Computational Text Analysis

Tools, Tricks, and Methods for Large Corpus Text Analysis

Sarah Connell
Cara Marta Messina
Alexis Yohros

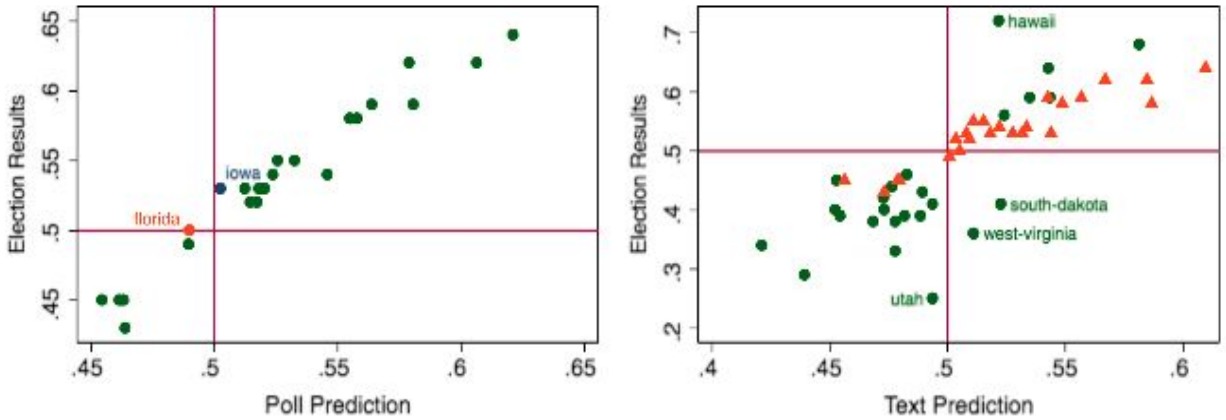
Northeastern University
NULab for Texts, Maps, and Networks

Big Data in the Modern World

Uses in Political Science

Beauchamp, N. (2017). Predicting and interpolating state-level polls using Twitter textual data. *American Journal of Political Science*, 61(2), 490-503.

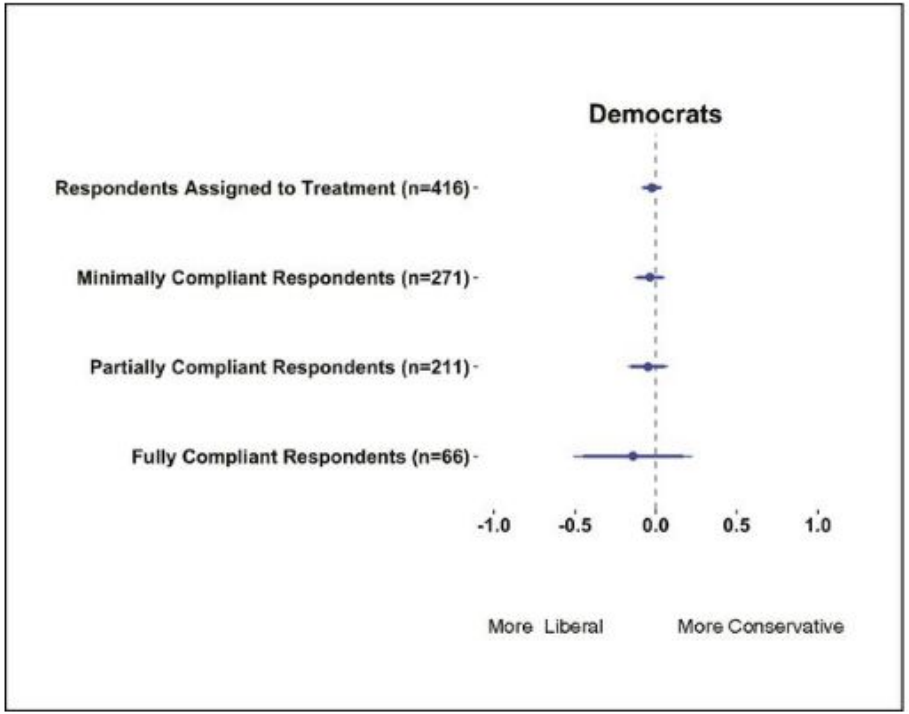
FIGURE 2 Polls at 11/4/12 vs. Election Results (Left) and Pure Text-Based Prediction on 11/4/12 from M1 (Right)

