

# CALIFORNIA PUBLIC LIBRARIES

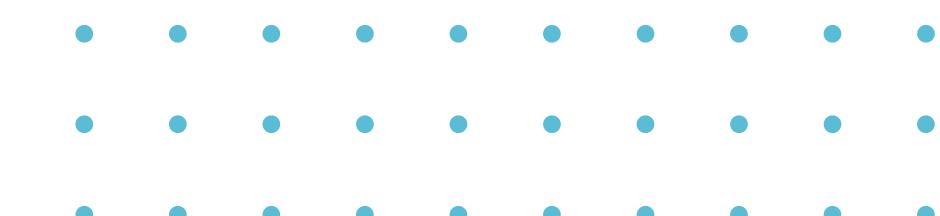


*Kathy Trieu  
Texas A&M University  
692 Statistical Consulting*

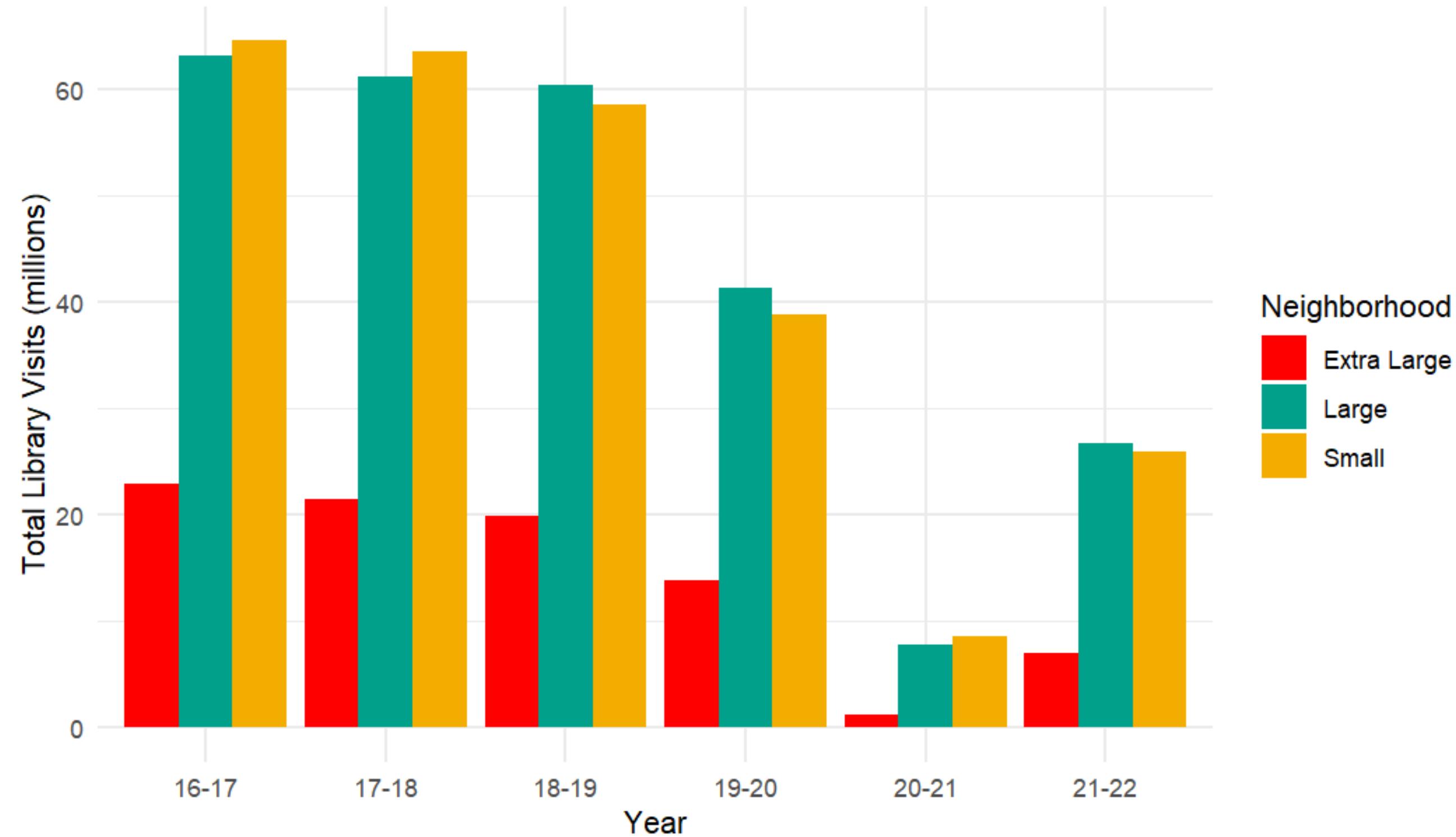


