

# Course Project Requirements

## Sequencing Legal DNA, Spring 2020

### [Assignment Submission Dropbox](#)

Course projects will undertake a new research analysis in the area of natural language understanding of legal text (broadly defined). Projects can be done individually or in small groups (up to four students). Students can earn 2 additional credit points by signing up for a more substantial project.

- **Topic (0%), due March 30th.** Please have a conversation with the instructor or a project advisor to select and confirm a topic, dataset, and approach.
  - We have a list of project ideas for a range of topics. Email [Claudia](#) for access.
- **Outline (10%), due May 4th.** Outline of the motivation, related literature, data, and approach. [Outline Requirements](#).
- **Presentation (10%), starting May 4th.** Students will give a short presentation about their project toward the end of the course. [Presentation Requirements](#).
- **Rough Draft (10%), due July 24th.** A rough draft of the Data, Methods, and Results sections must be submitted for feedback.
- **Final Draft (40%), due August 21st.** A polished paper reporting on the project's analysis and results (Introduction, Lit Review, Data, Methods, Results, and Conclusion).

### Picking a Topic

There should be some relation to the methods and substantive topics covered in the course, but overall we are very flexible with topics. We have a list of proposed topics available by request.

- [Google dataset search](#) (e.g., [search "corpus"](#))

### Submitting Work

Upload work in the [submission dropbox](#):

- **Outline:** Filename: *[LastName\_FirstName\_Outline.pdf]*. Should be anonymized.
- **Rough Draft:** Filename: *[LastName\_FirstName\_Rough.pdf]*.
- **Final Draft:** Filename: *[LastName\_FirstName\_Final.zip]*. Must include compiled PDF, document source files, and replication code.

### Literature Review

Check the [NLP Progress](#) repo for the best models by NLP task.

### Minimum Word Count Requirements

These are recommendations and will not be strictly enforced, but give an idea of the level of detail that is expected. Be careful about the differential requirements based on number of team members and associated course credits (three or five).

**Rough Draft: 1000 words, plus**

- 500 words per 3-credit student
- 1000 words per 5-credit student

**Final Draft: 1500 words, plus**

- 750 words per 3 credit student
- 1500 words per 5-credit student

So, for example, a single 3-credit student should have at least 1500 words in their rough draft and 2250 words in their final draft. The recommended section-level word counts below are for single-authored 3-credit papers and should be extended as appropriate.

**Recommended organizational structure of final drafts****1. Introduction (400+ words)**

- What is your research question or goal?
- Discuss the motivation for the project -- why is it interesting or necessary?
- Summarize your approach.
- Summarize results.
- What have we learned? Relate to previous literature.

**2. Literature Review (200+ words)**

- Review the literature on your research question. When appropriate, outline how your approach differs with the previous papers and how it addresses their shortcomings.

**3. Data and Summary Statistics (400+ words)**

- Describe the data sources, and report summary statistics. Include at least one figure/graph, for example plotting your outcome variable over time.

**4. Methods (800+ words)**

- Outline your approach. What tools will you use to solve your research problem?
- E.g., for machine learning: What model will you use? How will you evaluate and validate your model? How will you interpret or explain your model?
- For an econometric analysis, write out your regression equation. What are the empirical assumptions and how will you assess robustness of your results?
- If applicable, produce a diagram illustrating the approach.

**5. Results (400+ words)**

- Report your results. Include at least one table and at least one graph.
- ML/NLP Papers: Provide illustrations about how your system works and compare it to baselines. What useful information is learned?
- Econometrics Papers: For the graph, you can do an event study, or a regression discontinuity plot, coefplot, and/or binscatter. For the regression table, include multiple

specifications. (1) no fixed effects. (2) add fixed effects, (3) add trends, (4) add controls, (5) add lagged dependant variable, (6) etc.

6. Conclusion (50+ words)

- Discuss open questions or the limitations of your study. What could future work do to address these issues?

**Replication Code**

All papers must come with a replication package, with data, code, and documentation. Full credit will be given only if teaching assistants can reproduce all results in the paper.

**Tips**

[Cormac McCarthy tips on science writing](#)