

# The U.S. Presidential Race

## Comparing Coverage Across Three Major Sources

Galvanize, Inc.

September 2015

- Classification: *Is there a difference in coverage?*
- Topic Modeling: *What is the difference?*

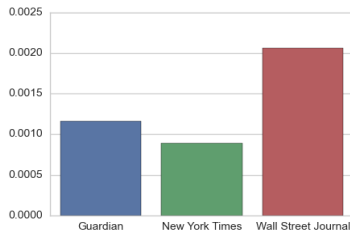
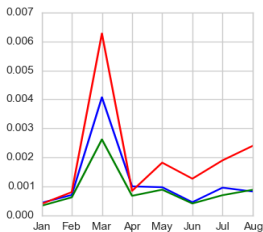
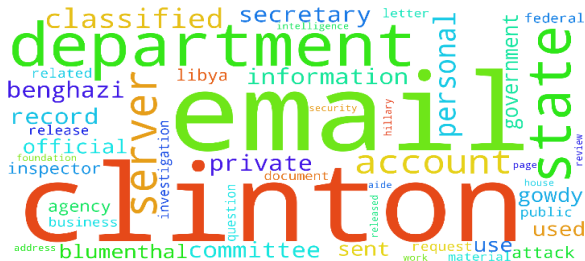
# Data collection

- # Scrape *Guardian*, *New York Times*, *Wall Street Journal*
- # January to August 2015
- # Raw data → MongoDB
- # Filter on candidates
- # TF-IDF, 2000 terms

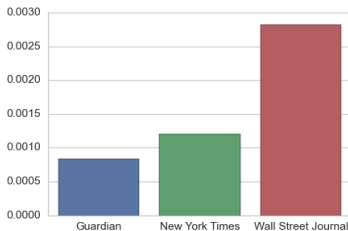
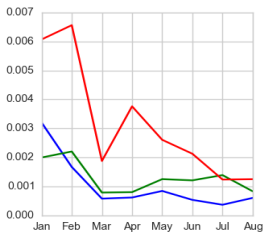
# Multiclass classification accuracy

	Global	Guardian	New York Times	Wall Street Journal
Naive Bayes	0.703	0.414	0.965	0.340
Random Forest	0.905	0.783	0.998	0.809
Gradient Boosting	0.921	0.856	0.983	0.833
<i>N</i>	3757	1018	2087	652

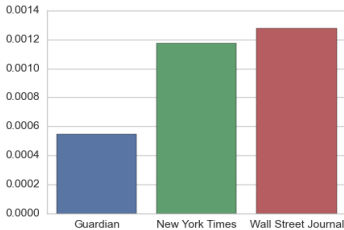
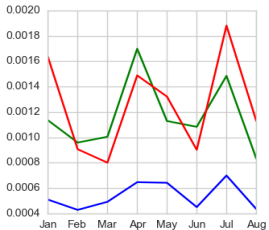
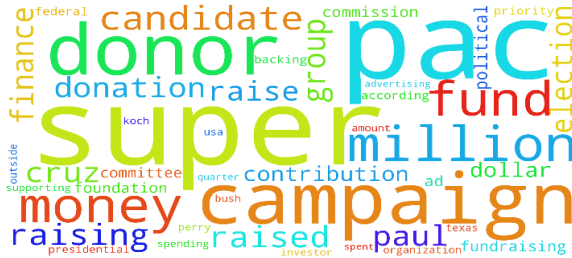
# Hillary's E-mail



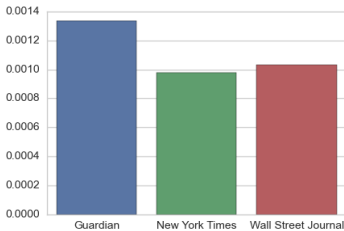
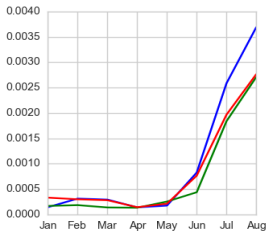
# Chris Christie



# Campaign Finance



# The Rise of Trump





# Social Politics

