Journal of Open Humanities Data

Seattle Public Library's Open Checkout Data: What Can It Tell Us About Readers and Book Popularity More Broadly?

DISCUSSION PAPER

NEEL GUPTA
DAVID CHRISTENSEN
MELANIE WALSH

*Author affiliations can be found in the back matter of this article

ubiquity press

ABSTRACT

The Seattle Public Library (SPL) publishes anonymized, open-access checkout data for every item in its collection, dating from 2005 to the present. To our knowledge, it is the only U.S. library to release checkout data by title with this level of temporal detail: one dataset records exact timestamps for print book checkouts, while another provides monthly aggregates across all formats (e.g., ebooks, audiobooks, print books). Because U.S. book sales data is largely inaccessible outside the publishing industry, SPL's open checkout data offers a rare and valuable alternative. But how well does it generalize beyond Seattle? Does it reflect book sales? And what can it tell us about readers more broadly?

In this paper, we introduce SPL's checkout data and evaluate its potential for humanistic and literary research. We specifically assess how well it: (1) corresponds with book sales, (2) and extrapolates to library checkout patterns elsewhere nationally. First, we compare SPL data against publishing revenue reports and prior research with access to sales figures. We find that SPL patrons embrace digital books more than general consumers, but the overall distribution of checkouts resembles broader book sales patterns. Second, we compare SPL's most checked out books per year to the New York Public Library's annual top 10 lists. We find general overlap, but also distinct regional preferences, suggesting geographic extrapolation should be approached with caution. We conclude that SPL's checkout data provides a rare window into library circulation and reading habits, with granular, time-series insights. However, its generalizability—particularly across regions and in relation to book sales—remains uncertain.

CORRESPONDING AUTHOR:
Neel Gupta
Information School, University of Washington, Seattle, USA
ngupta1@uw.edu

KEYWORDS:
readers; reception; libraries; literature; reader-response; circulation; popularity

TO CITE THIS ARTICLE:
Gupta, N., Christensen, D., & Walsh, M. (2025). Seattle Public Library's Open Checkout Data: What Can It Tell Us About Readers and Book Popularity More Broadly? *Journal of Open Humanities Data*, 11: 46, pp. 1–19. DOI: https://doi.org/10.5334/johd.332

DOI: 10.5334/johd.332

1 CONTEXT AND MOTIVATION

What books are people reading today versus a decade ago? How popular is this specific book, or that one? Answering these questions can help us track trends in reading, literature, and the publishing industry, as well as in culture, politics, and society more broadly. With the rise of the internet and widespread datafication, scholarly interest in these questions has only grown because more reading and reception data is now available (Price Lab for Digital Humanities, 2023)—though, crucially, it is not always *publicly* available.

Reading, typically a private act, is notoriously difficult to measure. But many proxy measurements for reading now exist. Some of these proxies are publicly available, like the number of online reviews published for a book, and some are private and proprietary, like the number of Kindle pages users have turned. Unfortunately, the most comprehensive, specific, and temporally sensitive book reception data remains proprietary and inaccessible to both the public and researchers.[1] Book sales data in the United States, for example, is owned by Circana's BookScan, and is currently inaccessible (even for pay) to individuals outside of the publishing industry (Walsh, 2022).

As a result, digital humanists have used publicly available sources like the *New York Times* bestseller list (Underwood, 2019) or Goodreads ratings (Porter, 2018) to model book popularity. These data sources are valuable, but they often lack measurement granularity or temporal specificity. To generate more powerful hypotheses about how literary objects function in the world—as commodities and as sources of entertainment, education, and meaning-making—we need better ways to measure constructs like book sales and readership.

We introduce a unique, freely available alternative to proprietary industry book data: the Seattle Public Library's (SPL) open checkout data. This data offers detailed, longitudinal insight into what people borrow—and, by extension, what they read—across genres, time periods, and even moments of social or political transformation. The SPL regularly releases two kinds of anonymized checkout datasets, which span from 2005 to the present:

1. Timestamped (rounded to the nearest minute) checkouts by title for **all physical items** in their collection;[2]
2. Monthly aggregated checkouts by title for **all physical and digital items** in their collection, broken down by **medium**.[3]

While many public libraries in the U.S. publish limited or high-level circulation data—such as lists of the 25 most popular fiction titles in a given year[4]—the Seattle Public Library (SPL) stands apart for the granularity and temporal specificity of the data it provides.

For example, in 2023, audiobook checkouts of Bram Stoker's *Dracula* (Duke Classics) spike in the month of Halloween, with aggregated checkouts rising from 32 in September to 64 in October, before falling back to half as many the next month. To our knowledge, SPL is the only library in the U.S. that openly publishes checkout data at this level of specificity.

In this paper, we introduce the SPL's **monthly checkout by title** dataset, discuss some of its affordances and limitations, and evaluate what kinds of research questions it could help answer. We specifically evaluate how well this data can be used as a proxy for book sales and for library checkouts in other geographic regions. We focus on the monthly checkouts dataset because it includes popular digital items like ebooks and audiobooks.

We provide three primary contributions:

---

1 Berglund (2024) represents an exceptional case because he was able to negotiate for access to private industry audiobook streaming data.

2 Checkouts by Title (2017). Retrieved from https://data.seattle.gov/Community-and-Culture/Checkouts-By-Title-Physical-Items-/5src-czff/about_data (last accessed: June 6, 2025).

3 Checkouts by Title (2017). Retrieved from https://data.seattle.gov/Community-and-Culture/Checkouts-by-Title/tmmm-ytt6/about_data (last accessed: June 6, 2025).

4 Popular Fiction Titles at the Chicago Public Library (2011). Retrieved from https://data.cityofchicago.org/Education/Libraries-Popular-Fiction-Titles-at-the-Chicago-Pu/nv46-bxa3/about_data (last accessed: June 6, 2025).

1. We provide an introduction to the Seattle Public Library's open checkout data, including suggested procedures and code for accessing it.
2. We discuss some idiosyncrasies and anomalies in the data, including ones that may influence long-term longitudinal modeling and comparisons.
3. We perform tests of *convergent validity*, in which we evaluate how well the checkout data serves as a measurement model (Jacobs & Wallach, 2021) for book sales and national readership.

2 DATASET DESCRIPTION

REPOSITORY LOCATION

Continuously updated data is available via the City of Seattle: https://data.seattle.gov/Community-and-Culture/Checkouts-by-Title/tmmm-ytt6/about_data. A copy of the data from April 2005 to February 2025 is available via Zenodo: https://zenodo.org/records/15792698.

REPOSITORY NAME

City of Seattle Open Data; Zenodo.

OBJECT NAME

"Checkouts by Title."

FORMAT NAMES AND VERSIONS

The data from the City of Seattle is available for download in the following formats: CSV, TSV, RDF, RSS, XML. The data is also available through an API endpoint as a CSV or JSON. The data from Zenodo is available as a CSV.

CREATION DATES

The resource was created on January 31, 2017, but the checkout information dates back to 2005.

DATASET CREATORS

David Christensen, Data Analysis Lead, Seattle Public Library. Data from 2005 to 2016 in this dataset is from the digital artwork, "Making Visible the Invisible," by studios of George Legrady.

LANGUAGE

English.

LICENSE

Public Domain.

PUBLICATION DATE

The data from the City of Seattle is regularly updated every month. The Zenodo dataset was published in June 2025.

3 SEATTLE PUBLIC LIBRARY (SPL) BACKGROUND

Because we evaluate the geographic specificity of the checkout data, it is important to provide context for this specific library and its users. Seattle, Washington is a major metropolitan city in the Pacific Northwest region of the United States, with a population of approximately 755,000 (Bureau, 2024). Seattle's population is diverse but predominantly white, with a racial composition of approximately 60% white, 17% Asian, 8% Hispanic or Latino, 7% Black, 10% multiracial, .6% American Indian and Alaska Native, and .3% Native Hawaiian or other Pacific

Islander (Bureau, 2024). More than 55% of adult residents identified as Democrat or leaning Democrat in 2024, tying for the sixth most liberal urban area in the U.S. (Balk, 2024). The SPL serves the city through 27 branch locations. In 2023, SPL had 293,000 active patrons and recorded 13.4 million checkouts (*2023 Statistical & Financial Summaries*, 2023).

