INFO 490C/690C Spring 2022 Schedule

Week 1: Introduction
Discussion Readings:
- Grimmer (2015). *We Are All Social Scientists Now*
- Underwood (2015). *Seven ways humanists are using computers to understand text*
- boyd & Crawford (2011). *Critical questions for big data*
Sessions:
- Wednesday (1/26): Introduction
- Friday (1/28): In-Class Discussion

Week 2: Tokenization
Discussion Readings:
- Schmidt & Fraas (2015). *The Language of the State of the Union*
- Daniels (2019). *The Largest Vocabulary In Hip Hop*
Technical Readings:
- RegexOne: Learn Regular Expressions
- Potts (2011). Sentiment Symposium Tutorial: Tokenizing
- Dombrowski (2020). Preparing Non-English Texts for Computational Analysis
- Optional: Church (adapted by Tyers). Unix for Poets
- Optional: Jurafsky & Martin (2021). "Regular Expressions, Text Normalization, Edit Distance"
Sessions:
- Monday (1/31): Regular Expressions
- Wednesday (2/2): Tokenization
- Friday (2/4): Cancelled

Week 3: Python & Counting
Discussion Readings:
- Herndon et al. (2019). *What Does Campaign Rally Music Say About the Candidates?*
- Davis (2020). *The Physical Traits that Define Men and Women in Literature*
Technical Readings:
- Review these Python resources/documentation:
  - Control Flow
  - Data Structures
  - Counters
  - Sorting
  - Reading and Writing Files
Sessions:
• Monday (2/7): Python & Counting
• Wednesday (2/9): Python & Counting (cont.)
• Friday (2/11): In-Class Discussion

Week 4: Sentiment Analysis
Discussion Readings:
• Kurt Vonnegut on Shapes of Stories [video]
• Jockers (2015). Revealing Sentiment and Plot Arcs with the Syuzhet Package
• Jockers (2015). That Sentimental Feeling
• Regan et al. (2016). The emotional arcs of stories are dominated by basic shapes
Sessions:
• Monday (2/14): Sentiment Analysis I
• Wednesday (2/16): Sentiment Analysis II
• Friday (2/18): In-Class Discussion

Week 5: Vector Space Model
Discussion Readings:
• Arnold et al. (2019). Visual Style in Two Network Era Sitcoms
Technical Readings:
• Polamuri (2015). Five Most Popular Similarity Measures Implemented in Python
Sessions:
• Tuesday (2/22): Vector Space Model
• Wednesday (2/23): Comparison
• Friday (2/25): Cancelled

Week 6: Clustering
Discussion Readings:
• Wilkens (2016). Genre, Computation, and the Varieties of Twentieth-Century U.S. Fiction
Technical Readings:
• Wednesday: Harris (2014). Visualizing K-Means Clustering
• Optional: Harris (2015). Visualizing DBSCAN Clustering
Sessions:
• Monday (2/28): Agglomerative Clustering
• Wednesday (3/2): K-Means Clustering & Visualization
• Friday (3/4): In-Class Discussion

Week 7: Classification
Discussion Readings:
• Klein & D’Ignazio (2020). "What Gets Counted Counts" from Data Feminism
• Long & So (2016). Literary Pattern Recognition
Technical Readings:
• Victor Powell, Conditional Probability: Explained Visually
• Arbital Guide to Bayes’ Rule

Sessions:
• Monday (3/7): Classification
• Wednesday (3/9): Classification (cont.) [held remotely]
• Friday (3/11): In-Class Discussion [held remotely]

Week 8: Final Projects & Datasets
Discussion Readings:
• Crawford & Paglen (2019). Excavating AI
Technical Readings:
• Wednesday: Krause (2017). Data Biographies
• Wednesday: Gebru et al. (2018/2021). Datasheets for Datasets
• Optional: Suresh (2019). The Problem with "Biased Data"

Sessions:
• Monday (3/21): Final Projects
• Wednesday (3/23): On Data
• Friday (3/25): In-Class Discussion

Week 9: Comparing Events
Discussion Readings:
• Broadwell et al. (2017). The Tell-Tale Hat
Technical Readings:

Sessions:
• Monday (3/28): Comparing Events
• Wednesday (3/30): Feature Analysis I
• Friday (4/1): In-Class Discussion

Week 10: Feature Analysis & Author Similarity
Discussion Readings:
• Storey & Mimno (2020). Like Two Pis in a Pod
Technical Readings:

Sessions:
• Monday (4/4): Feature Analysis II
• Wednesday (4/6): Author Similarity
• Friday (4/8): In-Class Discussion

Week 11: More on Hypothesis Testing & Final Project Peer Reviewing
Technical Readings:
• **Stray (2016). Solve Every Statistics Problem with One Weird Trick [video]**
• **Munroe (2011). Significant**

**Sessions:**
• Monday (4/11): Bootstrapping
• Wednesday (4/13): Multiple Hypotheses
• Friday (4/15): Final Project Peer Review

**Week 12: Project Check-Ins**
• Wednesday (4/20): Project Check-Ins
• *Friday (4/22): Cancelled*

**Week 13: More on Final Projects**

Discussion Readings:
• Optional: **Crawford & Paglen (2019). Excavating AI**

Technical Readings:
• **Jockers (2011). The LDA Buffet**
• **Boyd-Graber et al. (2017) Applications of Topic Models: Chapter 1.** Read for intuition.

Sessions:
• Monday (4/25): Project Check-Ins
• *Wednesday (4/27): Cancelled*
• Friday (4/29): Final Projects

**Week 14: Final Project Presentations**

Sessions:
• Monday (5/2): Final Project Presentations
• *Wednesday (5/4): Final Project Presentations*