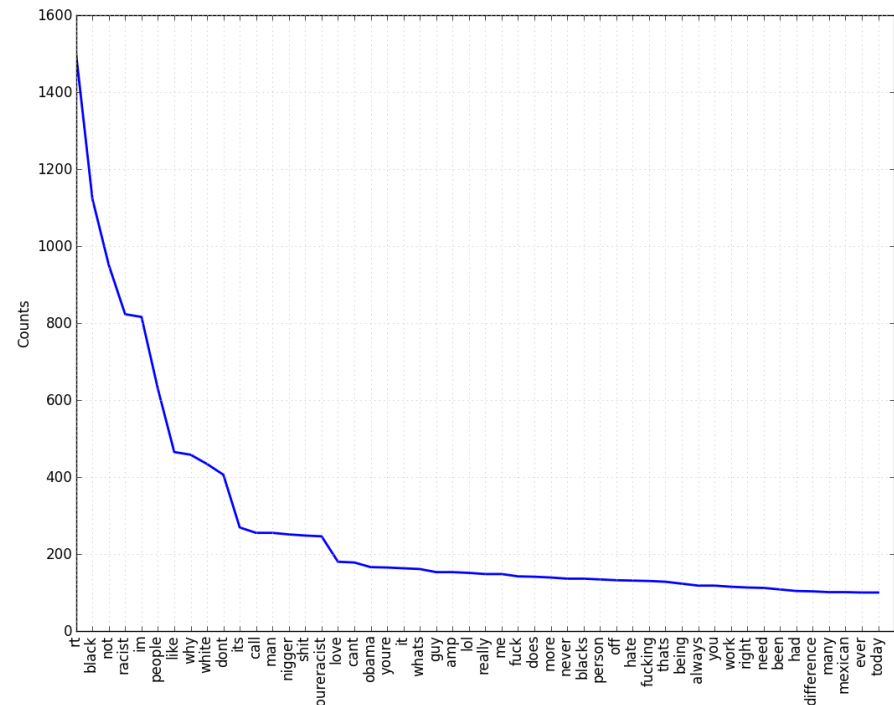# Detecting Racism in Tweets

Sean Noran, Fiete Botschen

- 15240 hand labeled tweets
- sources
    twitter stream (~8 %)
    racist jokes user (~ 87 %)
    racist tweets retweeter (~ 50 %)

- classifiers:
    NB, SVM, RF (similar results)

- features:
    TF-IDF, BoW, word2vec

- data analysis:
    Topic Modeling, feature extraction

Current state: generalizes not as good as we want. We hope to improve that quality.