4 DATASET DISCUSSION

The SPL checkout data is organized monthly, spanning from 2005 to the present. Metadata fields include:

- **UsageClass**: Denotes if the item is "physical" or "digital." *(Text)*
- **CheckoutType**: Denotes the vendor tool used to check out the item. *(Text)*
- **MaterialType**: Describes the type of item checked out (examples: book, song, movie, music, magazine). *(Text)*
- **CheckoutYear**: The 4-digit year of checkout for this record. *(Number)*
- **CheckoutMonth**: The month of checkout for this record. *(Number)*
- **Checkouts**: A count of the number of times the title was checked out within the "Checkout Month." *(Number)*
- **Title**: The full title and subtitle of an individual item. *(Text)*
- **ISBN**: A comma-separated list of ISBNs associated with the item record for the checkout. *(Text)*
- **Creator**: The author or entity responsible for authoring the item. *(Text)*
- **Subjects**: The subject of the item as it appears in the catalog. *(Text)*
- **Publisher**: The publisher of the title. *(Text)*
- **PublicationYear**: The year from the catalog record in which the item was published, printed, or copyrighted. *(Number)*

4.1 KEY PARTICULARITIES

There are several particularities in the SPL checkout data worth noting:

1. **ISBN field** – The *ISBN* field, which contains International Standard Book Numbers (ISBNs)—unique commercial identifiers for each edition or variation of a book—was added in June 2022. Checkout records before this date were not retroactively updated with ISBNs. This field will thus be blank for earlier data and may need to be merged with later records.
2. **PublicationYear** – The *PublicationYear* field does not always indicate the publication date (and certainly not the work's *first* publication date); it can also represent copyright, printing, or phonogram copyright dates, or even an estimated publication range. The formatting indicates the specific type of date recorded, including:
    - **2005** – Publication date
    - **c. 2005** – Copyright date
    - **[2005]** – Printing date
    - **p. 2005** – Phonogram copyright date
    - **2004, c. 2005** – Both publication and copyright dates
    - **2005–2007** – A range of years
    - **[2005?]** – Approximate date
3. **Subjects field inconsistencies** – The *Subjects* field is drawn from Library of Congress Subject Headings (LCSH). LCSH is a controlled vocabulary used for organizing library materials by subject. It is challenging to parse in this dataset because it is a long string

of text with an inherent hierarchical structure and complex syntax. Subject headings often include multiple levels of subdivisions, such as topical ("Climate change—Economic aspects"), geographic ("Education—United States—History"), chronological ("Feminism—19th century"), and form ("Science fiction—Bibliography"), each following specific ordering rules. LCSH terms also sometimes change due to political and social concerns, such as the shift from "Illegal aliens" to "Noncitizens." These changes are updated for current and future checkout records but are not updated for retroactive records.

4.2 HOW ARE CHECKOUTS COUNTED? HOW ARE USERS ANONYMIZED?

The SPL's methodology for recording checkouts is both highly granular and systematic in its protection of privacy. Each individual checkout is counted at the moment the patron initiates the borrowing process. SPL implements a two-pronged de-identification approach at the point of data capture. First, instead of storing personally identifiable information such as names or library card numbers, identifiers are generated through an irreversible tokenization process using a cryptographic "hash" algorithm. Second, the system stores records about items borrowed separately from any patron identifiers, with no common key to join. While internal researchers can track overall borrowing patterns, there is no possibility of re-identifying individual patrons. This dual approach of precise checkout counting and stringent user de-identification not only preserves the utility of the dataset for longitudinal and comparative research but also adheres to high standards of privacy and data protection. What users see in open data is essentially the item-level data without the tokenized identifiers. For more information on the technical details of the de-identification process, refer to Loter (2016).

While SPL's de-identification methods are designed to uphold strict privacy protections, it is worth acknowledging the broader ethical and technical challenges surrounding patron data. A notable example of the risks involved in handling library data is the inadvertent release of personally identifiable information (PII) in circulation records published through a 2023 *Code4Lib Journal* article (Board, 2023; Schuster, 2023; Swenson, 2021). This incident highlights how even well-intentioned data sharing can pose serious privacy risks if adequate safeguards are not implemented (Chin et al., 2023).

4.3 WHAT ISN'T COUNTED?

Important features that are not in the SPL checkout data include holds and physical book renewals. If an SPL patron wants to check out an in-use volume, they can place a "hold" on the item and join a waitlist. The number of holds for any given title is displayed on the SPL's website and catalog,[5] but this information is not included in the checkout data. The library's inventory limits and the invisibility of holds may contribute to divergences between library checkouts and book sales, which are not typically constrained by inventory.[6] Internally, the SPL uses hold-checkout ratios to manage and scale inventory for in-demand books, ensuring that inventory catches up with demand over time.[7]

Additionally, the checkout data only includes information about the first checkout for physical items, not any subsequent renewals. There is no information about how long a particular user has checked out any given item or book.

4.4 CHECKOUT DISTRIBUTION AND DATA COLLECTION HISTORY

There are a few anomalies in the SPL's checkout trends and data collection procedures important to consider. As shown in Figure 1, checkouts dropped dramatically in 2020 during the COVID-19 pandemic. Beginning on March 13, 2020, the SPL halted physical checkouts,

5 Seattle Public Library. (n.d). Braiding Sweetgrass. Retrieved from https://seattle.bibliocommons.com/v2/record/S30C2936388 (last accessed: 6 June 2025).

6 Publishers estimate demand pre-release to ensure they have enough inventory (Sinykin, 2023). While book sales aren't mediated by inventory as extensively as library checkouts, they are instead mediated by price. Neither book sales nor checkouts are perfect measurements of actual reader preferences; both measurements are partially determined through economic structures.

7 The SPL operates with a limited operating budget each year which a portion of is used to manage inventory. The 2025 budget was a record $100.3M. For longitudinal information on the SPL budget, see https://www.spl.org/about-us/the-organization/budget-and-operations/budget (last accessed: 30 June 2025).

and progressively reintroduced them over the next year until a full re-opening on July 1, 2021 (The Seattle Public Library, 2024a). As discussed in Section 7.1.2, patrons instead turned toward ebooks and audiobooks, a habit that stuck even after the SPL physically re-opened its doors.

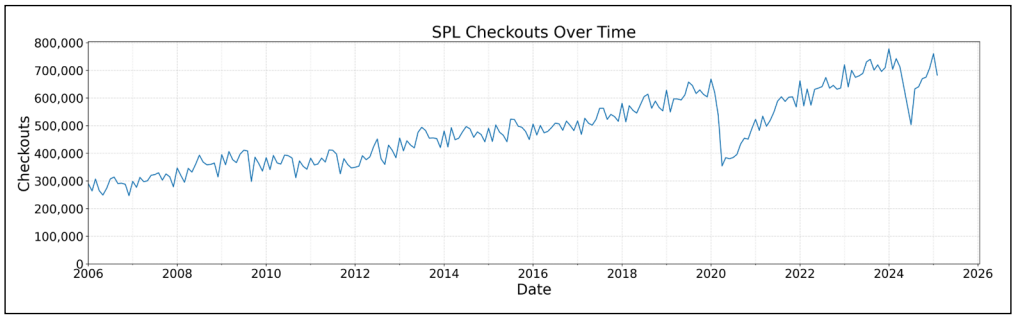

Figure 1 Total Seattle Public Library checkouts from 2005 to February 2025, filtered down to print books, audiobooks, and ebooks.

Recorded checkouts also plummeted between May and July of 2024, following a ransomware cyberattack on May 25 (Schlosser, 2024) that caused a prolonged service outage. As a response, the library shut down some of their data tracking systems. While checkouts continued via manual paper process, these checkouts do not show up exactly on the date of checkout in the SPL open data. Physical holds were not available from May 25 to August 8, and digital book checkouts were also unavailable for most patrons for almost three weeks of the ransomware outage.[8]

Lastly, across the twenty-year timeframe, the SPL made policy changes around fines, renewals, and digital holds that likely influence checkout behavior and trends. In January 2020, the library stopped fining patrons for late returns, leading to a slight lengthening of average checkout time (10 to 12 days for the 25th percentile). In July 2022, patrons were allowed to renew checkouts up to three times instead of two. Finally, in March 2024, the library reduced the maximum number of digital holds a specific patron can have from 25 to 10 (in order to cut costs for high demand volumes) (Library, 2023). We caution that researchers should not treat the longitudinal checkout data as homogeneously representing the same relationship to readership across the time series.

