1 Introduction

- **Group members**
  Aidan Swope, Brendan Hollaway, Harald Putterman
- **Team name**
  Sarcastic Gradient Dissent
- **Division of labour**

  Aidan worked on implementation and analysis of our basic visualizations, creating plots and reasoning about the dataset before latent-factor analysis. He also contributed to the analysis of other methods, writing some of the interpretation of many other visualizations in this report, and typeset the final report.

  Harry trained the basic matrix factorization algorithm and modified the code so that it worked with biases. He also optimized both of these algorithms using a gridsearch. Harry wrote code to do the SVD to reduce the dimensionality of the results. He also wrote the pipeline to plot the results and did the analysis of the results for all the plots produced using his methods.

  Brendan set-up and used an off-the-shelf latent-factor model by GraphLab to visualize the data. This required optimizing the model through a GridSearchCV-esque method, transforming data inputs and outputs to GraphLab to be compatible with the data given as well as Harrys SVD-based visualization, and finally visualizing in three-dimensions the test error.

  Harry, Aidan, and Brendan also split the work of analyzing the differences and similarities between the different visualizations produced by the different model types per section (e.g. Comedies or Most Popular Movies).

2 Models used

- **Matrix factorization without the use of a bias term**:

  This matrix factorization technique minimizes the loss below:

  $$\arg\min_{U,V} \frac{\lambda}{2}(||U||^2 + ||V||^2) + \sum_{(i,j)\in S}(Y_{i,j} - u_i^T v_j)^2$$

  We trained the model using stochastic gradient descent that updated the matrices U and V at every iteration.

- **Matrix factorization with a bias term**:

  This matrix factorization technique is very similar to the one above except bias terms are included.

  This model was also trained using SGD with the only difference being that the A and B values were updated for every iteration in addition to the U and V matrices.

  $$\arg\min_{U,V,a,b} \frac{\lambda}{2}(||U||^2 + ||V||^2) + \sum_{(i,j)\in S}(Y_{i,j} - (u_i^T v_j + a_i + b_j))^2$$

The predicted advantage of the model including the bias is that it focuses more on capturing the variability between the users and the movies.

- **Matrix factorization using the third party package GraphLab**:

  We used GraphLabs FactorizationRecommender, which is a latent-factor model that optimizes $\min_{\mathbf{w},\mathbf{a},\mathbf{b},\mathbf{V},\mathbf{U}} \frac{1}{|\mathcal{D}|} \sum_{(i,j,r_{ij})\in\mathcal{D}} \mathcal{L}(\mathrm{score}(i,j), r_{ij}) + \lambda_1(\|\mathbf{w}\|_2^2 + ||\mathbf{a}||_2^2 + ||\mathbf{b}||_2^2) + \lambda_2 \left(\|\mathbf{U}\|_2^2 + \|\mathbf{V}\|_2^2\right)$ where $\mathrm{score}(i,j) = \mu + w_i + w_j + \mathbf{a}^T\mathbf{x}_i + \mathbf{b}^T\mathbf{y}_j + \mathbf{u}_i^T\mathbf{v}_j$.

  Note that $\mathcal{D}$ is the observation dataset, r_{ij} is the rating that user i gave to item j. $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, ...)$ denotes the users latent factors and $\mathbf{V} = (\mathbf{v}_1, \mathbf{v}_2, ...)$ denotes the item latent factors. The loss function $\mathcal{L}(\hat{y}, y)$ is $(\hat{y} - y)^2$. λ_1 denotes the linear_regularization parameter and λ_2 the regularization parameter, which hereafter we shall be referred to as "standard regularization."[1]

  We chose to use their default implementation of SGD, which included additional modifications to improve convergence over standard SGD, in order to speed up the training process, since we trained over a large number of different potential parameter values, which we describe further in the Model Training and Scorer section.

- **SVD**:

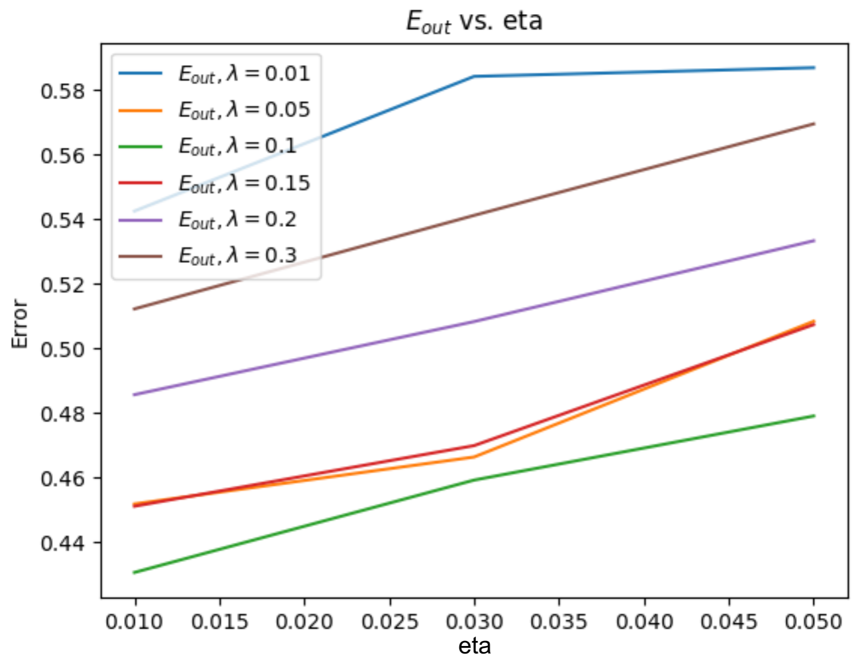
  For each of these models, after training, we applied SVD to the V matrix to get the two most important dimensions so we could plot them in 2D. In other words, we applied SVD to get $A\Sigma B^T$. After doing this, we performed a truncation $A_{1:2}^T V$ to get a matrix of the two most important latent factors for V (which represents the movies). At this point, we normalized and standardized $A_{1:2}^T V$ so that the plot would not be skewed. Note that any column in $A_{1:2}^T V$ represents one of the movies, so we simply picked out the columns we were interested in and plotted them.

3 Model Training and Scoring

- **Basic Matrix Factorization**:

  We were told for this project that the number of latent factors should be 20. This left the learning rate and regularization strength as the main factors to be determined through cross validation. Below is a plot of the results from doing this cross-validation. The metric used was mean squared error.

[1] https://turi.com/products/create/docs/generated/graphlab.recommender.factorization_recommender.FactorizationRecommender.html #graphlab.recommender.factorization_recommender.FactorizationRecommender



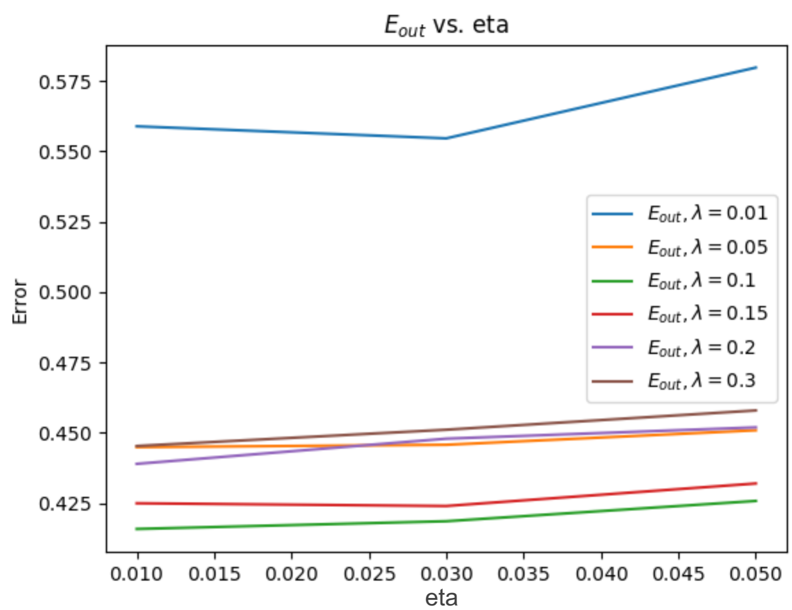
From this plot we determined an an optimal regularization constant of 0.1. Furthermore we see that an eta of 0.01 best minimizes the test error. Unfortunately, though, the lower learning rate also corresponded to a significantly longer training time so we decided that a good balance would be achieved between minimizing test error and training time by using eta = 0.03. When the regularization was too high the testing error was very high because the model was unable to fit the data at all. When the regularization was too low the test error increased because of of overfitting.

The plot indicates that training for more epochs would likely have increased the test error but this would have come at the cost of significantly increased training time.

Overall, using this model with the parameters we chose yielded a mean squared error on the test set of about 0.45.

- **<u>Matrix Factorization with Bias</u>**:

  Similar to the case without the bias terms, the parameters that needed to be determined were eta and the regularization strength. Below is a plot of the mean squared error of the model on the test set after being trained on the training set for different parameter values.



Interestingly, compared to the case without the bias, the learning rate makes much less of a difference for the training error. Furthermore, the spread in the different learning curves is much less linear than the case without the biases. We chose k to be 0.03 for this model for the same reason as the previous case. Regularization strength of 0.1 also minimized the test error for this model.

Overall, this model fit the data better than the model without biases. This was expected because the model with the biases has a better ability to capture the variability in the data; we are not significantly overfitting so increasing variance to lower bias is fine in this case. The test mean squared error was about 0.415. This was significantly lower than the .45 from the model without the biases.

