



Leveraging Social Media to Map Natural Disasters

Russell Garner, Michael Knight, Bruno Santos, Aakash Sharma
Data Scientists





Problem Statement

- How can we effectively utilize social media to determine how natural disasters have impacted a specific area, and subsequently map what areas require the most assistance?



Social Media Used

- Tweets from the CrisisLex Dataset – 40,000 tweets
 - 2012 Hurricane Sandy, 2013 Oklahoma Tornado, 2012 Alberta Floods, 2012 Queensland Floods
- Tweets gathered manually via scraping
 - 17,717 Hurricane Harvey tweets, 3,268 Omaha Floods tweets, 11,927 LA Earthquake tweets



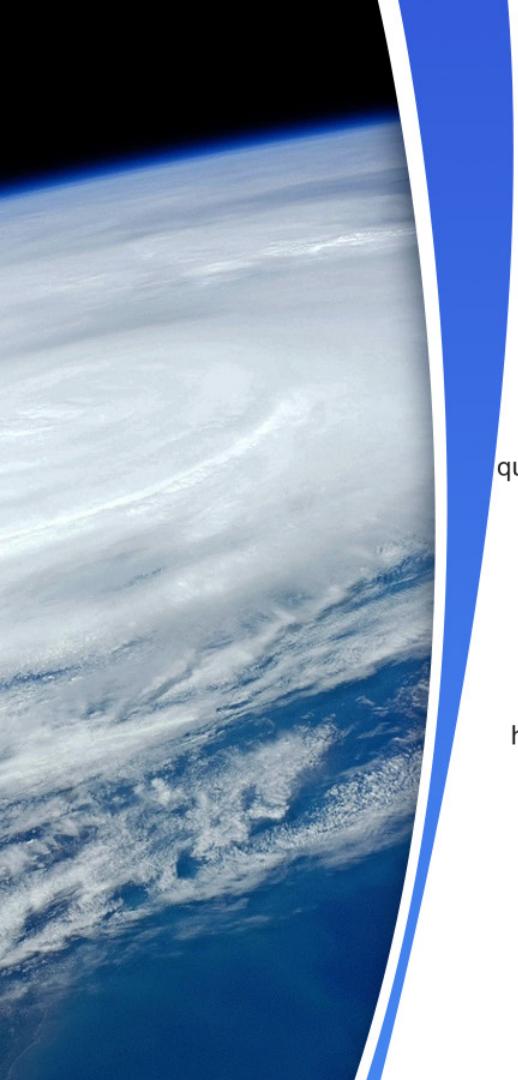
Model Training

- Utilizing the CrisisLex tweets, we built a Logistic Regression model to classify new tweets as being either “on-topic” or “off-topic” (**94%** accuracy)
- Utilized TfidfVectorizer to process the tweets
- Source: CrisisLex.org – CrisisLexT6

