# Twitter Disaster Behavior

April 30, 2020

**Group 15**

Kayley Bogemann, Shane Burchard, Jessie Butler, Austin Spencer, Taylor Thackaberry

**CS 4624: Multimedia, Hypertext, and Information Access**

Dr. Fox

Virginia Tech, Blacksburg, VA 24061

# Outline

Topics to be discussed:

1. Project Recap
2. Work Completed
   a. Project Timeline
   b. Topic Analysis
   c. Deliverable: Poster
3. Conclusions
4. Lessons Learned
5. Future Work
6. Acknowledgments
7. References

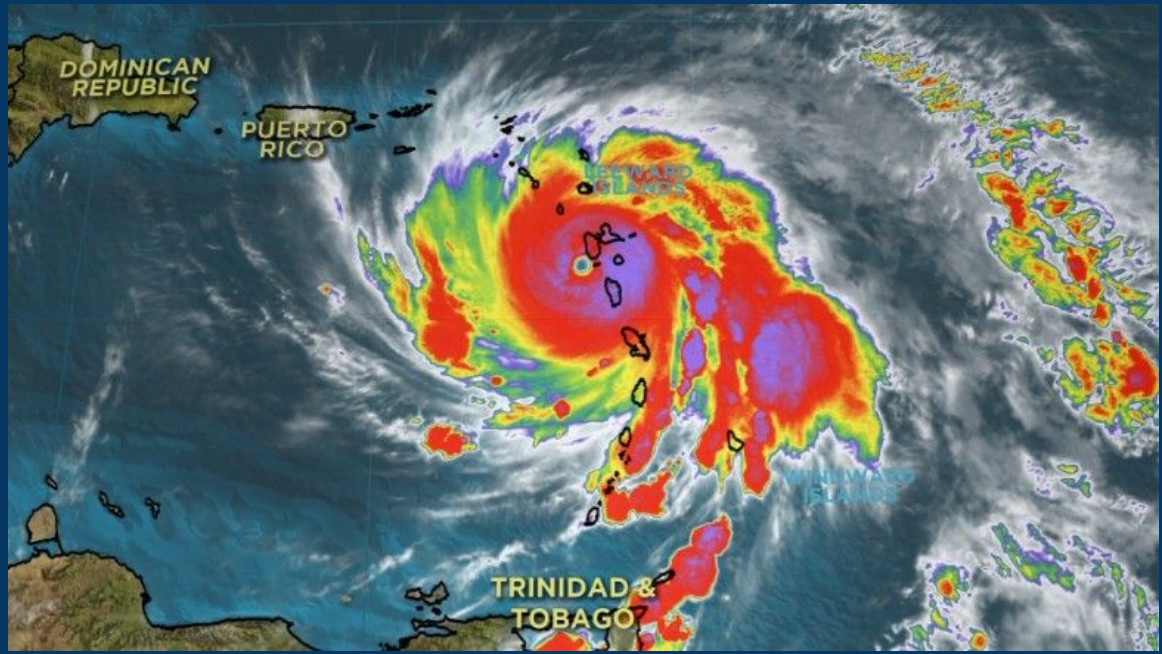



*Image Source: https://cdn.cnn.com/cnnnext/dam/assets/170919021311-hurricane-maria-screengrab-category-4-091917-0200-exlarge-169.jpg*