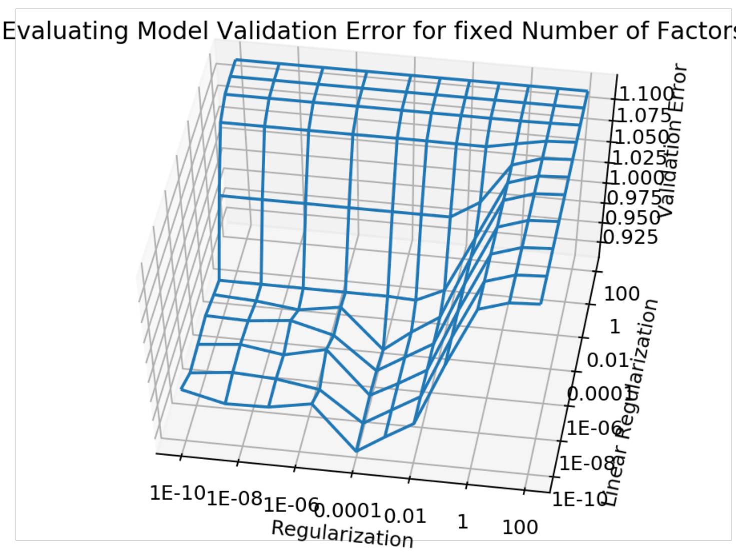
- **Matrix factorization using the third-party package GraphLab**:

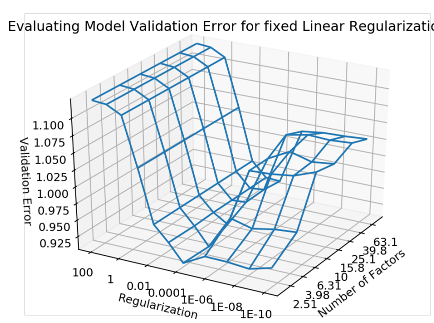
  To find the best model, we searched for the best validation-set error over the following space of parameters (code below is in Python):

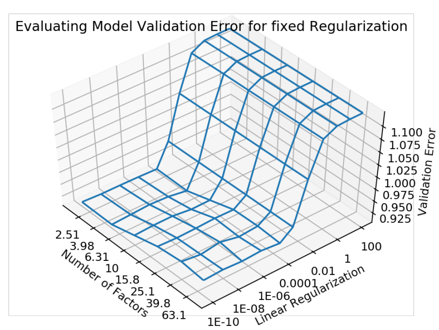
```
linear_regularization_values = [10 ** -10, 10 ** -8.5, 10 ** -7,
10 ** -5.5, 10 ** -4, 10 ** -3, 10 ** -2, 10 ** -1, 1, 10, 100]

# Standard regularization terms are those labelled "regularization" in the graphs.
regularization_values = [10 ** -10, 10 ** -8.5, 10 ** -7, 10 ** -5.5, 10 ** -4,
10 ** -3, 10 ** -2, 10 ** -1, 1, 10, 100]

number_of_factor_values = [2, 5, 10, 20, 40, 75]
```

  In a method analogous to SKLearns GridSearchCV, we iterated over all possible combinations of these three different parameters, trained the model using those parameters, and then evaluated the model on the test data set provided using root-mean-squared error. We then chose the model with the lowest RMSE, trained it on the full data set, and used SVD on this model to visualize it in two dimensions.

Below, we have plotted the test error of the model on the data set, using RMSE, over different combinations of two of the parameters. The fixed parameter in each case was chosen to be such that the optimal model used that parameter value; for example, when plotting standard regularization vs. linear regularization parameters, we fixed the number of latent factors to be 5, since the optimal parameters that minimized test error were with 5 latent factors, and a standard regularization and linear regularization term of 0.0001, which had a validation-set RMSE of 0.9095.







From these graphs, we see that high regularization values - i.e. when the parameter has a value of 1 or greater - both in the linear regularization and standard regularization terms, leads to higher validation error, which is almost certainly due to underfitting due to excess regularization. The high amount of regularization prevents the model from accurately fitting the data set. Also, for very low amounts of regularization, there appears to be an increase in validation error as well, although not nearly to the same scale as a very large amount of regularization. Likely, this is due to overfitting, as increased regularization often helps combat overfitting. Finally, we note that as the number of factors increased with a small amount (less than 0.0001) of standard regularization, there was a significant increase in the validation error. Dissimilarly, for a fixed standard regularization of 0.0001, there was little increase in validation error as the number of factors increased regardless of linear regularization value. Thus, it appears the standard regularization term has a much larger effect on preventing overfitting due to an increase in the number of factors than does linear regularization.

In comparison to the other two models, namely Basic Matrix Factorization, which had a RMSE of approximately .949, and Matrix Factorization with Bias Terms, which had a RMSE of approximately .911 (Note that these values were computed by taking the half-MSE values plotted in the above sections, multiplying them by two, then taking the square root), we see that the tuned GraphLab model was slightly more accurate on the test data set, with a RMSE of .9095, than the Matrix Factorization with Bias Terms, and significantly better than the Basic Matrix Factorization. However, the errors were very similar between the Matrix Factorization with Bias Terms, which is likely due to the fact that the optimization target was nearly the same, outside of the linear regularization parameter in the GraphLab model. Since we had this extra parameter to tune, we were able to better represent the data, or at least better fit the validation set. Obviously, we also have the additional bias terms over the non-bias model, which allowed for better fitting as well relative to that model.

4 Basic Visualizations

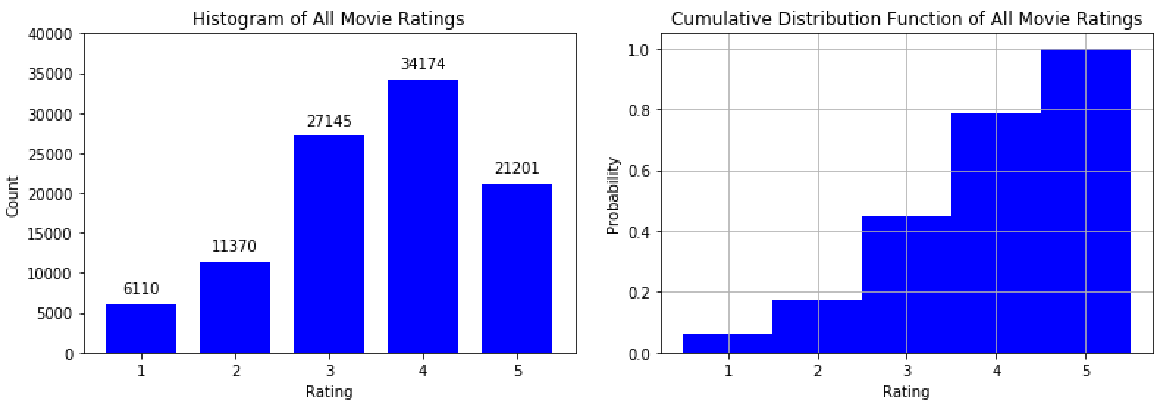
Despite their relative simplicity, the basic visualizations gave us much insight about the MovieLens dataset as a whole, as well as the genres we decided to investigate in detail.

- **MovieLens Dataset**:

  The MovieLens Dataset was collected through the MovieLens movie recommendation website and has been cleaned up through removal of data points from users who have rated fewer than 20 movies.[2] These facts are important to keep in mind when reasoning about the data.

  We visualized the entire set of ratings through a histogram of rating frequency and a cumulative distribution function of the same:

[2] https://www.kaggle.com/prajitdatta/movielens-100k-dataset



The cumulative distribution allows us to easily reason about trends in the ratings at a large scale. A negative skew is clearly visible – about 55% of the ratings are either 4 or 5. The data indicates a central tendency with about 73% of the ratings being from 2 to 4, but the tails are unequal. About 21% of movies are rated 5, while only about 6% are rated 1. The remaining ratings range from 2 to 4, though most are rated 3 or 4.

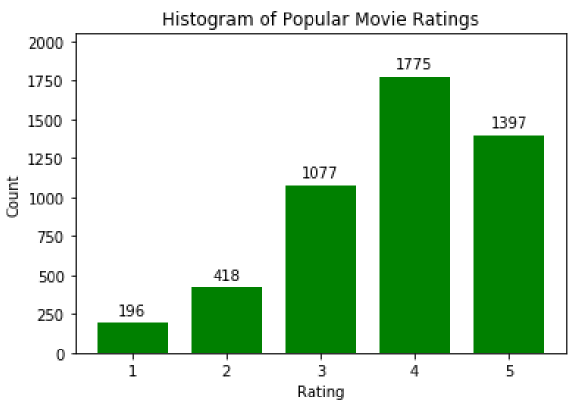
The histogram lets us examine more easily the frequencies of individual ratings. 3 and 4 are by far the most common ratings, making up about 61% of the ratings overall. 5 is a relatively common rating, comprising about 21% of the data. 2 is relatively uncommon at about 11%, and 1 is very uncommon at about 6%.

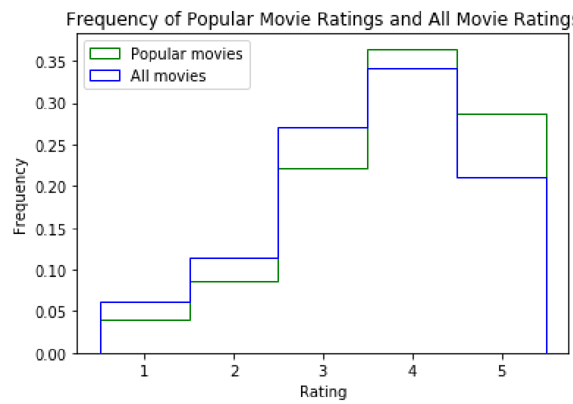
Overall, ratings tend to be mostly positive, with reviews unlikely to rate a movie below 3 and fairly likely to rate it at 4 or 5. Another competing trend is the datas central tendency; ratings of 5 are relatively unlikely given the datas negative skew, and ratings of 1 are very unlikely.

One possible explanation for the observed skew is that viewers are simply more likely to rate movies highly. However, sources of potential bias must be considered. Because the data is collected from a movie recommendation site, users are more likely to rate movies they enjoy because this is what the site is attempting to recommend. So, movies that a viewer strongly dislikes (potential ratings of 1) are very unlikely. In addition, viewers are drawn from the pool of people who use this service, and the data is further refined by removing users who have rated fewer than 20 movies, who perhaps stopped using the service after it failed to recommend movies they liked. These factors combine to form a set of viewers who perhaps enjoy movies in general more than is typical, and in particular those who enjoy movies recommended by this service. Therefore, conclusions drawn from this data about the likelihood of a person liking a given movie in general must be taken with some caveats.

- **Popular Movies**:

  We visualized the distribution of ratings of the ten movies which received the most ratings with a histogram of the raw ratings and a normalized histogram comparing this distribution to that of the dataset as a whole.



