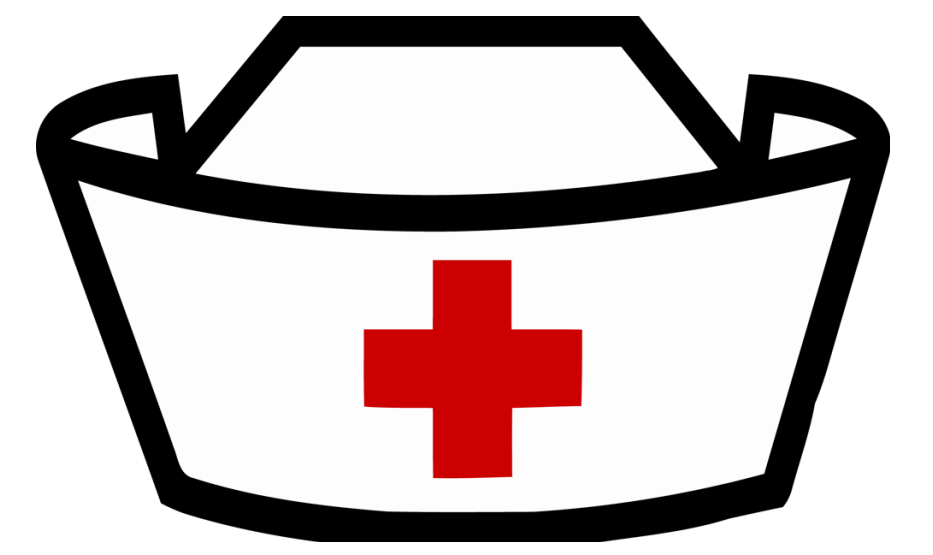# Exploring Emotions of Nurses During COVID Using Reddit

