Computational Social Science

(Data Analysis & Text as Data)

(**Tentative Syllabus** – Updated October 9, 2025)

Spring 2026

Instructor:	J.S. Truscott	**Email:**	jaketruscott@ufl.edu
Class Time:	3:00-6:00pm	**Class Location:**	Anderson 0032
Office:	Anderson 317	**Office Hours:**	W 1:00-3:00pm

Course Description

How can we leverage existing and emerging computing technologies for practical use in the social sciences? Once considered the domain of specialists with extensive technical training, computational methods are now more accessible than ever. This course provides an introduction to the tools and techniques that are transforming social science research, particularly in the realm of text analysis (and broader concepts related to text-as-data). Students will learn to leverage computational methods to analyze complex phenomena, develop reproducible workflows, and engage with the ever-growing intersection of data science and social inquiry.

Students should exit this course with the ability to (among other things):

- Compile and deploy complex computing routines in `R` and (to a lesser extent) `Python`, including multi-layered and hierarchical coding structures (e.g., iterative loops, data organization and manipulation, functions, package construction, etc.)
- Retrieve, process, and organize non-traditional data sources (i.e., text) for tasks using a combination of natural language processing techniques (e.g., supervised and unsupervised learning models)
- Create informative representations of data and other summary findings (e.g., tables and figures).
- Produce replicable workflows and documentation using `R`, `Python`, and LaTeX(`Beamer`).

Addendum re: Prerequisite(s)

This course assumes basic (*novice*) exposure to statistical packages and programming languages, namely `R` (and `Python`). Much of the course will focus on facilitating greater proficiency with these tools for social science research.

Moreover, while it is assumed that students will have previously satisfied the equivalent of *The Conduct of Inquiry* (POS6736) and *Political Data Analysis* (POS6737), necessary instruction and clarification of subjects and methodologies more advanced than the equivalent facilitated in those courses will be provided.

Technology Disclaimer

Given the focus of the course, it is expected that students will arrive to each class with a laptop capable of downloading `R` and its companion IDE (`RStudio`), as well as a `Python` interpreter and associated IDE (e.g., `Microsoft Visual Studio Code`). Necessary instructions to download both will be provided during the first class meeting.

Additional Notes Re: Technology

- We will principally conduct our work in `R` for much of the semester, though the content discussed in some weeks will lend itself better to being completed using `Python`.
- Instruction requiring `Python` will be demonstrated using `reticulate::()` in `R`, which provides an interoperability interface for using `Python` directly within `R`. However, students are welcome to complete the problem sets assigned in this course using a separate IDE (e.g., `Microsoft VSC`), if they so choose.

Textbooks

- Grimmer, J., Roberts, M. E., & Stewart, B. M. (2022). *Text as data: A new framework for machine learning and the social sciences*. Princeton University Press. **(GRS)**
- Silge, J., & Robinson, D. (2017). *Text mining with R: A tidy approach* (1st ed.). O'Reilly Media. **(SR)**

Additional Notes Re: Textbooks and Course Readings

- You are welcome to purchase a copy of **SR**, but we will be referencing material listed on the textbook's companion website.
- ***Other readings listed below will be made available via Canvas*** – Please have them completed **prior** to arriving for class.
- Topic notes for each week (with associated code and any datasets used for examples) will be made available on the course Github Repository.

Grading Policy

Weekly Problem Sets	40%
Participation	10%
Final Project & Presentation	50%

Grading Scale for Final Semester Grades

100-94 A	79-77 C+	63-60 D-
93-90 A-	76-74 C	59-0 F
89-87 B+	73-70 C-	
86-64 B	69-67 D+	
83-80 B-	66-64 D	

Evaluation

Weekly Problem Sets

Students will complete weekly homework assignments due on **Sundays at 11:59pm** to be submitted via a Canvas submission portal. These assignments will include a practice set related to the content discussed in class that week. Completion of each assignment will require submitting as a PDF rendered from `RMarkdown` (.R), `Jupter Notebook` (.ipynb), `Python` (.py), or LaTeX formats, such that the submission includes both the rendered figures (tables), as well as any associated workflows (code). A submission where the document was rendered using *Microsoft Word* (.docx) or similiar will not be accepted.

Notes Re: Weekly Problem Sets

- Apart from supplying your workflow with a completed problem set, you will also be required to provide documentation (comments) in the margins of your submission to demonstrate comprehension of the workflow's odds and ends. More information will be provided during our first class meeting.
- Each class meeting will begin with reviewing and discussing the prior week's assignment before continuing with any new material.

Participation

Participation will principally include weekly attendance of our classroom meetings, as well as active participation in our open discussions. Given that this is a graduate course and only meets once each week, it is expected that students attend every class meeting. Failure to attend more than two classes without excused absences will result in a 50% deduction of a student's course participation grade.

Final Project & Presentation

At the conclusion of the semester, students will present a project incorporating some element of the course material (*subject to the approval of the instructor*). These may include, but are not limited to, projects that incorporate text analysis or any other methodology related to studies of *big data* or parallel computing. Given the nature of the course, the focus is less on theory and substance than on demonstrating comprehension of the methods and strategies introduced throughout the semester.

