

# TOPIC MODELING AND ANALYSIS: COMPARING THE MOST COMMON TOPICS IN 19<sup>TH</sup>-CENTURY NOVELS WRITTEN BY FEMALE WRITERS

LAUREN RHA,  
SEAN SILVER (FACULTY ADVISOR)

## \* ABSTRACT

Women authors from the 19<sup>th</sup>-century have had a profound impact on the literary world due to their critical approach to and inclusion of various social phenomena within their work, such as women's rights, sexuality, and human psychology. This paper seeks to contribute to the discussion by quantifying thematic similarities in eight select novels by various female authors of the 19<sup>th</sup>-century. These novels were chosen due to their contribution to literature and their popularity, common use in college courses around the world, and the prominence of the female authors. This study included utilizing a programming environment known as R Studio to perform a topic model. Performing a topic model allowed for the discernment of ten main themes or topics that can be generally seen across all eight selected novels, and by extension, 19<sup>th</sup>-century literature by female authors. The research found initial evidence to support the general understanding of said literature as an endeavor of the themes of social critique and individual consciousness; however, the results were not absolute in conclusion because of the limited size of the corpus. A larger corpus of documents (novels) is necessary to reach further conclusions.

## 1 INTRODUCTION

In the 19<sup>th</sup>-century, iconic texts by female authors revolutionized prose fiction and imaginary work written in narrative form through their criticism of social expectations for women, commentary on the class system and expressions of sexuality, and unique focus on characters' psychological states and moral developments. This paper seeks to contribute to the discussion by quantifying and analyzing these thematic similarities in eight select novels by the following female authors of the 19<sup>th</sup>-century: Jane Austen, Charlotte Brontë, Emily Brontë, Anne Brontë, George Eliot, Louisa May Alcott, and Elizabeth Gaskell. Performing such a topic model on literary studies is not yet widely done, although it has proven to successfully extract themes and topics from word-frequency data. In doing so, perhaps computationally generated themes can be formed and present a new perspective on common and well-studied themes in the 19<sup>th</sup>-century literature of various female authors.

This research uses an approach known as "distant reading" for its analysis. "Distant reading" demonstrates the value of reading large numbers of text together and relies on computer-assisted modeling to analyze a larger amount of texts than can be read by any one individual.<sup>[15]</sup> "Distant reading" and the specific practice of topic modeling was performed within the R Studio programming environment. The R Studio programming environment is a development environment for R, a programming language used for statistical computing and graphics. The novels analyzed by the author included Jane Austen's *Emma*, Charlotte Brontë's *Jane Eyre*, Emily Brontë's *Wuthering Heights*, Anne Brontë's *The Tenant of Wildfell Hall*, George Eliot's *Middlemarch*, Louisa May Alcott's *Little Women*, Elizabeth Gaskell's *North and South*, and George Eliot's *The Mill on the Floss*. These eight novels were chosen for their popularity, common usage within college classes, contribution to the discipline of literature, and the prominence of their female authors. This research explains how the novels were prepared for topic modeling and produces a specific topic model visualization, an "interactive heatmap," to compare the frequency of topics found in each novel.



## 2 METHODOLOGY

### INSTALLATION OF PACKAGES & PREPARING THE CORPUS INTO A DOCUMENT TERM MATRIX

The first step, distant reading, involved the preparation of a corpus of documents into a document term matrix. A document term matrix represents the words inside a text as a table (or a matrix) of numbers. The document term matrix format was used to implement topic modeling. In doing so, the novels could be read as data instead of narratives.

The text file used included the following novels in one document: Jane Austen's *Emma*, Charlotte Brontë's *Jane Eyre*, Emily Brontë's *Wuthering Heights*, Anne Brontë's *The Tenant of Wildfell Hall*, George Eliot's *Middlemarch*, Louisa May Alcott's *Little Women*, Elizabeth Gaskell's *North and South*, and George Eliot's *The Mill on the Floss*. All of these texts had similar base content material; they all had a female character as their main protagonist. They are all also well-studied novels throughout the world, as many consider these texts "classics." These texts were accessed through the digital archive Project Gutenberg.

First, the document was reviewed and all the extraneous metadata was removed. The extraneous metadata included introduction pages, acknowledgments, publishing details, title pages, and the table of contents. The only text within the document was from the actual story of each novel. Next, each chapter in the document was labeled in a single continuous document. For example, *Emma*, which has 55 chapters in total, was numbered from 1-55. The following novel was *Jane Eyre*, which started at 56 to continue the document. This was done in order to make the coding process and analysis by R Studio easier. By combining all of the novels into one text document, there would only need to be one file scanned and coded into the program rather than eight individual ones.

