# Digital Patron Segmentation Study 2022

Note: as of 3/17/23, this study remains in draft form.

## Overview

The Washington Digital Library Consortium WDLC consists of 45 tiny-to-medium sized libraries who combine funds and resources to license digital books from OverDrive to serve 830,000 residents. The consortium is the third-largest OverDrive library in Washington State, expending nearly $1,000,000 in 2022 on the collection.  In 2022, it saw overall costs increase by 23% while circulation increased only 4%. The limited budget of its mostly small libraries does not allow WDLC to keep pace with costs, and for 2023 the consortium agreed on a service level cut.  Because we anticipate the possibility that costs will continue to increase, we are looking for ways to deliver more efficient service, which will enable more sustainable costs without a decrease in circulation.  We are analyzing our patron behavior to see if there are ways to provide service more efficiently.
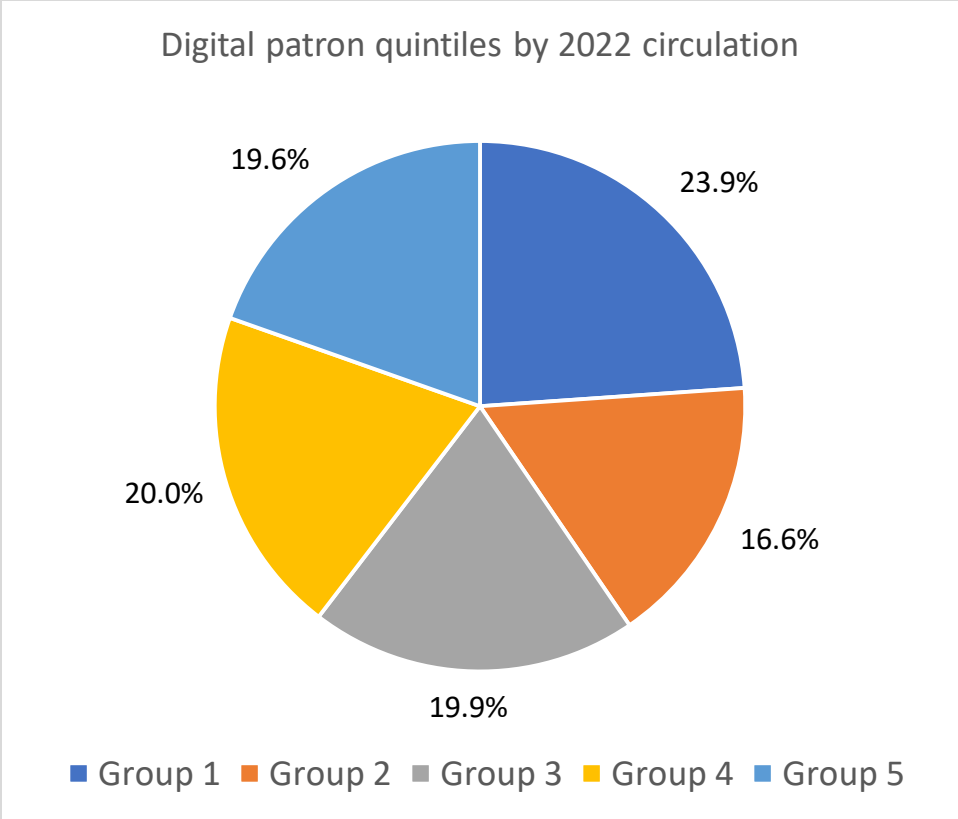
The study considers 1 year of digital checkout data from the largest library in the consortium, Whatcom County Library System (WCLS). We found that 11,899 patrons checked out a total of 350,863 eBooks, eAudiobooks and digital magazines, a representative sample for the consortium as a whole.  We segmented the patrons into approximate quintiles by circ count and compared the groups by patron age, format preference, audience preference, interest in popular books, and subject preference. We also compared the digital data to circulation on the WCLS physical collection.

## Digital patron groups

We broke the patron data into five groups of roughly equal size by the number of items they checked out in 2022.

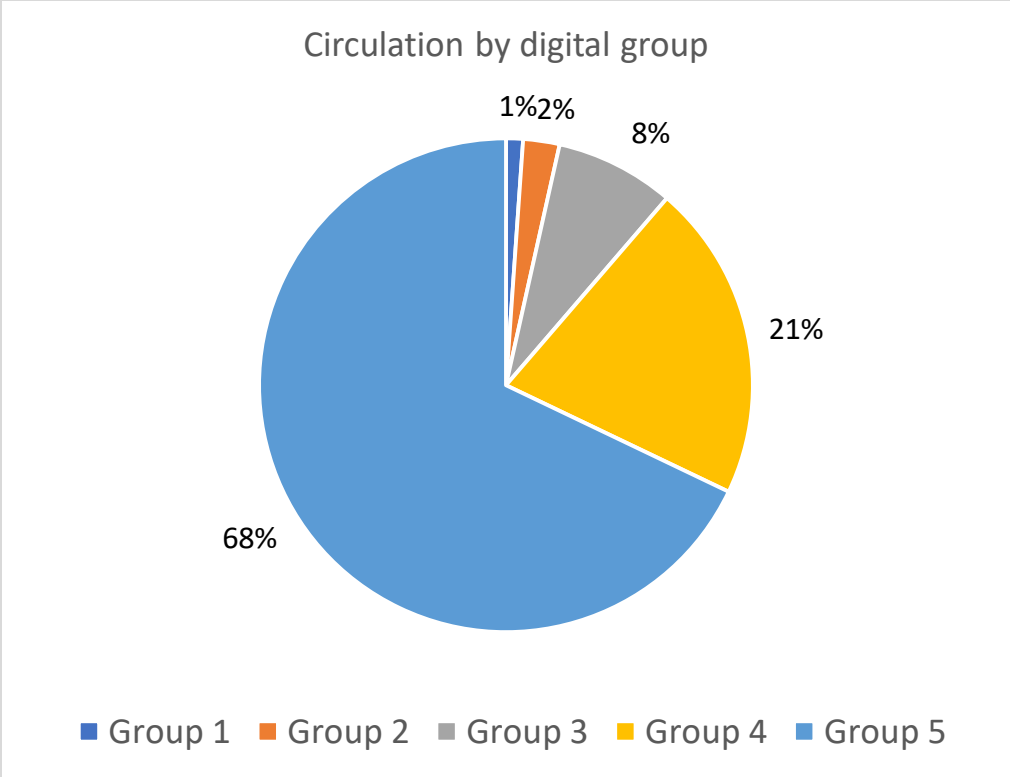| Groups | Items checked out in one year | % of total patrons |
|---|---|---|
| Group 1 | 1 – 2 | 23.90% |
| Group 2 | 3 – 6 | 16.60% |
| Group 3 | 7 – 18 | 19.90% |
| Group 4 | 19 – 46 | 20.00% |
| Group 5 | More than 46 | 19.60% |

Note: Group 1 is larger than 20% and Group 2 is smaller because strict adherence to 20% means that both Group 1 and Group 2 would include patrons who had checked out 2 items.

