

Introduction to Text Analysis

- ① Motivation
- ② Conceptual Overview
- ③ Extended Example
- ④ Additional Resources

① Motivation

Text is important for politics and political science

- Most of politics passes through language at some point
- When we have language, text is often a useful abstraction
- Most political data sets are derived from text
- Most qualitative work is text-based

① Motivation

Quantitative text analysis methods can assist both quantitative and qualitative modes of inquiry

- Systematizes rote tasks that a human can do
 - ▶ Equivalent to thousands of cheap RAs with moderate, known error rates (force multiplier)
 - ▶ My interest in text analysis stems from experience as a human RA (with moderate, unknown error rates)
- Makes helpful proposals for further reading, coding, etc
- Anyone who says it is magic is selling you something
 - ▶ Not a magic method for generating insights
 - ▶ Not a substitute for reading or deep knowledge

① Motivation

I've used it on almost every project, even “non-text” projects

- Book
 - ▶ Measured Jihadism using cleric fatwas
 - ▶ Explored topics in a Jihadi corpus
 - ▶ Characterized the population of Muslim clerics on Wikipedia
 - ▶ Improved speed of qualitative coding of CV data
- Classified NYT articles as being about human rights abuses or not (ISQ 2013)
- Searched for all instances of praise for ratification in US and EU press releases (ISQ 2015)

② Conceptual Overview

No math.

② Conceptual Overview

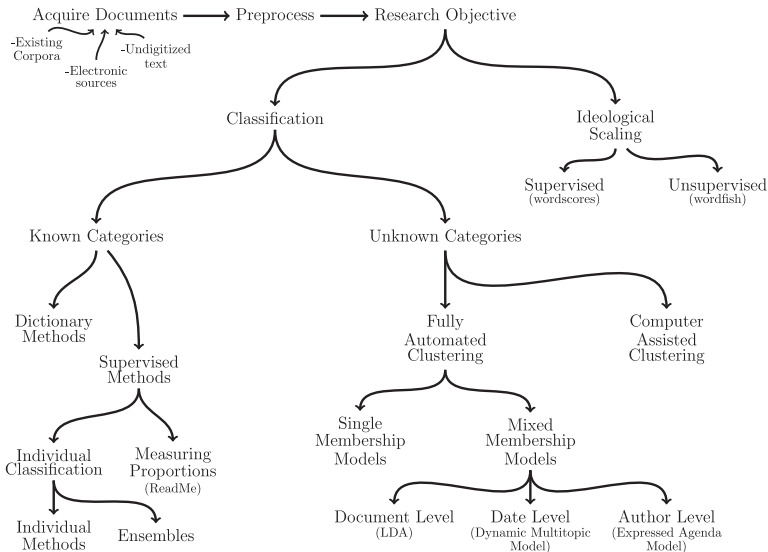
Four Principles of Quantitative Text Analysis (Grimmer and Stewart, 2013)

-
- (1) All quantitative models of language are wrong—but some are useful.
 - (2) Quantitative methods for text amplify resources and augment humans.
 - (3) There is no globally best method for automated text analysis.
 - (4) Validate, Validate, Validate.
-

My additions:

- (5) All text analysis is dimension reduction.
- (6) There is no free lunch.
(supervised methods require front-end work, unsupervised require back-end work)

② Conceptual Overview



③ Extended Example

④ Additional Resources

Textbooks:

- Tibshirani, Hastie, and Friedman. *ESL* (link)
- Bishop. *Pattern Recognition and Machine Learning*.

Papers:

- Text as Data (link)
- Computer Assisted Text Analysis for Comparative Politics (link)
- Work by Kevin Quinn, Arthur Spirling, Molly Roberts, Justin Grimmer, Gary King, Ben Lauderdale, Hanna Wallach, others.