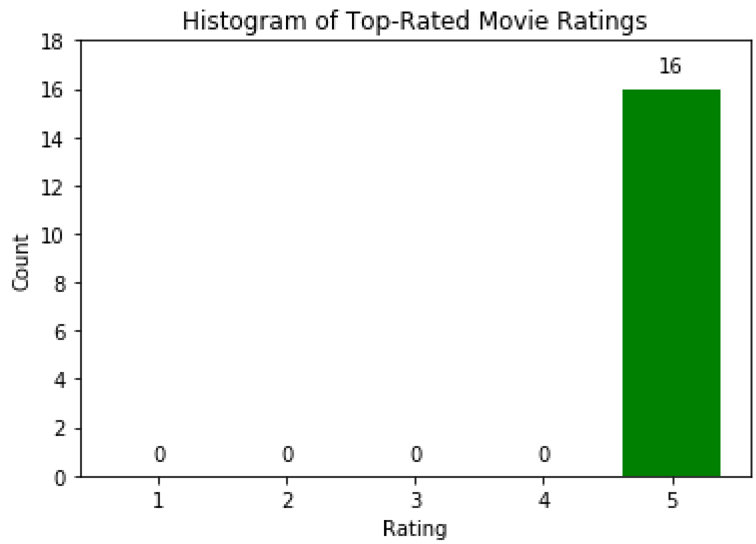


There seems to be a further negative skew relative to the dataset as a whole, with many more reviews of 5, some more of 4, and fewer other reviews. About 65% of ratings of these movies are either 4 or 5. The reduction in low rating frequency was much more pronounced with ratings of 3, with 1-star and 2-star ratings seeing smaller reductions. The jump in 5-star ratings was likewise much more pronounced than the increase in frequency of 4-star ratings. Overall, the proportion of ratings that are 1 or 5 increased from about 27% to about 33%. Therefore, another trend present in this data is an increase of ratings in the tails, with 1- and 5-star ratings becoming much more common for popular movies and, given the increased negative skew, ratings in the center becoming less common.

It is not surprising that the ten most popular movies tend to get higher reviews. However, the increase in the number of ratings in the tails is interesting. It appears that popular movies tend to be more polarizing, in that a given reviewer is more likely to either love the movie or hate it. However, we are judging a movies popularity by its review count, which might introduce biases of its own. Potentially, some other metric, such as view count, would be a better indication of popularity. In addition, the dataset for popular movie ratings is small, with only ten movies being considered. Therefore, though the most obvious trends in the data (increased average rating and spreading towards the tails) are likely to hold for other very popular movies, the effects scale and generality cannot be determined.

- **Highest-Rated Movies**:

  We visualized the distribution of ratings of the ten movies which were rated highly on average through a simple histogram.



With only 16 total reviews making up this dataset, each movie present was extremely unpopular, with only 1 to 3 reviews. Therefore, the high average ratings of these movies are very likely by chance. Owing to their low numbers individually, each movie was able to obtain only 5-star ratings, whereas a movie with more reviews would very likely get at least one rating of less than 5. For example, among these movies is Santa with Muscles, a movie with very low ratings on other sites such as imdb.com[3]. In addition, given the trend of higher ratings with more reviews, it is likely that these movies, if they received more reviews, would likely have much lower average ratings. Though these movies are rated on average much higher than the popular movies, this rating appears almost totally by variance.

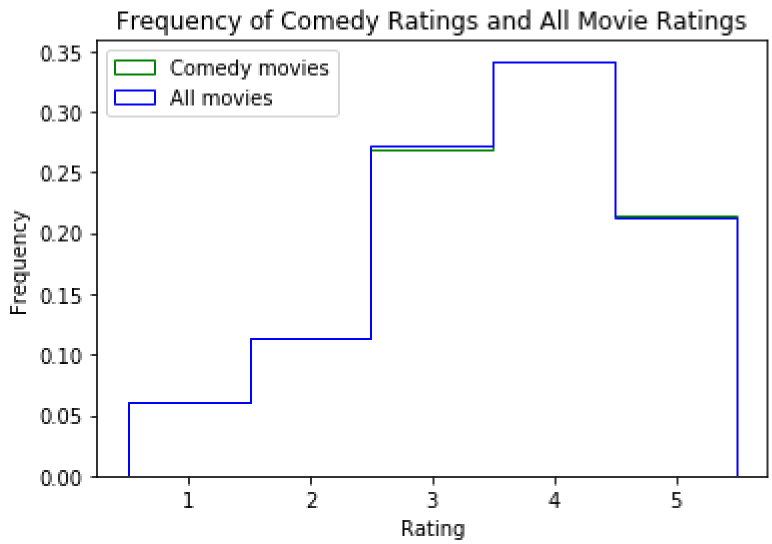
- **A Brief Note on the Generation of Genre-Specific Rating Plots**:

  Note that the plots presented for the three genres we investigated are different than those submitted for our "visualizations" post. This is because our code generating these plots had a bug which we unfortunately only discovered between submitting the visualizations and writing the report. The fix is small, but it changes the results and the conclusions we can draw from them, and since we value complete and correct analysis, we present the corrected graphs and analysis here.

- **Comedies**:

  We visualized the distribution of ratings of comedies by comparing its normalized rating histogram to that of the dataset as a whole.

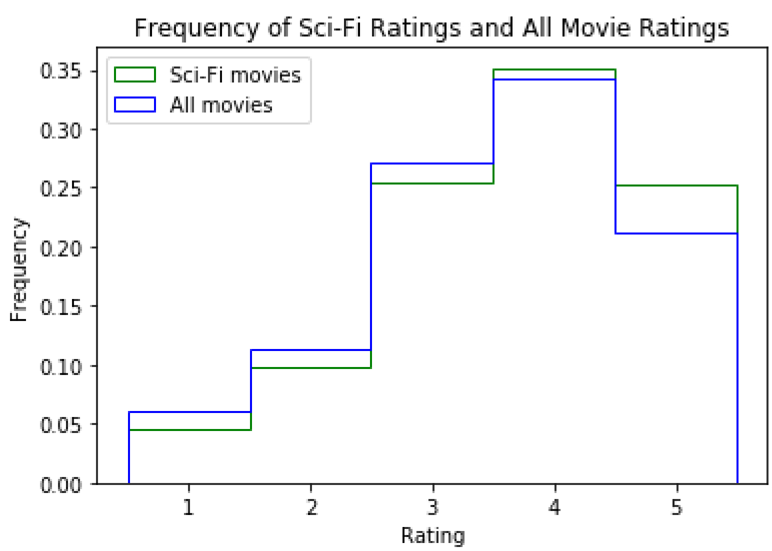
[3]http://www.imdb.com/title/tt0117550/



The ratings of comedies very closely tracked the ratings of other kinds of movies. For every individual rating, the frequency of that rating for comedies was very likely equal within variance to the frequency of that rating for other movies. One potential reason for this is that comedies make up a large portion (about 32%) of total ratings. Another possibility is that, because many movies feature an element of comedy, movies that are tagged "comedy" on MovieLens are not far from average.

- **Sci-Fi Movies**:

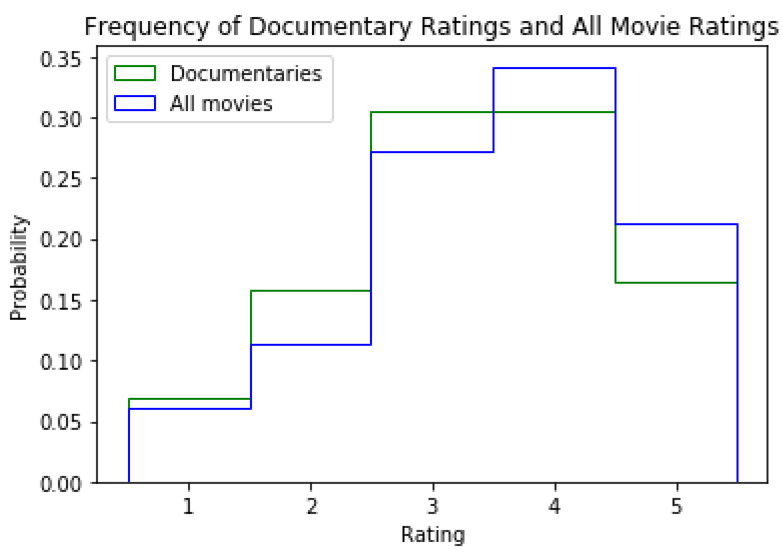
  We visualized the distribution of sci-fi movie ratings in the same manner as with comedies.



Sci-fi movies were rated higher on average than most movies, with a significant increase in the frequency of 5-star ratings, a very small increase in the frequency of 4-star ratings, and relatively small decreases in the frequencies of other ratings. Sci-fi movies made up about 10% of the movies considered, so the skew towards positive reviews is likely significant. The fact that most of the change in review frequency is towards 5-star reviews is interesting; it suggests that a viewer recommended a sci-fi movie is more likely to love it than simply like it relative to another movie.

- **Documentaries**:

  Again, we visualized the rating distribution of documentaries in the same way.



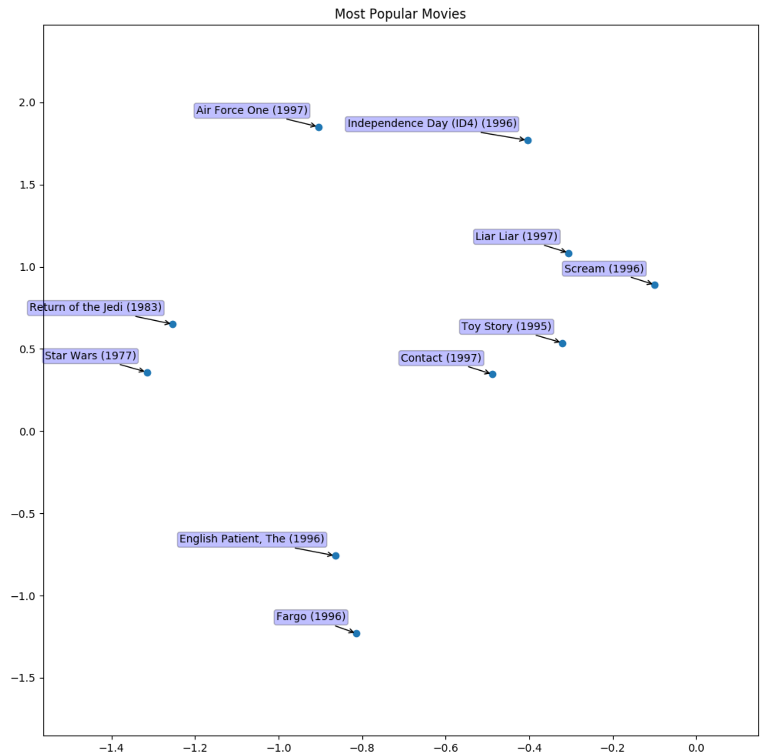
Documentaries, which made up only about 2% of the dataset, were rated on average much lower than other types of movies, with many fewer 4- and 5-star reviews (about 5% fewer), many more 2- and 3-star reviews (again, about 5% more), and very slightly more 1-star reviews. While the sample size of documentaries was relatively small, this strong trend suggests that viewers recommended a documentary are less likely to enjoy it than usual. However, documentaries are not hated significantly more than average; especially given the strong increase in negative rating frequency, the increase in 1-star ratings is very small.

