notice some unexpected clusterings. The results show that 'Seven', which is a crime thriller, is not close to 'The Usual Suspect', which is also a crime thriller. In addition, we notice that none of the Star Wars movies are close to each other and 'The Lion King' and 'Aladdin', both animated children movies, are not close to each other. Lastly, we also notice 'The Piano' and 'True Romance', both romance movies, are not together. However, we do notice that all movies in the purple grouping in the middle of the plot have at least one genre in common with each other besides 'The Lion King'. The grouping of this data is not as intuitive as the previous examples and that is likely due to the additional individual factor that is added to each component in the optimization objective. This is likely what creates the odd groupings that match some movies together ('True Romance', a crime romance, with 'Return of the Jedi', an action romance) based on implicit characteristics.

From the clusters resulting from Surprise NMF, we noticed that some of the adventure movies ('The Empire Strikes Back' and 'Return of the Jedi'), action romance movies ('Star Wars' and 'True Romance'), and thriller movies ('The Usual Suspects' and 'The Silence of the Lambs') are clustered together. We do however notice some unexpected results. One notably is that the adventure and thrillers movies from above are grouped together. In addition, we notice that 'The Lion King', an animated children's movie, is grouped with 'Seven', a crime thriller. We also notice that 'The Piano' is grouped by itself in the top right. In fact, 'The Piano' seems to be somewhat of an outlier from the other movies in all of the implementations, suggesting that there are implicit characteristics that set this movie apart. We also see that 'Aladdin', an animated children's movie, is grouped with 'True Romance' and 'Star Wars' rather than with 'The Lion King', the other animated children's movie. However, it does appear that 'The Lion King' and 'Aladdin' are close to each other, so if more movies were added and the clusters become more well-defined, it is possible that they could be clustered together. The same could be said of the thrillers ('Seven', 'The Usual Suspects', and 'The Silence of the Lamb').

Compared to our own unbiased and biased SVD implementations, the results of the off-the-shelf methods are less intuitive. We also noticed that there were patterns with each off-the-shelf methods, we noticed that it tends to groups certain movies together over many runs. The most notable being 'Aladdin' and 'Star Wars'. These two movies possibly something in common that we do not know about that is not related to the genre.

We also colored movies by their average ratings to visualize how ratings affect the projections in the different matrix factorization methods:

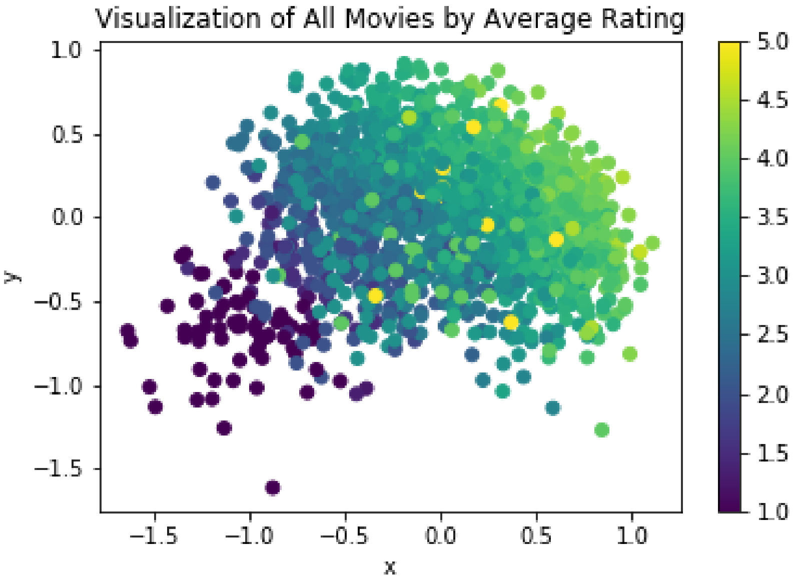


Figure 37: Movies Colored by Average Rating, Non-biased SVD

We observe that for the non-bias-incorporated SVD, the x-axis appears strongly correlated with rating, with higher x-values corresponding to higher average ratings, suggesting that the first principal component encodes the average rating of the movies. The y-axis appears weakly correlated with rating, as most of the lowest-rated movies also have lower y-values. However, we do observe some outliers, including movies with an average rating of near 5 (represented by yellow points) that are near the center of the plot.

