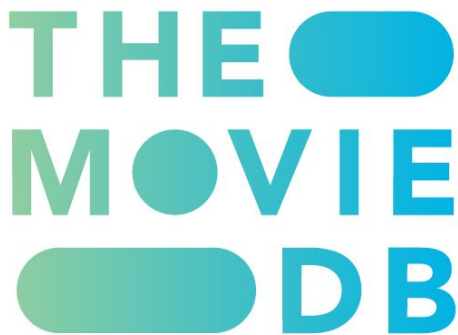


# TV episodes Reddit Discussion

Max Garber

# The Data

- Top (by score) 100 Reddit posts between 1 and 3 days after an episode premieres from the shows subreddit
- Information about each episode from The Movie Database (TMDb)
- Rating for each episode scraped from Rotten Tomatoes



# Processing the Text

1. Combined each post title and text.
2. Cleaned text
3. Tokenized
4. Parts of Speech tagging
5. Removed Stopwords
6. Stemming
7. Count Vectorizer
8. TF-IDF

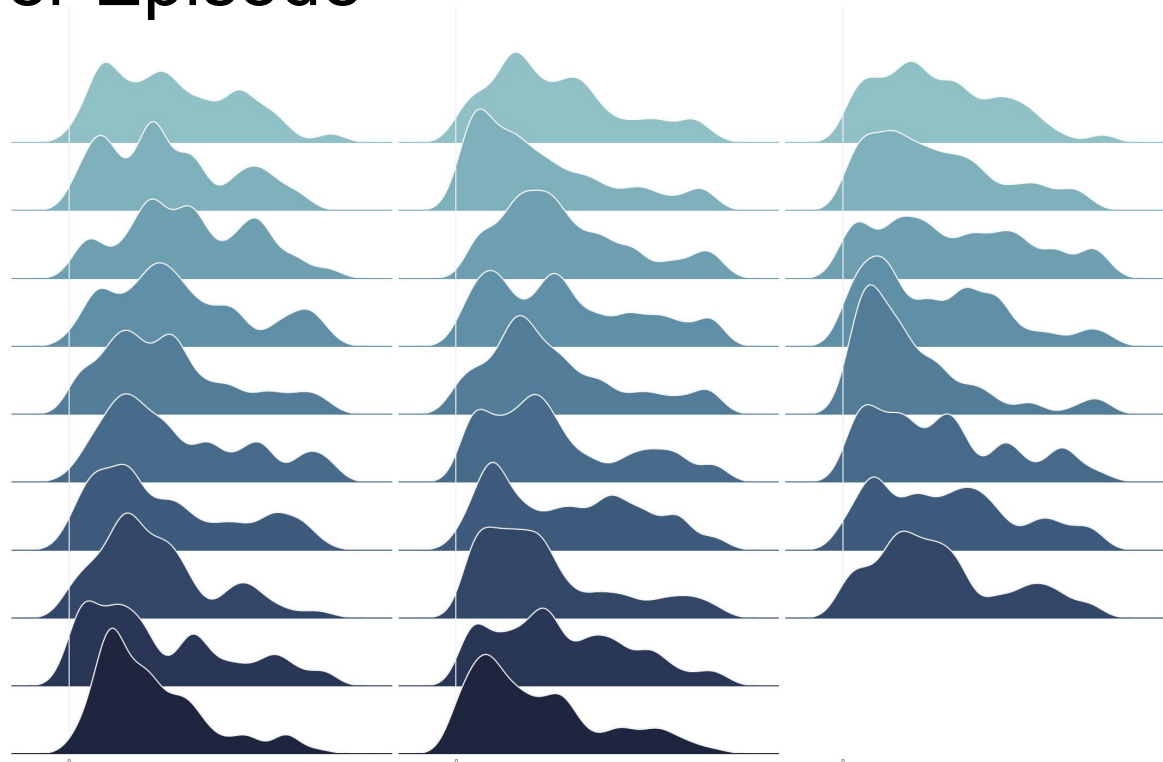
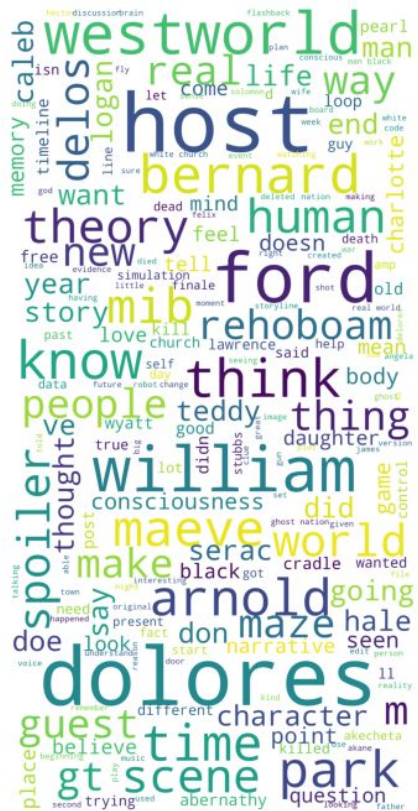