**Digital patron quintiles by 2022 circulation**

19.6%

23.9%

16.6%

19.9%

20.0%

- Group 1
- Group 2
- Group 3
- Group 4
- Group 5

## How much did each group circulate?

Group 5 has the most avid readers, circulating more than twice as much as all other groups put together. Group 4 and Group 5 together represent 89% of total circulation. Therefore, delivering service more efficiently to these groups should result in lower costs overall. But of course, any adjustment to the service level should not improve access for one group at the expense of another.

| Groups | Total digital circ | % of total circ |
|---|---|---|
| Group 1 | 3,898 | 1.10% |
| Group 2 | 8,361 | 2.40% |
| Group 3 | 27,425 | 7.80% |
| Group 4 | 72,949 | 20.80% |
| Group 5 | 238,230 | 67.90% |

## Circulation by digital group



The median checkout per year per group gives us an idea of how frequently patrons are circ'ing digital books. Median is the middle value in the range. In each group, half the patrons circ is below the median and half circ above.

| Groups | Digital cko in one year | Median |
|---|---|---|
| Group 1 | 1 – 2 | 1 |
| Group 2 | 3 – 6 | 4 |
| Group 3 | 7 – 18 | 11 |
| Group 4 | 19 – 46 | 30 |
| Group 5 | More than 46 | 80 |

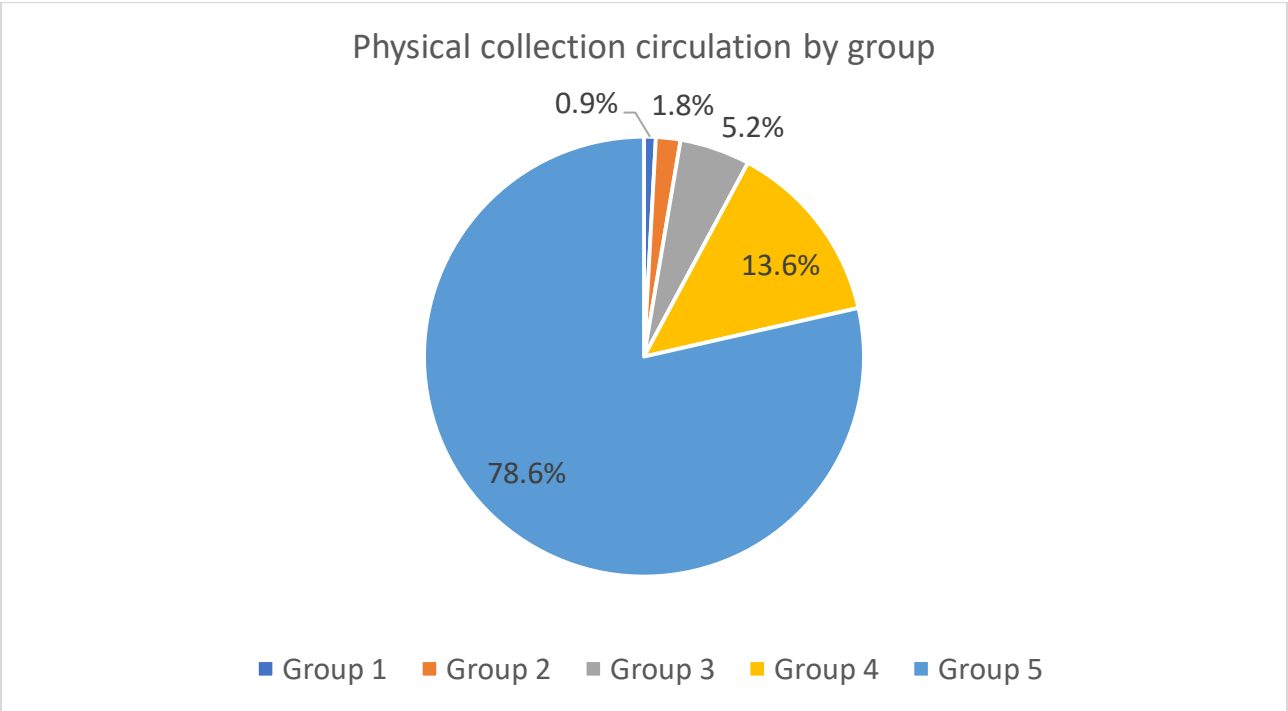# If we group patrons of the physical collection, does their circulation show a similar pattern?

To see whether checkout behavior is similar with the physical collection, we grouped those users by number of items checked out annually, and analyzed the percentage of total circ on books, audiobooks and magazines represented by each group. Circulation by group for digital patrons is repeated below for ease in comparison.

| Groups | Digital cko in one year | % of total digital circ | Physical cko in one year | % of total physical circ |
|---|---|---|---|---|
| Group 1 | 1 – 2 | 1.10% | 1 – 4 | 0.90% |
| Group 2 | 3 – 6 | 2.40% | 5 - 10 | 1.80% |
| Group 3 | 7 – 18 | 7.80% | 11 - 26 | 5.20% |
| Group 4 | 19 – 46 | 20.80% | 27 - 76 | 13.60% |
| Group 5 | More than 46 | 67.90% | More than 76 | 78.60% |

Note: Over 50% of total patrons use both the physical and digital collection, so these groups overlap.

Users of the physical collection checked out more items in every quintile than users of the digital collection, probably because the digital collection is limited to 5 items out at a time and 8 holds. By contrast, users of the physical collection may checkout 125 items at a time and place up to 75 holds. Library staff also shared anecdotally that a single patron often uses a single library account for multiple family members, which may lead to more physical circulation overall.
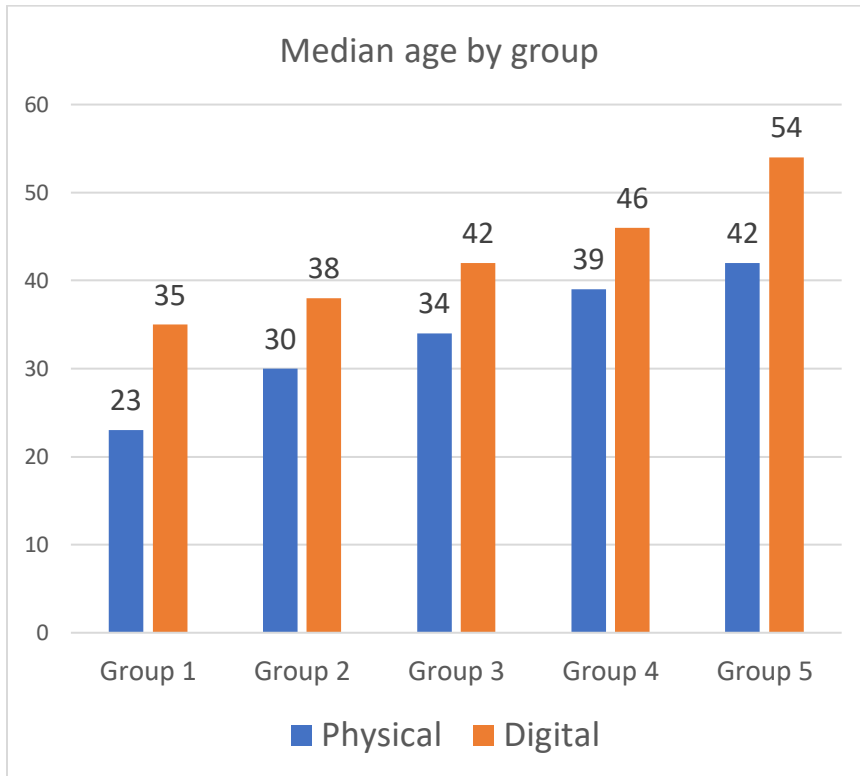
Regardless, it is notable that avid users of the physical collection are even more avid than digital book patrons. Physical groups 4 and 5 represent 92% of total circulation.

