

# Queer Encodings: An NLP Study of Challenged Books in the United States

Elliot Galvis  
Master's Thesis, Summer 2024

University of Chicago  
Digital Studies of Language, Culture, and History

## **Abstract**<sup>1</sup>

Attempts to ban books through complaints of “Sexually Explicit” content, “Body Description or Function,” or just “Homosexuality” have escalated dramatically in the United States in recent years. In response, many researchers and library organizations tag texts as LGBTQIA+ to illustrate the relationship between anti-trans panic and backlash to queer visibility with these book bans. But what makes a text queer? And how do both book banners and defenders constitute new definitions of queerness through their respective projects? Deploying a comparative approach of close reading and two “distant” reading methods represented by challenger complaints and computational topic modeling, this paper explores constructions of queer(ness in) literature in the hopes of critically assessing how classification around book bans protects or endangers queer people and cultures. This thesis draws from computational frameworks – and critiques – from queer and feminist literary scholars, as well as critical catalogers. Computational methods include analysis of metadata from the American Library Association and a BERTopic model trained on a subset of titles challenged from 2010-2023. Ultimately, I demonstrate some of the strengths and weaknesses in deploying topic models to analyze book bans and show that book challenges disproportionately target sex, gender, and queer topics as a source of harm in the texts even compared to more prevalent topics like violence.

---

<sup>1</sup> Content Warning: Some of the topics and tokens generated by large language models in this project reprint racial slurs found in the corpus they were trained on.

<b>Title and Abstract</b> .....	<b>1</b>
<b>Acknowledgements</b> .....	<b>3</b>
<b>I. Introduction</b> .....	<b>4</b>
<b>II. Background</b> .....	<b>7</b>
<b>III. Data Summarization</b> .....	<b>23</b>
<b>IV. Overview of computational methodology</b> .....	<b>28</b>
<b>V. Findings</b> .....	<b>33</b>
<b>VI. Limitations and Representational Shortcomings</b> .....	<b>47</b>
<b>VII. Conclusions</b> .....	<b>50</b>
<b>Appendix</b> .....	<b>53</b>
Figure 3-1a.....	53
Figure 3-1b.....	54
Figure 3-2.....	55
Figure 3-3.....	55
Figure 3-4a (All titles).....	56
Figure 3-4b (Corpus only).....	56
Figure 3-5a (All titles).....	56
Figure 3-5b (Corpus only).....	56
Figure 3-6a (All titles).....	57
Figure 3-6b (Corpus only).....	57
Figure 5-1.....	57
Figure 5-2.....	58
Figure 5-3.....	59
Figure 5-4.....	60
Figure 5-5.....	60
Figure 5-6a (Challenge Tags).....	61
Figure 5-6b (Model Topics).....	61
Figure 5-7.....	62
<b>Bibliography</b> .....	<b>63</b>

## Acknowledgements

I am grateful to several overlapping communities for making this project possible. First and foremost, thank you to my advisor Professor Hoyt Long for his support and critical feedback guiding this project and me since I first arrived at the University of Chicago. His patience and generosity in encouraging this thesis at every stage of its development alongside his welcoming introduction to the world of Cultural Analytics are the foundations on which this thesis rests.

My sincere thanks to the University of Chicago's Center for the Study of Gender and Sexuality for providing research and travel funding in support of this project. Thank you to Eric Strohane and the entire team at the American Library Association's Office for Intellectual Freedom for their work supporting libraries across the country and making data on banned books accessible and actionable. Sincere thanks to Professor Emily Knox at the University of Illinois Urbana-Champaign for sharing her insights on contemporary book banning as I conducted research for this project. Thank you to Taylor Faires and Filipa Calado for their advice and knowledge in considering computational possibilities for text analysis. Thank you to the participants of the 2024 Gender and Sexuality in Information Studies Colloquium and special thanks to Sara Howard and the other organizers for reassembling the colloquium. The community, ideas, and feedback of the scholars and library workers there reshaped this project in meaningful ways and sharpened my thinking on critical cataloging.

Thank you to the entire Forum for Digital Culture for creating such a robust program in which we could explore our interests through computational methods. Thank you to Professor Miller Prosser for introducing me to new ways of thinking about data and the importance of how metadata makes meaning, and to Professor David Schloen for leading the Digital Studies program. I am sincerely grateful to Professor Brooke Luetgert and Professor Clovis Gladstone: their generosity and teaching made data analysis and programming accessible in ways I could not have expected a year ago. Thank you to Professor Jeffrey Tharsen for his energizing teaching and ever-constructive feedback on machine learning, text analysis, and the natural language processing tools without which this project would not have been possible. His support and enthusiasm for this project's development have been driving forces in getting this work done. And thank you to my classmates and colleagues in Digital Studies and the University of Chicago Library. I learned so much from them and will always be grateful.

Finally, thank you to my family, partner, and friends for the ways in which they have taken care of me over the course of this thesis and beyond. Their curiosity, patience for my rants, and abiding love and care are the cornerstones of my life, and therefore, this work. I am thankful for them every day.

## I. Introduction

The United States has a book banning problem. From 2010-2023, 10,877 book titles were challenged in schools and libraries across the country. Organized campaigns by reactionary “parents’ rights groups” like Mom’s for Liberty caused the number of books challenged in 2021 to rise more than 800% from the previous year (1,858 compared to 223 in 2020) and shattered the record for individual titles challenged.<sup>2</sup> That record has continued to break every year since as attempts to ban books escalate nationwide.<sup>3</sup> Large language and federated models like ChatGPT have entered the fray as well, granting challengers a further layer of distance from the texts they target and desperate librarians a quick means of reviewing titles out of fear of legal punishment.<sup>4</sup> As a result of this massive escalation in challenges, there is unfortunately more data about attempts to censor texts in the United States than ever before.

Much of the research done on book bans in recent years has focused on the political, social, and cultural perspectives and goals of both challengers and targeted groups, especially queer and BIPOC authors, at the level of summary reporting on the trends in challenges. However, little research has been conducted to study corpora of challenged texts between the extremes of closely reading individual titles and extrapolating summary statistics of challenges, and it is between these poles that my project intervenes. Drawing from cultural analytics, critical cataloging practices, and queer theory, I employ several tools for Natural Language Processing (NLP) and Natural Language Understanding (NLU) to compare two kinds of “distant readings”<sup>5</sup>

---

<sup>2</sup> Ella Feldman, “Book Banning Attempts Are at Record Highs,” *Smithsonian Magazine*, March 18, 2024: <https://www.smithsonianmag.com/smart-news/book-banning-attempts-are-at-record-highs-180983964/>.

<sup>3</sup> American Library Association, Office for Intellectual Freedom, “Censorship by the Numbers,” March 4, 2024: <https://www.ala.org/bbooks/censorship-numbers>.

<sup>4</sup> Angela Watercutter, “How an Iowa School District Used ChatGPT to Ban Books,” *Wired*, August 18, 2023: <https://www.wired.com/story/chatgpt-ban-books-iowa-schools-sf-496/>.

<sup>5</sup> I borrow this term from Franco Moretti’s definition of “distant reading” for computational text analysis at scale in contrast to traditional humanistic “close reading.” Often accused of representing overly quantitative, anti-humanistic approaches to literary studies, distant reading nevertheless offers a helpful framework for describing categorical,

of challenged texts – one derived from the complaints of book challengers and the other from a topic model trained on challenged texts.

Several questions drive this project. What can a computational text analysis of individual books challenged in the United States reveal about the texts, authors, and topics being suppressed today? Given the fluid and contested nature of “queer” (as noun, adjective, and verb) posited by queer theory and across flexible and intersectional cultural contexts, how does the targeting of “queer” texts constitute new hostile understandings of and to queer people? And how do computational models that lack cognition of cultural and social phenomena represent these identities based on challenged books? In answering these questions, I aim to excavate and challenge computational derivations of what “queerness” means and what makes a “queer text” in the eyes of book challengers. Through this lens, I examine a subcorpus of books challenged between 2010-2023 to identify the deeper substance of texts being targeted in the larger context of contemporary anti-LGBTQIA+ backlash in the United States.

In order to operationalize this process, I have built a BERTopic model of 45 different topics from 200 books and compared its outputs with both the challenge labels assigned to titles by the American Library Association (ALA) and also my own close reading of the material. The original data for challenged books comes from the ALA’s Office for Intellectual Freedom, which has been recording book challenges since 1990, and I focus on 200 distinct titles from this dataset for developing my topic model and comparing challenge tags. Drawing out similarities between these distant readings of the texts, I identify how challenges collapse differences across these titles to target authors and books at the intersections of marginalized identities and

---

computational, and statistical analysis of text that can and should be combined with close reading methods. As Richard So advocates in his recuperation of Moretti, “statistical models are not just summaries of or reports about data, they are the mechanisms with which individuals reason and think”; Richard So, “All Models Are Wrong,” *PMLA/Publications of the Modern Language Association* 132, no. 3 (2017): 670.

disproportionately fixate on negative associations of gender, sexuality, and queerness as a source of harm.

Recently, several prominent coalitions of book challengers have faced public backlash that led some journalists to dismiss organized book banning as a fading front in the United States’ “culture wars.” The broad unpopularity of book bans,<sup>6</sup> recent failed political runs by familiar figures in the movement,<sup>7</sup> and the disbanding of several chapters of groups like Moms for Liberty all contribute to this perspective.<sup>8</sup> However, it is crucial to remember that challenges massively increased in 2023 – challenges against public libraries in particular nearly *doubled* compared to the previous record in 2022<sup>9</sup> – and that the escalation of austerity measures targeting public libraries has only increased alongside challenges.<sup>10</sup> It is likely that book censorship targeting marginalized communities will only get worse before it gets better. Given this urgency, I hope that this project will provide a tool to better understand books already being challenged to help stem the tide of book banning and to illustrate the ideas and queer knowledge systems that we lose when books are taken out of the hands of readers.

---

<sup>6</sup> American Library Association, “Voters Oppose Book Bans in Libraries,” March 21, 2022: <https://www.ala.org/advocacy/voters-oppose-book-bans-libraries>.

<sup>7</sup> Christopher Wiggins, “Moms for Liberty Candidates Rejected by Voters in Key School Board Elections,” *Advocate*, November 8, 2023: <https://www.advocate.com/politics/moms-for-liberty-rejected-election>.

<sup>8</sup> Tori Otten, “Moms for Liberty Completely Collapses in Former Strongholds,” *The New Republic*, February 8, 2024: <https://newrepublic.com/post/178879/moms-liberty-chapters-collapse-strongholds-pennsylvania-florida>.

<sup>9</sup> Raymond Garcia, “American Library Association reports record number of unique book titles challenged in 2023,” *American Library Association News*, March 14, 2024: <https://www.ala.org/news/2024/03/american-library-association-reports-record-number-unique-book-titles>.

<sup>10</sup> The New York Public Library was among the highest-profile cases of mass defunding; see Chris Sommerfeldt, “NYC libraries would end universal six-day service under Adams budget plan: testimony,” *New York Daily News*, March 12, 2024: <https://www.nydailynews.com/2024/03/12/most-nyc-libraries-would-have-to-end-weekend-hours-entirely-under-adams-budget-plan-testimony/>. New York City mayor Eric Adam reversed the budget cuts at the last minute before the 2025 fiscal year, but similar challenges have persisted nationwide; e.g., three libraries face complete closure in St. Charles County, Missouri.

## II. Background

### Book challenges and their hermeneutics

The formal banning of books in American schools and libraries has been in practice since the 19th century. Harriet Beecher Stowe's *Uncle Tom's Cabin*, Charles Darwin's *The Origin of Species*, and Walt Whitman's *Leaves of Grass* are staples of the country's history of contentious censorship and each respectively reflect the historical anxieties around Anti-Blackness and slavery, secularism as opposition to Christian fundamentalism, and expressions of gendered sexuality outside of hegemonic structures. Book challenges are not a new phenomenon, and as the American Library Association has documented since 1990, many of the aforementioned concerns presented above have endured in the 21st century.<sup>11</sup>

What is novel in book bans today is the scale and frequency of challenges. In the last three years, reactionary fears have ignited a firestorm of legislation and harassment by challengers to target public education and library services that PEN America has dubbed "Ed Scare."<sup>12</sup> A combination of historical and contemporary phenomena fuels these efforts: the steady dismantling of public services since the 1970s, the escalation of homophobic and transphobic bigotry, the coalescence of the libertarian and religious right with Donald Trump's election, the shutdown of in-person teaching during the onset of the Covid-19 pandemic, and the cementation of reactionary "parents' rights" networks in response.

The resulting escalation of book challenges across libraries and schools has been staggering. In 2000, the ALA recorded challenges to 378 individual titles, a count whose annual average decreased slightly over the next 20 years. In 2015, the number was only 190 books; in

---

<sup>11</sup> American Library Association, "Challenge Reporting," February 24, 2023: <https://www.ala.org/tools/challengesupport/report>.

<sup>12</sup> PEN America, "Ed Scare: Frequently Asked Questions": <https://pen.org/ed-scare-faq/>.

2020, still only 223.<sup>13</sup> However, that number has increased drastically each year since, as 1,858 individual titles were reported in 2021, 2,571 in 2022, and a staggering 4,240 in 2023. From an aggregate perspective, nearly 1,500 more titles were challenged in the three-year period of 2021-2023 than in the preceding eleven years.<sup>14</sup> Since then, a coalition of academics, the ALA, anti-censorship groups like PEN America, and even commercial publishers have taken up efforts to challenge and document book bans nationwide. Nonprofits like PEN America and the ALA have been tracking instances of book challenges through self-admittedly imperfect combinations of surveys, media analysis, and direct reporting mechanisms that unavoidably vastly underrepresent the actual numbers of books challenged.<sup>15</sup>

One of the preeminent scholars studying book challenges in the United States is Emily Knox, President of the National Coalition Against Censorship and Professor of Information Sciences at University of Illinois at Urbana-Champaign. In September 2023, Knox testified before Congress on the harmful nature of contemporary book challenges. She summarized her critique of these challenges as antithetical to intellectual and expressive freedoms, restrictive to the autonomy of readers from the minority of parents who challenge books, and catastrophic for the societal infrastructures that libraries provide in the United States. Moreover, the targeting of diverse books in particular threaten people of color and queer communities beyond the texts themselves when “attacks flatten the people’s humanity.”<sup>16</sup> Prior research by Knox has presented

---

<sup>13</sup> American Library Association, “Surge in Book Challenges Press Kit,” March 21, 2022:

<https://www.ala.org/news/mediapresscenter/presskits/surge-book-challenges-press-kit>.

<sup>14</sup> 6,222 from 2021-2023 compared to 4,654 from 2010-2020.

<sup>15</sup> The ALA estimates that they only capture “3-18% of book challenges” and PEN America notes that frequent forms of informal challenging and reporting, along with the opacity of “wholesale bans” in which entire libraries are shuttered or destocked, means their data is incomplete. See Kasey Meehan and Jonathan Freedman, “Banned in the USA: State Laws Supercharge Book Suppression in Schools,” PEN America, April 20, 2023:

<https://pen.org/report/banned-in-the-usa-state-laws-supercharge-book-suppression-in-schools/> and American Library Association, “Get Involved”: <http://www.ala.org/advocacy/bbooks/bannedbookweek/ideasandresources/activity>.

<sup>16</sup> Emily Knox, “Testimony before the U.S. Senate Committee on the Judiciary Hearing on ‘Book Bans: Examining How Censorship Limits Liberty and Literature,’” September 12, 2023: 1.

legal understandings of book challenges in opposition to First Amendment rights and the targeting of books about and by members of Black and queer communities.<sup>17</sup> Knox succinctly presents the benefit of representational material in libraries, writing that such books “allow all human beings to see themselves reflected in books, and...allow everyone to learn about people who are not like themselves,”<sup>18</sup> and conversely, that the erasure of these texts means that “the voices of members of diverse populations are removed from society.”<sup>19</sup>

Arguably the most relevant of Knox’s contributions for my project is the identification of challengers’ belief in a singular, literalist reading of texts that they often describe as “common sense.” First presented in Knox’s 2015 monograph *Book Banning in 21st-Century America*, this conceptualization of one indisputable interpretation of the written word aligns closely with fundamentalist Christian principles of literal interpretations of the Bible that are antithetical to the hermeneutical concept of “critical distance” widely accepted in modern reading principles, in which readers bring their own understandings to a text resulting in a plurality of meaning.<sup>20</sup> Nearly 10 years after the book’s publication, this principle of singular interpretation remains useful for understanding challenges today.

### **Anti-LGBTQIA+ activism**

As Knox illustrates, common sense reading plays a salient role in understanding (and informing) today’s anti-LGBTQIA+ readings of challenged texts, a role that becomes even clearer when viewed from the lens of queer scholarship. On the one hand, the erasure of a plurality of interpretive readings allows for the elevation of dominating – often white, Christian,

---

<sup>17</sup> Emily Knox, *Foundations of Intellectual Freedom*, (Chicago: ALA Neal-Schuman, 2023), 39. Emily Knox, “Silencing Stories: Challenges to Diverse Books,” *The International Journal of Information, Diversity, & Inclusion* 3, no. 2 (April 2019).

<sup>18</sup> *Ibid.*, 26.

<sup>19</sup> *Ibid.*, 34.