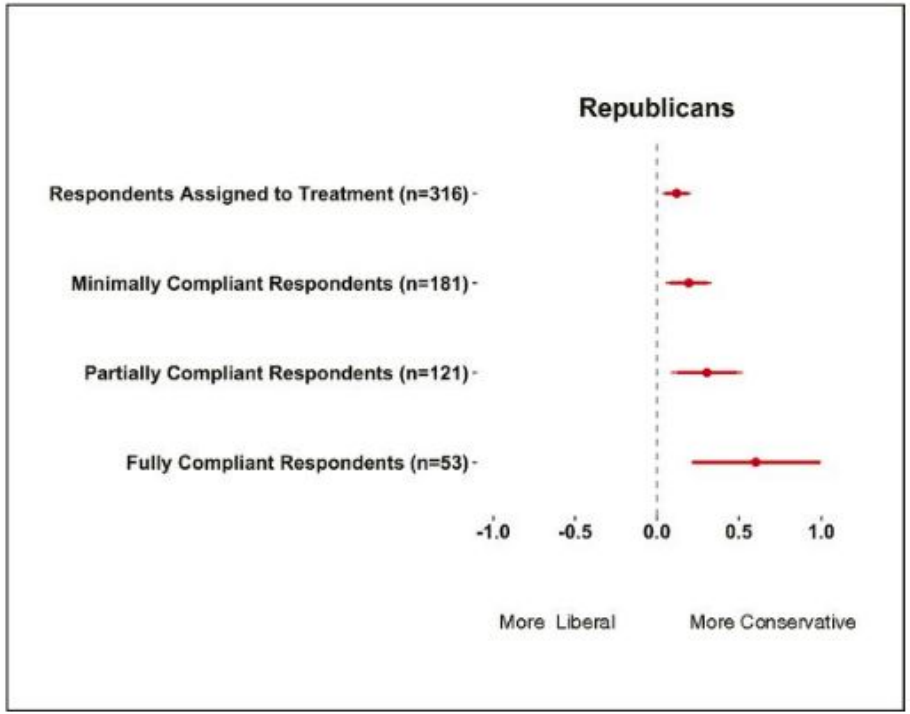


Note: Triangles are training states; circles are other states.

