Gaining Insight on Customer Topics of a US Airline via Twitter

Charlie Lew
November 15th, 2019
Metis

Introduction & Objective

- Social media popular medium for customer feedback in US airline industry

Introduction & Objective

- Social media popular medium for customer feedback in US airline industry
- Could topic modeling be used to offer new products & services?

Introduction & Objective

- Social media popular medium for customer feedback in US airline industry
- Could topic modeling be used to offer new products & services?
- Objective

Use Natural Language Processing to gain better insight into Customer Topics via Twitter

Dataset & Tools

- Pick Delta Airlines as a case study
  - *Near 20% domestic market share*
  - *High volume of daily twitter activity: ~1,500*

Dataset & Tools

- Pick Delta Airlines as a case study
  - *Near 20% domestic market share*
  - *High volume of daily twitter activity: ~1,500*
- Data source: *Twitter*
  - *Pulled 23,000 tweets covering a two week period*

Dataset & Tools

- Pick Delta Airlines as a case study
  - *Near 20% domestic market share*
  - *High volume of daily twitter activity: ~1,500*
- Data source: *Twitter*
  - *Pulled 23,000 tweets covering a two week period*
- Tools
  - *Python, Pandas, scikit-learn*
  - *Tweepy*
  - *Matplotlib, Seaborn*
  - *NLTK, Spacy, pyLDAvis*

Methodology Pipeline

Pull Tweets:

Tweepy

Methodology Pipeline

Pull Tweets:

Tweepy

→

Cleaner:

Tokenize,
Lemmatize
.etc

Methodology Pipeline

Pull Tweets:

Tweepy

→

Cleaner:

Tokenize,
Lemmatize
.etc

→

Vectorize:

TF-IDF
CountVectorizer

Methodology Pipeline

Pull Tweets:

Tweepy

→

Cleaner:

Tokenize,
Lemmatize
.etc

→

Vectorize:

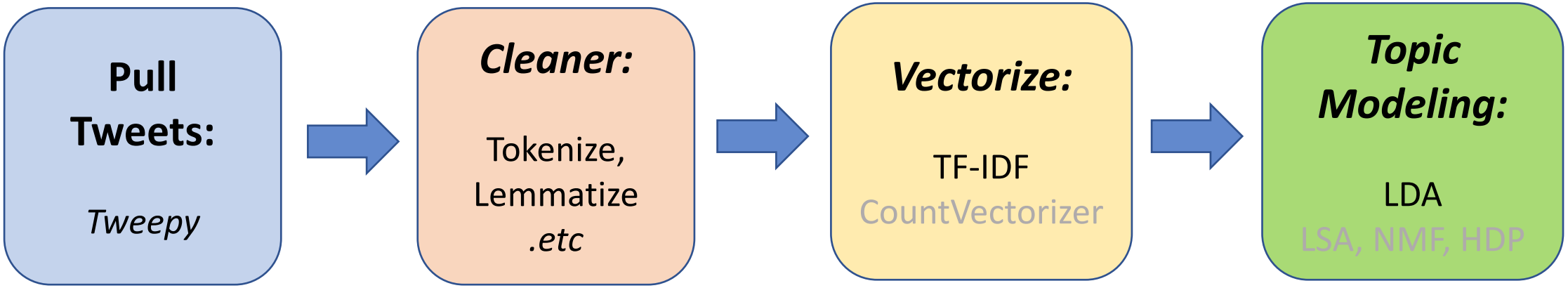
TF-IDF
CountVectorizer

→

Topic Modeling:

LDA
LSA, NMF, HDP

Methodology Pipeline



- Process above is repeated when new data is available
- Arrived at 20 topics
- Picked three that relayed clearest meaning

Selected Topic List (10/29 – 11/13)

Selected Topic List (10/29 – 11/13)

- **Hour Delay Minute Connection Waiting**

Selected Topic List (10/29 – 11/13)

- Hour Delay Minute Connection Waiting

- **Baby Pump Breast Bathroom Milk**

Selected Topic List (10/29 – 11/13)