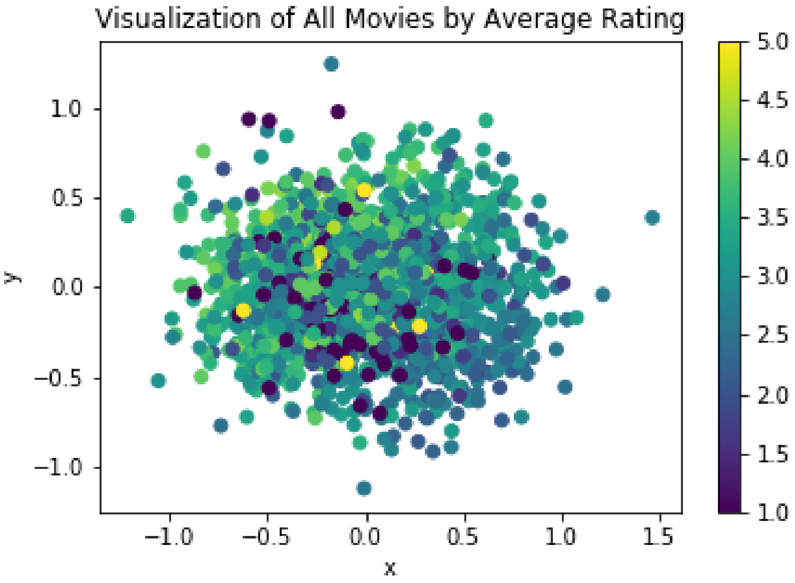


Figure 38: Movies Colored by Average Rating, Biased SVD

For the bias-incorporated SVD, there does not appear to be as obvious of a correlation between axis and rating. In general, the rating seems to decrease slightly with higher x-values and lower y-values (more light green points to the upper left of the plot, and more blue points to the lower right), but the trend is not as definitive as for the unbiased SVD. This suggests that the first two PCs may be capturing some other, more nuanced, characteristic of the data in addition to information given explicitly by the ratings. Moreover, it could be that this representation is too low-dimensional to provide the full picture of how this method encodes information.

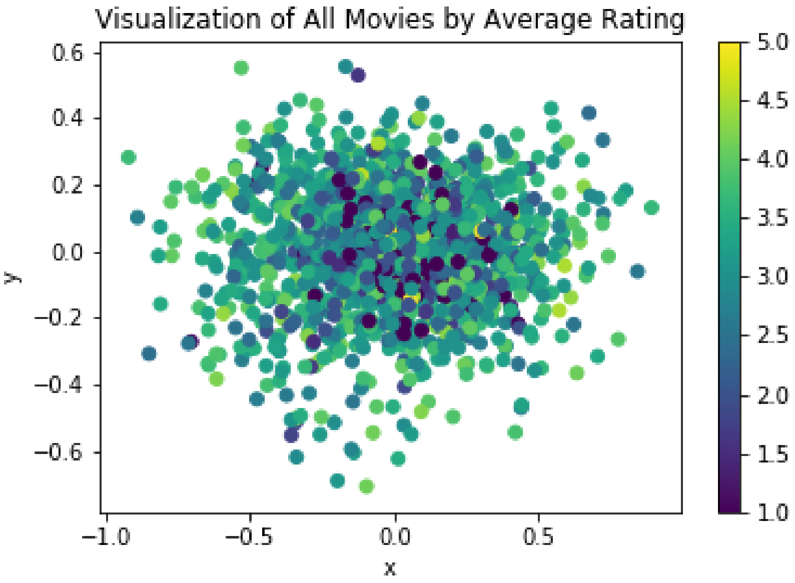


Figure 39: Movies Colored by Average Rating, SVD++

For the SVD++ from Surprise, there does seem to be a slight correlation between the axis and the average ratings. We see that in general, the lower rated movies are in the the center of the diagram while the rating become higher as we move out of the center of the plot. Like the previous plot, we see that this trend is not as obvious as our SVD method. However since there is a general trend, it possible that our graph does not have enough dimensions to convey the trend easily to us. As a result, there is a possibility that the rating of the movies affected the axis.

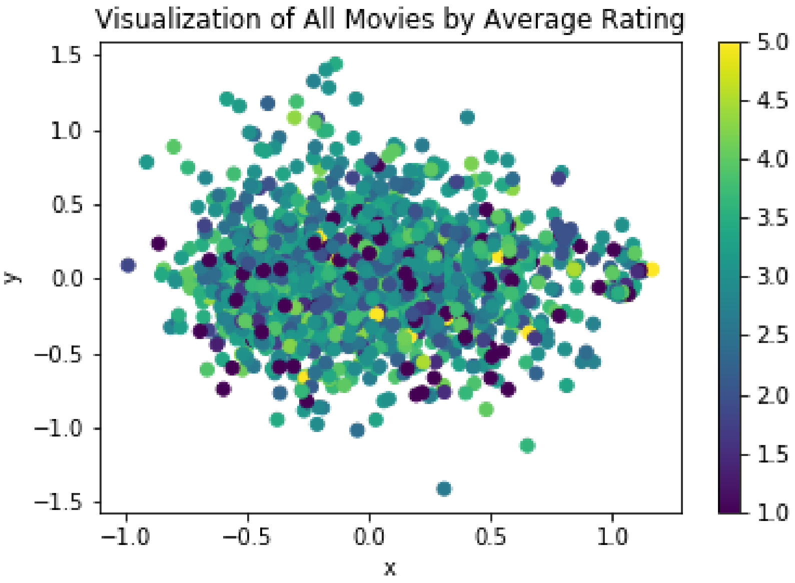


Figure 40: Movies Colored by Average Rating, NMF

From the NMF plot, we see that there is no real trend aside from the fact that a lower x tended to have a higher rating than movies with a higher x value. In addition to that we see that the extremely low rating and extremely high ratings of 1 or 5s are scattered across the plot. If we only take a look at the non extreme ratings. we see that the we see a general trend for lower rated scores with a higher x value and a higher score with lower x value. The y value does not seem to be correlated with the ratings. A possible reason that movies with extreme scores are scattered everywhere is that because of the low amounts of ratings for each movie that had an extreme average rating, there was not enough information to establish the same general trend as with average rated movies with more reviews. In addition to that, it could also be a lack of dimensions that does not allow us to display the true trend like the above plots.

