Leveraging Social Media to Map Disasters

By: Mimi Kim, Samuel Leadley, Jeremy Ondov, & Yisroel Len

Two Facts:

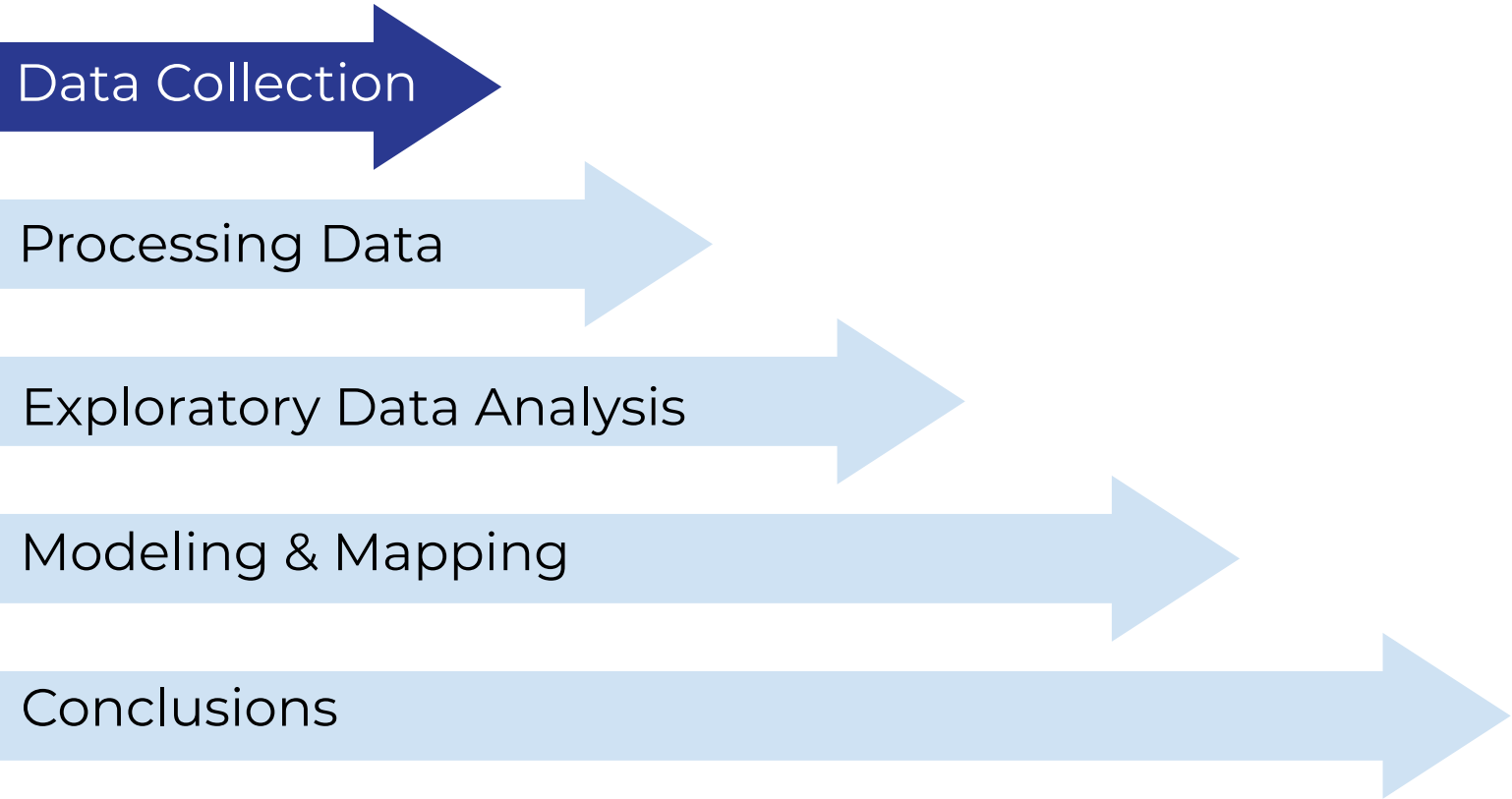
1. During a natural disaster it is essential for first responders to have access to victims' locations.

2. People often use social media to report their status.

The Goal:

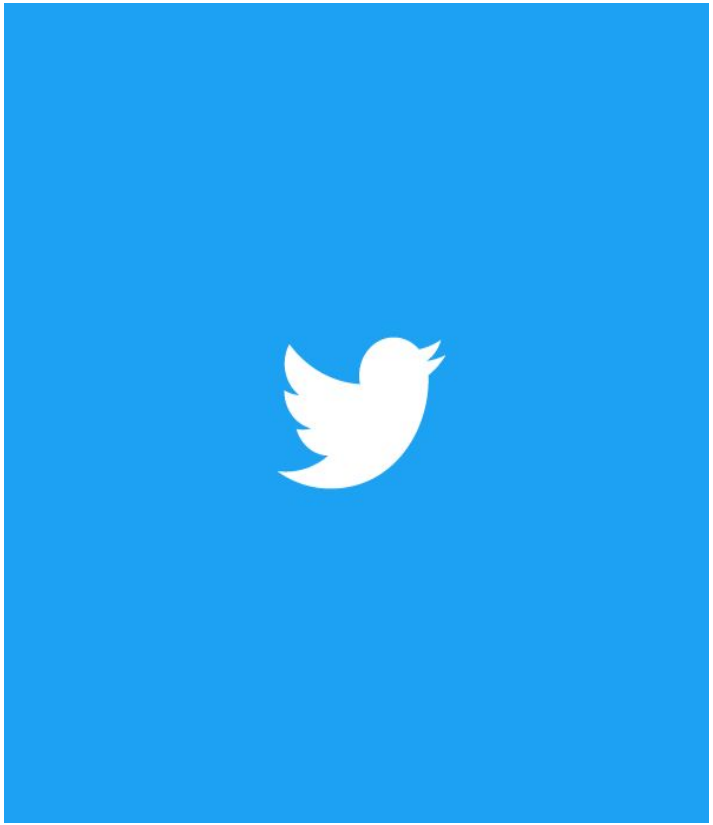
Leverage social media to locate people in need of emergency disaster relief.

