Best selling Books: Data Analysis

Weekly Update
https://github.com/sruthipsuresh/best-sellers-analysis

Weekly Summary

- Augmented dataset with new specific books (and kept all information).
- Ran all previous analyses (figured out better visualization for clustering).
- Ran LDA + STM with metadata (author).
- All new results uploaded to Github.

Data Collection and Cleaning

Book List and Terms

gBooksfinal.ipynb > gbooksprocessing.ipynb

Data Collection

- **NOTE:**

  ISBNs need to be lifted from Google Books (some variation, fixed it with notebook code).

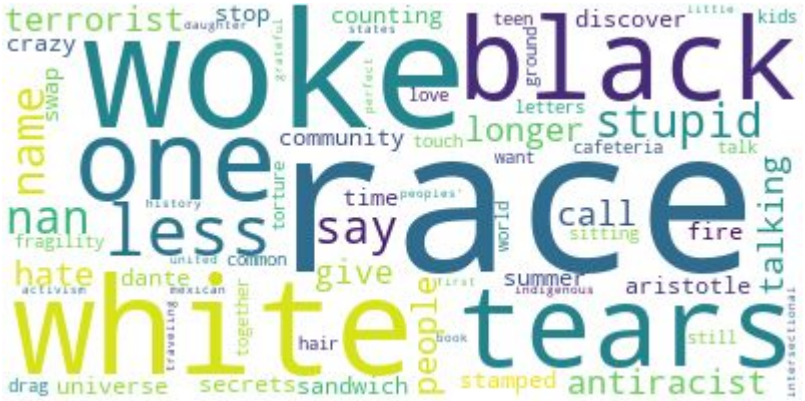
  A Mind's Voice: Helen Khan

  Same, But Different by Megan Curtis

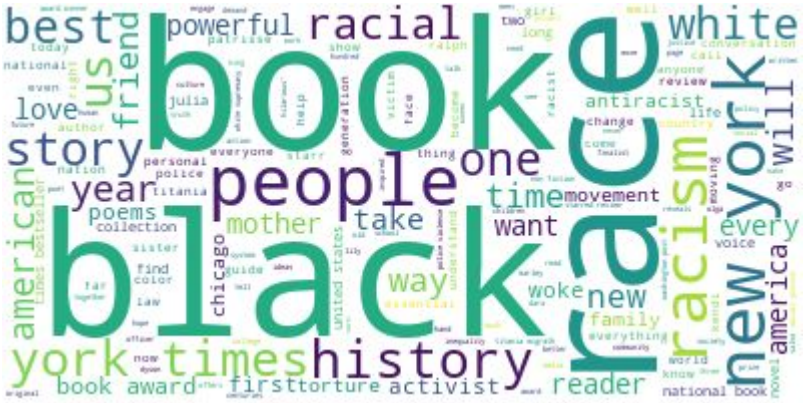
	kind	totalItems	items__kind	items__id	items__etag	items__selfLink	items__volumeInfo__title
0	books#volumes	1.0	books#volume	KE1EDwAAQBAJ	qWBLM38WKwE	https://www.googleapis.com/books/v1/volumes/KE1EDwAAQBAJ	How to Be Less Stupid About Race
1							
2	books#volumes	1.0	books#volume	C4EBwgEACAAJ	eyiTj6QNPVI	https://www.googleapis.com/books/v1/volumes/C4EBwgEACAAJ	Say Her Name
4	books#volumes	1.0	books#volume	JjvwvQEACAAJ	vNKstkMvPtI	https://www.googleapis.com/books/v1/volumes/JjvwvQEACAAJ	Woke
6	books#volumes	1.0	books#volume	6pNbDwAAQBAJ	KWz+4ssclI3c	https://www.googleapis.com/books/v1/volumes/6pNbDwAAQBAJ	How to Be an Antiracist
8	books#volumes	1.0	books#volume	LMovDgAAQBAJ	6L1dMtFNgdM	https://www.googleapis.com/books/v1/volumes/LMovDgAAQBAJ	Why I'm No Longer Talking to White People About Race
10	books#volumes	1.0	books#volume	AerMswEACAAJ	sRptP+HshEY	https://www.googleapis.com/books/v1/volumes/AerMswEACAAJ	And She Was
12	books#volumes	1.0	books#volume	6l4mDwAAQBAJ	nvr324fd2mI	https://www.googleapis.com/books/v1/volumes/6l4mDwAAQBAJ	When They Call You a Terrorist
14	books#volumes	1.0	books#volume	OTwYDQAAQBAJ	/wovxHoYBVI	https://www.googleapis.com/books/v1/volumes/OTwYDQAAQBAJ	The Hate U Give
16	books#volumes	1.0	books#volume	bqgHrgEACAAJ	IM14+H922vk	https://www.googleapis.com/books/v1/volumes/bqgHrgEACAAJ	Counting on Community

Wordclouds.