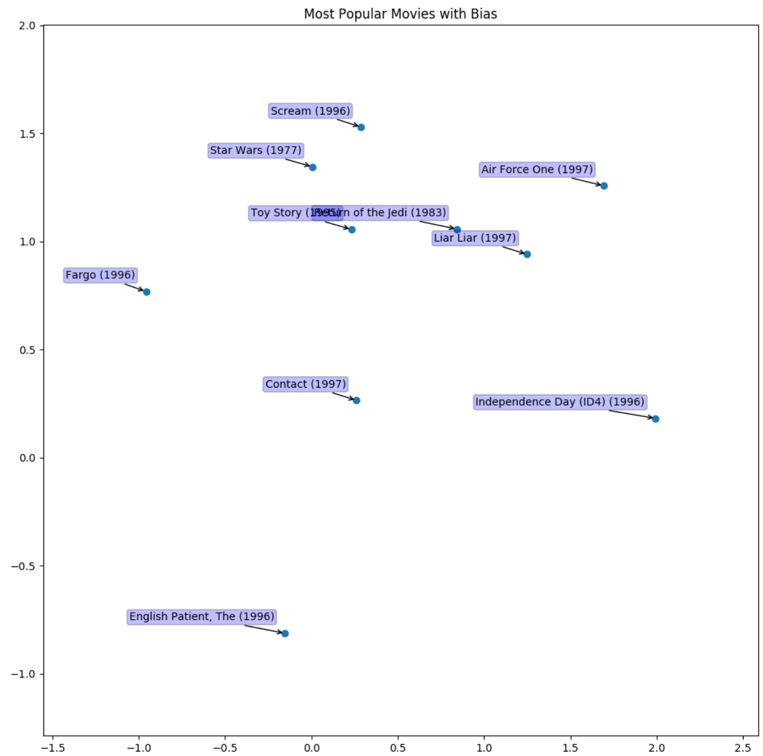
5 Matrix Factorization

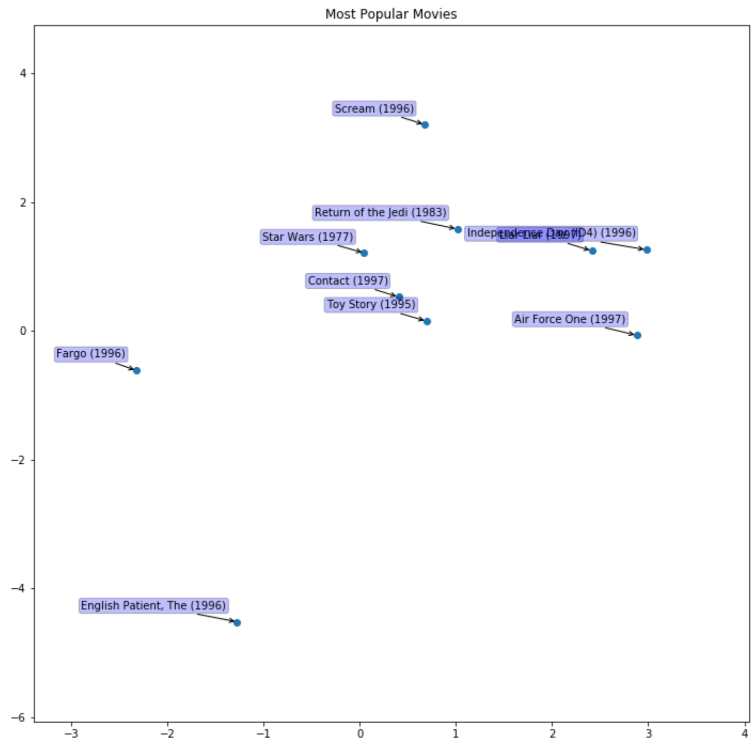
The 6 different plots that we made for each of the different matrix factorization models are a plot of the most popular movies, the highest rated movies, Star Wars and Star Trek movies, documentaries, comedies, and Sci-Fi movies.

- **Most Popular Movies**:

  We visualized the latent factors of the 10 most popular movies in the dataset. Below are the plots for each of the different models.


Most Popular Movies
Air Force One (1997)
Independence Day (ID4) (1996)
Liar Liar (1997)
Scream (1996)
Return of the Jedi (1983)
Toy Story (1995)
Star Wars (1977)
Contact (1997)
English Patient, The (1996)
Fargo (1996)
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−1.4
−1.2
−1.0
−0.8
−0.6
−0.4
−0.2
0.0


Most Popular Movies with Bias
Scream (1996)
Star Wars (1977)
Air Force One (1997)
Toy Story (1995)
Return of the Jedi (1983)
Liar Liar (1997)
Fargo (1996)
Contact (1997)
Independence Day (ID4) (1996)
English Patient, The (1996)



The first plot is the base model trained without bias, the second plot is the base model with bias included, and the bottom plot is from the third party package graphlab.

– Base model: The plot is segmented very clearly. The two movies at the bottom are both dramas. At the top, both of the movies are war-centric action and adventure movies. Towards the left are two Star Wars movies which have very similar latent factors. The group of 4 movies centered vertically and to the right are very non-uniform; scream and Contact are both horror movies while Toy Story is an animated movie for children and Liar Liar is a pg-13 comedy. It seems that Toy Story is a significant outlier which could be because it is ranked by adults and not children who are its primary audience. It seems that the vertical axis could be the action in the movie whereas the horizontal axis could be the content maturity level, with the exception of Toy Story which is an outlier.

– Base Model with Biases: Fargo and English Patient are both dramas and are on the left side of the figure. Towards the right side are war movies and action and adventure movies. In the middle, the movies seem to be less clearly distinguishable. The Star Wars movies are grouped together but Scream and Contact, two horror movies, are not. It seems that the horizontal axis could correlate to the amount of action in the movie.

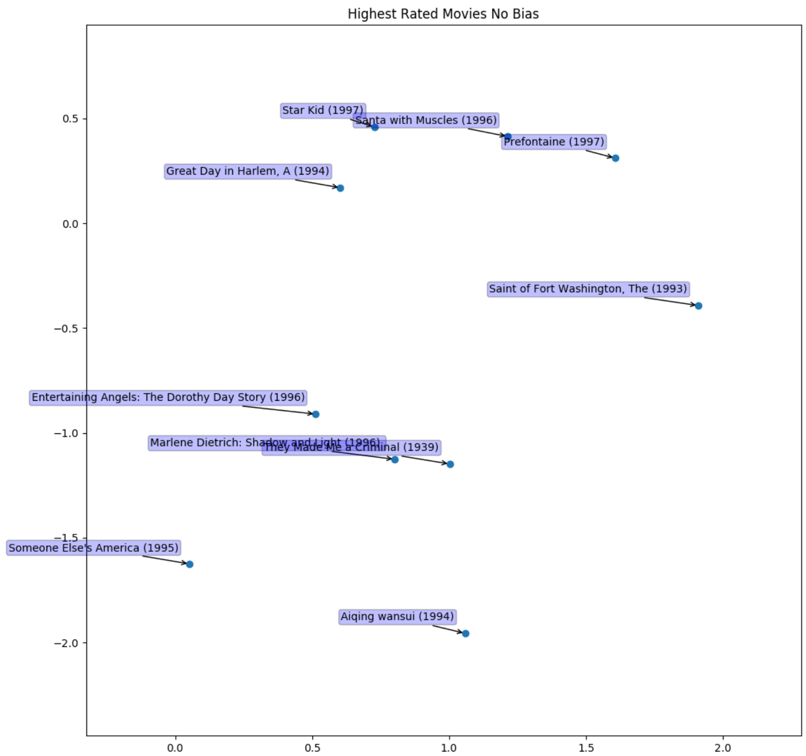
– Graphlab: On the right side of the figure are the action-adventure-war movies: Air Force One and Independence Day. Fargo and English Patient are both dramas and are on the left side of the figure. In the middle the movies seem to be less clearly distinguishable, with Star Wars, Contact, and Toy Story, all quite different movies, being grouped in the same general area. Additionally of interest, we see that the two Star Wars movies are grouped together but Scream and Contact, two horror movies, are not. Here the horizontal axis seems correlated with the amount of action in the movie.

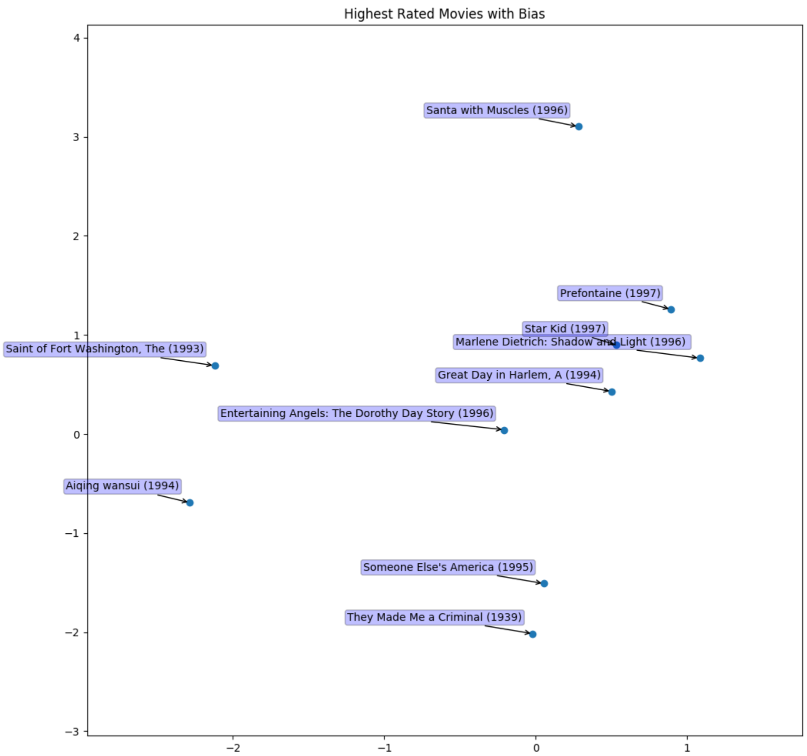
Overall, the plots produced by these different models have many similarities. The models all produce similar groupings for the dramas, Star Wars movies, and action, adventure, or war movies. The models also all have trouble classifying Toy Story which makes sense since it is a children's movie but is being rated by adults. Furthermore, the models do not consistently classify the horror movies correctly. Interestingly, the models also seem to produce a similar clustering towards the center of the plots with a less clear division between the different movies.

