Reading Poetry in the Age of *milk and honey*

Hannah Shlesinger
Dr Anderson
DigHum 100
Summer Sessions
30 May 2021

Projection Description:

In recent years, poetry sales have increased dramatically, notably after 2014 when *milk and honey* by Rupi Kaur hit the shelves and the internet. The goal of this project is to track how themes and reader demographics have changed over time, to see to what extent *milk and honey* revolutionized contemporary poetry.

Some call her an "insta-poet" for her sucess both as a writer and an influencer online. Part of the investigation will be to identify to what extent poetry has been commodified. My goal isn't to say whether these poets are 'good' or 'bad' but rather to analyze the types of poetry being consumed en masse by the public and to figure out what exactly draws people to these poems or poets.

[New] Guiding Questions:

- How has Rupi Kaur's work developed across her three published collections?
- How has her following developed and changed over time?
- How has she influenced the works of other contemporary poets?
- How has the 'book community' received her work over time?
- How are their poems structured?
- Do they contain visual elements (illustrations, drawings, photographs, etc.)?

[New] Process:

Based on interest and popularity, I have identified Rupi Kaur as a particularly influential poet. I am basing influence on social media following, Goodreads choice awards, and the Amazon best seller list.

First, I want to conduct analysis on Kaur's work to see how she presents her poetry and what kind of topics she discusses. I am interested in her incorporation of visuals, the length of her lines and poems and the themes she discusses in her work.

Next, I want to see how her work compares to other poets who have published after her. I am still in the process of selecting collections and works, but I am currently considering Lang Leav, Amanda Lovelace, Halsey, Lana Del Rey, and Lilli Reinhart.

Finally, I would like to analyze how reviews of Kaur's work has developed overtime to see if interest or general atmosphere in the book community has shifted since the publication of *milk and honey* . I am interested to see to what extent other authors or works are mentioned in Kaur's review sections. I will probably use either the Amazon review section or the Goodreads review section.

Sources of Data

- Text mined directly from selected works, converted into .csv files and assessed in Jupytr Notebooks
  [https://colab.research.google.com/drive/1EOSdBhcsj5Valjvmfoeqx5Ksl9JydMWA?usp=sharing]
- Twitter & Instagram folower demographic data extracted using APIs
- Goodreads & Amazon reviews of Rupi Kaur's work extracted using APIs

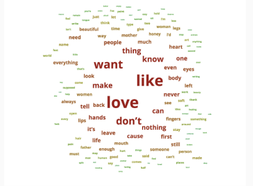
Initial Data Findings: I will probably select the poem collections that were past winners or nominated in the Poetry category for the GoodReads Choice Awards from the past few years. I may select other texts based on popularity on social media (determined by poet following or hashtags, but I still have to do more research o this)

GoodReads Choice Awards Poetry Winners

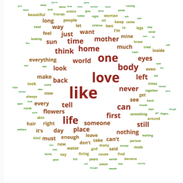
Year	Poet	Title	Votes	Total Votes
2011	Billy Collins	*Horoscope for the Dead*	1,260	7,711
2012	Mary Oliver	*A Thousand Mornings*	5,520	15,905
*2013	J.R.R Tolkien	*The Fall of Arthur*	15,096	47,441
2014	Lang Leav	*Lullabies*	9,571	75,474
2015	Trista Mateer	*The Dogs I have Kissed*	8,725	61,428
2016	Amanda Lovelace	*the princess saves herself in this one*	22,560	95,646
2017	Rupi Kaur	*the sun and her flowers*	35,856	118,052
2018	Amanda Lovelace	*the witch doesn't burn in this one*	25,298	141,883
2019	Laurie Halse Anderson	*Shout*	18,911	128,340
2020	Margaret Atwood	*Dearly*	22,811	164,701

*2013 Amazon acquires Goodreads

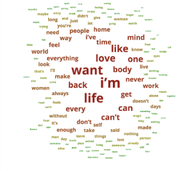
OverviewDocs Text Visualization of *milk and honey*



