Analyzing Book Checkout Trends at the Seattle Public Library: A 5-Year Data Study (2018-2022)

Sarvesh Fotedar, Ekrem Kizilkaya, Abbas Shaikh, Cody VanZandt, Andy Wang

March 1, 2023

Contents

1 Introduction

We present an preliminary examination of 11,749,255 checkouts from the Seattle Public Library system over a five-year period from 2018 to 2022. We intend to use this data – in concert with book review data from Goodreads – to examine Seattle's literary tastes and reading habits.

Animating this analysis are more general questions about the contours of contemporary American readership.

- Why do some books enjoy widespread acclaim while others fizzle?
- Do popular books explode onto the scene or accumulate readers more gradually?
- How do popular success and critical acclaim differ?
- How has recent publisher conglomeration of publishers changed the literary marketplace?
- And, ultimately, to what degree can the success or failure of a book be predicted?

The answers, we hope, will be illuminating for authors, publishers, critics, taste-makers, and readers alike.

2 Plots

2.1 Subject Breakdown by Publisher

Figure 1 shows the four most popular publishers based on the number of their books checked out over the 2018-2022 period. Each color on the bar represents the primary subject of a book that was checked out and the size of the colored region shows the proportion of the publisher's books with the same subject. The specific subjects were chosen from the most popular subjects of all books checked out from the library. The corresponding pie-chart was present in the previous draft of the report and can be found in the appendix of this draft.

This bar chart shows that fiction books are most frequently checked out from the top publishers at the library. We expected nonfiction to be the second most common subject based on our previous findings of subject popularity; however, nonfiction books do not seem to be as popular from these publishers. One explanation might be that nonfiction tends to show up as a secondary subject for most books, although further exploration would be needed to confirm.

