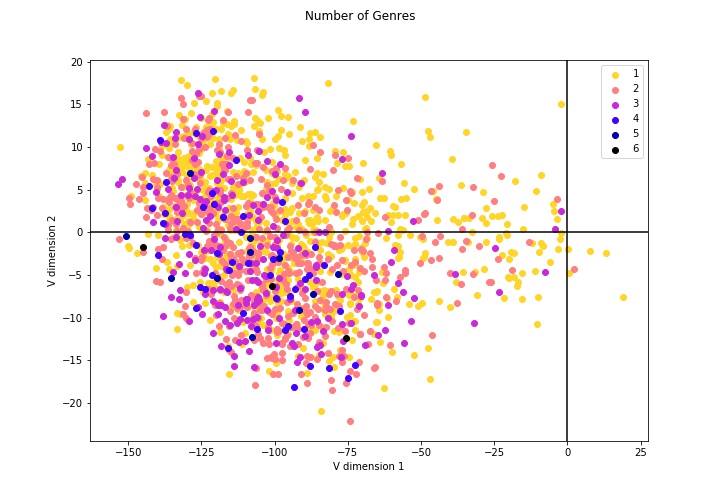
As we tried to make sense of our visualizations and considered different genres as the separating factor for movies, we realized that, perhaps, the projections we created shared greater insight beyond the scope of one or two genres.

We decided to plot the movies on V dimensions 1 and 2 (using a no bias matrix factorization model) with different colors representing how many genres a movie belonged to. From this, we can see that the more genres a movie belongs to, the more concentrated it is —there is less variation. However, because this could be attributed to the fact that there are more movies with only 1-2 genres, and fewer with 5-6 genres, we tried running the matrix factorization with the bias.



By running the matrix factorization with the bias term, we see a more centralized version of above:



By drawing connected lines to a mean point, we can see a better visualization of how the movies are plotted along these axes:



Since it’s hard to compare so many movies with one another, we decided to take a look at the “extremes” of each partition to get a smaller set of movies to analyze.

By comparing the “outliers” (movies with points farthest from the mean center of the partition) of the plotted movies, we see a clear divide between the movies, especially noticeable across dimension 2. The negative side has an abundance of thriller, dramatic movies, most of which are rated above PG-13, such as Pulp Fiction, A Clockwork Orange, The Crying Game. Meanwhile, the positive side has an abundance of comedic, family style movies such as Transformers, Space Jam, and Star Kid.