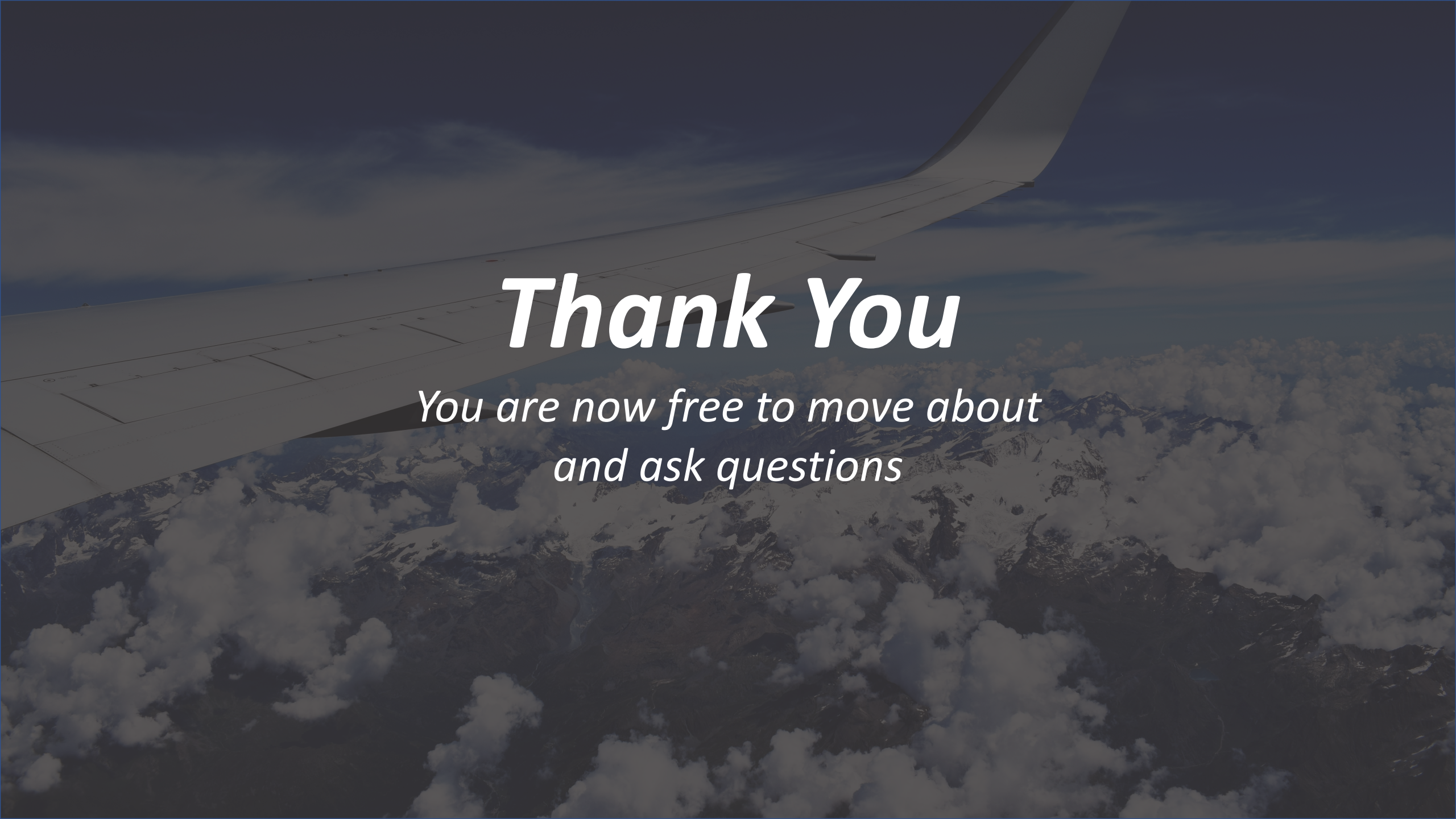
- Hour Delay Minute Connection Waiting
- Baby Pump Breast Bathroom Milk
- **Porn Child Movie Watch Showing**

Summary

- Topic Modeling a challenging subject especially on tweets
  - *Contains plenty of noise*
- How one ‘cleans’ could affect the outcome
- For Delta over two weeks
  - *Complaints of delays most prevalent*