OverviewDocs Text Visualization of *the sun and her flowers*

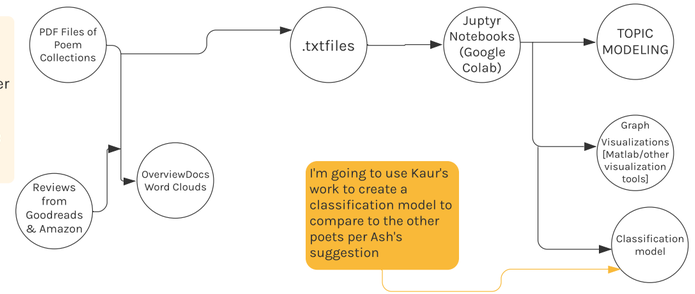


[New] OverviewDocs Text Visualization of *homebody*



Workflow

(process modeled after Jefferey Cheng's Poster 2020)



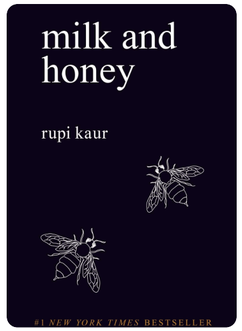


image source: https://www.amazon.com/Milk-Honey-Rupi-Kaur/dp/1449474256

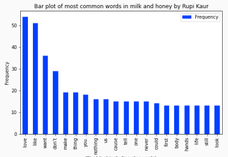
Topic Modeling for Amazon & Goodreads Reviews

This is where the visualization for the topic modeling of poetry reviews will go

[New]: This process will happen after my text mining/analysis, so it may end up being cut depending on how things develop

Stanza / Text Analysis

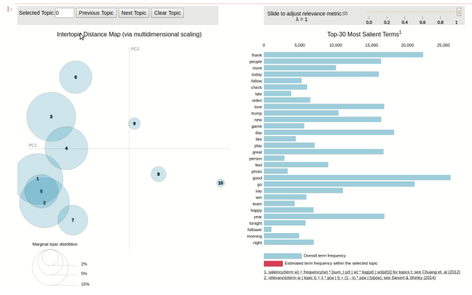
This is where visualization for the distribution of stanaza length, number of lines, etc will be included.



[New]first visualization attempt (word frequency visualization --> will probably find another more appropriate model)! :D

Topic Modeling for Poetry collections

This is where the visualization for the topic modeling of the actual texts would go to network out the specific use of words. I am also considering using WordEmbedding to compare the placement of keywords (love, like, want) on individual lines, although I will have to investigate this further. Below is an example of an LDA Mallet topic model that looked promising.



placeholder visualization from:
https://neptune.ai/blog/pyldavis-topic-modelling-exploration-tool-that-every-nlp-data-scientist-should-know

Results & Notes

I anticipate a lot of the poems will be short, perhaps no more than 4-8 lines in length. I also anticipate many of the poems will be about love and relationships based on the initial OverviewDocs word cloud visualization of *milk and honey*. I imagine popularity and follower demographics will indicate a young, primarily female audience of the poets, but I am also anticipating some outliers.

To an extent, popularity in numbers only tells us so much about how the poets and poems are being received and perceived by their audiences. In addition to this data, one hole I might try to fill is to run Amazon and Goodreads reviews through the same text mining process to see what the common responses to Kaur and other poets' work is.

Note: a lot of books sales data is private, so Goodreads and Amazon seem like the best alternatives to find book-related information, although I do plan on investigating this further

Sources & References:
https://www.theguardian.com/books/2019/jan/21/poetry-sales-soar-as-political-millennials-search-for-clarity
https://www.theatlantic.com/entertainment/archive/2018/08/when-poetry-isnt-poetry/567571/
https://www.theatlantic.com/entertainment/archive/2016/11/still-poetry-will-rise/507266/
https://www.theatlantic.com/technology/archive/2018/10/rupi-kaur-instagram-poet-entrepreneur/572746/
https://journals.library.columbia.edu/index.php/TMR/article/view/3456
https://www.goodreads.com/choiceawards/best-books-2020
https://www.katiejolly.io/blog/2018-01-05/random-rupi-markov-chain-poems
https://github.com/albertkx/rupi-kaur-poetry
https://pypi.org/project/amazon_scraper/
https://eds-a-ebscohost-com.libproxy.berkeley.edu/eds/pdfviewer/pdfviewer?vid=4&sid=94dc0324-2e03-475d-85ef-12b21e8dec3d%40pdc-v-sessmgr01
A Modern Analytical Comparison of Biblical Canon and Apocrypha Poseter Draft by Jefferey Cheng, June 7, 2020
https://towardsdatascience.com/a-complete-exploratory-data-analysis-and-visualization-for-text-data-29fb1b96fb6a
https://pypi.org/project/twitter-scraper/