Process


Data Collection
Processing Data
Exploratory Data Analysis
Modeling & Mapping
Conclusions

Data Collection

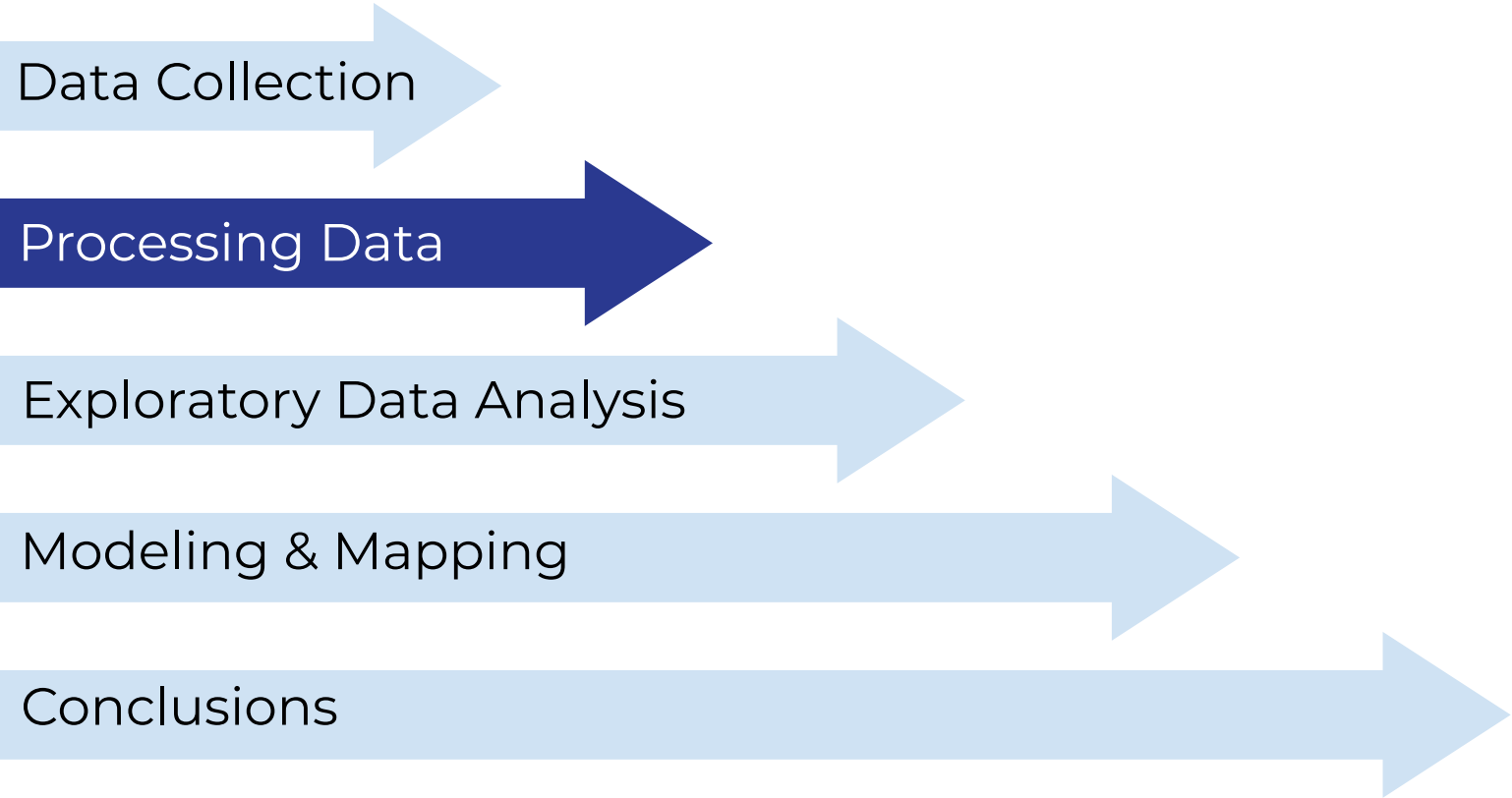
- Multiple social media options: Twitter, Facebook, Instagram
- Twitter’s basic API is restrictive
- Third party package, Twitterscraper was used



Disaster Sourcing

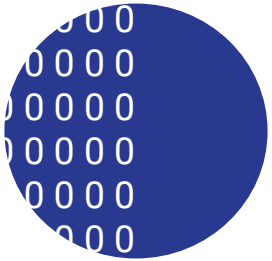
- To make a robust model, multiple disaster events were targeted
  - Hurricane Harvey (Aug-Sept 2017)
  - Montecito Mudslides (Jan 2018)
  - Southern Tornados (April 2019)
  - Noreaster (March 2018)
  - Floods (July 2019)
- Total tweets collected: 22,862

Process

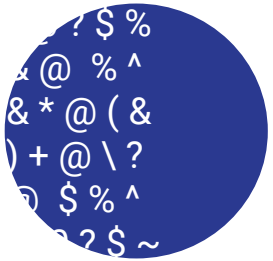

Data Collection
Processing Data
Exploratory Data Analysis
Modeling & Mapping
Conclusions