Next Steps…

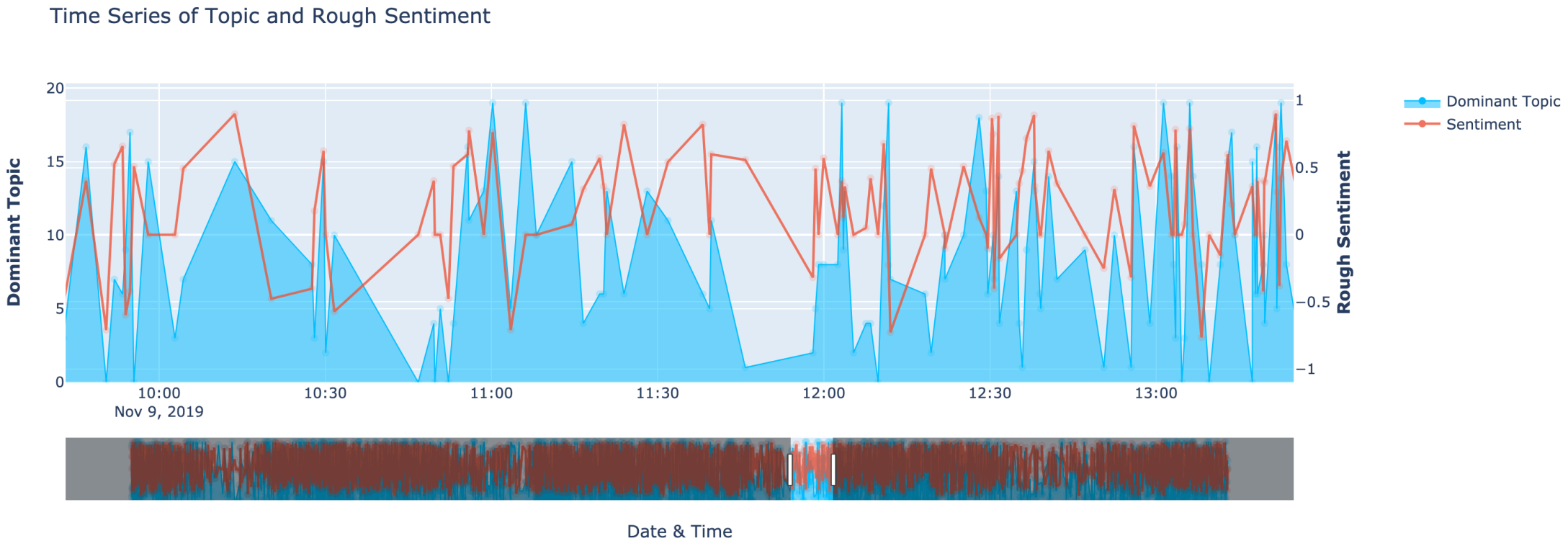
- Perform further cleaning & modeling
- Get more data over a longer period
  - *Topics can change over time*
- Real-time visualization of topics and sentiment
  - *Useful tool for airlines to anticipate outcomes & react*
- Could deep learning help filter through the noise?



