**Overall Information**

* You must submit a report with code and interpretation within the same document. This report should be created within markdown or jupyter notebooks.

**Proposed Text**

* Import the text into your report from your proposal.
* If the text is one big, long string, first break into sentence segments and store it in a Pandas DataFrame.

**Fix Errors**

* Examine the text for errors or problems by looking at the text.
* Clean the data with examples from class (either impurity or example provided in NER section).

**Processing Text Summary**

* Write a paragraph explaining the process of cleaning data for a your NLP pipeline. You should explain the errors you found in the dataset and how you fixed them. Why did you think these things were important to fix for this project?

**Part of Speech Tagging**

* Tag your data with spacy’s part of speech tagger.
* Convert this data into a Pandas DataFrame.
* Use the dataframe to calculate the most common parts of speech.
* What is the most common part of speech?
* Do you see words that are multiple parts of speech?
* What can you learn about the text from examining the most common nouns and verbs?

**KPE**

* Use textacy to find the key phrases in your text.
* Using textacy utilities, combine like key phrases.
* What did you learn about your text by using keyphrase analysis?

**NER**

* Use spacy to extract named entities.
* Create a summary of your named entities.
* Apply Snorkel to your data to show any relationship between names.
* What kinds of relationships did you explore? Did you find any?

**Knowledge Graphs**

* Based on the chosen text, add entities to a default spacy model.
* Add a norm\_entity, merge\_entity, and init\_coref pipelines.
* Update and add the alias lookup if necessary for the data.
* Add the name resolver pipeline.
* Create a co-occurrence graph of the entities linked together in your text.

**Summary IE**

* Write a summary of the results from your information extraction. What did you learn about your text? What sort of relationships and entities did you find in the text? What might you consider adding?