Comparing the best vs most popular movies, we observe that when using SVD without bias, the best movies tend to have a smaller range of x- and y-values than the most popular movies. This occurs because the best movies all have high ratings, so they tend to be found in a more narrow spread due to the effect that average rating has on clustering, as we discussed above. On the other hand, popular movies are not necessarily the most highly rated, since many more people have seen them and there may be more dissenting/negative opinions. Thus, the spread of the most popular movies is higher, as we can see in the following figures:

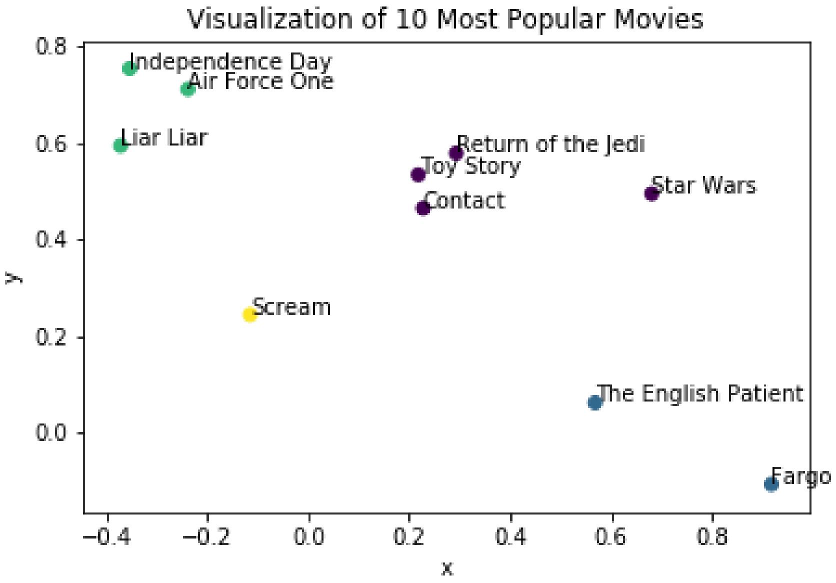


Figure 41: Visualization of 10 Most Popular Movies, Unbiased SVD

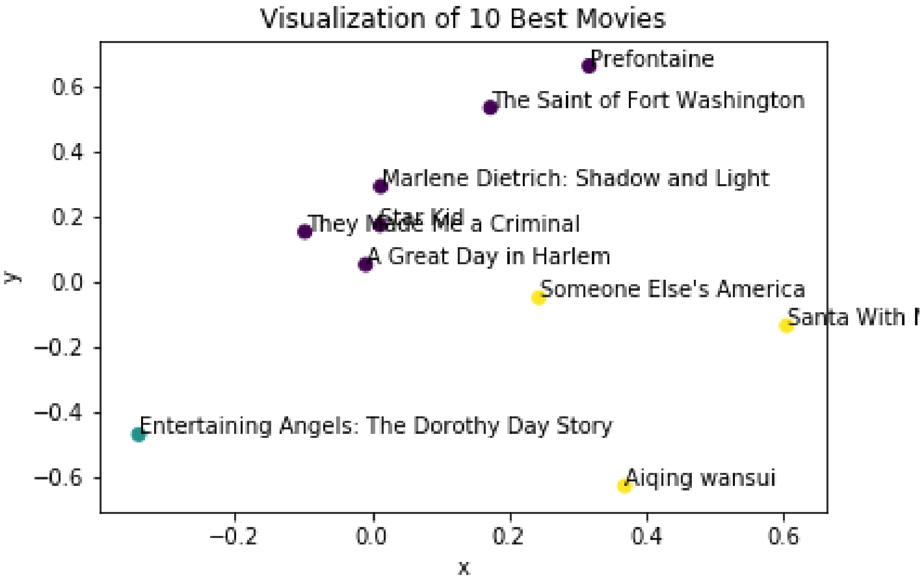


Figure 42: Visualization of 10 Best Movies, Unbiased SVD

We also notice that there are no obvious visual clusters among the 10 best movies, suggesting that high ratings are not necessarily indicative of other shared characteristics. In fact, we observe that the best movies span many genres, from children's to crime to adventure to documentary and more. On the other hand, there are more visual clusters apparent in the plot for the 10 most popular movies, suggesting that movies that have the most ratings also share other common characteristics. For example, 'Star Wars' and 'Return of the Jedi' are from the same popular franchise, so we would expect them to be clustered together. We also observe an abundance of thriller, action, and war movies among these 10, including 'Independence Day' and 'Air Force One', which are both action movies; this may explain the appearance of more distinct clusters for the most popular movies.