Each chapter was arranged into a contingency table that organized the most frequent words found to the least. Then, a relative frequency table was generated by dividing the total count of every word (token) in the chapter text by the total word count in a chapter. A function, *stopword()*, which was coded by a professor in the English Department at

Rutgers University - New Brunswick, Professor Sean Silver, was used to remove words such as "the," "a," and "and," as well as character names. This was done because otherwise, the most frequently appearing words would be these words, which do not represent any theme that can be applied across all eight texts. The text was then transformed into a document term matrix, ready to be prepared and used through topic modeling.

### TOPIC MODELING: A DESCRIPTION

Topic modeling refers to a coding technique that is able to process and analyze large-scale amounts of data (corpus) into key themes or topics. When analyzing novels through a technological lens, there are often two problems: the ambiguity of definition of certain words (i.e. a word like "fan" holds two meanings: a cooling apparatus and an enthusiast for a particular subject), as well as processing large data sets. The former is understood situationally; the human mind is able to identify, through context and critical thinking, the proper definition of such words. However, the computer often struggles with - to call on the aforementioned example - separating the two types of "fans." In this case, a method known as "key word in context" is used. This method allows the programmer to see the word within the context of the text, so as to better discern what the proper definition for the word is. Conversely, unlike the computer, it is very difficult for humans to read and interpret incredibly large amounts of data.<sup>[3]</sup> As a result, it was imperative that the technique known as topic modeling was used to address and offer solutions for both of these concerns.

Within topic modeling, the identified large-scale document is "read" by the computer and separated into key themes or topics. These topics are determined by the likelihood that certain words will appear together within the novel; each topic consists of words that have similar meanings or often emerge in context of each other. This is known as "collocation" and is essentially equivalent to concepts or themes. After the identification of major themes, the estimated proportions of the topics within the documents are identified. In this research, a visualization was then generated: an interactive heatmap. The in-

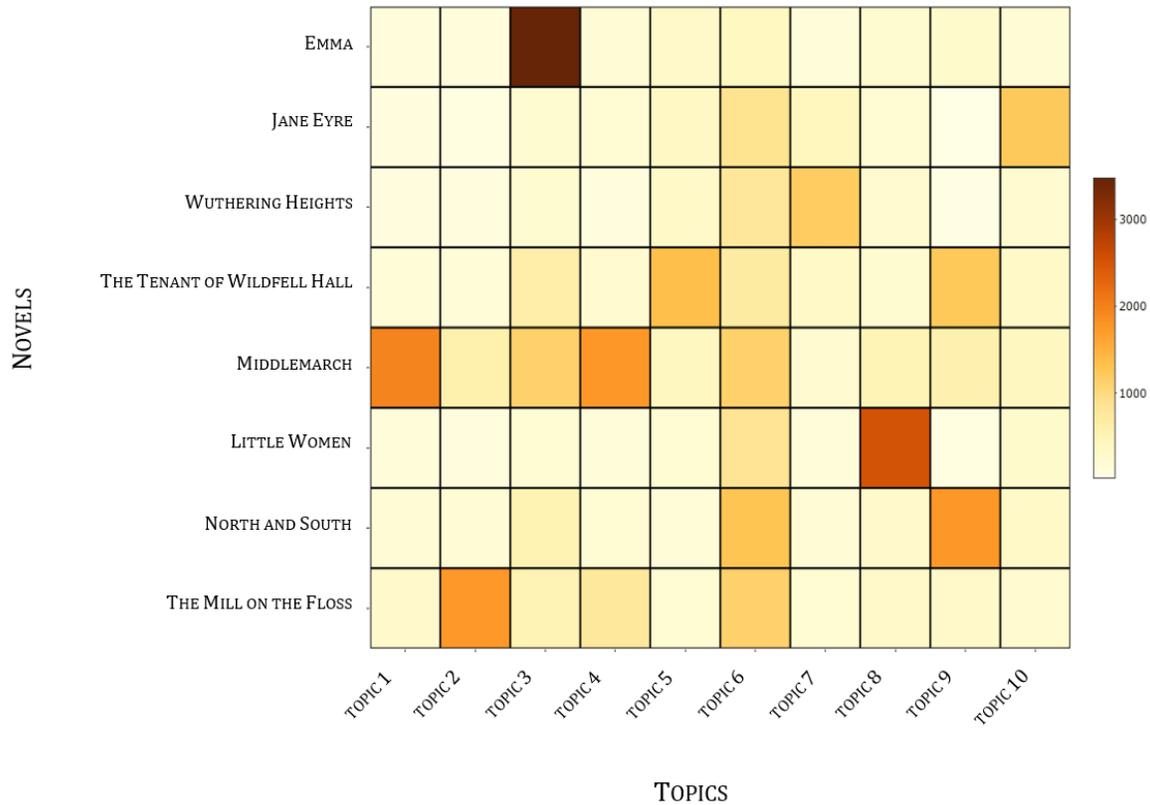


FIGURE A: INTERACTIVE HEATMAP OF TOPICS MOST FREQUENTLY FOUND IN 19<sup>TH</sup>-CENTURY NOVELS BY FEMALE AUTHORS

Generated interactive heatmap of topics most frequently found in 19<sup>th</sup>-century novels.

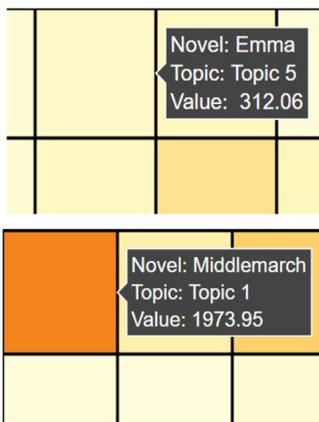


FIGURE B: Examples of the interactive square. Each interactive square displayed the name of the novel, the topic, and a value that represented the sum of all of the probability values generated in the topic modeling sequence.

Interactive heatmap is a graphical representation of the data where each of the individual values that are contained within the matrix is represented as different colored squares. When hovering over each colored square, the row label, column label, and calculated value of the data appears. In this way, topic modeling hovering allows for the organization and summarization of “electronic archives at a scale that would be impossible by human annotation.”<sup>[3]</sup>

The number of topics/themes one can generate (“k”) is arbitrary and up to the discretion of the programmer. If the “k” value chosen was too small, the topics generated would be too broad to classify as “main themes.” However, if the “k” value chosen was too large, there would be too many overlapping topics.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Term 1	sir	<e2>	<e2>	life	hand	came	master	little	<e2>	sir
Term 2	men	father	little	felt	love	eyes	door	mother	i<e2>	room
Term 3	money	things	quite	mind	leave	away	answered	old	don<e2>	little
Term 4	little	money	time	husband	heart	face	exclaimed	dear	time	looked
Term 5	old	sister	sure	feeling	little	look	house	girls	mother	eye
Term 6	things	uncle	dear	looking	continued	hand	entered	away	yo<e2>	time
Term 7	work	little	really	self	wish	heart	kitchen	time	it<e2>	heard
Term 8	question	bit	day	marriage	company	room	took	things	you<e2>	day
Term 9	going	em	father	sort	course	love	continued	people	father	house
Term 10	fellow	looking	friend	sense	husband	voice	returned	happy	speak	saw
Term 11	having	mother	wish	quite	return	felt	horse	looked	room	night
Term 12	mean	mind	having	possible	returned	knew	ds	came	day	hall
Term 13	raffles	poor	ds	strong	wife	night	person	day	o<e2>	dark
Term 14	time	pullet	body	consciousn	time	saw	heights	going	mamma	light
Term 15	town	family	certainly	world	sister	little	grange	look	men	round
Term 16	people	head	happy	position	friend	day	cousin	face	th<e2>	ladies
Term 17	family	old	woman	nature	thoughts	mind	mistress	father	kind	black
Term 18	fine	fathers	poor	people	fear	turned	chamber	work	away	st
Term 19	opinion	business	pleasure	dorotheas	affection	stood	head	took	face	large
Term 20	aunt	half	certain	answered	looked	bid	help	poor	school	school
Term 21	reason	want	believe	future	pleasure	door	blood	head	believe	quite
Term 22	medical	children	subject	experience	conduct	gone	having	won	life	high
Term 23	featherston	going	came	living	matter	feel	papa	pretty	he<e2>	white
Term 24	sense	tone	going	light	child	bed	ye	didn	papa	old
Term 25	horse	wi	letter	soul	heaven	half	window	boy	sure	door