<sup>20</sup> Emily Knox, *Book Banning in 21st-Century America*, (Lanham: Rowman & Littlefield, 2015), 40-41.

and cisgender/heterosexual – perspectives to claim center stage as the only acceptable form of meaning-making. On the other, the logic of common sense challenges has gotten more complex. Rather than relying on explicit literalism as a metric of truth, current right-wing coalitions increasingly advance traditional homophobia from the religious right with tools of the conspiracy-minded anti-vaccination and QAnon movements. The result is that singular “truth” found in texts has shifted from being framed as universally true *and* universally legible, to being universally true but obscured by mainstream media and political perspectives. This has calcified into a theory that public institutions like schools and libraries, bolstered by national and local LGBTQIA+ advocacy groups and greater queer representation in media,<sup>21</sup> have become state agents for “indoctrinat[ing] our kids,” “sexualizing children,” and employing “pedophiles and groomers” in the form of teachers and librarians.<sup>22</sup> Challenges against queer books and library workers are not therefore conceptualized as attacks on freedom of speech or information, but rather about “protecting the innocence of our children” and combating “woke gender ideology,” broadly construed as anything ranging from representation of genderqueer and nonbinary people to any attention to the use of gendered pronouns.<sup>23</sup>

The pitting of queer people and their visibility against the safety and presumptive (cis)gender identity and (hetero)sexuality of children are familiar from the postwar history of

---

<sup>21</sup> “Gender Resource Guide”, Minnesota Family Council, 2019, 31: <https://genderresourceguide.com/wp-content/themes/genderresource/library/documents/NPRGFullDocumentPrintV17.pdf>.

<sup>22</sup> Elizabeth A. Harris and Alexandra Alter, “A Fast-Growing Network of Conservative Groups Is Fueling a Surge in Book Bans,” *The New York Times*, updated January 10, 2023: <https://www.nytimes.com/2022/12/12/books/book-bans-libraries.html>; Amanda Waldroupe, “Violent, aggressive, and personal attacks — language used to challenge books has changed dramatically, research presented at Oregon Library Association conference shows,” *Oregon Arts Watch*, April 30, 2024: <https://www.orartswatch.org/violent-aggressive-and-personal-attacks-language-used-to-challenge-books-has-changed-dramatically-research-presented-at-oregon-library-association-conference-shows/>.

<sup>23</sup> Harris and Alter, “A Fast-Growing Network of Conservative Groups Is Fueling a Surge in Book Bans.” The labeling and derision of “woke ideology” by conservatives today in many ways parrots the deployment and targeting of “political correctness” by conservatives in the 1980s-1990s as undermining American culture by a minority group of intellectual elites. Eve Sedgwick illustrates this precedent quite neatly: Eve Kosofsky Sedgwick, “Queer and Now,” *Tendencies*, (Durham: Duke University Press, 1993), 18-20.

LGBTQIA+ persecution in the United States. The demonization of gay people, and especially gay educators, as corrupting or sexually dangerous threats to children was a common tactic in the homophobic rhetoric at the end of the twentieth century.<sup>24</sup> Anita Bryant’s infamous “Save Our Children” coalition was based on the exact premise that legal and social acceptance of gay people were attempts to corrupt children since “homosexuals cannot reproduce – so they must recruit. And to freshen their ranks, they must recruit the youth of America.”<sup>25</sup> As predecessors to the Parents Rights advocates of book bans today, Bryant and her allies did not view their attacks on gay rights legislation as contradictory to civil liberties, but rather as necessary defenses for “the civil rights of parents: to save their children from homosexual influence.”<sup>26</sup> This rhetoric has not disappeared in the twenty-first century despite legal and social victories for gay, lesbian, and sexually minoritized communities.

Rather, it has been redeployed in the targeting of trans and gender-nonconforming communities and drag performers. Manhattan Institute strategist Christopher Rufo and Florida Governor Ron DeSantis have been among the most prominent of today’s Anita Bryant analogs, though perhaps with a higher degree of cynicism than Hill herself. Rufo brags about the susceptibility of media to right-wing targeting of anti-racist and queer-affirming efforts and wrote in 2022 that “when parents, voters, and political leaders understand the true nature of Drag Queen Story Hour and the ideology that drives it, they will work quickly to restore the limits that

---

<sup>24</sup> Melissa Block, “Accusations of ‘grooming’ are the latest political attack – with homophobic origins,” *NPR*, May 11, 2022: <https://www.npr.org/2022/05/11/1096623939/accusations-grooming-political-attack-homophobic-origins>. Note that the “Lavender Scare” of the 1950s, elevated by historian David K. Johnson, laid much of the groundwork for Bryant and her contemporaries’ work in the latter half of the century. However, the rhetoric of the Lavender Scare stigmatized gay men for their imagined risk to the American nuclear family and for presenting a point of cultural and moral “weakness” in the fight against Communism during the Cold War. Rather than focusing on the imagined dangers of queer people to children, this framework imagined queer citizens as un-American and thereby a risk to national security and identity.

<sup>25</sup> Anita Bryant, *The Anita Bryant Story*, (Old Tappan, NJ: 1977): 62.

<sup>26</sup> Gillian Frank, “‘The Civil Rights of Parents’: Race and Conservative Politics in Anita Bryant’s Campaign against Gay Rights in 1970s Florida,” *Journal of the History of Sexuality* 22, no. 1 (2013): 127; Advertisement, *Miami News*, March 21, 1977, 3A.

have been temporarily—and recklessly—abandoned. They will draw a bright line between adult sexuality and childhood innocence, and send the perversions of ‘genderfuck,’ ‘primitivism,’ and ‘degeneracy’ back to the margins, where they belong.”<sup>27</sup> This explicit call for the exclusion of queer people and knowledge-making from society illustrates core elements of the perspectives driving book challenges today, with a presumed risk that children are made vulnerable to the “evils” of LGBTQIA+ people and ideas through public education and libraries.

Unsurprisingly, the most prevalent terms in the complaints of book challenges studied by the ALA are “Sexually Explicit,” “Children,” and “Pornographic,” and “LGBTQIA” is listed by the ALA as the fourth most frequent reason given for a book’s challenge.<sup>28</sup> The association of queerness with danger to children and sexual immorality illustrates how book challenges can construct a distant reading of books, in which the labeling of a book as LGBTQIA+ implies a form of harm and becomes a proxy term for social immorality and danger. More than 30 years ago, Eve Sedgwick offered a haunting and prescient summary of these efforts: “Teachers are fired, routinely, for so much as intimating the right to existence of queer people, desires, activities, children...Seemingly, this society wants its children to know nothing; wants its queer children to conform or (and this is not a figure of speech) die; and wants not to know that it is getting what it wants.”<sup>29</sup> Rather than the illusory claim that book bans protect the “innocence” of children, the epistemicide carried out especially against queer texts and histories must be understood as part of the civil and hermeneutic injustices against children and LGBTQIA+ adults today.

---

<sup>27</sup> Trip Gabriel, “He Fuels the Right’s Cultural Fires (and Spreads Them to Florida),” *The New York Times* (April 24, 2022); Christopher Rufo, “The Real Story Behind Drag Queen Story Hour,” *City Journal* (Autumn 2022): <https://www.city-journal.org/article/the-real-story-behind-drag-queen-story-hour>.

<sup>28</sup> See Figures 3-2 and 3-6a.

<sup>29</sup> Sedgwick, “Queer and Now,” 2-3.

## Queer Theory, Critical Cataloging, and LGBTQIA+ Labels

Despite the hostility of the aforementioned perspectives, other frameworks for understanding gender and sexuality based identity remain abundant, with queer theory among the most polysemous of such perspectives. This project was inspired in part by the evident contradictions at play between the fluidity and temperamental definitions of identity presented by queer theorists and the rigidity of labeling that book challengers – and library practices themselves – impose when categorizing a text as “LGBTQIA.” When the atomized, digital nature of information in computation comes into play, this contradiction can heighten without proper contextualization and critical interpretation. After all, in machine learning’s translation of books into sentences, then words, then quantified tokens and back again, there are risks of abstracting meaning beyond useful interpretation. However, it is in the productive tension of these contradictions that I explore new understandings of the corpus texts I have selected. By combining computational analysis with categorical challenge labels and my close reading applying a queer lens, I hope to shed light on the nature of the books challenged better than any one interpretive strategy could alone. Before grappling with these contradictions, more framing on how queer theory encodes and decodes meaning is necessary.

The field of queer theory originally emerged at the end of the twentieth century as an expansion on and deviation from women’s studies or orientation-based disciplines like gay or lesbian studies. Eve Kosofsky Sedgwick’s *Epistemology of the Closet*, considered one of the founding texts of the field despite never using the term “queer,” grapples with some of the binarisms and limitations that she sees in preceding terms, particularly in the labeling of “homosexuality” and “heterosexuality” as categorically distinct or in the ambiguous slippage

between terms like gay and lesbian.<sup>30</sup> Sedgwick's intervention draws from the post-structuralist ideas of Michel Foucault, who famously argued in the *History of Sexuality* that the nineteenth century saw the transformation of understandings of "homosexual" behavior to become a fixed category or identity for people themselves, thereby fixing a taxonomic label to what had previously been more fluidly articulated: in his words, "the homosexual was now a species."<sup>31</sup>

By 1993, Sedgwick had begun using "queer" as a response to the question "What if the richest junctures weren't the ones where *everything means the same thing?*" to mean "the open mesh of possibilities, gaps, overlaps, dissonances and resonances, lapses and excesses of meaning when the constituent elements of anyone's gender, of anyone's sexuality aren't made (or *can't be* made) to signify monolithically... anyone's use of 'queer' about themselves means differently from their use of it about someone else."<sup>32</sup> From this offering, queer theory emerges as an expansive, tangled means of navigating gender and sexuality beyond binary frameworks related (even in opposition) to heterosexuality and cisgender identity. "Queer" in this light became a hermeneutical method as much as a personal identifier: one can queer a text in the same stroke as someone may identify as queer. Scholars took this offering and ran with it, to push expansive definitions of queerness even further. Many, like José Esteban Muñoz, began to push back against the twentieth century framing of gender and sexual minorities pitted in opposition to children (and the future they represent) by expanding queer thinking as a utopian enterprise belonging to a collective liberated future.<sup>33</sup>

---

<sup>30</sup> Eve Kosofsky Sedgwick, *Epistemology of the Closet*, (Berkeley: University of California Press, 1990), 158-159; 17-18.

<sup>31</sup> Michel Foucault, *The History of Sexuality Volume 1: An Introduction*, (New York: Random House, 1978), 43.

<sup>32</sup> Sedgwick, "Queer and Now," 6; 8-9.

<sup>33</sup> José Esteban Muñoz, *Cruising Utopia*, (New York: NYU Press, 2009).

A distinct but important concept for understanding definitions of “queer” within queer theory is that of intersectionality. Legal scholar Kimberle Crenshaw first coined the term to describe how antidiscrimination law’s orthodox approach to racism or sexism treats these categories as mutually exclusive, with the effect that people who are marginalized at the intersections of multiple axes of oppression are “theoretically erased,” which “creates a distorted analysis of racism and sexism because the operative conceptions of race and sex become grounded in experiences that actually represent only a subset of a much more complex phenomenon.”<sup>34</sup> Sociologists and Black feminists quickly expanded intersectionality to methods and subjects beyond the legal sphere to include social, political, and economic hierarchies alongside other categories like sexuality, wealth, disability, and nationality.<sup>35</sup>

Though historically and theoretically distinct, intersectionality and queer studies are important to consider together. Queer theory posits that there is no singular, stable version of queer that defines everyone who might claim it, and intersectionality helps illustrate the diversity of queerness as it is always constructed in relation to other social identities. Understanding queerness, and any LGBTQIA+ identities, therefore requires consideration of race, economic status, disability, and the other identities that people embody simultaneously. Queer theory allows us in turn to “de-naturalise taken for granted categories of analysis,” as Aristeia Fotopoulou writes, “even beyond issues of sexuality and gender,” indicating how the labels at play in intersectionality can themselves be queered to generate new forms of understanding.<sup>36</sup> Most pertinently for this project, considering intersectionality and queer theory in tandem allows

---

<sup>34</sup> Kimberle Crenshaw, “Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics,” *The University of Chicago Legal Forum*, (1989), 139-140.

<sup>35</sup> E.g., see Leslie McCall, “The Complexity of Intersectionality,” *Signs* 30, no. 3 (Spring 2005), 1771-1800 and Patricia Collins, “Intersectionality’s Definitional Dilemmas,” *Annual Review of Sociology* 41, (2015), 1-20.

<sup>36</sup> Aristeia Fotopoulou, “Intersectionality of Queer Studies and Hybridity: Methodological Frameworks for Social Research,” *Journal of International Women’s Studies* 13, no.2 (2012), 25.

for a more nuanced understanding of books and their discrete labeling, whether by book challengers, catalogers, or computational classification. This combination allows us to usefully expand our approach to categorical thinking while minimizing the abstraction and loss of nuance that comes with it.

One such expansion of particular import to this project and library spaces is the development of “critical cataloging” practices. One facet of critical cataloging is the practice by library workers to challenge existing subject headings or categorizing structures of books on the basis that they are dated, inaccurate, or reflect harmful or unnecessarily limited perspectives. As President of the American Library Association Emily Drabinski puts it, to this kind of critical cataloger, “[some] materials are cataloged incorrectly, and they can be cataloged correctly with the correct pressure.”<sup>37</sup> Examples of this approach deployed by gay and lesbian library workers include Barbara Gittings’ efforts to declassify homosexuality as a mental disorder by the American Psychiatric Association and Samuel Berman’s work petitioning the Library of Congress to avoid pathologizing homosexuality and queer sexual practices through legacy subject headings.<sup>38</sup> These campaigns represent a meaningful identification of how hegemonic powers can normatively assert semantic meaning through labeling books, and such efforts therefore undermine the purported objectivity of readings like the “common sense” interpretations by book challengers.

However, a queer approach to critical cataloging goes further and questions the limitations of substituting one label for another. Queer hermeneutics informed by Sedgwick’s work present an inherent contradiction between the principles of queer theory and library

---

<sup>37</sup> Emily Drabinski, “Queering the Catalog: Queer Theory and the Politics of Correction,” *The Library Quarterly* 83 (2), April 2023: 95.

<sup>38</sup> Melissa Adler, *Cruising the Library*, (New York: Fordham University Press, 2017), 43-45.

cataloging, even as libraries are often bastions of queer communities and knowledge formation.

For such catalogers, how can the idea of “queer” as polysemous, unstable, and flexibly interpretable, be encapsulated by a regulated, centrally maintained, and prescribed system like the Library of Congress Subject Headings that libraries rely on to make meaning legible?

Drabinski argues that “the entire project of library classification and cataloging is at odds with queer ideas about historicity, contingency, and the impossibility of a fixed system of linguistic signs.”<sup>39</sup> Taking up Drabinski’s challenges and Sedgwick’s theory, Melissa Adler traces how the deployment and circulation of controlled vocabularies like the Library of Congress Subject Headings expand the suppression of information even as they aspire to render texts legible.

Taken to its extreme, Adler writes that “the terms for norms and their binary oppositions are reiterated over and over, naturalizing and normalizing discourses, creating a cycle of prohibition that authorizes certain ways of speaking and disallowing unacceptable speech and subjects.”<sup>40</sup>

For such critics, the scrutiny of individual terms alone is insufficient. In order to grapple with the interpretability of texts in library spaces, users and librarians need to interrogate the very foundations and problems inherent in labeling schema in the first place.

These concerns are familiar to Cultural Analytics and Digital Humanities scholars working with texts.<sup>41</sup> Laura Mandell points to the reliance and perpetuation of gender stereotypes as an interpretive strategy for quantitative text analysis. She criticizes work by Matthew Jocker for shunning sociohistorical contingencies of gender in favor of “stereotyping,” and contrasts it with scholarship by Mark Olsen that grapples with the constructed and contingent nature of

---

<sup>39</sup> Drabinski, “Queering the Catalog,” 101.

<sup>40</sup> Adler, *Cruising the Library*, 100.

<sup>41</sup> For presentations and responses to some of the more pertinent of these critiques, see So, “All Models Are Wrong,” 668-669; Dong Nguyen et. al., “How we do things with words: Analyzing text as social and cultural data,” *Frontiers in Artificial Intelligence* 3 (2019): 8-9.

linguistic style and gender.<sup>42</sup> Without the latter's attention to polysemous interpretive strategies and the contingent construction of social categories, natural language processing tools can simply reproduce harmful hegemonic perspectives as disembodied readings. Without careful evaluation and interpretation, deploying computational methods can therefore replicate or simply substitute the monosemic, "common sense" readings of texts that book challengers deploy when identifying texts for removal.

Nevertheless, abstracted descriptions, labels, and interpretations of text *do* serve practical functions that are enormously important for readers to access texts and identify helpful concepts. As Angela Chen writes, "A shared vocabulary makes ideas more accessible while a lack of language can render an experience illegible."<sup>43</sup> One such shared vocabulary is The Homosaurus, an online linked data thesaurus designed to supplement general controlled vocabularies with crowd-sourced, peer-edited language specific to LGBTQIA+ terms and histories.<sup>44</sup> By presenting itself as eternally in-progress, The Homosaurus recognizes many of the challenges posed by queer theorists and critical catalogers while providing the scaffolding for mutual intelligibility. Drabinski concedes that "in order to be accessible to users, materials must be fixed in place and described using controlled vocabulary. However, this fixing is always fundamentally fictive."<sup>45</sup>

In navigating these fictions, scholars like Drabinski propose shifting more of the weight of navigating semantics from the domain of catalogers to the discursive, on-site work of public services librarians and teaching librarians.<sup>46</sup> Such an approach ultimately helps to both reveal the constructed nature of text labeling but also allows readers and library workers to navigate the

---

<sup>42</sup> Laura Mandell, "Gender and Cultural Analytics: Finding or Making Stereotypes," *Debates in the Digital Humanities 2019*, (Minneapolis: University of Minnesota Press, 2019), 6-8.

<sup>43</sup> Angela Chen, *Ace: What Asexuality Reveals About Desire, Society, and the Meaning of Sex*, (Boston: Beacon Press, 2020), 17.

<sup>44</sup> The Homosaurus, "About": <https://homosaurus.org/about>.

<sup>45</sup> Drabinski, "Queering the Catalog", 105.