# "PALACES FOR THE PEOPLE"

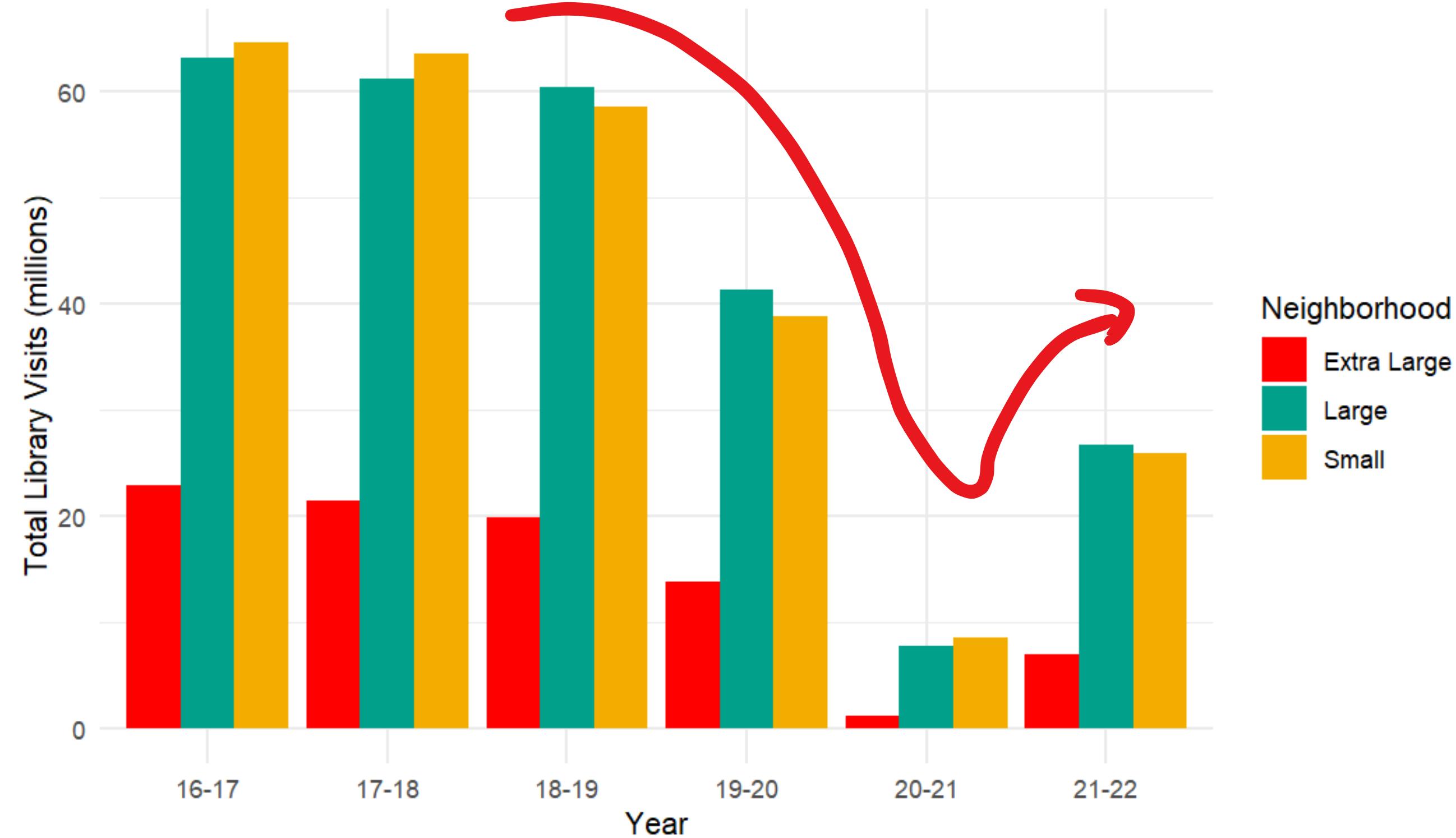
“Everyone is welcome at the library, regardless of whether they’re a citizen, a permanent resident, or even a convicted felon. And all of it [...] is free.” - Eric Klinenberg