## Physical collection circulation by group

0.9%   1.8%
5.2%
13.6%
78.6%

■ Group 1   ■ Group 2   ■ Group 3   ■ Group 4   ■ Group 5

## How old are the patrons in each group?

To understand patron interests and behavior, it can be useful to get an idea of their stage in life. Therefore, we determined the median age per group for both the digital and physical collections, where we observe two patterns:

- For both, the more a group circulates, the older they are likely to be
- Patrons of the digital collection have a median age 7-12 years older than patrons of the physical collection

**Median age by group**

| Group | Physical | Digital |
|-------|----------|---------|
| Group 1 | 23 | 35 |
| Group 2 | 30 | 38 |
| Group 3 | 34 | 42 |
| Group 4 | 39 | 46 |
| Group 5 | 42 | 54 |

It would appear that readers are more likely to read more books as they age. We speculate that this may have to do with an increase in leisure time for older adults as opposed to younger adults who may be pursuing an education or have a young family.

Likewise, the tendency of younger people to use the physical collection may have to do with a preference on the part of families to minimize screen time for children.
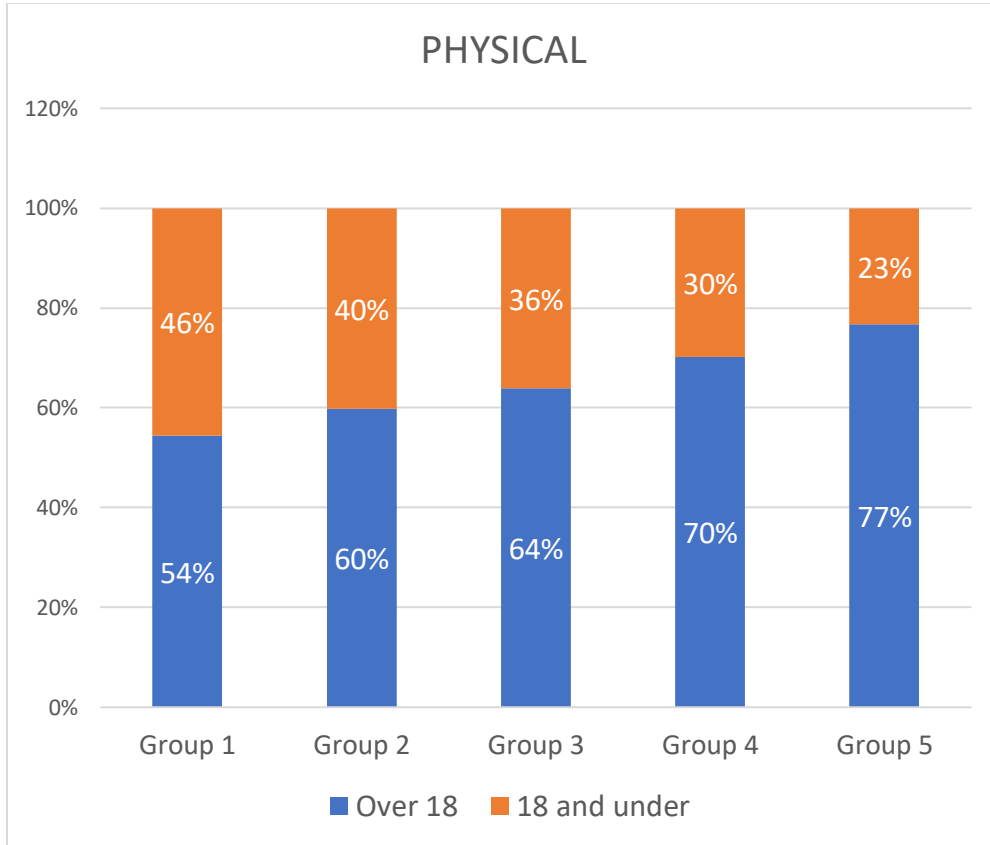
Another possibility is that older adults, who are more vulnerable to COVID-19, may have learned during the pandemic to prefer digital reading as safer than visiting a public library location. Future studies may want to investigate whether older adults eventually return to physical books.
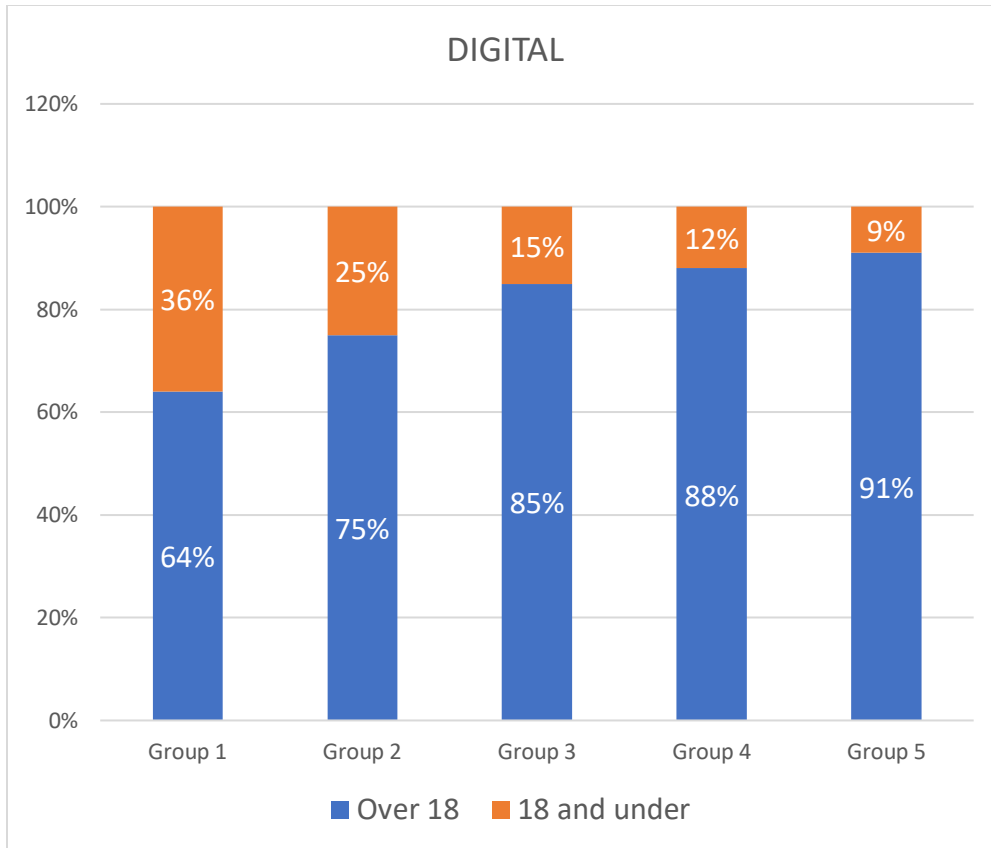
Note: all groups ranged from an age of zero to an age of 100 or more, differing only in their median.