4.5 THE "WORK" CHALLENGE

The last challenge that we would like to discuss is the lack of persistent "work" identifiers in the data, a challenge that is not unique to the SPL but one that nevertheless makes analyzing some trends difficult. To stick with our opening example, let's say we are interested in analyzing engagement with Bram Stoker's *Dracula*. There are many different variations and editions of *Dracula*, but a literary scholar might be interested in considering the overarching "work" in the Functional Requirements for Bibliographic Records (FRBR)[9] sense—that is, all the editions and variations clustered together. Unfortunately, clustering all relevant editions and variations of *Dracula* is not easy. There are many deviations in title and author name formatting due to differences in publishing conventions, and in the digital and physical checkout systems. Consider all the title variations that refer to the same work of *Dracula*:

- *Dracula*
- *Dracula (Abridged)*
- *Dracula (Unabridged)*
- *Dracula / Bram Stoker; illustrated by Tudor Humphries; [abridgment, Jo Fletcher-Watson]*
- *Dracula / Bram Stoker; edited with an introduction and notes by Maurice Hindle; preface by Christopher Frayling*
- *Dracula: (Penguin Classics Deluxe Edition)*

8 More information about the ransomware outage can be found in the SPL's 2024 "Levy Report" (The Seattle Public Library, 2024b).

9 For more information, refer to https://www.oclc.org/research/activities/frbr.html (last accessed: June 30 2025).

To add to the complication, there is also author name variability: “Bram Stoker”; “Stoker, Bram, 1847–1912”; “Stoker, Bram, 1847–1912”.

While the SPL checkout data now includes ISBNs—unique commercial identifiers—they apply only to specific editions and formats of a work. Persistent identifiers at the work level do not currently exist, making large-scale analysis of literary works challenging. Different editions and variations can often be clustered manually or semi-manually for targeted subsets, as is the case in our analysis in Section 7.2. Scaling this kind of clustering remains an important open research challenge.

5 REUSE POTENTIAL

The SPL checkout data holds significant potential for a range of communities and research objectives. Research questions around book popularity have long been central to libraries and literary studies, especially subfields like reader’s advisory, reader-response theory, and book history (Felsenstein & Connolly, 2015; Pawley, 2001; Pierce, 2020; Radway, 1997; Saricks, 2005). They have also been addressed in disciplines like history to understand how national culture is reflected in popular books (Clair, 2007; Rose, 2003), sociology to argue that reading is a constitutive activity for distinct demographic classes (Griswold, 2008), and business to understand the variables that determine books’ economic success (Wang et al., 2019; Yucesoy et al., 2018).

We outline some potential research questions for specific communities below:

Library and Information Science

- How do borrowing habits differ between library systems and regions?
- What formats (e.g., print, ebooks, audiobooks) do patrons prefer, and how have these preferences shifted over time?
- How can longitudinal borrowing data support collection development and resource allocation?

Literary Studies

- How, and why, do genres and literary forms rise or fall in popularity?
- Which books or authors demonstrate long-term popularity, and which trends are more ephemeral?
- How might circulation data complement reader-response theory or historical models of reading?

Other Humanities and Social Sciences

- How do cultural or political events influence borrowing behavior?
- How do reading patterns reflect local or national identities and values?
- What factors predict a book’s sustained popularity or cultural impact, beyond initial commercial success?

In the next section, we specifically evaluate how well the checkout data aligns with book sales and broader library readership to investigate its potential for future use.

6 METHOD

6.1 CONVERGENT VALIDITY TESTS

To assess the SPL data’s suitability for future research, we use *convergent validity* tests to explore the relationship between SPL checkouts and 1) book sales 2) library checkouts in other regions. Because comprehensive book sales data in the U.S. is not available to researchers or the general public, and because other libraries do not release detailed open checkout data, making this comparison is difficult. We thus turn to two sources that serve as starting points.

6.1.1 Book Sales

To compare SPL's checkouts to book sales, we draw on prior scholarship on sales data. We specifically investigate whether the distribution of checkouts in the SPL differs from the analogous relationship with sales of best-selling books (Hackett, 1967; Newman, 2005; Sorensen, 2007).

We also draw on select, aggregated sales information from a series of Publishing Perspective's articles that summarize the Association of American Publisher's (AAP) monthly and annual statistics reports, or "StatShots" (Anderson, 2023, 2024, 2025). Although the AAP's StatShot reports are pay-walled—for example, the 2023 AAP StatShot Annual can be purchased for $495—these articles give us some of the basic statistics from those reports, such as aggregate yearly book sales by format.

6.1.2 Geographic Specificity

To compare SPL's checkouts to library checkouts in another region, we draw on the NYPL's published list of its top 10 most checked out books from 2021, 2022, and 2023. New York is another major metropolis—indeed the most populous city in the U.S.—so our comparison does not shed much light on how SPL checkouts might compare to checkouts in smaller cities or towns. Nevertheless, New York is on the opposite coast from Seattle with its own distinct demographic composition.[10]

6.2 ACCESSING AND ANALYZING THE SPL CHECKOUT DATA

Users can search, filter, and download the SPL's checkout data from Seattle's Open Data web interface.[11] The data is also available through an API endpoint. We gathered all available monthly SPL checkout data (April 2005–February 2025) by developing an R script to work with the API (Seattle Public Library, 2024). Given the size of the dataframe, we employed Polars in Python for efficient data manipulation.[12]

7 RESULTS

7.1 DO SPL CHECKOUTS LOOK LIKE BOOK SALES?

7.1.1 Yes, Library Checkouts Offer a Window Into Books as Market Goods

We find that library checkouts resemble book sales in the overall pattern of relative consumption among popular titles. Both in library checkouts and in book sales, we see readers flock to a set of ultra-popular books, with consumption dropping off steeply before flattening out into a long tail.

To compare overall library checkout trends with book sales patterns, we draw on prior work that analyzed BookScan data before the company changed their policies barring researchers from accessing the data. Sorensen (2007) leveraged BookScan data to investigate sales patterns of top bestsellers, noting that "the distribution of sales across books is heavily skewed." As demonstrated in Figure 2, there is a steep drop in book sales across the top 100 most popular books from a dataset of 1.2k hardcover bestsellers from 2001 and 2002. Similarly, Hackett (1967) collected data on the 633 bestselling books that sold 2 million or more copies between 1895 and 1965. Newman (2005) analyzed this data and fit a power law distribution to it, demonstrating that the relationship between book sales and book rank is consistent with other social phenomenon, like citations of scientific papers (Price, 1965), sales of popular music recordings (Cox et al., 1995), and market shares of commercial brands (Kohli & Sah, 2006).

10 For more thorough statistical comparisons between Seattle and New York, refer to https://www.census.gov/quickfacts/fact/table/newyorkcitynewyork,seattlecitywashington/PST045224 (last accessed 30 June 2025).

11 Checkouts by Title (2017). Retrieved from https://data.seattle.gov/Community-and-Culture/Checkouts-by-Title/tmmm-ytt6/explore/query/ (last accessed: June 6, 2025).

12 Code can be found at https://github.com/neelgupta2112/Seattle-Public-Library-Library-Checkout-Data.

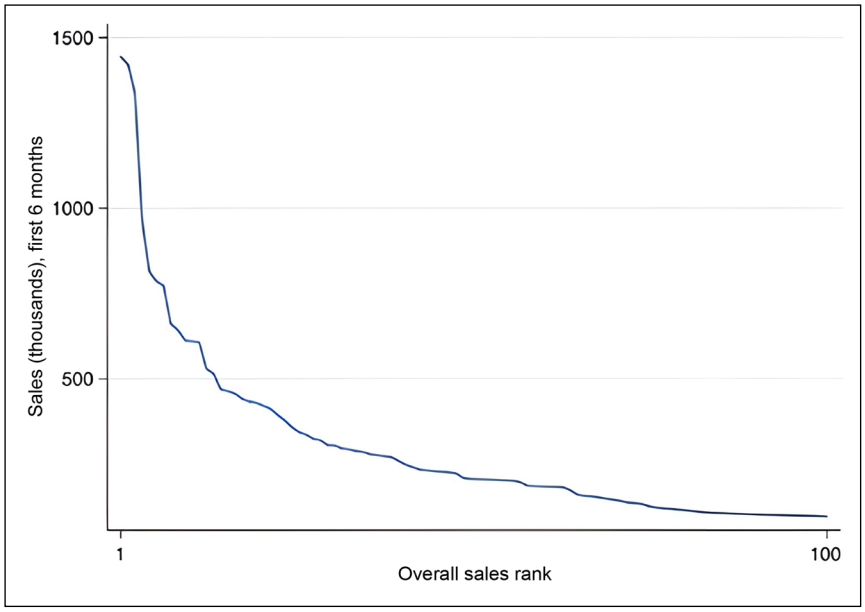

Figure 2 Sorensen (2007) plots the top 100 bestselling books in 2001–2002, drawing from a dataset of 1.2k hardcover fiction titles.