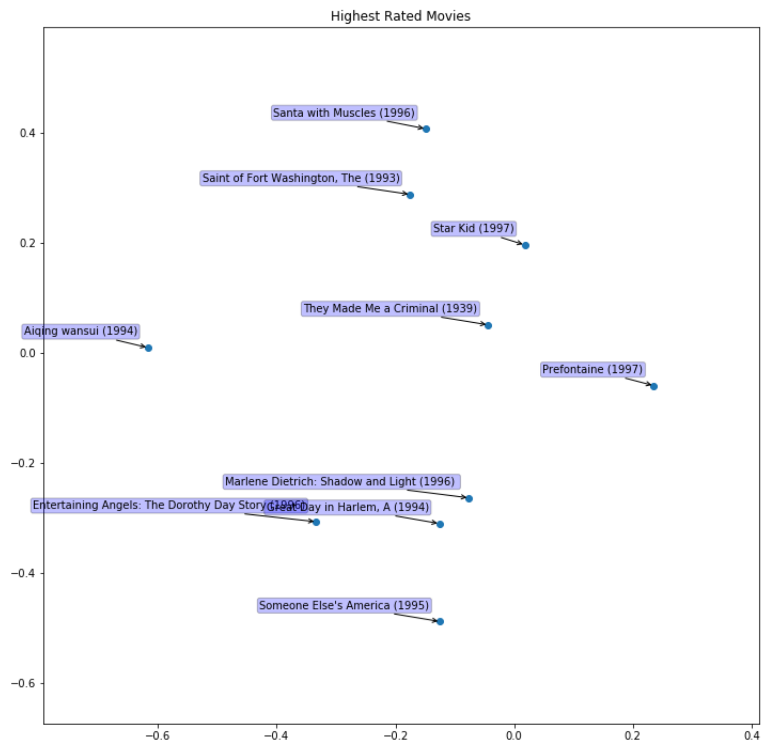
The two models that produce the most similar results are the graphlab third-party package and the base model with the biases added. For many of the movies, such as Fargo, English Patient, Scream, and Star Wars, the latent factors are almost exactly the same. Compared to the other models, the base model without biases produced a clustering more consistent with the similarities between the movies. The model without biases performed as expected, with greater variability between the points that were in the clusters of the first model.

- **Highest-Rated Movies**:

  We visualized the latent factors of the 10 highest-rated movies in the dataset. Below are the plots for each of the different models.


Highest Rated Movies No Bias
Star Kid (1997)
Santa with Muscles (1996)
Prefontaine (1997)
Great Day in Harlem, A (1994)
Saint of Fort Washington, The (1993)
Entertaining Angels: The Dorothy Day Story (1996)
Marlene Dietrich: Shadow and Light (1996)
They Made Me a Criminal (1939)
Someone Else's America (1995)
Aiqing wansui (1994)
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0.0
0.5
1.0
1.5
2.0


Highest Rated Movies with Bias
Santa with Muscles (1996)
Prefontaine (1997)
Star Kid (1997)
Marlene Dietrich: Shadow and Light (1996)
Saint of Fort Washington, The (1993)
Great Day in Harlem, A (1994)
Entertaining Angels: The Dorothy Day Story (1996)
Aiqing wansui (1994)
Someone Else's America (1995)
They Made Me a Criminal (1939)
4
3
2
1
0
−1
−2
−3
−2
−1
0
1



The first plot is the base model trained without bias, the second model is the base model with bias included, and the bottom plot is from the third party package graphlab.

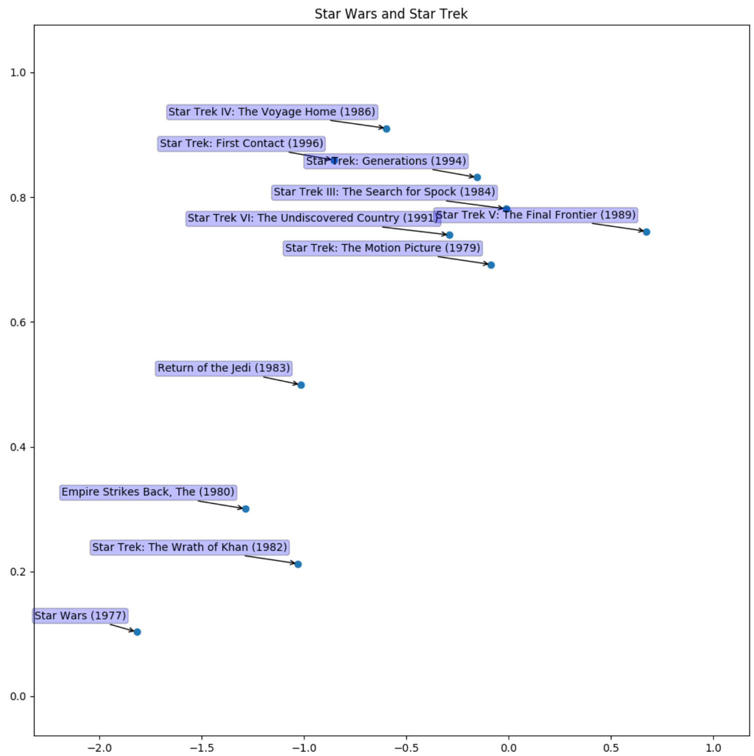
In each of the plots above, there was not much information to be ascertained. As was mentioned in the basic visualizations section, the reason that these movies have such high ratings is variance. Since these movies have so few reviews it is difficult for the models to correctly classify them.

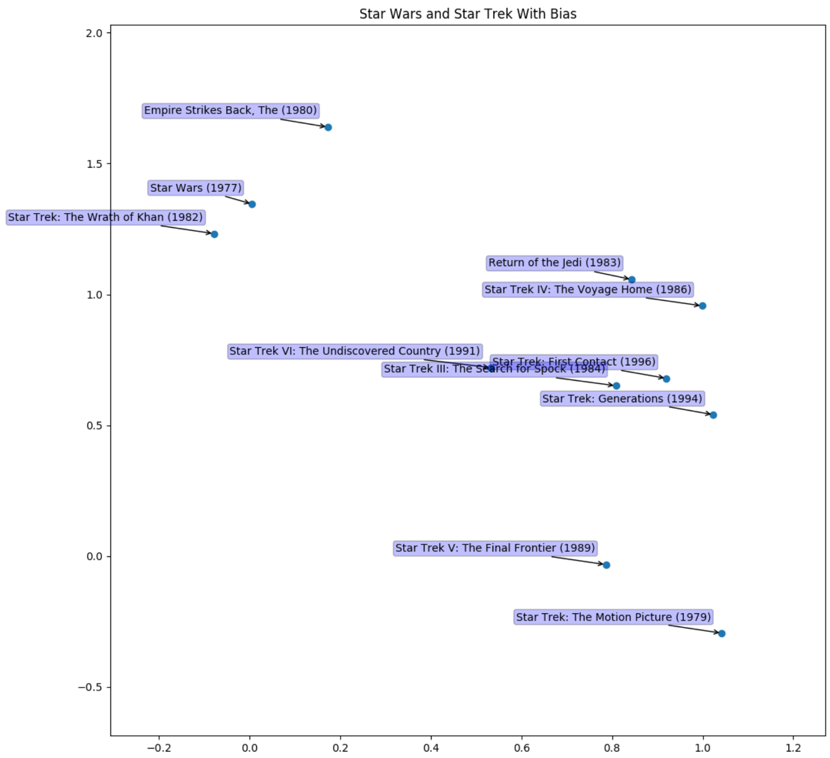
In fact, the plots from the different models are completely different from each other. In other words, since there is so little data the matrix factorization becomes highly variable between runs. Overall, without good data, even the better models are unable to make good headway.

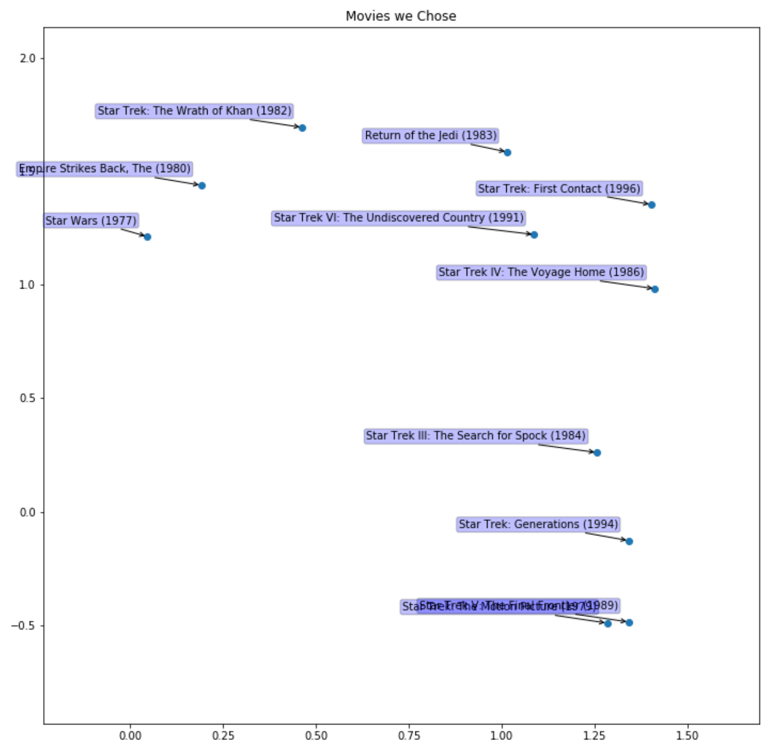
There isnt much to say when comparing this to the most popular movies, since these plots are completely random and of different movies.