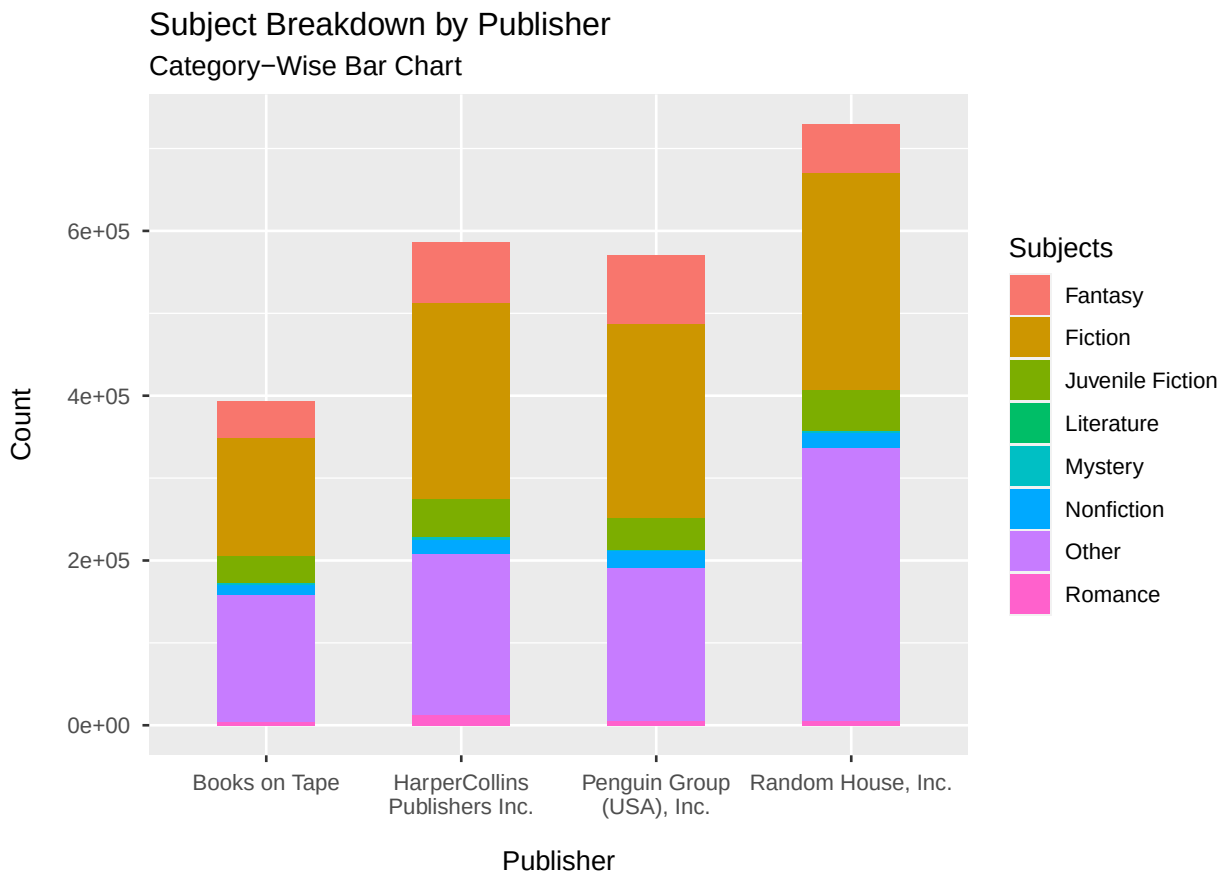

Figure 1: Subject Breakdown by Publisher

2.2 Logarithmic Scatter Plot for all Monthly Normalized Increase in Checkouts for Textbooks vs All Books

Figure 2 shows the monthly result from each year for the normalized increase in checkouts over the minimum month for that year, on a logarithmic scale. The idea behind the normalized increase method is to highlight the changes between months in the same year as each year is compared to the same reference value, rather than compare the absolute values between months, which may not differ by much. This can also serve as a way to compare other years, as each shape of the point corresponds to a different year, 2018-2022.

This logarithmic scatter plot is able to give a useful comparison between years, as months for most years have relatively the same magnitude in their normalized increase in checkouts over their minimum checkout month, for both textbooks and all books. But there is an exception for some months of 2020, as the normalized increase over the minimum month is many orders of magnitude greater than for any other years. The explanation for this is that at the beginning of the COVID-19 pandemic, the library suffered a severe lack in checkouts, making the monthly minimum number of books checked out significantly lower than usual, which when compared to the normal pre-COVID-19 months through the normalized increase method describes the normal months as having an increase that is orders of magnitude greater rather than a smaller ratio increase as was seen for the other years.

