



# Diagnosing the Disaster Tweets



An NLP Project by Jon Rosenberg  
conducted in participation in Springboard



# Preprocessing

---

- Working in NLP is largely preprocessing and then model application
- In preprocessing we:
  - Reduce tokens to lemmas
  - Remove stop words
  - Remove excess spaces and punctuation (but are careful to keep emoticons for our use case)
  - Clean out some garbled characters and out of vocabulary words
- We create an TF-IDF representation of our corpus for traditional machine learning methods.

# Exploration

- There's not a lot of exploration that can be performed in NLP problems.
- When classifying though as in our case, there is the traditional word clouds:

