

# Natural Language Processing of Musical Theater Song Lyrics

Nora May  
Metis, Project Fletcher



# Process and Methods



- **Scraped lyrics**
  - Over 100 “popular” musicals
  - 1934 - 2015
  - Over 2000 songs → Documents
- **Topic Modeling**
  - TF-IDF, NMF
- **Sentiment Analysis**
  - Vader
- **Recommender**
  - Word2vec similarity measures
  - K-Means for Spotify API for audio features

# Topic Modeling

```
fdist.most_common(10)
```

```
[('love', 2379),  
 ('know', 2291),  
 ('oh', 2115),  
 ('go', 1938),  
 ('got', 1685),  
 ('get', 1589),  
 ('see', 1580),  
 ('never', 1576),  
 ('time', 1387),  
 ('come', 1384)]
```

```
get_topics(more_musicals_df_real['song_lyrics_pro'], 8, st_words, 0.5, 5)
```

Topic #0:

know go time say could would want tell think see

Topic #1:

love want heart true never tell care without always tomorrow

Topic #2:

got babi gonna yeah get come gotta take big caus

Topic #3:

dream true star come seem believ everi could boy wonder

Topic #4:

girl littl beauti get boy danc miss big pretti wait

Topic #5:

goodby hello go farewell long bye america spell old bell

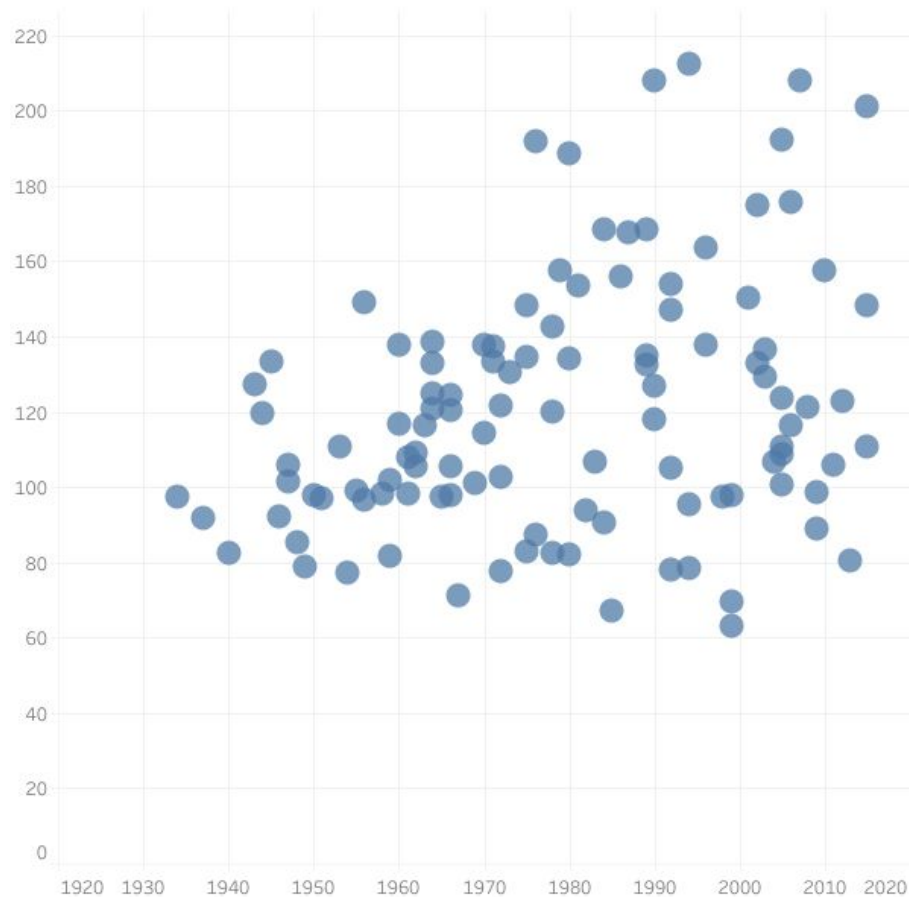
Topic #6:

day life world new night light live sing come song

Topic #7:

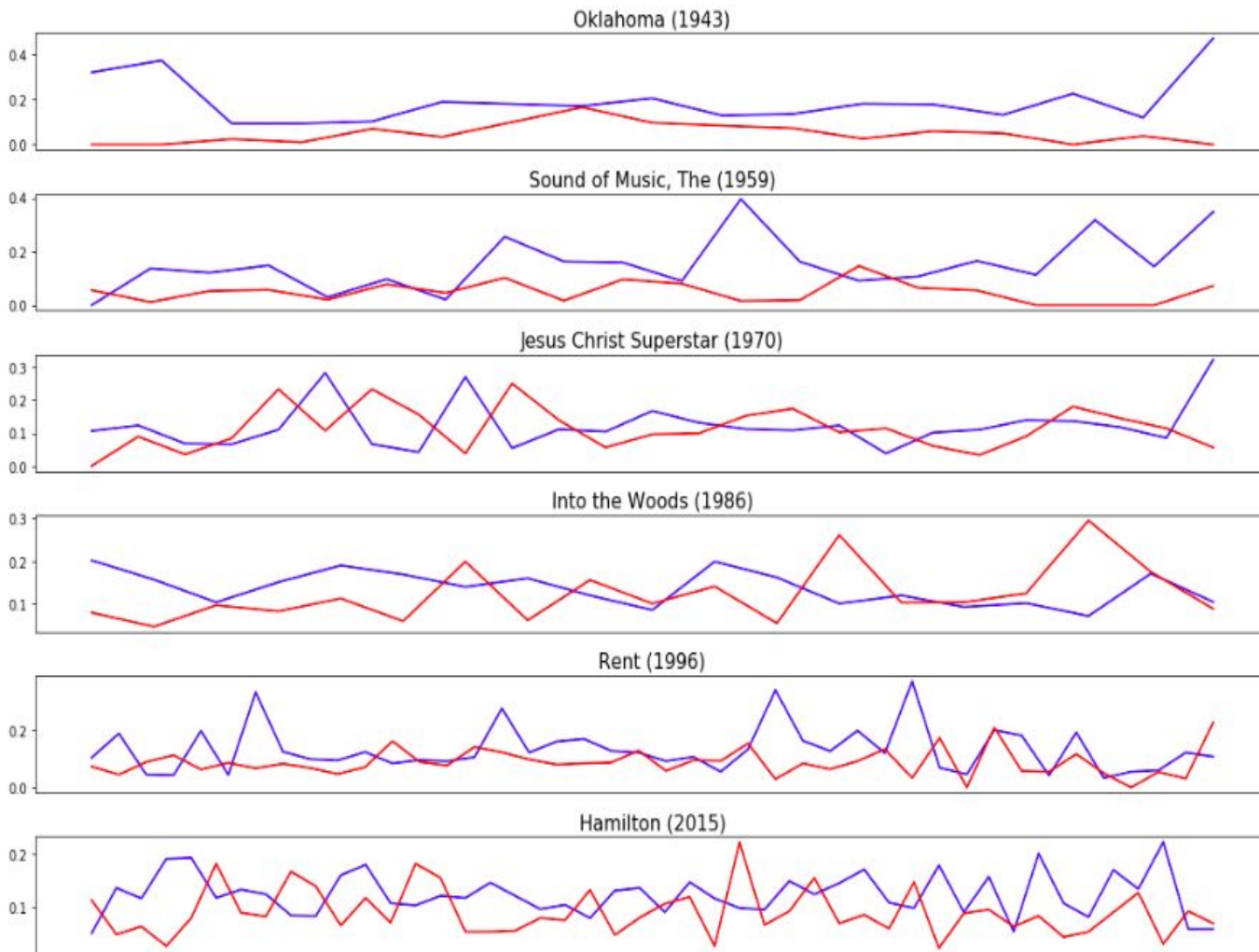
man woman old king men ladi never lord big marri