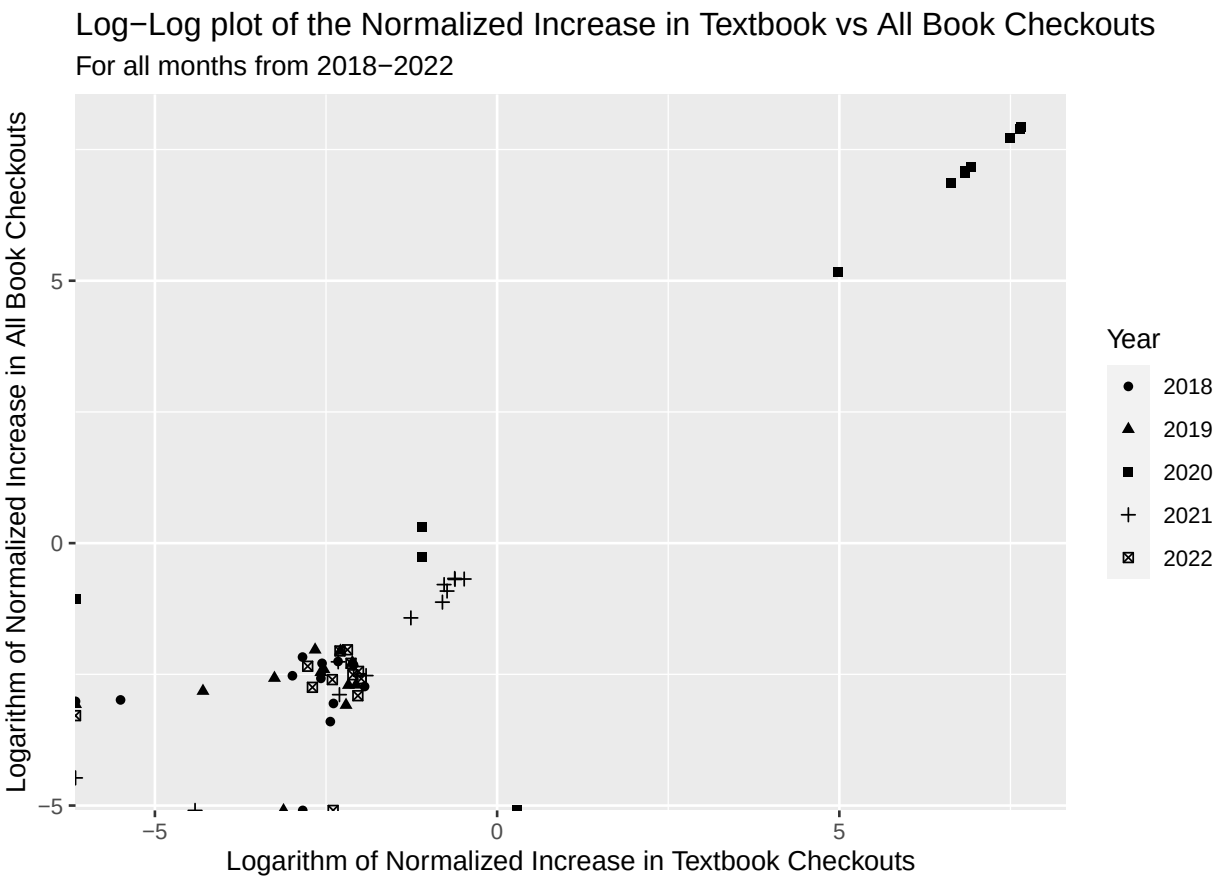

Figure 2: Logarithmic Scatter Plot of Normalized Increase in Book Checkouts

2.3 Top Publisher Audiobook Checkouts

Figure 3 shows number of audiobook checkouts for the 8 top audiobook publishers, 2018-2022. Even in this rather cursory glance, a couple of trends catch the eye. First, international publishing conglomerates – known in the trade as the "Big Five" – dominate. Four of the Big Five make the list by name: Hachette, HarperCollins, Macmillan, and Simon & Schuster. The fifth Big Five publisher, Penguin Random House, also makes the list through its imprint, Books on Tape. In fact, the only independent publisher on the list is Blackstone, as conglomerate RBMedia operates both Tantor Media and Recorded Books.

It seems without coincidence that the only independent publisher on the list, Blackstone, is also the only top publisher to see its checkouts decline over the period. There is, then, perhaps a larger story to be uncovered about the conglomeration of audiobook publishing and its deleterious effects on independent operators.

As an aside, Hachette Audio's near-vertical rise in checkouts bears further study. One wonders whether the massive publisher began a new acquisition or distribution strategy in early 2020.