NLP Preprocessing

Drop Duplicates



Check for Nulls



Remove Unnecessary Punctuation

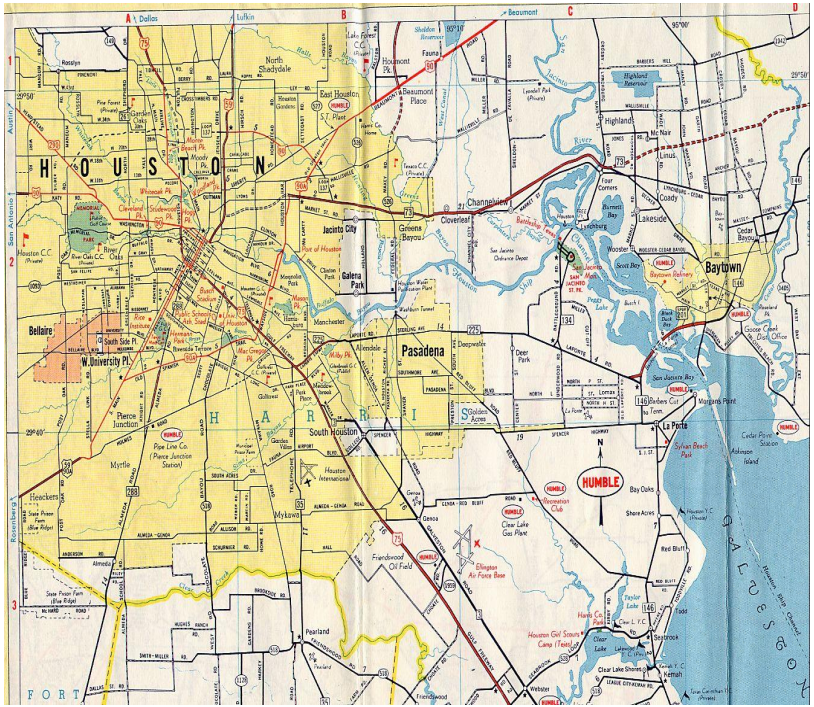
Creating Target Variable

- Create critical bag of words

  E.g. medivac, sos, & save me

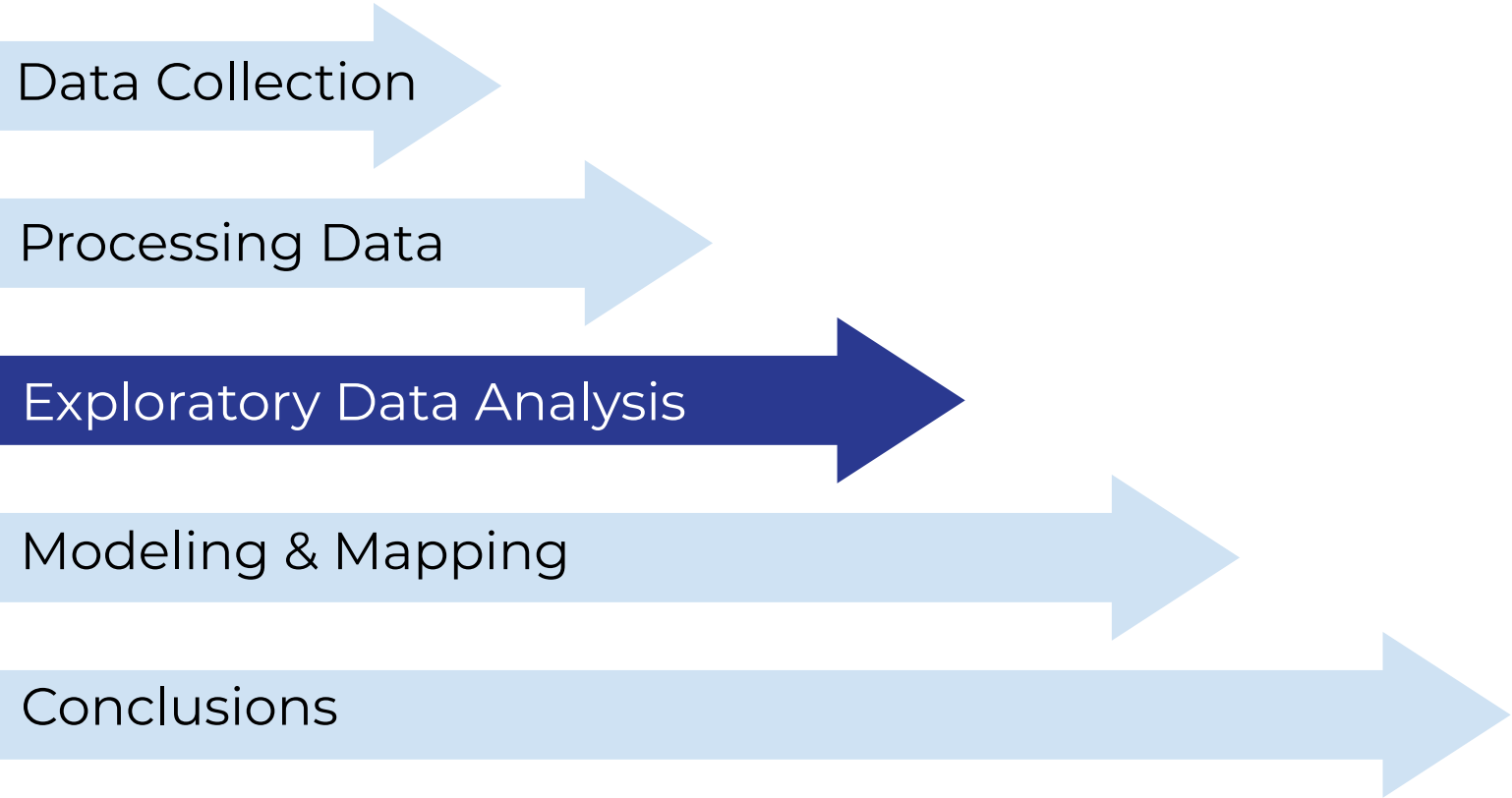
- Mapped to tweets

Assigning Locational Data

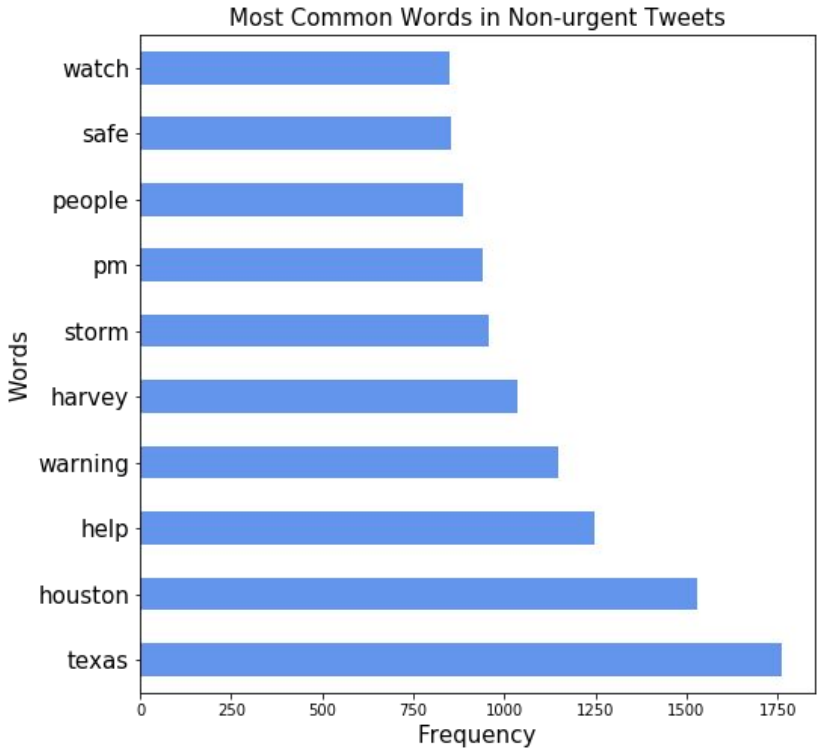