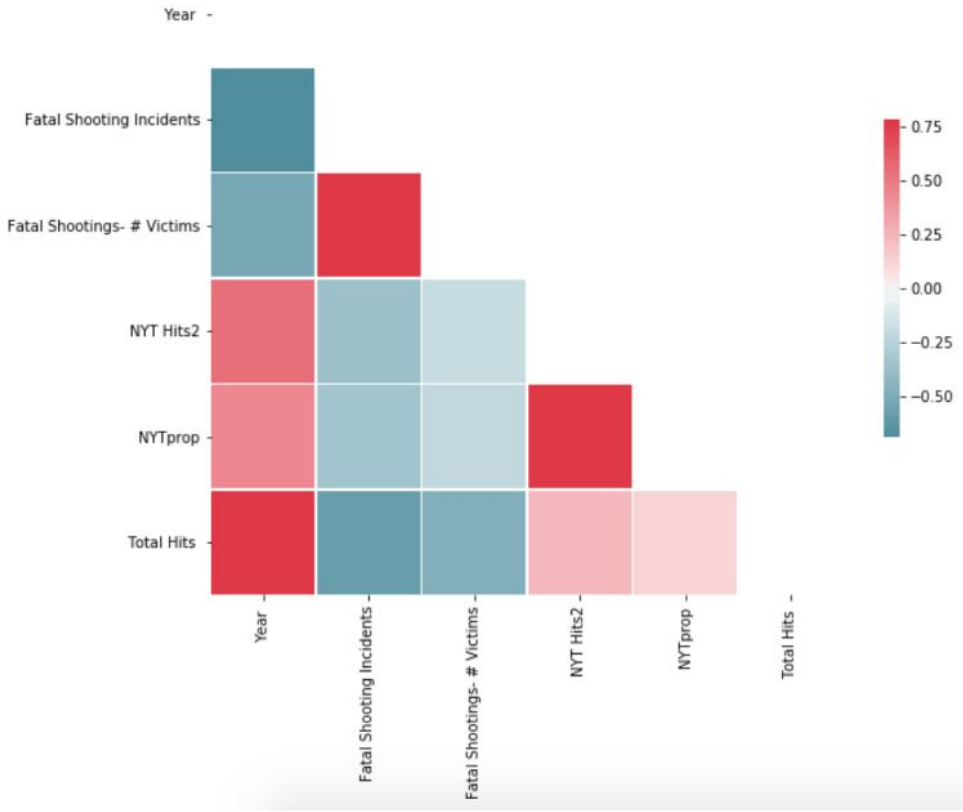
ii. Does exposure to opposing viewpoints change your opinion?

Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. *PNAS*, published August 28, 2018 https://doi.org/10.1073/pnas.1804840115

