

## **Who Uses Libraries In-Person and Online?**

**Kate Criswell, William Lee, Sofia Zarazagas Del Saz, Alexis Hancz**

### **Introduction**

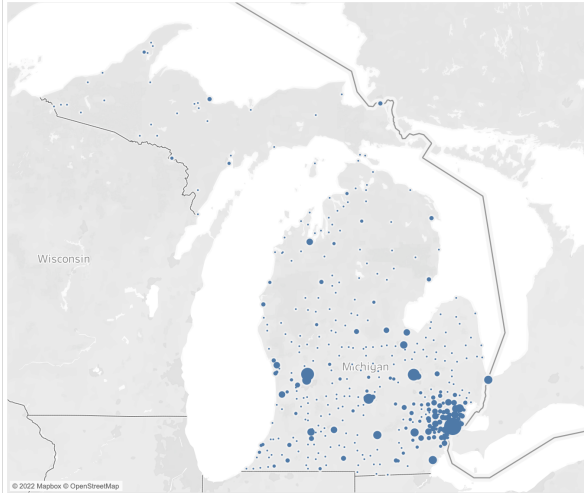
As a team we have been given data on different demographics' access and use of public libraries and their resources. We are evaluating how these different subsets of the population interact with these resources to better inform analysts on how to allocate funding towards improving access and quality of public library internet resources. Comparing averages within demographics we will evaluate significant inequalities in usage and access to information and the internet provided by libraries.

### **Description of Data Sources**

Our team is using two data sets that present information about library location, resource level and usage. Data about location and resource level are provided by the Institute of Museum and Library Services which is a federally supported and funded organization. They collected data through voluntary consensus in the year 2020. Data about library usage was collected by Pew Research Center which is an independent nonprofit organization funded by the Pew Charitable Trust. For this data set specifically, they collected data through surveys from 2011 to 2016. We looked at 8 variables from the usage data provided by Pew Research Center for our analysis on demographic relationships. We used library location and resource level from the Institute of Museum and Library Services to map where libraries are located in Michigan.

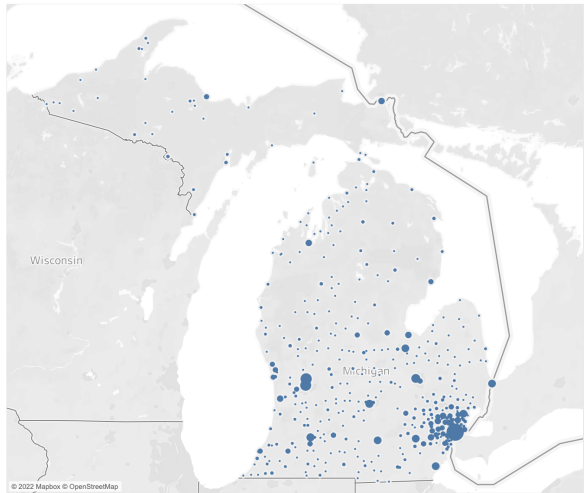
## Resource Analysis

Populations Surrounding Established Library Systems



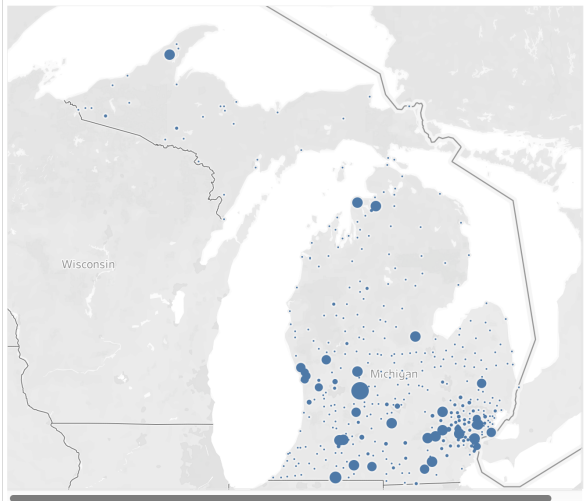
Summary	
Count:	396
SUM(Popu Lsa)	
Average:	24,892.92
Median:	9,705.50
Standard deviation:	52,951
Population per Established Library System	
+	492
●	200,000
●	400,000
●	600,000
●	713,777

Book Volume of Libraries Across Michigan



Summary	
Count:	396
SUM(Bkvol)	
Average:	63,944.41
Median:	33,260.50
Standard deviation:	117,454
Book Volume per Library	
+	6,006
●	500,000
●	1,000,000
●	1,500,000
●	1,760,316

Electronic Book Volume of Libraries Across Michigan



Summary	
Count:	396
SUM(Ebook)	
Average:	59,130.19
Median:	18,430.00
Standard deviation:	139,349
Electronic Book Volume per Library	
+	0
●	500,000
●	1,000,000
●	1,494,387

Using the data provided by the Institute of Museum and Library Services we compared the populations, volume of books and volume of Ebooks for the state of Michigan. When comparing each variable, we found that the median book volume (33260 books) was greater than the median ebook volume (18430 Ebooks) within Michigan. We also observe a cluster of library locations around the Detroit area which has a larger population than other areas in Michigan. When looking at the difference in Ebook volume for library systems within this cluster, we observe that physical books are more easily accessible than Ebooks as seen through dot size on our map. Additionally, we observe that Ebook distribution across the state is not representative of the spread of population. Meaning, areas with access to higher amounts of Ebooks are not necessarily areas with higher populations. However, when observing physical book volume we see that the spread of access is similar to that of the spread of population across the state. Lastly, we looked at the standard deviation of book and Ebook volume. The standard deviation for book volume is 117453 which is lower than the standard deviation for Ebook volume which was 139348. Meaning, that Ebook volume has a larger distribution than book volume. This can be observed in the maps by the larger number and median size of dots for Ebook volume compared to physical book volume.

