Plan of Attack:

Scrap 3 different comic books: Saga, V for Vendetta, Watchman

figure out what stuff to include or just use entire doc

Scrape reviews?

NLP them: can reduce clusters to go down a tree

Done: Get rid of tags in string

Data product:

Send box where you can see where text came from to see synopysis

Send link to buy

Questions: figure what other NLP techniques to use for recommendation

Done: Hierarchical clustering

Extra features:

For words for typed input, you can also generate popular words from the corpus

FP growth/KNN in that select on a couple of comics and then it will tell you closest neighbor: instead us average of comic vectors with NMF; or just average word counts and then similar

Recommend more than 1 comic book

Type without \_ but with space; accepts lower case

Voting system if multiple recommenders

Do graph of how all comics connects

Use song lyrics to make predictions

Present as if you are on an island, and you want to read the must read books before you die, but don’t want to read long tomes, then you can get a mapping of books to comic books

Done: Recommendation from one comic most similar to another

Remove titles that are over-recommended

Randomly select a comic book and then select a similar one

Unlikely:

Search by comic book authors; publishing house