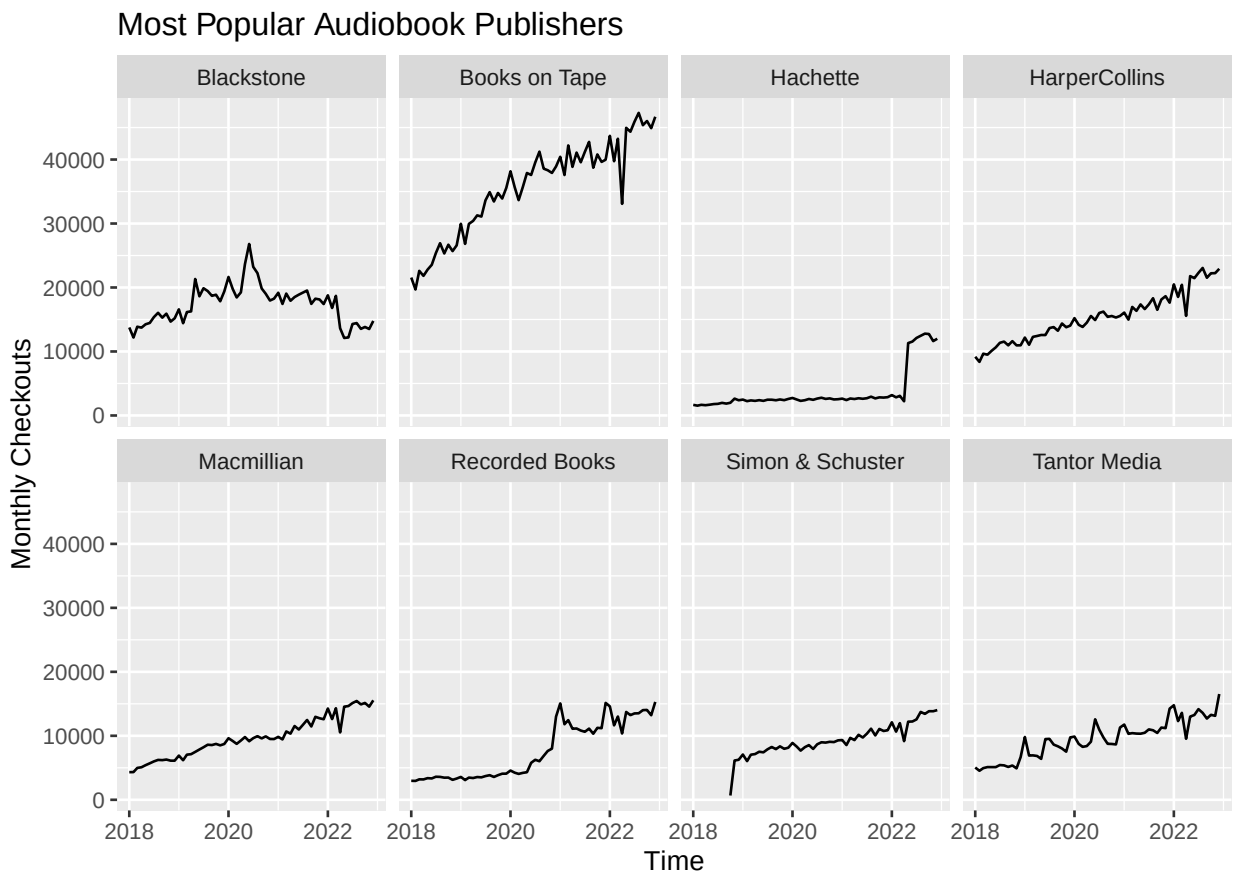

Figure 3: Checkouts for the top eight audiobook publishers, 2018-2022

2.4 Checkouts by Material Type

Figure 4 shows the number of checkouts from 2018 to 2022 by material type. The material type classification splits materials checked out from the Seattle Public Library into common categories including books, e-books, audio-books, songs, television, comics, movies, etc. Additionally, the usage class of each material type, i.e. whether it is physical or digital, is indicated by the color of each bar. The majority of checkouts were understandably books, e-books, and audio-books, however nearly 1 million checkouts each were also from video and sound discs. Apart from these 5 material types there is a significant decrease in checkouts, with no other material types exceeding 50000 checkouts. The most checked out of these less popular material types are songs, music, prints, television, and movies.