## What percentage of circulation is attributed to people 18 and under?

Libraries often group patrons by age, offering different library accounts for minors than they offer to adults. To shed light on collection usage, therefore, we broke circulation down by group and how many patrons within the group are 18 and under. Unsurprisingly, given the data above, we see a higher

percentage of older adults correlate with higher circulation. However, it is notable that almost 1 in 4 Physical Group 5 patrons and almost 1 in 10 Digital Group 5 patrons are 18 and under. Clearly, our most avid readers include many young people and any adjustments to digital services need to account for them.

## PHYSICAL

| Group | Over 18 | 18 and under |
|-------|---------|--------------|
| Group 1 | 54% | 46% |
| Group 2 | 60% | 40% |
| Group 3 | 64% | 36% |
| Group 4 | 70% | 30% |
| Group 5 | 77% | 23% |

**DIGITAL**

(Chart showing stacked bar percentages by group)

- Group 1: Over 18 = 64%, 18 and under = 36%
- Group 2: Over 18 = 75%, 18 and under = 25%
- Group 3: Over 18 = 85%, 18 and under = 15%
- Group 4: Over 18 = 88%, 18 and under = 12%
- Group 5: Over 18 = 91%, 18 and under = 9%

Legend: ■ Over 18   ■ 18 and under

## How many digital patrons also use the print collection?

Three years out from the pandemic, we were curious to know how many of our digital patrons use our physical collection presently or have done so in the past.  Our ILS, which tracks physical circ only, enables us to see per patron:
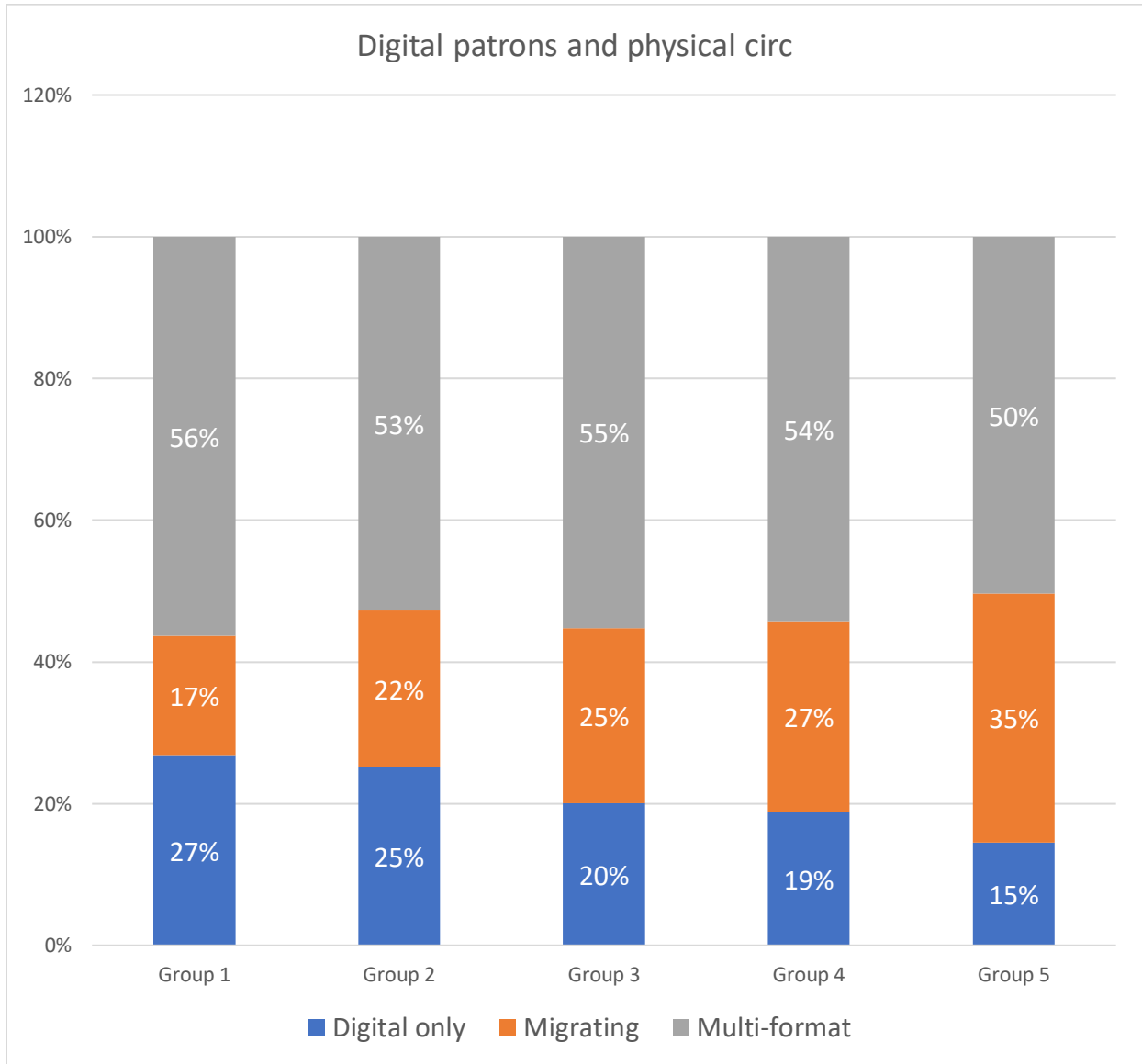
- YTD circ (which was evaluated in the end of December and represents most of the year)
- Previous Year Circ
- Lifetime Circ

Using this data, we were able to determine what percentage of digital patrons in each group:

- Digital only: Have no physical circulation at all and therefore only use digital books
- May be migrating: Have a record of physical circulation, but not in the past two years
- Multi-format patrons: Have a record of physical circulation in the past two years

The multi-format patrons, who do have circulation within the last two years, represent between 50-56% across groups, which means that 44-50% of patrons have either never visited a library location to check out a book or have stopped doing so.  Most digital only patrons are in Group 1, which circulates only 1-2 digital items per year. So these are patrons who are experimenting with library service. Group 1 also has the highest percentage of multi-format patrons suggesting that they may be regular users of the physical collection who tried digital books, but prefer print overall.