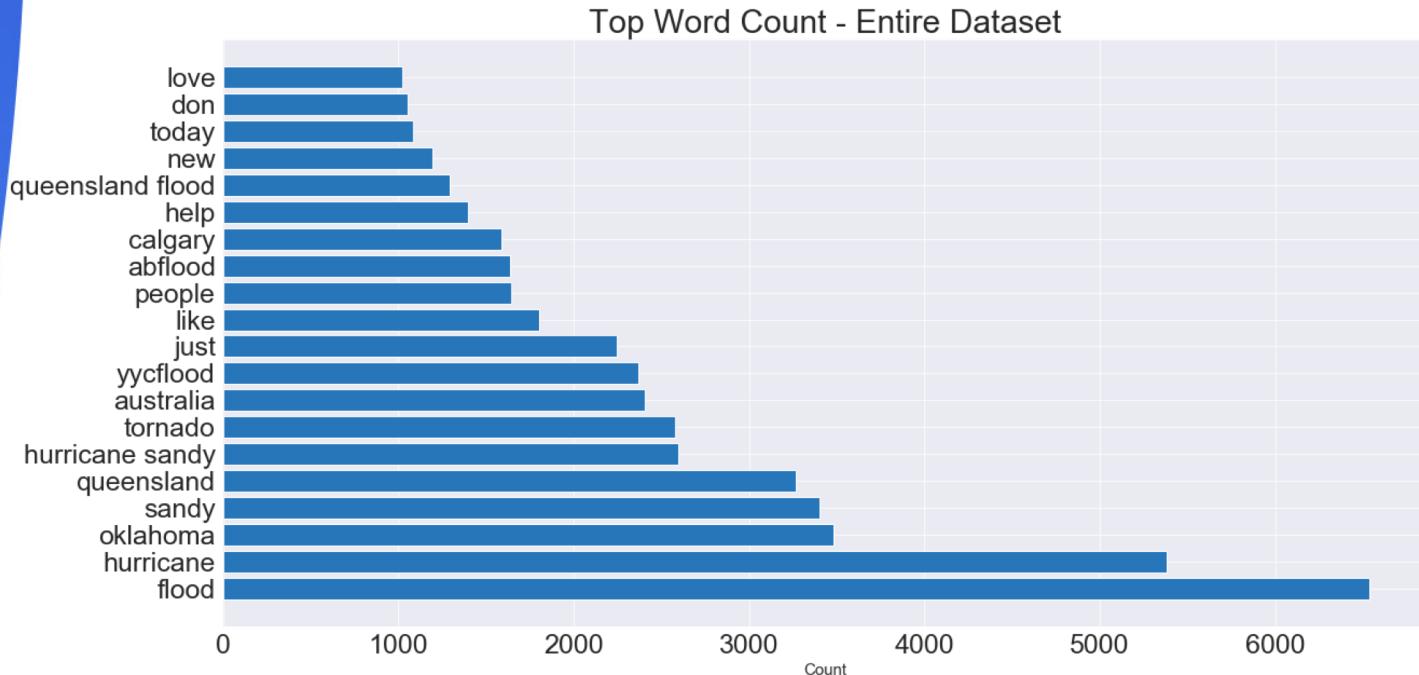


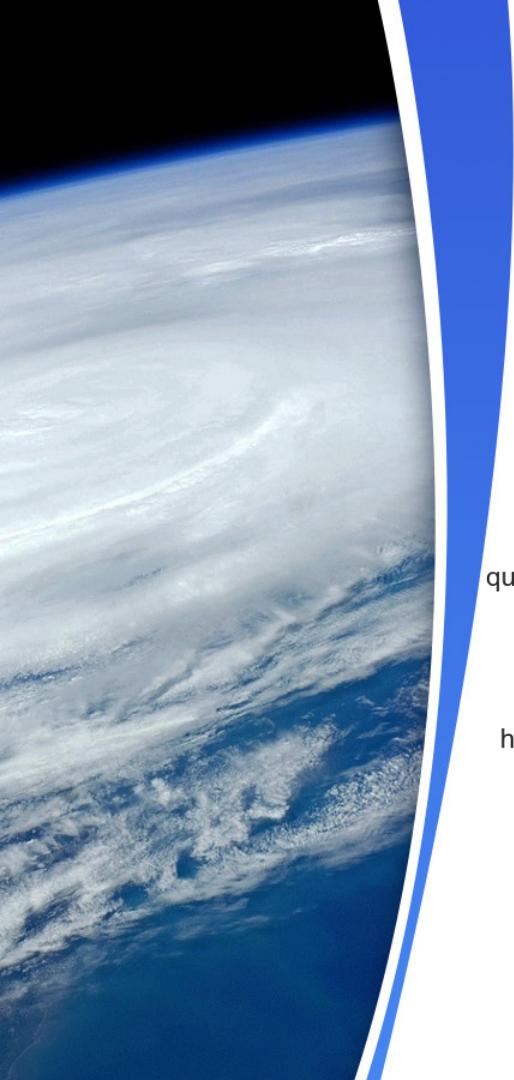
Tweet Processing

- Clean and process our raw tweets
 - Remove unnecessary characters, tokenize
- Classify new tweets as on-topic or off-topic
- Of the on-topic tweets, classify them as urgent or non-urgent using NLP, Word2Vec, Cosine Similarities
 - Word2Vec model pre-trained on Google News
- Explore most common words

