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' 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.

Guiding Questions:

- Who are the most popular poets in the last ten years?
- Who are their readerbases?
- What types of following do they have?
- What are these poets writing about?
- How are their poems structured?
- Do they contain visual elements (illustrations, drawings, photographs, etc.)?

Process:

First, I want to identify the most popular poems and poets in the past few years. I will find multiple sources of data to create a network of 'popularity.' I have started with initial *Goodreads* data from the winners of the *Goodreads* Choice Awards. I am using Rupi Kaur as my main model since her numbers across platforms in my preliminary research stand out significantly.

Using text mining and computational analysis, I am seeking to find which themes and words are most prevalent in Kaur's poetry and how her work compares to the other bestsellers since the publication of *Milk & Honey*. Specifically, I will be investigating the stanza length and form and common imagery or descriptive language.

Although it depends on my scope, I would additionally like to select a few poets with strong social media followings and use instagram/twitter API services to get a better sense of their readerbase and map out the connections between their respective followings online.

Sources of Data

- Goodreads Choice Awards 2012-2019 scraped using API (Alternatively, this Keggle data set)
- Amazon Bestsellers List 2012-2019 scraped using API (Alternatively, this Keggle dataset)
- Text mined directly from selected works, converted into .csv files and assessed in Python
- Twitter & Instagram folower demographic data extracted using APIs

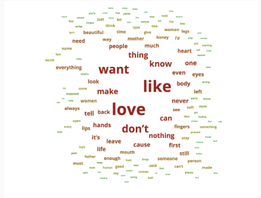
Initial Data Findings: I will probably select the poem collections that were past winners or nominated in the Poetry category for the GoodReads Choice Awards from the past few years. I may select other texts based on popularity on social media (determined by poet following or hashtags, but I still have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

GoodReads Choice Awards Poetry Winners

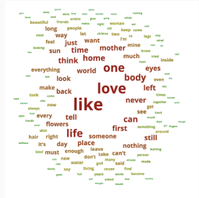
Year	Poet	Title	Votes	Total Votes
2011	Billy Collins	*Horoscope for the Dead*	1,260	7,711
2012	Mary Oliver	*A Thousand Mornings*	5,520	15,905
*2013	J.R.R Tolkien	*The Fall of Arthur*	15,096	47,441
2014	Lang Leav	*Lullabies*	9,571	75,474
2015	Trista Mateer	*The Dogs I have Kissed*	8,725	61,428
2016	Amanda 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*

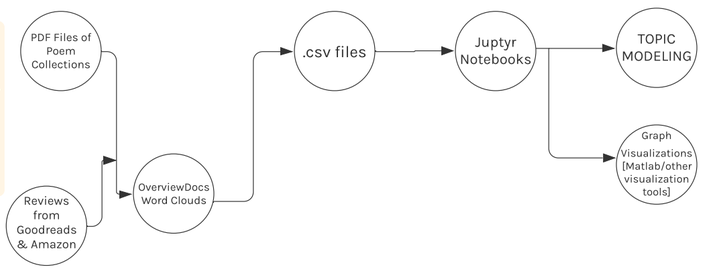


OverviewDocs:

I want to make OverviewDocs word clouds for *milk and honey* as well as a wordcloud with all of the popular poem collections in one cloud

Workflow

(process modeled after Jefferey Cheng's Poster 2020)



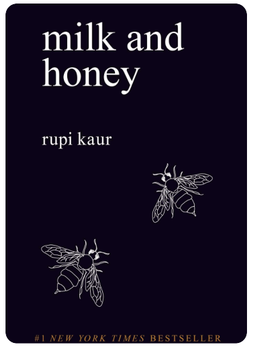


image source: https://www.amazon.com/Milk-Honey-Rupi-Kaur/dp/1449474255

Topic Modeling for Amazon & Goodreads Reviews

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

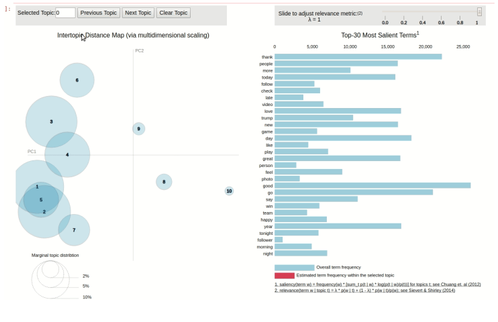
Stanza / Text Analysis

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

place-holder visualization from: https://towardsdatascience.com/a-complete-exploratory-data-analysis-and-visualization-for-text-data-29fb1b96fb6a

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/