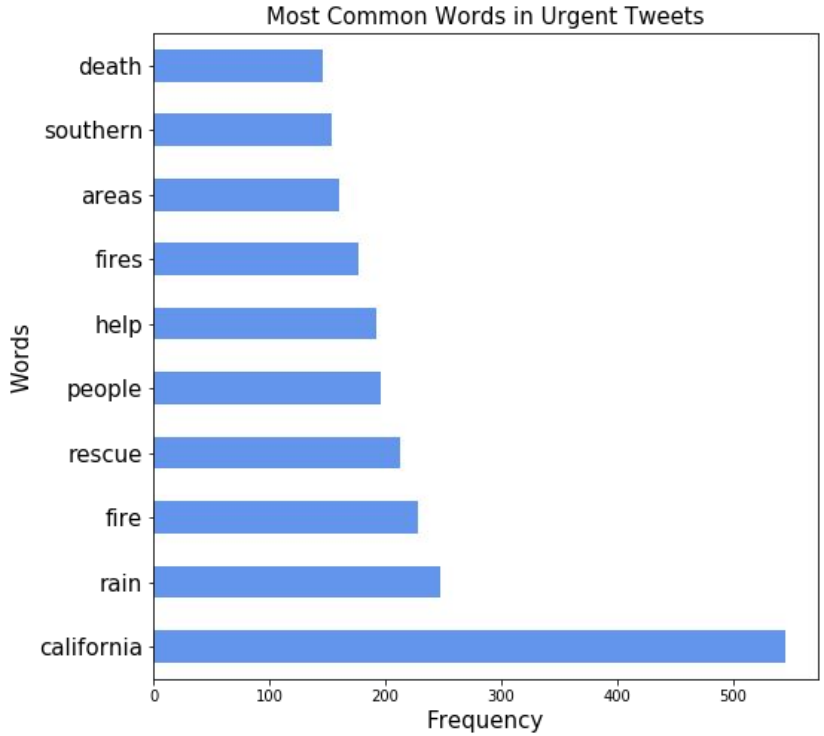
- Geolocation was not available for our tweets
- Locations were randomly assigned to tweets
- Five areas were created to simulate concentrated areas

Process

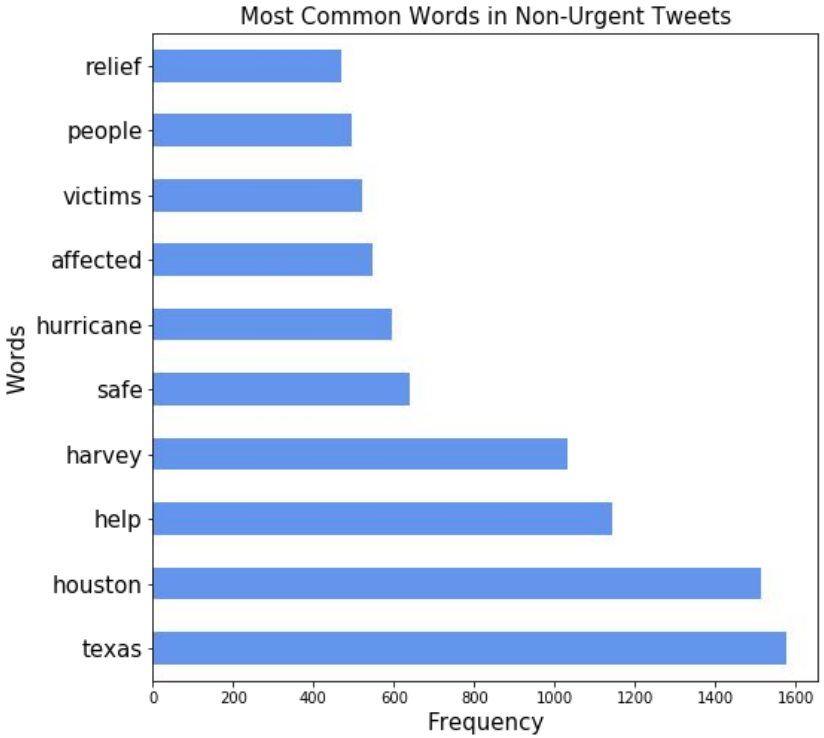

Data Collection
Processing Data
Exploratory Data Analysis
Modeling & Mapping
Conclusions