FIGURE C: The 10 generated topics in TRIAL 5.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Term 1	quite	life	father	master	<e2>	sir	little	love	sir	came
Term 2	little	felt	money	answered	i<e2>	men	old	hand	little	face
Term 3	time	mind	sister	house	don<e2>	money	mother	leave	room	eyes
Term 4	sure	feeling	things	exclaimed	mother	time	away	little	looked	heart
Term 5	dear	looking	mind	door	little	old	dear	company	light	away
Term 6	friend	husband	bit	kitchen	time	little	girls	return	hall	hand
Term 7	really	sense	family	night	yo<e2>	going	time	husband	eye	time
Term 8	having	self	little	took	father	fellow	things	heart	saw	look
Term 9	day	marriage	em	having	it<e2>	having	came	fear	black	door
Term 10	ds	quite	uncle	returned	room	people	looking	heaven	dark	room
Term 11	pleasure	away	poor	turn	you<e2>	sort	look	wife	day	night
Term 12	body	nature	looking	continued	day	want	help	time	ladies	day
Term 13	half	things	head	papa	poor	raffles	happy	child	round	looked
Term 14	morning	dear	pullet	mistress	speak	question	day	returned	white	voice
Term 15	father	consciousn	old	entered	men	town	face	wish	door	heard
Term 16	certainly	end	children	stairs	sure	opinion	going	feelings	night	head
Term 17	woman	work	business	road	away	things	work	affection	school	felt
Term 18	wish	possible	fathers	heights	look	reason	looked	answered	passed	love
Term 19	poor	position	school	chamber	believe	fact	boy	continued	st	house
Term 20	subject	feel	wi	grange	wish	making	eyes	smile	heard	knew
Term 21	obliged	kind	small	bid	hear	work	people	half	eyes	life
Term 22	heard	look	quite	cousin	life	family	won	meet	house	turned
Term 23	coming	world	large	room	aunt	<eb>	pretty	determined	table	gone
Term 24	happy	hands	look	window	looking	pounds	took	length	stood	stood
Term 25	going	dorotheas	aunt	chair	o<e2>	stone	didn	friends	drawing	saw

FIGURE D: The 10 generated topics in TRIAL 6 for comparison.

### TOPIC MODELING: IMPLEMENTATION

Latent Dirichlet allocation (LDA), the simplest topic model, was employed within this research, as the “intuition behind LDA is that documents exhibit multiple topics [themes].”<sup>[3]</sup> In LDA, “all the documents [in this case, the novels] in the collection share the same set of topics, but each document exhibits those topics with different proportion.”<sup>[11]</sup> I chose to have LDA generate ten topics/themes ( $k = 10$ ) and to return the top 25 terms from each topic and the estimated proportions of topics in the documents. Because the corpus of documents was divided into chapters and not novels, I had to create a separate code to differentiate certain chunks of chapters as certain novels. For example, I coded for chapters 1-55 to be for *Emma*, chapters 56-93 to be *Jane Eyre*, and so forth. Using this data, I then plotted the generated values that show the estimated proportions of topics in each novel.

There were ten total trials of the interactive heatmap and topic modeling performed.

### 3 RESULTS

After performing ten trials of the interactive heat map and topic modeling, although there was a slight variation in the order of words and changes in a few number of terms each trial, results displayed that the generated topics and terms were generally the same. **TRIAL 5** was randomly chosen. Any trial would have sufficed because there was stability between the various trials and the generated themes. In general, there were not many topics that were found prominently amongst all of the novels - only **TOPIC 6** was found averagely within all of the novels. Instead, there were concentrations of certain topics in certain novels. This can be seen in **TOPIC 3**'s prevalence in *Emma* and **TOPIC 1** and **TOPIC 4**'s prevalence in *Middlemarch*, and so forth. Experimentation on this topic has led to the conclusion that the most popular word in a topic changes about forty percent of the time between essentially identical topic models.

The most significant and comprehensive themes amongst the novels included **TOPIC 4** and **TOPIC 6**. **TOPIC 4** touched upon much of women's worth within the social constructions of the 19<sup>th</sup>-century (“husband,” “life,” “marriage,” “self”), while **TOPIC 6** was

focused more on the mind and body (“eyes,” “face,” “look,” “heart,” “hand,” “felt,” “voice”). The rest of the topics - although they were found in some capacity amongst all the novels, were generally only concentrated within one or two novels and did not show a real, significant, or comprehensive theme. Instead, it can be understood that the remaining themes mainly consisted of a general melting pot of the most commonly appearing words within each respective novel (perhaps due to the plot, characterization of certain characters, and/or an individual writers' writing style as opposed to chosen themes).