We find that SPL library checkout patterns follow the same trend (Figure 3). Top ranked books are significantly more checked out than books even a couple of ranks below them. To measure the precision of the mathematical relationship, we model the distribution of SPL library checkouts from 2021 to 2023 with a power-law function:

$$\text{checkout\_share} = a \cdot \text{checkout\_rank}^{b}$$

where a is a scaling constant and b is the exponent characterizing the distribution.

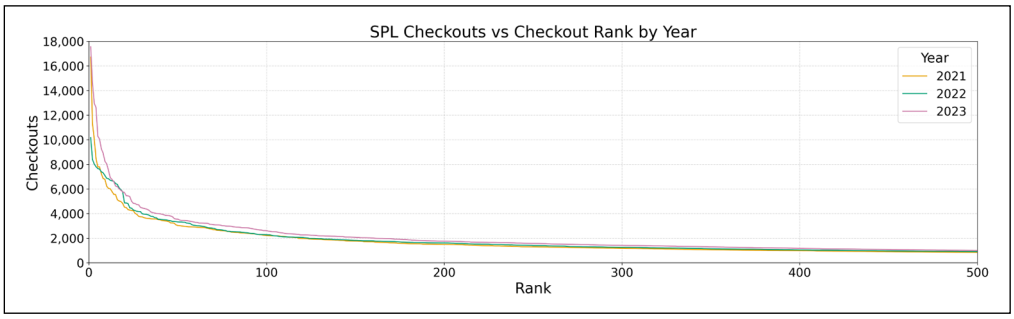

Figure 3 Checkouts of books dramatically drop outside top-ranked volumes.

For our model, we focus on the top 1k most popular volumes for each year from 2021-2023. For our dependent variable, we use checkout share, rather than checkouts, to normalize against different yearly dynamics in overall book demand.

As shown in Table 1, the regression fits the data tightly with an R^2 value of 0.99. We visualize this fit compared to the observed data in Figure 4. The model is especially reliable for books ranked outside the top 20 in the long tail, where borrowing patterns are more stable. However, at the very top of the distribution, we see wider variance. For example in 2021, Brit Bennett's *The Vanishing Half*—the most-checked-out book—accounted for twice the share of checkouts compared to 2022's top-ranked book, Anthony Doerr's *Cloud Cuckoo Land*. This is consistent with the behavior of power laws in nature in general, which rarely follow the distribution precisely over their entire range (Newman, 2005).

	COEFFICIENT	STANDARD ERROR
Intercept (a)	0.3937	—
Slope (b)	–0.5572	(0.0012)
R^2	0.987	
p-value	$\leq$ 0.001	
Observations	3000	

Table 1 This table shows the results of our Log-Log regression of checkout share by checkout rank.

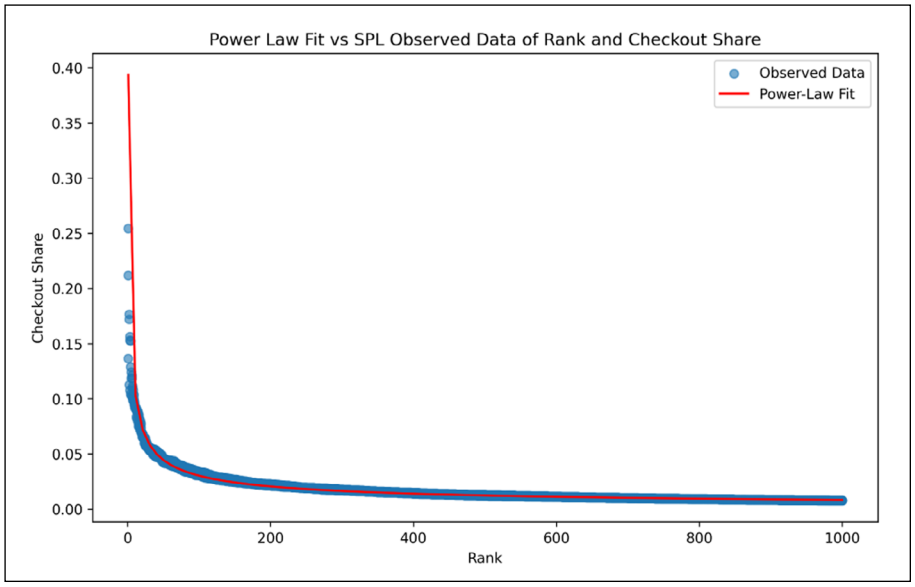

Figure 4 This figure shows the line of best fit graphed against the observed SPL Rank and checkout share data of top volumes from 2021–2023.

The persistence of the power law relationship across a huge temporal frame (the 20th century, 2001–2002, 2021–2023) demonstrates that both library checkouts and book sales register how literary objects function as commodities in competition, and suggests library checkouts may be an adequate proxy for book sales and readership more generally.

7.1.2 No, Library Patrons Embrace More Digital Books Than Regular Consumers

We identify at least one way that book sales and library checkouts seemingly diverge: ebook and audiobook use. Digital books have risen consistently in checkout share since their introduction to the SPL in the early 2010s, now outstripping physical checkouts (Figure 5). As Figure 6 shows, the most popular books are checked out digitally more than physically. The same sharp increase is not evident in book sales figures (Table 2). For example, in 2022 at the SPL, there were 2.8M[13] print checkouts and 4.7M digital checkouts, meaning that digital checkouts made up **63%** of total checkouts. By contrast, in 2022, the AAP (Anderson, 2023) reported that digital books were responsible for $1.9B in revenue compared to $6.9B for physical books, only **21%** of the total revenue.

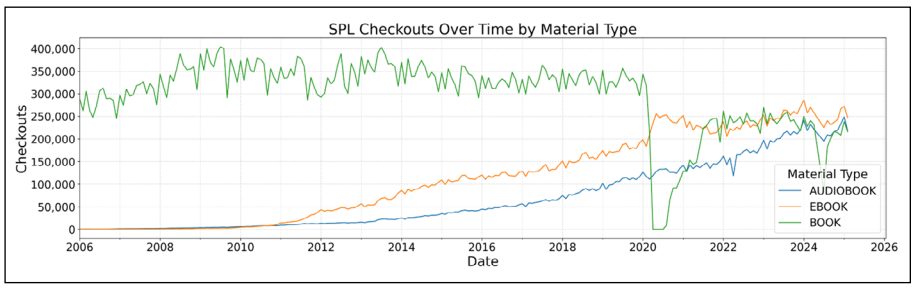

Figure 5 This figure shows checkouts from the Seattle Public Library from 2006 to 2025, broken out by audiobook, ebook, and print book.

Total revenue does not directly reflect the total number of book *purchases*. Popular commercial ebooks typically have lower prices than physical books, so we would expect that ebook revenue would be lower given the same amount of units. But we believe the gap is too wide to be explained by price alone (*Printed Books vs eBooks Statistics, Trends and Facts [2024]*, 2024; Wischenbart, 2012), and the upward trend in digital books is much stronger in the SPL checkout data than in the revenue reports. Also contributing to digital book numbers in 2023 and onward is the SPL Books Unbanned Program,[14] which provides free nationwide ebook access for teens and young adults up to age 26. In 2023, Books Unbanned was responsible for 84K checkouts, and in 2024, 245K checkouts, a 5% share of all digital checkouts.

13 Digital book checkouts include renewals while print checkouts do not, presenting a potentially unfair comparison in the numbers. However, in 2022, renewals made up only 5% of all checkouts on the digital side, a small enough percentage that it does not alter the directional read. We stick with the numbers in the dataset rather than scaling for ease of replication.

14 Refer to https://www.spl.org/programs-and-services/teens/books-unbanned (last accessed 30 June 2025) for more information.

