

Wizard of Oz

Clusters

	total_char	syl_sum	sentiment
cluster			
0	29.780802	9.312321	-0.035960
1	126.460800	39.316800	0.141200
2	62.884279	19.733624	0.130329

	total_char	sentiment
cluster		
0	29.780802	-0.035960
1	59.172854	0.151742
2	126.116348	0.113151

When comparing novels I found that generally when I excluded the syllable count that most novels followed this pattern.

Short Sentences to Low Sentiment

Long Sentences to Medium Sentiment

Medium Length Sentences to High Sentiment

*

Frankenstein and Sherlock Holmes were the two that didn't follow this pattern. However, Sherlock's sentiment clusters did not have a wide range and Frankenstein's sentence length did not have a wide variance.

I am in the process of exploring more novels.

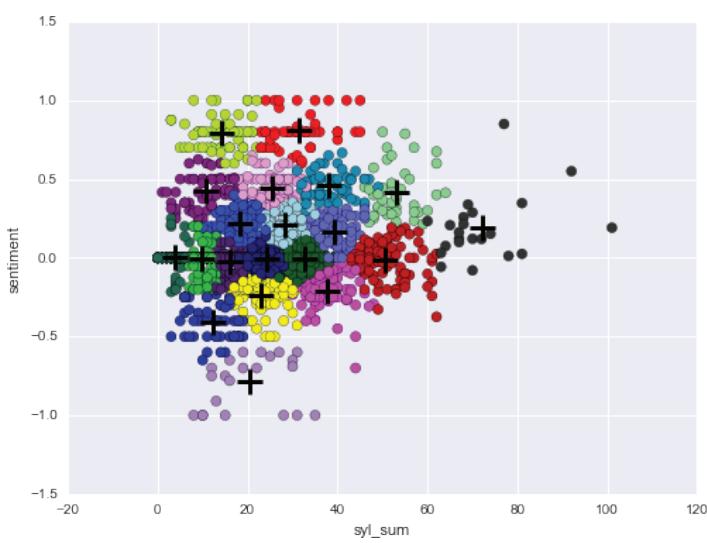
More Clusters

When I left in the syllable column and upped the number of clusters more similarities in sentence patterns arose. I'm currently working on the code to cycle through twenty piece clusters for each novel to see where similarities arise.

	total_char	syl_sum	sentiment
cluster			
0	167.555000	52.155000	0.172032
1	29.500719	9.220144	-0.032877
2	55.243506	17.240260	-0.010549
3	108.111345	33.489496	0.053553
4	71.119048	22.813492	0.603102

Ideal Cluster Number

Twenty Clusters Seemed to be good amount for the Wizard of Oz. Breaking the novel up in this way gave me a unique look at the type of sentences the book is made up of.



	total_char	syl_sum	sentiment
cluster			
0	57.745098	18.245098	0.222249
1	121.710843	37.602410	-0.207736
2	31.404762	9.845238	-0.007241
3	99.106383	31.361702	0.806418
4	75.204545	22.965909	-0.236287
5	76.884211	24.068421	-0.006626
6	79.022222	25.488889	0.444507
7	167.348837	53.093023	0.417853
8	163.597938	50.608247	-0.015971
9	44.138462	14.169231	0.790872
10	38.377358	12.433962	-0.405454
11	33.537313	10.641791	0.422822
12	105.981595	32.748466	-0.009319
13	50.706897	15.922414	-0.019478
14	12.948127	3.881844	-0.000583
15	118.142857	38.125000	0.458200
16	128.229167	39.343750	0.167051
17	61.681818	20.318182	-0.786591
18	237.954545	72.136364	0.193617
19	90.806452	28.338710	0.209662

Individual Web Pages

In the next month I want to create webpages for several novels that display artfully how the novel is unique and similar to other novels. The pages will be interactive, using d3, allowing the user to explore the different types of sentences. To help with this I will look for the right number of clusters for each novel.

Sentence Examples - Wizard Cluster

The Discovery of Oz the Terrible .

When you get to Oz do not be afraid of him, but tell your story and ask him to help you.

"For," he said, "if I should get caught in the rain, and rust again, I would need the oil-can badly."

"I am terribly afraid of falling, myself," said the Cowardly Lion, "but I suppose there is nothing to do but try it.

"They must be dreadful beasts."