### 4 DISCUSSION

Although it is difficult to come to concrete conclusions about shared themes/topics because the corpus was small in comparison to analyzing thousands of novels written by female authors in this century, I still found some support through the generation of topics based on word frequency for our understanding of 19<sup>th</sup>-century literature by female authors as an endeavor of social critique and individual consciousness of women during the 19<sup>th</sup>-century.

Using the technique of topic modeling (interactive heatmap), I assessed eight texts by seven unique Anglophone female writers. Multiple trials were performed in order to distinguish and observe which themes appeared in highest frequency, as well as to assess whether there was a notable difference in the generated themes for each respective trial. Although there was slight variation in the order of words on the spreadsheet and a few number of terms changed each trial, the results repeatedly showed the keywords to be virtually the same (**FIGURE C** compared to **FIGURE D**). For example, if one trial had the word “family” within a topic about agreeableness and neighborly interaction (**TOPIC 1** in **FIGURE C**), another trial may not have the specific word of “family,” but still had keywords such as “quite,” “father,” “woman,” and “little.” This is due to the nature of topic modeling. Within topic modeling, a label is automatically assigned to each topic based on its top term. Then, the code changes the topic model in various ways to see if the decided label turns up again.<sup>[16]</sup>

19<sup>th</sup>-century literature is often characterized

by its understanding of the psychological state within stressful and restrictive social environments or when confronted by moral dilemmas.<sup>[13]</sup> 19<sup>th</sup>-century literature dealt with ideas of moral decline and development and addressed the different modes and extent of control within certain psychological, social, cultural, and moral frameworks. It “delve[d][into] the internal life of the characters... through a focus on the thoughts and motivations of the characters rather than their occupations and external settings alone.”<sup>[13]</sup>

**TOPIC 6** is indicative of this; it included words such as “felt,” “mind,” “love,” “eyes,” “face,” “stood,” and “looked.” As seen from the interactive heatmap, this topic is found averagely in all of the texts. **TOPIC 6**’s words have an emphasis on mind and body. Words such as “felt,” “mind,” and “love” represent the psychological state and emotion, while words such as “eyes,” “face,” “stood,” and “looked” are grounded to the physical body; this connection between mind and body can be interpreted as representations of self-control. Within Sally Shuttleworth’s study, *Charlotte Bronte and Victorian Psychology*, she meditates on this idea of self-control, writing: “rigorous control and regulation of the machinery of mind and body would offer a passport to autonomous selfhood and economic liberty.”<sup>[14]</sup> However, she also importantly notes that there are “two conflicting models of psychology found in Victorian economic discourse: the individual is figured both as an autonomous unit, gifted with powers of self-control, and also as a powerless material organism, caught within the operations of a wider field of force.”<sup>[14]</sup> In considering this, **TOPIC 6**, - rich with words such as “felt” and “look,” - strengthens the general understanding and characterization of 19<sup>th</sup>-century literature and its preoccupation with the consciousness and social standing in society.

Within these novels, “males hold positions of political, institutional, and sometimes of economic power denied to females,” which can be seen in **TOPIC 4**, with words such as “husband,” “feeling,” “self,” “marriage,” “world,” “living,” and “experience.”<sup>[14]</sup> This topic touches on the aspect of 19<sup>th</sup>-century society that defined much of a woman’s legal and social worth. Although this connection is in line

with our understanding of Victorian society, it was significant that **TOPIC 4** on women’s worth based on social construct was not as prominent as **TOPIC 6** on the mind and body in all of the texts. This showed the authors’ intent to put an emphasis on the psychology and inner world of these female characters, rather than their marriage, in defining them and giving them a voice as individuals. The prominence of **TOPIC 6** demonstrated that “females hold a kind of psychological and moral power that is exemplified in their status as paradigmatic protagonists.”<sup>[12]</sup>