# Project Recap

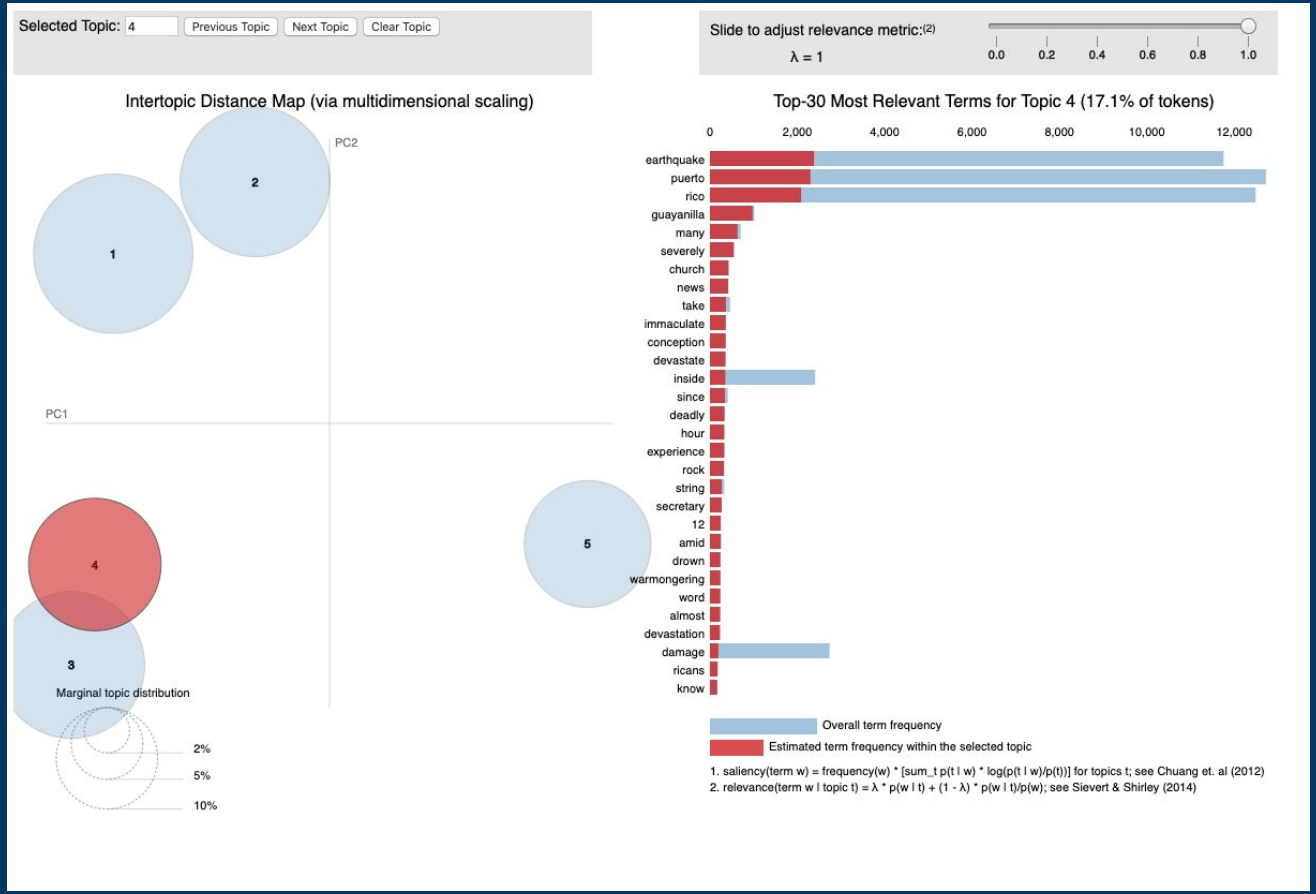
- Overall Goal: Identify patterns in behavior during and following natural disasters
  - Organize this data with graphs and maps
  - Communicate data to researchers
- Gather data from social media, specifically Twitter
- Our research specific to recent Puerto Rico earthquakes and Hurricane Maria