## Usage Analysis

### Univariate Descriptive Statistical Analysis

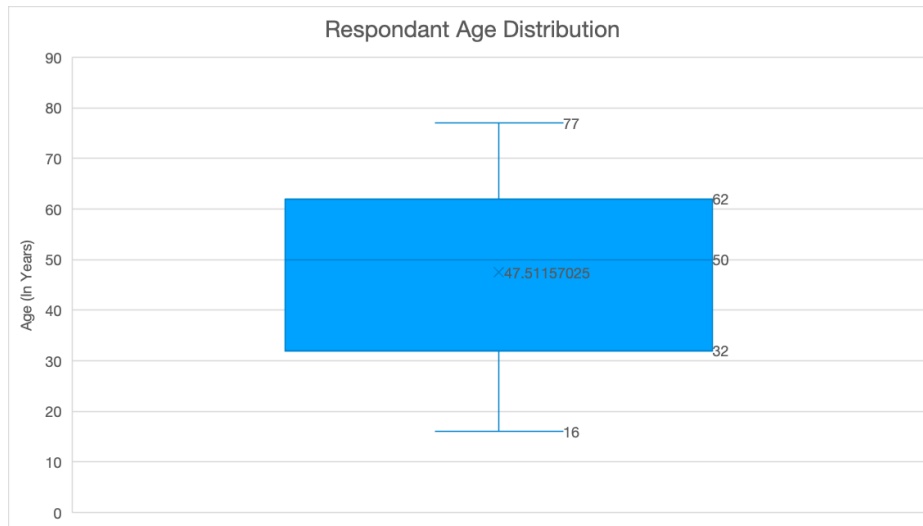
As seen through the pivot table below, more people within our sample did not use a public library website ( $x=809$ ) than those who did use a public library website ( $x=792$ ). Additionally, more people within our sample did not use a public library in person ( $x=1146$ ) than those who did use a public library in person ( $x=455$ ). However, we can see that more people used public library websites ( $x=792$ ) than those who used public libraries in person ( $x=455$ ).

		Use of public library or bookmobile in person over the past 12 months		
		Y	N	TOTAL
Use of public library website over the past 12 months	Y	389	403	792
	N	66	743	809
	TOTAL	455	1146	1601

Within our sample we looked at the variable Sex. 52.03% of our sample is male and 47.97% is female. This distribution is fairly even and will allow use to examine sex as a characteristic of those using library resources.

Row Labels	Count of Respondent Sex	Percent of Total
Female	768	47.97%
Male	833	52.03%
<b>Grand Total</b>	<b>1601</b>	<b>100.00%</b>

Within our sample we looked at the age of participants. The median age was 50 years old. Ages ranged from 16 to 77 years old. The middle 50% of participants were between the ages 32 and 62. This spread of ages will allow for a good evaluation of age as a factor in library usage.



Next, we will look at the number of books respondents have read in the last 12 months. As seen in the box plot, the median is 4 books. This variable has values that are the most widely distributed compared to the other characteristics being considered. This variable will allow us to see how more frequent readers compare to non frequent readers in their access to libraries.



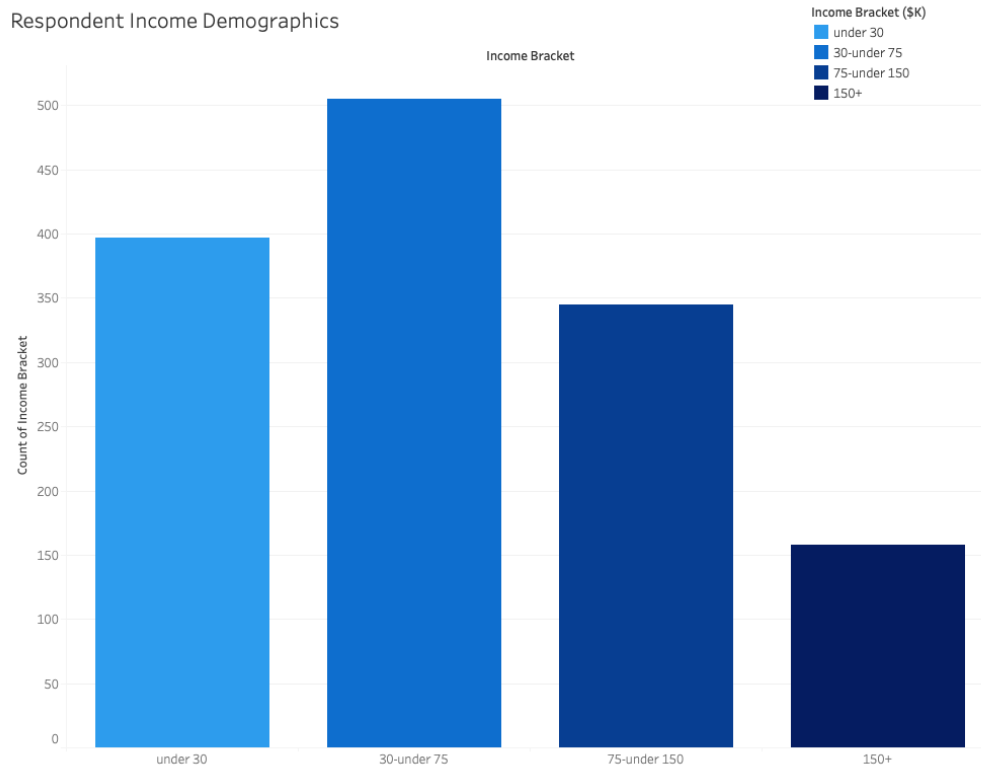
We also looked at the number of respondents that were parents. This looked at participants that were parents to any child under the age of 18. 75.58% of respondents were not parents and 24.42% of respondents were parents. This characteristic was considered important because having children could influence a participant's want or need for library resources.

Row Labels	Count of Respondent Parental Status	Percent of Total
No	1210	75.58%
Yes	391	24.42%
<b>Grand Total</b>	<b>1601</b>	<b>100.00%</b>

Next we will look at community type. We gave respondents a range from living in a large city (1) to living in a rural area (4). As seen above, the majority of participants live in a small city or town. This variable will allow us to look at geographic location and its effects on library access and usage.