In addition, within a quantitative study conducted by John A. Johnson, Joseph Carroll, Jonathan Gottschall, and Daniel Krueger, the data concluded that within 19<sup>th</sup>-century Victorian literature, female protagonists scored the “highest of Agreeableness,” among other characteristics. Agreeableness was defined as a “pleasant, friendly disposition and tendency to cooperate and compromise, versus a tendency to be self-centered and inconsiderate.”<sup>[12]</sup> This tendency toward altruism is “more strongly associated with maturity, [and]... as vehicles for improving society.”<sup>[12]</sup> Although this could be understood as a great emphasis on the improvement of society as a whole, it also demonstrates the female characters’ need to be overwhelmingly “pleasant” within society. It was unsurprising that **TOPIC 3** of my research included words that evoke some sort of community or degree of agreeableness amongst neighbors, such as “quite,” “little,” “time,” “sure,” and “dear.” As I was interested in understanding the proper context of these words, I performed a keyword in context. Words like “quite” were used to express full and absolute agreeableness on a subject; examples include: “to feel quite sure,” “something quite fresh,” and “you do quite right.” Similarly, the keyword “sure” evoked ideas of obliging mannerisms; examples include: “To be sure,” “I am sure of having their opinions with me,” and “I am sure you are a great deal too kind.”

The heavy concentration of this “agreeableness” theme within the first novel, *Emma*, is quite appropriate of the text, as it is a novel about a young female and her social life, experiences, and relationships within the small town of Highbury. Although the novel is an exploration of Emma’s relationship

with her mind and body (ideas of maturity and development), she comes to develop her understanding of self-control and her emotional state through her interactions with others.

As mentioned before, I still found certain words and, thus, topics that were, understandably, more concentrated within certain novels. As just discussed, **TOPIC 3** was very concentrated within *Emma*. Although this was the most prominent case, in general, there seemed to be topics that were strongly associated with one particular novel over the others.

## 5 CONCLUSION

This research aimed to add to the current discussion surrounding 19<sup>th</sup>-century literature by female authors by quantifying and analyzing various thematic elements within eight select novels. These themes included social expectations for women, commentary on class and expressions of sexuality, and a focus on the protagonist's psychological state or moral development. This research is significant because it allowed a new method of analysis of 19<sup>th</sup>-century literature by female authors. The method of topic modeling utilized within the R Studio coding environment is effective in generating themes based on word-frequency data, but it must be noted that there are limitations to this method of analysis as it completely rejects the traditional format of reading novels. Within novels, there is a plot and specific storyline. The novel is organized and divided in a way that best supports the development of this plotline. However, topic modeling does not discern nor care for the sequence of the story and is instead only interested in quantifying the text - or "data" - and extracting and putting certain groups of words into "themes" or "topics" so as to quantitatively display the author and novel's thematic intent.

This research would have benefited from a larger corpus of works in order to come to more conclusive, specific, and generally applicable themes to be applied to novels during this time period. In addition, although words such as "you," "I," and "it" were included within the *stopword()* function, they still appeared in the results. They were joined by a strange symbol: `<e2>`, which I was regrettably unable to remove. This was most likely due to oversight

within the coding script. Ideally, the analyses would be redone with a fully cleaned corpus because the appearance of the `<e2>` symbol limits the ability to draw conclusions and undermines the interpretability of the topic wordlists.

Although the results of this analysis on the famous novels of 19<sup>th</sup>-century female writers were telling, they were not absolute in their conclusion and lacked a large enough corpus to properly determine the main themes amongst a large number of Victorian novels. If repeated, this research should take on a larger number of novels. There was some evidence of the societal restrictions imposed on women during this time period (**TOPIC 4**), yet conversely, there was also some evidence of the authors' intent to dispel these societal, political, and legal factors in defining the female individual by defining them through their mind, consciousness, and body (**TOPIC 6**). Evidence showed the importance of psychology, the body, and the inner worlds of their female protagonists in relaying the stories of women and giving them the opportunity to voice their thoughts and feelings through literature ■

## 6 ACKNOWLEDGEMENTS

Thank you to Professor Sean Silver of the English department at Rutgers University–New Brunswick for his assistance and encouragement of this research.

