Reading Poetry in the Age of milk and honey: A computational text analysis of Rupi Kaur's poetry

Hannah Shlesinger Dr Anderson DigHum 100 30 May 2021

Projection Description:

In recent years, poetry sales have increased dramatically, notably after 2014 when *milk* and honey, a poetry collection by Rupi Kaur, hit the shelves and the internet. The goal of this project is to understand what makes Kaur's poetry so popular.

To do so, I am going to be taking a distant reading approach to Kaur's three collections: milk and honey (2014), the sun and her flowers (2017), and homebody (2020). This poster is part of a larger investigation to understand what draws contemporary audiences to poets and their work. The rest of the investigation will focus primarily on community reaction quantified by reviews of Kaur's work on sites like Goodreads and Amazon.

Guiding Questions:

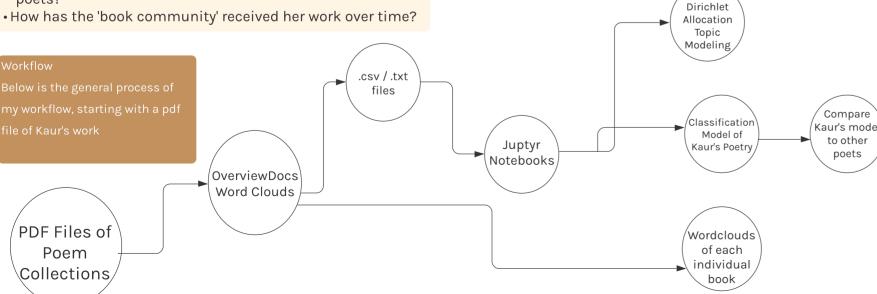
- How has Rupi Kaur's work developed across her three published collections?
- What are the major themes of her work and how are they related?
- What are the unique structural elements of her work?
- How has she influenced the works of other contemporary poets?

Process:

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.



Sources of Data

• pdfs of Rupi Kaur's poetry (milk and honey, the sun and her flowers, homebody) which sum to over 500 individual poems

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

<u>Year</u>	<u>Poet</u>	<u>Title</u>	<u>Votes</u>	<u>Total Votes</u>
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

Poets and collections chosen for potential comparison*:

- Lang Leav
- · Amanda Lovelace
- Halsey
- Lili Reinhart
- *subject to change because of access and copyright issues [if I can't get access to a large enough size, I may instead contain scope to just looking at Kaur's work]

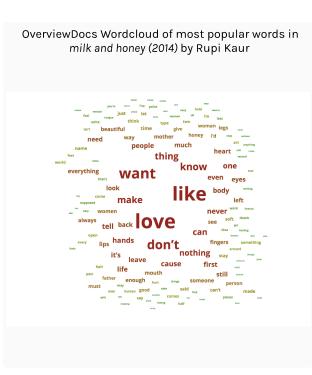
Selection process and justification:

I disregarded both Margaret Attwood and Laurie Halse Anderson since they are already acclaimed novelists. I was particularly interested in looking at popular collections published by poets or influential non-writers who published collections.

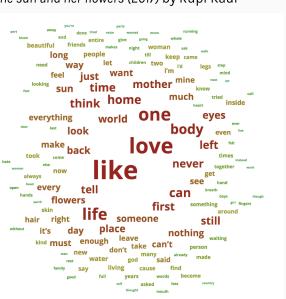
Page 2: Visualizations & Results

OverviewDocs:

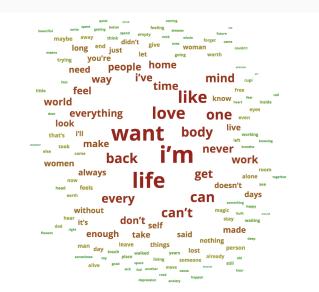
To the right are wordlclouds of each of Rupi Kaur's collections, displaying the words that show up with the highest frequency in each collection.



OverviewDocs Wordcloud of most popular words in the sun and her flowers (2017) by Rupi Kaur



OverviewDocs Wordcloud of most popular words in homebody (2020) by Rupi Kaur



Results & Notes

This investigation incorporates a distant reading approach which emphasizes using data and techniques performed by computers that could not reasonably be done by a person.

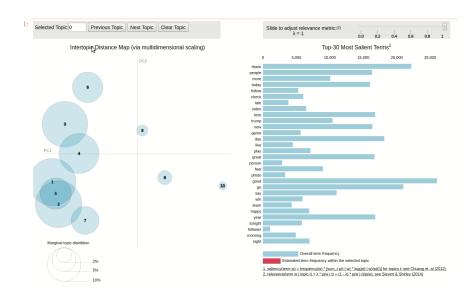
As I somewhat expected, my analysis revealed that a lot of common themes in Kaur's work relate to love in desire with the overwhelming top words being "want," "like," and "love."

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.

More results & analysis[TBD]

Topic Modeling for Poetry collections

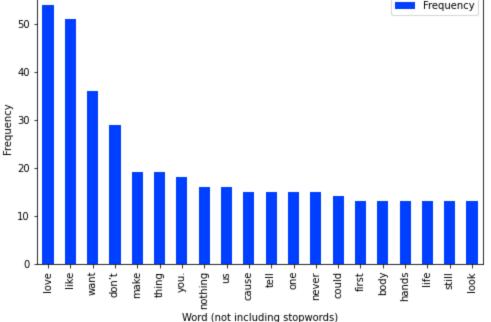
This is where the visualization for the topic modeling of the actual texts would go to network out the specific use of words. I am also considering using WordEmbedding to compare the placement of keywords (love, like, want) on individual lines, although I will have to investigate this further. Below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

Bar plot of most common words in milk and honey by Rupi Kaur



This is also just a placeholder visualization I created in Google CoLab. I will probably visualize for all of the collections in a bar plot and just use the wordclouds for individual collections

Expand on the concept of close vs distant reading?



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/