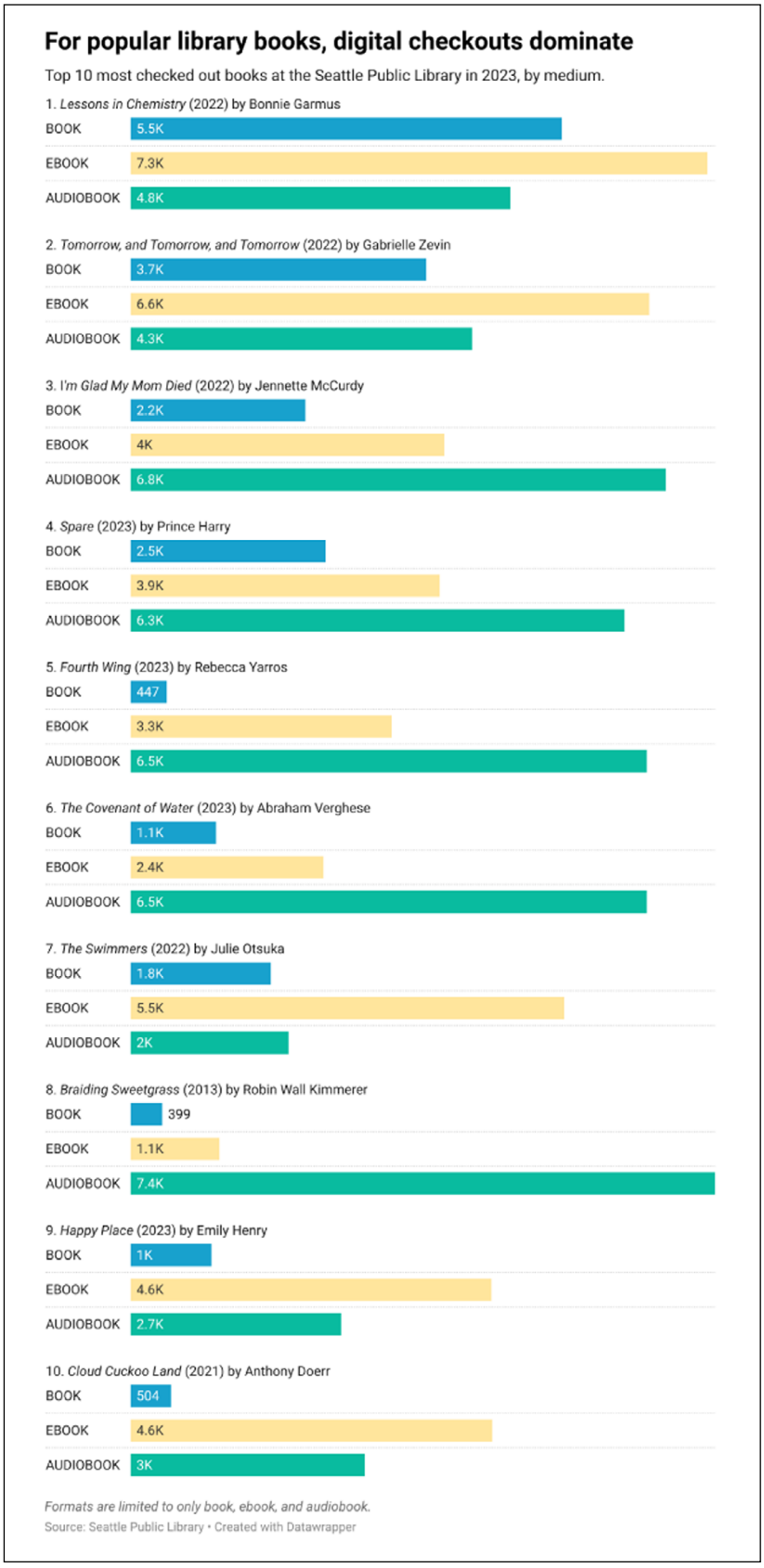

Figure 6 This figure shows the top-10 Checked out Books in the SPL in 2023 by Format.

YEAR	DIGITAL BOOK SALES (% OF TOTAL)	DIGITAL BOOK LIBRARY CHECKOUTS (% OF TOTAL)
2022	21.59%	61.98%
2023	21.97%	65.21%
2024	23.66%	69.61%

Table 2 This table displays the proportion of digital book sales revenue nationally—as reported by the Association of American Publisher's annual statistics reports (Anderson, 2023, 2024, 2025)—and the proportion of digital checkouts from the Seattle Public Library based on their open checkout data.

The difference between SPL checkouts and the larger publishing industry may be partially explained by stringent library rules regarding the checkout of physical books during the pandemic (*A Look Back at the Library's COVID-19 Response*, 2020). Ebooks gained a temporary checkout boost over those couple of months, as print-preferring customers were forced to switch formats in the absence of alternatives. That behavioral shift seemed to stick, as after the lockdowns concluded, physical checkout share dropped to 30% by 2024.

There is evidence of libraries' outsized embrace of digital books beyond the SPL checkout data. For example, in an interview with Lamdan et al. (2023), a Harper Collins executive reported that the Big 5 publishing house made twice as much ($19 million) from library ebook licenses in 2021 as 2020. Libraries including and beyond SPL, likely responding to the constraints of the pandemic, seem to have shifted their inventory toward digital book options. Another Hachette executive noted that while libraries made up 50% of Hachette's total ebook units downloaded between 2017 and 2020, they only accounted for 13% of total ebook revenue, emphasizing both libraries' significant share of the digital book market, and how sales data inadequately accounts for library checkouts.

7.2 DO SPL CHECKOUTS LOOK LIKE NYPL CHECKOUTS?

To compare checkouts between the SPL and NYPL, we draw on the NYPL's annually published lists of their top-10 most checked out books, focusing on 2021–2023 (The New York Public Library, 2021, 2022, 2023). We compare those titles to their ranks in the SPL checkout dataset in order to look for meaningful similarities and differences (see Table 9.1 in the Appendix for all top-10 lists).

7.2.1 Yes, SPL and NYPL Share Popular Books; No, They Likely Differ in Actual Checkouts

There is consistent yet not overwhelming overlap between the top 10 most checked out books in Seattle and New York City. In 2021, five of the ten most popular volumes were shared between the two library systems, and three were shared in 2022 and 2023. Every single NYPL top-10 book lies in the SPL top-60 for the same year, and falls in the top 0.01% of most checked out volumes. In other words, the most popular books remain popular in both library systems.

However, by just looking at rank, we lose information on actual checkout numbers. We hypothesize that the *extent* to which popular books are checked out often differs between the two systems. From the distribution in Section 7.1.1, we know that small differences in checkout ranks correspond to large differences in absolute checkouts and checkout share. For example, *Mexican Gothic* by Silvia Moreno-Garcia was the second-most checked out book in the NYPL in 2021, but it was the 24th most checked out book in the SPL. If the NYPL were to have a similar distribution to the SPL, the checkout share of *Mexican Gothic* in the SPL may be less than half of its checkout share in the NYPL. This suggests that while a book might be broadly popular in multiple library systems, it might be *much* more popular in one system than another.

7.2.2 Some Books Are More Popular in A Specific Place, But Why?

Characterizing why books are more popular in one place than another is challenging. Is it based on preferences in a given community, or the myriad particularities of one library system versus another? We consider a few examples to discuss these complexities.

In 2023, Emily Henry's novel *Happy Place* (2023) had 8,254 checkouts in the SPL, 74% more checkouts than her recent previous novel, *Book Lovers* (2022). The relative ranking of these two books is flipped in the NYPL in the same year, with *Book Lovers* placing 4th in the rankings and *Happy Place* entirely out of the top 10. We hypothesize that *Book Lovers* particularly appeals to the New York community because of its content. *Book Lovers* is narrated by a New Yorker with a plot (spoiler alert!) that sees the protagonist return to the city to further her ambitions in publishing. By contrast, *Happy Place* is narrated by a struggling surgery resident in San Francisco who reconnects with her ex-fiance. While geographic-specific content may explain the difference in popularity between these two library systems, *Book Lovers*'s popularity decline (Figure 7) in the SPL is likely explained by library programming.

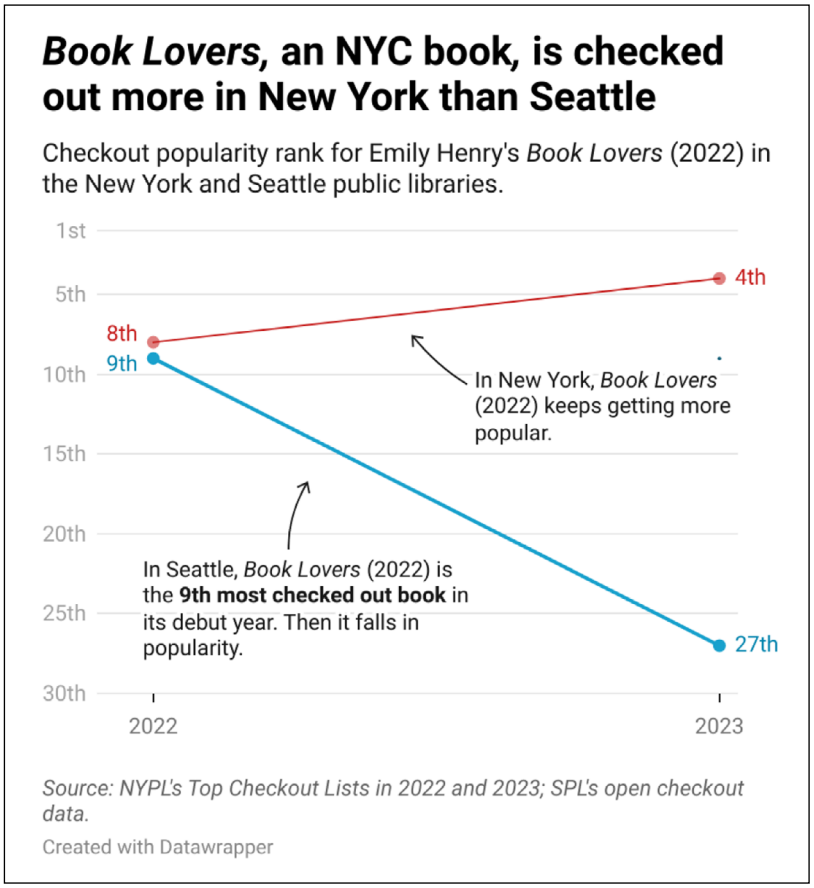

Figure 7 This figure shows Book Lover's rank in both the NYPL and SPL in 2022 and 2023.