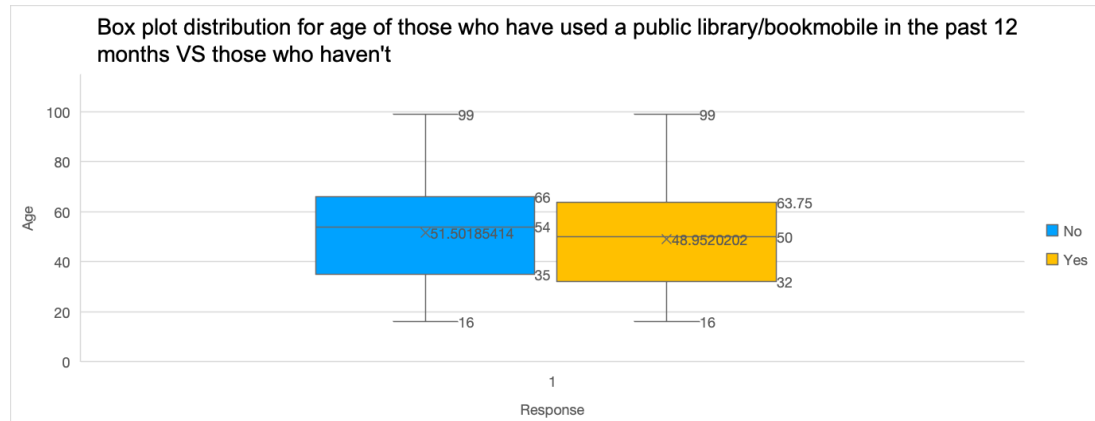
Row Labels	Count of Respondent Rurality / Community Type	Percent of Total
Small city or town	599	37.41%
Suburb near large city	355	22.17%
Large city	353	22.05%
Rural area	294	18.36%
<b>Grand Total</b>	<b>1601</b>	<b>100.00%</b>

Lastly, we looked at the variable income. We grouped participants into four groups. As shown in the pivot table the majority of participants are in the bracket \$30,000 to under \$75,000. This variable will allow us to analyze socioeconomic status within participants. We can look at how annual income affects access and library usage.



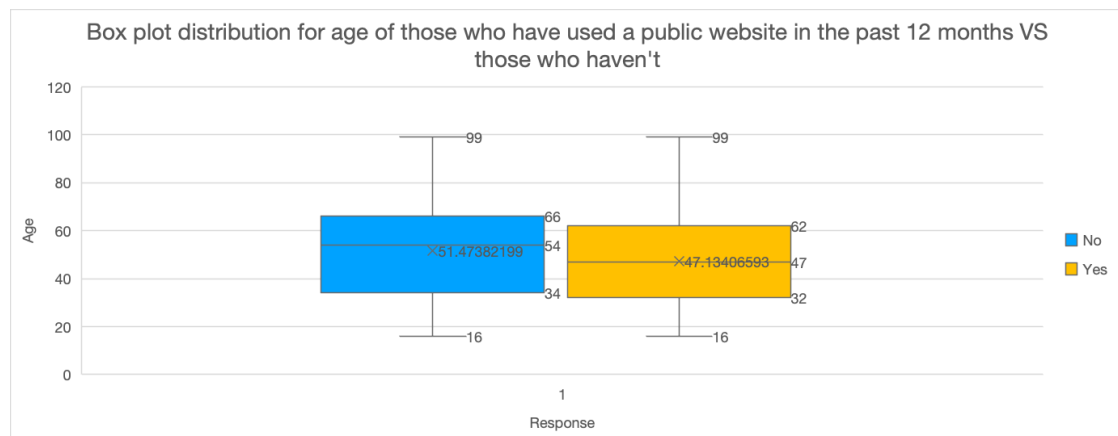
## Bivariate Relationship Analysis

### T Test



p=0.0835

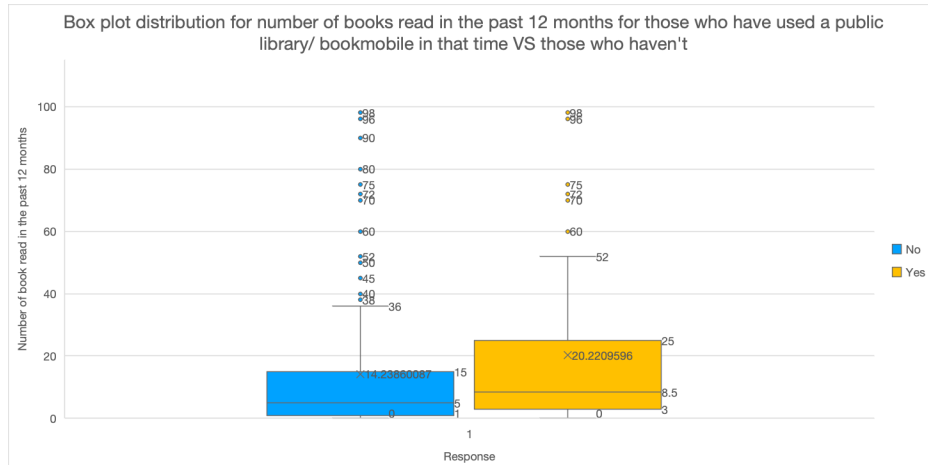
We believe age will be related to in person library usage such that older people will be the highest users. We expect younger people to have more knowledge on how to use the internet, meaning they would not need to come in person to a library for books or internet access. Older persons may not have the same level of knowledge and may need in person resources to access books and the internet. The model containing the variables “age” and “public library/book mobile usage in the past 12 months” is not statistically significant (p-value=0.0835). This suggests no relationship between age and in person public library usage. (t=1.73, df=1, p-value=0.0835)



p=0.372

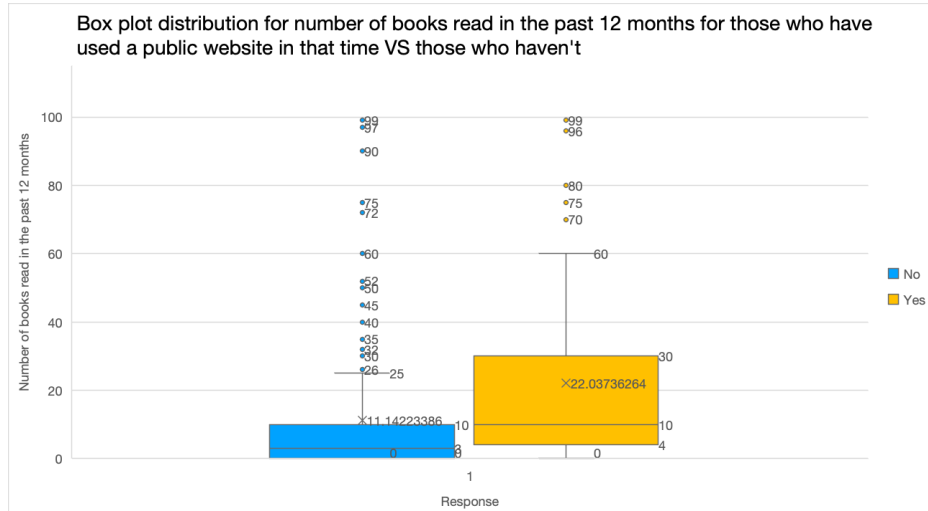
We believe age will be related to public website usage such that younger people will be the highest users. We expect older people to have less knowledge on how to use the internet, meaning they would need to come in person to a library for books or internet access. We expect younger people to have a higher level of knowledge and not need in person resources to access books. The model containing the variables “age” and “online library usage in the past 12 months” is not statistically significant (p-value=0.372). This suggests no relationship between age and in person public library usage. (t=0.893, df=1, p-value=0.372)