<sup>46</sup> *Ibid.*, 107.

multimodal interpretive and organizing processes used by libraries in a more honest and nuanced way than strict reliance on labels and other distant reading.

This is not to say that cataloging and computational classification by extension do not provide a crucial role in grappling with the scale of collections and challenged book lists. Adler recognizes that libraries and public-facing scholarship are in many ways best equipped to responsibly employ these tools when compared to commercial entities like Chat-GPT or Google.<sup>47</sup> Indeed, many Cultural Analytics scholars have already been engaging with computational work with exactly such considerations in mind.

### **Computation and Topic Modeling**

Scholars in the humanities and social sciences who apply computational methods in their research are familiar with balancing the useful clarity of computation and quantitative methods with the critical thinking needed to interpret the outputs of these tools as representative and interpretable, but not fundamentally “true.” Emily Knox’s aforementioned framework of “literalism” in book challenges also becomes useful for comparison with the occasionally perceived “objectivity” of computational reading methods. Computational objectivity is a fallacy that can be dangerously misleading. For one, computers do not generate some “impure,” unbiased truth untouched by human interpretation: human actors at every stage of a large language model’s deployment (data collection, training, testing, implementation, and interpretation) bring their own hermeneutics to a model’s final representation of texts. Furthermore, while NLP and NLU tools like BERTopic are extremely useful for reframing books for further interpretation, such models cannot fully contextualize literature beyond training data explicitly provided to them, unlike the cognitive processes of information retrieval and

---

<sup>47</sup> Adler, *Cruising the Library*, 9.

contextualization that humans do constantly. A study of language models that treats their results as fixed risks the same artificially universal perspective as “common sense” literalist interpretations of books. In this way, Knox’s illustration of book challengers’ hermeneutics similarly illustrates a risk for deploying machine learning techniques uncritically, and foregrounding these concerns pushes scholars to provide further context and informed interpretation to their studies to fully grasp the benefits of computational text analysis.

Two projects that grapple with these considerations and bear heavily on this paper are Matt Warner’s “A Queer Way of Counting” and Laura Mandell’s “Gender and Cultural Analytics.” Both recognize the constructed nature of gender and sexuality over time, and both present cases in which unsupervised machine learning techniques generate topic models that can help interrogate or buttress what exactly is being characterized when labels such as “queer” are deployed to text.

As part of his explorative project to represent queer novels computationally, Warner generates an LDA topic model with MALLET that effectively illustrates the expansive and mercurial nature of queerness. When observing that his model fails to draw clear delineations between imagined differences between texts in his “queer” corpus, Warner is unsurprised. For him, the abstract nature of queer, as he puts it, as “a category defined oppositional” to cisgender and heterosexual norms, does little to actually delineate content-based understandings of text.<sup>48</sup>

Mandell takes a different approach in that she compares other scholars’ stylometric analyses of texts with topic modeling as a means of revealing gender stereotypes perpetuated by unscrutinizing approaches to computational text analysis. By reconsidering patterns of text content in terms of topic clusters, rather than authors’ genders, Mandell illustrates a key tension

---

<sup>48</sup> Matt Warner, “A Queer Way of Counting: Bibliography and Computational Approaches to the Queer Novel,” *New Literary History* 54 (1), 2023: 951.

in how the successful deployment of such tools to study gender require critical reframing, and how approaches lacking these considerations are often doomed to entrench existing biases with an added veneer of objectivity.<sup>49</sup> This recursive, comparative approach to computational studies of text are crucial when centering polysemous understandings of challenged books, and I try to maintain these perspectives throughout this study.

Beyond these critical approaches to understanding gender and sexuality with natural language processing, the use of topic modeling for examining text corpora has been well-tested for more than a decade in examining news content, diachronic trends in journals, and indeed, banned books. An early example of such work is Andrew Goldstone and Ted Underwood’s study of the *Publications of the Modern Language Association* (PMLA), in which the authors create LDA topic models (using MALLET and an original implementation respectively) as a means of tracing the content and rhetoric of the journal over time. Goldstone and Underwood illustrate that topic models, given their variance when tweaking parameters, provide a powerful tool for comparative studies of literature as they present snapshots of text “at different distances, and with different levels of granularity...[granting a] multiplicity of perspective [that] should be understood as a strength rather than a limitation of the technique.”<sup>50</sup> This capacity for polysemous interpretation, within the constraints of computational approaches, make topic modeling a useful method in my study of book bans by serving as a center for comparison between the monosemic readings of book challengers and the pluralistic interpretations of queer reading.

---

<sup>49</sup> Mandell, “Gender and Cultural Analytics: Finding or Making Stereotypes?” 7.

<sup>50</sup> Andrew Goldstone and Ted Underwood, “What Can Topic Models of PMLA Teach Us About the History of Literary Scholarship?” *Journal of Digital Humanities* 2 (1), 2012: <https://journalofdigitalhumanities.org/2-1/what-can-topic-models-of-pmla-teach-us-by-ted-underwood-and-andrew-goldstone/>.

Since Goldstone and Underwood’s piece, the tools for topic modeling have only been further refined, particularly by scholars like David Mimno with MALLET (originally built by Andrew McCollum for Java and with enhancements, including a Python wrapper, by Maria Antoniak),<sup>51</sup> Matthew Jockers with his work *Macroanalysis* and its accompanying materials,<sup>52</sup> and Maarten Grootendorst’s BERTopic<sup>53</sup>—the tool which I use to develop my model and explore in greater detail below.

Others have already begun work with topic modeling to represent the effects of book banning: Marcelo Goncalves, Isabella Langrock, Jack LaViolette, and Katie Spoon deployed an abbreviated topic model in their study of the electoral geographies of book banning by generating an R-based topic model of challenged books’ descriptions using Google Books API.<sup>54</sup> Though their analysis is brief and only references their model in the appendix, the authors’ use of topic modeling and metadata analysis allows them to examine how LGBTQIA+ and BIPOC subjects are specifically targeted in the context of local culture war mobilizations, finding that historically conservative regions that have become increasingly diverse are more likely to experience book bans than their neighbors.<sup>55</sup> Delving into the electoral demographics across which book challenges take place, Goncalves et. al.’s research serves as a useful comparative point not only for the deployment of topic modeling but also for the increasingly detailed scholarship on the scope and impact of book challenges over the past several years. My project

---

<sup>51</sup> David Mimno, “Topic Modeling”: <https://mimno.github.io/Mallet/topics>; Andrew Kachites McCallum, “MALLET: A Machine Learning for Language Toolkit”, 2002: <http://mallet.cs.umass.edu>. Maria Antoniak, “little-mallet-wrapper”: <https://github.com/maria-antoniak/little-mallet-wrapper>.

<sup>52</sup> Matthew Jockers, *Macroanalysis: Digital Methods and Literary History*, (Champaign: University of Illinois Press, 2013): <https://www.matthewjockers.net/macroanalysisbook/>.

<sup>53</sup> Maarten Grootendorst, “BERTopic: Neural topic modeling with a class-based TF-IDF procedure,” *arXiv preprint*, 2022.

<sup>54</sup> Marcelo S.O. Goncalves, Isabelle Langrock, Jack LaViolette, and Katie Spoon, “Book bans in political context: Evidence from US schools,” *PNAS Nexus* 3, no. 6 (June 2024): 6-7.

<sup>55</sup> *Ibid.*, 5; 15.

contributes to this scholarship by delving into the content of challenged texts through a comparison of computational and queer readings.

### **III. Data Summarization**

#### **Corpus overview**

This project relies on datasets from the American Library Association's (ALA's) Office for Intellectual Freedom (OIF) to examine a subcorpus of 200 challenged books targeted in challenges between 2010-2023. The tables in Figures 3-1a and 3-1b provide sample metadata for the 10 most frequently challenged books in the total dataset and in my selected corpus respectively. In all, the 200 titles featured in my corpus were chosen from the total 10,877 challenged texts given their relative accessibility, frequency of challenges during the target period, and their representativeness of challenge motivations and targeted groups.

The digital editions of books in my corpus were obtained through manual digitization, conversion of PDFs and TXT files using optical character recognition (OCR) and Python scripts, and extractions from TEI files kept by the University of Chicago's Textual Optics Lab. 29 of these texts were selected from the dataset for being either challenged more than 50 times or labeled by the ALA as containing LGBTQIA+ content. These were digitized at the University of Chicago Library or retrieved as EPUB, PDF, or TXT files from online repositories including the Internet Archive and Project Gutenberg. I then converted these files to TEI and TXT documents using original Python scripts that preserved structural tags for paragraph divisions and corrected basic errors generated by OCR software during the conversion from PDF formats. The remaining 171 texts were originally collected by the Textual Optics Lab from the History of Black Writing Novel Corpus hosted at the University of Kansas and the US Novel Corpus hosted at the



education, or depictions of certain bodies in general constitute “pornography” unsuitable for the readership where challenges are levied, especially for children and teenagers. Evidence of this trend is apparent when presented with a word cloud of the language used in the complaints for all titles in the dataset (Figure 3-2).

Surprisingly, Figure 3-3 tempers this focus on illicit imagery to some extent since the vast majority of titles are not Graphic Novels or Comics and therefore are unlikely to contain any visual material. This is important to note for my later analysis, as none of the titles within my

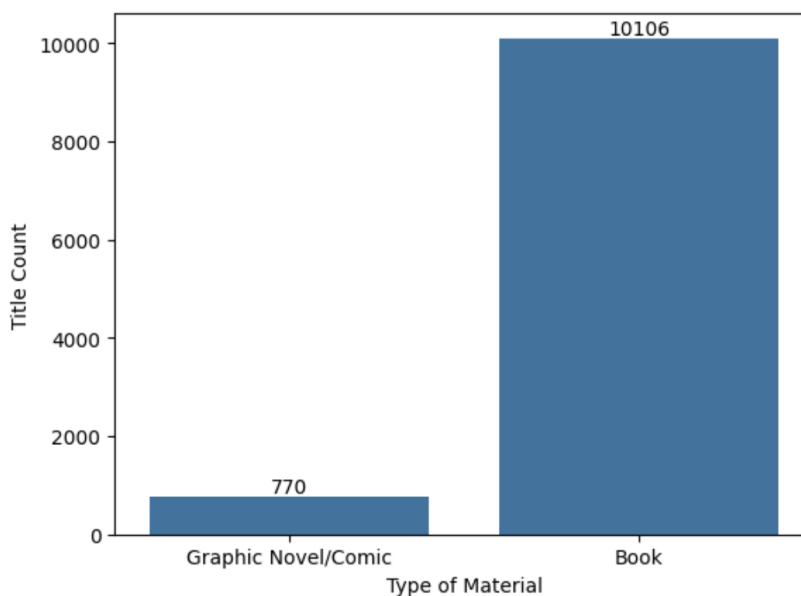


Figure 3-3

corpus contain illustrations or graphic material in their raw text forms, evident by the representation of Children’s titles at only 2% of my corpus.

Figures 3-5a and 3-5b illustrate the distributions of the Title Tags adapted by the ALA from WorldCat’s OCLC subject classifications to give a better sense of the texts targeted along the intersections of race and queer identity. In the total dataset, texts tagged as LGBTQIA+ vastly outnumber the other fields, while my corpus has 46% of titles labeled as BIPOC Fiction,

followed closely by 44% labeled as LGBTQIA+ (with significantly smaller proportions for BIPOC Biography, BIPOC Non-Fiction, and BIPOC/LGBTQIA+ that more closely match the

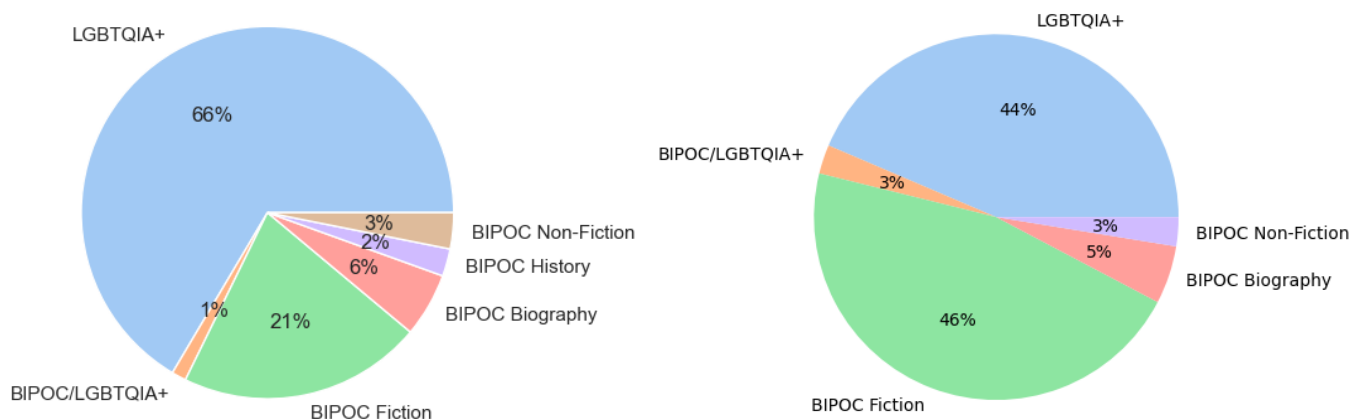


Figure 3-5a (All titles)

Figure 3-5b (Corpus only)

proportions of the overall dataset). These discrepancies in OCLC tags between the larger dataset and my corpus make sense given that one of the primary repositories for my corpus dataset is the History of Black Writing Novel Corpus and that many of the “LGBTQIA+” books were published between 1990 and 2022, outside the date range for the US Novel Corpus. However, it is worth interrogating the ALA’s relatively small number of titles labeled BIPOC/LGBTQIA+, and I will examine the implications of this lack of hybrid labeling further in my findings.

Attention to the distribution of challenge tags illustrates the hostile perceptions of the challenged books and their content by people challenging titles. As can be seen in Figures 3-6a and 3-6b, the proportions of challenges along the three most frequent complaint tags, “Sexually Explicit,” “Offensive Language,” and “Violence” are consistent between datasets. However, the following challenge tags shift in order between the larger dataset and corpus subset:

“LGBTQIA” (7%), “Religious Viewpoint” (7%), and “Occult/Satanism” (6%) become “Religious Viewpoint” (6%), “Racist Content” (5%), and “LGBTQIA” (5%), indicating a slight

shift in my corpus’ representation from the total – and most importantly a slight underrepresentation of texts targeted on the basis of being LGBTQIA+. Furthermore, tags for “Body Description or Function,” “Death,” “Grammar,” “Humor,” and “Supernatural” drop out for my corpus entirely.

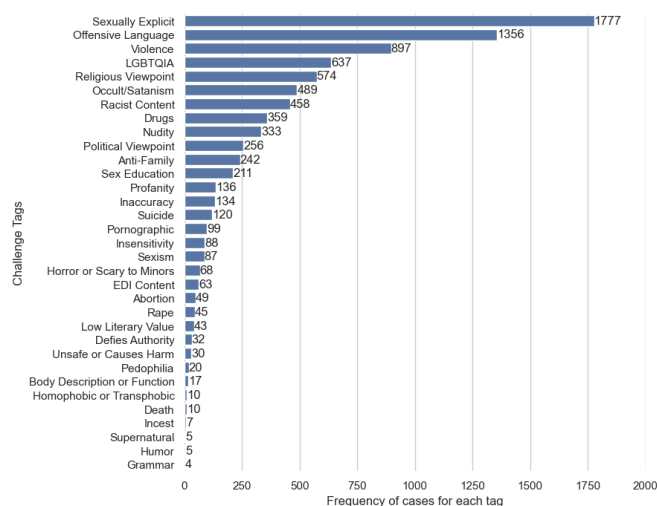


Figure 3-6a (All titles)

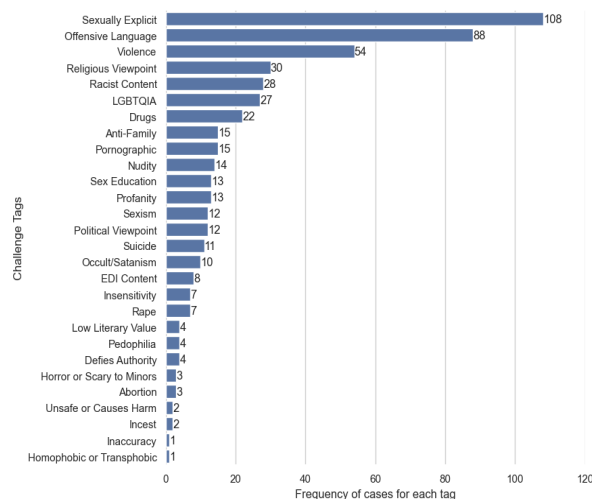


Figure 3-6b (Corpus only)

An examination of encoded senses of queerness becomes all the more pressing given these statistics illustrating the targeting of LGBTQIA+ and BIPOC titles in book challenges. However, it is prudent to expand this examination from the strict challenge of “LGBTQIA” to consider how texts regarding sex and gender are being characterized broadly in the challenges. Especially given the focus of individual challenge complaints on “pornographic” material and sexuality in general, tags such as “Sexually Explicit,” “Sex Education,” “Homophobic/Transphobic,” “Nudity,” “Incest,” and dogwhistles like “Pornographic,” “Anti-family,” and “EDI Content” are important to consider and contextualize in their own right.

