Quantifying the origins and impacts of book bans in U.S. public schools, 2021 - 2022

Keywords: book bans, politics, education, censorship, polarization

Extended Abstract

In the 2021-2022 school year, more books were banned in U.S. school districts than in any previous year [1]. Book banning, like other forms of censorship, can have serious consequences for public expression [2], and is an especially relevant concern in the current polarized U.S. political climate. While the media has been quick to cover instances of book bans [see 3, 4 for recent examples], there is little empirical research on book bans and the impact of these censorship attempts. In particular, it is unknown if banning a book effectively decreases interest in the book, or if there is a backlash to the ban, in which censorship increases interest in the book, author, or theme - the so-called "Streisand Effect" [5, 6]. Here, we describe the results of a unique large-scale, systematic study of book bans in the U.S., in which we leverage a dataset of 2,532 book bans that occurred across the U.S. in the 2021-2022 school year [7]. We reveal trends in the school districts that ban books, the types and themes of books, demographic information of the authors, and the response to bans across several different interest indicators. Our study yields a more complete understanding of the origins and impacts of book banning in the U.S., an increasingly salient political issue and a growing form of censorship for our digitally mediated world.

We use the PEN America's Index of School Book Bans [7] to study bans at three levels: (i) the book, including major themes, genres, and author demographics; (ii) the district, particularly the political and socioeconomic environments that produce bans, and (iii) public interest in banned books. We build a large database of information about each banned book and banning district that combines information about each book from Goodreads, Google Books, and public biographies of authors, with district level information from U.S. school district meta-data [8], U.S. Census data [9], and election returns [10], and, finally, public interest indicators from Bookshop.com book sales, Google Trends search data, and Seattle Library checkout data.

To begin with, we provide a descriptive overview of the corpus of banned books across themes and genres. This is achieved in two complementary ways: a topic model of book descriptions from Google Books, and a clustering procedure based on a feature space derived from Goodreads book genres. The induced clusters allow us to identify high-level categories of books which inform subsequent analyses. Broadly, these are: 1) Young adult queer romance novels in contemporary, realistic settings; 2) Fantasy/Sci-fi with frequent queer themes; 3) Age-diverse fiction focused on gender and sexuality, with greater emphasis on women than on queerness; 4) Non-fiction books about social movements, cultural issues, and important related historical figures; 5) Fiction for older readers with more mature themes, but not queer- or romance-centric; and 6) Children's literature with culturally diverse characters. **Fig. 1A** displays the relationship of each cluster with the four most prevalent book banning states.

Secondly, to evaluate the relationship between book bans and various social and political factors, we link each school district to its respective county and examine the evolution of demographic and political data in these localities [10]. Notably, bans take place in counties where the Republican Party faces increasingly more competitive elections, measured in terms of voting share in presidential elections. Compared to counties in the same

commuting zone, the counties with bans are electoral battlegrounds (**Fig. 1B**). This suggests that book bans might be part of a political strategy to drive attention and support to Republicans in increasingly competitive elections. Observable demographic changes in the last two decades are similar for both groups of counties, suggesting that the likelihood of a ban is not a function of socioeconomic factors but political ones.

Finally, we evaluate the impact that book bans have on the public interest in the banned books. We are particularly interested in whether bans have either (i) a chilling effect, where censorship decreases interest, or (ii) a notable backlash, where interest in the book increases. Because book sales data are heavily embargoed by the publishing industry [11], we use three proxies for national interest to identify traces of either effect: (1) Bookshop.com, an online retailer of books, sales data, (2) Google Trends search data; and, (3) Seattle Public Library Checkout data (which is available through the city's open data effort). Across all indicators we see no change in interest across the 3 month period directly preceding and following the ban (**Fig. 1C, 1D**).

Book bans are an increasingly common political tactic in the U.S. Our large-scale study identifies that bans are frequently targeted at diverse children's books and nonfiction accounts of social movements. Additionally, we identify that increasingly competitive electoral races are the greatest predictor that a district will ban books, across many political and socioeconomic factors. Despite the growing prevalence of book bans, they seem to have little impact on public interest in the books, with interest in the banned books remaining similarly low pre- and post-ban. Taken together, these results suggest that book bans are a political stunt to capture attention around salient issues (like Critical Race Theory and the politics of sex and gender). Our study presents one of the first substantial analyses of book bans and the consequences on the larger national information environment.

References

[1] Gabbatt, A. (2022 , April 7). 'Unparalleled in intensity' – 1,500 book bans in US school districts. *The Guardian*. https://www.theguardian.com/us-news/2022/apr/07/book-bans-pen-america-school-districts
[2] Roberts, M. E. (2018). *Censored: Distraction and diversion inside China's great firewall.* Princeton University Press.
[3] Harris, E. A. & Alter, A. (2022, July 6). With rising book bans, librarians have come under attack. *The New York Times.* https://www.nytimes.com/2022/07/06/books/book-ban-librarians.html
[4] Natanzan, H. & Rozsa, L. (2022, August 17). Students lose access to books amid 'state-sponsored purging of ideas'. *The Washington Post.* https://www.washingtonpost.com/education/2022/08/17/book-ban-restriction-access-lgbtq/
[5] Jansen, S. C. & Martin, B. (2015). The Streisand effect and censorship backfire. *International Journal of Communication*, *9*(0), 656-671.
[6] Hobbs, W. R. & Roberts, M. E. (2018). How sudden censorship can increase access to information. *American Political Science Review*, *112*(3), 621-636.
[7] PEN America. (2022). *PEN America's Index of School Book Bans (July 1, 2021 - June 30, 2022)* [Data set]. https://pen.org/banned-in-the-usa/
[8] National Center for Education Statistics. (2000-2020). *American Community Survey – Education Tabulation (ACS-ED)* [Data set]. https://nces.ed.gov/programs/edge/Demographic/ACS
[9] United States Census Bureau. (2000-2020). *Small Area Income and Poverty Estimates (SAIPE) Program* [Data set]. https://www.census.gov/data/datasets.html
[10] MIT Election Data and Science Lab. (2018). *County Presidential Election Returns 2000-2020* (Version 11) [Data set]. https://doi.org/10.7910/DVN/VOQCHQ
[11] Walsh, M. (2022). Where is all the book data? *Public Books*. https://www.publicbooks.org/where-is-all-the-book-data/