Furthermore, you are free to continue with the development of a project already in production, so long as the submission (i) meets the necessary requirements, and (ii) permission is granted by the student's (a) principal advisor – *if the project is part of their comprehensive exam or dissertation activities*, or (b) previous instructor – *if the project began in another graduate course*).

Note: Grades for this assignment will draw from both the instructor and a peer assessment. More information will be provided during the course of the semester.

Communication with Instructor

If you need to contact me for any reason, I ask that you do so using your UF email address (rather than Canvas). I am generally very good at responding to emails quickly, but please understand that I likely will not respond until the next day if the message is sent late in the night.

Email: jaketruscott@ufl.edu

University Policies & Resources

This course complies with all UF academic policies. For information on those polices and for resources for students, please see HERE.

Student Evaluations (GatorEvals)

Students are expected to provide professional and respectful feedback on the quality of instruction in this course by completing course evaluations online. Students can complete evaluations in three ways: (1) The email they receive from GatorEvals; (2) Their Canvas course menu under GatorEvals; or (3) The central portal at `https://my-ufl.bluera.com`. Guidance on how to provide constructive feedback is available at `https://gatorevals.aa.ufl.edu/students/`. Students will be notified when the evaluation period opens. Summaries of course evaluation results are available to students at `https://gatorevals.aa.ufl.edu/public-results/`.

Note: Starting in Spring 2025, the central portal changed to `https://my-ufl.bluera.com`. The previous GatorEvals link (`https://ufl.bluera.com/ufl/`) will redirect to the updated portal for one year.

Acknowledgments

Given the scope of the content and the fact that it is not as simple as writing a course within a substantive, theory-driven subfield, I want to make sure to give special thanks to Dr. Joseph Ornstein (University of Georgia) and Dr. Bryce Dietrich (Purdue University) for both facilitating my education of these concepts, as well as allowing me to ~~steal~~ borrow from their existing syllabi.

Key Dates

- January 19 (*No Class – Holiday*)
- February 8 (*Final Project Topic Selection Due*)
- March 16 (*No Class – Spring Break*)
- April 13 & April 20 (*Presentations*)

Course Schedule

Class 1 (January 12)

Topic: Syllabus & Introduction
Material: Review Syllabus, Download `R`/`RStudio`, and Discuss Data Organization & Management.
Homework: Play Around in `R`!

Class 2 (January 26)

Topic: Intermediate Programming in `R`
Material: `R` Functions, Iterative Loops, `ggplot()`, etc.
Homework: Class 2 Practice Set

Class 3 (February 2)

Topic: Intermediate Programming in `R` (Cont.) & Parallel Computing
Material: Parallel Computing in `R`
Homework: Final Project Topic Selection

Class 4 (February 9)

Topic: Introduction to Text Analysis
Material: Harvesting & Text Retrieval, Corpus Creation, and Text Pre-Processing
Homework: Class 4 Practice Set
Readings:

- **GSR** Ch. 1-4
- **SR** Ch. 1
- Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. Political analysis, 26(2), 168-189.

Class 5 (February 16)

Topic: The Bag of Words
Material: Word Counts, Tokenization, Complexity Reduction, and Document Feature Matrices (Part I)
Homework: Class 5 Practice Set
Readings:

- **GSR** Ch. 5 & 16
- **SR** Ch. 5

Class 6 (February 23)

Topic: Modeling the Bag of Words
Material: DFM (Part II), Similarity and Dictionary Classification
Homework: Class 6 Practice Set
Readings:

- **GSR** Ch. 6-7

Class 7 (March 2)

Topic: Clustering and Topic Discovery
Material: Latent Dirichlet Allocation (LDAs) and Structural Topic Models (STMs)
Homework: Class 7 Practice Set
Readings:

- **GSR** Ch. 10 & 12-13
- **SR** Ch. 6

Class 8 (March 9)

Topic: Word and Sentence Embeddings
Material: `word2vec`, `GloVe`, and `BERT`
Homework: Class 8 Practice Set
Readings:

- **GSR** Ch. 8
- Rodriguez, P. L., & Spirling, A. (2022). Word embeddings: What works, what doesn't, and how to tell the difference for applied research. The Journal of Politics, 84(1), 101-115.

Class 9 (March 23)

Topic: Supervised Classification & Sentiment Models
Material: Naive Bayes and Support Vector Machines (SVMs)
Homework: Class 9 Practice Set

- **GSR** Ch. 17-19
- Truscott, J. S. (2024). Analyzing the rhetoric of Supreme Court confirmation hearings. Journal of Law and Courts, 12(1), 45-66.

Class 10 (March 30)

Topic: Unsupervised Methods & Scaling
Material: Text Models for Measuring Ideology (Wordscores, Wordfish, and Wordshoal)
Homework: NA

- **GSR** Ch. 14, 16.3, 20, 21.3
- Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. American political science review, 97(2), 311-331.
- Truscott, J. S., & Romano, M. K. (2025). Measuring Judicial Ideology Through Text. Journal of Law and Courts, 1-18.

Class 11 (April 6)

Topic: Free Day (*or Deep Learning/Transformers*)
Material: TBD
Homework: NA

Class 12-13 (April 13 & April 20)

Topic: Presentations