Similarly, we can examine the plots of the most popular and best movies resulting from bias-incorporated SVD:

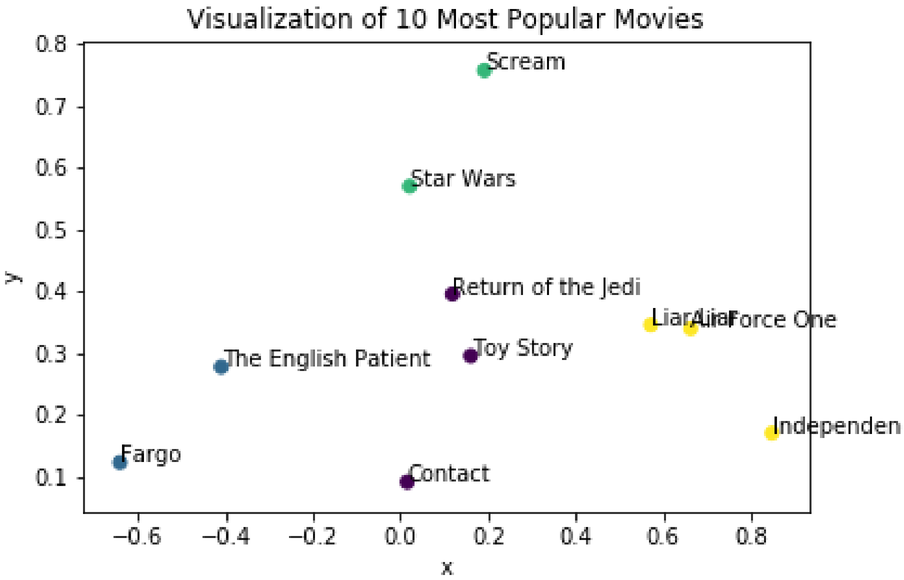


Figure 43: Visualization of 10 Most Popular Movies, Biased SVD

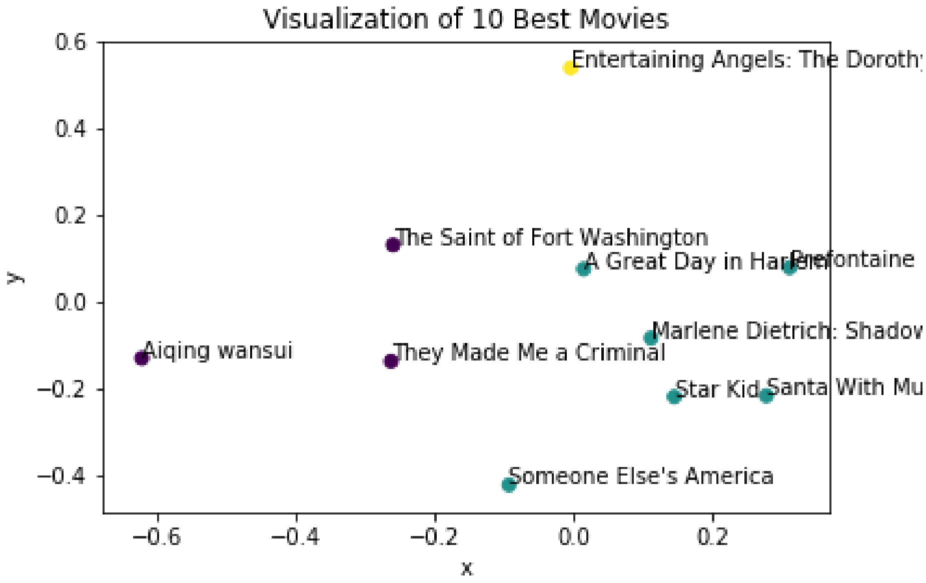


Figure 44: Visualization of 10 Best Movies, Biased SVD

Once again, we observe that there are no obvious visual clusters among the 10 best movies, whereas clusters are more apparent among the 10 most popular movies. Similarly to the unbiased version, we see that ’Air Force One,’ ’Liar Liar,’ and ’Independence Day’ are clustered together, while ’The English Patient’ and ’Fargo,’ while in the same cluster, are not particularly close together. This may have occurred simply because these two movies are farther from the rest; perhaps it would be better to separate them each into their own individual cluster, since each belongs to three genres and only one overlaps between them. Moreover, in both versions, it seems that ’Scream’ is farther from the rest of the movies. This observation is in line with the fact that ’Scream’ is the only horror movie among with 10 most popular.

In the figures below, we plotted the most popular and the best movies together. We observe that these groups appear to be fairly distinct from each other, which suggests that the most popular movies and the best movies do not necessarily have many characteristics in common.