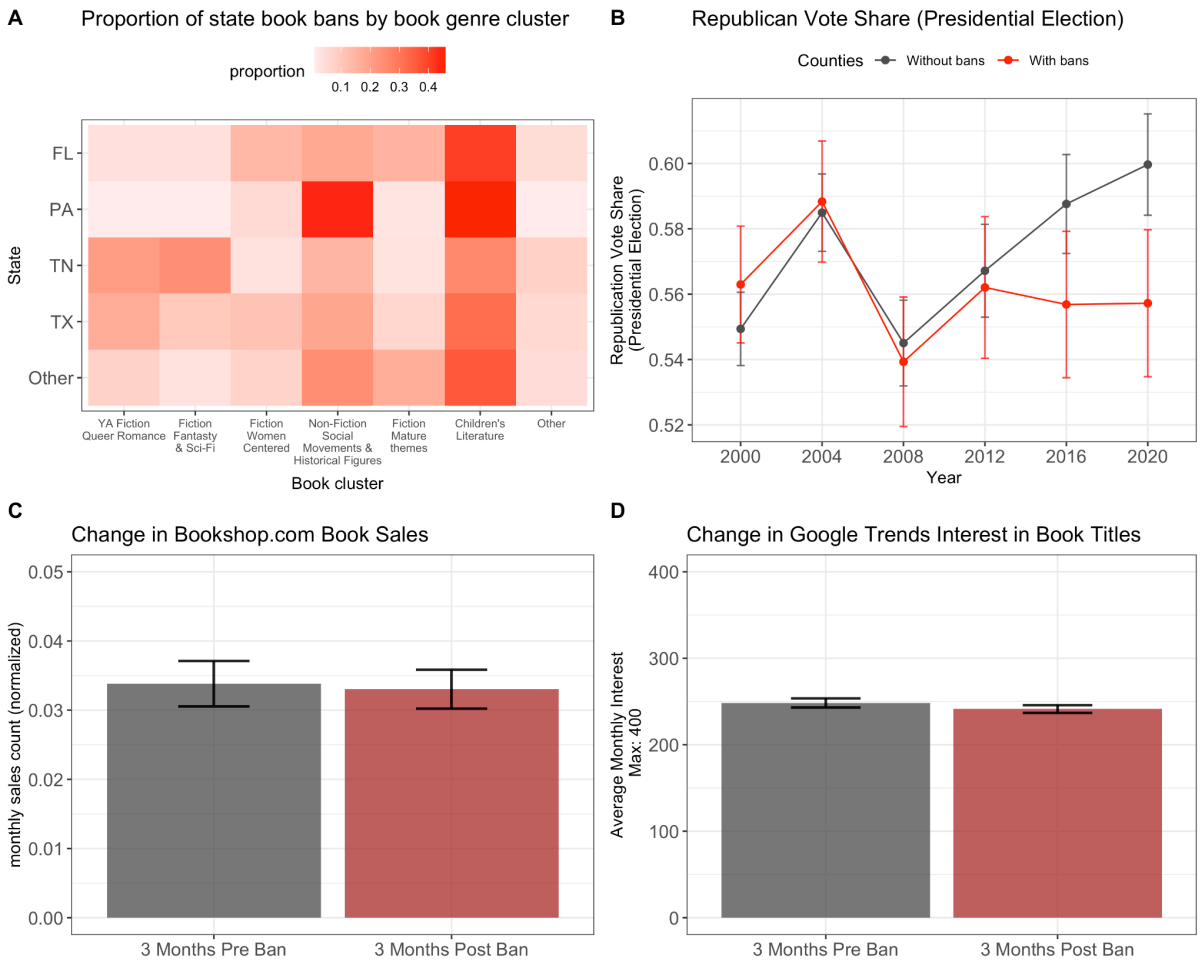

Figure 1. (**A**) The fraction of books banned within each state, split by book genre cluster. Rows sum to 1. Florida, Pennsylvania, Tennessee, and Texas are presented alphabetically and broken out from the remaining states, as together they represent over 85% of all book bans in the data. Six book clusters were induced from the individual genres assigned to each book by Goodreads users, with some books falling in multiple genres. Treating these genre designations as a feature space, books were clustered according to their genres with a combination of the HDBSCAN and UMAP algorithms. (**B**) The Republican vote share in presidential elections from 2000-2020, for counties that banned books in 2021-2022 vs. counties in the same commuting zone that did not ban books. The counties most likely to ban books are increasingly more competitive electoral districts than counties in the same commuting zone. Finally, we plot the null change in interest after books were banned across two interest indicators: (**C**) book sales via Bookshop.com and (**D**) Google searches of book titles from Google Trends. Book sales are normalized for each book to maintain privacy of Bookshop's data, and Google Trends data are only available normalized across the ban period. Interest remains low for most books in our sample. Figures include 95% confidence intervals.