The highest percentage of patrons who may have migrated from physical to digital formats are in Group 5, with 35%. So, the pandemic may have "hooked" avid readers on digital books.
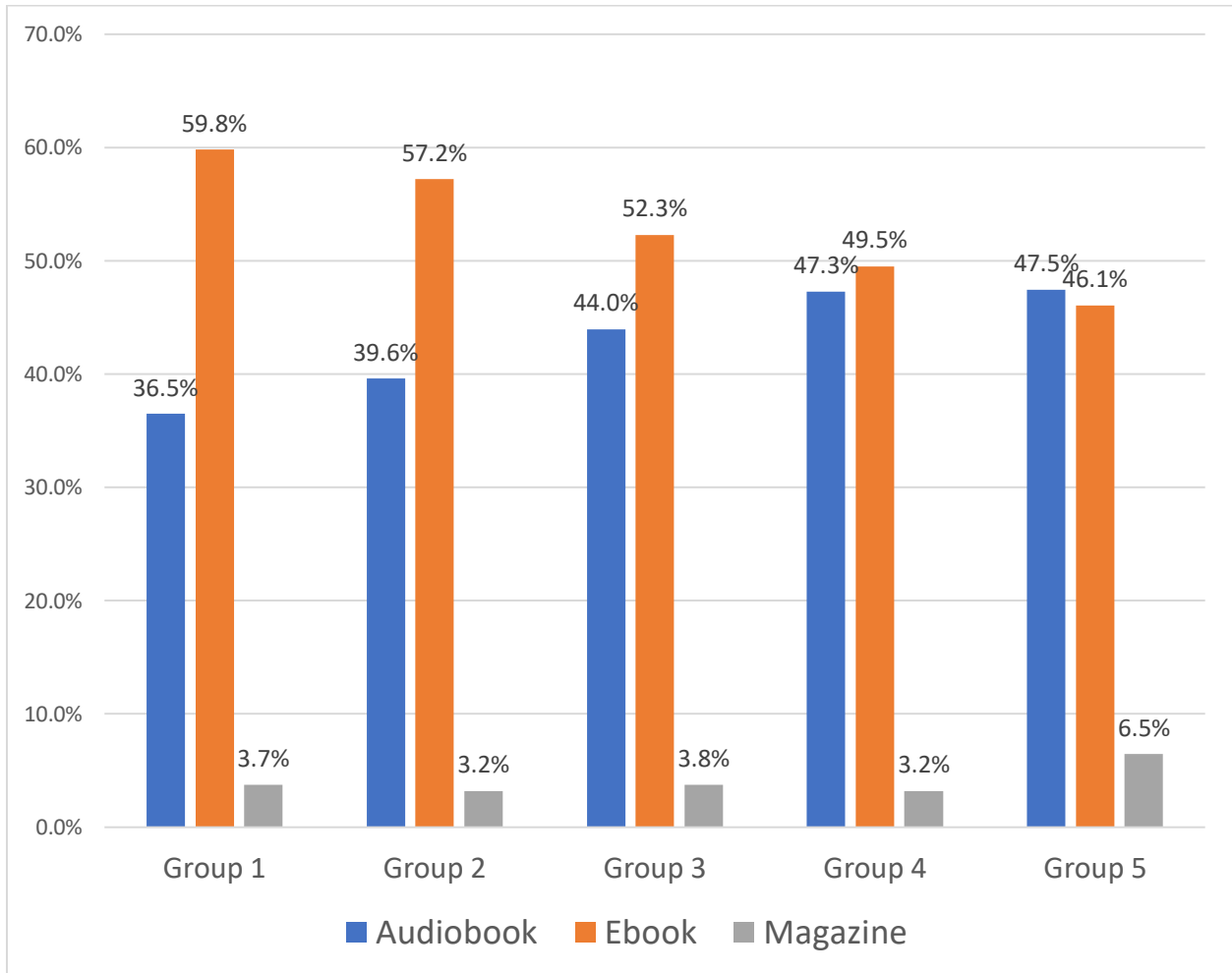


Digital patrons and physical circ

## What formats do digital groups prefer?

WDLC offers patrons eBooks, eAudiobooks, and digital magazines, so we analyzed circulation for each format for each group. As with patron age, we see a slope by group for eBooks and eAudiobooks. The more people checkout books the more they prefer eAudio.

We cannot usefully speculate on why this might be since it is not necessarily intuitive that a person who checks out 8 books per year would prefer eBooks and a person who checks out 48 per year would prefer eAudio. Possibly the more avid a reader is, the more inclined they are to a format that allows them to multi-task, reading at the same time that they exercise, commute or do household tasks.

This data is worth watching as eAudio usage continues to grow.

The other datapoint of note here is that digital magazines are of approximately equal interest to Groups 1 – 4, while Group 5 circulated almost twice as many as other groups. Digital magazines are always available to patrons, with no wait times, and also their checkouts do not count against the consortium limit.  So Group 5's unusual interest in digital magazines may suggest that they are more interested in readily available content than other groups.



Note: we did not pull this data for the physical collection because Audiobooks on CD represent a small and shrinking portion of our collection.

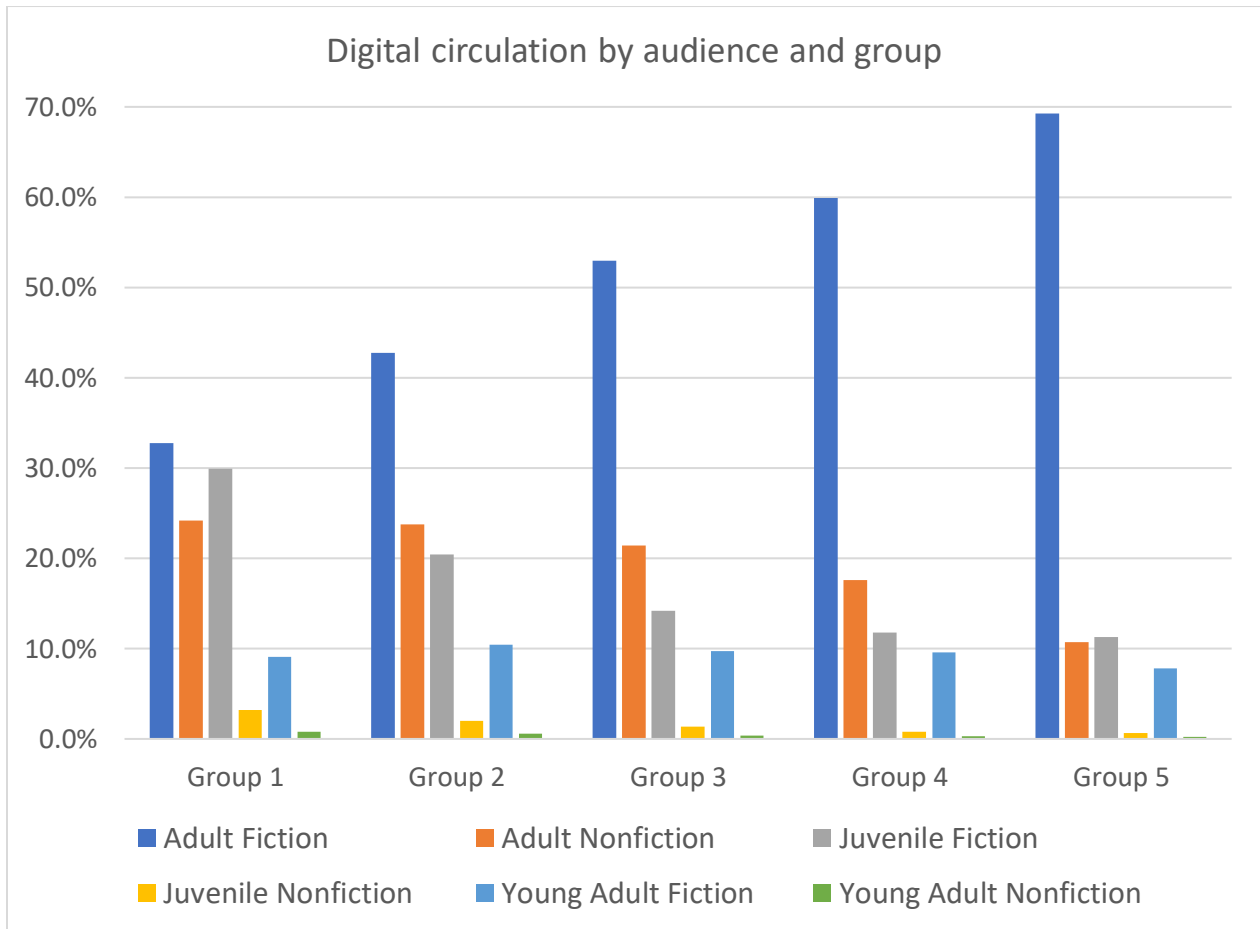## What audiences do the digital groups prefer?