Title

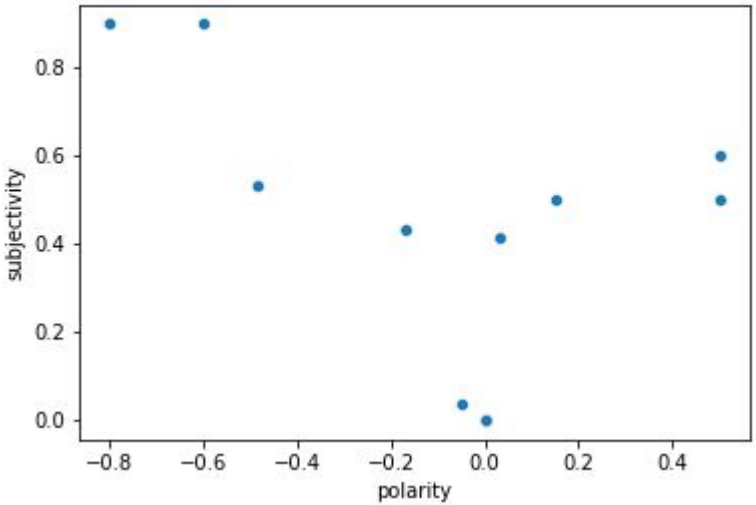


Description

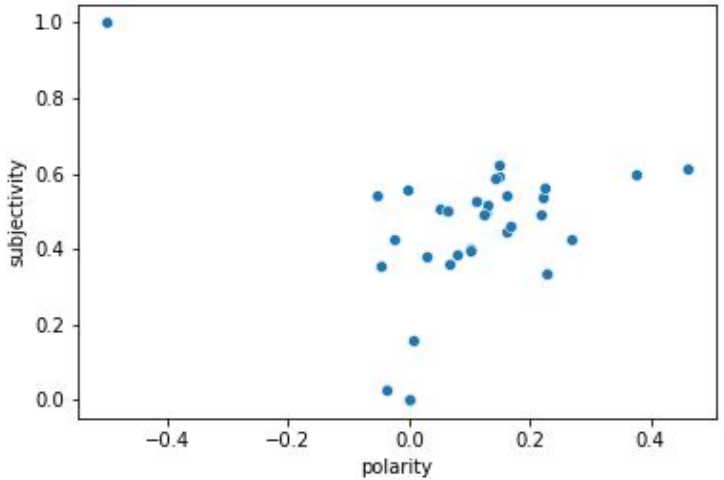


Sentiment Analysis.

Title



Description



Clustering.

Title

cluster3	cluster4	cluster5
		how to be less stupid about race
		how to be an antiracist
	why i'm no longer talking to white people about race	
the hate u give		
aristotle and dante discover the secrets of the universe		
the sandwich swap		
the fire this time		
	white fragility	
why are all the black kids sitting together in the cafeteria		
the torture letters		
		so you want to talk about race
between the world and me		
		an indigenous peoples' history of the
traveling black		
	white tears	
we are grateful		