Those creatures frightened me so badly that my heart is beating yet!"

he called after them, and they were very sorry to leave him.

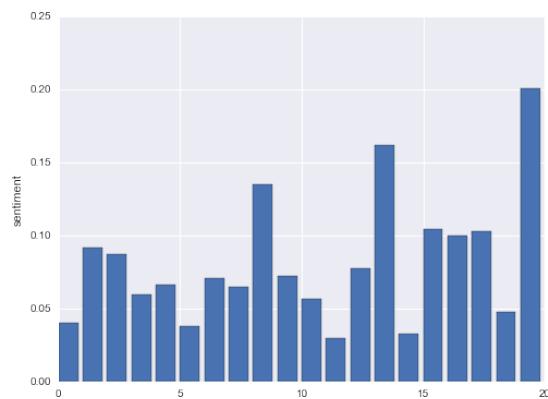
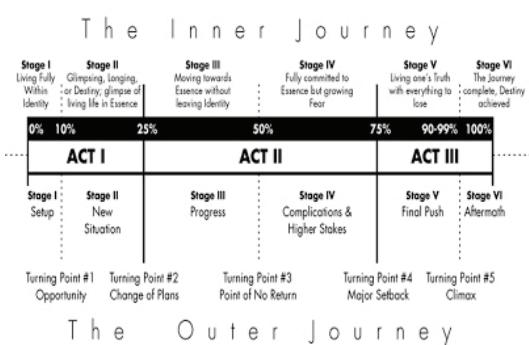
Of course this was a bad thing for the Scarecrow.

I am afraid I shall never have any brains, after all!"

"Why do you wish to see the terrible Oz?"

Sentiment Patterns

Michael Hauge's "Six Stage Plot Structure"



Sentences from the 75% Mark, Turning Point 4, and Major Setback

"We have lost our way," said Dorothy.

"Can you tell us where the
Emerald City is?"

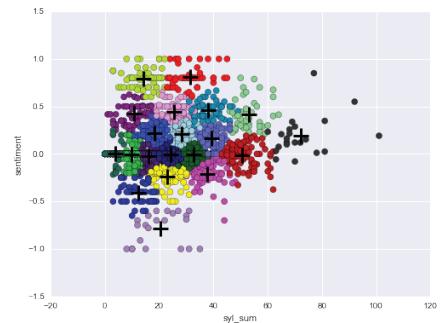
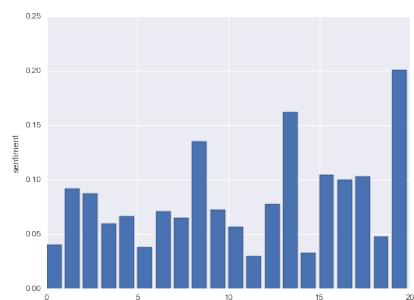
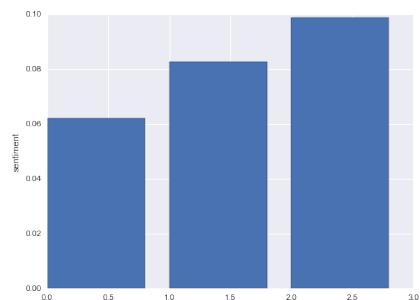
"Certainly," answered the Queen; "but it is a great way off, for you
have had it at your backs all this time."

Then she noticed Dorothy's
Golden Cap, and said, "Why don't you use the charm of the Cap, and call
the Winged Monkeys to you?
They will carry you to the City of Oz in
less than an hour."

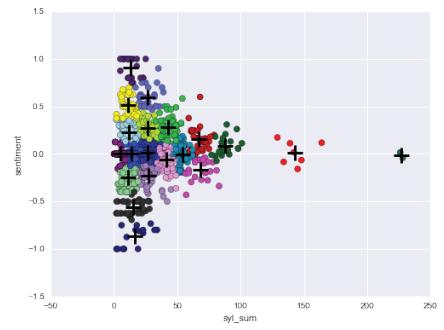
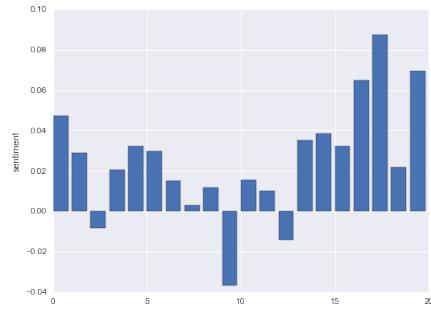
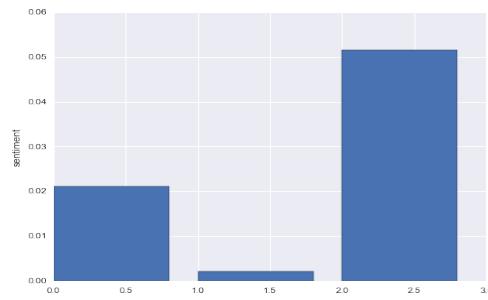
Future - add in parts of speech & content clustering .

