* Scrap the Basic/Advanced User modes
* Let’s not group our user tasks strictly into “projects” or “workflows”
  + Just allow a user to import text data and then do whatever they want
* User data storage
  + User accounts for people working with their data?
  + How to authorize user accounts - LDAP, OAuth, RIT accounts only? No accounts at all? etc.
  + Allow unauthorized users to only work with preselected corpora, and not any APIs or the ability to upload theirs
* Importing text data
  + From user uploaded files (TXT, DOC, PDF, LATEX)
  + From pre-stored corpora (Project Gutenberg, etc)
    - Look into best way to acquire Gutenberg books
  + From social media API scraping (Twitter, Facebook, Reddit)
    - Allow a user to OAuth their social accounts to get expanded access to text data
    - Check with Cissi on how she thinks about that
  + **Minimum viable product:** Project Gutenberg, Twitter without OAuth, and uploading TXT
* Text preprocessing
  + Word/Sentence tokenization (breaking raw text stream into lists by word or sentence boundary)
  + Strip extra punctuation
  + Convert contractions/abbreviations to full words
  + Make non-proper nouns full lower case
  + re[gex].search for word patterns within text
  + Stemming algorithm
  + Lemmatization algorithm
* Text data storage format
  + “Raw” preprocessed text in order
  + Unigram, bigram, trigram lists
  + Frequency Distribution of uni/bi/trigrams
* Common output format
  + Some sort of JSON for describing the type of analysis performed and the output given so the visualizer can analyze