# Westworld: Different Plot Lines

# Latent Semantic Analysis



# Westworld Topic per Episode



### Distribution of Topic over Posts

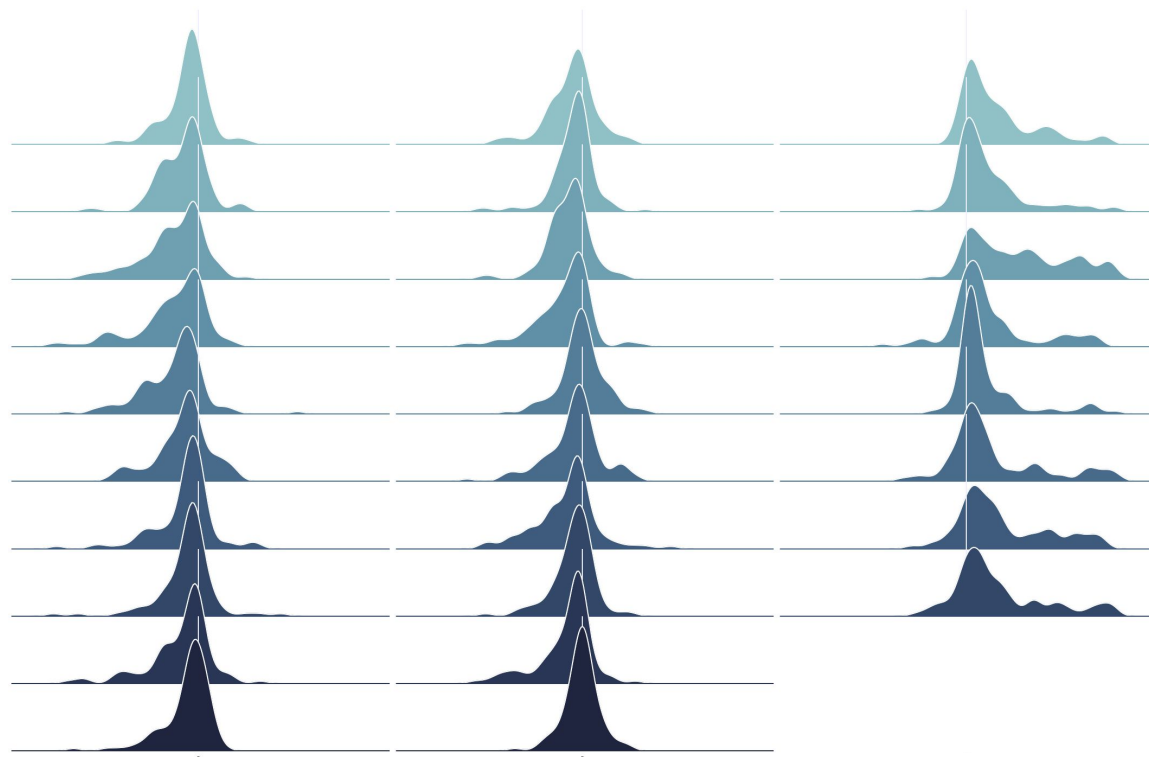
Season



# Topic per Episode



## Episode

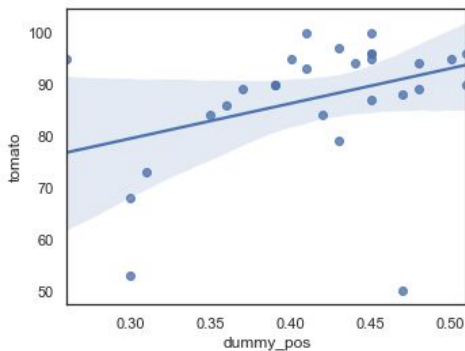


### Distribution of Topic over Posts

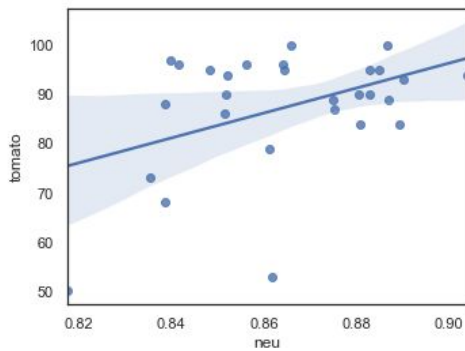
Season