How the Novels Compare

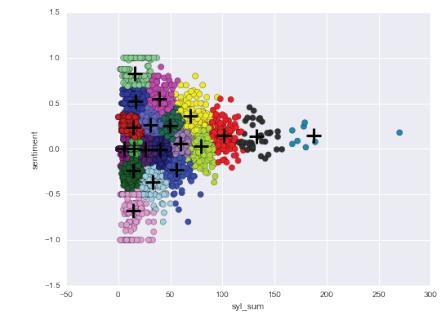
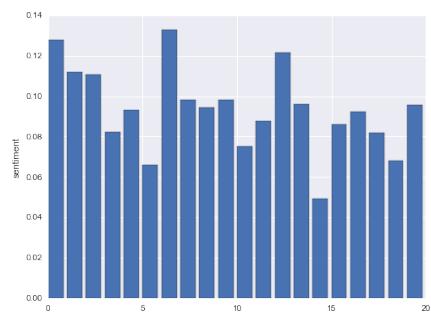
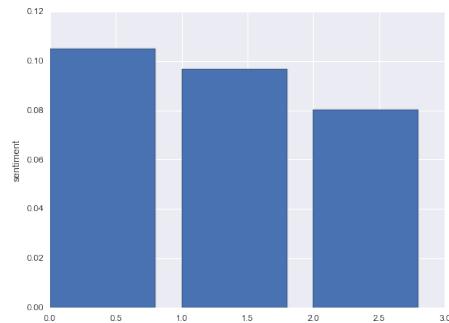
Wizard of Oz



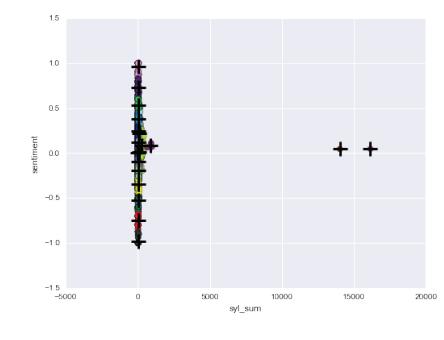
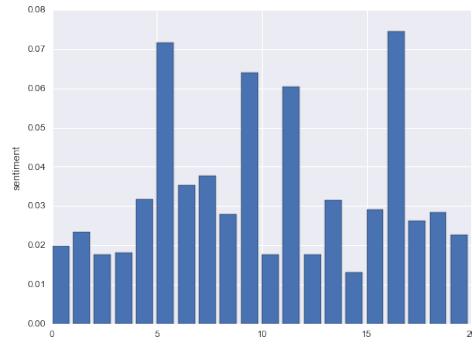
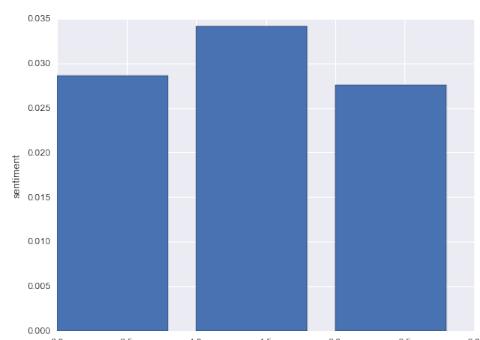
Alice in Wonderland