$p=6.86e-28$

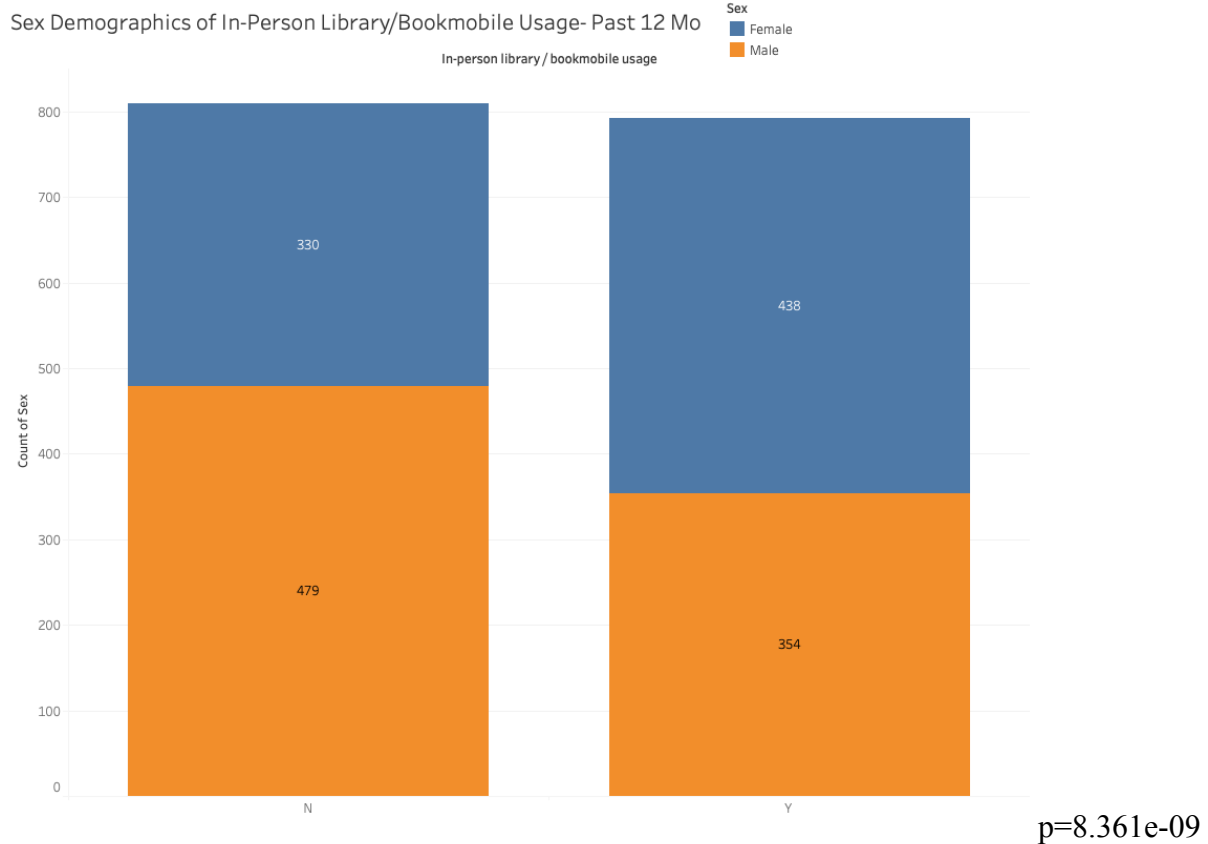
We expect the number of books read in the last 12 months to be related to in-person library usage such that people who read more books would have the highest usage. We expect people who read more to need more books and thus go in-person to libraries. Additionally, we expect those who enjoy reading to want to see and converse with other like minded people. The model containing the variables “number of books read in the past 12 months” and “public library/book mobile usage in the past 12 months” is statistically significant ( $p\text{-value}=6.86e-28$ ). This suggests some relationship between number of books read and in-person library usage. ( $t=-11.2$ ,  $df=1$ ,  $p\text{-value}=6.86e-28$ )