- **Star Wars and Star Trek Movies**:

  We visualized the latent factors of the Star Wars and Star Trek movies in the dataset. Below are the plots for each of the different models.


Star Wars and Star Trek
Star Trek IV: The Voyage Home (1986)
Star Trek: First Contact (1996)
Star Trek: Generations (1994)
Star Trek III: The Search for Spock (1984)
Star Trek VI: The Undiscovered Country (1991)
Star Trek V: The Final Frontier (1989)
Star Trek: The Motion Picture (1979)
Return of the Jedi (1983)
Empire Strikes Back, The (1980)
Star Trek: The Wrath of Khan (1982)
Star Wars (1977)
1.0
0.8
0.6
0.4
0.2
0.0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0


Star Wars and Star Trek With Bias
Empire Strikes Back, The (1980)
Star Wars (1977)
Star Trek: The Wrath of Khan (1982)
Return of the Jedi (1983)
Star Trek IV: The Voyage Home (1986)
Star Trek VI: The Undiscovered Country (1991)
Star Trek: First Contact (1996)
Star Trek III: The Search for Spock (1984)
Star Trek: Generations (1994)
Star Trek V: The Final Frontier (1989)
Star Trek: The Motion Picture (1979)
2.0
1.5
1.0
0.5
0.0
−0.5
−0.2
0.0
0.2
0.4
0.6
0.8
1.0
1.2



The first plot is the base model trained without bias, the second plot is the base model with bias included, and the third plot is from the third party package graphlab.

- Base Model without Bias: The Star Trek movies are almost entirely distinct from the Star Wars movies, demonstrating that our model was able to differentiate the two similar series. The Star Trek movies are much more tightly clustered than the Star Wars movies, potentially indicating that Star Wars is broader in scope than Star Trek, which tends to more consistent in content. This agrees with our experience that Star Wars tends to appeal to more types of people but attract less devoted fans than does Star Trek. Another interesting element of this visualization is the substantial deviation of Star Trek: The Wrath of Khan from the other Star Trek movies, which are otherwise tightly clustered. One possible explanation for this is that The Wrath of Khan is often considered a significant departure from its predecessor, the original Star Trek.
- Base Model with Bias: The groupings in this plot are more spread out than those without the bias term. Unlike the previous case where all the Star Wars movies and Star Trek: The Wrath of Khan were grouped together now only two of the Star Wars movies and Star Trek: The Wrath of Khan were grouped together. It seems to be that the movies on the left are darker than the movies on

the right. On the vertical axis I dont know enough about the different Star Trek movies to make any certain conclusions.

– Graphlab: First, we note that the Star Wars movies and Star Trek: The Wrath of Khan are all grouped together in the upper left, a pattern we have also seen from other analyses. We also note that all the star trek movies outside of Star Trek: The Wrath of Khan are at relatively the same X value, meaning that there is likely some core component of star trek that this SVD has projected along the X-Axis. Unfortunately, on the vertical axis I dont personally know enough about the different Star Trek movies to make any conclusions about the relatively even distribution of the various Star Trek movies throughout the Y-Axis.

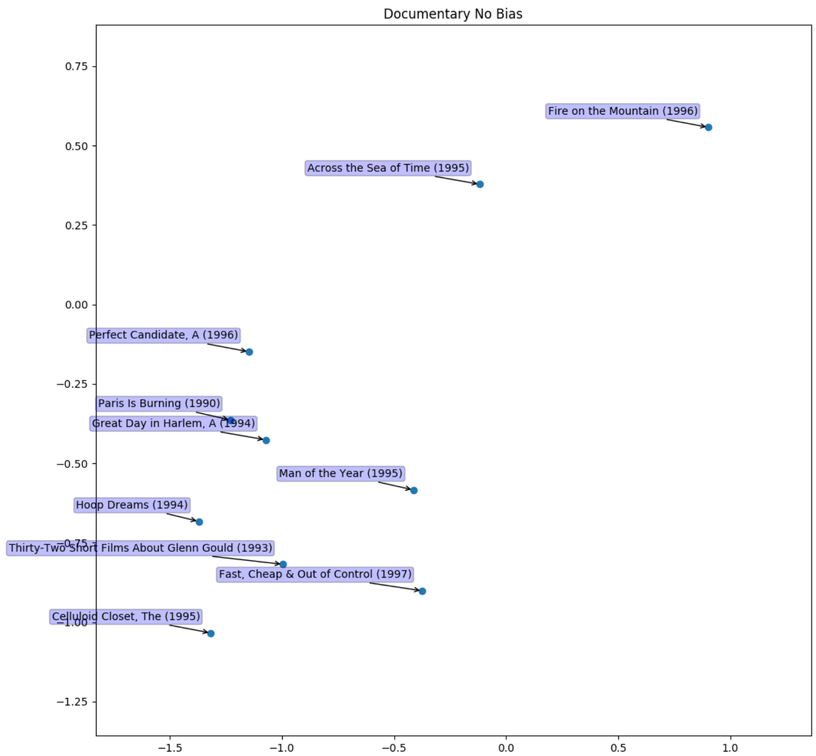
These plots all showed similar behavior, though there was some rotation about the origin in the results. All of the plots grouped Episode 4, Episode 5, and Star Trek: The Wrath of Khan. Furthermore, in all of the plots, Return of the Jedi was close to the cluster of the aforementioned movies.

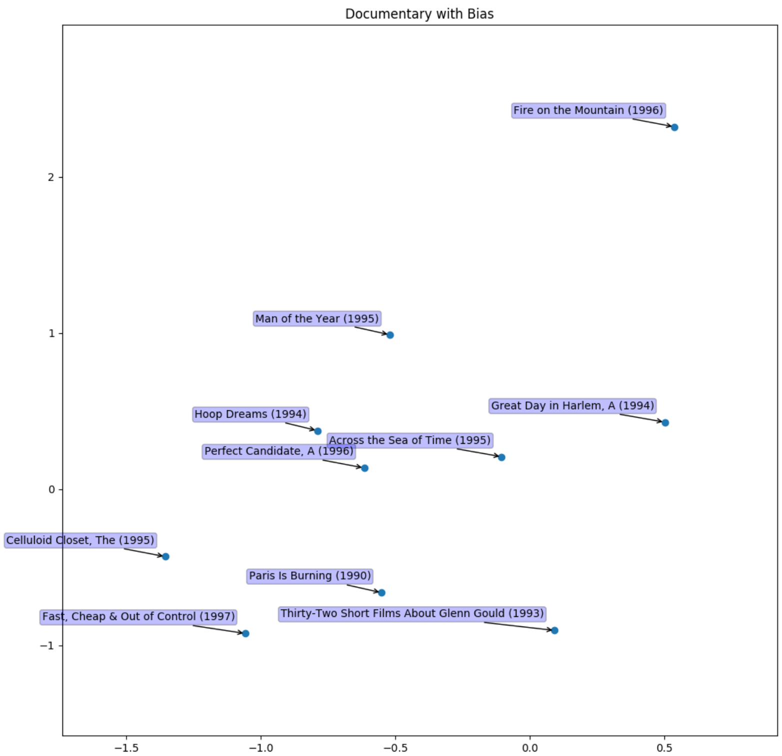
Within the Star Trek movies, there seems to be less of a clear ordering in all of the plots. In other words, if you look at the position of the different Star Trek movies relative to the cluster of the Star Wars movies there is no consistent pattern between the different models.

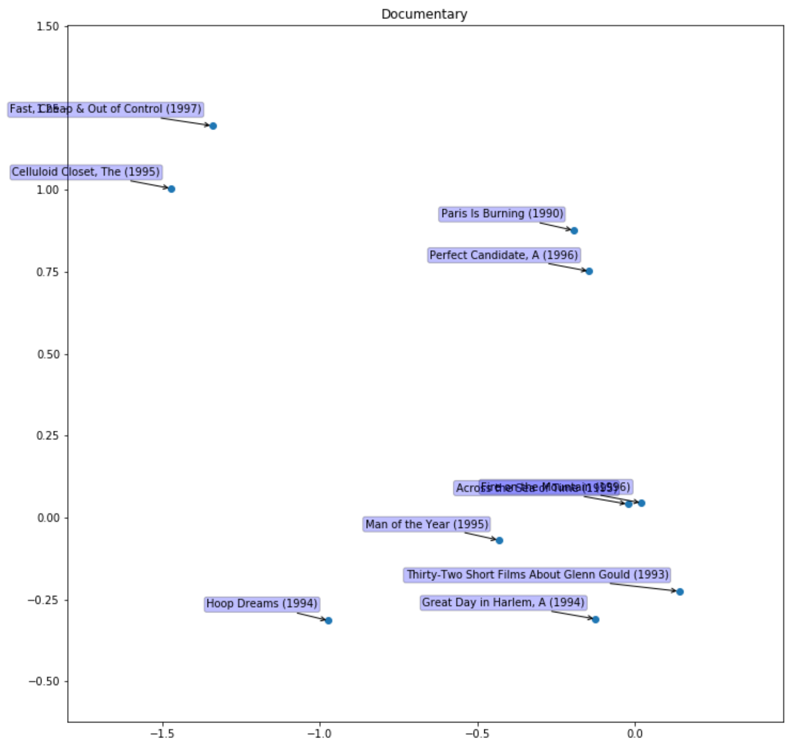
The base model that includes the bias is better at separating the Star Trek movies from each other, as was expected, because it is able to learn more variability.

- **Documentaries**:

  We visualized the latent factors of a few select Documentaries from the dataset in two-dimensions using SVD, the graphs of which are below for each of the different models.


Documentary No Bias
Fire on the Mountain (1996)
Across the Sea of Time (1995)
Perfect Candidate, A (1996)
Paris Is Burning (1990)
Great Day in Harlem, A (1994)
Man of the Year (1995)
Hoop Dreams (1994)
Thirty-Two Short Films About Glenn Gould (1993)
Fast, Cheap & Out of Control (1997)
Celluloid Closet, The (1995)
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
−1.5
−1.0
−0.5
0.0
0.5
1.0


Documentary with Bias
Fire on the Mountain (1996)
Man of the Year (1995)
Hoop Dreams (1994)
Great Day in Harlem, A (1994)
Across the Sea of Time (1995)
Perfect Candidate, A (1996)
Celluloid Closet, The (1995)
Paris Is Burning (1990)
Fast, Cheap & Out of Control (1997)
Thirty-Two Short Films About Glenn Gould (1993)
2
1
0
−1
−1.5
−1.0
−0.5
0.0
0.5



The top plot is the base model trained without bias, the second plot is the base model with bias included, and the bottom plot is from the third party package graphlab.