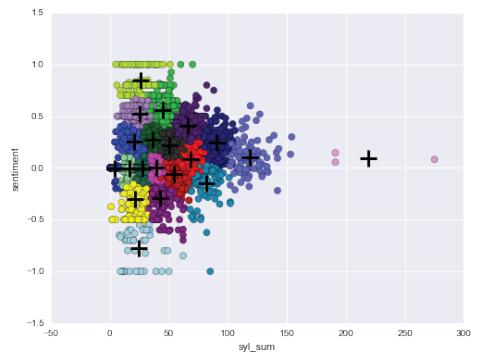
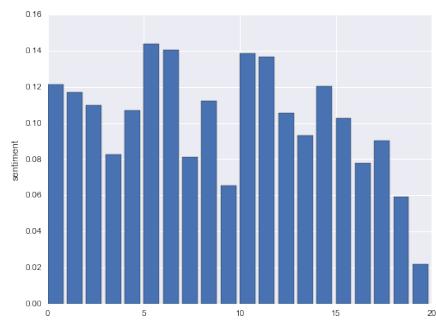
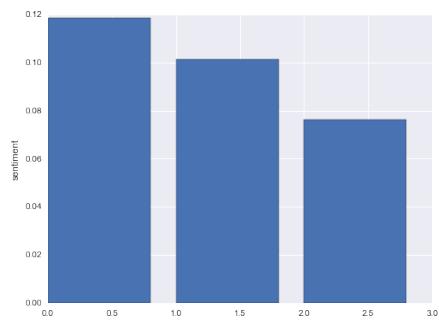
Pride and Prejudice



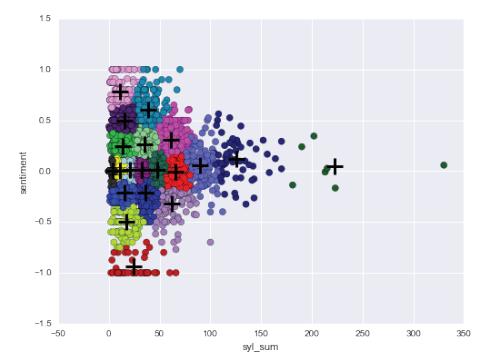
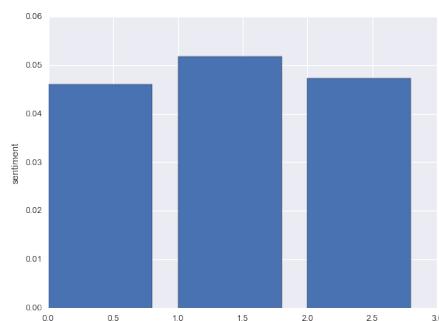
Ulysses



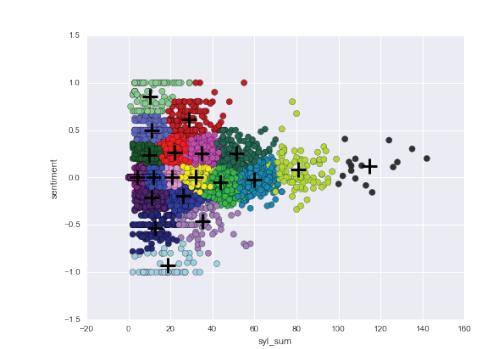
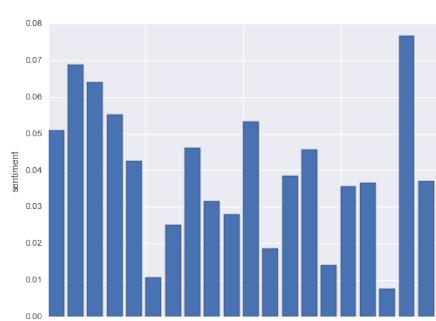
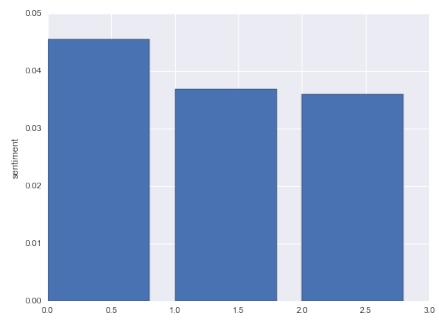
Odyssey