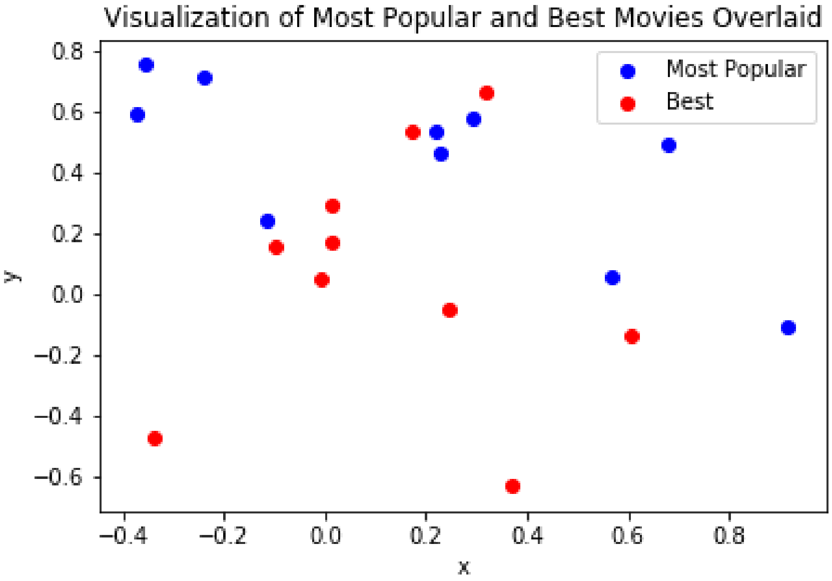


Figure 45: Visualization of 10 Best Movies and 10 Most Popular Movies Overlaid, Unbiased SVD

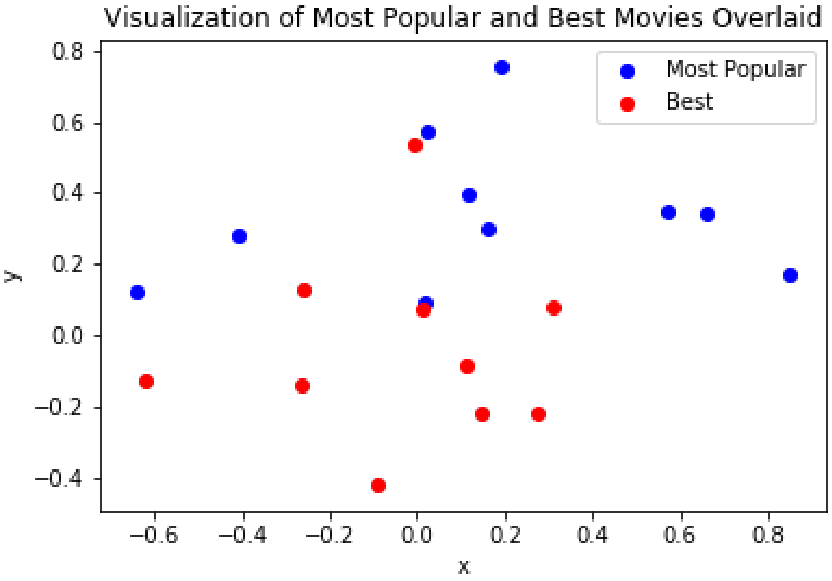


Figure 46: Visualization of 10 Best Movies and 10 Most Popular Movies Overlaid, Biased SVD

When we take a look at at the most popular and best movies using the projections from SVD++, we see that the most popular movies are generally in a linear pattern going from a low x and y to a high x and y with some other movies that are spread out near the y = -0.1 value. This shows that there is some pattern connecting the most popular movies. On the other hand, the visualization of the best movies is a lot more scatter. This could be because of how few ratings these movies have and thus they have no real correlation with anything. From this we also do not see a clear relationship between the most popular and the best movies.