## IV. Overview of computational methodology

### Pre-processing

Pre-processing and cleaning of my corpus' metadata took place over a period of nine months. The majority of cleaning for the challenged titles metadata took place in OpenRefine given its relative transparency and easy manipulation, though some minor changes were made using Google Sheets and through the use of Python scripts for editing CSV and TXT files. The first steps of metadata cleaning were conventional stripping of leading and trailing whitespaces, standardizing facets and field values (e.g., "LGBTQIA+" and "LGBTQ" challenge tags became "LGBTQIA"), and clustering equivalent values so that they are read identically. Publication and authorship data were manually added where missing, and these changes were incorporated into the TEI editions of my corpus texts. The most meaningful step in metadata pre-processing was to implement a one-hot encoding of my Challenge Tag features in order to more easily quantify their values. This helped with one of the primary difficulties of computational approaches to the dataset, namely scaling issues due to the fact that the majority of titles have been challenged 10 or fewer times, while a tiny percentage have been highly targeted more than 100 times. In order to illustrate trends in challenge labels, I added features for "Pornography" and "Pedophilia" after creating a document term matrix for the specific complaints. I also employed a backward-elimination approach to feature selection by eliminating strictly identifying features in the Titles dataset that lack descriptive value (e.g., ISBN numbers).

The remaining pre-processing of the 200 texts in my corpus was largely conducted manually. I generated TEI and/or TXT editions of the texts depending on their source using the Tesseract OCR library and my own Python pipelines supplemented by the Calibre desktop application. Some of the editions included prefaces, conclusions, and author biographies I did not

want to include, while others only contained the text body. I then performed manual quality control of the text files that I had obtained using OCR and corrected numerous spelling mistakes and several obvious elisions. Despite preliminary attempts to remove front and back matter for each text, several files still contained copyright material, which I then removed along with other unwanted text while developing my topic models.

### **Computational Analysis and Operationalization**

My computational approach to analyzing the 200-text corpus was based around the creation of a BERTopic model to generate representative topics for comparison with challengers' "distant reading" of texts. The training of this model was conducted using Python scripts and Slurm batch jobs run using the University of Chicago Research Computing Center's (RCC) Midway 3 cluster for High Performance Computing (HPC). Once the models were trained, I used a Jupyter Notebook with Python libraries including Pandas, Numpy, Seaborn, Matplotlib, and HuggingFace's Transformers for interpreting the models and generating visualizations.

Building the topic model was arduous given the size of the corpus and BERTopic's pretrained models. After quickly failing to train the BERTopic model locally due to memory limitations and corpus scale, I was able to construct two sample topic models using an allocated 164 GB of memory on a CPU partition of the Midway 3 cluster. I adapted code from Jeffrey Tharsen into Python scripts that could be submitted as in a batch job and required breaking the corpus texts into sentence documents totaling 1,746,806 sentences among all 200 texts. Initially, I constructed two sample topic models and provided tokenization parameters to fit the corpus into 40 topics and save the top 30 word-tokens for each topic. I otherwise adopted the default

(unseen) pipeline built into the BERTopic module:<sup>57</sup> 1) map sentence embeddings into a 384-dimensional vector space using a SentenceTransformer (SBERT) model, 2) reduce the embeddings' dimensionality using Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) and cosine similarity, 3) cluster these reduced vectors using HDBSCAN, 4) tokenize these clusters in turn as topics using sci-kit learn's CountVectorizer, and 5) generate a document term matrix of topics based on class-based total frequency/inverse document frequency (c-TF-IDF) weights so that topics can be spatially represented in comparison to one another.

The first of these models retained every word during tokenization, while the second model filtered out stopwords selected from the corpus as well as sci-kit learn's default list of stopwords. It took nearly 25 hours to successfully create these two 11.5 GB models, the upper time and size limit for future builds. Clearly benefiting from the exclusion of stopwords, the second model generated much more coherent topics than the first. Promisingly, one topic in this second model collapsed several sexual, racial, and gender identities (the top tokens were “gay,” “white,” “black,” and “racial”) that could be meaningfully interpreted as a “social identity” topic. However, there was clear room for improvement since the model included one topic devoted to copyright terms and another composed of basic verbs and demonstrative pronouns (“know,” “dont,” “thats”).

I then cleaned the corpus' text files to remove copyright material and added a custom list of stopwords to my vectorizer to include frequently occurring particles, prepositions, and additions from the NLTK English stopwords list, as well as adapting portions of the list of

---

<sup>57</sup> Maarten Grootendorst, “The Algorithm,” *BERTopic*: <https://maartengr.github.io/BERTopic/algorithm/algorithm.html#code-overview>.

common proper names provided by Matthew Jockers.<sup>58</sup> In order to draw out greater nuance from the previous model's topics, I increased the specified number of topics from 40 to 60 in the next run and filtered using the expanded stopwords list. While the new model included more granular social categories in its topics, many of them contained too few tokens to be significant.

Specifying 50 topics in the next run improved results slightly, but several topics were still too small. However, several sociological and religious topics began to appear that were comparable to challenge tags, a development that made me reluctant to reduce the number too much.

Ultimately, I settled on a topic model with 45 topics. For this model, I added to my Python code in order to retain the topic classifications for each of the 1,746,806 sentences (the appropriate unit for “documents” required by SBERT for embeddings) used to train the topic model, and then saved the resulting Pandas DataFrame as a local output.<sup>59</sup> This last step was a crucial addition given that BERTopic does not by default retain the documents with which each model is trained in order to reduce storage space. Only by saving the original document classifications was I able to later aggregate the classifications to represent topic distributions with the books themselves rather than their sentences.

Once I had built and stored the topic model, I explored various approaches to representing and synthesizing its output locally in Jupyter Notebook. I first specified a subset of the 45 topics that included the twelve most coherent topics with any relative bearing on book challenge concerns (shown in Figure 5-1).<sup>60</sup> The labels for these topics were assigned based on

---

<sup>58</sup> Matthew Jockers, “Expanded Stopwords List”: <https://www.matthewjockers.net/macroanalysisbook/expanded-stopwords-list/>.

<sup>59</sup> Maarten Grootendorst, “get\_document\_info(),” *BERTopic*: [https://maartengr.github.io/BERTopic/api/bertopic.html#bertopic.BERTopic.get\\_document\\_info](https://maartengr.github.io/BERTopic/api/bertopic.html#bertopic.BERTopic.get_document_info).

<sup>60</sup> It should be noted that several more of the generated topics were coherent and, in another study, may prove interesting for analysis. I had previously considered including topics labeled “Age/aging” and “Physical metrics” to try and include proxies for sociocultural identifiers like age, height, and weight, but ultimately found that the topics were outliers and not particularly useful in my analysis. Additionally, topics related to events and time, vernacular

their 30 most representative tokens retained by the model. Further elaboration on the assignment of these labels and my interpretation of the topics generated in comparison with the labels assigned to book challenges will be provided in Section V.

With these topics selected, I loaded my corpus texts into Jupyter Notebook and loaded the previously saved Pandas Dataframe with topic information for each sentence in the corpus. I instantiated a new Dataframe in which each sentence retained a metadata field specifying its source text. From here, it was easy to compile all of the topic classifications for each title's sentences in order to identify not the distribution of topics within each title and which sentences in each title correspond to a given topic. This process was instrumental for rendering the data useful for my analysis below.

With text aggregation complete, I was able to compare the topics present in each text to determine topic classifications for books based on raw and relative frequency and to rank the corpus titles according to their topics. The raw classifications, while interesting, unsurprisingly often characterized the longest texts as the "most" representative of topics: this is evident by the appearance of *Ulysses*, *From Here to Eternity*, and *Shogun* among the top 15 most frequent texts for the Gender/Sexuality/Queer topic in the raw classifications contrasted with their significantly lower rankings in the weighted classifications since the topic makes up less than 1% of their content. The weighted classifications proved more useful, and I assigned each topic a percentage score to reflect its proportional distribution within each book. From this process, I was able to calculate more precise rankings of each topic for each book and change displays to represent only those twelve topics previously illustrated as cogent and most relevant for book challenger concerns. The final step, for legibility's sake, was to create another DataFrame with indexes for

---

and onomatopoeic words, computation, astronomy, physics, and communication were all clearly distinguished in the model but were not considered to have particular relevance in comparison to challenge tags.

all 200 titles and columns representing the top 5 representative topics for each book. The completion of these steps finally allowed for the interpretive analysis of the topic model and presented a computational understanding of challenged texts according to these parameters.

## V. Findings

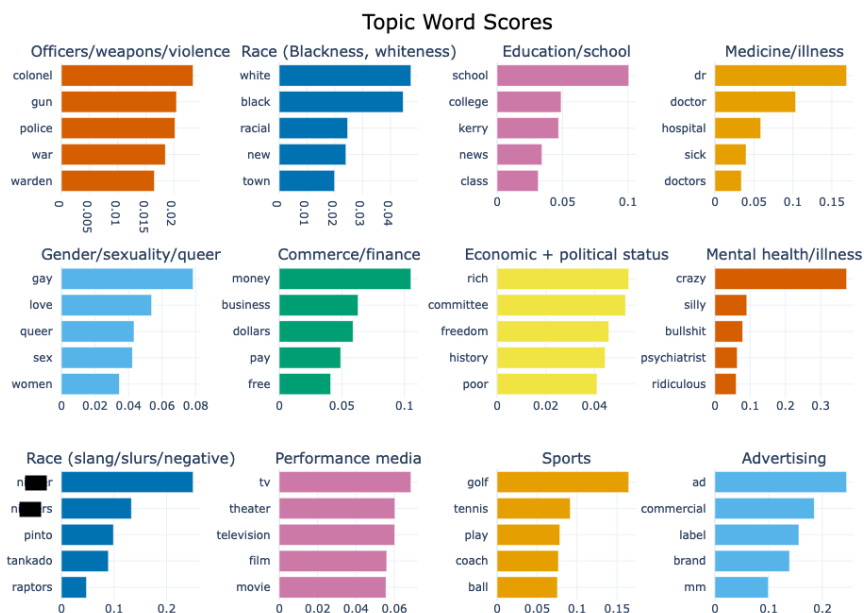


Figure 5-1

Among the twelve topics identified in Figure 5-1, I will most closely examine the topic related to gender, sexuality, and LGBTQIA+ topics – hereafter referred to as “Gender/sexuality/queer” or topic number 9 – but will contextualize it with reference to other topics related to sociopolitical categories and cultural symbols. Principal among these, listed in order of frequency within the corpus, are “Officers/weapons/violence,” “Race (Blackness, whiteness),” “Economic + political status,” and “Mental health/illness.” In the analysis that follows, I illustrate three main findings: first, that the model contradicts an understanding of challenges that would classify “queer” or LGBTQIA+ as deviant or even distinct from sexuality

or gender in general; second, that the model struggles to represent topics intersectionally; and third, that the topic model underscores book bans' disproportional targeting of texts depicting sex, gender, or queer characters compared to more prominent topics like violence.

Before examining the topics in greater depth, it is important to illustrate some of the interpretations made while determining labels for the topics based on their token probabilities. After all, this process of labeling and assigning coherent meaning to the tokens is itself a human-derived, interpretation of the “reading” conducted by BERTopic. I sought to capture the representative nature of each topic based on the top 30 tokens preserved during model construction. Figure 5-1 represents the top 5 of these tokens, but I found it useful to delve further into the list in order to more precisely represent the meaning of each topic.

As a result, the groupings of tokens for several of these topics fundamentally contradict some of the interpretive views of book challengers, particularly with the topic I have labeled “Gender/sexuality/queer.” Based on previous model results and my own presumption that LGBTQIA+-related tokens would emerge as a distinct topic, I saw the three most prominent tokens (“gay,” “love,” and “queer”) for this topic and was primed to label it “LGBTQIA” to match the label for the corresponding challenge tag.

However, it became very clear as I read further that the model had built the topic not only from “queer” terms, but out of a range of tokens related to gender and sexuality including – and at times, even more pertinent to – cisgender and heterosexual identities as well: “sex,” “marry,” “married,” “wife,” “straight,” “wedding,” “condom,” and “virgin,” to name only a few. This *lack* of distinction between LGBTQIA+ and non-LGBTQIA+ concepts of gender and sexuality represents a crucial finding insofar as the topic model does not classify queer notions of sexuality, gender, and romance as clearly distinct from hetero- and cisnormative sociocultural

expectations, in clear contradiction of the “commonsense” queerphobic readings offered by many of the challenge cases against the corpus texts.<sup>61</sup> In labeling the topics, I tried to take a similarly broad and representative view of the available 30 tokens and their accompanying probability scores to more precisely characterize titles for deeper analysis. In doing so, I found that the model’s reflection of gender and sexuality included queerness at face, thus refuting a basic premise of nearly 15% of book challenges that LGBTQIA+ identity is distinct and intrinsically harmful compared to dominant concepts of sexuality and gender.

The remaining labeled topics were helpful for further characterizing the “Gender/sexuality/queer” texts within the context of the topic model and book challenges themselves. As mentioned, BERTopic’s built-in methods allowed for comparing topics relative to one another using c-TF-IDF weighting. From the resulting topic embeddings, Figures 5-2 and

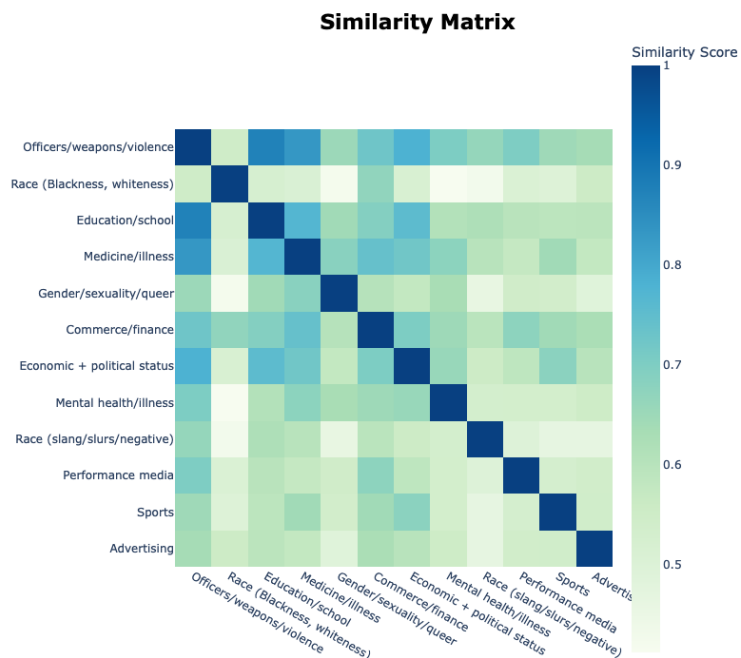


Figure 5-2

<sup>61</sup> Though not the focus of my analysis, I similarly found it interesting that for the topic “Race (Blackness, whiteness),” “white” is the predominant token, in a reflection of how many of the texts in the corpus turns a racializing and interrogative lens on whiteness itself, when most frequently and historically, racialization is turned by oppressive groups outwards as a tactic of domination.



Figure 5-3

5-3 illustrate the similarity of all of the labeled topics to one another using a heatmap and spatial representations respectively that allows us to map topics in relative mathematical space. Hoyt Long describes this measurement technique as a way to “read relationally,” in which studying the proximities and distances of embedding clusters – in this case, topics – helps inform our interpretation of the embeddings themselves, a strategy closely in line with principles of contemporary sociology and cultural studies, as Austin Kozlowski, Matt Taddy, and James

Evans argue.<sup>62</sup> Visualizing a topic's proximity to others based on the distance between their embeddings is therefore useful for holistically contextualizing each topic's representation of the corpus texts.

When viewed from this lens, some interesting trends emerge in topic comparison. In particular, "Race (Blackness, whiteness)" forms the center of three outlying topics relating to "Economic + political status" and "Race (slang/slurs/negative)."<sup>63</sup> The clustering of these three topics makes sense, particularly given the semantic relationship between the first and third of these topics in natural language and the sociopolitical approximations of race, particularly in regard to white and Black communities and their histories, in relation to economic and political power. The clustering of "Gender/sexuality/queer" with the nearest topic "Mental health/illness" (with which it shares a similarity of 68%) is also illustrative and concerning for understanding how social categories are represented by the topic model. Unfortunately, this proximity is unsurprising given the representation of many characters in the corpus experiencing trauma in navigating gender and sexuality in titles like *The Perks of Being a Wallflower* along with real-world statistics relating to anti-LGBTQIA+ discrimination affecting mental health.<sup>64</sup>

What is unexpected, however, is the distance the model places between the topics for "Gender/sexuality/queer" and "Race (Blackness, whiteness)": as Figure 5-3 illustrates, the two topics are among the furthest apart within the twelve selected, and "Race (Blackness, whiteness)" is the *least* similar topic to "Gender/sexuality/queer" with a similarity score of only 43%. This result is surprising for several reasons. Given that previous versions of the topic

---

<sup>62</sup>Austin C. Kozlowski, Matt Taddy, and James A. Evans, "The geometry of culture: Analyzing the meanings of class through word embeddings," *American Sociological Review* 84, no. 5 (2019): 922.

<sup>63</sup> Notably, the model distinguished this latter topic from the larger "Race (Blackness, whiteness)" but grouped the two quite closely together, maintaining a clear proximity between racist slurs and other race-related tokens.

<sup>64</sup> The Trevor Project, "2023 U.S. National Survey on the Mental Health of LGBTQ Young People": [https://www.thetrevorproject.org/survey-2023/assets/static/05\\_TREVOR05\\_2023survey.pdf](https://www.thetrevorproject.org/survey-2023/assets/static/05_TREVOR05_2023survey.pdf).

model clustered similarly-constituted topics quite closely – even grouping tokens related to race and queerness under a single “social identity” topic in my first model – I expected this model to generate similarly approximate topics.