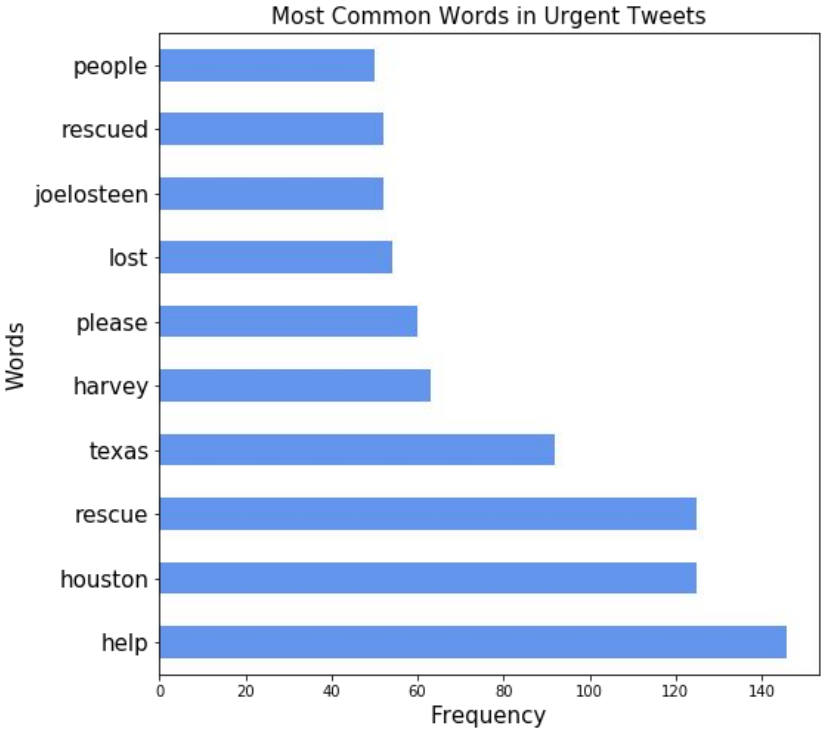
Total Dataset





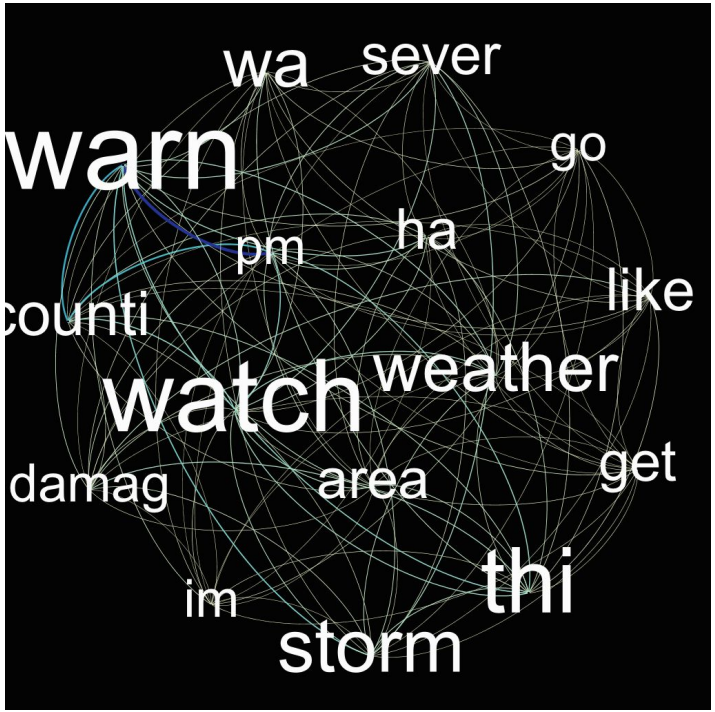
Hurricanes


Most Common Words in Non-Urgent Tweets
relief
people
victims
affected
hurricane
safe
harvey
help
houston
texas
Words
0
200
400
600
800
1000
1200
1400
1600
Frequency

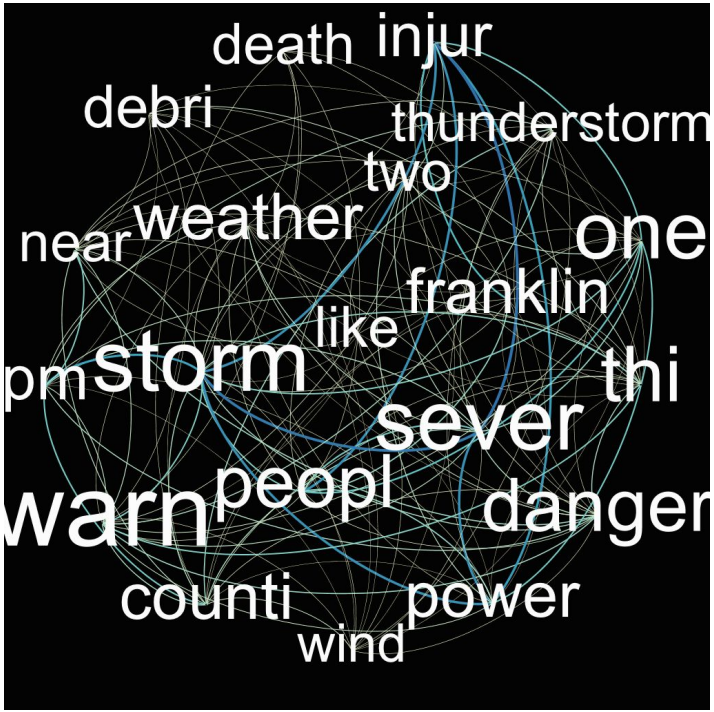

Most Common Words in Urgent Tweets
people
rescued
joelosteen
lost
please
harvey
texas
rescue
houston
help
Words
0
20
40
60
80
100
120
140
Frequency