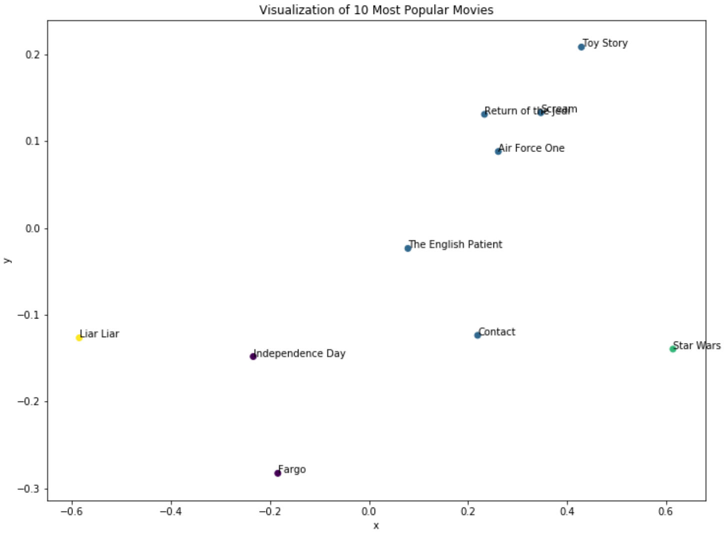


Figure 47: Visualization of 10 Most Popular Movies, SVD++

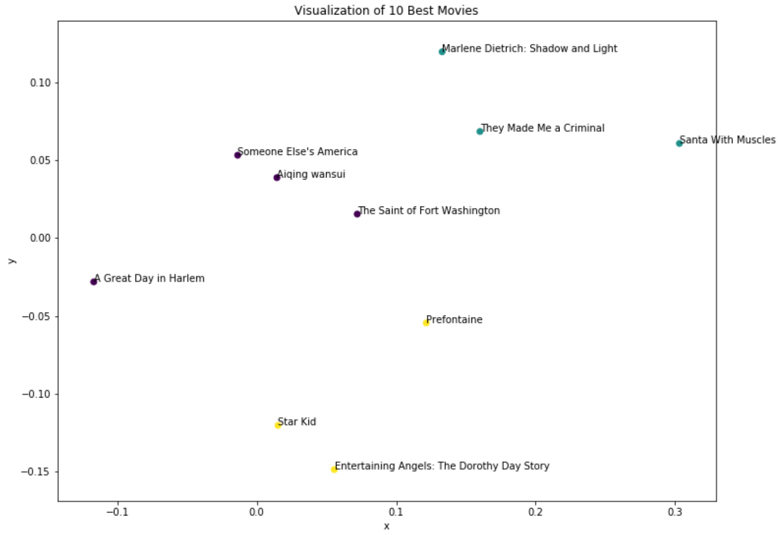


Figure 48: Visualization of 10 Best Movies, SVD++

From our visualizations using results from NMF, we see that a lot of popular movies tend to have lower x-values than the best movies. Overall, the best movies have a much wider spread in x than the most popular movies, and they also have a slightly wider spread in y. This may be a reflection of the wide variety of genres that the best movies fall into, and the fact that they do not seem to have much in common with each other.

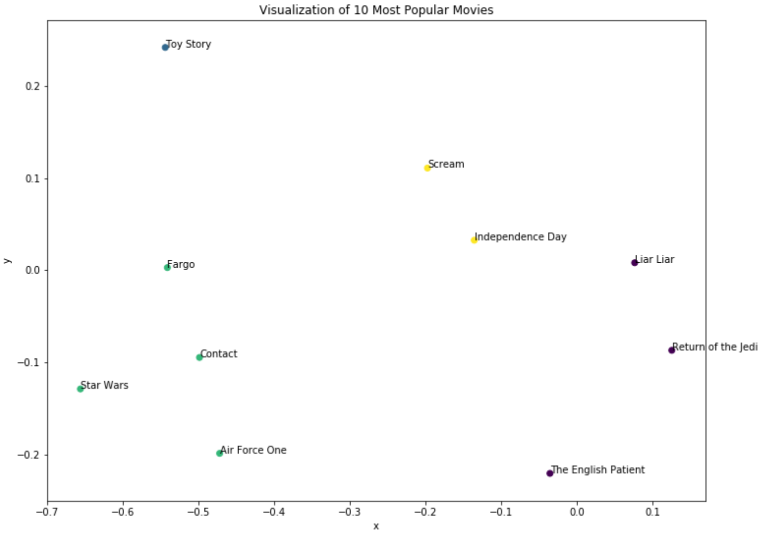


Figure 49: Visualization of 10 Most Popular Movies, NMF

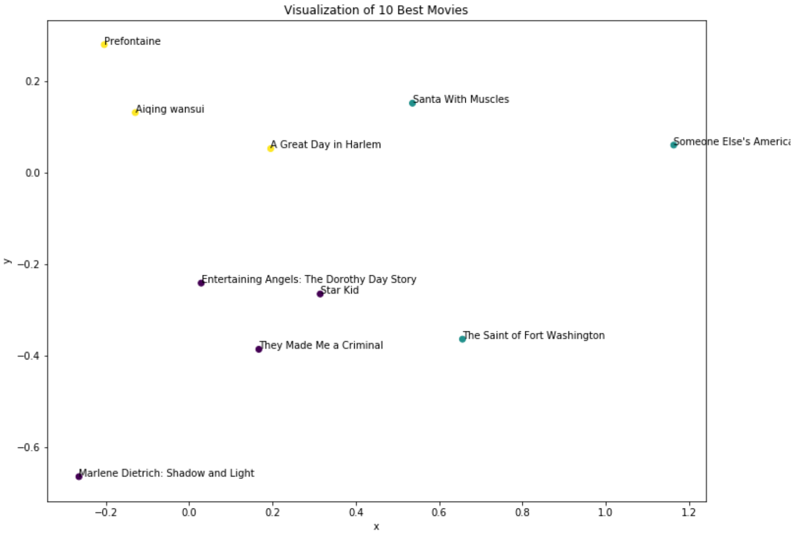


Figure 50: Visualization of 10 Best Movies, NMF

From our visualizations of movies of three different genres (Animated movies, Western movies, and Horror movies) in Section 3, we observe a few trends.

For the unbiased SVD plot of animated movies, many Disney movies like 'Aladdin,' 'Toy Story, 'and 'The Lion King' are clustered together. The same is true for the biased SVD plot. This trend only somewhat holds for the off-the-shelf implementations that we used. One outlier that seems far from the other movies in the unbiased and biased SVD plots is 'The Aristocats'; this may be due to the fact that it is one of the oldest among the animated movies that we chose, so there may be elements of the animation style that cause a difference in the ratings.

For the Western movies, we see that there are two distinct groups resulting from the unbiased SVD implementation: ['Maverick', 'The Apple Dumpling Gang', 'Young Guns', 'Legends of the Fall', and 'Tombstone'] in one cluster with lower x-values and ['Butch Cassidy and the Sundance Kid', 'The Magnificent Seven', 'The Good, the Bad, and the Ugly', and 'Unforgiven'] in the other cluster with higher x-values. One outlier is 'Dances with Wolves', which does not obviously belong to either group. One obvious distinction between the two groups is that the first consists mostly of movies made in the 1980s and 1990s, while the second consists mostly of movies made in the 1950s and 1960s. This indicates that there may be a relationship between the time period of the movie and its quality/appeal to various audiences, which may affect rating. Perhaps older Western movies are considered more "classic" and thus have higher x-values (which correspond to higher ratings on average) than newer Western movies. Although this distinction is not as apparent from the off-the-shelf plots, we see almost the exact same clustering in the plot from the biased SVD, which provides more support for this interpretation.