Moreover, to have these two topics not only clearly distinguished, but so vastly distant in the model – with only “Race (Blackness, whiteness)” and “Mental health/illness” having a lower similarity score (41%) of all pairs of categories compared by their c-TF-IDF weights – seems to indicate a surface-level contradiction with principles of intersectionality. Relatedly, the distance draws separations from racial categorization and its neighboring topic of “Economic + political status” from categorization according to gender and sexuality and its nearest neighbor “Mental health/illness” – distinctions that similarly do not translate into reality in which mental health, socioeconomic, and political power affect one another and in turn affect people in salient ways based on how race, gender, and sexuality intersect. It is crucial to remember that conceptual tools for understanding power and relationships in human identity are mutually constitutive and intersecting, and the model’s lack of representation for this is even more surprising when considering the arguments of scholars like Laura Nelson and Kozlowski, Taddy, and Evans, who have effectively demonstrated how vectorized word embeddings can capture various facets of personal identity through an intersectional lens.<sup>65</sup>

Elements of token-level effectiveness in representing intersectionality do appear in the BERTopic model, particularly in how the model distributes gendered tokens across sociological topics. One example indicates a masculinization of race and a feminization of sexuality, queerness, and gender itself within the model: “girls” and “women” appear as top tokens for

---

<sup>65</sup> Kozlowski, Taddy, and Evans, “The geometry of culture,” 36. Nelson references their work in her own article: see Laura K. Nelson, “Situated Knowledge and Partial Perspectives: A Framework for Radical Objectivity in Computational Social Science and Computational Humanities,” *New Literary History* 54, no. 1 (Winter 2023), 864-869.

“Gender/sexuality/queer” but neither appear in “Race (Blackness, whiteness)”, while “men” and “man” both appear among the top tokens for “Race (Blackness, whiteness).” Considering the additional degree of abstraction in c-TF-IDF compared to traditional TF-IDF – generating relative weights for topics of clustered tokens aggregated into a single document rather than the weights of individual token embeddings – it may be that BERTopic embeddings are less effective at representing intersectionality on a topic level as has been shown on a token level. This initial observation requires further interrogation to better understand how the model constructs gendered, sexual, and queer concepts in relation to other sociological concepts. Keeping this sociological lens at the front, I can now excavate the distinctions between intersectionality found through close reading of the challenged books from distant reading of the corpus by BERTopic.

Examining these titles most proportionally representative of “Gender/sexuality/queer” further illustrates the model’s characterization of these concepts, especially when keeping in mind the statistics for challenges against them. By order in which the highest proportion of their content relate to this topic (illustrated in Figure 5-4), six titles are worth examining more closely based on their other representative topics and their representativeness of “Gender/sexuality/queer”: *This Book is Gay*, *Queer: The Ultimate LGBT Guide for Teens*, *Beyond Magenta*, *Being Jazz*, *All Boys Aren’t Blue*, and *Invisible Life*. By way of introduction to the texts, all six of these books center queer protagonists and are classified as “Young Adult” books. Interestingly, the 5 of these most representative of the “Gender/sexuality/queer” topic are non-fiction books published after 2010. The top 2 – *This Book is Gay* and *Queer: The Ultimate LGBTQ Guide for Teens* – are both framed as manuals or guides to serve as information resources for queer teenagers, while the next three titles are either memoirs or collected

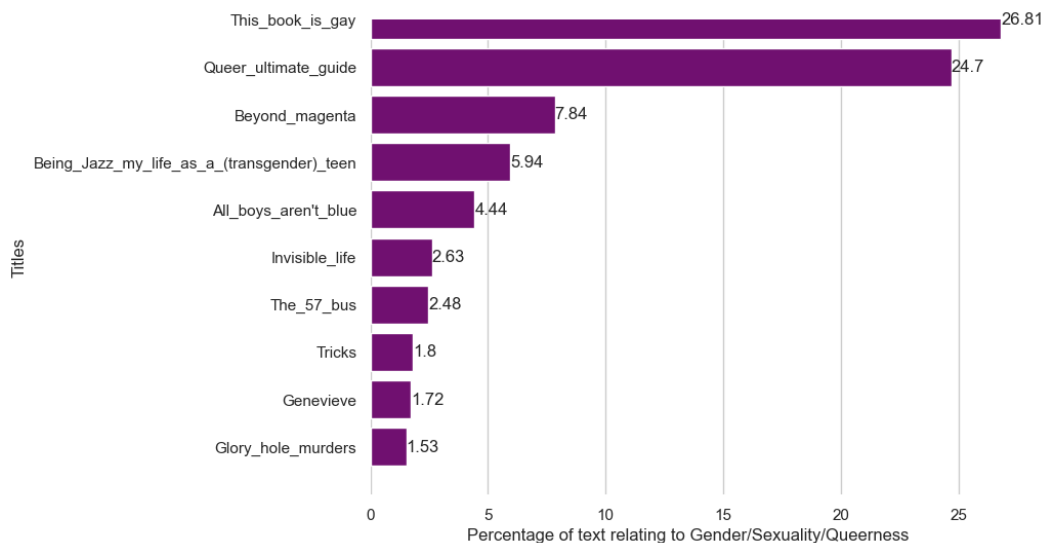


Figure 5-4

testimonials. *Invisible Life* is the only novel, the only one published in the twentieth century, and the least frequently targeted on the list with only 3 challenge cases from 2010-2023.

The importance of these distinctions become even clearer when comparing the statistical results from the topic model with the challenge tags assigned to these queer titles. In the table seen in Figure 5-7, I provide the percentage of each of the six books that falls under the “Gender/sexuality/queer” topic, the number of times each title has been challenged between 2010-2023, the top 3 most represented topics from the BERTopic model for each title, and the top 3 most frequent challenge tags for each title. All six of these titles have been challenged most frequently on the basis of being “LGBTQIA.” Through comparison, their values reveal two main findings for how genre and sociocultural categories affect the frequency of book challenges.

First, the genre of each of the six titles corresponds closely with the proportions of their content that the topic model labels “Gender/sexuality/queer” and the quality and quantity of challenge tags leveled against them. The guide books contain between 24%-27% of their text relating to “Gender/sexuality/queer” and have both been challenged as “Sex Education” materials; the three personal accounts fall within 4%-8% and all were labeled by the model with

the “Education/school” topic; and only 2.63% of the novel is represented by “Gender/sexuality/queer”—despite this being *Invisible Life*’s most represented topic among those labeled and being *exclusively* challenged on the basis of being an “LGBTQIA” text. Given BERTopic’s reliance on token-level classifications and SBERT’s use of sentences as the document size for training, it is unsurprising that genre influences the percentages of the topic that the model determines. After all, a “how-to” guide is more likely to repeat consistent, explicit language to present its content than a memoir or experience-based narrative, while a novel that repeated the same topic-heavy tokens ad nauseum would be criticized for its prose. Furthermore, the three memoirs all include chapters describing (or are entirely centered on) queer experiences at school, and it therefore makes sense that the education-related topic would represent such a high proportion of content within these texts, whereas *This Book is Gay* does not extensively feature school settings and *Invisible Life*’s protagonist Raymond occupies the professional world. These correlations of genre with secondary topics and challenge tags demonstrate how effectively the model’s generation of topics map to linguistic conventions and how book challengers react in turn.

Second, though the percentage of each text classified as “Gender/sexuality/queer” does not directly correlate with how frequently it was targeted by challenges, these six texts are all within the top 50% of most challenged books in the corpus, the five non-fiction texts fall within the top 20%, and *All Boy’s Aren’t Blue* is the single most challenged title in the corpus and the second most targeted title of all challenged books recorded during the 2010-2023 period. Given that texts related to sex and queer identities are clearly targeted at high rates, it could be expected that the proportion of “Gender/sexuality/queer” representation in a given text might elevate the frequency at which it is challenged. However, *All Boy’s Aren’t Blue*, only the fifth highest in the

list of representative texts for “Gender/sexuality/queerness,” was challenged over 200 times more than the second most and more directly educational – and from the perspective of challengers, perhaps more directly indoctrinating – *Queer: The Ultimate LGBTQ Guide for Teens*. Moreover, the largest labeled topic within *All Boy’s Aren’t Blue* is not actually “Gender/sexuality/queer,” but rather “Race (Blackness, whiteness).” As to be expected, some of this topic’s most representative texts are also among the most frequently challenged in the corpus, but at overall lower percentiles than “Gender/sexuality/queer.” How to explain this?

Once again, intersectionality becomes crucial to consider. What a strictly topic-based reading of the model fails to account for is the intersectionality at work in George Johnson’s memoir, and depressingly, the intersection of anti-Blackness and homophobia that motivates challengers to target so many of the texts in the corpus. In his introduction, Johnson relates this dynamic well:

“This book is an exploration of two of my identities— Black and queer— and how I became aware of their intersections within myself and in society. How I’ve learned that neither of those identities can be contained within a simple box, and that I enter the room as both of them despite the spaces and environments I must navigate. In the white community, I am seen as a Black man first—but that doesn’t negate the queer identity that will still face discrimination. In the Black community, where I more often find myself, it is not the Black male identity that gets questioned immediately. It is that intersection with queerness that is used to reduce my Blackness and the overall image of Black men.”<sup>66</sup>

Johnson echoes many of the concerns of critical catalogers like Melissa Adler and Emily Drabinski in arguing that neither of his identities can be neatly classified or disembodied from his simultaneous experience of them – or the persecution he faces for them. It is insufficient to characterize the surge in book bans as distinctly queerphobic, sex-negative, or racist in nature:

---

<sup>66</sup> George M. Johnson, *All Boy’s Aren’t Blue: A Memoir-Manifesto*, (New York: Farrar, Straus and Giroux, 2020), 3.

rather, all of these priorities fuel one another and inform the basis for which certain books get targeted again and again at the intersections of various marginalized facets of humanity.

The intersectionality present throughout many of the corpus texts also helps explain this model’s alienation of racialized identity from those pertaining to gender, sexuality, and queerness, as discussed above and represented in the intertopic distance map of Figure 5-3. The model weighs probabilities for which topics a sentence document falls into and can display the distribution for this weighting to show how much each token in a sentence pulls that sentence towards a discrete topic.

For example, Figure 5-5 illustrates how one sentence from *This Book is Gay* features tokens that the model identifies as likely to fit into “Race (Blackness, whiteness),” “Gender/sexuality/queer,” or “Economic + political Status.” This reveals a level of nuance built into the functionality of BERTopic that in its best implementations may in fact allow for intersectionality-oriented data visualization. Unfortunately, the reduction of each sentence to a single topic in its general display – in this case, to “Gender/sexuality/queer” – undermines the

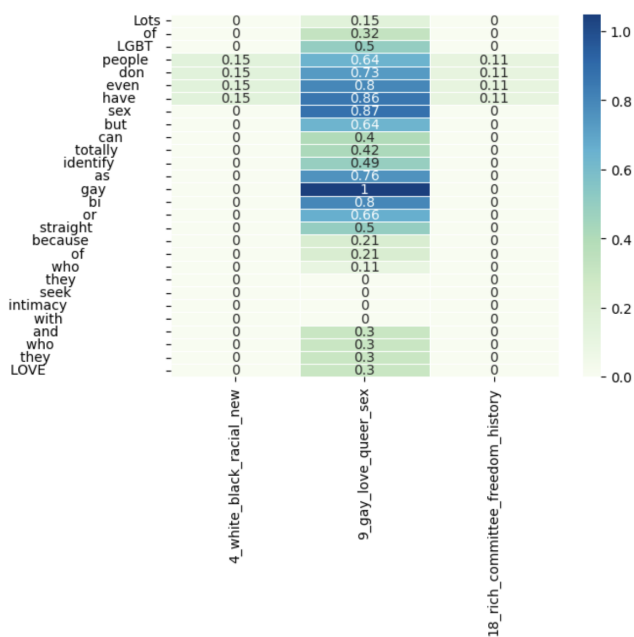


Figure 5-5

possibilities for intersectional analysis and representation within BERTopic’s architecture that is necessary for more precisely understanding these texts. An interface that retained the probabilistic ambiguity of the model’s classifications would be a useful representation of the messiness of intersectionality and semantic meaning and avoid the erasure carried out by the assignment of a single, discrete topic label. In a more drastic example, the model classifies the sentence “No wonder so many kids of color and queer kids don’t feel they have the opportunity to speak for themselves”<sup>67</sup> from *All Boys Aren’t Blue* solely under the topic “Gender/sexuality/queer” because it does not evidently register the phrase “kids of color” as an indication of race as strongly as it does “queer”—despite the equal weighting of these terms under close reading. On a technical level, it is less likely that the model will recognize the tokens “kids” and “of” as related to race, even given BERTopic’s ability to relate tokens in sequence bidirectionally so that “color” informs the prior two tokens. They are simply too likely to be represented in non-racializing ways throughout the corpus for the model to weigh them appropriately here. These close readings illustrate the limitations of categorical readings of texts like topic models even when deployed carefully and with best intentions. This caution becomes all the more crucial when comparing the model’s topics with challenge tags as another “distant reading” of the corpus.

Finally, in directly comparing the labels for the topics identified within the corpus with the challenge tags assigned by the ALA, we can immediately see how “Gender/sexuality/queer” correspond with the aforementioned anxieties over queer visibility and sexual material at the heart of book bans. In considering the challengers’ tags in parallel to these topic labels, the tags for “Sexually Explicit,” “Pornographic,” “Nudity,” and “LGBTQIA” provide obvious analogs.

---

<sup>67</sup> Ibid., 64.

Similarly, the topics, “Race (Blackness, whiteness),” “Officers/weapons/violence,” and “Mental health/illness,” closely resemble the challenge tags “Racist Content” and “EDI Content,” “Violence,” and “Suicide” respectively. I provide Figures 5-6a and 5-6b to illustrate the distributions of top challenge tags alongside representative topics respectively for the corpus.

Several divergences appear when comparing the statistics for these topic-tag cognates. Most notably, “Sexually Explicit” represents the most frequent challenge tag for the corpus, yet “Gender/sexuality/queer” only appears as the fourth most prevalent topic (equal to “Commerce/finance,” which has no cognate among the challenge tags). Conversely, “Violence”

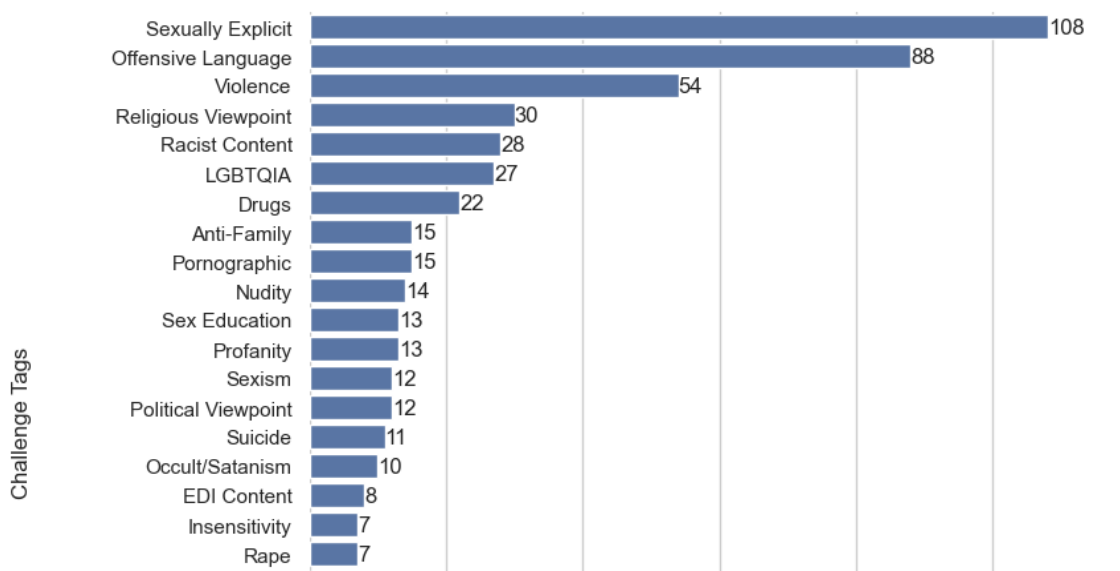


Figure 5-6a (Challenge Tags)

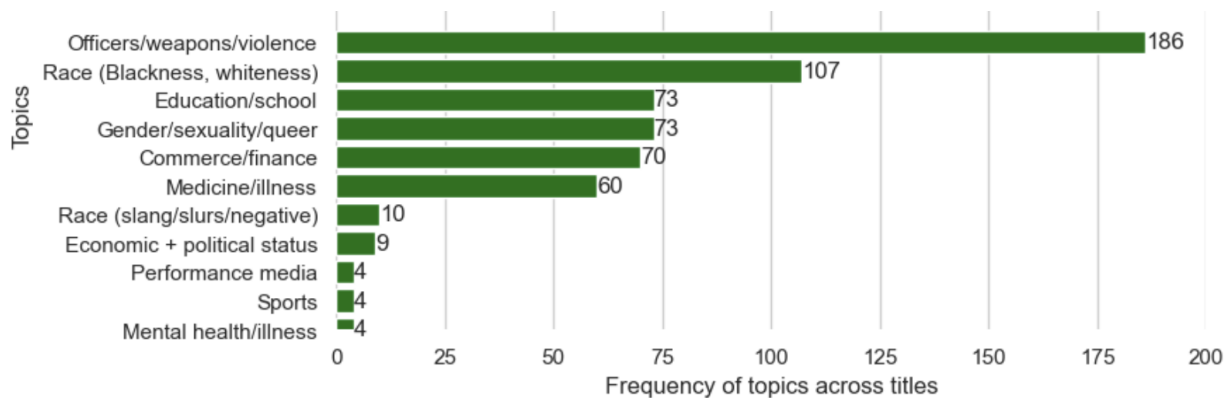


Figure 5-6b (Model Topics)

appears as the third most frequent Challenge Tag and applied to 27% of the corpus, while “Officers/weapons/violence” was by far the most prevalent topic and appeared as one of the three most representative topics for 93% of texts.