## 7 REFERENCES

- [1] Alcott, Louisa May, *Little Women*. (Melbourne; London; Baltimore: Penguin Books, 1953)  
[HTTPS://WWW.GUTENBERG.ORG/EBOOKS/514](https://www.gutenberg.org/ebooks/514)
- [2] Austen, Jane, *Emma*. (London; New York: Penguin Books, 2003)  
[HTTPS://WWW.GUTENBERG.ORG/EBOOKS/158](https://www.gutenberg.org/ebooks/158)
- [3] Blei, David. "Introduction to Probabilistic Topic Models." (2011)  
[HTTPS://WWW.EECIS.UDEL.EDU/~SHATKAY/COURSE/PAPERS/UIINTROTO-TOPICMODELSBLEI2011-5.PDF](https://www.eecis.udel.edu/~shatkay/course/papers/UIINTROTO-TOPICMODELSBLEI2011-5.PDF)
- [4] Brontë, Anne, *The Tenant of Wildfell Hall*. (Oxford: Blackwell, 1931)  
[HTTPS://WWW.GUTENBERG.ORG/EBOOKS/969](https://www.gutenberg.org/ebooks/969)
- [5] Brontë, Charlotte, *Jane Eyre*. (New York: Penguin Books, 1985)  
[HTTPS://WWW.GUTENBERG.ORG/EBOOKS/1260](https://www.gutenberg.org/ebooks/1260)
- [6] Brontë, Emily, *Wuthering Heights*. (London; New York: Penguin Books, 2003)  
[HTTPS://WWW.GUTENBERG.ORG/EBOOKS/768](https://www.gutenberg.org/ebooks/768)
- [7] Burrma, Rachel. "The Fictionality of Topic Modeling: Machine Reading Anthony Trollope's Bassetshire Series." *Big Data and Society* 2, no 2 (2015)  
[HTTPS://WORKS.SWARTHMORE.EDU/FAC-ENGLISH-LIT/286](https://works.swarthmore.edu/fac-english-lit/286)
- [8] Eliot, George, *Middlemarch*. (London: Vermont: J. M. Dent; Charles E. Tuttle, 1997)  
[HTTPS://WWW.GUTENBERG.ORG/EBOOKS/145](https://www.gutenberg.org/ebooks/145)
- [9] Eliot, George, *The Mill on the Floss*. (Edinburgh; London: W. Blackwood and Sons, 1860)  
[HTTPS://WWW.GUTENBERG.ORG/EBOOKS/6688](https://www.gutenberg.org/ebooks/6688)
- [10] Gaskell, Elizabeth Cleghorn, *North and South*. (Harmondsworth, Penguin, 1970)  
[HTTPS://WWW.GUTENBERG.ORG/EBOOKS/4276](https://www.gutenberg.org/ebooks/4276)
- [11] Jockers, M. L., and Mimno, D. "Significant themes in 19th-century literature." *Poetics* 41 (2013): 750-769.
- [12] Johnson, J. A., et al. "Portrayal of personality in Victorian novels reflects modern research findings but amplifies the significance of agreeableness." *Journal of Research in Personality* (2010)  
[HTTP://PERSONAL.PSU.EDU/FACULTY/J/5/J5J/PAPERS/VICTORIANPERSONALITY.PDF](http://personal.psu.edu/faculty/j/5/J5J/PAPERS/VICTORIANPERSONALITY.PDF)
- [13] Sen, Debashish. *Psychological Realism in 19th Century Fiction: Studies in Turgenev, Tolstoy, Eliot and Brontë*. (United Kingdom: Cambridge Scholars Publishing, 2020).
- [14] Shuttleworth, Sally. *Charlotte Brontë and Victorian Psychology*. (New York: Cambridge University Press, 1996).
- [15] Underwood, Ted. "A Genealogy of Distant Reading." *Digital Humanities Quarterly*, 11, no. 2, (2017)  
[HTTP://WWW.DIGITALHUMANITIES.ORG/DHQ/VOL/11/2/000317/000317.HTML](http://www.digitalhumanities.org/dhq/vol/11/2/000317/000317.html)
- [16] Yang, Y., et al. "The Stability and Usability of Statistical Topic Models." *ACM Transactions on Interactive Intelligent Systems*, 6, no.2, article 14 (2016)  
[HTTPS://DL.ACM.ORG/DOI/ABS/10.1145/2954002?DOWNLOAD=TRUE](https://dl.acm.org/doi/abs/10.1145/2954002?download=true)



Lauren Rha is a rising first-year graduate student at the Rutgers Graduate School of Education. She hopes to become an ESL teacher for immigrants and refugees coming into the United States. Her interest in and study of 19th-century literature motivated her to research common themes across eight famous Victorian novels. The research presented was done as a final project within Professor Sean Silver's English 315 class during her second semester as a junior, where she learned the fundamentals of basic coding for literary analysis within the R Studio environment.