Aly Ibrahim, Eric Larson, Haven Penn, and Shreya Shanker
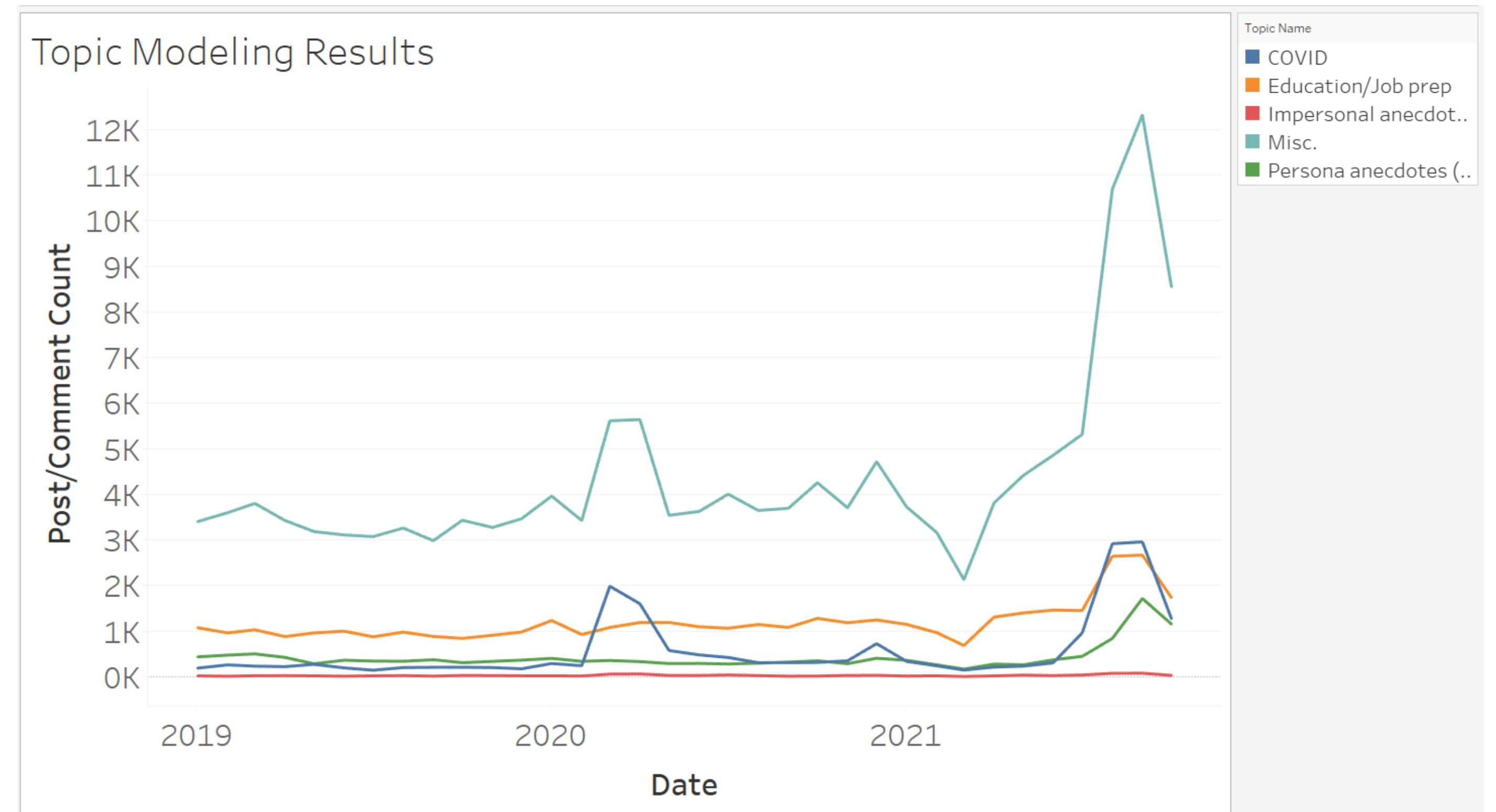