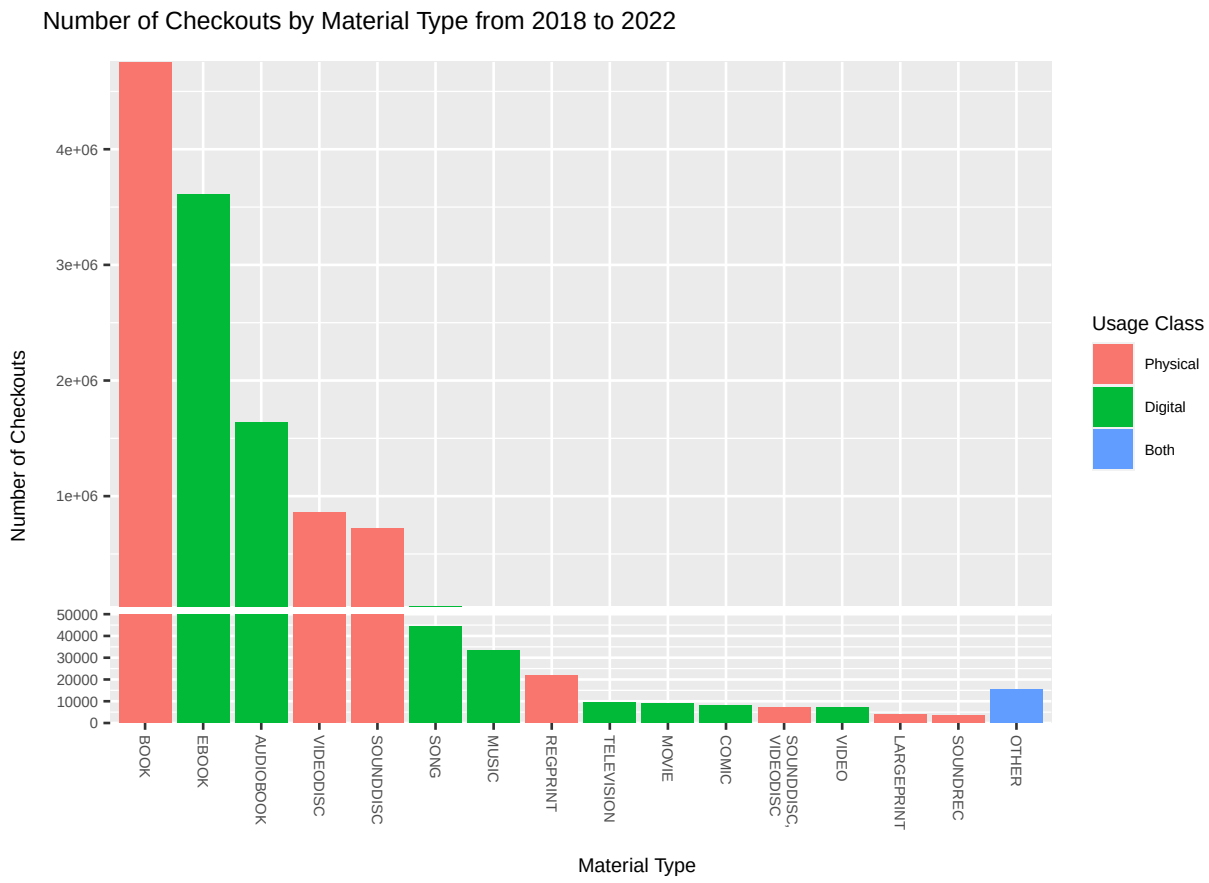

Figure 4: Checkouts by Material Type

2.5 Physical Book Checkout Over Time

Figure 5 is a seasonal polar plot that shows the physical book checkout count over time from 2018 to 2022 in the Seattle Public library. The values range from 20 to 30000 total checkouts per month. The effect of covid can be seen clearly as the checkout quantity from March 2020 to 2021 decreases significantly until July 2021 where it recovers to roughly half the amount of 2018 and 2018. The most popular months seem to be January, August, and October. A most likely explanation is due to the start of school, the library becomes more active within those months. February and December are the months with the least checkouts most likely due to February being the month with the least days and December being where most students being on vacation and not going to the library.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0    2011    2015    2013    2018    9820
   ## `summarise()` has grouped output by 'CheckoutMonth'.  You can override using the
                                    ## `.groups` argument.
## Error in ggseasonplot(x = toPlot.ts, polar = TRUE, continuous = FALSE): could not find
                                    function "ggseasonplot"
```

Figure 5: Physical Book Checkout Over Time

2.6 Total Checkout Heatmap

Figure 6 is a heatmap that shows the total checkouts by month and year. The figure also clearly shows the effect of covid starting March 2020 with the effects lasting until mid 2021. This heatmap more clearly shows which months are least active which is Febuary, September, November, and December. A possible explanation might be that Febuary is the month with the least days and the other months being when students are on vacation. The months with the most activity is during August, January, and July. A likely explanation for this would be students getting ready for the semester to start.