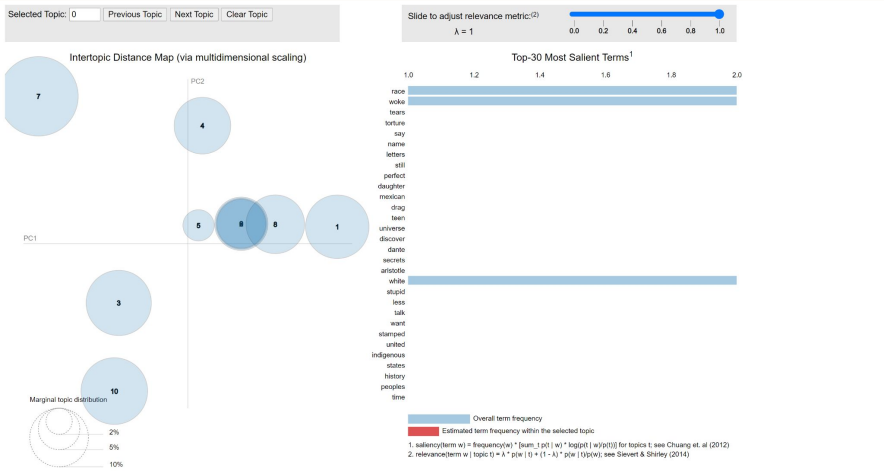
Description

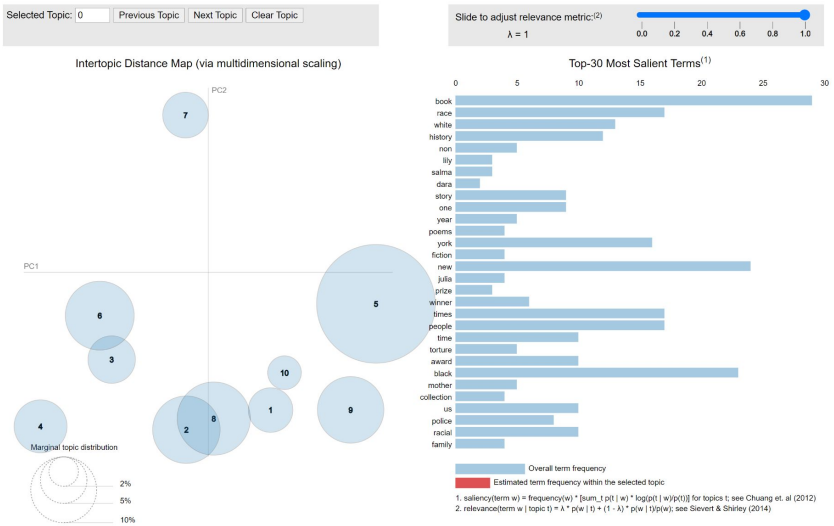
cluster2
how to be less stupid about race
the fire this time
why are all the black kids sitting together in the cafeteria
so you want to talk about race

Topic Modeling (STM, Python).

