**Reflective Paragraph on Methodology -- Kimutis**

Adding and correcting the metadata for my dataset was a relatively straightforward process, but one that nevertheless involved a number of judgement decisions. In trying to clearly account for how these decisions were made, I hope to account for some of the discrepancies between my dataset and Jessica and Ted’s, as well as to highlight some of the difficulties of a project like this.

**Category**

My first task for each entry was to determine whether or not the volume was a novel or not. For this project, I considered any longer work of fiction, including those that might typically be called a novella, a novel. Determining whether a given text was a novel was usually as straightforward as simply opening it up and skimming a few pages in HathiTrust, but even then there were a few murky cases where the lines between history, memoir, and fiction were not always clear cut. In these cases, the presence of written dialogue was usually enough for me to consider a text a **novel**.

Texts that were not novels were listed as either **shortstories**, **poetry**, or **nonfic**. I considered collections of folk myths or legends, such as Thomas Wentworth Higginson’s “Tales of the Enchanted Islands of the Atlantic,” as shortstories, even in cases where the stories were part of a cultural or religious tradition and could not be said to have been authored per se.

In some cases, a text was too recent to be in the public domain, and it was unclear based off the title whether or not it was a work of fiction. In these instances, I tried to find more information on Wikipedia and Google. In the rare case where I could not find any information about a text, I had to resort to using the metadata already supplied by HathiTrust as well as my own best guess.

**Publication Date**

In determining the date of publication for a given text, I defaulted to Wikipedia, but also used Google Books and a variety of other resources for some of the lesser-known or forgotten texts. The *At the Circulating Library* database was particularly useful for finding information on minor Victorian works.

**Author Biographical Info**

If an author had a Wikipedia page, I trusted that for their birth and death dates, nationality, and gender. If not, as with the publication date of minor texts, I did my best trying to find the information on Google. In some cases, this proved impossible. In some cases where I could not find biographically information on an author, I put in my best guess for gender based off their name.

An author’s nationality was not always clear cut. An author may be an immigrant or expatriate, or, for some older fiction, their nation of origin may not exist in its current form anymore. In the latter cases, I went with what they would be considered today: Kafka, for example, would be Czech, not Austro-Hungarian. For cases where an author moved between different countries during their life, I placed the most weight in making my decision on where their writing took place: Kazuo Ishiguro, for example, an English-language novelist who was born in Japan but moved to the UK at a young age, would be listed as from the UK in my dataset. This was not a hard and fast rule, however. In cases where I was not confident making a decision, I simply left the nationality field blank.