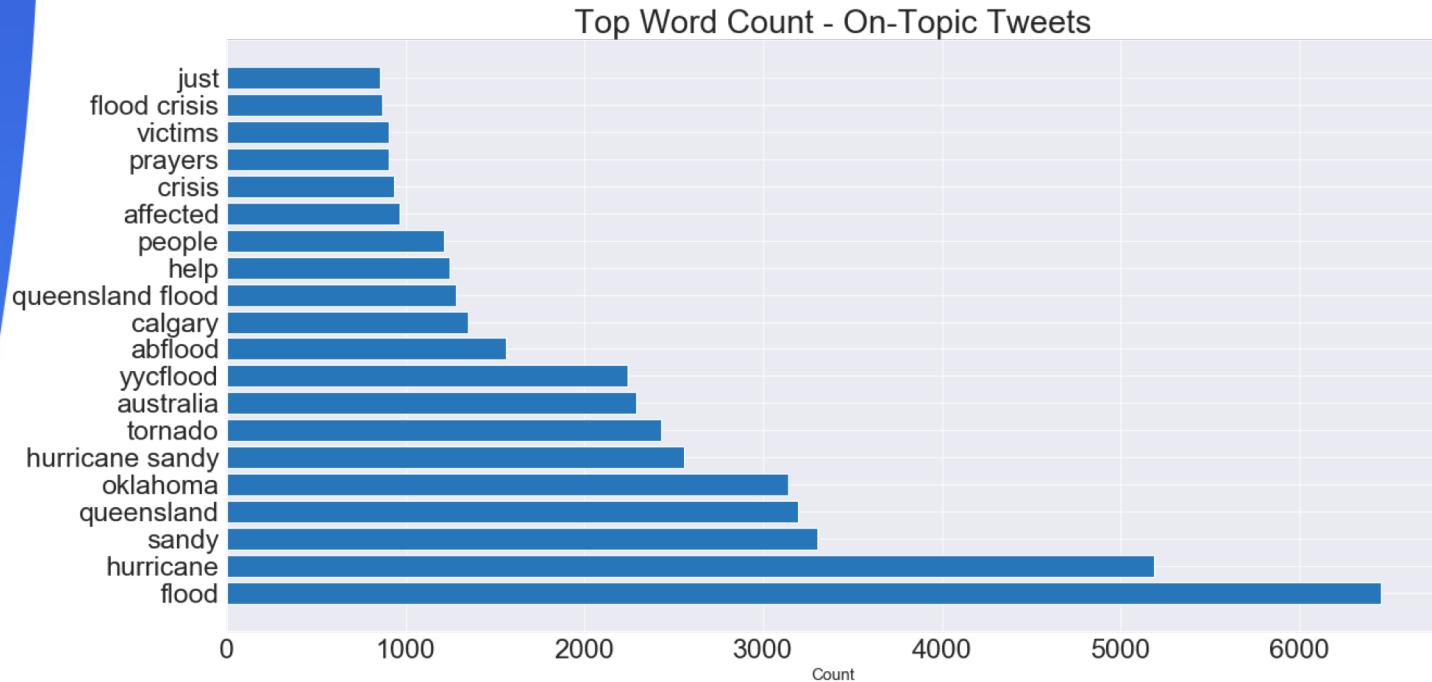


Top Features – CrisisLex Dataset





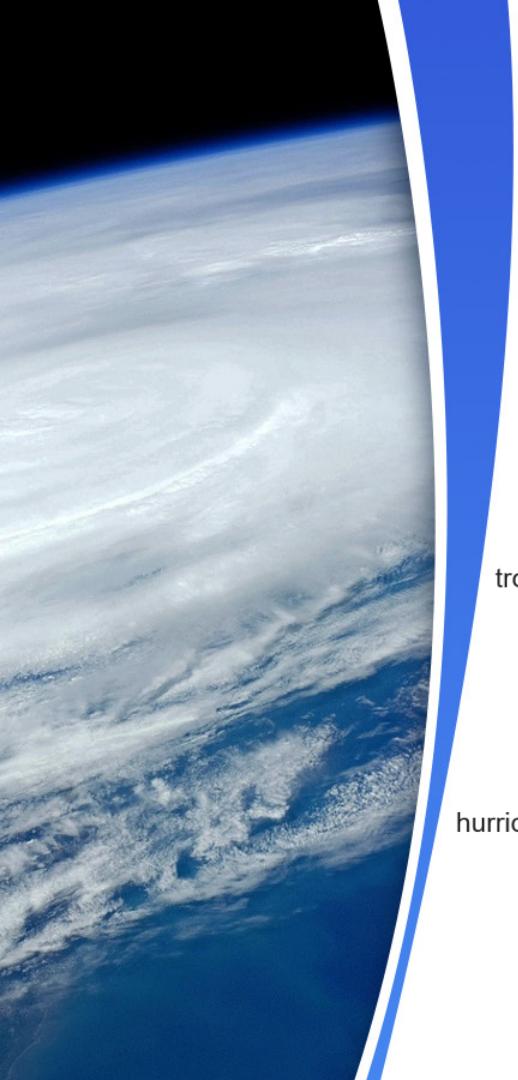
Top Features – On-Topic Tweets



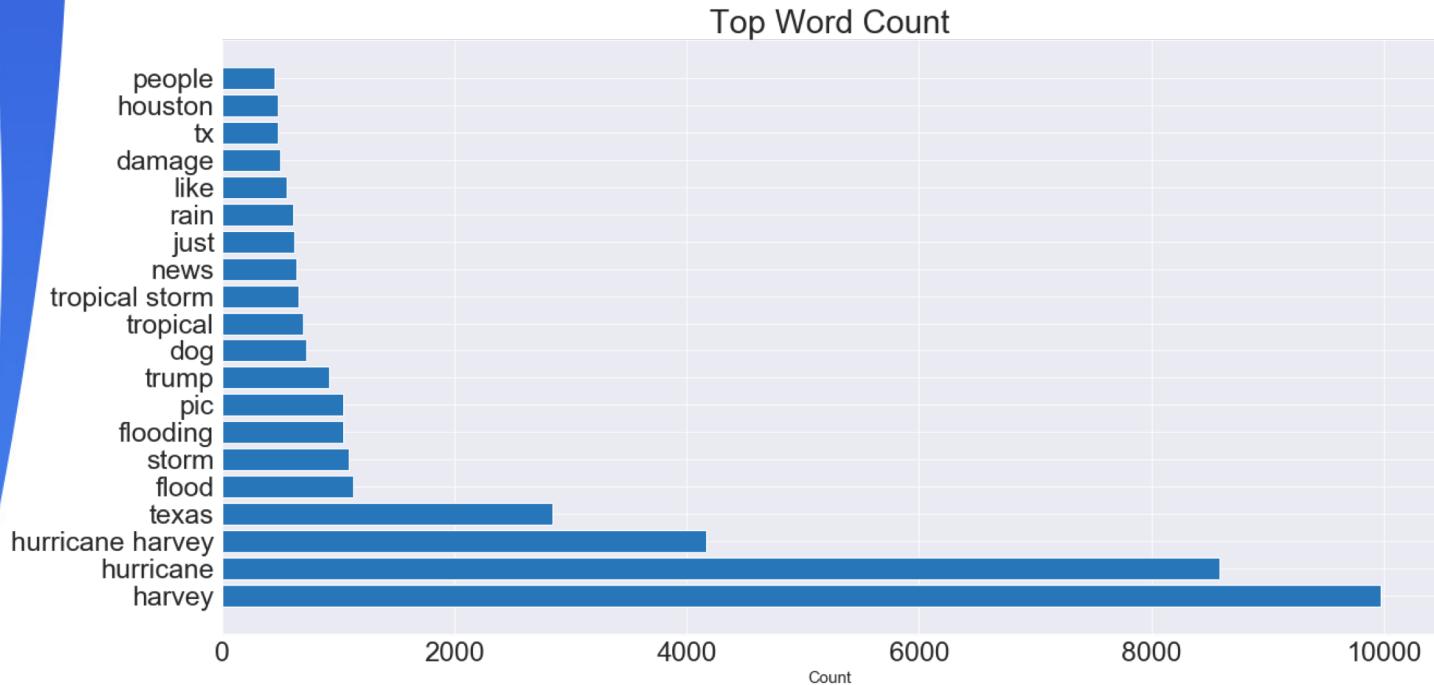


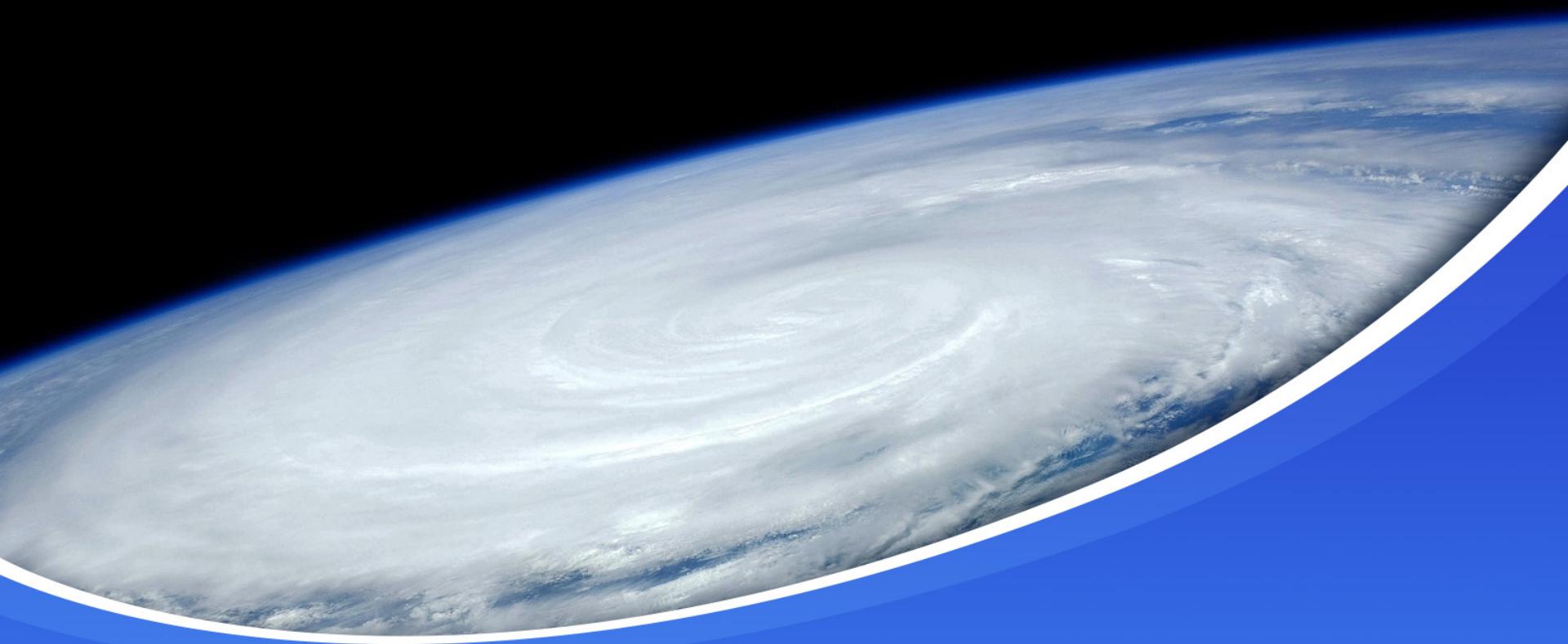
Model Testing

- Using our trained Logistic Regression model, we aimed to classify our newly collected tweets
 - 11158 on-topic, 6559 off-topic for Harvey
 - 191 on-topic, 3077 off-topic for Omaha
 - 96 on-topic, 11,831 off-topic for LA
- So why the discrepancies?