In terms of horror movies, "classic" horror movies like 'Psycho', 'The Shining', and 'Alien' tend to be clustered together, which may be a reflection on their collective popularity. 'Psycho' and 'The Shining' in particular are clustered together by all of our implementations, which may result from the fact that they both have elements of psychological thrillers as well.

Compared to each other, the three genres that we decided to visualize have different characteristics.

From the results of unbiased SVD, we observe that the horror movies tend to have lower y-values than the animated and western movies. This trend is more apparent when we plot all three genres on the same figure, as below:

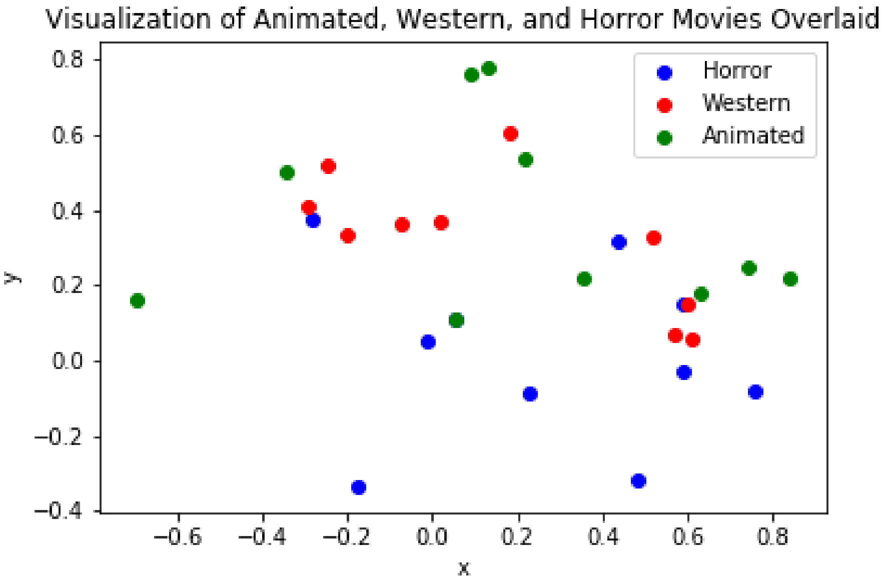


Figure 51: Visualization of 10 Horror, 10 Western, and 10 Animated Movies, Unbiased SVD

Since we observed above that for the unbiased SVD implementation, y-axis is loosely correlated with rating, this suggests that these horror movies are generally rated lower than the animated and western movies. This conclusion is supported by our basic visualizations, from which we observed that horror movies had far more low ratings as a genre than animated or western movies. This plot also suggests that animated movies and western movies have more in common with each other than with horror movies.

We can examine the same plot for the bias-incorporated implementation:

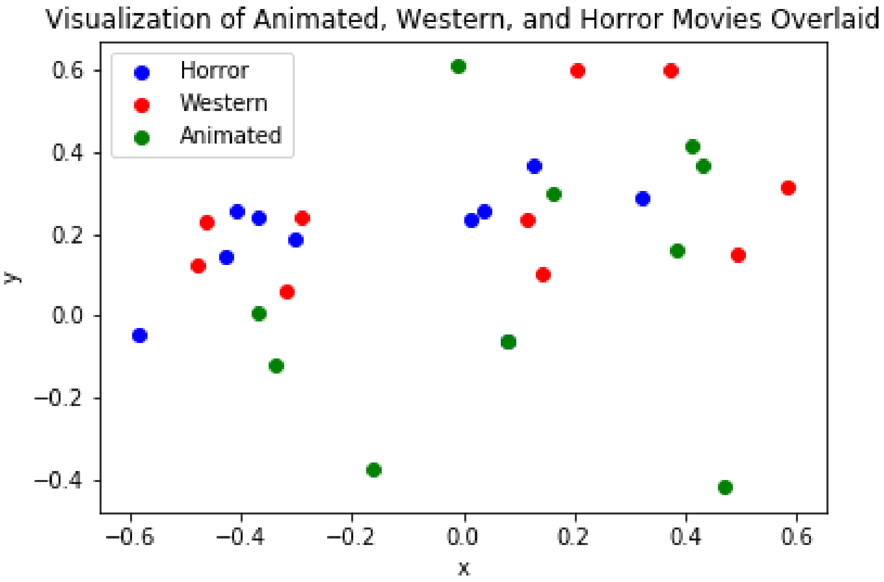


Figure 52: Visualization of 10 Horror, 10 Western, and 10 Animated Movies, Biased SVD

This time, it seems that the horror movies and the western movies have more in common with each other than with the animated movies. This was at first rather unexpected given that western and animated movies seem more similar in terms of average ratings; however, it does seem reasonable that we would not observe the same behavior as in the unbiased SVD case, since neither of the axes in the biased case have an extremely clear correlation with rating. This newly observed phenomenon suggests that western and horror movies are similar in ways that go beyond ratings, which is supported by the fact that these two genres often share similar elements of action and violence.