These discrepancies reveal a stark divide between sexual, gender-focused, and queer content – highlighted as especially undesirable in book challenges but relatively infrequent in the corpus – and far-more prevalent and far-less challenged “Violence.” Such a distinction in the framing of these texts emphasizes the dramatically queerphobic and sex-averse nature of these challenges, in which representation of such issues and their relation to people is considered more objectionable than weapons or physical violence – and importantly, the state agents associated with them. Among the top 30 tokens for the “Officers/weapons/violence” topic are “gun,” “war,” “fight,” “shot,” and “pistol,” as well as specifically governmental and military officials (“colonel,” “police,” “cop,” “guard,” and “cops”). Evidently from challenge tags, despite the prominence of gun violence and police violence in the United States that is reflected in the corpus texts in turn, the far more pressing danger perceived by challengers is sexuality or queer people and their ideas.

The real world parallels of this finding are disturbing: four days into 2024, a teenager shot eight people at Perry High School in Iowa before dying by suicide; barely a month later, a woman fired an AR-15 into a church in Houston.<sup>68</sup> Both attacks were elevated on social media and Fox News – not for the horrific violence of the attacks, but because the shooters were (in one case, mistakenly) identified as LGBTQIA+ by right wing social media accounts. Chaya Raichik, founder of the Libs of TikTok Twitter/X handle with more than 3.2 million followers as of July

---

<sup>68</sup> R.G. Cravens, “Anti-Trans Rhetoric Helps the Right Pivot to Gun Violence Myths, Sell Guns,” *Hatewatch*, The Southern Poverty Law Center, March 24, 2024: <https://www.splcenter.org/hatewatch/2024/03/22/anti-trans-rhetoric-helps-right-pivot-gun-violence-myths-sell-guns>.

2024, took the incidents as evidence of “the LGBTQ movement turning youth into violent extremists.”<sup>69</sup> Thus, we can understand how book challenges’ relative lack of concern for violence itself compared to the perceived danger of queer topics does not exist in a vacuum: it is in line with a particular cultural fixation on LGBTQIA+ people as inherently dangerous initiated by Anita Bryant and stewarded by her counterparts today. Raichik is also well-known for railing about “pornographic” books in libraries, and more than thirty of the schools and public libraries she has named in her posts have subsequently been targeted with violence against workers or bomb threats.<sup>70</sup> Ironically, in January 2024, the head of the Oklahoma State Department of Education named Raichik to the state’s library advisory board, where she now has a direct role in removing and barring books from circulation within public libraries – despite having no background in librarianship or education.<sup>71</sup> Given these real-world trends surrounding anti-queer legislation and the targeting of libraries and public education, it is more important than ever to understand the precise ways in which book challenges distort and restrict queerness and LGBTQIA+ representation in literature.

## VI. Limitations and Representational Shortcomings

Before concluding, I must address the limitations of my dataset, methods, and analysis for this project. First and foremost, the corpus I have presented here is limited in its scope. This is firstly due to the incompleteness of the original list of challenged titles. As detailed by the

---

<sup>69</sup> Christopher Wiggins, “Fox News and GOP politicians falsely claim Lakewood Church shooter was transgender,” *Advocate*, February 12, 2024:

<https://www.advocate.com/news/false-transgender-claim-texas-shooter>; <https://x.com/libsoftiktok>.

<sup>70</sup> David Ingram, “After Libs of TikTok posted, at least 21 bomb threats followed,” *NBC News*, February 7, 2024: <https://www.nbcnews.com/tech/internet/libsoftiktok-x-chaya-raichik-bomb-threat-twitter-of-libsoftiktok-rcna102784>.

<sup>71</sup> Matt Laviertes, “Libs of TikTok creator accused of inspiring school bomb threats named to Oklahoma library board,” *NBC News*, January 23, 2024: <https://www.nbcnews.com/nbc-out/out-news/libsoftiktok-bomb-threats-oklahoma-library-committee-rcna135369>.

ALA, the challenges they track are identified only through a voluntary reporting mechanism and journalism-based data collection conducted by the OIF. This means that many libraries and schools whose challenges do not reach media attention or who are unaware of the ALA's survey system are categorically unable to report titles challenged.<sup>72</sup> Secondly, I selected the corpus of 200 texts in large part due to the accessibility of digital copies and the availability of books for digitization. The process of digitizing books, conducting OCR, and writing code to convert texts between formats is valuable but time-consuming academic labor, and I had the significant benefits of access to high-quality scanners, the University of Chicago Library's collections, and the aforementioned repositories of the Textual Optics Lab. Even so, more time would allow for an expansion of the corpus for computational analysis beyond a mere 2% of the challenged texts. Thirdly, my corpus does not include any of the graphic narratives or illustrated children's books. As shown in section III, such titles do not make up a majority of challenged books but are among many of the most *frequently* challenged titles over the last fifteen years. Challengers' fixation on "pornographic" and explicitly visual materials demands that these texts receive due attention. Future projects may very well combine the methods I have presented here with computer vision technologies. Particularly when it comes to representations of embodied characters or authors in narrative, a study of graphic materials challenged in the United States is crucial for understanding not just the words but also the images that challengers find so objectionable.

The next limitation broadly applies to computational and quantitative methods when applied to textual material. Especially when grappling with socially constructed categories like those represented in the dataset's labels (BIPOC and LGBTQIA+ classifiers, distinctions

---

<sup>72</sup> Furthermore, libraries engaging with the ALA face increased scrutiny or punishment in many parts of the United States as of 2024, as new legislation efforts seek to criminalize libraries or their workers for having an affiliation with the ALA. E.g., see the recently thwarted HB 777 in Louisiana: <https://legiscan.com/LA/text/HB777/2024>.

between age groups, what qualifies as ban-worthy sexually explicit content versus “appropriate” sexual content), it bears repeating that the abstraction of information removes nuance. This risks objectifying facets of a text that are subjectively selected through a series of processes by data entrants (challengers, their reporters, and library workers), data managers (the ALA), and data processors and analysts (myself) – all of whom might understand the same features and categorical values differently but relay them nonetheless. Thus, I hope that my approach with these data representations – and with my project generally – will be at least as much a critique and warning against the risks that accompany computational analysis as it is a deployment of such methods.

Third, in addition to the aforementioned digital text resources, I had the rare and inaccessible benefit of using a high performance computing system through the Midway 3 cluster at the University of Chicago. Without the aid of this resource-intensive system, it would have been much more difficult if not impossible to build my BERTopic model on a corpus of this size. This presents a fundamental problem for the replicability and accessibility of this thesis project given that only a small proportion of researchers interested in book bans are likely to have the training and access to HPC systems for validating or challenging this research. In adapting this project and my methods going forward, I hope to develop a less-resource intensive and more easily reviewable method for conducting this and similar analyses. In the meantime, my project and its findings must be considered with an additional degree of scrutiny given the additional hurdles presented for peer review and computational verification.

Finally, this project, like much of the current academic research and especially media attention on book bans, neglects the examination of restrictions within the carceral systems of the United States, where book censorship is and has been far more common throughout the rise of

mass incarceration since the last third of the twentieth century.<sup>73</sup> The combined lack of reporting, arbitrary nature of carceral censorship, and sheer ubiquity of book banning and disruptions of text transmission within prisons make this a difficult object of study on any scale. However, the fact that the techniques of censorship and population control deployed in the United States today are often tested on incarcerated peoples first means due attention must be given to the history of censorship and resistance to book banning behind bars. I hope that future expansions of this project incorporate such histories and analysis throughout.

## VII. Conclusions

Research on the trends and effects of book banning in the United States is ever evolving and requires computational approaches to undertake at scale. The problem of book bans is not going away. In keeping with the nature of censorship, the expansion of challenges across the country makes understanding book bans even more difficult. Critics and librarians groaned at the irony that books criticizing bans are now targets themselves, creating a dystopian loop in which criticism of censorship becomes censored in turn. Most notably, the 2017 children’s book *Ban This Book* was removed from the shelves of the Indian River County, Florida school library in May 2024 for “teaching rebellion of [sic] school board authority.”<sup>74</sup> As a personal anecdote, my copy of *Book Banning in 21st Century America* that I cite throughout this project was itself removed from library shelves: its title page bears a stamp reading “Withdrawn from the records of the Mid-Continent Public Library.” Though it is unlikely that my copy was removed as a

---

<sup>73</sup> Moira Marquis and Juliana Luna, “Reading Between the Bars,” PEN America, October 25, 2023: <https://pen.org/report/reading-between-the-bars/>; Hope Corrigan, “Used paperbacks change lives behind bars, even with growing prison book bans,” *The Washington Post*, June 3, 2024: <https://www.washingtonpost.com/lifestyle/2024/05/29/prison-book-ban-through-bars/>.

<sup>74</sup> Douglas Soule, “‘Challenges our authority’: School board in Florida bans book about book bans,” *Tallahassee Democrat*, June 11, 2024: <https://www.tallahassee.com/story/news/politics/2024/06/11/florida-school-board-bans-book-about-book-bans/73970418007/>.

result of a book challenge given the circulation of additional copies within the Mid-Continent library system, this speaks to a central concern when combating book challenge campaigns: how can we understand book bans when the tools for doing so are themselves under attack?

My hope with this project has been to deploy computational approaches to knowledge formation in order to expand upon and bolster the traditional methods of sociological and literary analysis in light of these epistemic attacks – particularly in their focus on queer and BIPOC epistemologies. Topic modeling provides one such method for exploring these texts, both to engage challenger concerns as an approach to distant readings and to compare with my own close reading informed by queer theory and critical cataloging. Given my results, it's possible to reframe the claims of book challengers and their conceptions of what makes these books harmful or distinguishes “queer” from other ways of being, while also underscoring how computational methods can only take us so far without careful validation and comparison with other frameworks. The pipeline for topic modeling and interpretation that I present above can and should be refined further, particularly through experimentation with the parameters and components for modeling allowed for by BERTopic. In recognizing the ongoing nature of this project's development, I hope to expand the study of challenged books and to push back against uses of distant readings – both human and computational – that would try to claim the final word on these books and the queer threads within them.

In considering his approach to intersectionality in his writing, George Johnson offers a mission pertinent for this research. “I use my history as a tool to fight against my marginalization. The greatest tool you have in fighting the oppression of your Blackness and queerness and anything else within your identity is to be fully educated on it.”<sup>75</sup> It is precisely

---

<sup>75</sup> Johnson, *All Boy's Aren't Blue*, 65.

these liberatory tactics that contemporary book banning seeks to erase. Hopefully, projects like this represent a stepping stone in the education process and tool development that assists marginalized people, queer or otherwise, in the journey to survive and thrive.

## Appendix

Full Title	Author(s)	# Cases since 2010	Booklooks/ RatedBooks Rating	Type of Material	Fiction/ NonFiction	Age Group	Publisher	Publication Year	OCLC Tag
Gender Queer: A Memoir	Maia Kobabe	408	4.0	Graphic Novel/ Comic	Non-Fiction	Young Adult	Oni Press	2019.0	LGBTQIA+
All Boys Aren't Blue: A Memoir-Manifesto	George M. Johnson	249	4.0	Book	Non-Fiction	Young Adult	Penguin Books	2021.0	BIPOC/ LGBTQIA+
The Bluest Eye	Toni Morrison	239	4.0	Book	Fiction	Adult	Plume Book	1994.0	BIPOC Fiction
The Perks of Being a Wallflower	Stephen Chbosky	190	4.0	Book	Fiction	Young Adult	Pocket Books	1999.0	LGBTQIA+
The Absolutely True Diary of a Part-Time Indian	Sherman Alexie	181	2.0	Book	Fiction	Young Adult	Little, Brown	2007.0	BIPOC Fiction
It's Perfectly Normal	Robie H. Harris	171	3.0	Book	Non-Fiction	Children	Candlewick Press	1994.0	NaN
Looking for Alaska	John Green	165	3.0	Book	Fiction	Young Adult	Dutton Books	2005.0	NaN
This Book is Gay	Juno Dawson	161	4.0	Book	Non-Fiction	Young Adult	Sourcebooks	2015.0	LGBTQIA+
Flamer	Mike Curato	158	3.0	Graphic Novel/ Comic	Fiction	Young Adult	St. Martin's Press	2020.0	LGBTQIA+
Lawn Boy	Jonathan Evison	155	3.0	Book	Fiction	Young Adult	Algonquin Books of Chapel Hill	2018.0	BIPOC Fiction

Figure 3-1a

Full Title	Author(s)	# Cases since 2010	Booklooks/ RatedBooks Rating	Type of Material	Fiction/ NonFiction	Age Group	Publisher	Publication Year	OCLC Tag
All Boys Aren't Blue: A Memoir-Manifesto	George M. Johnson	249	4.0	Book	Non-Fiction	Young Adult	Penguin Books	2021.0	BIPOC/ LGBTQIA+
The Bluest Eye	Toni Morrison	239	4.0	Book	Fiction	Adult	Plume Book	1994.0	BIPOC Fiction
The Perks of Being a Wallflower	Stephen Chbosky	190	4.0	Book	Fiction	Young Adult	Pocket Books	1999.0	LGBTQIA+
The Absolutely True Diary of a Part-Time Indian	Sherman Alexie	181	2.0	Book	Fiction	Young Adult	Little, Brown	2007.0	BIPOC Fiction
Looking for Alaska	John Green	165	3.0	Book	Fiction	Young Adult	Dutton Books	2005.0	NaN
This Book is Gay	Juno Dawson	161	4.0	Book	Non-Fiction	Young Adult	Sourcebooks	2015.0	LGBTQIA+
Me and Earl and the Dying Girl	Jesse Andrews	138	3.0	Book	Fiction	Young Adult	Amulet Books	2012.0	NaN
The Hate U Give	Angie Thomas	131	3.0	Book	Fiction	Young Adult	HarperCollins	2017.0	BIPOC Fiction
Out of Darkness	Ashley Hope Pérez	130	3.0	Book	Fiction	Young Adult	Lerner Publishing Group	2015.0	BIPOC Fiction
I Know Why the Caged Bird Sings	Maya Angelou	127	3.0	Book	Non-Fiction	Adult	Random House	1969.0	BIPOC Biography

Figure 3-1b

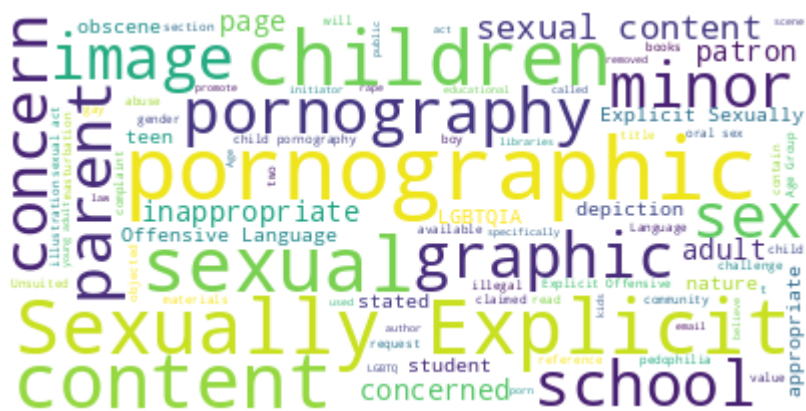


Figure 3-2

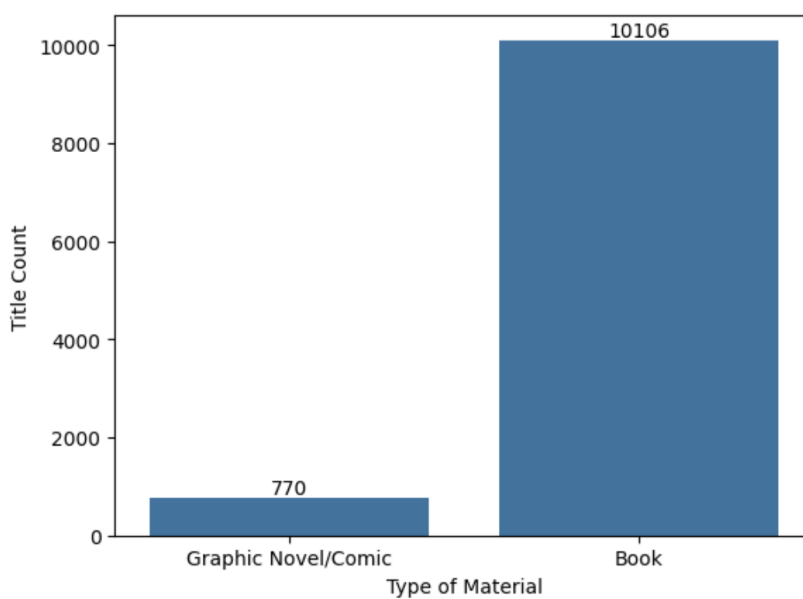


Figure 3-3

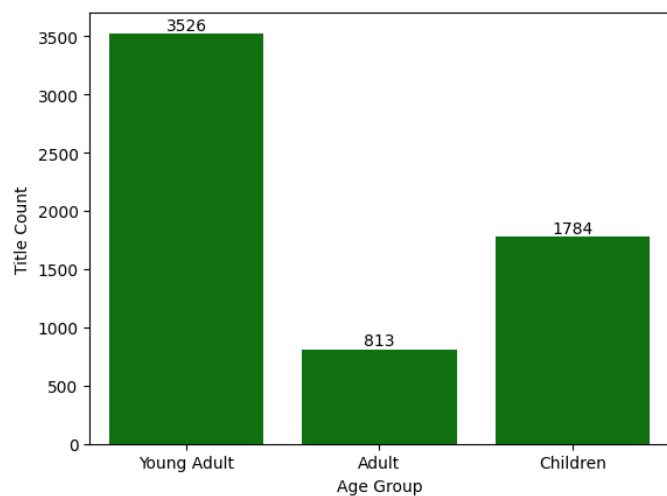


Figure 3-4a (all titles)