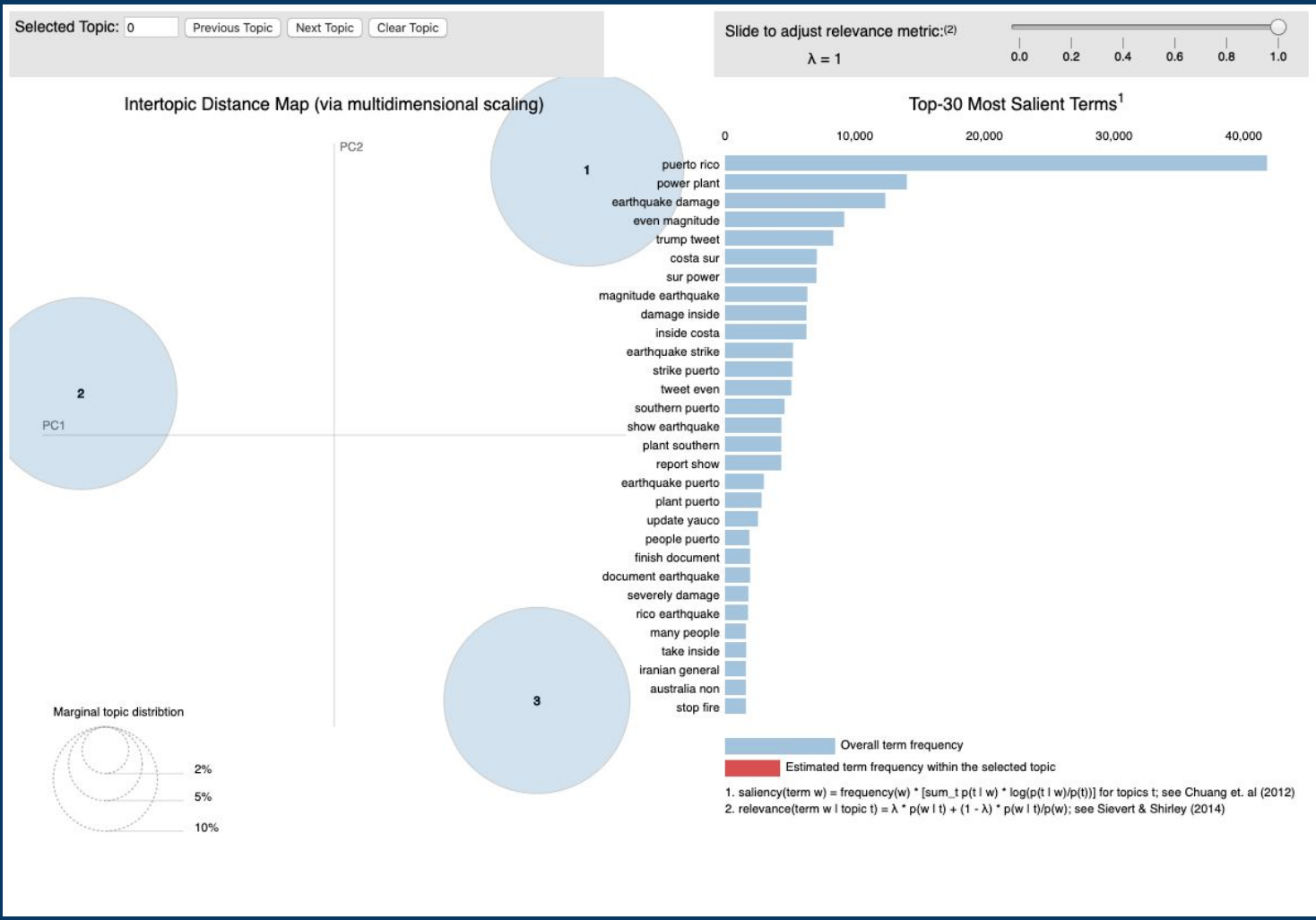
# Project Timeline

| Date | Description |
|---|---|
| January 29, 2020 | Began word frequency analysis on an initial, smaller set of data |
| February 4, 2020 | Created smaller dataset of tweets containing geotags |
| March 2, 2020 | Completed word frequency analysis for final datasets of earthquake and hurricane twitter data |
| March 4, 2020 | **Completed draft of ISCRAM poster** |
| March 30, 2020 | Created map of geotag location changes during and after earthquake and hurricane |
| April 6, 2020 | Conducted topic analysis on all tweets related to earthquake |
| April 20, 2020 | Completed analysis on topics discussed by individuals vs organizations |
| April 26, 2020 | **Initial submission to VTechWorks** |
| April 30, 2020 | **Final presentation** |

# Topic Analysis

- Identified most common topics in tweets
- Separated tweets from individuals vs. brands/companies
- Visualization identifies 3-5 topic groups for 1-word, 2-word, and multi-word topics