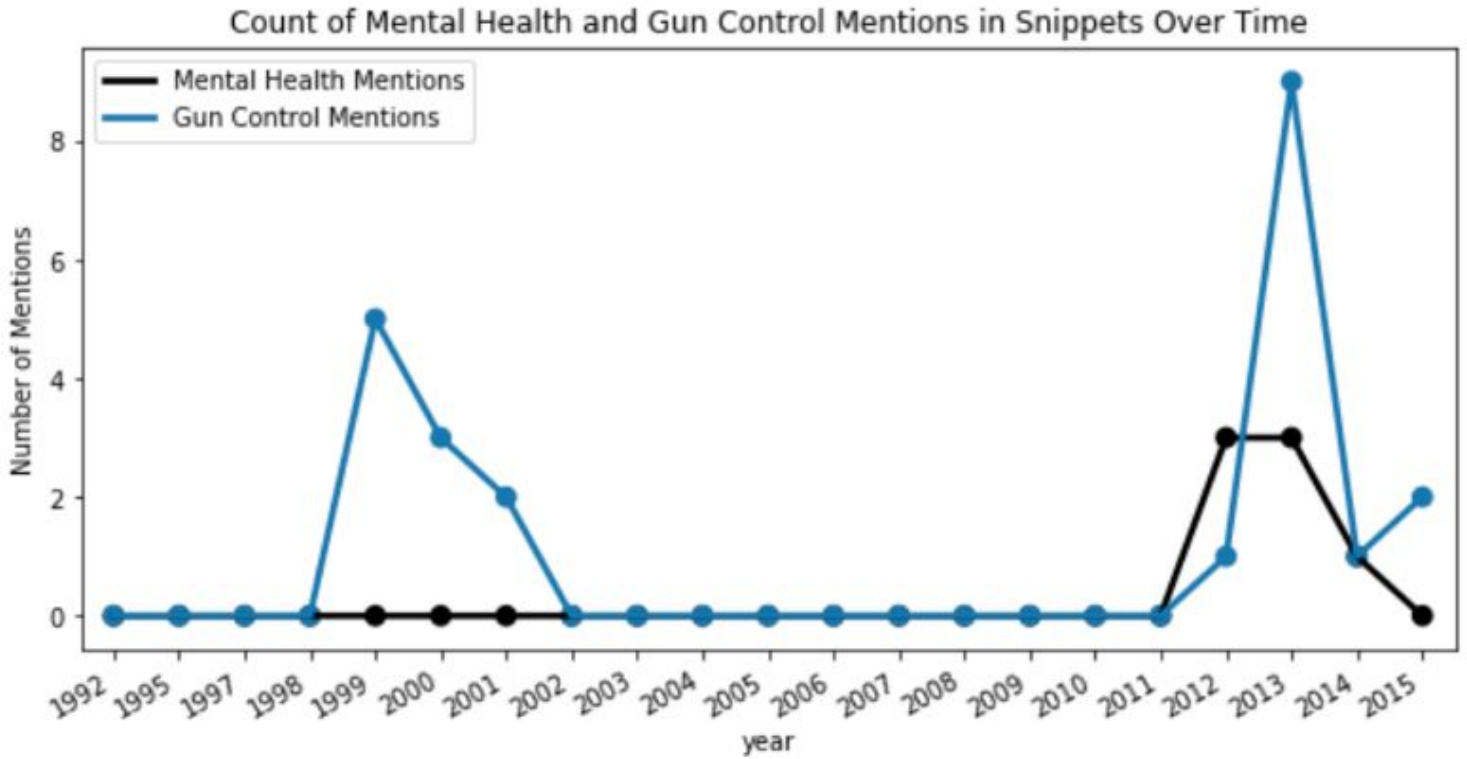




Applying it to our own Research



School Shootings and the Media



Northeastern University
NULab for Texts, Maps, and Networks

Follow Along!

Please check your emails for a link that has all the resources you need to follow along!

You should have: a document with links + data

Northeastern University
NULab for Texts, Maps, and Networks

Twitter Text Analysis

This mini-lecture covers (follow at https://bit.ly/2D9L1i3):

- What does Twitter data look like? How have researchers used Twitter data?
- Several approaches to analyzing Tweets from a week in October using the #midterms hashtag
  - Word count
  - Location using "user_location," which is filled in by the user
  - Word vector models
- The promises and pitfalls of using Twitter data

Northeastern University
NULab for Texts, Maps, and Networks

Web-Browser Computational Text Analysis Tools

You don’t necessarily need to know code to perform text analysis! While coding can be more flexible in its allowances, there are plenty of text analysis tools you can use to your advantage to explore your corpus.

Northeastern University
NULab for Texts, Maps, and Networks

Text Analysis Tools Links

In the shared class folder, click the document “Links to Browser Text Analysis Tools”

Also, open the “data” folder to look at some of the data we have compiled for you from

Northeastern University
NULab for Texts, Maps, and Networks

SameDiff

SameDiff: https://databasic.io/en/samediff/

SameDiff compares **two** texts and displays the words that appear in both texts and words that appear in each individual text.

Northeastern University
NULab for Texts, Maps, and Networks

WordTree

WordTree: https://www.jasondavies.com/wordtree/

Explore contextual patterns of particular words in individual texts. You can use SameDiff as a starting place to search words in the WordTree to see how contexts differ or particular contexts.

Northeastern University
NULab for Texts, Maps, and Networks

In-Class Activity!

Using SameDiff and WordTree, take some time to explore the different texts we have made available to use or explore your own data online.

If you are interested, you can also use the Drag and Drop Sentiment Analysis tool. Sentiment Analysis uses a dictionary that weighs words as “positive” or “negative” based on a dictionary. Be wary that these are not always “correct” because SA does *not* measure context.

Northeastern University
NULab for Texts, Maps, and Networks

Thank you!

https://web.northeastern.edu/nulab/

Sarah Connell
NULab Assistant Director
sa.connell@northeastern.edu
Cara Marta Messina
NULab Coordinator
messina.c@husky.neu.edu
Alexis Yohros
Digital Teaching Initiative RA
yohros.a@husky.neu.edu