## Average unique words per song in a musical over time



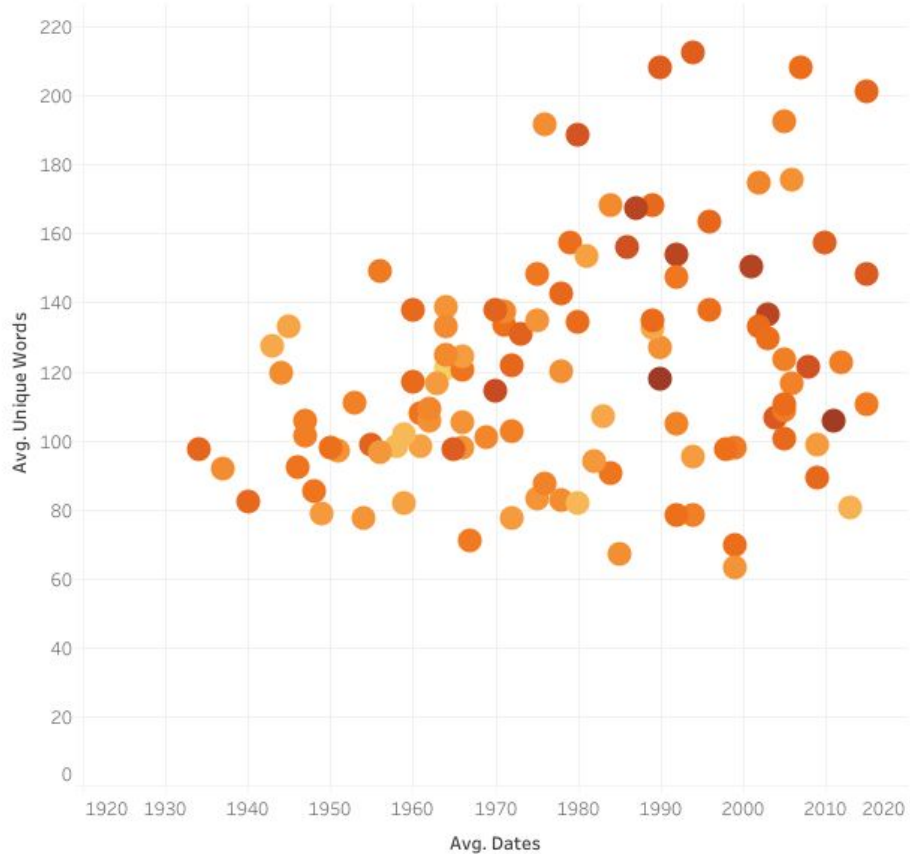
# Sentiment Analysis

Pos  
Neg

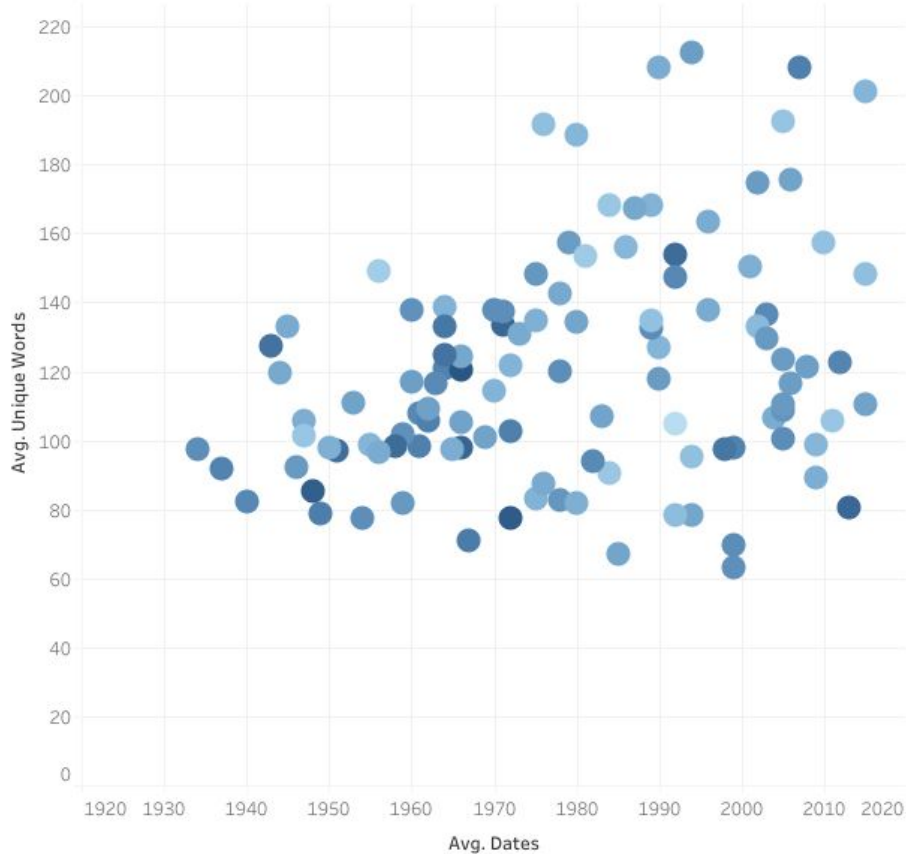


# Average Unique Words Per Song in a Musical Over Time

color indicates strength of negative sentiment



color indicates strength of positive sentiment



# Recommender

- Word2Vec
  - Generate similarity index over whole corpus
- Spotify Data
  - 8 Audio Features
  - K-means and DBScan to use as a filter
- Flask App
  - Thanks Blake!

The gensim logo, featuring the word "gensim" in a blue, lowercase, sans-serif font. A red L-shaped line is positioned to the right of the logo.

# Future Directions

- Add all musicals; go back to Operas/Operettas
- Analysis of whole Libretti for further understanding of how songs are used in storytelling
- Make a more detailed similarity model and more palatable website for recommender







Questions?