```
## `summarise()` has grouped output by 'CheckoutMonth'.  You can override using the## `.groups` argument.
```

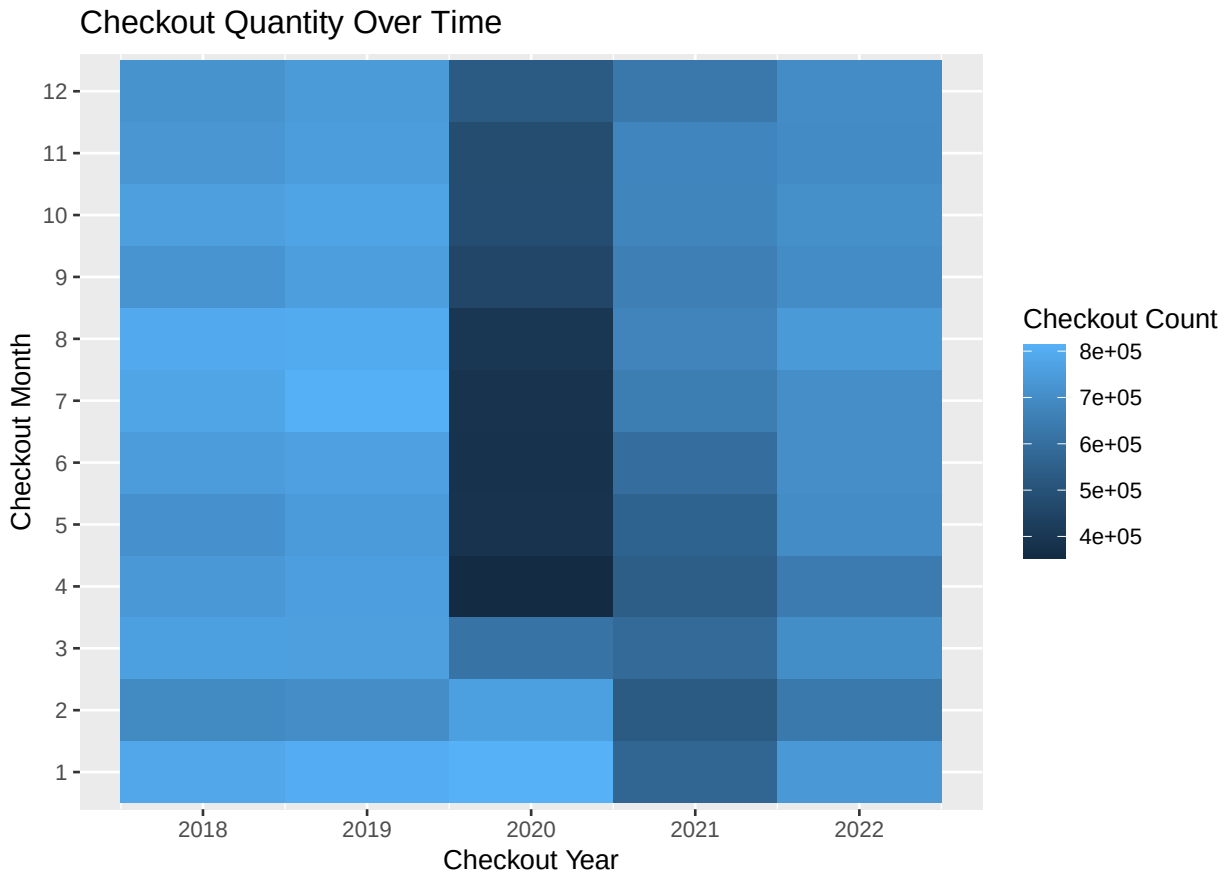

Figure 6: Total Checkout Heatmap