



CRIM 3300: Punishment in the Age of Mass Incarceration
Megan Denver
Data Ethics and Introduction to Text Analysis

Key Words

- **Big Data:** Volume, variety, velocity and veracity. Collects a large and diverse amount of data quickly and uses algorithms to categorize and re-categorize data.
- **Algorithms:** A set of procedures to be followed by certain technologies (computers, cell phones, etc). Algorithms typically rely on data and a set of instructions to “read” that data in some way
- **Algorithmic Bias:** The implicit biases that are reinforced and perpetuated through algorithms creation, the data collection, data analysis, the results, and the impact of these results
- **Computational Text Analysis (CTA):** Text analysis is making inferences based on textual data. CTA involves a computer drawing out patterns in a text, and then the researcher interprets those patterns. CTA includes methods such as word count frequency, nGrams, and sentiment analysis. CTA is similar to statistical analysis, but the data are texts.
- **Corpus (plural-corpora):** A collection of texts or data used for analysis and research purposes.
- **Stop words:** words that appear frequently in a language, like pronouns, prepositions, and basic verbs. These are often removed for basic computational analysis. Some English stopwords include: a, the, she, he, I, me, us, of, is, would, could, should, etc.

Web-Browser Computational Text Analysis Tools

These browser GUI (Graphical User Interface) text analysis tools are not particularly powerful or sophisticated, but they can show word frequencies and patterns in language. While using coding languages like Python and R can open up other types of analysis (such as word embedding models and topic modeling), these GUI tools allow you to do more basic analysis to help begin examining your texts computationally. We will be working with:

- Lexos: <http://lexos.wheatoncollege.edu/upload>
- Voyant: <https://voyant-tools.org/>

Sample Corpus

The data are several State of the Union addresses given by the current and former Presidents, going back to George Bush Sr. Data retrieved from

<https://www.presidency.ucsb.edu/documents/app-categories/spoken-addresses-and-remarks/pridential/state-the-union-addresses>

Find this handout and more at <http://bit.ly/diti-fall2019-denver>

Questions? Contact us!

Cara Marta Messina, messina.c@husky.neu.edu

Garrett Morrow, morrow.g@husky.neu.edu