OverDrive applies one of six audiences to most eBooks and eAudiobooks.

- Adult Fiction
- Adult Nonfiction
- Juvenile Fiction
- Juvenile Nonfiction
- Young Adult Fiction

- Young Adult Nonfiction

We segmented circulation data by audience and group to see if there were useful patterns. We found a distinct sloping preference for five of the six audiences.  Group 1 is the most balanced of the groups in the range of audiences it selected whereas Group 5 is the least balanced with an overwhelming preference for Adult Fiction.

Young Adult Fiction is the only audience without a distinct slope and appears equally interesting to all groups.
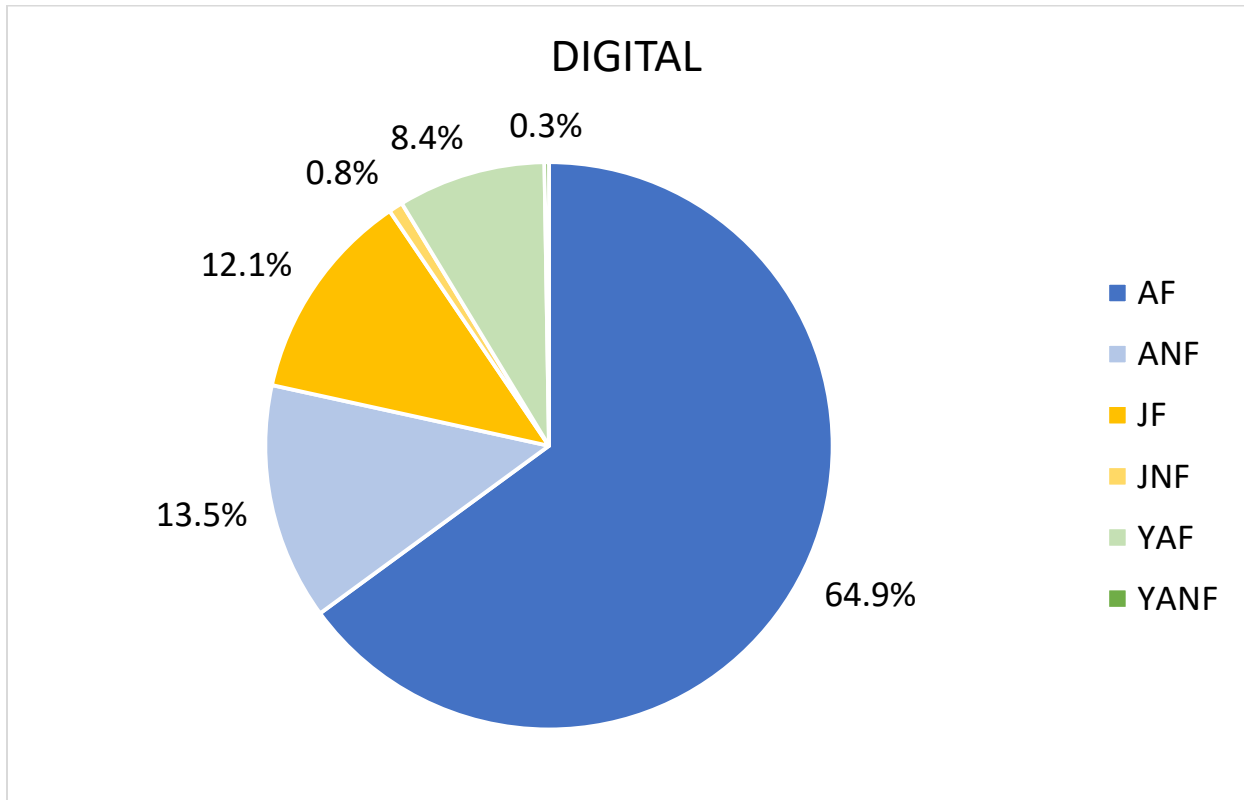


Digital circulation by audience and group

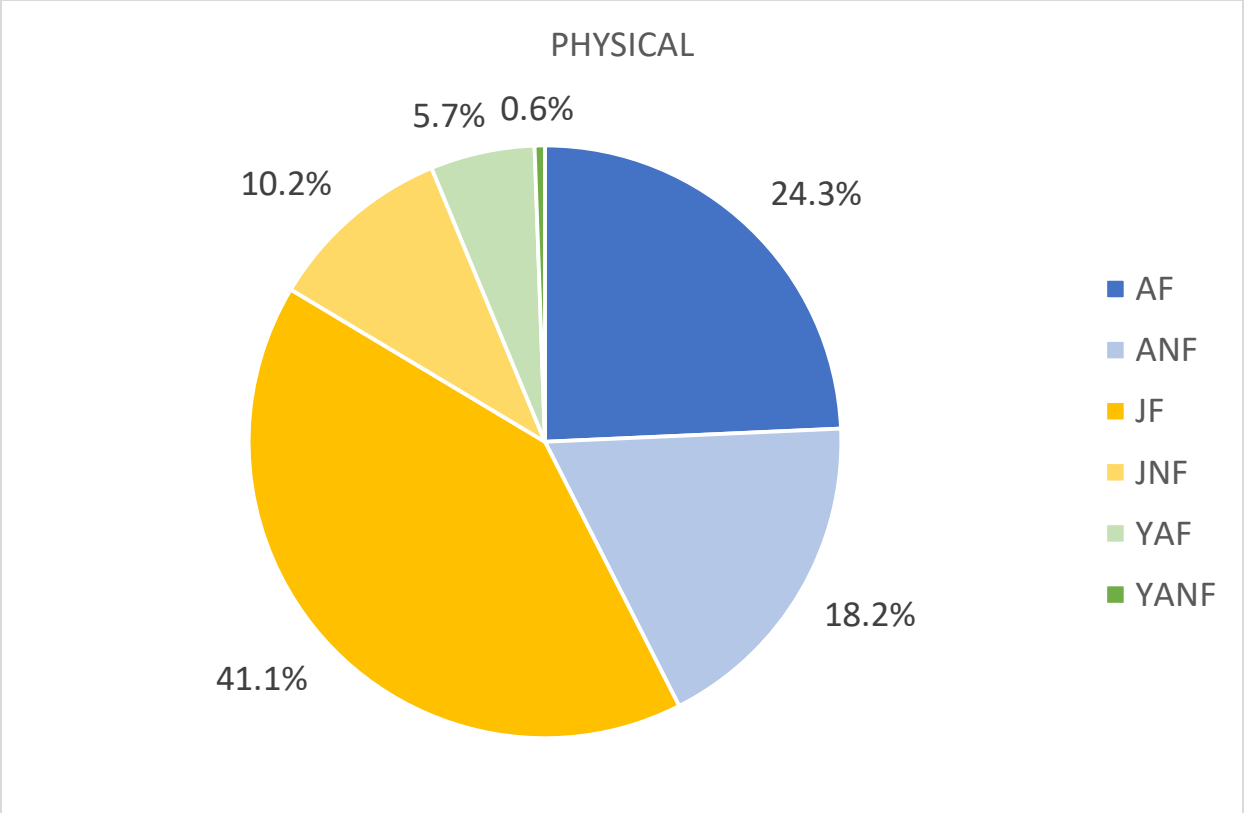## How does digital circulation by audience compare to physical circulation by audience?