Tornado

Non Urgent

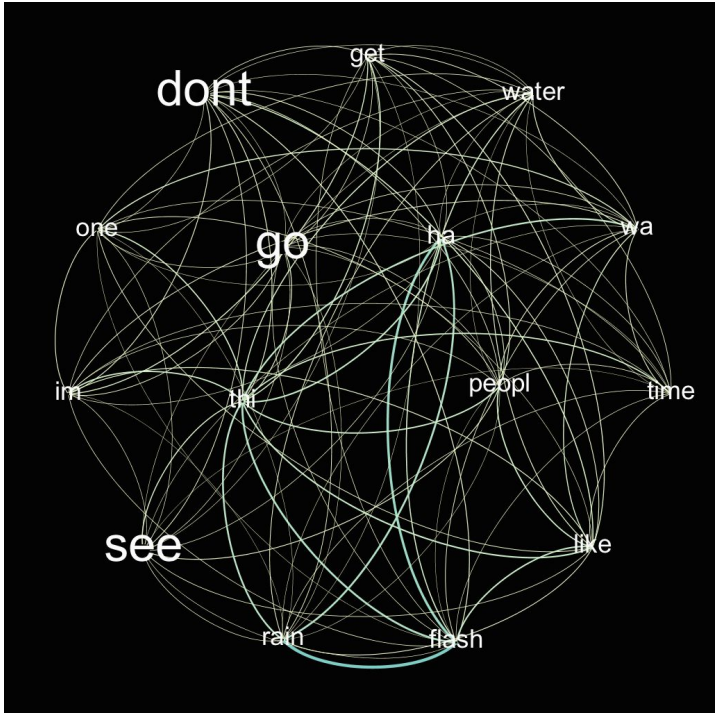


Urgent



Floods

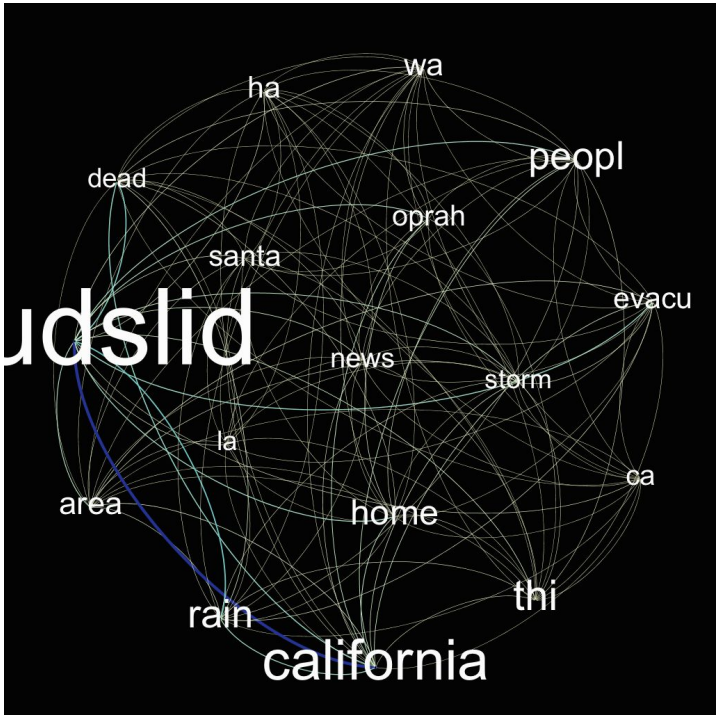
Non Urgent



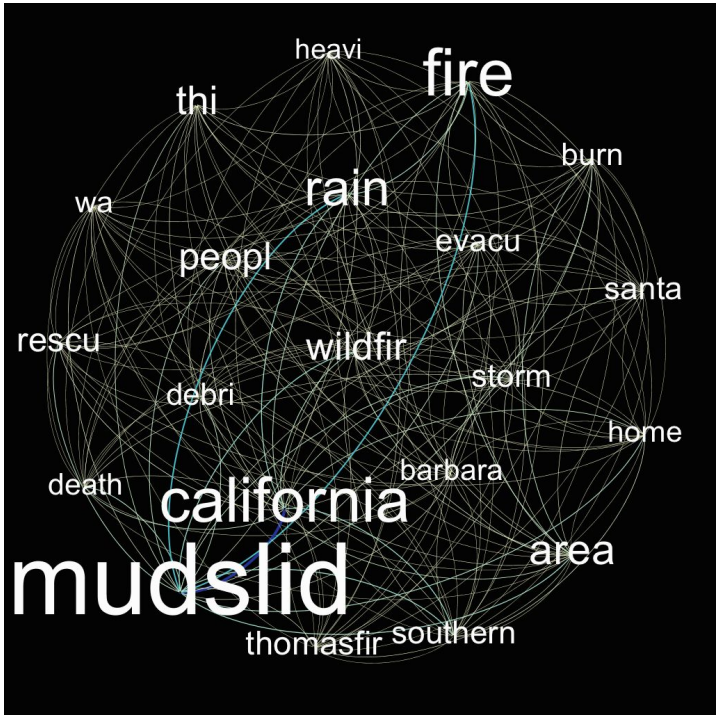
Urgent

Mudslides

Non Urgent

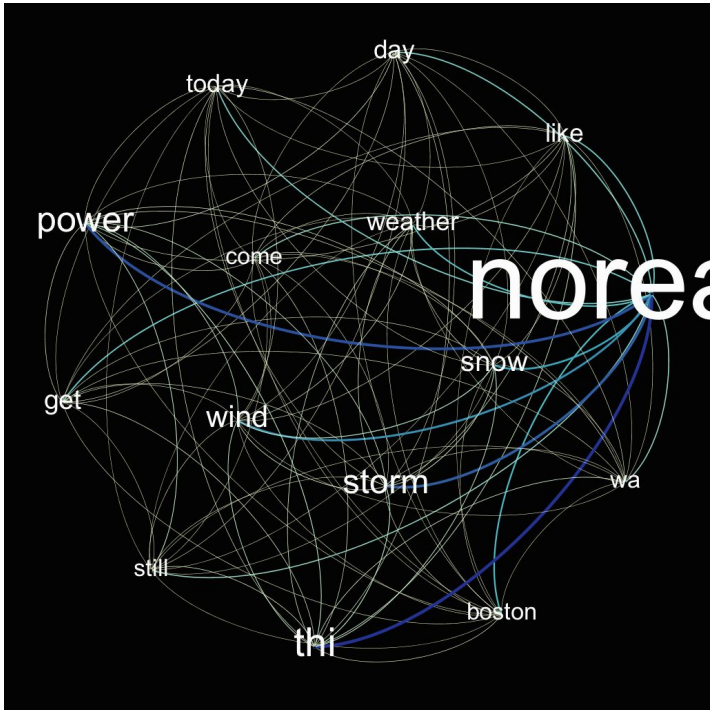


