

INFO 490C/690C Spring 2021 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:

- Tuesday (2/2): Introductions & review of course syllabus
- Thursday (2/4): Setting up local computing environments

Week 2: Text to Data

Discussion Readings:

- [Dimson \(2015\). Emojineering Part I](#)
- [Mukerjee \(2015\). I Can Text You A Pile of Poo, But I Can't Write My Name](#)

Technical Readings:

- [Zentgraf \(2015\). What Every Programmer Absolutely, Positively Needs To Know About Encodings And Character Sets To Work With Text](#)
- [RegexOne: Learn Regular Expressions](#)
- [Potts \(2011\). Sentiment Symposium Tutorial: Tokenizing](#)

Sessions:

- Tuesday (2/9): Character Encodings
 - Reading: Zentgraf 2015
 - Extended Notebook
- Thursday (2/11): Tokenization
 - Reading: RegexOne; Potts 2011
 - Session Notebook & Recording

Week 3: Counting

Discussion Readings:

- [Schmidt & Fraas \(2015\). The Language of the State of the Union](#)
- [Daniels \(2019\). The Largest Vocabulary In Hip Hop](#)

Sessions:

- Tuesday (2/2): Python Fundamentals
 - Reference: [Python Documentation](#)
 - Session Notebook & Recording
- Thursday (2/4): Counting
 - Session Notebook & Recording

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:

- Tuesday (2/16): Sentiment Analysis I
 - Session Notebook & Recording
- Thursday (2/18): ~~Sentiment Analysis II~~ Cancelled.

Week 5: 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](#)
- [Francisco Jacobelli, Text Classification Using Naive Bayes \[video\]](#)

Sessions:

- Tuesday (3/2): Classification I
 - Session Notebook & Recording
- Thursday (3/4): Classification II
 - Session Notebook & Recording

Week 6: Review

Discussion Readings:

- [So & Roland \(2020\). Race and Distant Reading](#)

Sessions:

- Tuesday (3/9): Tokenization Revisited
 - Session Notebook & Recording
- Thursday (3/11): Revisiting Course Concepts
 - Session Notebook & Recording

Week 7: No Class

Week 8: Similarity & Distance

Discussion Readings:

- [Barron et al. \(2018\). Individuals, institutions, and innovation in the debates of the French Revolution](#)

Technical Readings:

- [Polamuri \(2015\). Five Most Popular Similarity Measures Implemented in Python](#)

Sessions:

- Tuesday (3/23): Representations & Similarity

- Session Notes & Recording
- Thursday (3/26): Comparing Texts Continued
 - Session Notes & Recording

Week 9: Clustering

Discussion Readings:

- [Wilkins \(2016\). Genre, Computation, and the Varieties of Twentieth-Century U.S. Fiction](#)

Sessions:

- Tuesday (3/30): Agglomerative Clustering I
 - Session Notes & Recording
- Thursday (4/1): Agglomerative Clustering II
 - Session Notebook & Recording

Week 10: Data Revisited

No Discussion Readings

Technical Readings:

- [Krause \(2017\). Data Biographies](#)
- [Gebu et al. \(2020\). Datasheets for Datasets](#)
- [Suresh \(2019\). The Problem with "Biased Data"](#)

Sessions:

- Tuesday (4/6): Data Revisited
 - Session Slides & Recording
 - [Collaborative Document: Dataset questions](#)
- *Thursday (4/8): Cancelled*

Week 11: Clustering

Discussion Readings:

- [Madrigal \(2014\). How Netflix Reverse-Engineered Hollywood](#)

Technical Readings:

- [Harris \(2014\). Visualizing K-Means Clustering](#)
- [Harris \(2015\). Visualizing DBSCAN Clustering](#)

Sessions:

- Tuesday (4/13): K-Means Clustering
 - Clustering Notebook (See Final Project's' Reference Materials folder)
 - Session Notes & Recording
- Thursday (4/15): Clustering & Visualization
 - Clustering Notebook (See Final Project's' Reference Materials folder)
 - Session Recording
 - Session Announcements

Week 12: Feature Analysis

Discussion Readings:

- [Storey & Mimno \(2020\). Like Two Pis in a Pod](#)

Technical Readings:

- [Dunning \(2008\). Surprise and Coincidence](#)
- [Broadwell et al. \(2017\). The Tell-Tale Hat](#)

Sessions:

- Thursday (4/22): Comparing Events
 - Session Notes
 - Session Notebook & Recording

Week 13: Feature Analysis

Sessions:

- Tuesday (4/27): Feature Selection
 - Feature Analysis Notebook (See Final Project's' Reference Materials folder)
 - Session Recording
- Thursday (4/29): Feature Ablation
 - Feature Analysis Notebook (See Final Project's' Reference Materials folder)
 - Session Recording

Week 14: Final Project

Sessions:

- Tuesday (5/4): Open Session
 - This session will be an opportunity to review any topics of interest and answer questions. It will not be recorded.