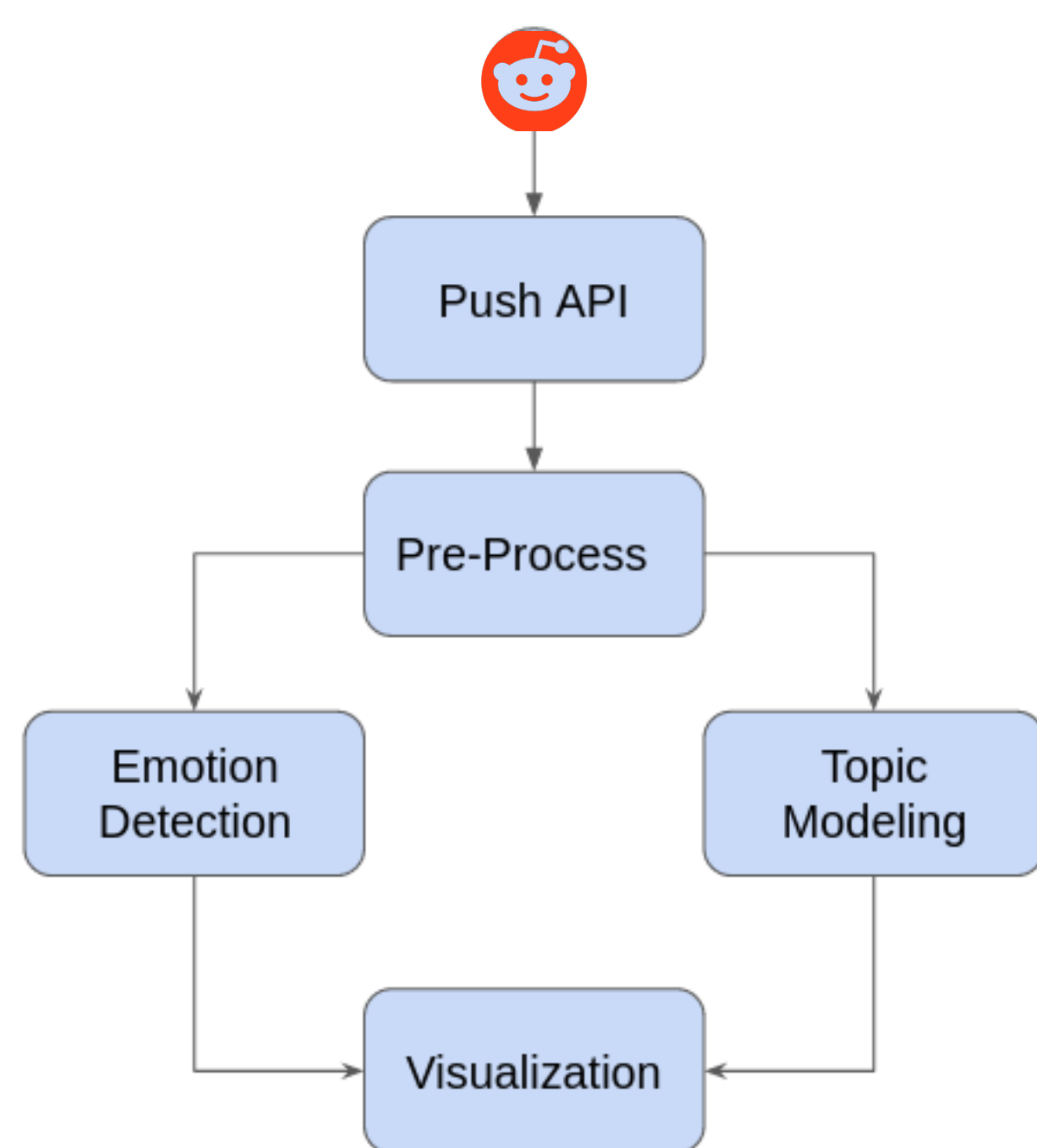
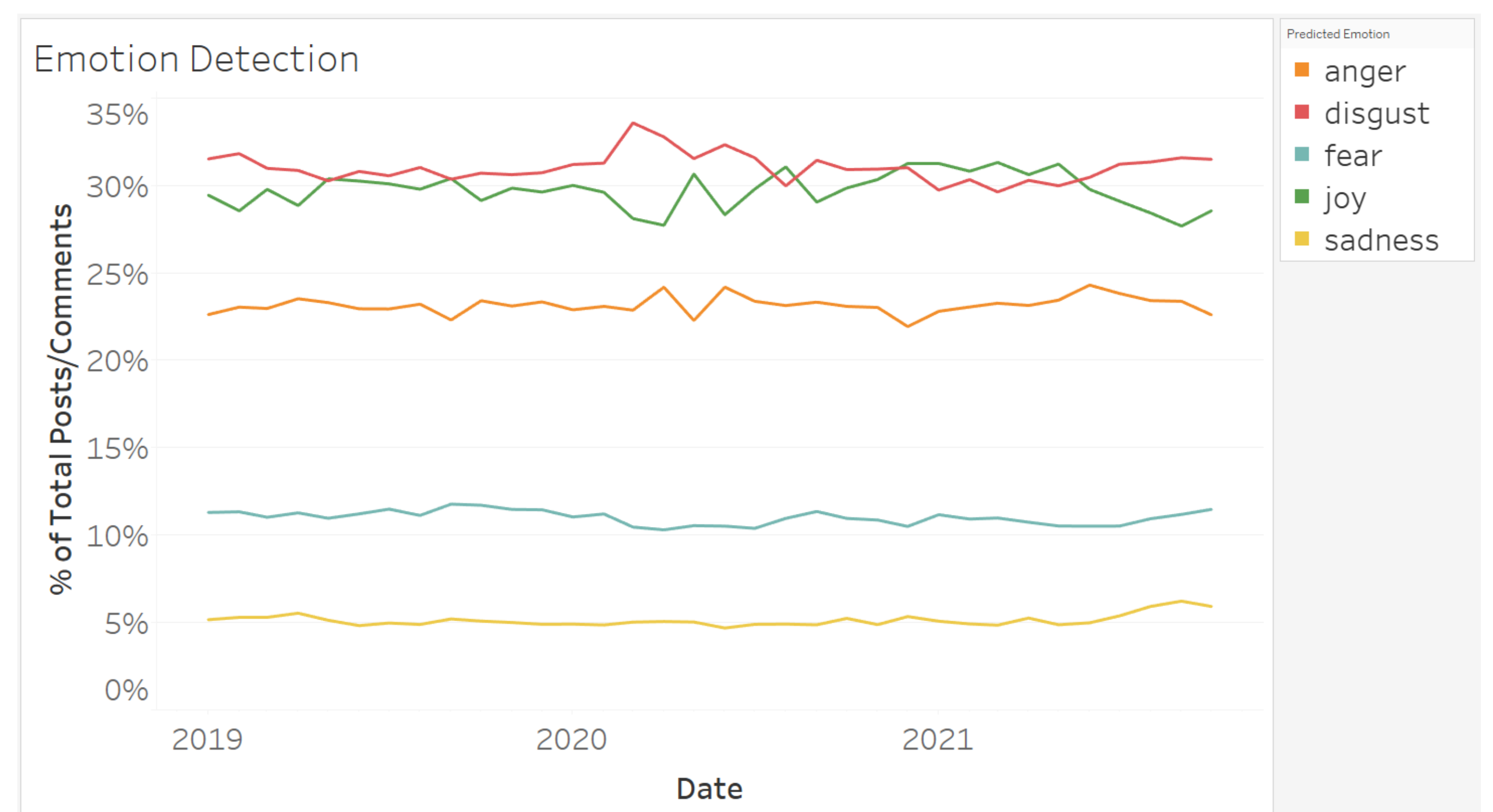**GT Georgia Tech**