Urgent

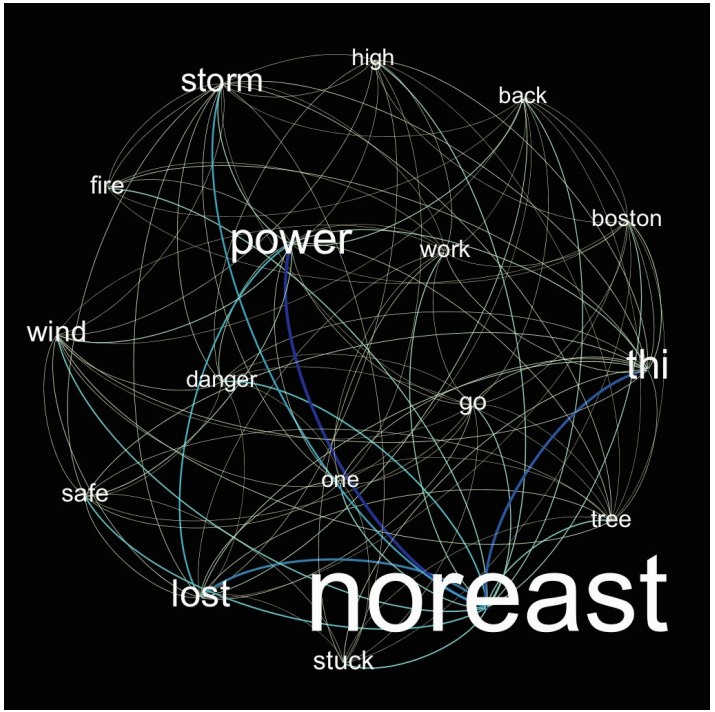


Noreaster

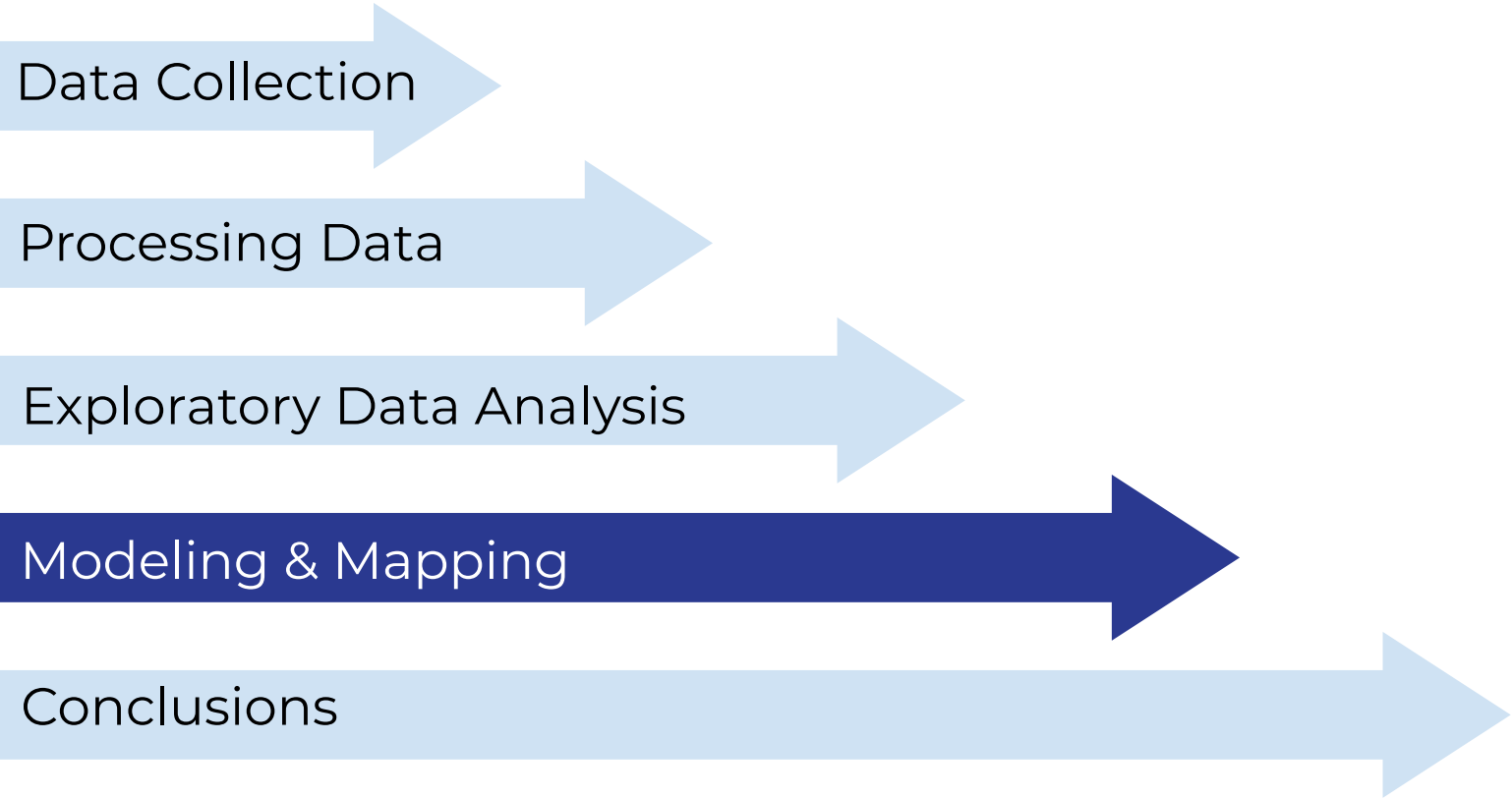
Non Urgent



Urgent



Process


Data Collection
Processing Data
Exploratory Data Analysis
Modeling & Mapping
Conclusions