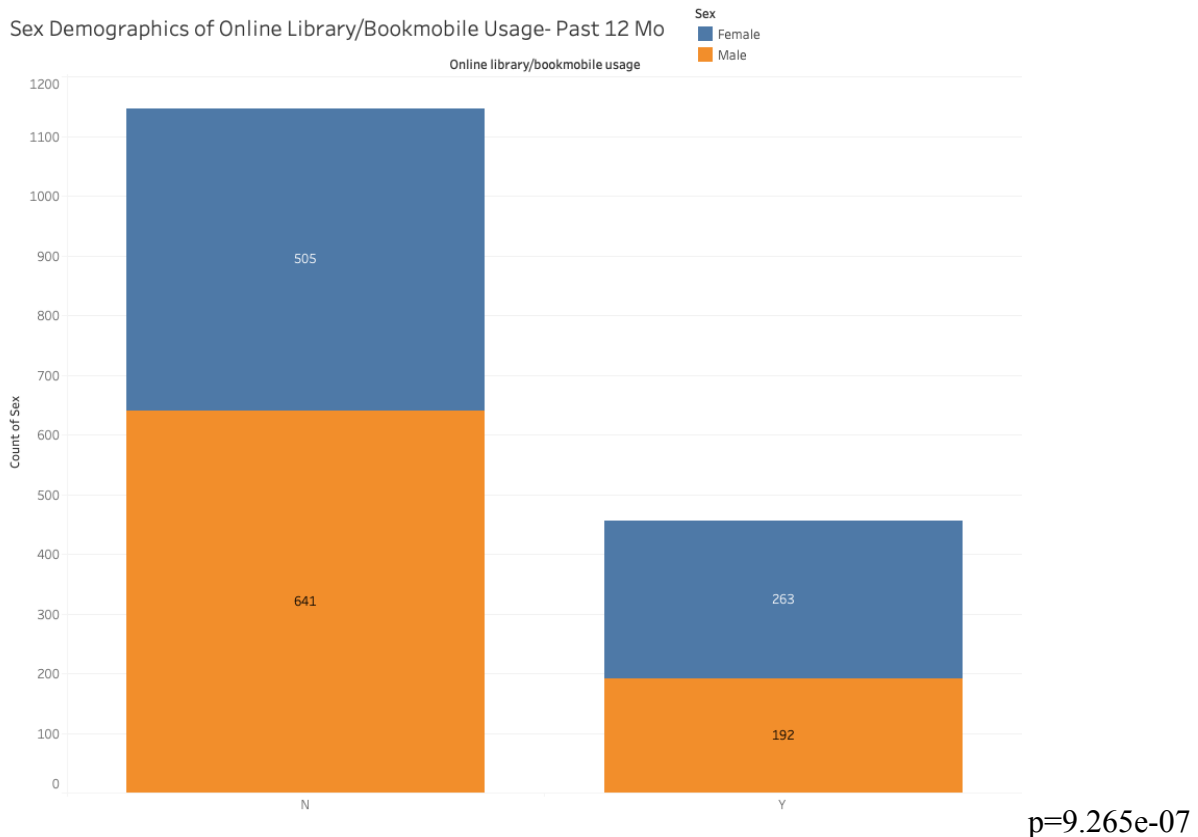
$p=1.96e-07$

We expect the number of books read in the last 12 months to be related to online library usage such that people who read more books would have the highest usage. We expect people who read more to need more books and thus use online library resources. Additionally, we expect those who enjoy reading to find online resources more convenient than going in person. The model containing the variables “number of books read in the past 12 months” and “online usage in the past 12 months” is statistically significant ( $p\text{-value}=1.96e-07$ ). This suggests some relationship between number of books read and online library usage. ( $t=-8.75$ ,  $df=1$ ,  $p\text{-value}=1.96e-07$ )

## Chi-Squared Test

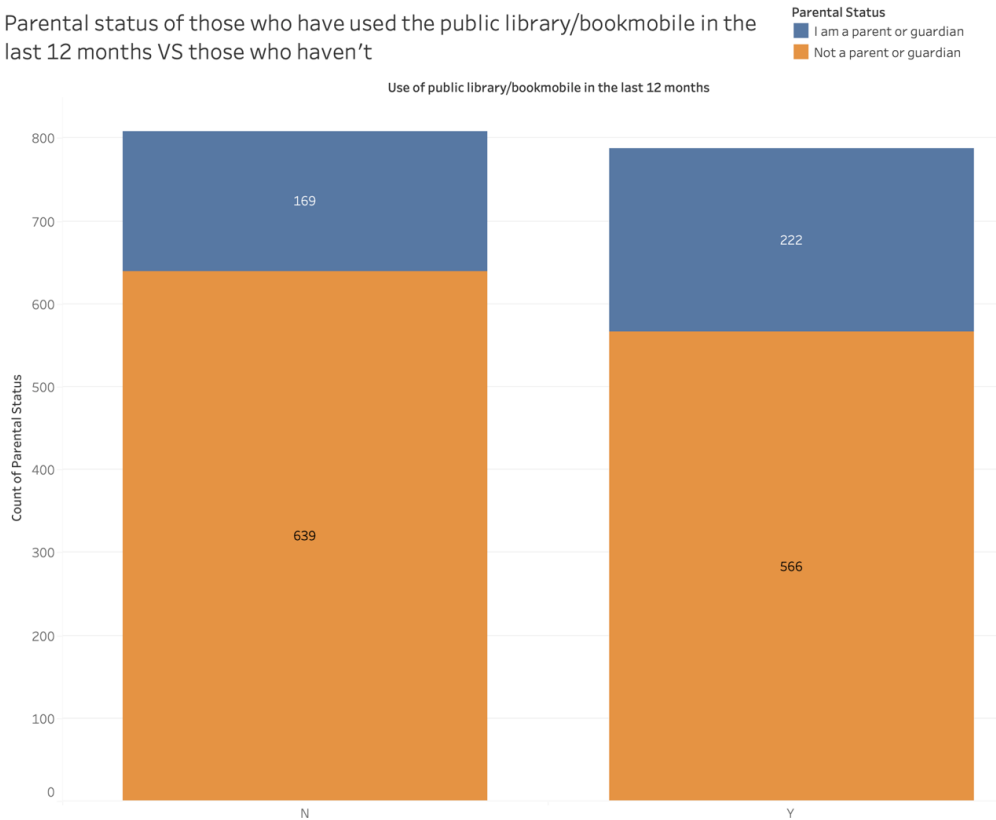


We believe that sex will be related to in-person library usage such that females will have the highest usage. We expect women to have a higher interest in reading and therefore seek out library resources more frequently. Additionally, we expect men to need library resources but not seek out in-person library resources. The model containing the variables “sex” and “public library/book mobile usage in the past 12 months” is statistically significant. This suggest a relationship between sex and in person library usage. (x-squared= 33.189, df=1, p-value= 8.361e-09)



We believe that sex will be related to online library usage such that males will have the highest usage. We expect women to have a higher interest in reading and therefore seek out in-person library resources more frequently. Additionally, we expect men to need library resources but not seek out in-person library resources, thus having a higher amount of online usage. The model containing the variables “sex” and “online usage in the past 12 months” is statistically significant. This suggest a relationship between sex and online library usage. (x-squared= 24.073, df=1, p-value= 9.265e-07).