We can examine the same plot for the Surprise SVD++ implementation:

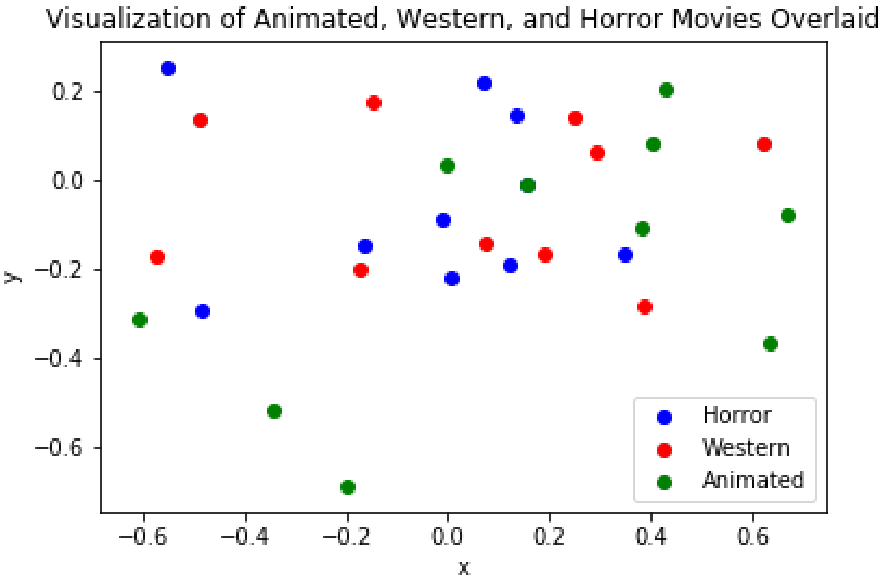


Figure 53: Visualization of 10 Horror, 10 Western, and 10 Animated Movies, SVD++

We can see that in this plot, the general trend is that a majority of the animated movies are located on the right and/or bottom of the plot while most of the horror movies occupied the top and/or left of the plot. This shows a clear distinction between these two genres and their differences, since most animated movies are not horror movies and are typically targeted towards children, so those genres do not overlap much. Lastly, we see that the Horror and Western movies both have high y values and do not have low y values. In addition, it is hard to determine a clear distinction between the locations of horror and western movies since they are mixed together. This suggests that SVD++ was able to separate these chosen animated movies from the movies we chose from the other 2 genres.

We can examine the same plot for the Surprise NMF implementation:

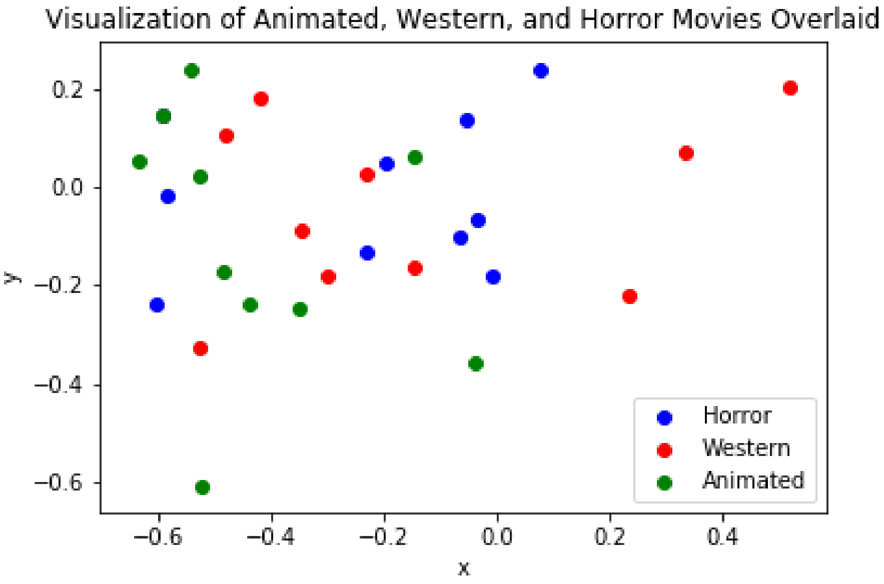


Figure 54: Visualization of 10 Horror, 10 Western, and 10 Animated Movies, NMF

We find that the data seems to be very scattered, but we notice that animated movies occupy the left of the plot. Additionally, we notice that the horror movies have a more moderate range of values in both axes while the Western and animated movies have more spread in the x and y axes, respectively. Overall, however, it seems that the three genres occur similar areas near the center of the plot, and as a result, it can be harder to distinguish between all the movie types using this model.

Overall, we found that the various different implementations of matrix factorization gave different and sometimes surprising visualization results. The biased version of SVD gave the most easily interpretable results when projected into 2 dimensions compared to the other model; however, the other models could have found more implicit relationships between movies that are simply not noticeable when classifying by genre.