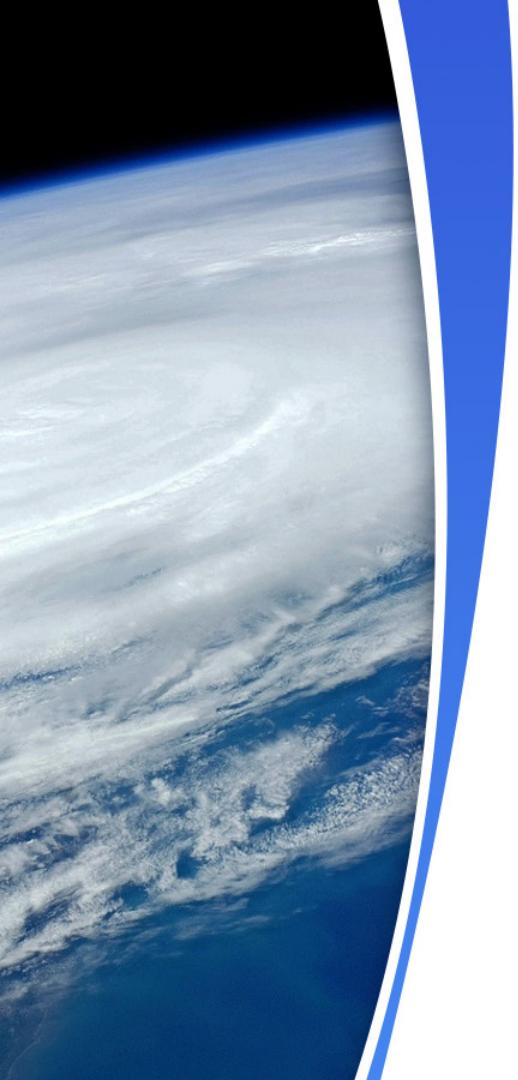


Top Features – Hurricane Harvey



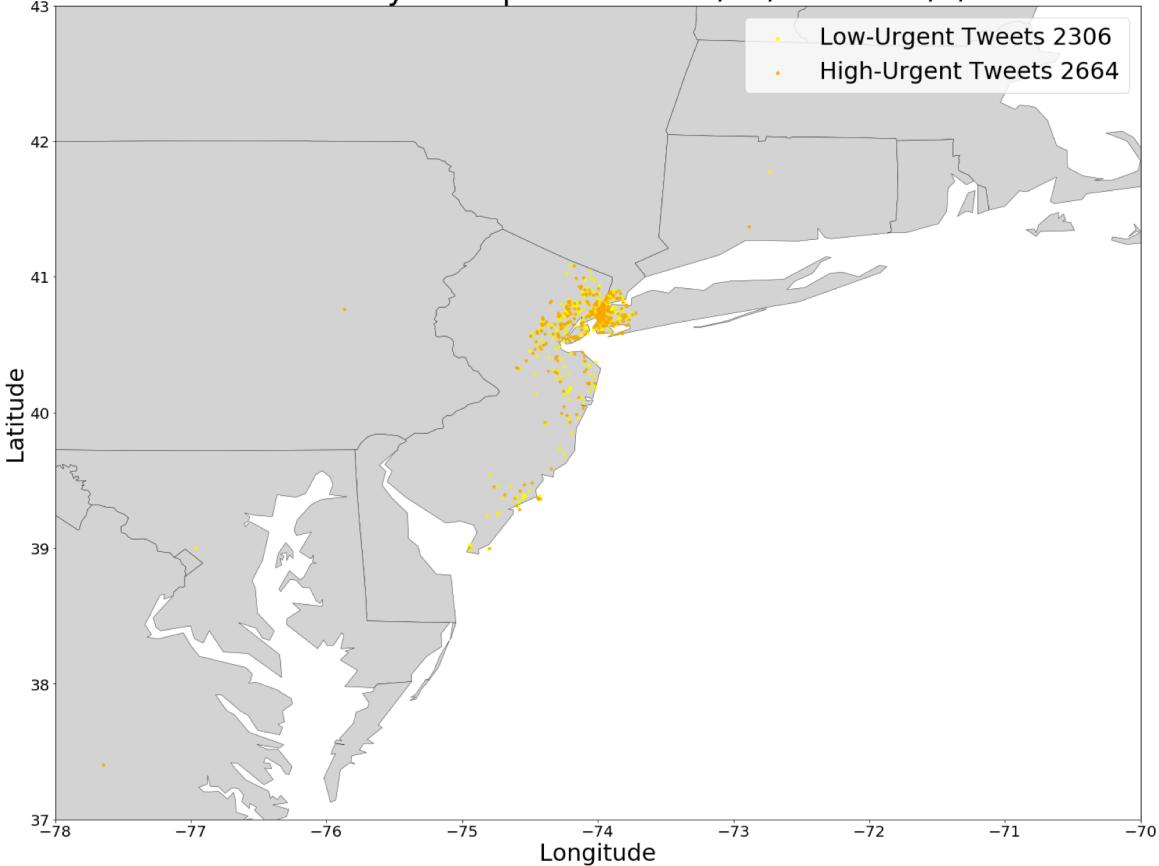


Mapping Posts



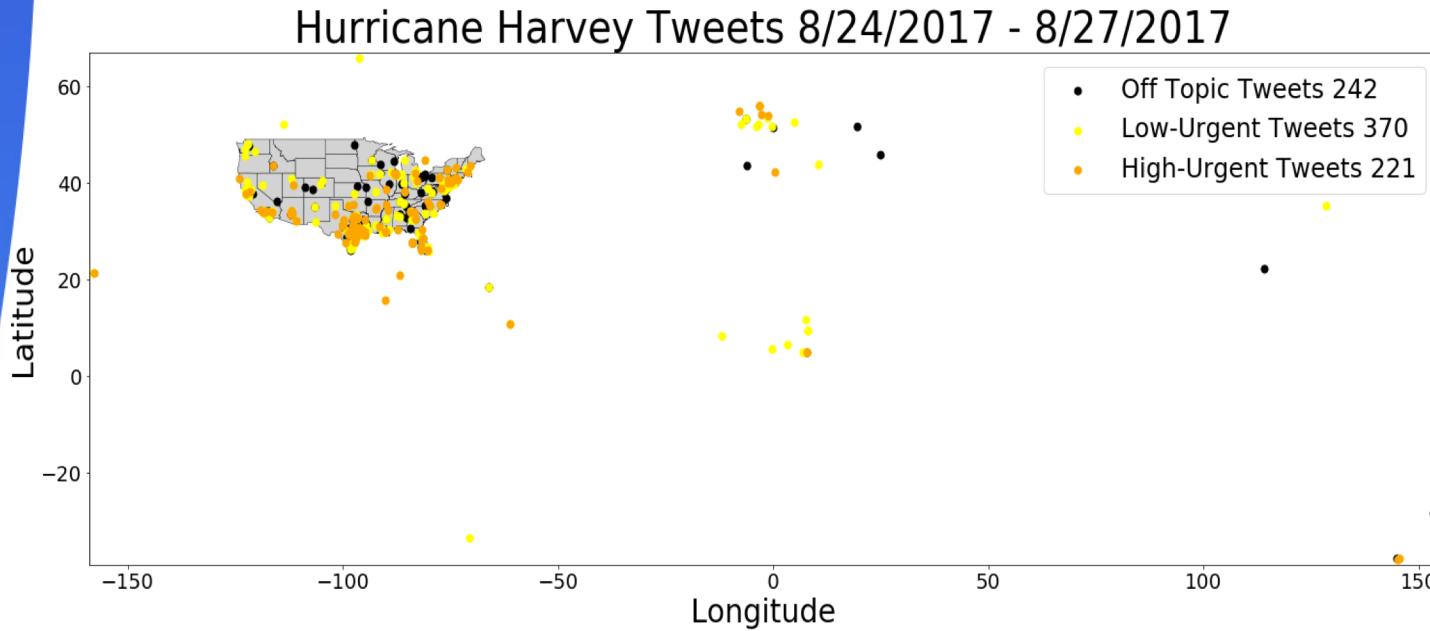
Disaster Mapping - Sandy

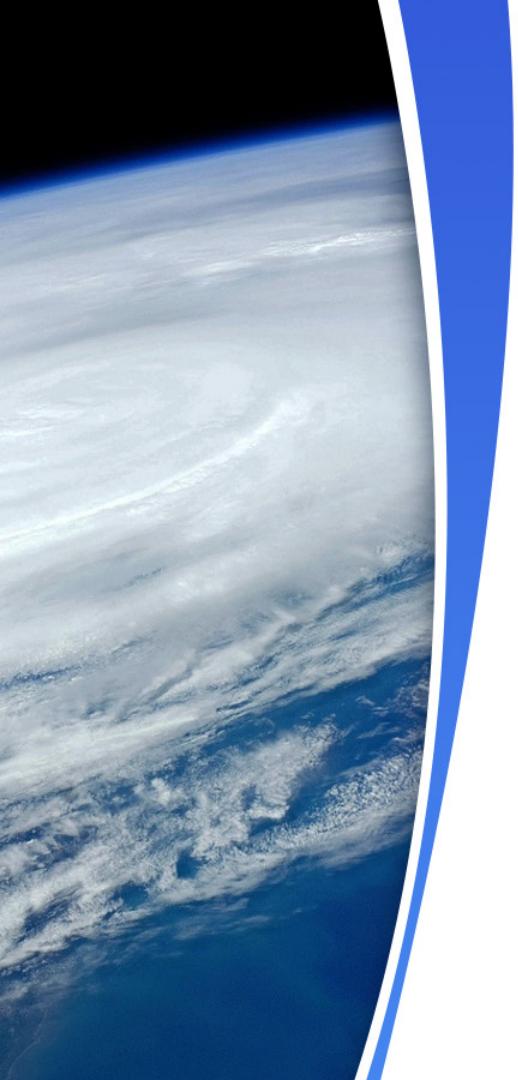
Hurricane Sandy on Topic Tweets 10/22/2012 - 11/2/2012