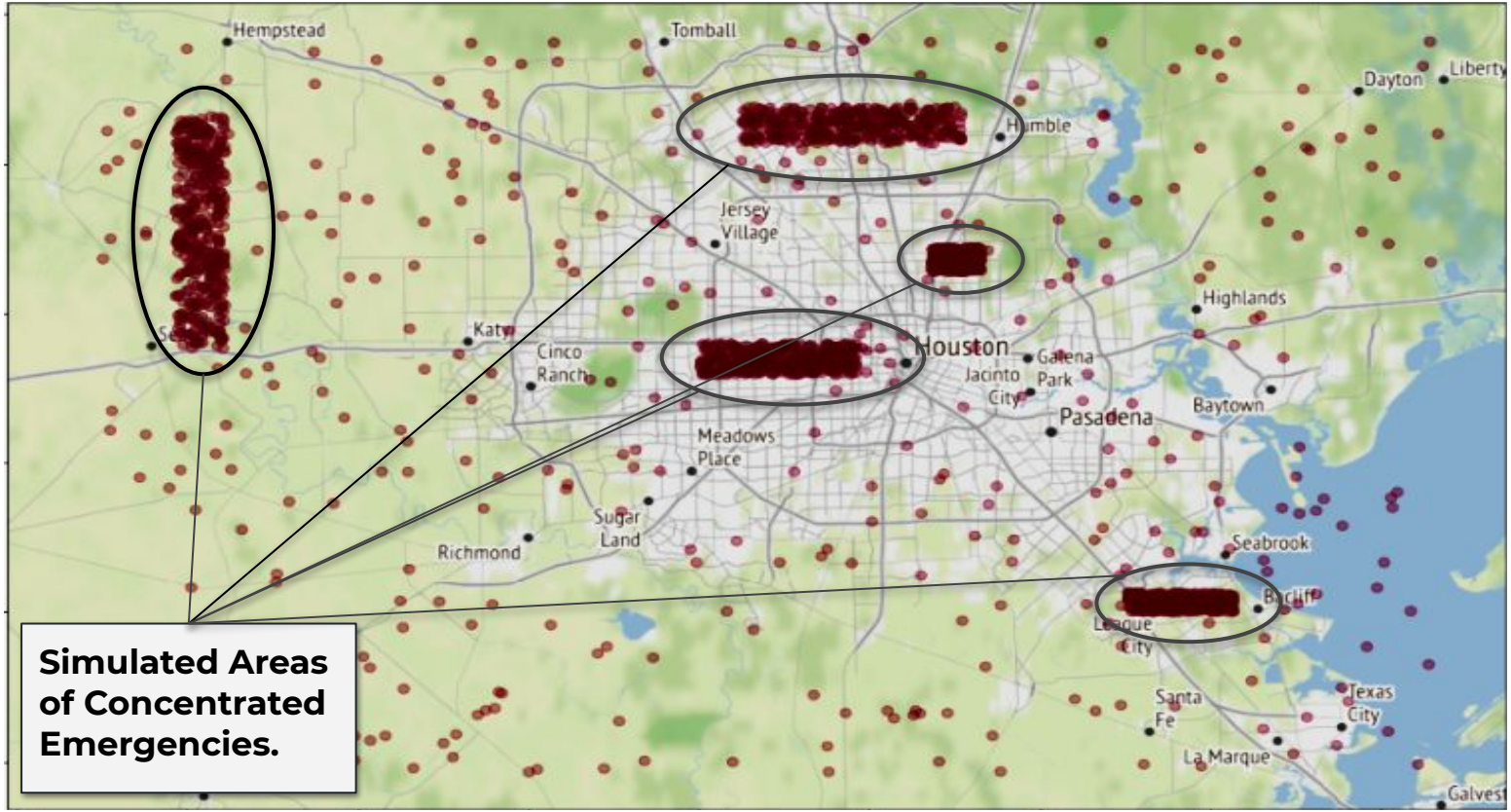
Model Selection

- Several classification models were used
  - Logisitc regression, random forest, adaptive boosting
- CountVectorizer was used with a custom set of stop words
  - Including top words shared by positive/negative classes
- GridSearchCV was used to search through large sets of parameters
- Best model was determined to be AdaBoost

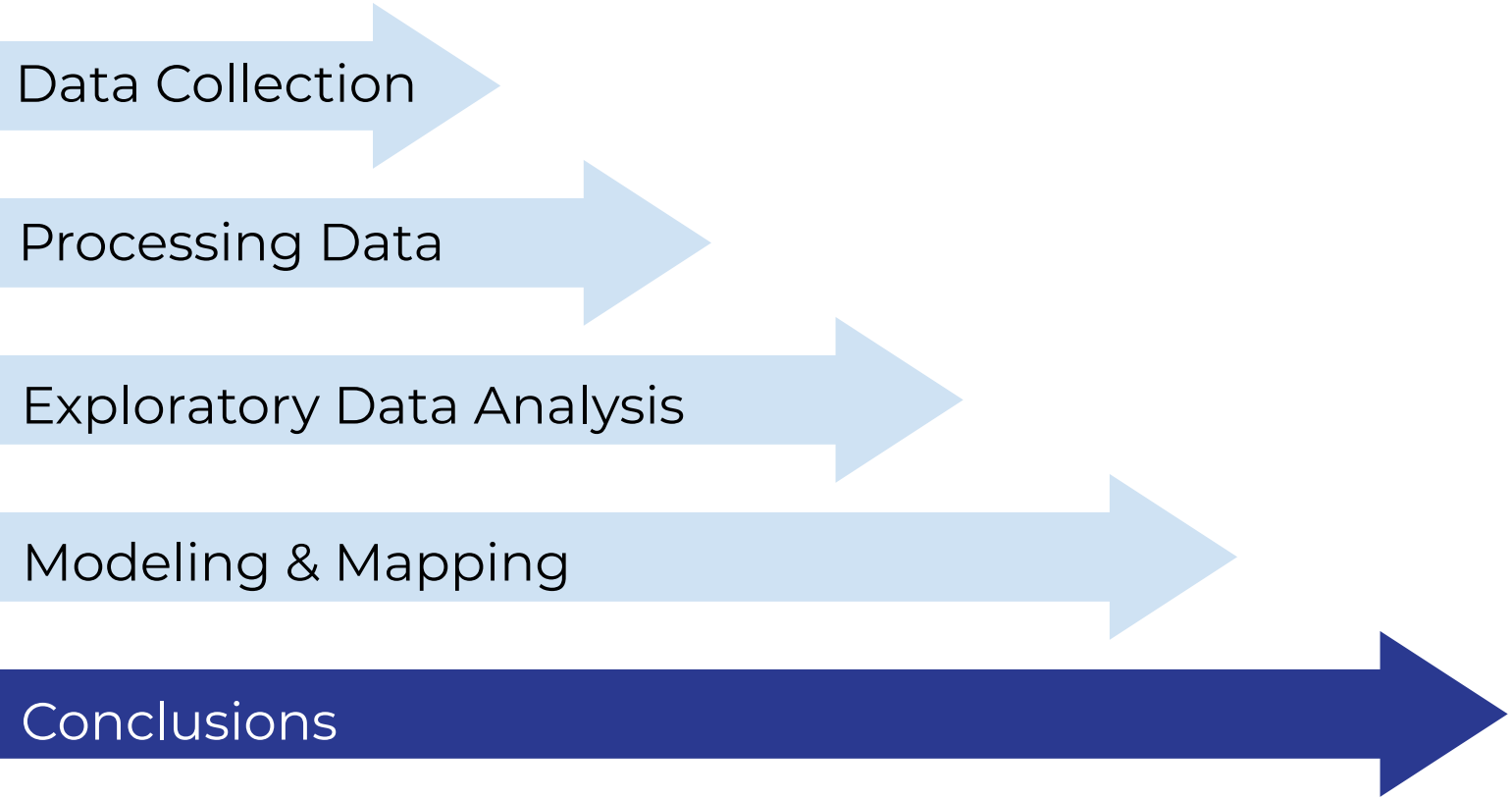
Model Evaluation

- Recall (sensitivity) was selected as the target metric
  - False negatives should be minimized
- Final model performance: 89% recall
  - 489 Predicted Positives / 551 Actual Positives
- Positive tweets are then visually represented

Mapping Urgent Tweets



Process


Data Collection
Processing Data
Exploratory Data Analysis
Modeling & Mapping
Conclusions

Conclusions

Limitations

- Identification Issues
- Data Issues
- Privacy Issues

Recommendations

- Broader social media access.
- Real time geo-locations.
- Concentration of each post's timing.
- Greater variation and quantity of posts.

Thank You

Questions?

Sources

- https://www.smh.com.au/world/nuclear-war-extreme-weather-top-list-of-2018-global-threats-20180118-h0k0ce.html
- https://about.twitter.com/en_us/company/brand-resources.html
- https://www.mnn.com/earth-matters/climate-weather/blogs/tornado-alley-forecast-more-tornadoes
- https://houstorian.wordpress.com/old-houston-maps/
- https://cohgis-mycity.opendata.arcgis.com/datasets/city-of-houston-etj
- https://blog.twitter.com/en_sea/topics/insights/2018/5-Tips-for-using-Twitter-during-emergencies-and-natural-disaster.html