In 2022, *Book Lovers* was featured as a Peak Picks[15] title—a special SPL program that promotes high-demand books by making them more accessible. Peak Picks titles are available on a first-come, first-served basis with shorter checkout periods, no holds, and no renewals, allowing more patrons to access them quickly. In other words, checkout trends don't reflect reader preferences in a vacuum; they are also shaped by library-level decisions around distribution and promotion. In fact, prior research has shown that marketing often complicates extrapolating normative preferences (what people actually prefer) from revealed preferences (what they choose) (Beshears et al., 2008)—or, in our context, extrapolating the books patrons prefer from the books patrons check out.

A similar case is Seattle's sustained engagement with Robin Wall Kimmerer's *Braiding Sweetgrass: Indigenous Wisdom, Scientific Knowledge, and the Teachings of Plants* (2013), which likely reflects both the tastes of a particular community and the influence of library programming. *Braiding Sweetgrass* is the only book that made SPL's top 10 list in every year that we examined (2021–2023). It never made the NYPL top-10 list in those years, and while the book did appear on the *NYT's* nonfiction bestseller list, it was not an across-the-board bestseller.[16]

The enduring appeal of *Braiding Sweetgrass* (Figure 8) at the SPL may be tied to the Pacific Northwest's deep connections to Indigenous communities, and to regional, environmental, and cultural initiatives that resonate with the book's themes. Seattle is located on the traditional land of the Coast Salish peoples, and local institutions have made visible efforts to acknowledge and engage with Indigenous knowledge systems.

But another factor that contributes to *Braiding Sweetgrass's* popularity is library inventory. *Braiding Sweetgrass* was added to the SPL's "Always Available" audiobook collection in December 2020, which means that any patron can borrow it instantly without placing a hold.[17] *Braiding Sweetgrass'* popularity in the SPL not only reflects sustained interest but also a unique access model that enables that interest to be consistently realized.

15 A Peek at Peak Picks for May 2022. (2022) Retrieved from https://blog.spl.org/2022/04/25/a-peek-at-peak-picks-for-may-2022/ (last accessed: 9 June 2025).

16 *Braiding Sweetgrass* has made top-10 lists at other libraries sporadically, like San Francisco and Amherst (Ulaby, 2024).

17 The Always Available collection can be found at https://spl.overdrive.com/library/avail/collection/1457015 (last accessed 30 June 2025).

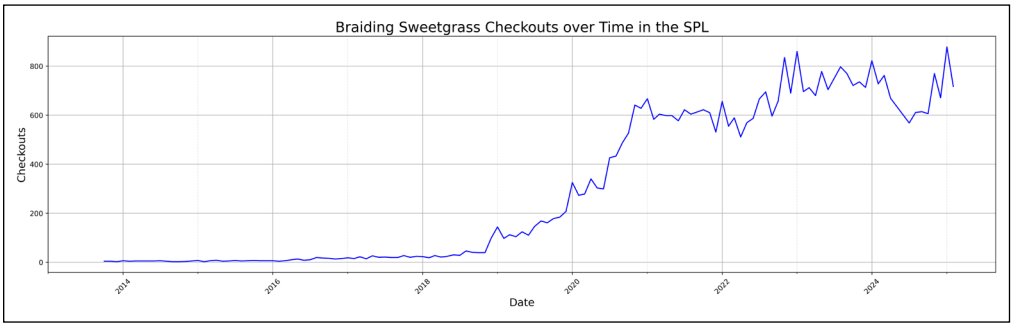

Figure 8 This figure shows the growth in checkouts of Robin Wall Kimmerer's *Braiding Sweetgrass* (2013).

8 CONCLUSION

The Seattle Public Library's open checkout data offers an unusually rich, granular, and freely accessible window into contemporary library borrowing and reading habits. Our analysis evaluates the extent to which this dataset can serve as a proxy for broader constructs like book sales and national library borrowing and readership, offering two primary findings:

1. **Book Sales:** SPL checkouts replicate a key structural pattern seen in book sales—namely a power-law distribution among popular titles—and consistently reflect interest in widely read books. However, they diverge in meaningful ways, particularly in patrons' strong uptake of digital formats and the influence of library-specific programming, highlighting limits to using checkout data as a direct proxy for sales behavior.
2. **National Library Borrowing and Readership:** In comparison with other library systems like the NYPL, SPL shares a core set of high-demand titles, but exhibits differences in relative checkout volumes and rankings. These variations—likely driven by regional preferences, institutional policy, and access models—suggest that SPL data cannot be assumed to reflect national readership uniformly.

We conclude that SPL's open checkout data is a powerful and underutilized resource for studying contemporary literary demand and reception, especially for questions that benefit from longitudinal, item-level, and format-specific insights. At the same time, we emphasize that its interpretive power depends on understanding the institutional, geographic, and infrastructural forces that shape circulation patterns in public libraries and in Seattle.

9 APPENDIX

9.1 TOP 10 MOST CHECKED OUT BOOKS — SPL VS. NYPL

Table 3 2021 NYPL Top 10.

NYPL RANK	TITLE	AUTHOR	SPL NUMBER OF CHECKOUTS	SPL CHECKOUT RANK
1	The Vanishing Half	Brit Bennett	16,720	1
2	Mexican Gothic	Silvia Moreno-Garcia	4,287	24
3	Klara and the Sun	Kazuo Ishiguro	6,851	9
4	A Promised Land	Barack Obama	11,319	2
5	Caste: The Origins of Our Discontents	Isabel Wilkerson	7,796	7
6	The Guest List	Lucy Foley	6,227	11
7	Where the Crawdads Sing	Delia Owens	6,046	12
8	Maybe You Should Talk to Someone	Lori Gottlieb	3,492	41
9	The Other Black Girl	Zakiya Dalila Harris	2,919	57
10	Malibu Rising	Taylor Jenkins Reid	4,489	22

SPL RANK	TITLE	AUTHOR	SPL NUMBER OF CHECKOUTS
1	The Vanishing Half	Brit Bennett	16,720
2	A Promised Land	Barack Obama	11,319
3	The Midnight Library	Matt Haig	10,070
4	Anxious People	Fredrik Backman	8,478
5	The Four Winds	Kristin Hannah	7,797
6	Caste: The Origins of Our Discontents	Isabel Wilkerson	7,796
7	Braiding Sweetgrass	Robin Wall Kimmerer	7,229
8	Klara and the Sun	Kazuo Ishiguro	6,851
9	Nomadland: Surviving America in the 21st Century	Jessica Bruder	6,794
10	The Guest List	Lucy Foley	6,227

Table 4 2021 SPL Top 10.

NYPL RANK	TITLE	AUTHOR	SPL NUMBER OF CHECKOUTS	SPL CHECKOUT RANK
1	The Midnight Library	Matt Haig	6,668	13
2	Lessons in Chemistry	Bonnie Garmus	6,089	17
3	The Lincoln Highway	Amor Towles	7,809	4
4	Malibu Rising	Taylor Jenkins Reid	3,857	34
5	People I Meet on Vacation	Emily Henry	4,441	24
6	This Time Tomorrow	Emma Straub	4,236	26
7	The Seven Husbands of Evelyn Hugo	Taylor Jenkins Reid	6,835	11
8	Book Lovers	Emily Henry	7,095	9
9	Verity	Colleen Hoover	3,360	47
10	It Ends With Us	Colleen Hoover	4,158	29

Table 5 2022 NYPL Top 10.