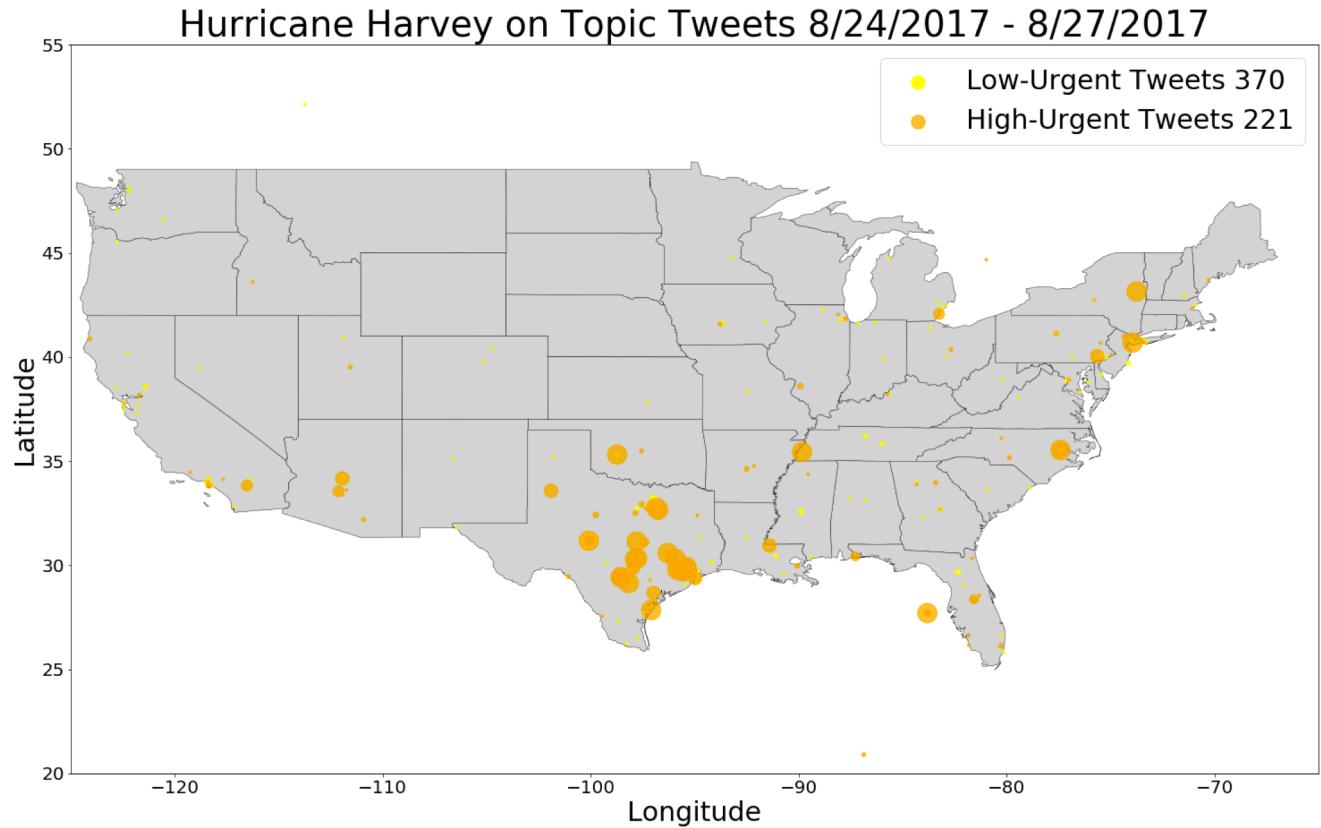


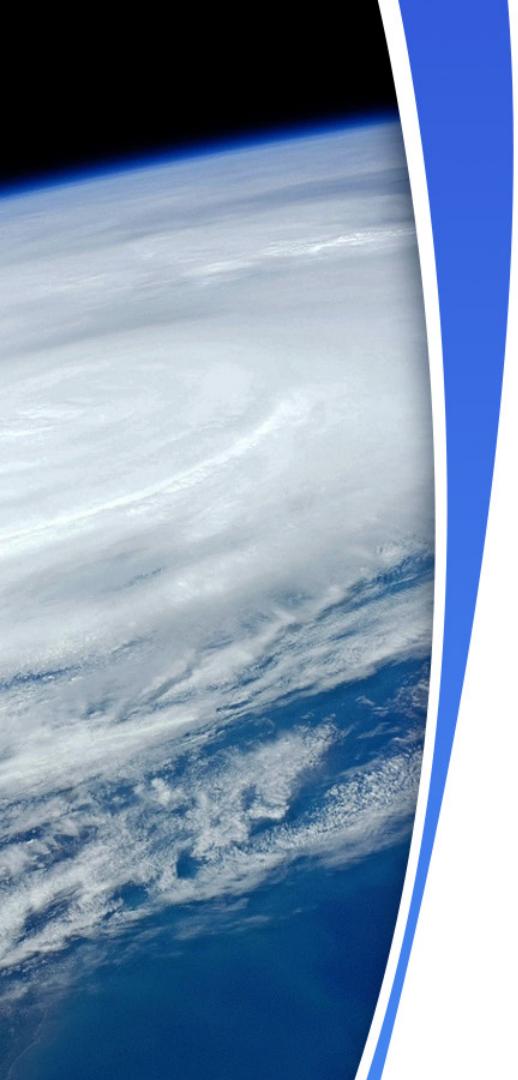
Disaster Mapping - Harvey





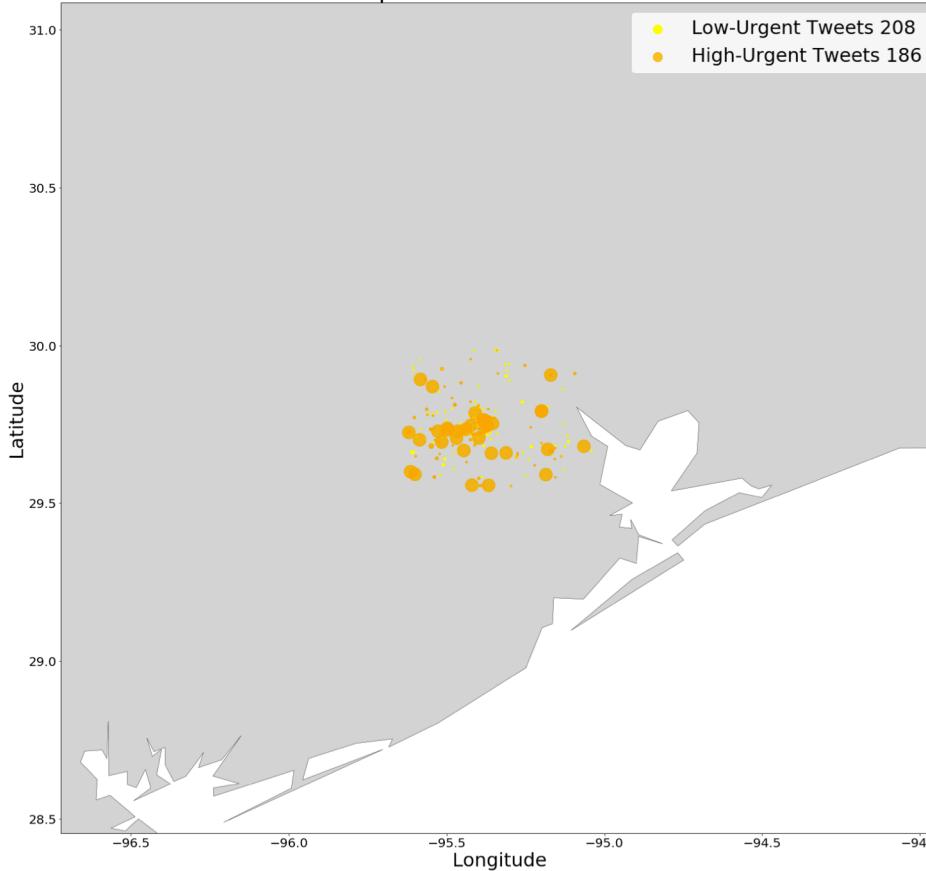
Disaster Mapping - Harvey