– Base Model without Bias: It seems that as the vertical position is slightly correlated with whether the movie relates to politics and government. The biggest outlier of all of the data points seems to be Fire on the Mountain. One possible reason for this is that this documentary is much more like an action or adventure movie than the rest of the movies shown.

– Base Model with Bias: In this plot the clearest separation seems to be the topic of the documentaries. All 4 of the bottom documentaries are very "artsy" and counter-culture whereas the top 6 movies are about political and socioeconomic characterizations. Thus it seems like an increase in the vertical latent factor corresponds to how mainstream the topic of the documentary is.

– Graphlab: In this figure, we see that Thirty-Two Short Films About Glenn Gould and A Great Day in Harlem are clustered closely in the bottom right, which could be related to the fact that they are both documentaries about musicians, namely a pianist and a group of jazz musicians, respectively. Unfortunately, the connections between other close groupings, such as Paris is

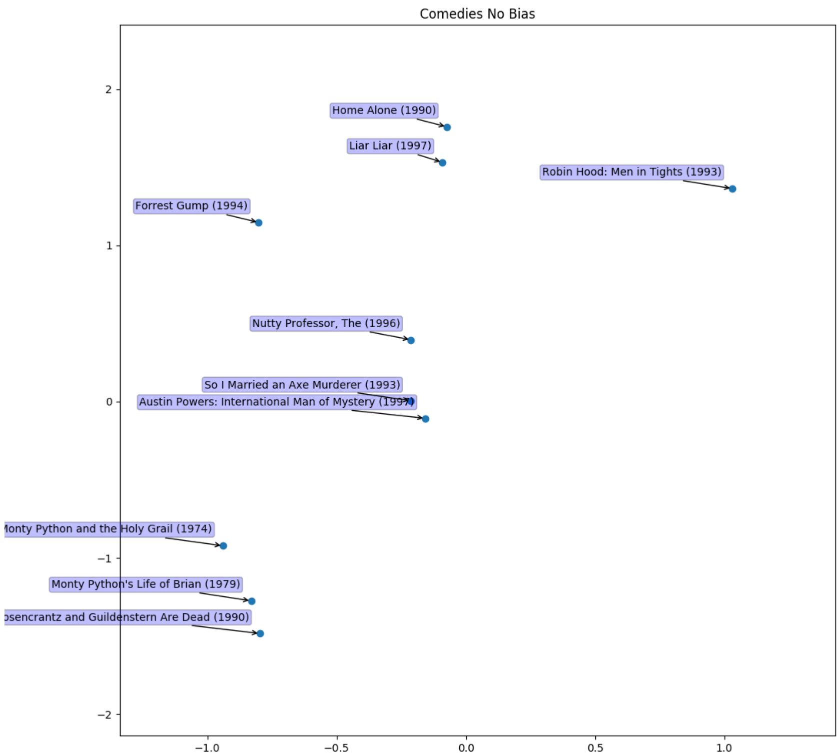
Burning and The Perfect Candidate or Across the Sea of Time and Fire on the Mountain are not as imminently obvious, and no general trends readily emerge.

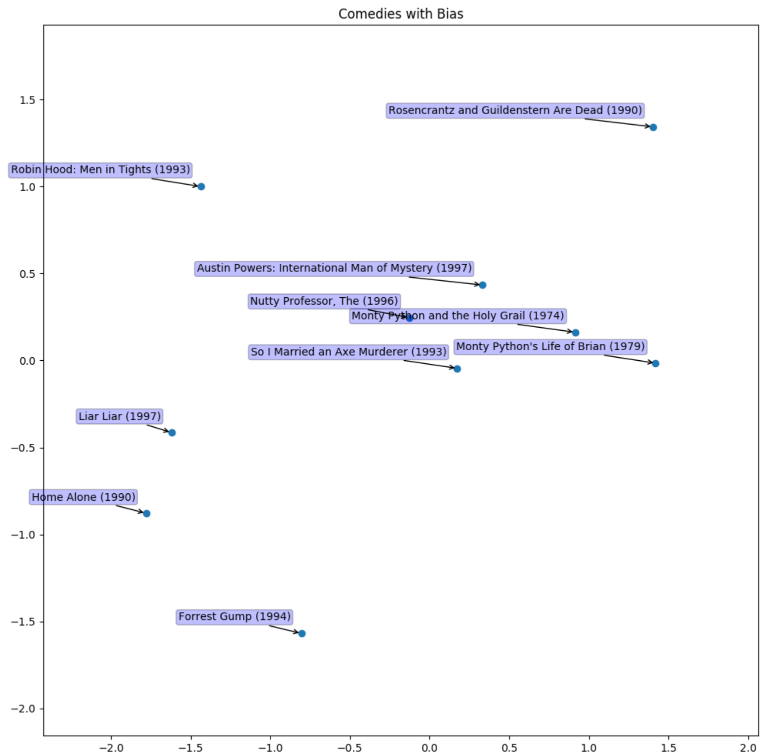
These graphs were all relatively different for the different models, although there are a few interesting similarities between the three. Namely, we see that The Celluloid Closet and Fast Cheap & Out of Control were closely grouped in all three. Although at first glance these movies do not appear to be overly similar - one is about four men with various relatively unrelated jobs, whereas the other is about hollywood depictions of hollywood - there might be some hidden relation that is being picked up upon by the various models. However, it is also possible that it is just a quirk of the data set, or is a result of overfitting in some regard.

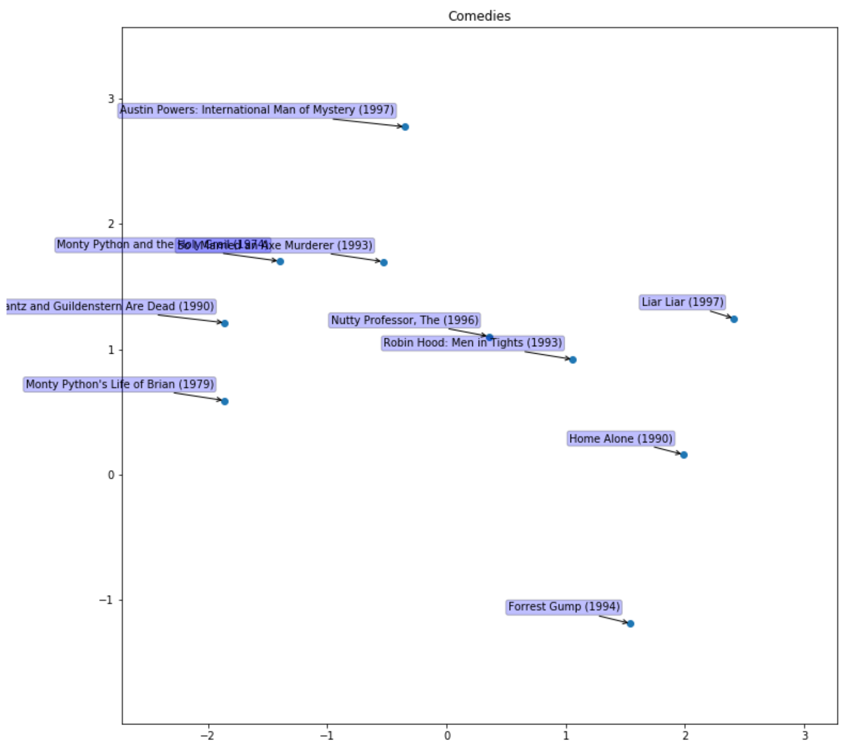
Additionally, we see that in terms of easily-noticed patterns, the Base Model With Bias appeared to have been the best, with its vertical latent factor corresponds to how mainstream the topic of the documentary is, whereas the other graphs did not have such clear patterns.

- **Comedies**:

  We visualized the latent factors of a few select Comedies from the dataset in two-dimensions using SVD, the graphs of which are below for each of the different models.


Comedies No Bias
Home Alone (1990)
Liar Liar (1997)
Robin Hood: Men in Tights (1993)
Forrest Gump (1994)
Nutty Professor, The (1996)
So I Married an Axe Murderer (1993)
Austin Powers: International Man of Mystery (1997)
Monty Python and the Holy Grail (1974)
Monty Python's Life of Brian (1979)
Rosencrantz and Guildenstern Are Dead (1990)
2
1
0
−1
−2
−1.0
−0.5
0.0
0.5
1.0


Comedies with Bias
Rosencrantz and Guildenstern Are Dead (1990)
Robin Hood: Men in Tights (1993)
Austin Powers: International Man of Mystery (1997)
Nutty Professor, The (1996)
Monty Python and the Holy Grail (1974)
So I Married an Axe Murderer (1993)
Monty Python's Life of Brian (1979)
Liar Liar (1997)
Home Alone (1990)
Forrest Gump (1994)
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.0
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0



The top plot is the base model trained without bias, the second plot is the base model with bias included, and the bottom plot is from the third party package graphlab.

- Base Model without Bias: The division in this plot are pretty clear. Both the Monty Python movies and Rosencrantz and Guildenstern Are Dead have similar latent factors. This makes sense since they both have a similar historical setting and absurd sense of humor. Nutty Professor, So I Married an Axe Murderer, and Austin Powers: International Man of Mystery all have a more raunchy sense of humor. Home Alone and Liar Liar have similar latent factors which makes sense since they both appeal to children. Overall it seems like the vertical axis is correlated with how intellectual and mature the sense of humor is and the horizontal axis is correlated with the maturity level required for the content.
- Base Model with Bias: In the bottom left corner of this plot are more child-oriented comedies like Home Alone, Liar Liar, and Forrest Gump. Towards the middle are movies that have either a more intellectual or raunchy sense of humor. The main outlier on this plot is Rosencrantz and Guildenstern are Dead which makes more sense to be clustered with movies like Monty Python. Perhaps one reason for this is that this movie is based off of side characters in Shakespeare and

as a result requires more prior knowledge to fully understand the humor. It seems as if the vertical latent dimension corresponds to how intellectual the humor is while the horizontal axis corresponds to the maturity level of the humor.