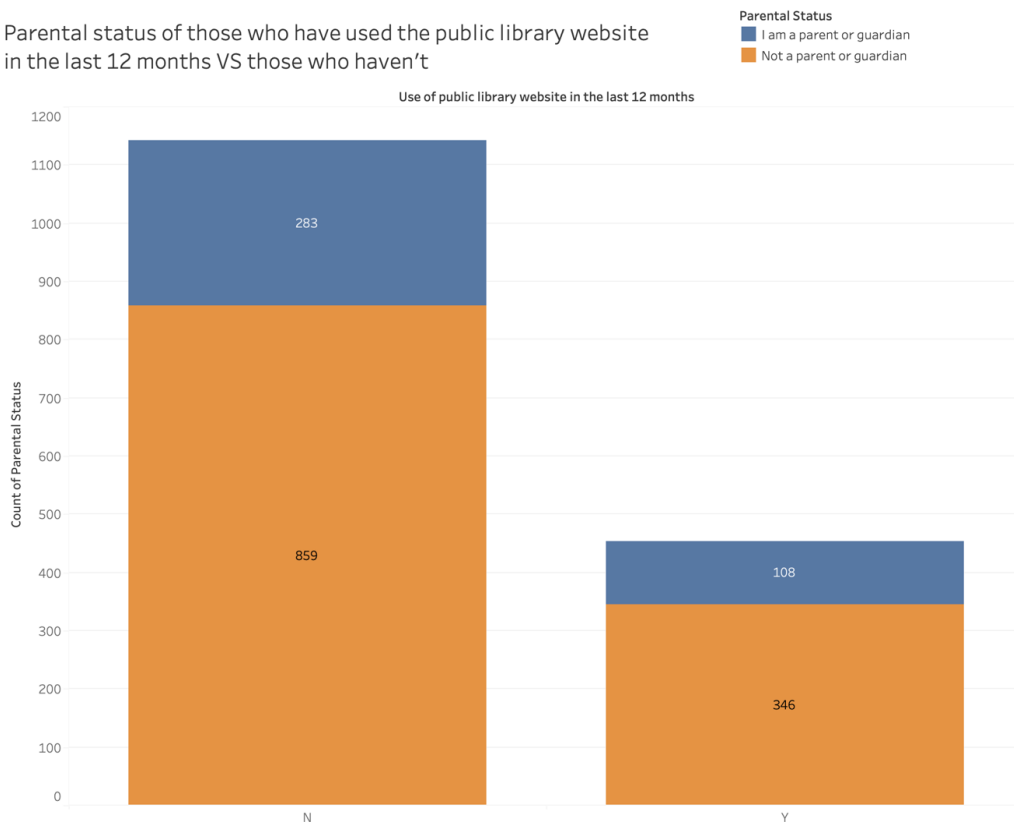
Parental status of those who have used the public library/bookmobile in the last 12 months VS those who haven't



p=0.0009266

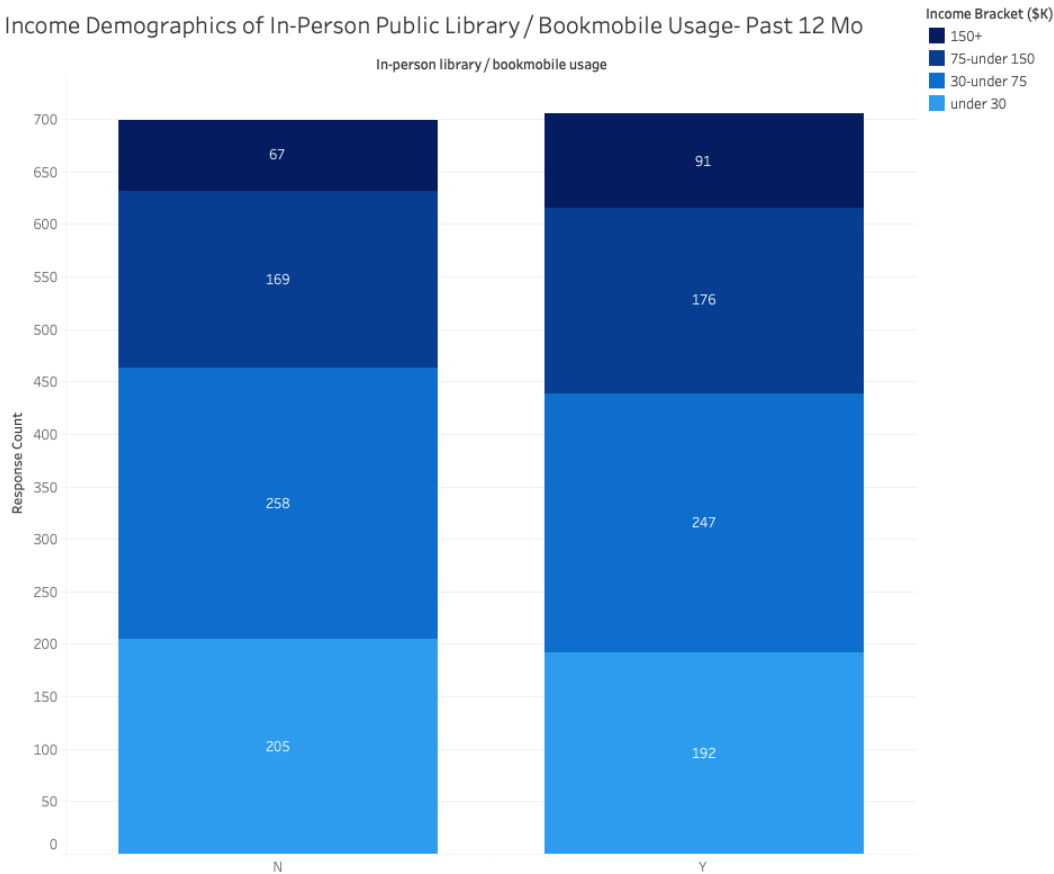
We believe that parental status will be related to in-person library usage such that parents will have the highest usage. We expect parents to want increased access to books due to having children. Additionally, libraries often provide areas for children to play which we expect to be an additional draw for parents. The model containing variables “parental status” and “public library/book mobile usage in the past 12 months” is statistically significant. This suggests a relationship between parental status and in-person library usage. (x-squared= 10.969, df=1, p-value= 0.0009266).