Introduction

- **Problem:** How to evaluate emotional stress on nurses during COVID pandemic
- **Standard Method:** Using surveys to gather data
- **Weaknesses:** Limited participants, long time to collect data, and possible survey bias
- **Solution:** Use topic modeling and emotion detection algorithms to analyze Reddit posts and comments, searching for trends over pre-COVID and post-COVID conversations.

Data

- **Source:** r/Nursing posts/comments from 2019 to present
- **Retrieval:** Retrieved from Reddit archive using PushShift API
- **Size:** ~1,000,000 posts and comments total

Topic Modeling

- **Algorithm:** LDA algorithm - a probabilistic model
- **Size:** ~200,000 posts and comments after filtering
- **Topic number:** Picked number of topics using Coherence

Emotion Detection

- **Algorithm:** Linear SVM
  - 63.5% accuracy classifying text into 5 emotions
    - 20% points higher than model in comparable research
  - Trained using ISEAR dataset, 8000 labelled documents
- **Features**
  - NRC Emotion Intensity Lexicon ratings,
    - Approximate intensity of 8 emotions
  - Doc2Vec vectors
    - Approximate the concepts
  - Latent Sentiment Analysis vectors
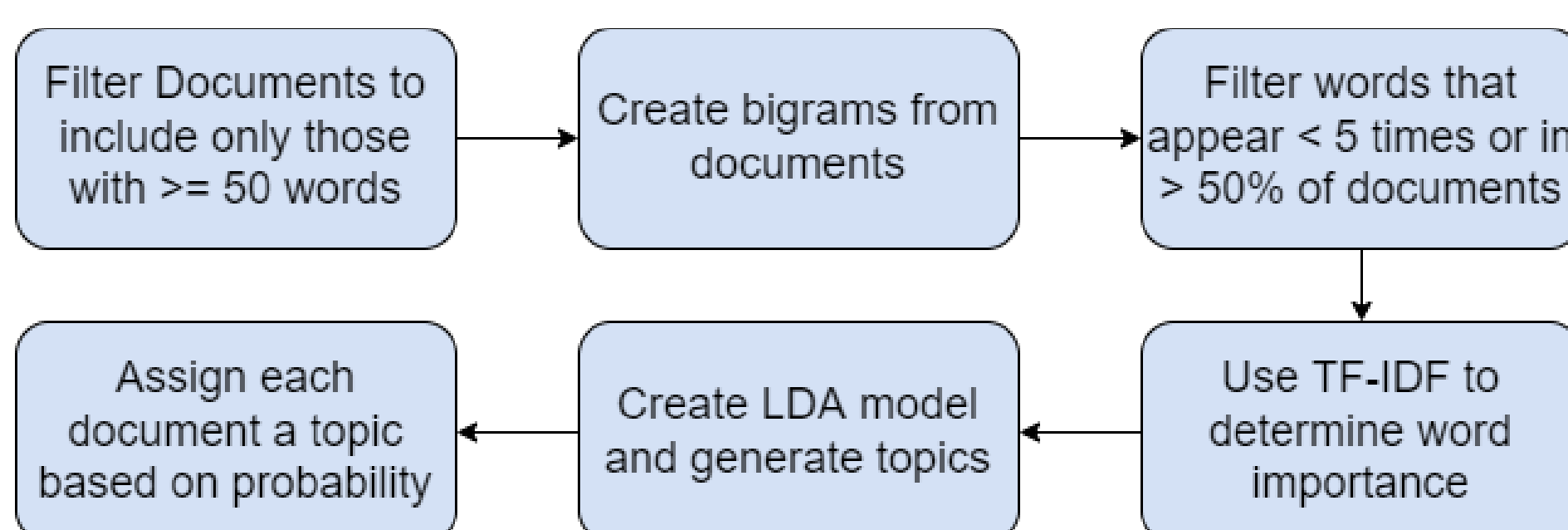   - Emphasize unique and important words

- The COVID topic spikes in March-April 2020 (beginning of pandemic), and there is another peak in December 2020 (vaccine introduction)
- Education/Job prep remained reletively popular throughout the timeline

- Volatility increases after beginning of pandemic
- Joy is at its lowest levels during 1st wave (Mar - May 2020) and 4th wave (Aug - Oct 2021); the 3rd wave (Nov 2020 - Jan 2021) began a period of the highest levels of joy

Conclusion

- **Takeaways**
  - The algorithms seem to work! The results generally follow the COVID timeline
  - The 1st and 4th COVID waves had lowest levels of joy and highest counts of content related to COVID, potentially indicating higher levels of emotional stress

- **Areas for Improvement**
  - More recent and topical training dataset for emotion detection model
  - Bi-grams significantly improved topic modelling results; emotion detection results, largely based on 1-grams, could see a similar improvement
  - Minor interference from Reddit bots, which could be filtered out using a bot database

To see more insights visit our interactive dashboard!