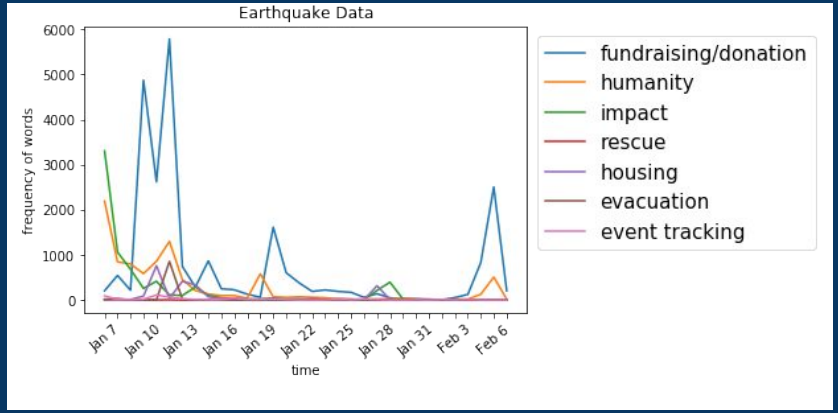

Selected Topic: 0
Previous Topic
Next Topic
Clear Topic
Slide to adjust relevance metric:(2)
λ = 1
0.0
0.2
0.4
0.6
0.8
1.0
Intertopic Distance Map (via multidimensional scaling)
PC2
PC1
1
2
3
Marginal topic distribtion
2%
5%
10%
Top-30 Most Salient Terms1
0
10,000
20,000
30,000
40,000
puerto rico
power plant
earthquake damage
even magnitude
trump tweet
costa sur
sur power
magnitude earthquake
damage inside
inside costa
earthquake strike
strike puerto
tweet even
southern puerto
show earthquake
plant southern
report show
earthquake puerto
plant puerto
update yauco
people puerto
finish document
document earthquake
severely damage
rico earthquake
many people
take inside
iranian general
australia non
stop fire
Overall term frequency
Estimated term frequency within the selected topic
1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

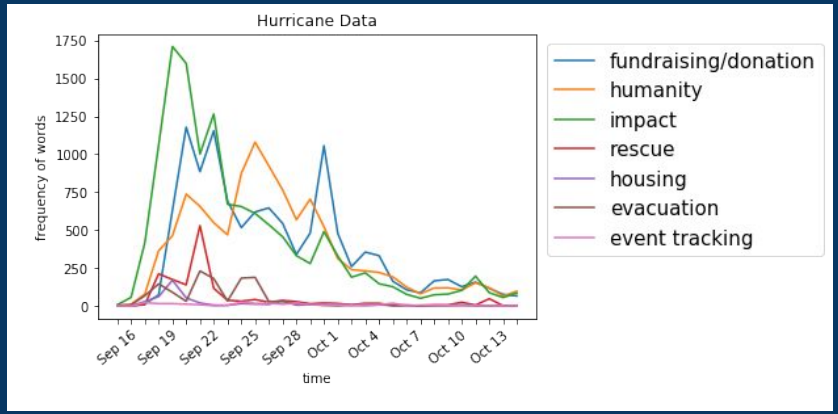
# Topic Analysis - Earthquake

| | Individual | Brands / Official Accounts |
|---|---|---|
| One Word | power, electricity, magnitude, aftershock, devastation, health | emergency, damage, deadly, response, tremor, electricity |
| Two Word | church guayanilla, people homeless, trump administration, worst earthquake | power plant, without power, state emergency, damage inside, public health, |
| Multi Word | earthquake aftershock power, thousands residents sleep car, waiting federal aid | public health emergency response, thousands people homeless, power outage effects, withhold aid |

# Deliverable: Poster

- Small visual updates
- Updated graphs to highlight interesting trends
  - “Prepare” didn’t make sense for an earthquake
- Submitted findings in document form to ISCRAM website
- Poster submitted along with final report to VTechWorks



# Tweet Comparison for Puerto Rico Earthquake and Hurricane Maria

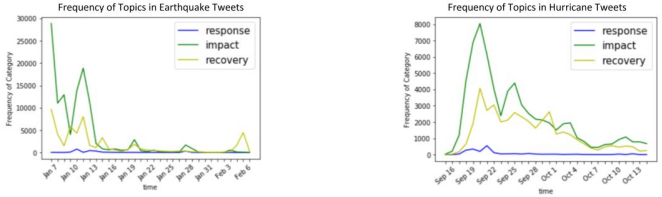
Kayley Bogemann, Shane Burchard, Jessie Butler, Austin Spencer, Taylor Thackaberry

## Introduction

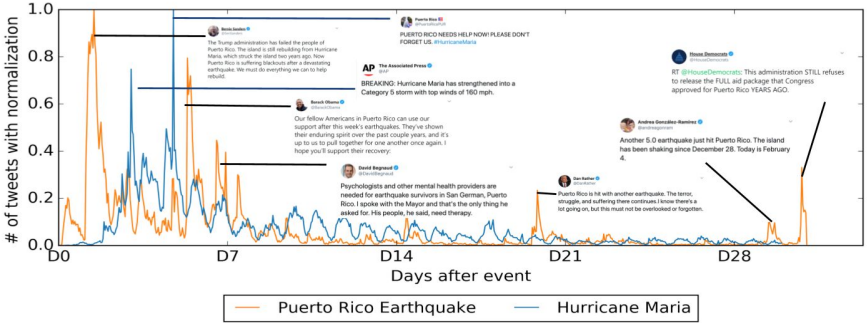
- Since December 2019, Puerto Rico experienced several **earthquakes** > 5.0 magnitude.
- By **analyzing over 300,000 tweets**, we plan to **identify behavioral patterns.**
- We consider the **geotags** associated with tweets and **key words or topics in tweets.**
- Similar analysis is run on tweets from **Hurricane Maria** to compare the behavioral patterns of tourists and locals of each event.