Parental status of those who have used the public library website in the last 12 months VS those who haven't



We believe that parental status will be related to online library usage such that non-parents will have the highest usage. We expect non-parents to have less need for in-person resources. The model containing variables “parental status” and “online usage in the past 12 months” is statistically significant. This suggests a relationship between parental status and in-person library usage. (x-squared= 9.2191, df=1, p-value= 0.002395),

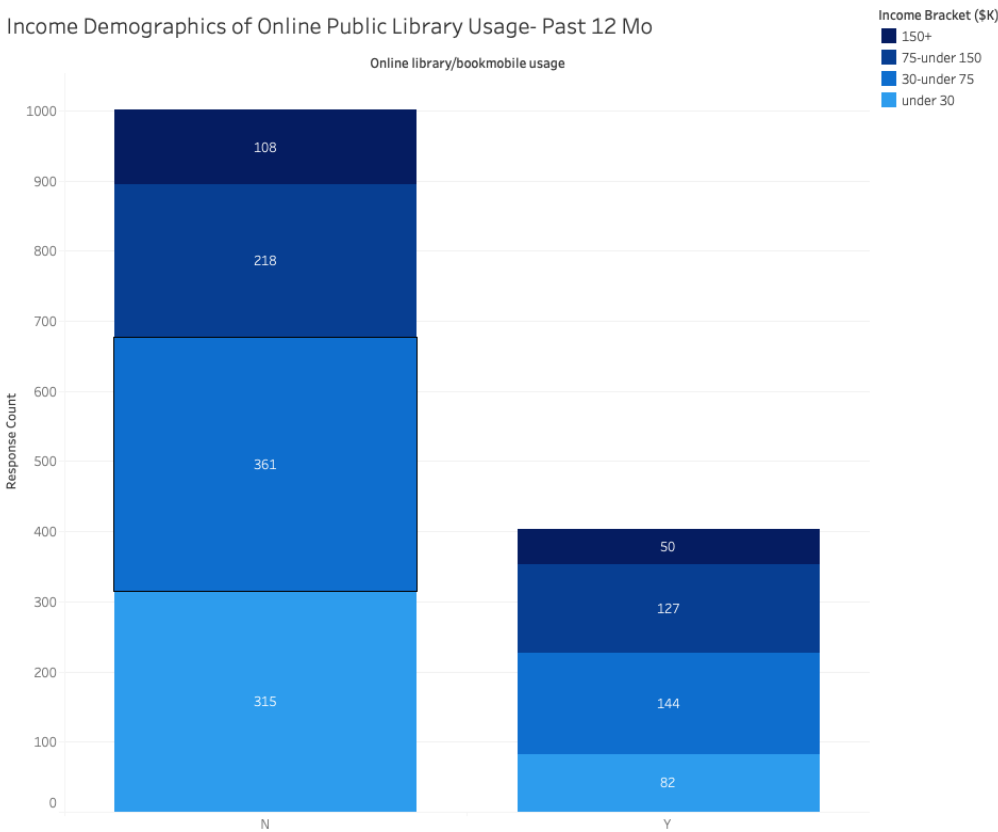
Income Demographics of In-Person Public Library / Bookmobile Usage- Past 12 Mo



p=0.2197

We believe that income will be related to in-person library usage such that people with lower incomes will have the highest usage. We expect those with lower incomes to need library resources to access the internet and books. On the other hand, we expect those with higher incomes to have access to the internet and books at home. The model containing the variables “income” and “public library/book mobile usage in the past 12 months” is not statistically significant. This suggests no relationship between income and in-person library usage. (x-squared= 4.4181, df=3, p-value= 0.2197).

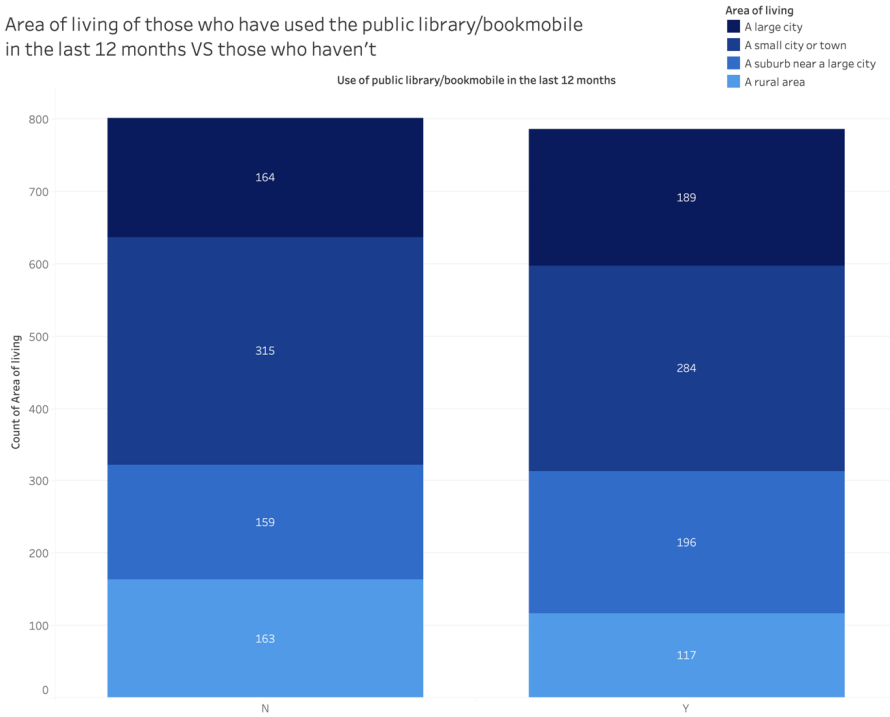
Income Demographics of Online Public Library Usage- Past 12 Mo



$p = 2.125e-05$

We believe that income will be related to online library usage such that people with higher incomes will have the highest usage. We expect those with higher incomes to have at home access to the internet and thus access to online library resources. On the other hand, we expect those with lower incomes to not have access to the internet and would need library resources. The model containing the variables “income” and “online usage in the past 12 months” is statistically significant. This suggests a relationship between income and in-person library usage. (x-squared= 24.337, df=3, p-value= 2.125e-05)

Area of living of those who have used the public library/bookmobile in the last 12 months VS those who haven't