# Sentiment : Westworld

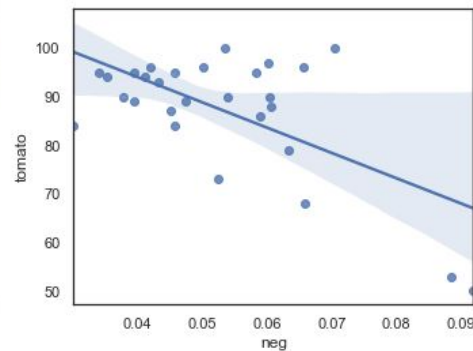
Tomatometer



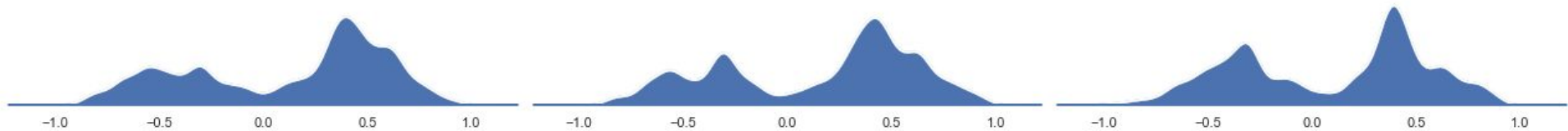
Positive Sentiment  
of post titles



Neutral Sentiment  
of post titles + text



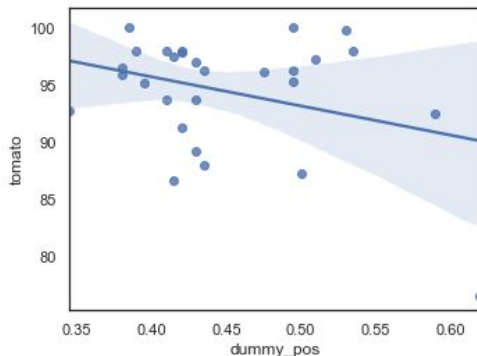
Negative Sentiment  
of post titles + text



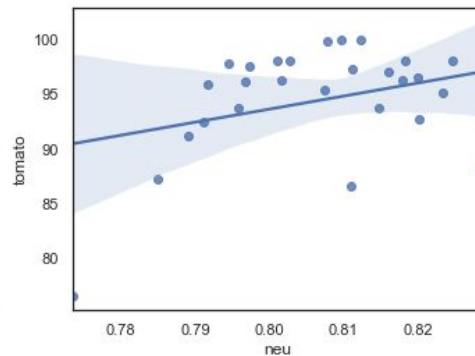
Distribution of Sentiment over Posts per Season