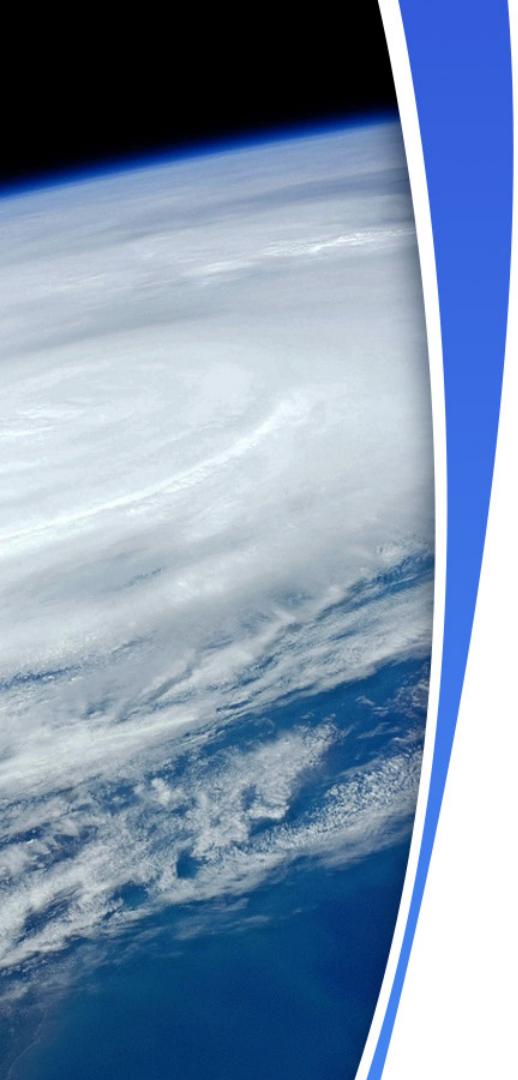




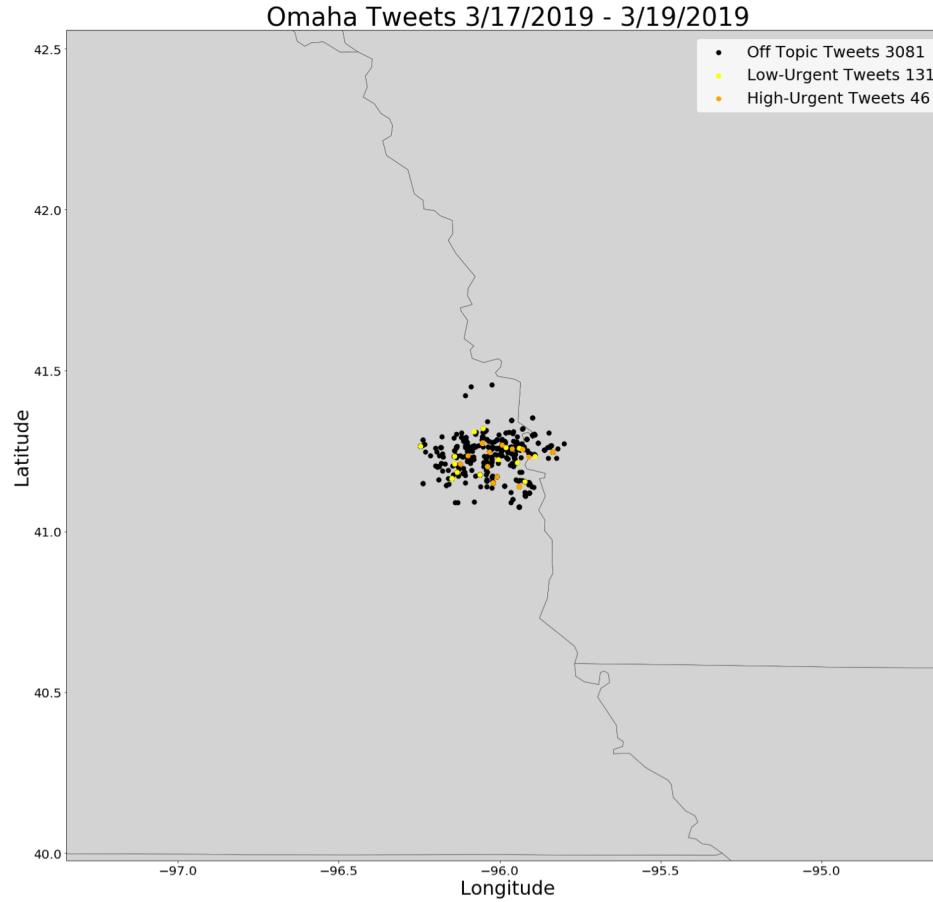
Disaster Mapping - Harvey

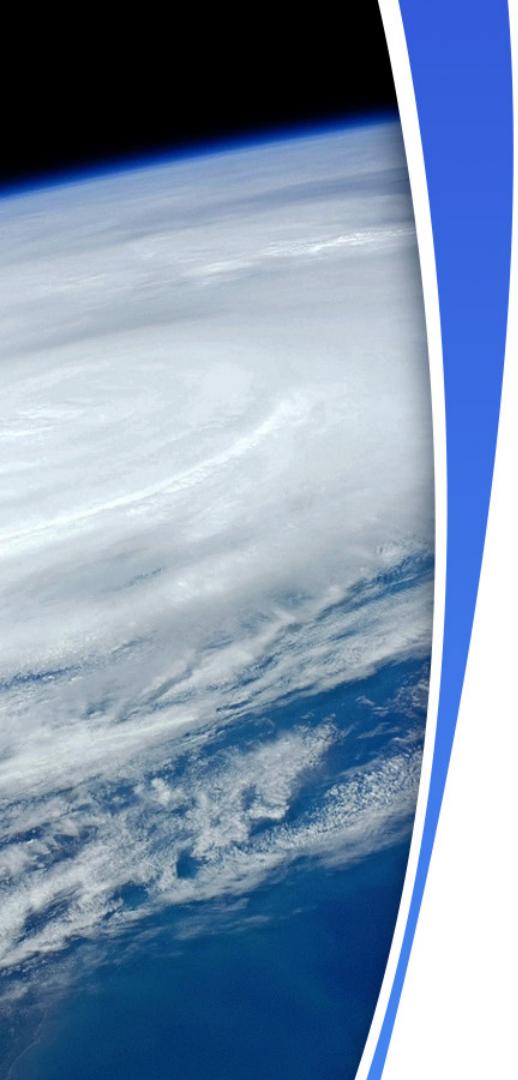
Houston on Topic Tweets 8/25/2017 - 8/28/2017





