Applied Computing Asisgnment #5: Not Reading a Boston Biography

Due: **Over Blackboard to both Professors Schmidt and Cordell**, TBD at 5:00pm.

The idea for this assignment was adapted from our colleague Paul Fyfe of North Carolina State University. Prof. Fyfe describes his version of the assignment in "How Not to Read a Victorian Novel, *Journal of Victorian Culture* 16, no. 1 (April 2011). Here's how Prof. Fyfe introduces the assignment for his students:

> Franco Moretti was dissatisfied with how literary scholars accept just a handful of possible texts as representative of cultural eras. Even if those texts are diverse and interesting, how can they possibly represent broader trends at scale? Moretti wants to change our sense of literary history by enlarging it, or by increasing our critical distance from it. He coined the phrase "distant reading" as an approach to analyzing lots and lots of texts instead of an unrepresentative few. Distant reading uses other modes of analysis and models of interpretation than the "close reading" we are familiar with. In his own work, Moretti compiles textual information from lots and lots of novels into maps, graphs, and logical trees. Seen this way, texts can reveal new patterns and language trends than we could otherwise discover close up. An array of digital visualization and text analysis tools now make Moretti's methods more accessible to the casual user. The first paper will be an experiment in using these tools. We will consider "distance" not only as the subject of our course but also as a potential mode of reading and interpretation. What does literary criticism and analysis look like if we accept distance "as a condition of knowledge"?

In this class, for instance, we only had time to read a few chapters each from the biographies of Henry Adams and Malcolm X; we learned a bit about these two important historical figures, but not nearly as much as is printed in just two–admittedly substantial–books. This challenge would only grow if we wanted to understand a larger set of biographies. Even if we only wanted to study biographies with some portion set in Boston, we would need to navigate a rather large corpus!

Before you read any of Adams or X, we experimented with computational text analysis techniques, asking what such methods could teach us about the first chapters of each book. We then read a bit of the books, thinking about how our ideas generated through machine reading helped us understand the books as human readers— or how we were mislead. For this assignment, we will build on that in-class work and "not read" the rest of these biographies. The point of this assignment will be to play, to experiment. The goal of this assignment is not, ultimately, to reconstruct a missing story, but to "read" books in a fundamentally new way, and to think through the intellectual implications of doing so.

Here are the steps you should follow:

1. **Make Some Predictions**. Choose either Adams or X for sustained analysis. Based on the chapters of these books we read in class, what do you think the rest of the books will be about? Who will be the major characters? What will be the important settings? What central themes or ideas will the book explore? Are there any stylistic characteristics you expect to continue? Be sure to write your ideas down in a document you can refer back to later.

2. **Read Bios Into R**. Read the .txt files for *The Education of Henry Adams* or *The Autobiography of Malcolm X* into R. (You can download [Malcolm X's autobiography from here](https://archive.org/stream/EnglishIslamicEbooks/BiographyC

3. Generate word clouds.

- How you might "read" this? Come up with a few different observations. What kinds of words are there? Are there patterns or inconsistencies in the words, or in what is relatively more or less frequent?

4. Compare word frequencies.

   - Does focused attention to word frequency change your opinions about your book? Which words are most common, and what might they say about the book's substance? What about scarce or infrequent words?
   - What keywords or patterns do you observe?

5. **Explore Ngrams**. Move from word frequencies to 2-, 3-, 4-, or more ngrams. What do these pairs and phrases help you understand about the major themes and ideas of the book? Compare these with the same pairs or phrases in Google's Ngram Viewer, which displays the frequency of worlds over time by drawing on the massive Google Books corpus. Look at the frequency of your biographies' prominent Ngrams through time, paying particular attention to their frequency at the times the biography was published.

   - Do any of them stand out, either as particularly common words during their time or, perhaps as interestingly, as particularly uncommon words during their time?
   - Try a few more words from the frequency lists you generated earlier. Then, try comparing some of the keywords from your chosen work with some keywords from your key work—do any interesting comparisons emerge?
   - The big question here: can a tool like the Ngrams viewer, which analyzes so many texts, help you understand anything about the historical place of a book you have never (entirely) read?

6. **Generate topic models**. Using RMallet, generate a topic model for your chosen biography. What topics does it draw from the most heavily, and what words constitutes those topics. Can you discern anything from these computationally-derived topics that helps you understand the book's themes or ideas (more colloquial "topics," if you will).

7. **Break it down**. Try breaking the book into chapters or sections. Try many of the above steps for individual sections to see what you can regenerate from a "distant" perspective. Particularly, compare the sections we read in class against the sections we did not.

8. **Compare and contrast**. In the steps above, you tried "distant reading" using a book you were at least a little bit familiar with—having read a few chapters probably made the output of these computational techniques sensible. Now you should choose another Boston biography—which you haven't read at all— from the list below. Go through all the steps above using this new text, working to glean what you can about its contents without glancing at a single line (at least a line as printed in the codex itself) from the actual book. How does your purely computational understanding of this new book compare and contrast with your understanding of a book you read partially as a narrative and partially through computational methods? Here are some titles from which you could choose; you are responsible for finding the data, though you can email us if you run into a problem.

9. **Write a Reflection**. Finally, you will write a report (in R Markdown) describing what you did and what you learned. Include at least five or six code blocks you generated in the above efforts in the flow of the text–these may generate graphs or tables that you can refer to. Please keep the emphasis on *what you learned*: a) about your chosen text, b.) about this kind of "distant reading." We're interested in your

speculations, your thoughtful reflections on text analysis. Grades will be based on how thoughtfully you engage with the assignment and how clearly those thoughts are expressed in prose. You do not need a central argument (although it is fine if you have one.) The goal of this assignment is to ruminate on what kinds of knowledge a distant reading can or cannot produce. In other words, it encourages you to think about how textual analysis changes our attention to texts. A good paper can have lots of unanswered questions. Good questions are evidence of thoughtfulness.

10. **Compile your code and submit**. Compile your R markdown reflection (with code), and export as an HTML file. Submit the file over Blackboard.

––––––––––––––––––––––––––––––––––

Possible sources for your second book

Your second book can be any book *about a person* dealing with the city of Boston. (For these purposes, it's OK if they come from a surrounding time like Cambridge, Concord, Salem, or Foxborough; please stick to Massachusetts, though). You must have not previously read it; it must have originally been printed as a **book**. One good place to find e-books to download is `archive.org` (and its sister site, `openlibrary.org`; you can plug in search terms there and find text files to download. You may need to read a little bit of a bio somewhere to ensure that your figure *is* connected to Boston.

It will be much, much easier to find full-text biographies written before 1922. That said, if you are able to find a full-text biography from *after* 1922 of a more modern figure (John Kennedy, Whitey Bulger, Elizabeth Warren, whoever) you should feel free to use it.

Some options are:

- Other Adams family biographies and autobiographies.

- Henry David Thoreau, *A week on the Concord and Merrimac Rivers*.

- William Dean Howells, *The Rise of Silas Lapham*.

- Biographies published of some of the following figures

- John Singleton Copley

- Frederick Law Olmstead

- James Michael Curley

- Isabella Stuart Gardiner

- John Winthrop

- William Lloyd Garrison

- Basically anyone name-checked in the *Education of Henry Adams*; most have biographies of their own.

- Find a figure so obscure you've never heard of him or her. One good method would be to take one of Boston's place or street names (right around here we have things named after Ruggles, Parker, Dudley), see if you can learn who they are named after, and then search Open Library for a biography.