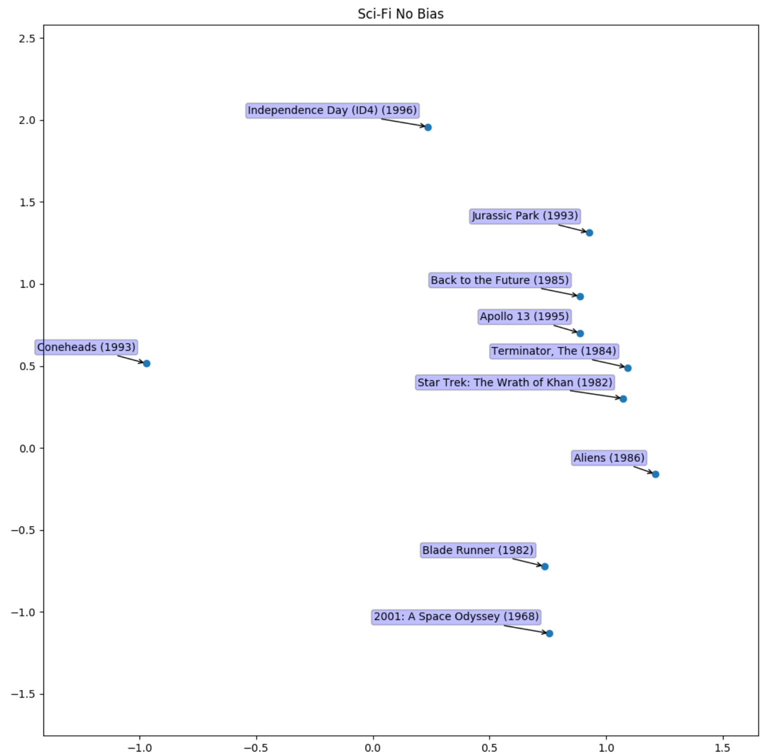
- Graphlab: First, we note that the two Monty Python movies are located relatively-close together, as expected, and further that Rosencrantz and Guildenstern are Dead is between them. Also, in the middle-right of the graph, we see several sci-fi/fantasy comedies together: Liar Liar, The Nutty Professor, and Robin Hood, Men In Tights. Thus, although not all similar movies were clustered together, e.g. Forrest Gump and Rosencrantz and Guildenstern are Dead, which are both comedy-dramas, many clusterings that reflect the movie type are present.

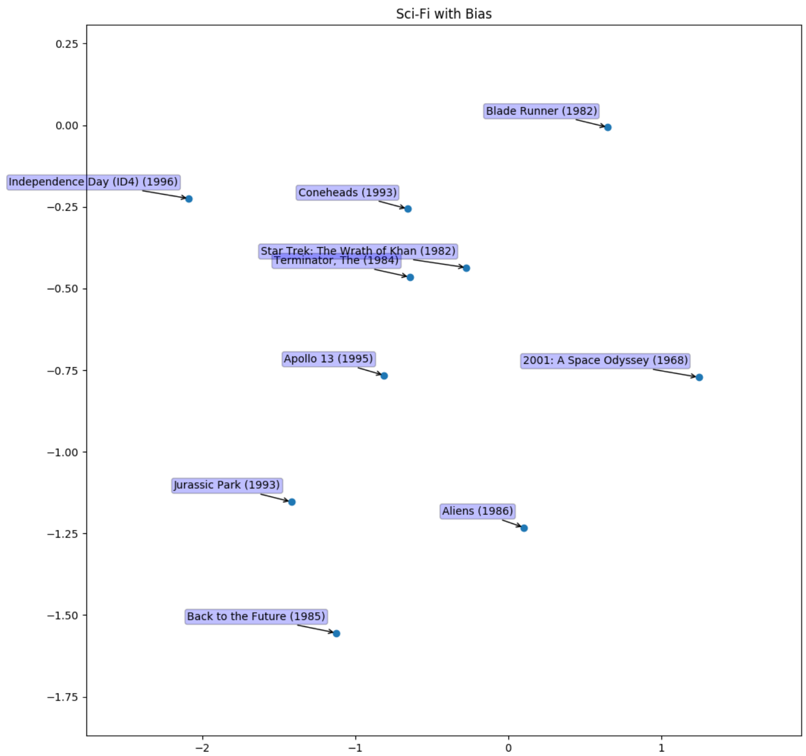
In all three models, we saw a strong correlation between the Monty Python movies, and in two of the models also Rosencrantz and Guildenstern are Dead, which was to be expected from the movies qualities. Additionally, Liar Liar and Home Alone were relatively close to each other in each graph, although they were closest in the No-Bias graph, which again is a reasonable prediction.

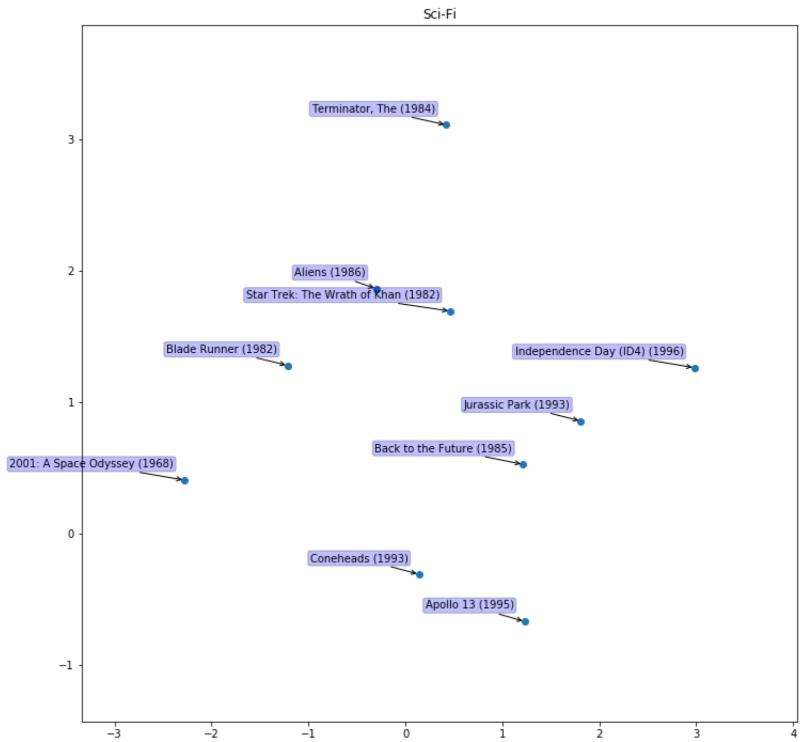
Although the rest of the groupings between graphs had several differences, there were reasonable trends in each, leading us to believe that the graphs simply modeled different latent factors that all reasonably approximated the movies, but in different manners. This is one of the expected difficulties of projecting high-dimensional data down to 2-dimensions.

- **Sci-Fi Movies**:

  We visualized the latent factors of a few select Sci-FI movies from the dataset in two-dimensions using SVD, the graphs of which are below for each of the different models.


Sci-Fi No Bias
Independence Day (ID4) (1996)
Jurassic Park (1993)
Back to the Future (1985)
Apollo 13 (1995)
Coneheads (1993)
Terminator, The (1984)
Star Trek: The Wrath of Khan (1982)
Aliens (1986)
Blade Runner (1982)
2001: A Space Odyssey (1968)


Sci-Fi with Bias
Blade Runner (1982)
Independence Day (ID4) (1996)
Coneheads (1993)
Star Trek: The Wrath of Khan (1982)
Terminator, The (1984)
Apollo 13 (1995)
2001: A Space Odyssey (1968)
Jurassic Park (1993)
Aliens (1986)
Back to the Future (1985)
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.25
−1.50
−1.75
−2
−1
0
1



The top plot is the base model trained without bias, the second plot is the base model with bias included, and the bottom plot is from the third party package graphlab.

- Base Model without Bias: Coneheads is a massive outlier on this plot of the latent factors. This makes sense, since it is a comedy and all of the other movies are more serious. A pretty clear pattern seems to be that as the vertical position increases the films become more action centric. In other words, movies towards the bottom are more intellectual while the ones at the top are have more nonsensical action scenes (i.e. like Michael Bay Movies). Independence Day was one of the most popular movies so it makes sense that it is slightly further away from the main cluster since it has a broader appeal.
- Base Model with Bias: The horizontal axis on this plot seems to correspond to the amount of action in the movie. Movies like Independence Day are filled with nonstop action whereas movies like 2001: A Space Odyssey are more thematically complex. The outliers in this plot seem to be the comedies like Coneheads and Back to the Future. The vertical latent factor doesnt seem to correspond to an obvious trend in the data.

  - Graphlab: We notice that the vertical latent factor appears to at least partially measure the amount of action in the film; Coneheads, Apollo 13, Back to the Future, and 2001: A Space Odyssey are all relatively low-action, in the context of the other movies in this diagram. We also note that some of the more alien-centric movies are closely clustered, namely Star Trek: The Wrath of Khan and Aliens.

  In all three plots, we noticed a correlation between one of the latent factors and the amount of action present in the films. Thus, it appears that the amount of action is a factor that, directly or indirectly, can be quite useful in classifying films and is often indicative of how much a given viewer will enjoy them.

  We also note that 2001: A Space Odyssey was always distinct, even to the point of being an outlier, from the other movies, perhaps indicating that it had significant differences from other movies of its genre.

- **Genre Comparisons**:

  Across our visualizations of the three genres, we noticed several interesting points.

  Documentaries, which are thematically very different from comedies and sci-fi movies (which generally aim to entertain rather than inform) and which make up a small portion of the dataset overall, had visualizations with few similarities to our others.

  However, sci-fi movies and comedies followed similar trends to our other visualizations. The level of action in the movies was often an important factor in which viewers rated it highly, as well as the movies maturity or complexity, which weve generally been considering inversely related with action. Overall, for most popular movies this metric seems very important and nearly universal. Similarly, setting and context seem to be important in both comedies and sci-fi films, with historical comedies such as Rosencrantz and Guildenstern are Dead and several Monty Python movies being grouped closely, and

  Beyond that, the latent factor clusterings seem to be based mainly on genre-specific factors.