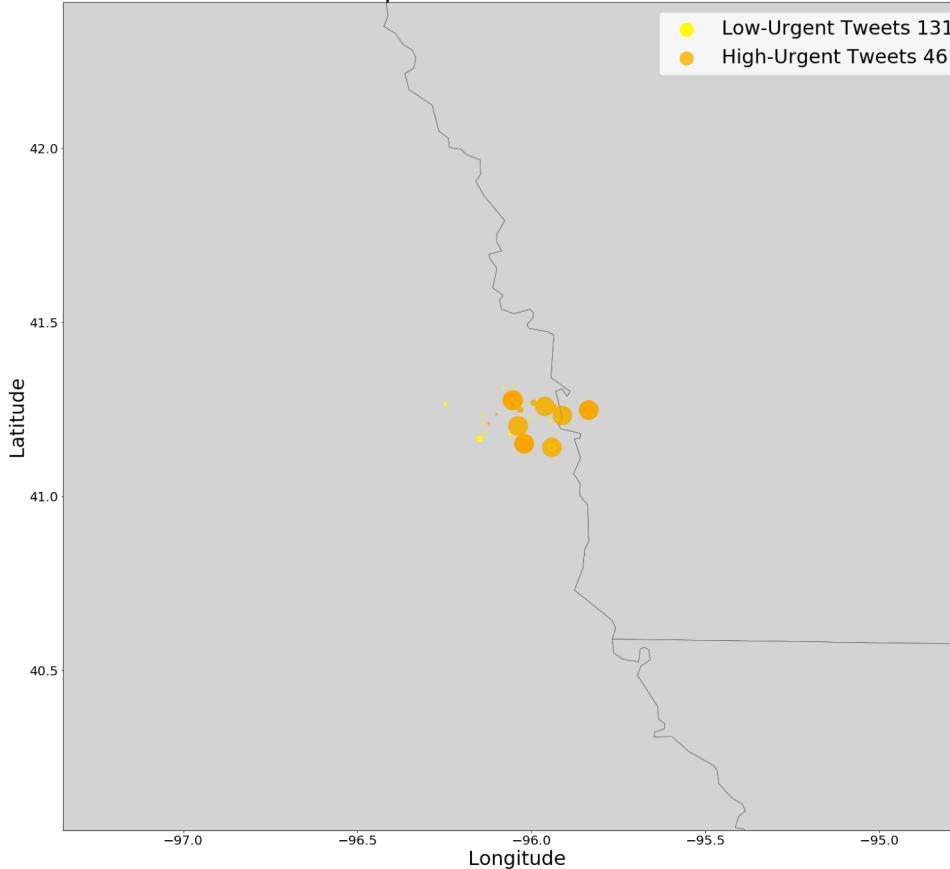
Disaster Mapping - Omaha





Disaster Mapping - Omaha

Omaha on Topic Tweets 3/17/2019 - 3/19/2019





Main Takeaways

- Natural Language Processing is extremely potent and useful when analyzing social media
- Social media can be a very beneficial tool
- Mapping social media posts is possible
- Building accurate and reproducible models takes time



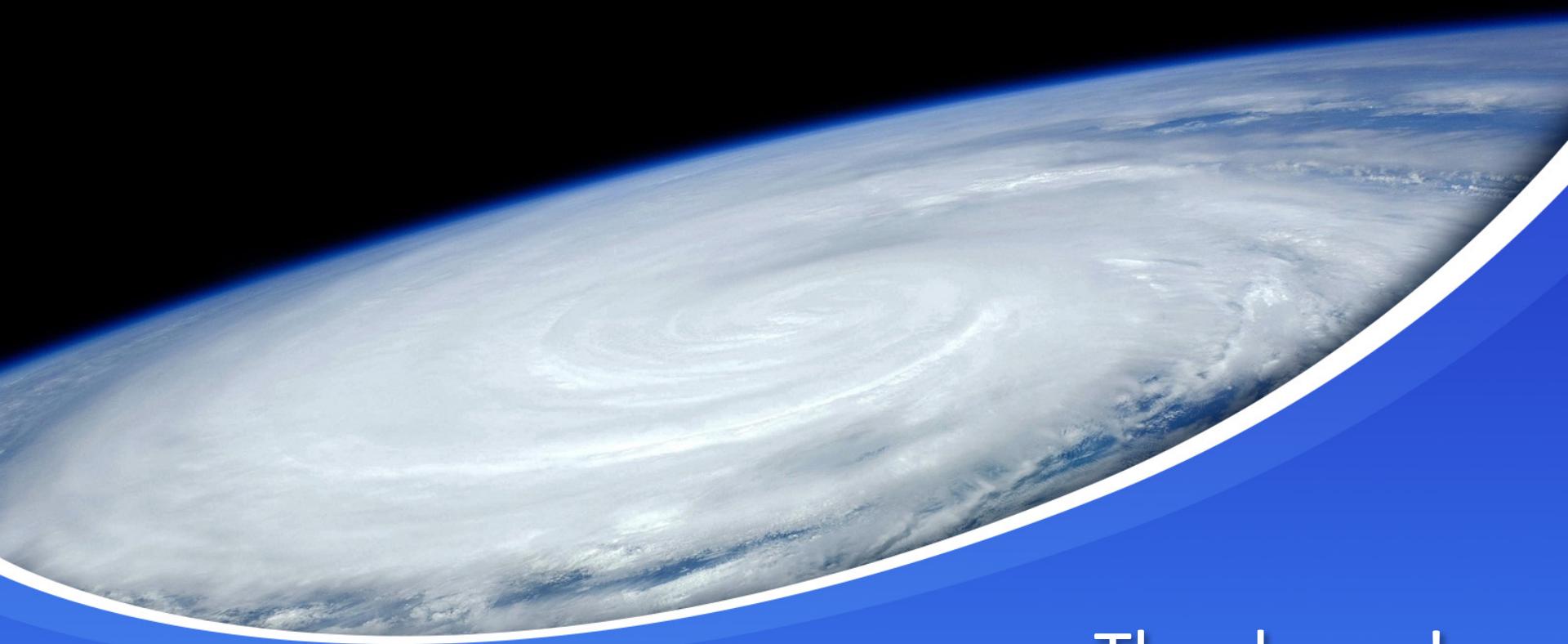
Biggest Hurdles

- Twitter's basic API provides limited information
- Models need to be trained on a specific disaster to provide ultimate efficacy
- Subject matter experts should weigh in with their expertise for maximum accuracy
- Timing



Improvements

- Use Subject Matter Experts to help craft more appropriate urgent / non-urgent lists of words
- Create a model for each specific natural disaster to help predict tweets being on-topic / off-topic
- Use a grid search to pull tweets from a wider range of areas to better map



Thank you!
Questions?