SPL RANK	TITLE	AUTHOR	SPL NUMBER OF CHECKOUTS
1	Cloud Cuckoo Land	Anthony Doerr	10,178
2	The House of Broken Angels	Luis Alberto Urrea	8,396
3	The Last Thing He Told Me	Laura Dave	8,024
4	The Lincoln Highway	Amor Towles	7,809
5	Crying in H-Mart	Michelle Zaun	7,660
6	Braiding Sweetgrass	Robin Wall Kimmerer	7,607
7	Apples Never Fall	Liane Moriarty	7,398
8	The Maid	Nita Prose	7,315
9	Book Lovers	Emily Henry	7,095
10	Anxious People	Fredrik Backman	6,867

Table 6 2022 SPL Top 10.

NYPL RANK	TITLE	AUTHOR	SPL NUMBER OF CHECKOUTS	SPL CHECKOUT RANK
1	Lessons in Chemistry	Bonnie Garmus	17,557	1
2	Tomorrow, and Tomorrow, and Tomorrow	Gabrielle Zevin	14,632	2
3	Spare	Prince Harry	12,616	4
4	Book Lovers	Emily Henry	4,738	27
5	Verity	Colleen Hoover	4,833	25
6	Yellowface	R. F. Kuang	6,009	18
7	The Heaven and Earth Grocery Store	James McBride	4,117	35
8	The Seven Husbands of Evelyn Hugo	Taylor Jenkins Reid	5,763	19
9	It Ends With Us	Colleen Hoover	5,004	24
10	Daisy Jones and the Six	Taylor Jenkins Reid	3,837	45

Table 7 2023 NYPL Top 10.

SPL RANK	TITLE	AUTHOR	SPL NUMBER OF CHECKOUTS
1	Lessons in Chemistry	Bonnie Garmus	17,557
2	Tomorrow, and Tomorrow, and Tomorrow	Gabrielle Zevin	14,632
3	I'm Glad my Mom Died	Jennette McCurdy	12,967
4	Spare	Prince Harry	12,616
5	Fourth Wing	Rebecca Yarros	10,288
6	The Covenant	Abraham Verghese	10,051
7	The Swimmer	Julie Otsuka	9,258
8	Braiding Sweetgrass	Robin Wall Kimmerer	8,917
9	Happy Place	Emily Henry	8,254
10	Cloud Cuckoo Land	Anthony Doerr	8,043

Table 8 2023 SPL Top 10.

DATA ACCESSIBILITY STATEMENT

Data supporting the findings of this study is openly available in Zenodo at https://zenodo.org/records/15792698. Besides the Zenodo snapshot that we created for data preservation purposes, the SPL checkout data is also stored through the city's data portal, the SPL's internal data warehouse, and on the artist George Legrady's servers (Legrady created an art installation that first collected the checkout data).[18] The SPL expects the continuously updated data to remain available despite the ambiguity around federal funding at this time.

The open checkout dataset continues to grow (currently 11.36 GB) and can be challenging to download or manipulate on computers without ample storage and memory. We used a high-performance computing package (specifically the Python library Polars) to efficiently process and analyze the data without overwhelming local resources. This may pose accessibility issues for those without computational expertise or resources.

ETHICS AND CONSENT

Patron privacy is a core ethical commitment of libraries, grounded in the values of intellectual freedom and confidentiality. Any work involving library usage data must take seriously the risks of exposing user data, even unintentionally. In recent years, incidents such as the inadvertent

18 For more information on Legrady's installation, refer to https://blog.spl.org/2022/12/06/how-a-digital-artwork-helped-the-seattle-public-library-lead-in-book-data/ (last accessed 30 June 2025).

publication of patron-linked circulation records in *Code4Lib* have underscored the need for caution and care in data sharing (Board, 2023; Schuster, 2023; Swenson, 2021).

The dataset discussed in this paper reflects the Seattle Public Library's rigorous privacy-preserving practices. All checkout data is de-identified at the point of capture using irreversible tokenization and disassociated storage of user and item records. By making this dataset available and documenting its structure and reuse potential, we aim to promote ethical, privacy-conscious research while upholding the library community's longstanding commitments to user confidentiality and responsible stewardship of public data. The ethical stakes of sharing library usage data remain complex, and we recognize that decisions about data openness must continue to be made with care, context, and evolving best practices in mind.

ACKNOWLEDGEMENTS

We would like to thank our anonymous reviewers for their helpful feedback and suggestions. We are also grateful to Ben Lee for feedback in the writing stage, and to Matt Salah and Meira Gilbert for feedback at the conceptualization stages of the project.

COMPETING INTERESTS

David Christensen works for the Seattle Public Library and manages the open checkout data. Views expressed in this piece do not necessarily reflect those of the SPL. Both Melanie Walsh and Neel Gupta are not affiliated with the Seattle Public Library and have no competing interests to report.

AUTHOR CONTRIBUTIONS

Neel Gupta: Conceptualization, Writing, Methodology, Visualization, Investigation, Formal Analysis.

David Christensen: Conceptualization, Writing, Supervision.

Melanie Walsh: Conceptualization, Writing, Visualization, Supervision.

AUTHOR NOTE

David Christensen is an employee of the Seattle Public Library and manages the library's open checkout data. This article was written as part of an independent research project. The views expressed are those of the authors and do not necessarily reflect those of the Seattle Public Library.

AUTHOR AFFILIATIONS

Neel Gupta orcid.org/0009-0000-1263-4009
Information School, University of Washington, Seattle, USA

David Christensen
Seattle Public Library, Seattle, USA

Melanie Walsh
Information School, University of Washington, Seattle, USA

REFERENCES

2023 Statistical & Financial Summaries. (2023). Retrieved 2025-03-11, from https://www.spl.org/about-us/library-impact/2023-impact-report/2023-impact-statistics

A Look Back at the Library's COVID-19 Response. (2020). Retrieved 2025-06-30, from https://web.archive.org/web/20241209044138/https://www.spl.org/about-us/library-impact/2020-impact-report/a-look-back

Anderson, P. (2023). AAP StatShot: Overall US Market Down 6.4 Percent in 2022. Retrieved 2025-06-30, from https://publishingperspectives.com/2023/02/aap-statshot-overall-us-market-down-6-4-percent-in-2022/

Anderson, P. (2024). *AAP StatShot: In 2023, US Revenues Were $12.6 Billion.* Retrieved 2025-06-30, from https://publishingperspectives.com/2024/03/aap-statshot-in-2023-us-revenues-were-12-6-billion/ (Publishing Perspectives)

Anderson, P. (2025). AAP's December StatShot: US Book Market Up 6.5 Percent. Retrieved 2025-06-30, from https://publishingperspectives.com/2025/03/aaps-december-statshot-us-market-up-6-5-percent-year-to-date/

Balk, G. (2024). Seattle Keeps Getting More Liberal, Survey Data Shows. Retrieved 2025-03-10, from https://www.seattletimes.com/seattle-news/data/seattle-keeps-getting-more-liberal-survey-data-shows/

Berglund, K. (2024). *Reading Audio Readers: Book Consumption in the Streaming Age*. Abingdon New York (N.Y.): Bloomsbury Academic. https://doi.org/10.5040/9781350358393

Beshears, J., Choi, J. J., Laibson, D., & **Madrian, B. C.** (2008). How are preferences revealed? *Journal of Public Economics*, *92*(8), 1787–1794. Retrieved from https://www.sciencedirect.com/science/article/pii/S0047272708000728. https://doi.org/10.1016/j.jpubeco.2008.04.010

Board, C. J. E. (2023). Extra Editorial: On the Release of Patron Data in Issue 58 of Code4Lib Journal. *The Code4Lib Journal, 58*. Retrieved 2025-06-08, from https://journal.code4lib.org/articles/18040

Bureau, U. C. (2024). *U.S. Census Bureau QuickFacts.* Retrieved 2025-03-10, from https://www.census.gov/quickfacts/fact/table/seattlecitywashington#

Chin, C.-L., Wang, T., Wang, E.-H., & **Bashir, M.** (2023). Patron Privacy Protections in Public Libraries: IT Professionals' Points of View. *Journal of Intellectual Freedom & Privacy*, *8*(3), 28–39. https://doi.org/10.5860/jifp.v8i3.7965

Clair, W. S. (2007). *The reading nation in the romantic period* (1st edition ed.). Cambridge: Cambridge University Press.