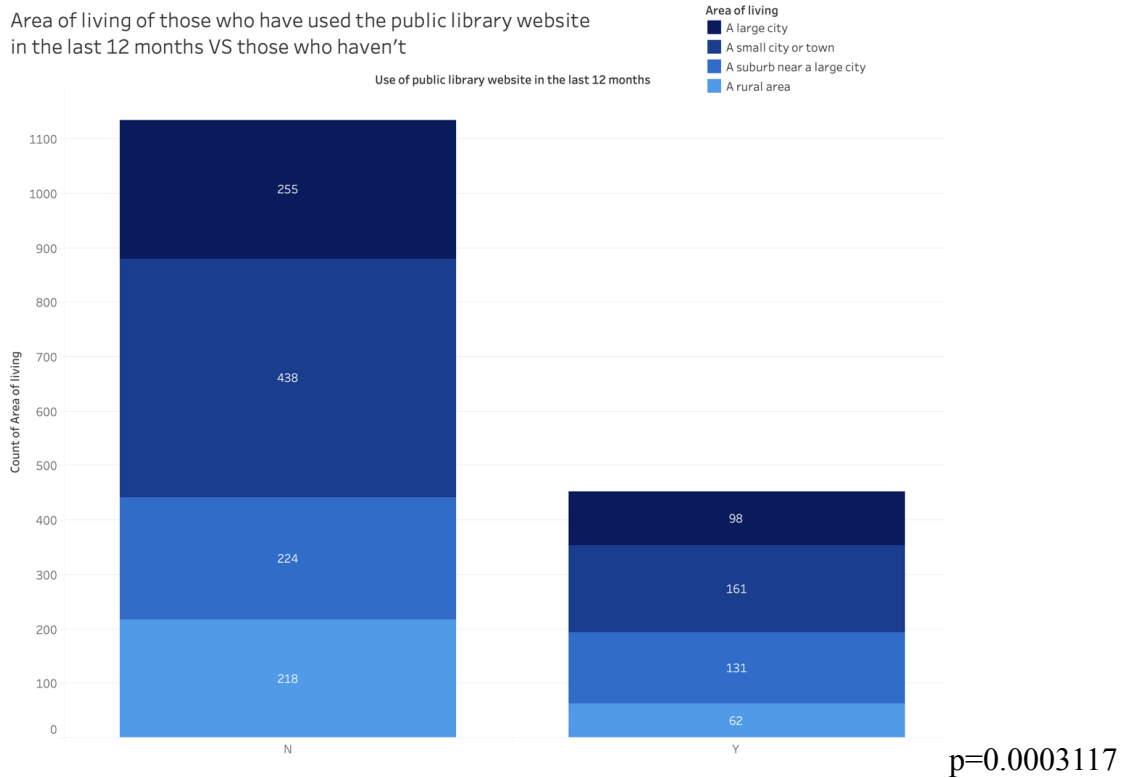


$p=0.002144$

We believe that community type will be related to in-person library usage such that people who live in a large city or a suburb near a large city will have the highest usage. We expect those who live in or live close to a large city will have more libraries within a comfortable distance. We expect those who are in a rural area to be far away from libraries and therefore not travel long distances to go in person. The model containing the variables “community type” and “public library/book mobile usage in the past 12 months” is statistically significant. This suggests some relationship between income and in-person library usage. (x-squared= 14.648, df=3, p-value= 0.002144).



Area of living of those who have used the public library website in the last 12 months VS those who haven't



We believe that community type will be related to online library usage such that people who live in rural areas will have the highest usage. We expect those in rural areas to not have as convenient access to in-person library resources and therefore utilize online libraries. We expect those who are closer to large cities to have a large amount of libraries to use in-person and will not need online resources as much. The model containing the variables “community type” and “online usage in the past 12 months” is statistically significant. This suggests some relationship between income and online library usage. (x-squared= 18.724, df=3, p-value= 0.0003117).

## Binomial Regression

	Pr(> z )
(Intercept)	0.8381
library.data\$age	0.0997 .
library.data\$Income30-under 75	0.1526
library.data\$Income75-under 150	0.2946
library.data\$Incomeunder 30	0.6159
library.data\$RuralityA Rural Area	0.0082 **
library.data\$RuralityA Small City or Town, OR	0.1011
library.data\$RuralityA Suburb Near A Large City	0.9959
library.data\$NumbersofBooksRead	<2e-16 ***
---	
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1	

	OR	2.5 %	97.5 %
(Intercept)	1.0487062	0.6646807	1.6574902
library.data\$age	0.9959102	0.9910538	1.0007777
library.data\$Income30-under 75	0.7483386	0.5021720	1.1125133
library.data\$Income75-under 150	0.8009689	0.5280905	1.2122774
library.data\$Incomeunder 30	0.9002990	0.5965251	1.3567718
library.data\$RuralityA Rural Area	0.6079422	0.4195788	0.8780239
library.data\$RuralityA Small City or Town, OR	0.7763467	0.5732602	1.0504903
library.data\$RuralityA Suburb Near A Large City	0.9991137	0.7113905	1.4030713
library.data\$NumbersofBooksRead	1.0552320	1.0438600	1.0675649

The coefficient on rural area rurality and number of books read were statistically significant. From the model, we see that living in a rural area is associated with a decrease in the odds of using a library in-person (coefficient is negative). We see that living in a rural area is associated with a 0.608 decrease in odds of going to a library in-person. From this model, we see that the number of books read in the last 12 months is associated with an increase in odds of using a library in-person (coefficient is positive). We see that every 1 extra book read in the last 12 months is associated with a 1.055 increase in odds of going to a library in-person.

The coefficients on age, income, rurality of a small city/town or suburb near a large city are not statistically significant. The model is not statistically significant (p-value=0.8381) so age is not useful in explaining the odds that a person uses in-person libraries. The model is not statistically significant (p-value = 0.6159, 0.1526, 0.2946) so income is not useful in explaining the odds that a person uses in-person libraries. The model is not statistically significant (p=0.1011, 0.9959) so living in a small city/town or near a large city, respectively, are not useful in explaining the odds that a person uses a library in-person.

## **Executive Summary**

- The number of books someone reads, Gender, Parental Status and Community Type have a significant relationship with in-person and online library usage
- Income has a significant relationship with only online library usage
- Living in a rural area will decrease the odds you use a library in person
- The more books you read the higher the odds you will use a library in person
- For future analysis on how to allocate funds for library location and resource levels, statisticians should consider gender, parental status, income, and specifically focus on community type and the amount of books read by individuals.