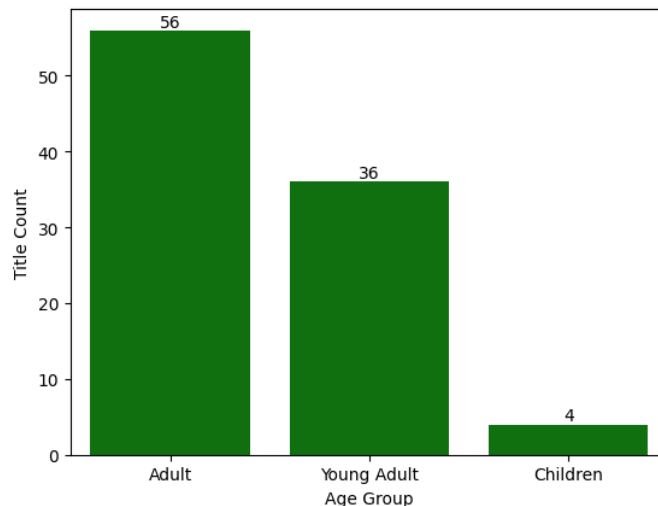


Figure 3-4b (corpus only)

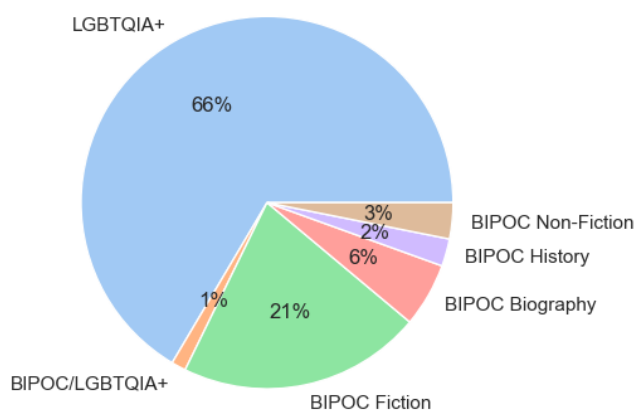


Figure 3-5a (All titles)

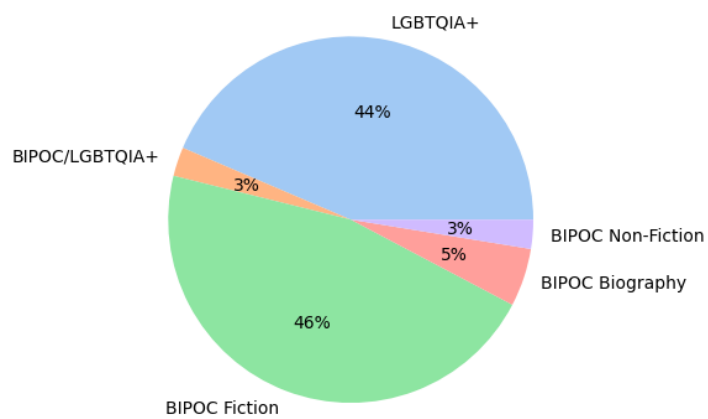


Figure 3-5b (Corpus only)

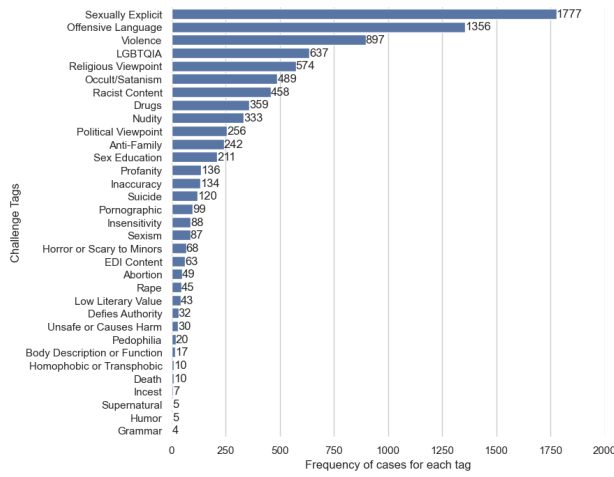


Figure 3-6a (All titles)

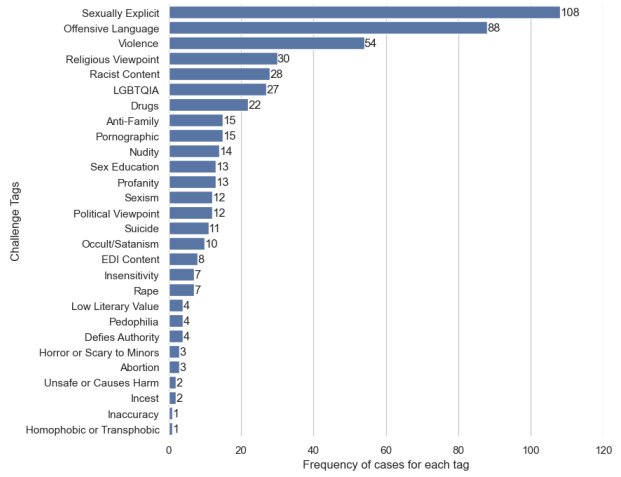


Figure 3-6b (Corpus only)

Topic Word Scores



Figure 5-1

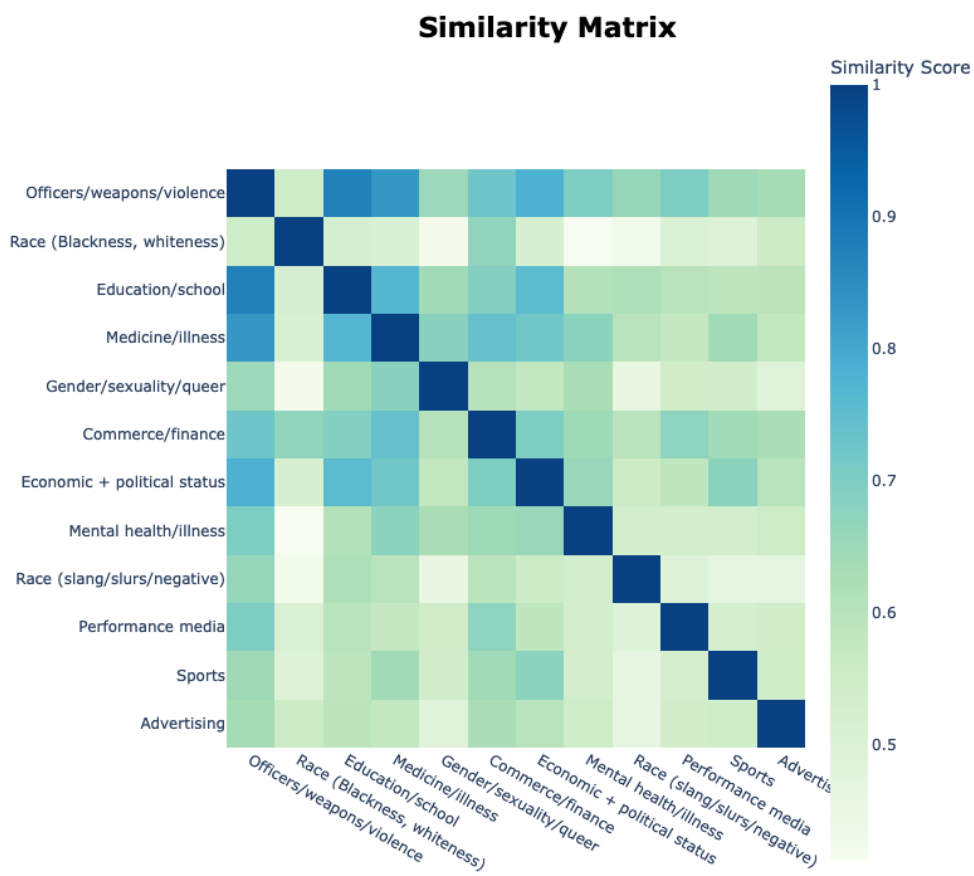


Figure 5-2

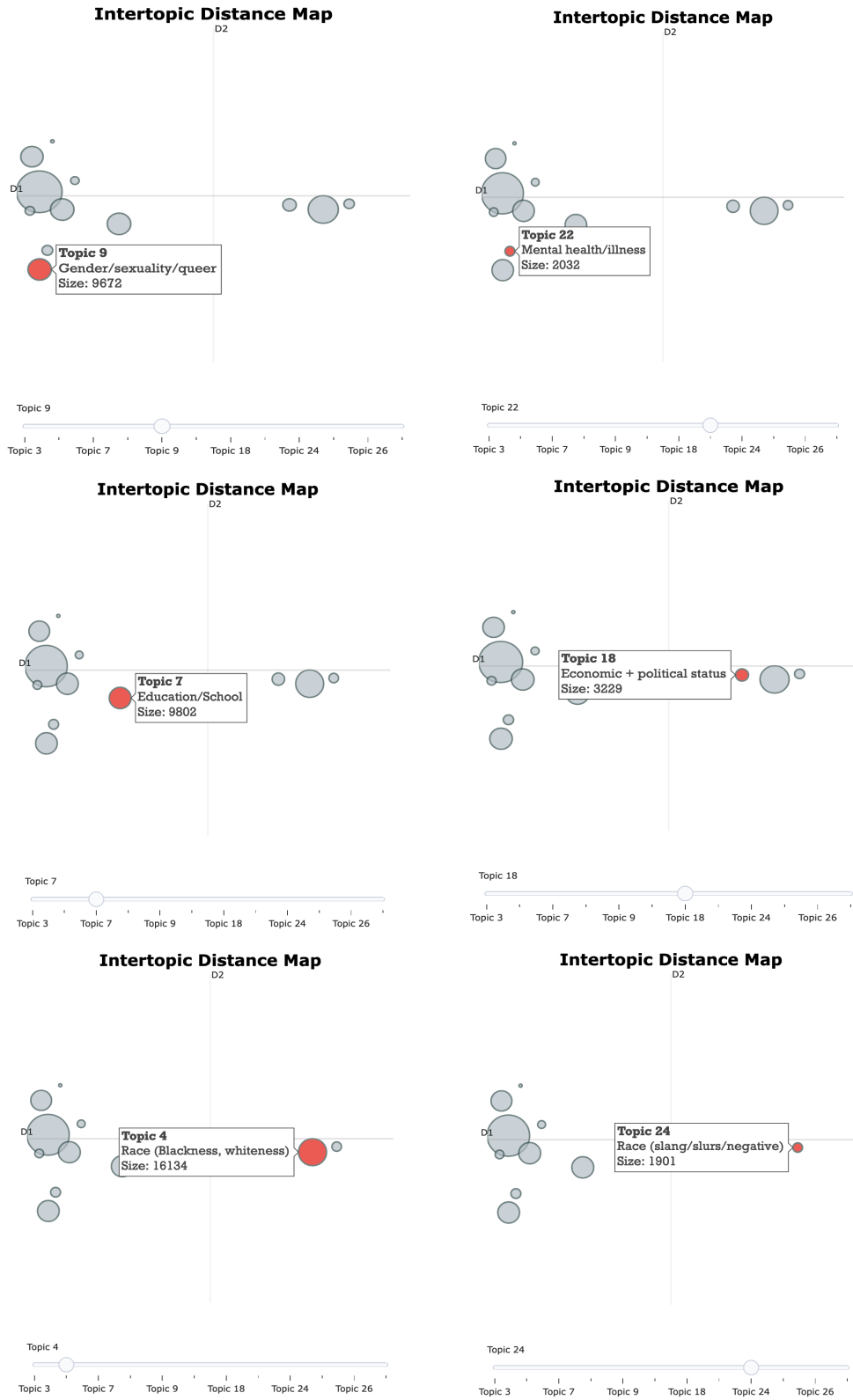


Figure 5-3

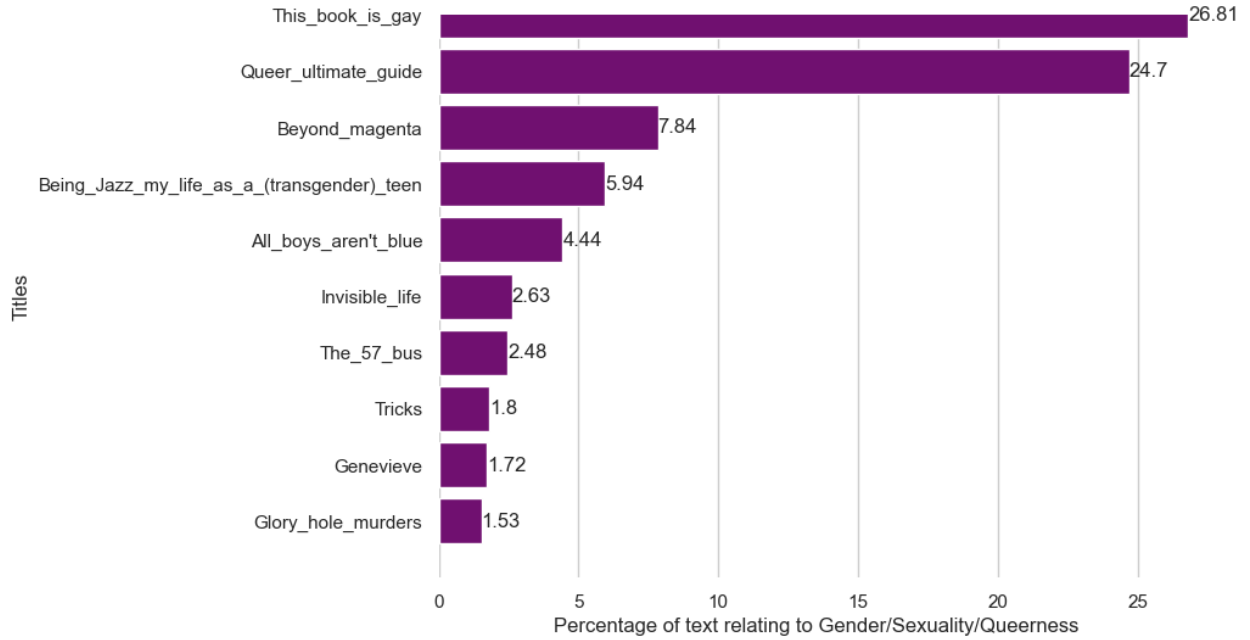


Figure 5-4

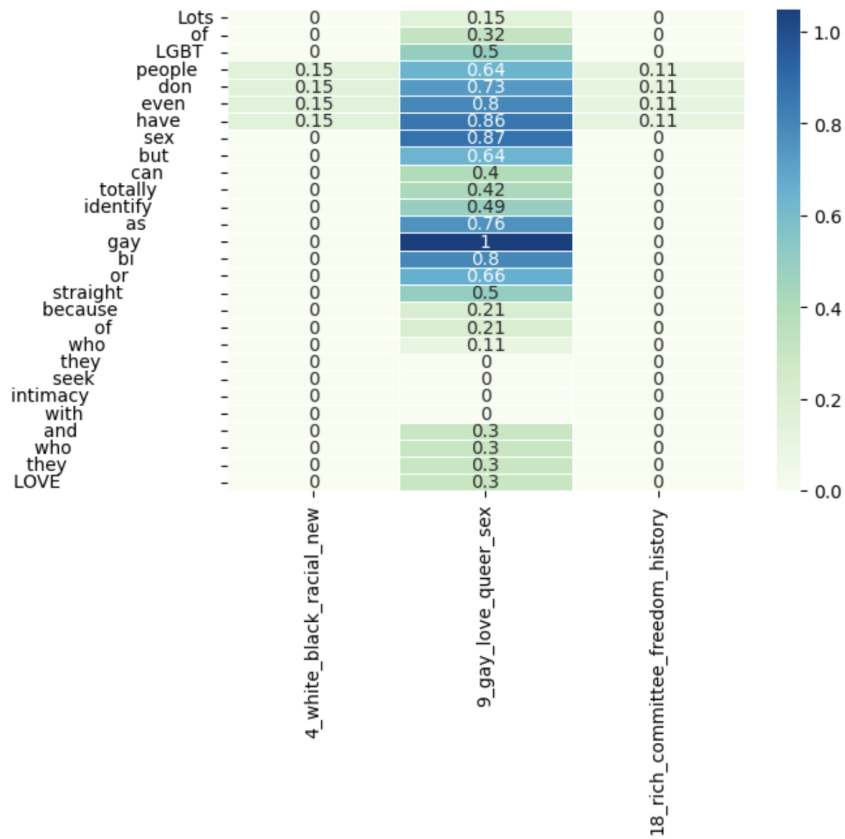


Figure 5-5

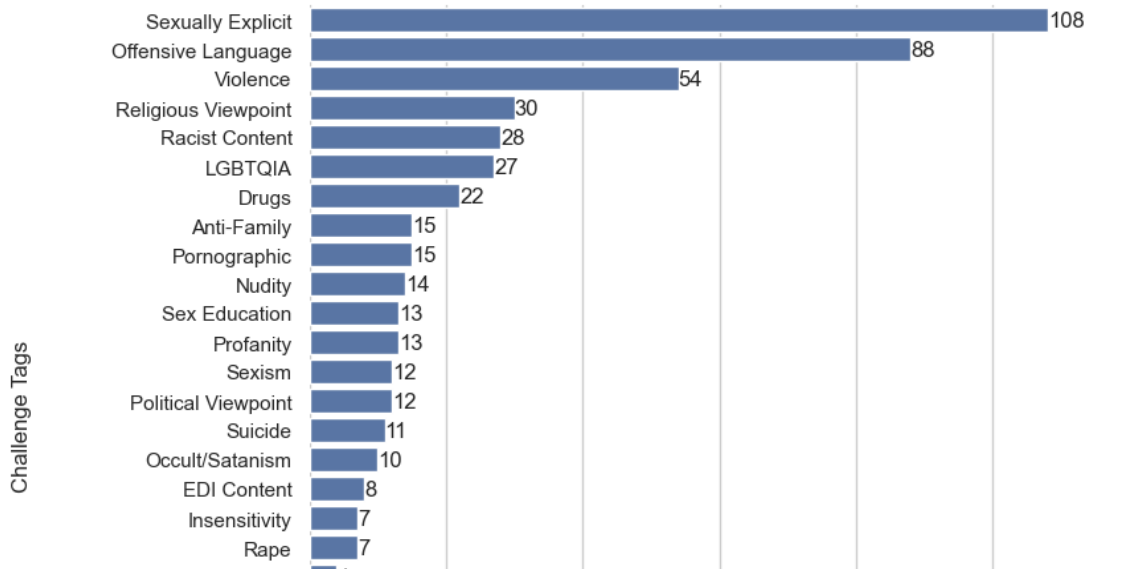


Figure 5-6a (Challenge Tags)