Cox, R. A. K., Felton, J. M., & **Chung, K. H.** (1995). The concentration of commercial success in popular music: An analysis of the distribution of gold records. *Journal of Cultural Economics*, *19*(4), 333–340. Retrieved 2025-06-30, from http://link.springer.com/10.1007/BF01073995. https://doi.org/10.1007/BF01073995

Felsenstein, F., & **Connolly, J. J.** (2015). *What Middletown Read: Print Culture in an American Small City* (First Edition ed.). Amherst: University of Massachusetts Press.

Griswold, W. (2008). *Regionalism and the reading class*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226309262.001.0001

Hackett, A. P. (1967). *70 years of best sellers, 1895–1965* (First Edition ed.). Chatham: R. R. Bowker Company.

Jacobs, A. Z., & **Wallach, H.** (2021). Measurement and Fairness. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 375–385). ACM. Retrieved 2024-10-04, from https://dl.acm.org/doi/10.1145/3442188.3445901. https://doi.org/10.1145/3442188.3445901

Kohli, R., & **Sah, R. K.** (2006). Market shares: Some power law results and observations. *Management Science*, *52*(11), 1792–1798. (Originally published as University of Chicago, Harris School of Public Policy Working Paper Series 04.01 (November 1, 2003)). https://doi.org/10.1287/mnsc.1060.0572

Lamdan, S., Weinberg, M., Schultz, J., & **Woodcock, C.** (2023). *The Anti-Ownership Ebook Economy: How Publishers and Platforms Have Reshaped How We Read in the Digital Age* (Tech. Rep.). Engelberg Center on Innovation Law & Policy, NYU School of Law. Retrieved 2025-06-30, from https://www.nyuengelberg.org/news/new-white-paper-reveals-complex-dynamic-between-ebook-publishers-and/. https://doi.org/10.33682/s4qr-ytrm

Library, S. P. (2023). *The Seattle Public Library Is Reducing Our Maximum Digital Holds. Here's Why.* Retrieved 2025-06-30, from https://shelftalkblog.wordpress.com/2024/03/13/the-seattle-public-library-is-reducing-our-maximum-digital-holds-heres-why/

Loter, J. (2016). Gaining insights and protecting privacy: De-identifying patron data at the seattle public library. *Alki: The Washington Library Association Journal*, *32*(1), 11–13. Retrieved 2025-06-30, from https://assets.pubpub.org/f9054k8d/01679859527513.pdf

Newman, M. (2005). Power laws, pareto distributions and zipf's law. *Contemporary Physics*, *46*(5), 323–351. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/00107510500052444. https://doi.org/10.1080/00107510500052444

Pawley, C. (2001). *Reading on the Middle Border: The Culture of Print in Late-Nineteenth-Century Osage, Iowa* (First Edition ed.). Amherst: University of Massachusetts Press.

Pierce, J. B. (2020). *Narratives, Nerdfighters, and New Media* (1st edition ed.). Iowa City: University Of Iowa Press. https://doi.org/10.2307/j.ctv16x2c1r

Porter, J. (2018). *Popularity/prestige* (No. 17). Retrieved 2025-06-30, from https://litlab.stanford.edu/assets/pdf/LiteraryLabPamphlet17.pdf

Price, D. J. D. S. (1965). Networks of scientific papers: The pattern of bibliographic references indicates the nature of the scientific research front. *Science*, *149*(3683), 510–515. Retrieved 2025-06-30, from https://www.science.org/doi/10.1126/science.149.3683.510. https://doi.org/10.1126/science.149.3683.510

Price Lab for Digital Humanities. (2023). *Reader studies in the digital age.* Symposium abstract, University of Pennsylvania. Retrieved 2025-06-30, from https://pricelab.sas.upenn.edu/reader-studies-digital-age (Symposium hosted by the Price Lab for Digital Humanities).

Printed Books vs eBooks Statistics, Trends and Facts [2024]. (2024). Retrieved 2025-02-26, from https://www.tonerbuzz.com/blog/paper-books-vs-ebooks-statistics/

Radway, J. A. (1997). *A feeling for books: the book-of-the-month club, literary taste, and middle-class desire*. Chapel Hill: University of North Carolina Press.

Rose, J. (2003). *The Intellectual Life of the British Working Classes* (New edition ed.). New Haven: Yale University Press.

Saricks, J. G. (2005). *Readers' Advisory Service in the Public Library*. Chicago: American Library Association.

Schlosser, K. (2024). *Seattle public library recovers from ransomware attack, makes moves to strengthen tech security.* Retrieved 2025-06-30, from https://www.geekwire.com/2024/seattle-public-library-recovers-from-ransomware-attack-makes-moves-to-strengthen-tech-security/

Schuster, D. (2023). Bringing it All Together: Data from Everywhere to Build Dashboards. *The Code4Lib Journal, 58*. Retrieved 2025-06-08, from https://journal.code4lib.org/articles/17693

Seattle Public Library. (2024). *Checkouts by title dataset.* Seattle Open Data Platform. Retrieved 2025-06-30, from https://data.seattle.gov/Community-and-Culture/Checkouts-by-Title/tmmm-ytt6/about_data

Sinykin, D. (2023). *Big fiction: How conglomeration changed the publishing industry and american literature*. New York, NY: Columbia University Press. https://doi.org/10.7312/siny19294

Sorensen, A. T. (2007). Bestseller Lists and Product Variety. *Journal of Industrial Economics, 55*(4), 715–738. Retrieved 2024-10-30, from https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-6451.2007.00327.x. https://doi.org/10.1111/j.1467-6451.2007.00327.x

Swenson, M. (2021). Editorial : The Cost of Knowing Our Users. *The Code4Lib Journal* (52). Retrieved 2025-06-08, from https://journal.code4lib.org/articles/16208

The New York Public Library. (2021). *NYPL's Top Checkouts of 2021*. Retrieved from https://www.nypl.org/spotlight/top-checkouts-2021 (Accessed: 2024-12-09).

The New York Public Library. (2022). *NYPL's Top Checkouts of 2022.* Retrieved from https://www.nypl.org/spotlight/top-checkouts-2022 (Accessed: 2024-12-09).

The New York Public Library. (2023). *NYPL's Top Checkouts of 2023.* Retrieved from https://www.nypl.org/spotlight/top-checkouts-2023 (Accessed: 2024-12-09).

The Seattle Public Library. (2024a). *A look back at the library's covid-19 response.* Retrieved 2025-06-30, from https://web.archive.org/web/20241209044138/https://www.spl.org/about-us/library-impact/2020-impact-report/a-look-back (Archived from the original on December 9, 2024).

The Seattle Public Library. (2024b). *Q2–Q3 2024 Levy Report* (Tech. Rep.). The Seattle Public Library. Retrieved 2025-06-30, from https://www.spl.org/Seattle-Public-Library/documents/about-us/levy/reports/Q2-Q3%202024%20Levy%20Report.pdf

Ulaby, N. (2024). *These were the most-borrowed books from public libraries in 2024*. NPR. Retrieved 2025-06-30, from https://www.npr.org/2024/12/29/nx-s1-5234258/most-borrowed-library-books-2024

Underwood, T. (2019). The long arc of literary prestige. In *Distant horizons: Digital evidence and literary change* (pp. 68–110). Chicago: The University of Chicago Press. https://doi.org/10.7208/chicago/9780226612973.001.0001

Walsh, M. (2022). Where is all the book data? *Public Books*. Retrieved 2025-06-30, from https://www.publicbooks.org/where-is-all-the-book-data/

Wang, X., Yucesoy, B., Varol, O., Eliassi-Rad, T., & **Barabási, A.-L.** (2019). Success in books: predicting book sales before publication. *EPJ Data Science*, *8*(1), 31. Retrieved 2025-06-30, from https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-019-0208-6. https://doi.org/10.1140/epjds/s13688-019-0208-6

Wischenbart, R. (2012). *The global eBook Market Current Conditions and Future Projections* (revised October 2012. ed.). Sebastopol: O'Reilly.

Yucesoy, B., Wang, X., Huang, J., & **Barabási, A.-L.** (2018). Success in books: a big data approach to bestsellers. *EPJ Data Science*, *7*(1), 7. Retrieved 2025-06-30, from https://epjdatascience.springeropen.com/articles/10.1140/epjds/s13688-018-0135-y. https://doi.org/10.1140/epjds/s13688-018-0135-y

TO CITE THIS ARTICLE:
Gupta, N., Christensen, D., & Walsh, M. (2025). Seattle Public Library's Open Checkout Data: What Can It Tell Us About Readers and Book Popularity More Broadly? *Journal of Open Humanities Data*, 11: 46, pp. 1–19. DOI: https://doi.org/10.5334/johd.332

Submitted: 16 April 2025
Accepted: 16 June 2025
Published: 13 August 2025

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.