Title

Description

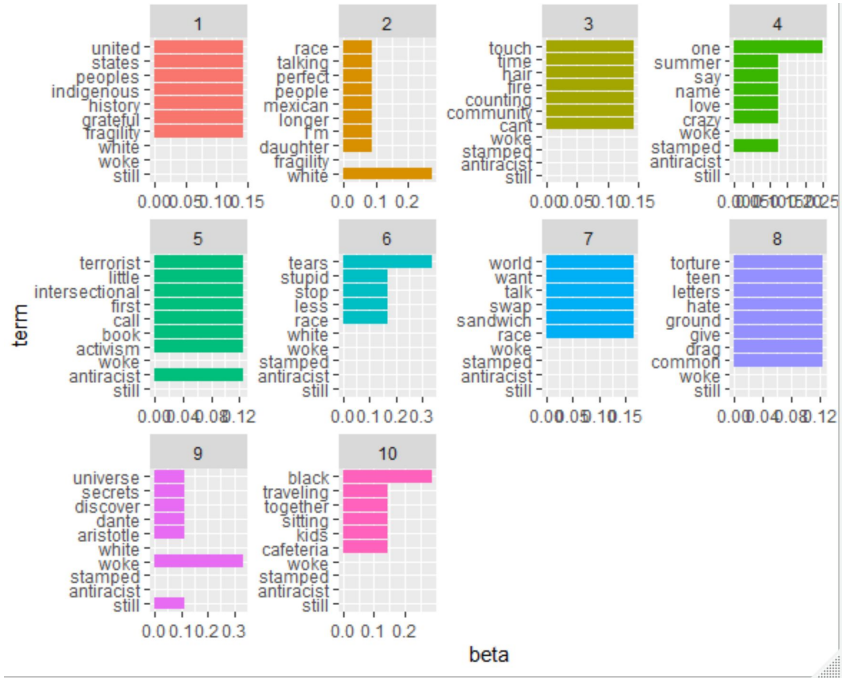




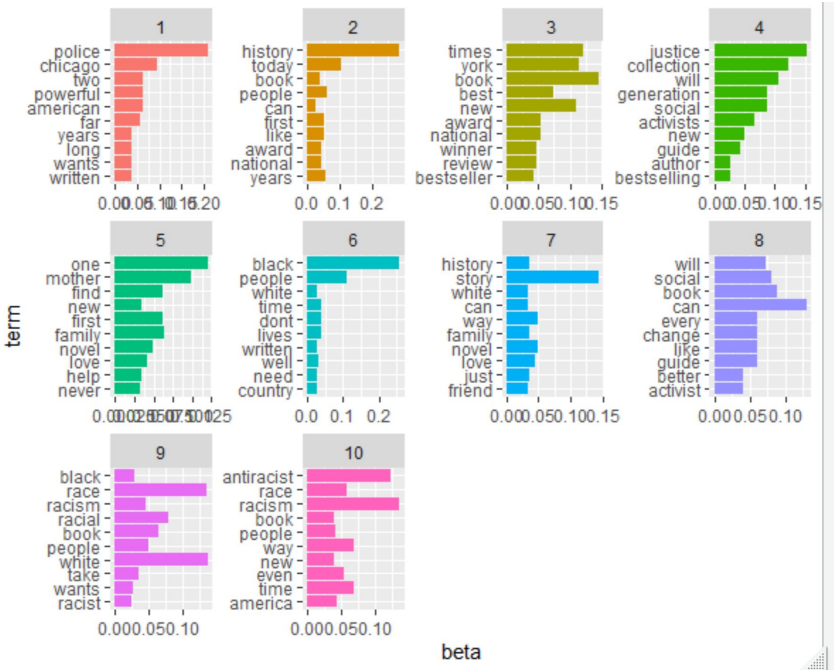
Modeling using R

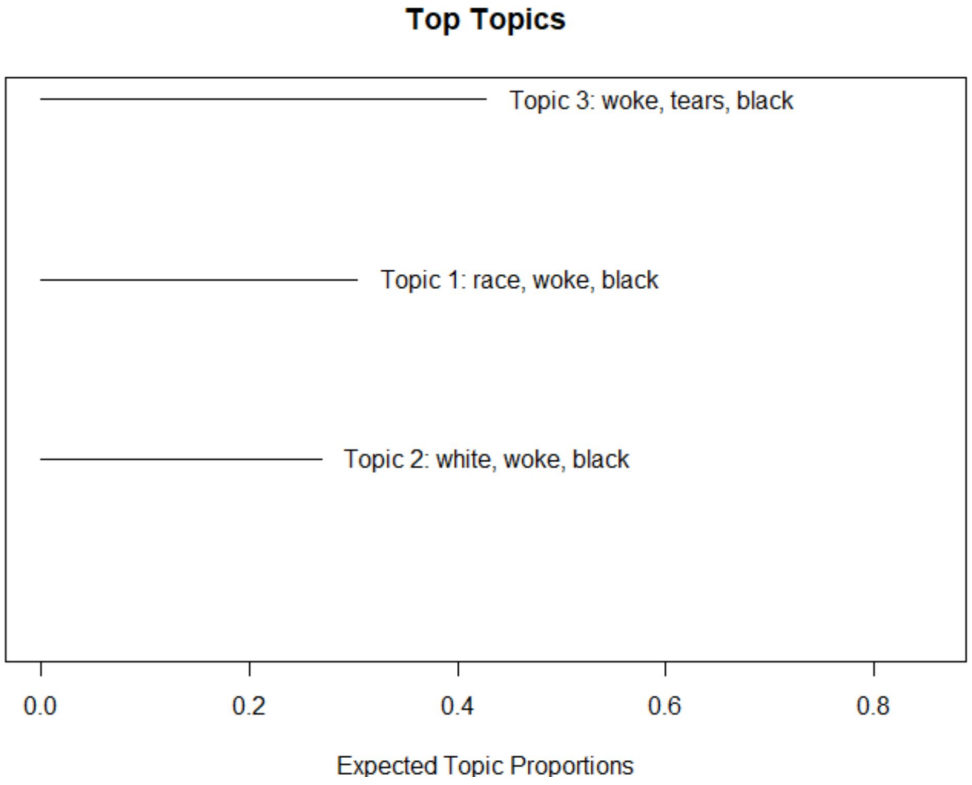
ldamodeling.R
stmmodeling.R

TITLE

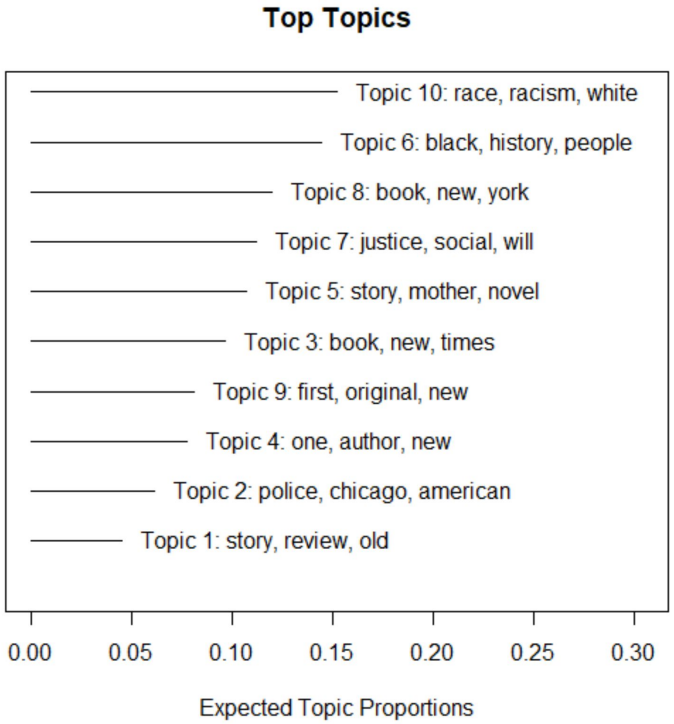


DESCRIPTION



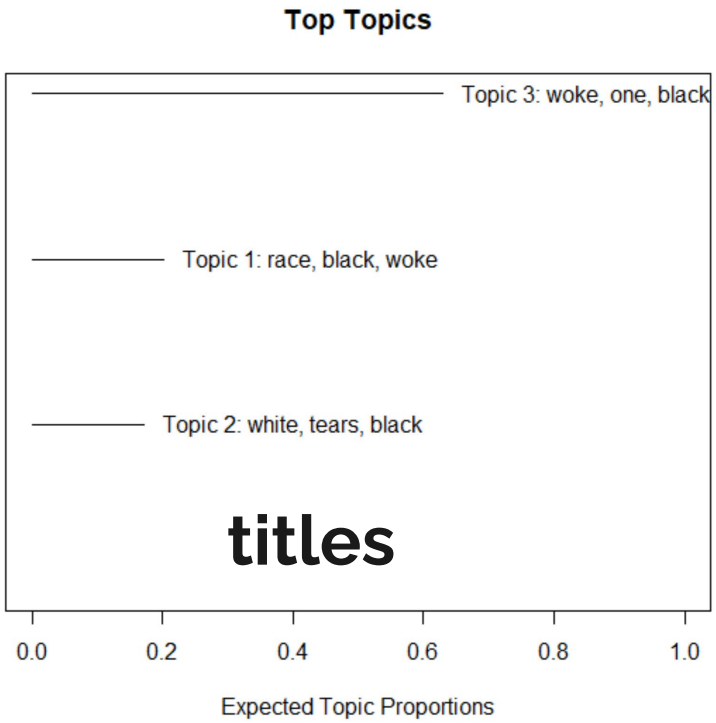


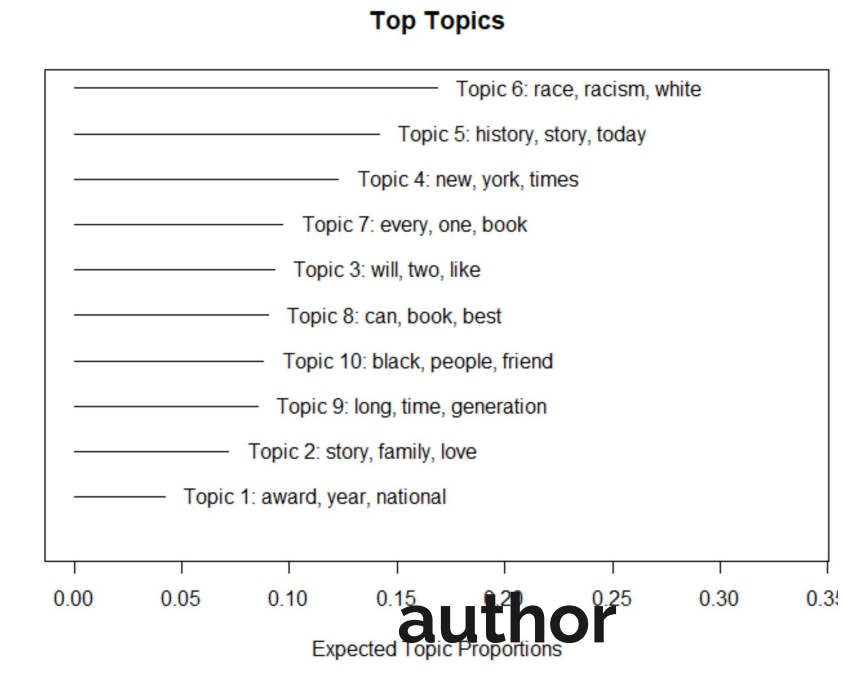
titles



description

With author metadata





Conclusion

Plans for this week:

- Manually categorize/label books with clusters/topics.
- Compare to clusters made by LDA/STM.
- Write code to get years.
- Read other STM model topics and write code.
- Add stop words (such as book).
- Arbit. Position book (need negative books)
- Add bigram (New York) for better textual analysis.
- 7 topics LDA.
- Positional Scaling
- Check out quanteda.
- Visualize > visualization for hypothesis formation > structural analysis