# Sentiment : Westworld - Comments

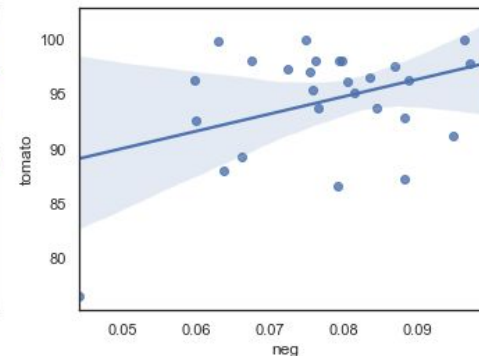
Tomatometer



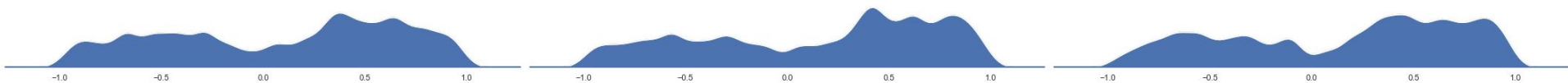
Positive Sentiment  
of comments



Neutral Sentiment  
of comments



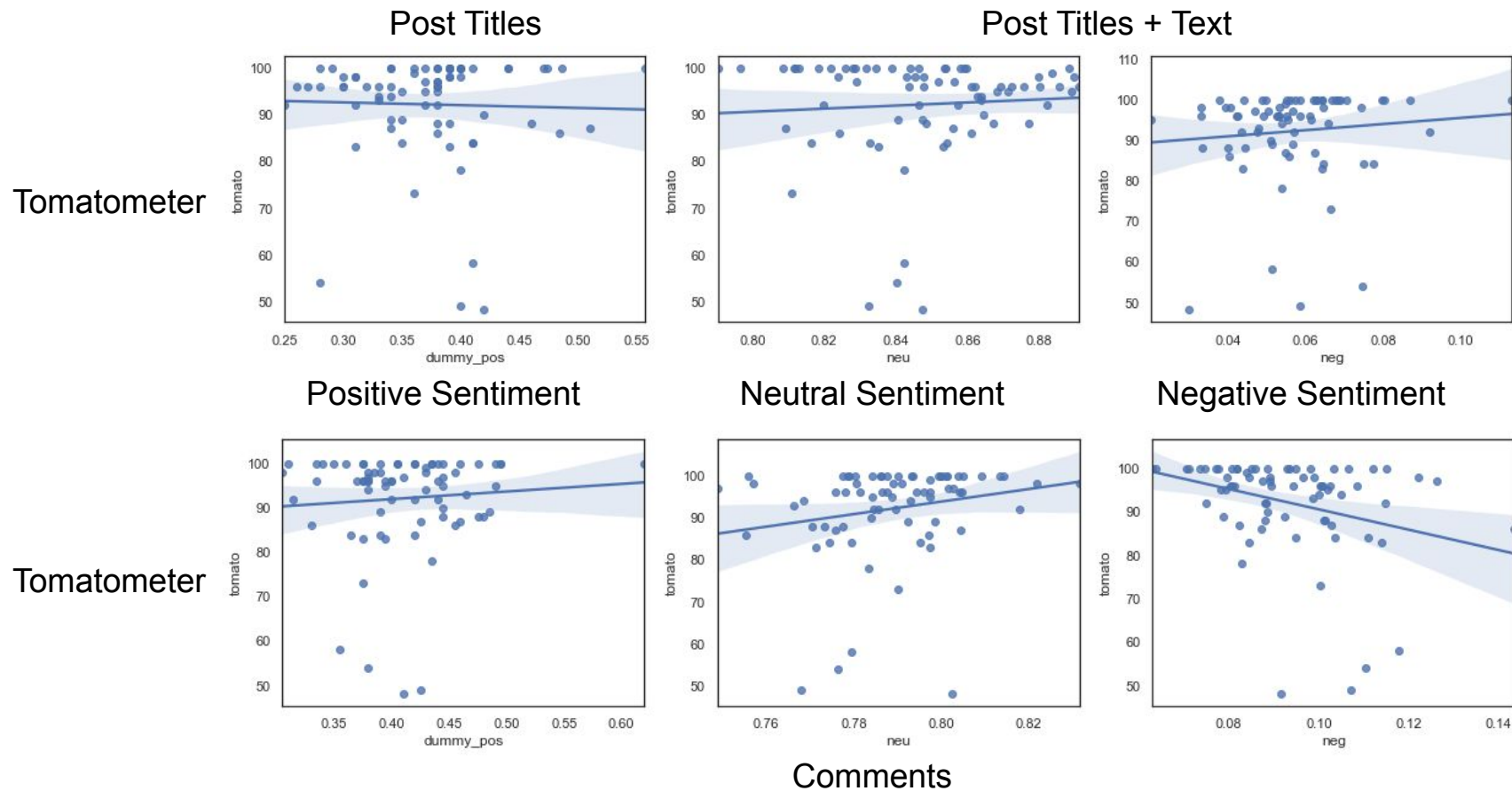
Negative Sentiment  
of comments



Distribution of Sentiment over Comments per Season



# Sentiment : Game of Thrones



Questions?

## Words common across multiple shows



## High Rating



## Low Rating