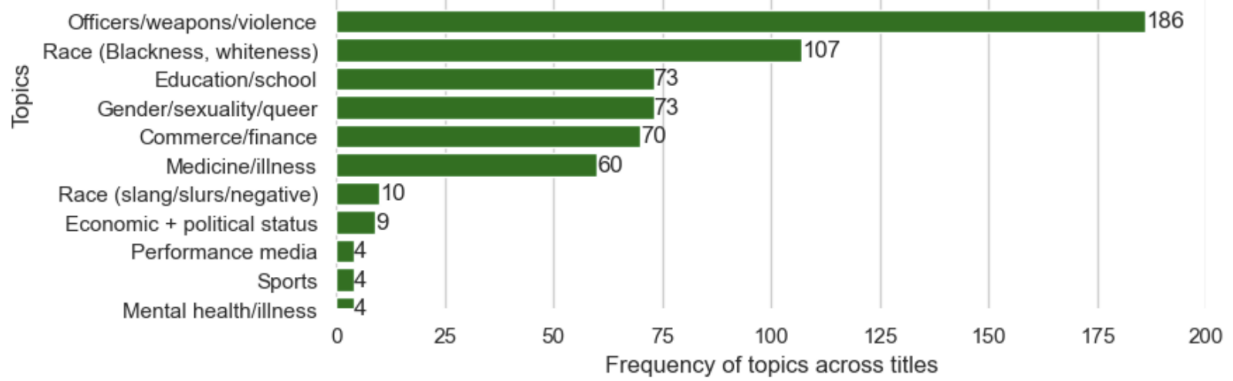


Figure 5-6b (Model Topics)

<b>Title</b>	<b>% “Gender/sexuality/queer”</b>	<b>Challenge Count and Percentile</b>	<b>1st Topic ----- 1st Challenge Tag</b>	<b>2nd Topic ----- 2nd Challenge Tag</b>	<b>3rd Topic ----- 3rd Challenge Tag</b>
<i>This Book is Gay</i>	26.81%	161 (Top 2.5%)	Gender/sexuality/queer	Medicine/illness	Race (Blackness, whiteness)
			LGBTQIA	Sexually Explicit	Sex Education
<i>Queer: The Ultimate LGBTQ Guide for Teens</i>	24.7%	26 (Top 16.5%)	Gender/sexuality/queer	Education/school	Economic + political status
			LGBTQIA	Political Viewpoint; Religious Viewpoint; Sex Education	
<i>Beyond Magenta: Transgender and Nonbinary Teens Speak Out</i>	7.84%	123 (Top 6.5%)	Gender/sexuality/queer	Medicine/illness	Education/school
			LGBTQIA	Sexually Explicit	Anti-Family; EDI Content; Offensive Language; Political Viewpoint; Religious Viewpoint
<i>Being Jazz: My Life as a (Transgender) Teen</i>	5.94%	37 (Top 14.5%)	Gender/sexuality/queer	Education/school	Sports
			LGBTQIA	Religious Viewpoint	
<i>All Boys Aren't Blue: A Memoir-Manifesto</i>	4.44%	249 (Most Challenged)	Race (Blackness, whiteness)	Gender/sexuality/queer	Education/school
			Sexually Explicit	LGBTQIA	EDI Content; Incest; Rape
<i>Invisible Life</i>	2.63%	3 (Top 46.5%)	Gender/sexuality/queer	Race (Blackness, whiteness)	State officials, weapons, violence
			LGBTQIA		

Figure 5-7

## Bibliography

- Adler, Melissa. *Cruising the Library*. New York: Fordham University Press, 2017.
- American Library Association. "Get Involved." January 3, 2013. Accessed July 18, 2024. <https://www.ala.org/bbooks/bannedbooksweek/ideasandresources/activity>.
- American Library Association. "Surge in Book Challenges Press Kit." November 22, 2021. Accessed July 18, 2024. <https://www.ala.org/news/mediapresscenter/presskits/surge-book-challenges-press-kit>.
- American Library Association. "Voters Oppose Book Bans in Libraries." March 21, 2022. Accessed July 18, 2024. <https://www.ala.org/advocacy/voters-oppose-book-bans-libraries>.
- American Library Association. "Challenge Reporting." February 24, 2023. Accessed July 18, 2024. <https://www.ala.org/tools/challengesupport/report>.
- American Library Association, Office for Intellectual Freedom. "Censorship by the Numbers." March 4, 2024. Accessed July 18, 2024. <https://www.ala.org/bbooks/censorship-numbers>.
- Antoniak, Maria. "little-mallet-wrapper." Accessed July 18, 2024. <https://github.com/maria-antoniak/little-mallet-wrapper>.
- Block, Melissa. "Accusations of 'grooming' are the latest political attack – with homophobic origins." *NPR*, May 11, 2022. Accessed July 18, 2024. <https://www.npr.org/2022/05/11/1096623939/accusations-grooming-political-attack-homophobic-origins>.
- Bryant, Anita. *The Anita Bryant Story*. Old Tappen: 1977.
- Chen, Angela. *Ace: What Asexuality Reveals About Desire, Society, and the Meaning of Sex*. Boston: Beacon Press, 2020.
- Collins, Patricia. "Intersectionality's Definitional Dilemmas." *Annual Review of Sociology* 41, (2015): 1-20.
- Corrigan, Hope. "Used paperbacks change lives behind bars, even with growing prison book bans." *The Washington Post*, June 3, 2024. Accessed July 18, 2024. <https://www.washingtonpost.com/lifestyle/2024/05/29/prison-book-ban-through-bars/>.
- Cravens, R.G. "Anti-Trans Rhetoric Helps the Right Pivot to Gun Violence Myths, Sell Guns." *Hatewatch*, The Southern Poverty Law Center, March 24, 2024. Accessed July 18, 2024. <https://www.splcenter.org/hatewatch/2024/03/22/anti-trans-rhetoric-helps-right-pivot-gun-violence-myths-sell-guns>.
- Crenshaw, Kimberle. "Demarginalizing the Intersection of Race and Sex: A Black Feminist Critique of Antidiscrimination Doctrine, Feminist Theory and Antiracist Politics." *The University of Chicago Legal Forum*, (1989): 139-167.

Drabinski, Emily. "Queering the Catalog: Queer Theory and the Politics of Correction." *The Library Quarterly* 83 (2), April 2023: 94-111.

Feldman, Ella. "Book Banning Attempts Are at Record Highs." *Smithsonian Magazine*, March 18, 2024. Accessed July 18, 2024. <https://www.smithsonianmag.com/smart-news/book-banning-attempts-are-at-record-highs-180983964/>.

Fotopoulou, Aristeia. "Intersectionality of Queer Studies and Hybridity: Methodological Frameworks for Social Research." *Journal of International Women's Studies* 13, no.2 (2012): 19-32.

Foucault, Michel. *The History of Sexuality Volume 1: An Introduction*. New York: Random House, 1978.

Frank, Gillian. "'The Civil Rights of Parents': Race and Conservative Politics in Anita Bryant's Campaign against Gay Rights in 1970s Florida." *Journal of the History of Sexuality* 22, no. 1 (January 2013): 126-160.

Gabriel, Trip. "He Fuels the Right's Cultural Fires (and Spreads Them to Florida)," *The New York Times*, April 24, 2022.

Garcia, Raymond. "American Library Association reports record number of unique book titles challenged in 2023," American Library Association News, March 14, 2024. Accessed July 18, 2024. <https://www.ala.org/news/2024/03/american-library-association-reports-record-number-unique-book-titles>.

"Gender Resource Guide." Minnesota Family Council, 2019. Accessed July 18, 2024. <https://genderresourceguide.com/wp-content/themes/genderresource/library/documents/NPRGFullDocumentPrintV17.pdf>.

Goldstone, Andrew and Ted Underwood. "What Can Topic Models of PMLA Teach Us About the History of Literary Scholarship?" *Journal of Digital Humanities* 2, no. 1 (Winter 2012). Accessed July 18, 2024. <https://journalofdigitalhumanities.org/2-1/what-can-topic-models-of-pmla-teach-us-by-ted-underwood-and-andrew-goldstone/>.

Goncalves, Marcelo S.O., Isabelle Langrock, Jack LaViolette, and Katie Spoon, "Book bans in political context: Evidence from US schools", *PNAS Nexus* 3, no. 6 (June 2024): 1-37.

Grootendorst, Maarten. "The Algorithm." *BERTopic*. Accessed July 18, 2024. <https://maartengr.github.io/BERTopic/algorithm/algorithm.html#code-overview>.

Grootendorst, Maarten. "BERTopic: Neural topic modeling with a class-based TF-IDF procedure." *arXiv preprint*, 2022. Accessed July 18, 2024. <https://arxiv.org/pdf/2203.05794>

Grootendorst, Maarten. "get\_document\_info()." *BERTopic*. Accessed July 18, 2024. [https://maartengr.github.io/BERTopic/api/bertopic.html#bertopic.\\_bertopic.BERTopic.get\\_document\\_info](https://maartengr.github.io/BERTopic/api/bertopic.html#bertopic._bertopic.BERTopic.get_document_info).

Harris, Elizabeth A. and Alexandra Alter. "A Fast-Growing Network of Conservative Groups Is Fueling a Surge in Book Bans." *The New York Times*, updated January 10, 2023. Accessed July 18, 2024. <https://www.nytimes.com/2022/12/12/books/book-bans-libraries.html>.

The Homosaurus. "About." Accessed July 18, 2024. <https://homosaurus.org/about>.

Ingram, David. "After Libs of TikTok posted, at least 21 bomb threats followed." *NBC News*, February 7, 2024. Accessed July 18, 2024. <https://www.nbcnews.com/tech/internet/lib-tiktok-x-chaya-raichik-bomb-threat-twitter-of-libsoftiktok-rcna102784>.

Jockers, Matthew. *Macroanalysis: Digital Methods and Literary History*. Champaign: University of Illinois Press, 2013. <https://www.matthewjockers.net/macroanalysisbook/>.

Jockers, Matthew. "Expanded Stopwords List." Accessed July 18, 2024. <https://www.matthewjockers.net/macroanalysisbook/expanded-stopwords-list/>.

Johnson, George M. *All Boy's Aren't Blue: A Memoir-Manifesto*. New York: Farrar, Straus and Giroux, 2020.

Knox, Emily. "Testimony before the U.S. Senate Committee on the Judiciary; Hearing on 'Book Bans: Examining How Censorship Limits Liberty and Literature.'" September 12, 2023. Accessed on July 18, 2024. [https://www.judiciary.senate.gov/imo/media/doc/2023-09-12\\_-\\_testimony\\_-\\_knox.pdf](https://www.judiciary.senate.gov/imo/media/doc/2023-09-12_-_testimony_-_knox.pdf).

Knox, Emily. *Foundations of Intellectual Freedom*. Chicago: ALA Neal-Schuman, 2023.

Knox, Emily. "Silencing Stories: Challenges to Diverse Books." *The International Journal of Information, Diversity, & Inclusion* 3, no. 2 (April 2019): 24-39. <https://doi.org/10.33137/ijidi.v3i2.32592>.

Knox, Emily. *Book Banning in 21st-Century America*. Lanham: Rowman & Littlefield, 2015.

Kozlowski, Austin C., Matt Taddy, and James A. Evans. "The geometry of culture: Analyzing the meanings of class through word embeddings." *American Sociological Review* 84, no. 5 (2019): 769-981.

Lavietes, Matt. "Libs of TikTok creator accused of inspiring school bomb threats named to Oklahoma library board." *NBC News*, January 23, 2024. Accessed July 18, 2024. <https://www.nbcnews.com/nbc-out/out-news/lib-tik-tok-bomb-threats-oklahoma-library-committee-rcna135369>.

- Marquis, Moira and Juliana Luna, "Reading Between the Bars." PEN America, October 25, 2023. Accessed July 18, 2024. <https://pen.org/report/reading-between-the-bars/>.
- McCall, Leslie. "The Complexity of Intersectionality." *Signs* 30, no. 3 (Spring 2005): 1771-1800.
- McCallum, Andrew Kachites. "MALLET: A MACHine Learning for LanguagE Toolkit." 2002. Accessed July 18, 2024. <https://mimno.github.io/Mallet/index>.
- Meehan, Kasey and Jonathan Friedman. "Banned in the USA: State Laws Supercharge Book Suppression in Schools." PEN America, April 20, 2023. Accessed July 18, 2024. <https://pen.org/report/banned-in-the-usa-state-laws-supercharge-book-suppression-in-schools/>.
- Mandell, Laura. "Gender and Cultural Analytics: Finding or Making Stereotypes." *Debates in the Digital Humanities 2019*. Minneapolis: University of Minnesota Press, 2019: 3-26.
- Mimno, David. "Topic Modeling." Accessed July 18, 2024. <https://mimno.github.io/Mallet/topics>.
- Moretti, Franco. *Distant Reading*. London: Verso, 2013.
- Muñoz, José Esteban. *Cruising Utopia*, New York: NYU Press, 2009.
- Nelson, Laura K. "Situated Knowledge and Partial Perspectives: A Framework for Radical Objectivity in Computational Social Science and Computational Humanities." *New Literary History* 54, no. 1 (Winter 2023): 853-877.
- Nguyen, Dong et. al., "How we do things with words: Analyzing text as social and cultural data." *Frontiers in Artificial Intelligence* 3, article 62 (August 2020).
- Otten, Tori. "Moms for Liberty Completely Collapses in Former Strongholds." *The New Republic*, February 8, 2024. Accessed July 18, 2024. <https://newrepublic.com/post/178879/moms-liberty-chapters-collapse-strongholds-pennsylvania-florida>.
- PEN America. "Ed Scare: Frequently Asked Questions." Accessed July 18, 2024. <https://pen.org/ed-scare-faq/>.
- "Provides relative to the use of public funds." Louisiana HB 777, May 2024. Accessed July 18, 2024. <https://legiscan.com/LA/text/HB777/2024>.
- Rufo, Christopher. "The Real Story Behind Drag Queen Story Hour." *City Journal* (Autumn 2022). Accessed July 18, 2024. <https://www.city-journal.org/article/the-real-story-behind-drag-queen-story-hour>.

Sedgwick, Eve Kosofsky. "Queer and Now." *Tendencies*. Durham: Duke University Press, 1993.

Sedgwick, Eve Kosofsky. *Epistemology of the Closet*. Berkeley: University of California Press, 1990.

So, Richard. "All Models Are Wrong." *PMLA/Publications of the Modern Language Association of America* 132, no. 3 (May 2017): 668-673.

Sommerfeldt, Chris. "NYC libraries would end universal six-day service under Adams budget plan: testimony." *New York Daily News*, March 12, 2024. Accessed July 18, 2024. <https://www.nydailynews.com/2024/03/12/most-nyc-libraries-would-have-to-end-weekend-hours-entirely-under-adams-budget-plan-testimony/>.

Soule, Douglas. "'Challenges our authority': School board in Florida bans book about book bans." *Tallahassee Democrat*, June 11, 2024. Accessed July 18, 2024. <https://www.tallahassee.com/story/news/politics/2024/06/11/florida-school-board-bans-book-about-book-bans/73970418007/>.

Textual Optics Lab. "About Us." Accessed July 18, 2024. <https://textual-optics-lab.uchicago.edu/about>.

The Trevor Project. "2023 U.S. National Survey on the Mental Health of LGBTQ Young People." Accessed July 18, 2024. [https://www.thetrevorproject.org/survey-2023/assets/static/05\\_TREVOR05\\_2023survey.pdf](https://www.thetrevorproject.org/survey-2023/assets/static/05_TREVOR05_2023survey.pdf).

Waldroupe, Amanda. "Violent, aggressive, and personal attacks — language used to challenge books has changed dramatically, research presented at Oregon Library Association conference shows." *Oregon Arts Watch*, April 30, 2024. Accessed July 18, 2024. <https://www.orartswatch.org/violent-aggressive-and-personal-attacks-language-used-to-challenge-books-has-changed-dramatically-research-presented-at-oregon-library-association-conference-shows/>.

Warner, Matt. "A Queer Way of Counting: Bibliography and Computational Approaches to the Queer Novel." *New Literary History* 54, no. 1 (Winter 2023): 939-966.

Watercutter, Angela. "How an Iowa School District Used ChatGPT to Ban Books." *Wired*, August 18, 2023. Accessed July 18, 2024. <https://www.wired.com/story/chatgpt-ban-books-iowa-schools-sf-496/>.

Wiggins, Christopher. "Fox News and GOP politicians falsely claim Lakewood Church shooter was transgender." *Advocate*, February 12, 2024. Accessed July 18, 2024. <https://www.advocate.com/news/false-transgender-claim-texas-shooter>.

Wiggins, Christopher. "Moms for Liberty Candidates Rejected by Voters in Key School Board Elections." *Advocate*, November 8, 2023. Accessed July 18, 2024.  
<https://www.advocate.com/politics/moms-for-liberty-rejected-election>.