To determine if the audience preference above is specific to digital lending or is a pattern we can expect in physical lending too, we compare total circ by audience for the digital and physical collections.  The blue colors represent Adult Fiction and Nonfiction, yellow represents Juvenile Fiction and Nonfiction, and green represents Young Adult Fiction and Nonfiction.

There is a dramatic difference between the proportion of circulation dedicated to Adult Fiction and Nonfiction, which is 78% of total digital circulation and only 42.5% of physical circulation. Alternately,

Juvenile Fiction and Nonfiction dominate the physical circulation with 51% and just 13% of digital circulation. Young Adult titles are a comparable percentage at 8.7% digital and 6.1% physical.

Since the patrons of the physical collection tend to be younger and represent more youth accounts, it makes sense in one way that Juvenile Fiction and Nonfiction see so much more usage in the physical collection.  Anecdotally, library staff reviewing the data reported that many families prefer to minimize screen time for their young people.



DIGITAL

- AF
- ANF
- JF
- JNF
- YAF
- YANF

8.4%  0.3%
0.8%
12.1%
13.5%
64.9%

PHYSICAL

- AF — 24.3%
- ANF — 18.2%
- JF — 41.1%
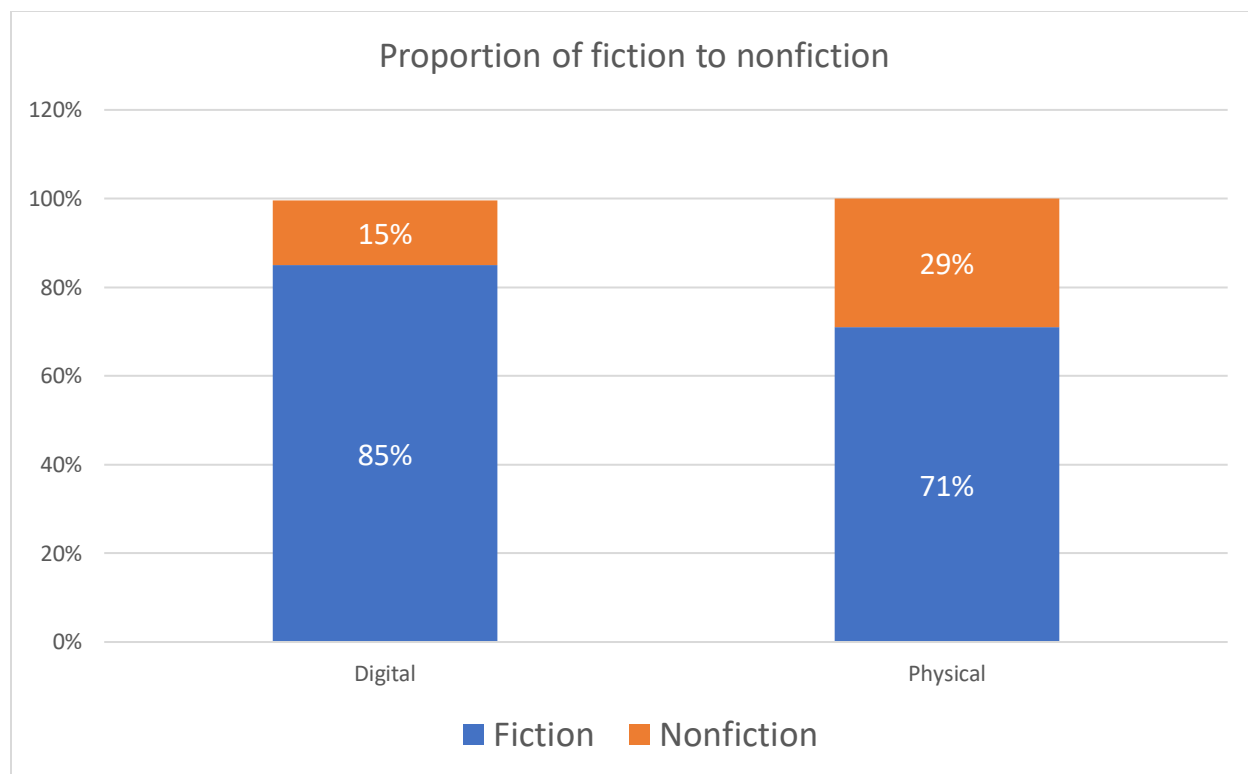- JNF — 10.2%
- YAF — 5.7%
- YANF — 0.6%

## How does the proportion of fiction to nonfiction compare?

Both digital and physical patrons prefer fiction at all audience levels. However, digital patrons have a stronger preference for fiction, again at all audience levels.

We speculate that a preference for digital fiction may relate to the scientific studies showing that digital books are not as efficient at conveying information. People who are reading for pleasure may find the digital book adequate for their reading experience, but will lean to print when reading for information.

Proportion of fiction to nonfiction

## What subjects or genres do the digital groups prefer?

A preference for fiction genres is also apparent for all five digital groups when we look at the circulation data associated with the most common subjects that OverDrive uses to classify digital books. Specifically, the heatmap shows what percentage each group's circulation includes titles with each subject. So, for example, all groups have approximately the same level of interest in Christian Fiction, which represents 1.5-1.6% of total circulation across groups.

Note: OverDrive subject metadata is applied programmatically based on the BISAC headings applied to a work by its publisher. OverDrive subjects are not selected by librarians and their quality and accuracy is not up to the standard set for most print collections and any analysis of group subject preferences should keep that in mind.

Also, every title has multiple subjects.

Given the shortcomings of the data quality, some observations are as follows:

- As noted above, fiction subjects are by all groups over nonfiction, with Biography & Autobiography, the first nonfiction subject, ranking 12th overall.
- Groups 4 and 5 have roughly an equal interest in "Literature" (a subject generally applied to literary fiction).
- Group 5 shows a preference for genre fiction: romance, mysteries, and thrillers, and it would be interesting to study further whether there is a correlation between avid reading and genre fiction preference.