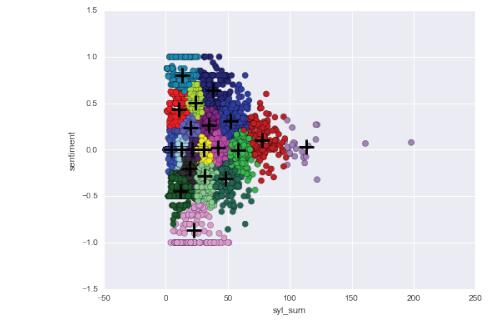
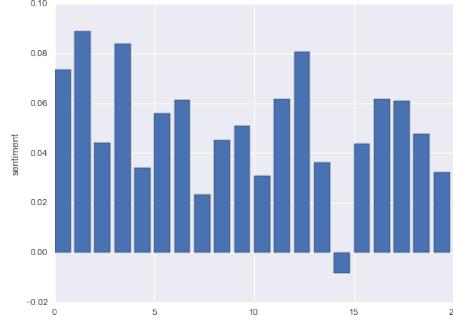
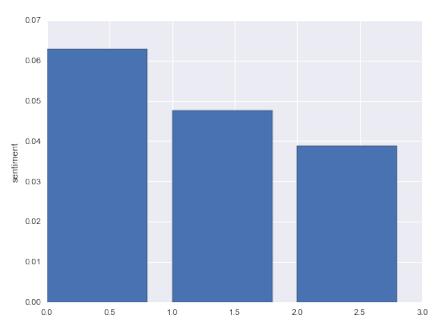
Tale of Two Cities



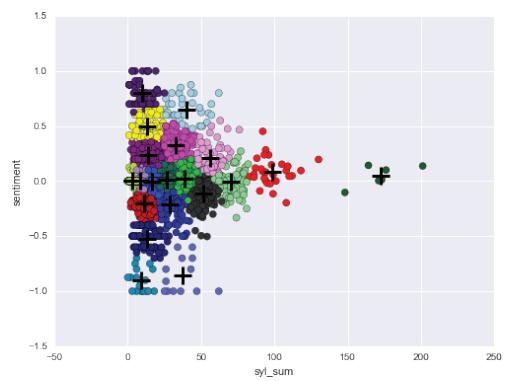
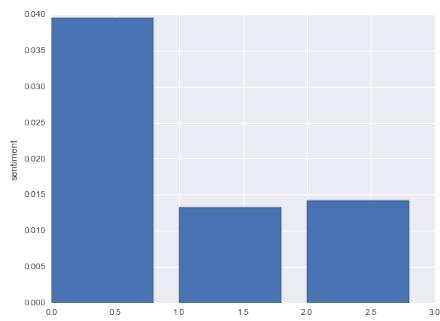
Sherlock Holmes



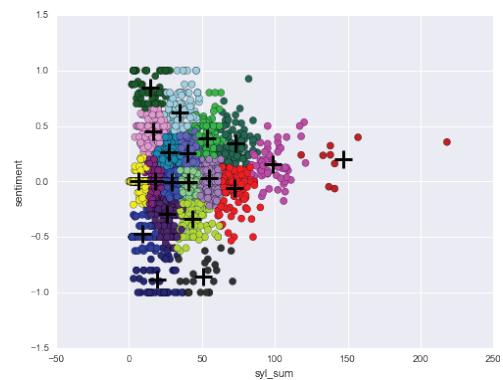
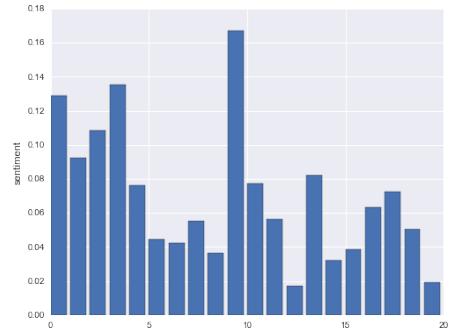
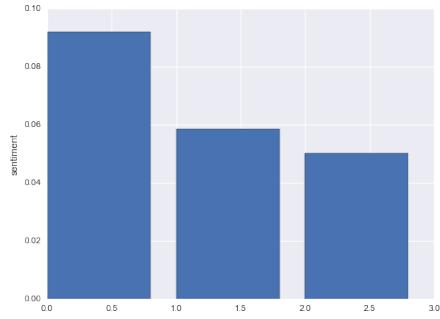
Dracula



Jungle Book



Frankenstein



Syllable Count

I will work on improving the syllable count code and then display the unique syllable counts for each novel on their web pages using d3 as their story finger print or snow flake - something like that.

Future



Future goals for the project:

Learn how novels are written.

Predict which new novels will be popular
with the aide of social media and online reviews.

Write.