## Methods

- **Collecting Tweets** pertaining to disaster topics
- Using **Python** to analyze data using a Tweet parser
- **Graphing** data, such as
  - Frequency of Tweets
  - Frequency of Topics
- Analyzing **popular topics** mentioned in Tweets




*Frequency of Topics for the Puerto Rico Earthquakes and Hurricane Maria*




## Data

- **402,016 Tweets** related to Puerto Rico Earthquake between January 7, 2019 and February 6, 2019
- **317,214 Tweets** related to Hurricane Maria between September 15, 2017 and October 14, 2017

## Conclusions

- **Hurricane Maria's arrival was forecasted**, resulting in a larger corpus of tweets about it occuring the days leading up to and during the event.
- **The earthquakes were not predicted**, and sporadic aftershocks meant that sharp spikes of activity started every time there was another earthquake.
- The overall decrease in discussion over the time period is **likely due to external (mainland US) lack of interest or relevance.**

## Acknowledgement

NSF CMMI-1638207 CRISP: Collaborative Research: Coordinated, Behaviorally-Aware Recovery for Transportation and Power Disruptions

*The above maps illustrate Hurricane Maria's path compared to a heatmap of Twitter activity*

# Conclusions

- Difficult to filter out irrelevant information on Twitter
- Forecasted events, such as Hurricane Maria
  - Much more data leading up to the event
  - Geolocation of tweets showed the path Hurricane Maria followed
- Unpredicted events, such as earthquakes
  - Sharp spikes of activity with the sporadic aftershocks
- Tweet activity overall decreases over the time period

*The above maps illustrate Hurricane Maria's path compared to a heatmap of Twitter activity*

# Lessons Learned

- Importance of communication
  - GroupMe
  - Important to give updates as soon as possible
- Transition to Zoom meetings was difficult
  - Varying levels of internet quality
  - Not all team members could attend all meetings
- Creating clearer project goals and deliverables
  - Planning ahead

# Future Work

- Analysis we conducted could be expanded
  - COVID-19, future natural disasters
- Storage of data into SQL Database
  - Python analysis ran slow
- Implementation of other Natural Language Processing Techniques
  - Our topic analysis used pyLDAvis.gensim, nltk
  - spaCy
  - Prepare the dataset for deep learning

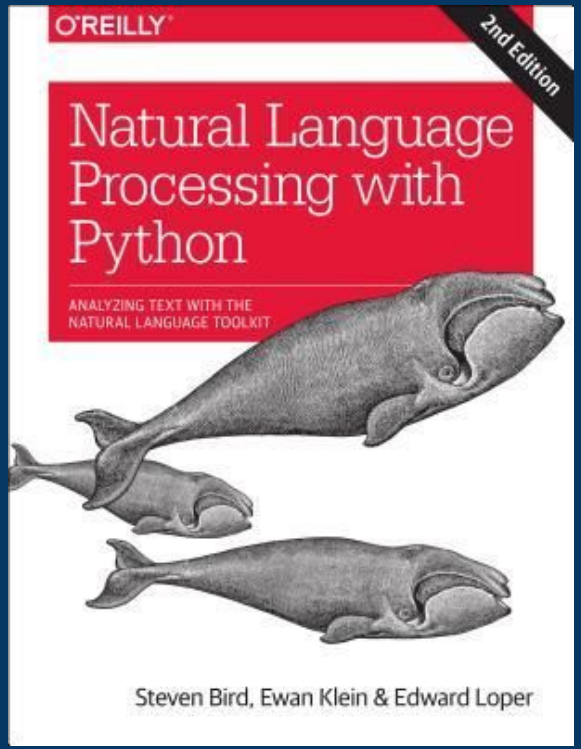



Image: https://www.amazon.com/Natural-Language-Processing-Python-Analyzing/dp/0596516495

# Acknowledgements

- Professor: Dr. Edward A. Fox
- Client: Alice Song
- “Collaborative Research: Coordinated, Behaviorally-Aware Recovery for Transportation and Power Disruptions” (NSF-CMMI-1638207)


# References


- Work from previous semester with Hurricane Irma (Google Drive)