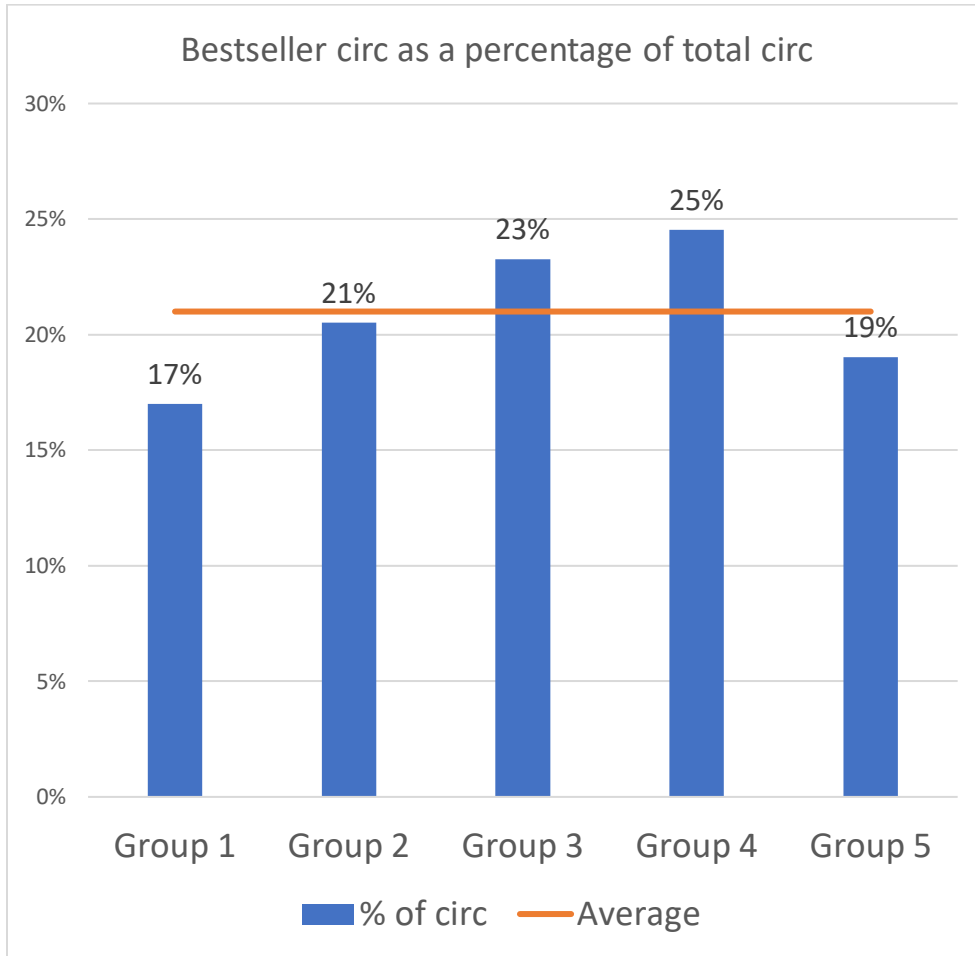|  | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 |
|---|---|---|---|---|---|
| Literature | 7.0% | 9.4% | 11.1% | 13.0% | 12.9% |
| Thriller | 5.6% | 6.4% | 8.7% | 9.6% | 11.1% |
| Romance | 4.3% | 6.8% | 7.8% | 9.5% | 11.4% |
| Mystery | 4.5% | 4.6% | 6.0% | 6.8% | 10.0% |
| Suspense | 4.0% | 4.0% | 6.1% | 6.9% | 8.5% |
| Historical Fiction | 4.4% | 5.3% | 6.3% | 6.9% | 6.9% |
| Fantasy | 7.1% | 8.3% | 8.2% | 7.8% | 6.6% |
| Juvenile Fiction | 15.2% | 10.3% | 7.2% | 5.8% | 5.4% |
| Young Adult Fiction | 4.6% | 5.3% | 4.9% | 4.7% | 3.7% |
| Humor (Fiction) | 5.3% | 4.7% | 3.9% | 4.1% | 4.4% |
| Juvenile Literature | 12.1% | 7.8% | 5.3% | 4.4% | 3.6% |
| Biography & Autobiography | 4.0% | 3.7% | 3.5% | 3.0% | 1.9% |
| Science Fiction | 2.8% | 3.6% | 3.4% | 2.9% | 2.5% |
| Young Adult Literature | 2.7% | 3.1% | 2.7% | 2.6% | 2.1% |
| History | 2.0% | 2.5% | 2.1% | 1.8% | 1.2% |
| Comic and Graphic Books | 3.0% | 2.5% | 1.6% | 1.2% | 1.8% |
| Self-Improvement | 2.6% | 2.5% | 2.4% | 1.8% | 0.9% |
| Classic Literature | 1.6% | 2.1% | 1.9% | 1.6% | 0.9% |
| Sociology | 1.9% | 1.6% | 1.5% | 1.3% | 0.8% |
| Christian Fiction | 1.5% | 1.6% | 1.7% | 1.6% | 1.6% |
| Science | 1.1% | 1.1% | 1.0% | 0.7% | 0.6% |
| Psychology | 1.4% | 1.6% | 1.5% | 1.0% | 0.5% |
| Mythology | 1.3% | 1.1% | 1.0% | 1.1% | 0.7% |

## Which digital groups are most interested in popular books?

We are interested in behavior related to the most popular books because they are the most expensive and may present opportunities for improving the efficiency of the service. "Bestseller" is here defined as a title in the top 10% of the digital collection by hold count at the time the circulation statistics were compiled. The average price of a bestseller license is $42.78 while the overall average price of a license is $30.30.

For all groups, bestsellers represented 21% of total circ, but percentages by group vary. The usual slope that we see across the groups is disrupted as Group 5, which circulates the most books, is the second least likely to choose bestsellers.

Importantly, Group 5 circulates more bestsellers than any other group simply because they circ so much, but a book's popularity is not as important to them as it is to Groups 2, 3, and 4. Group 5 may actually

prefer bestsellers, but they are avid readers, they may busy themselves with less popular books while they wait on the holds queue for bestsellers.

**Bestseller circ as a percentage of total circ**

| Group | % of circ |
|-------|-----------|
| Group 1 | 17% |
| Group 2 | 21% |
| Group 3 | 23% |
| Group 4 | 25% |
| Group 5 | 19% |

Legend: ■ % of circ — Average

## What does the data tell us about the problem we are trying to solve?

Is it possible to provide service more efficiently?  In library terms, is it possible to maintain or grow circ levels while reducing cost per circ?

Group 5 is the group to consider targeting for more efficient service since they circulate so much. Any change to the service that improves it for them will improve efficiency overall.  However, any change to the service must carefully be evaluated as to its impact to other groups.

Still, we are exploring some optimizations with Groups 4 and 5 in mind and monitoring our costs to see whether the optimizations help.  The two Groups share an interest in eAudio formats and a preference for Adult Fiction, especially genre fiction.  Group 4 is interested in popular items while Group 5 seems to be compelled more by what is available than what is popular.  As such we have:

- Licensed the Blackstone and Tantor eAudio Simultaneous Use packages, 50 titles each, with an emphasis on adult fiction.  We promote them in a curated list that is always in the top 10.  After 6 months, the cost per circ is at or below our average cost per circ for the collection.
- Used curated lists more intentionally, including adult fiction titles, lots of eAudio, and at least 300 titles in available status when the list runs.  We run the lists for two weeks, adding more if available titles drop under ~50.  We use the Insights > Patron Interests area on Marketplace to track the performance of each curated list, noting the total checkouts and holds the list generates.  This area of Marketplace only tracks transactions in Libby, but can be used to compare different lists "Apples to Apples."  Our most successful lists as measured by a high ratio of checkouts to holds
- Use curated lists overtly to promote digital magazines, since their instant availability is appealing to Group 5, updating every two weeks.  Curated lists work better for magazines than automated lists, but it is possible to put together a list quickly using magazine subject headings.

We also plan to explore with selectors whether the collection should rebalance somewhat in accordance with patron behavior. For example, if Adult Nonfiction represents an estimated 30% of our collections total copies but 13.5% of the circ.  Given that print formats may convey information more effectively, it might make sense to licenses fewer titles and or copies of nonfiction titles as a proportion of the collection and put that money toward portions of the collection where the percentage of circ exceeds the percentage of copies.