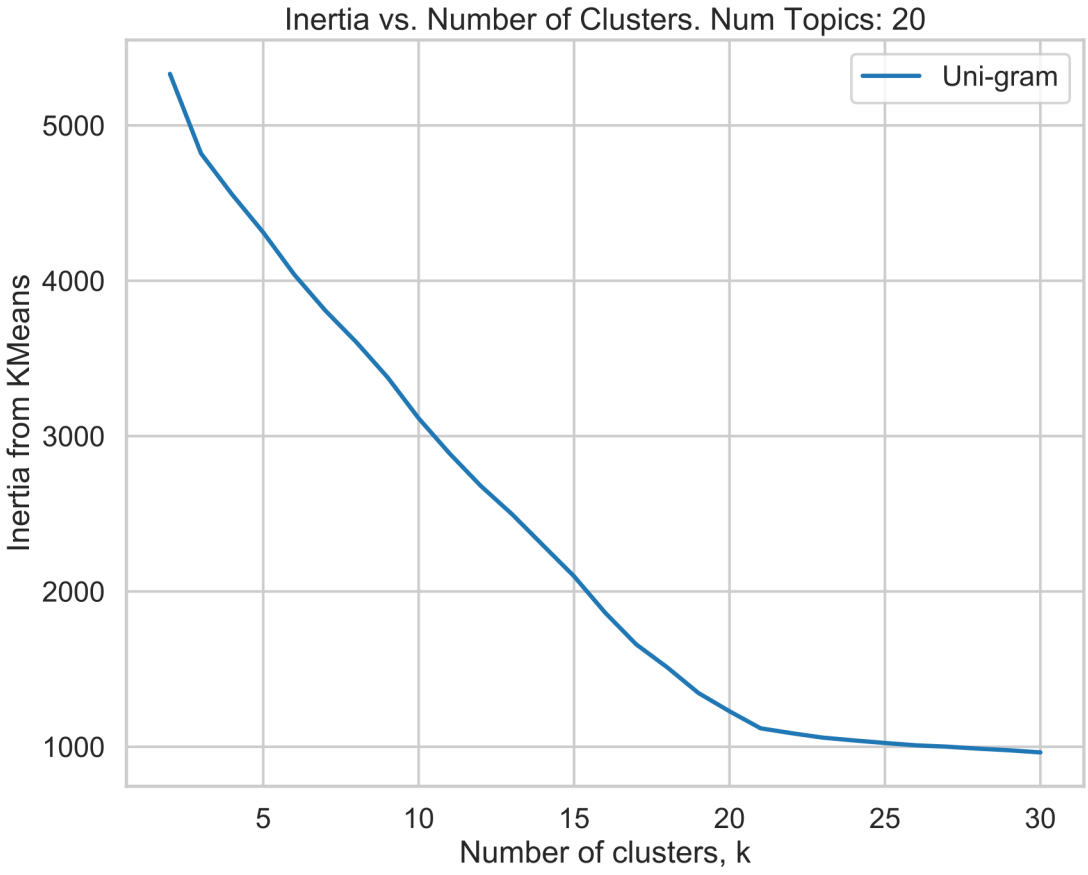
Thank You

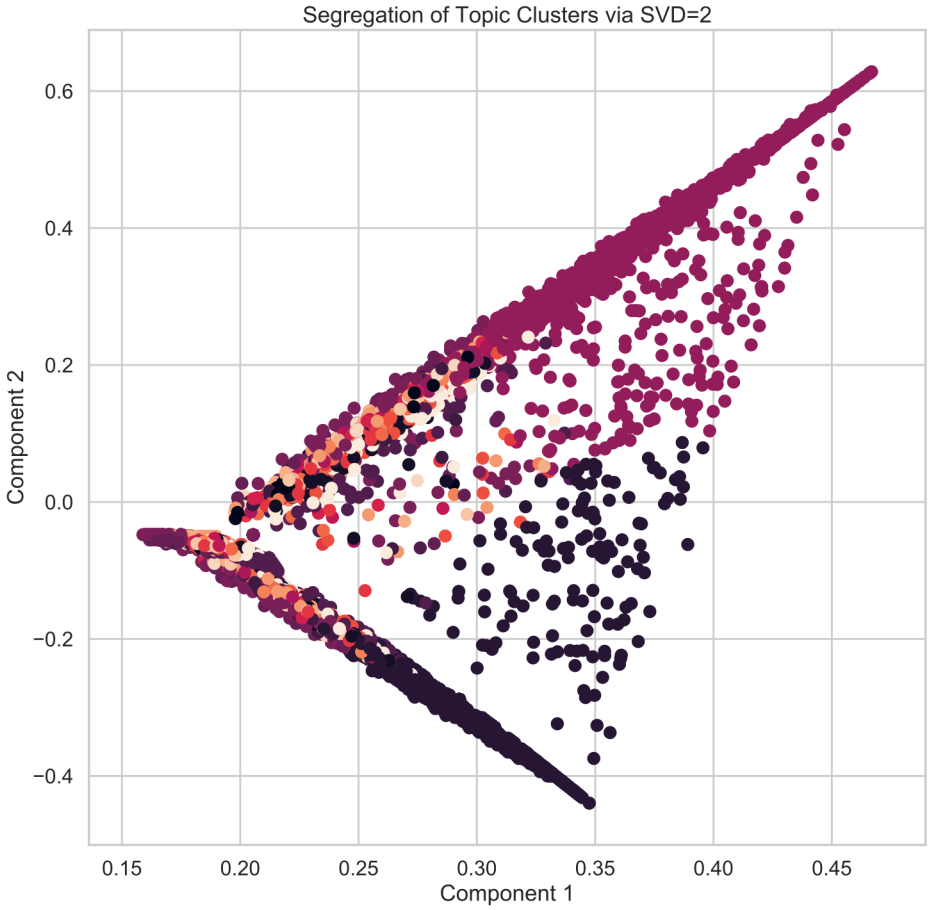
You are now free to move about and ask questions

Time Series of Topic and Sentiment

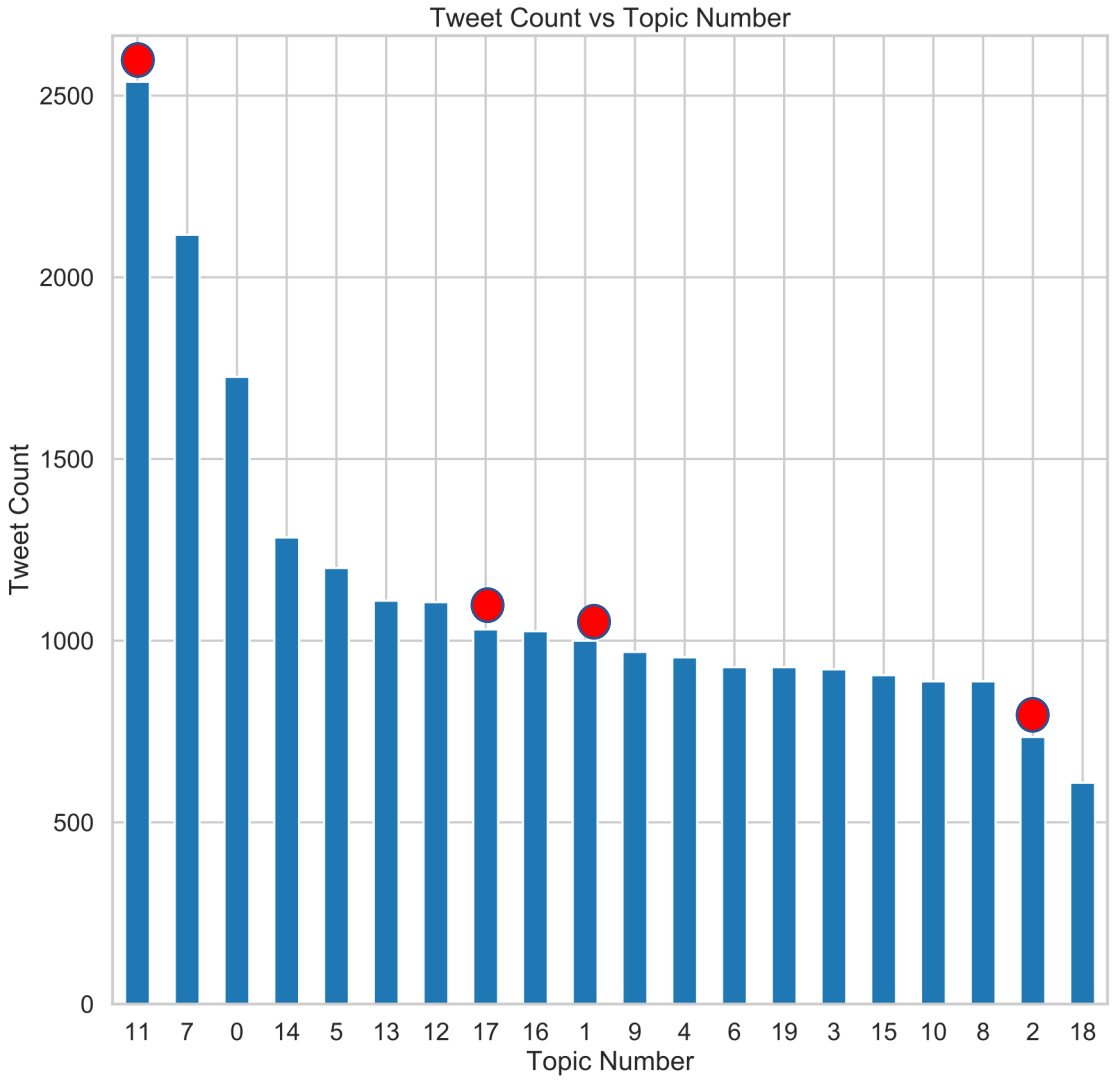
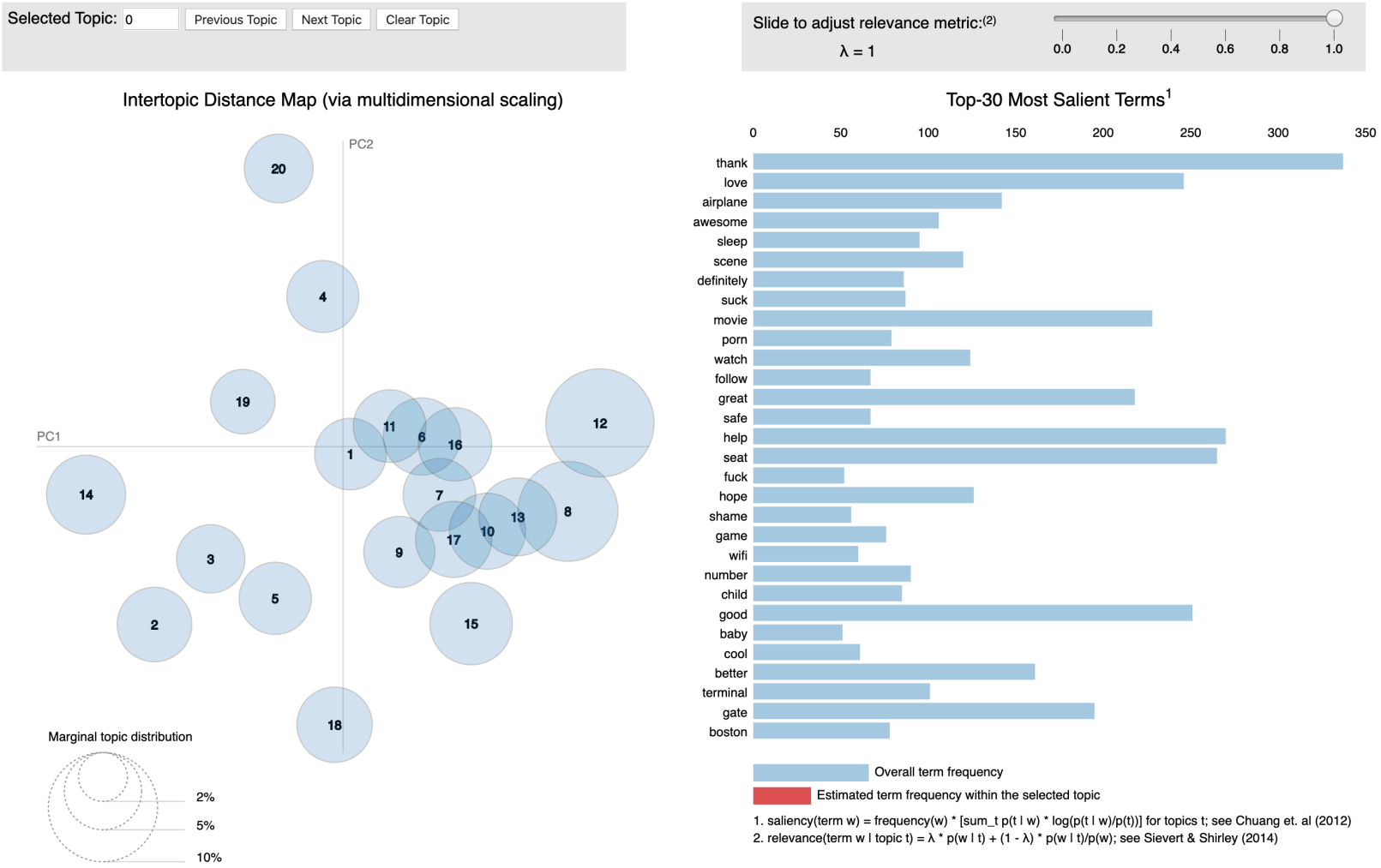


Extra Slides (K-Means for LDA)





Topic Distribution



pyLDAviz for 20 Topics