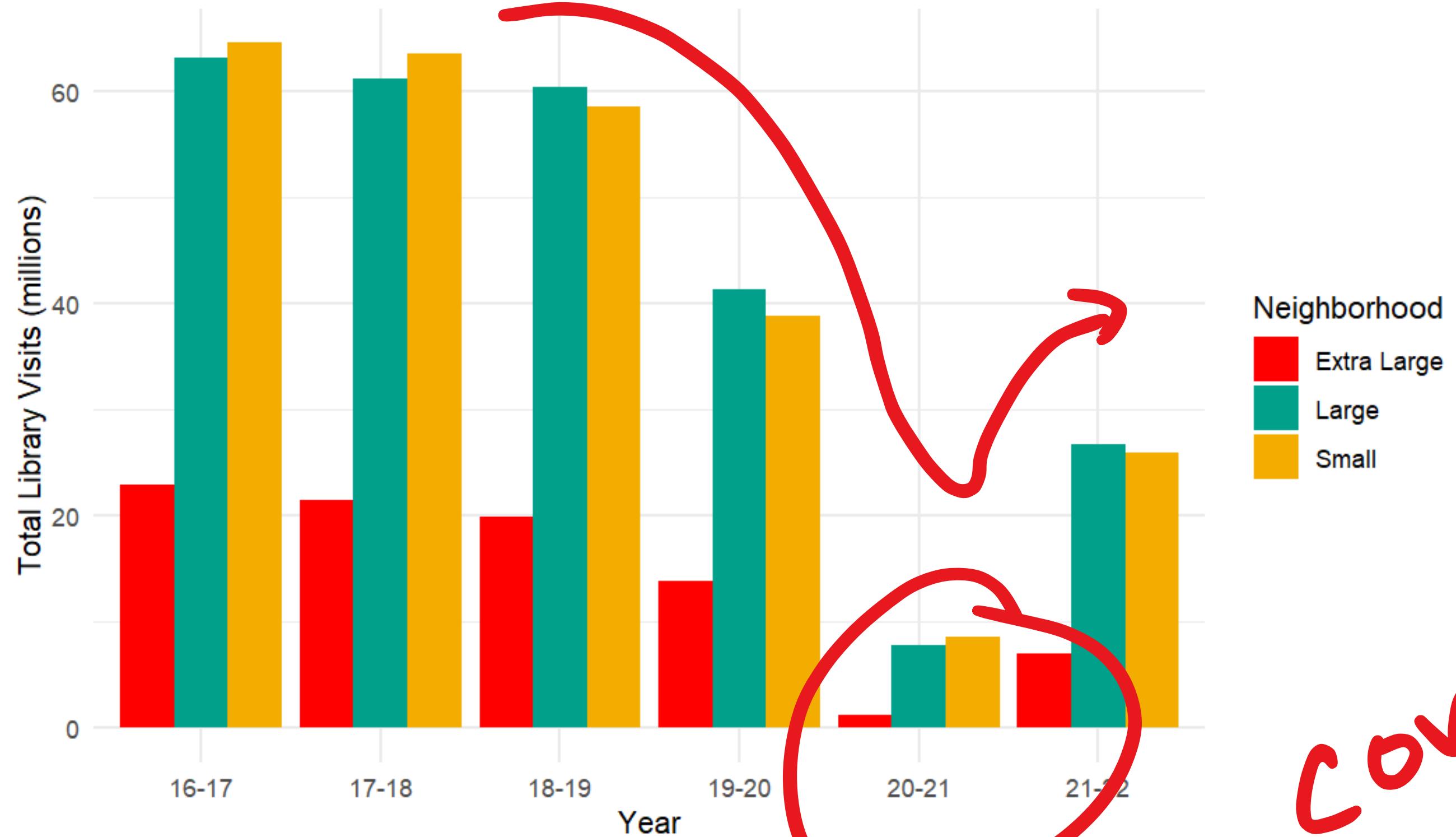
## Visits to All Libraries



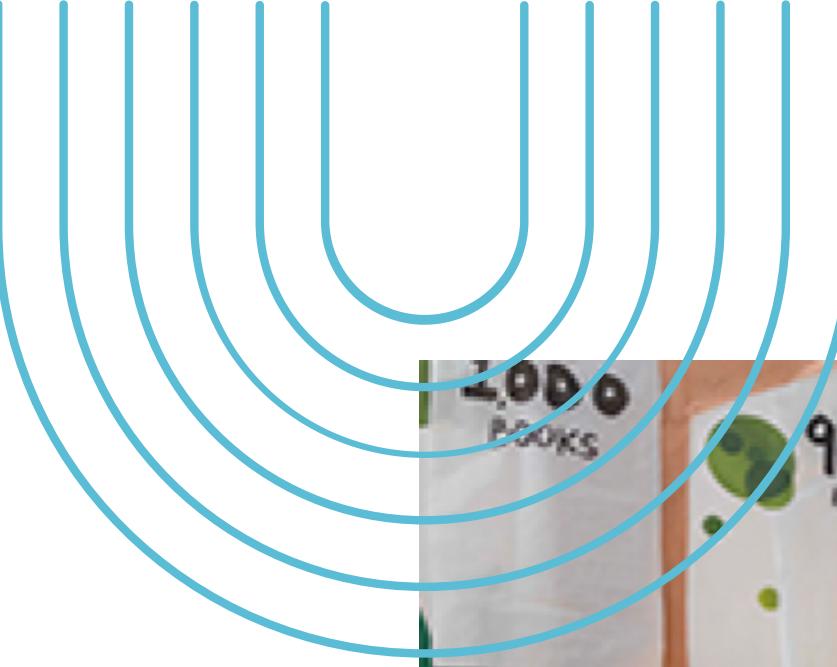
## Visits to All Libraries



## Visits to All Libraries



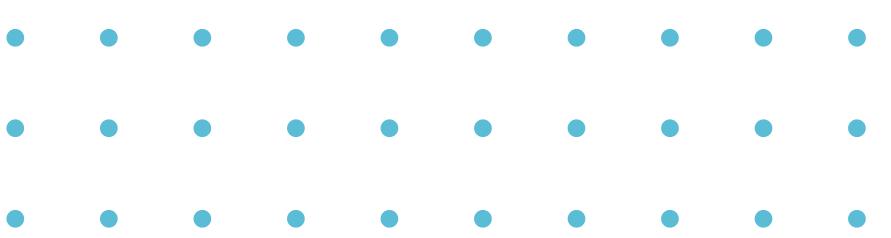
Covid-19



Fountain Valley, Orange County, CA



How can we fill up these seats?





01.

## RESEARCH PROBLEM

*Define research problem*

02.

## DATA DESCRIPTION

*Dataset and variables used*

03.

## DESCRIPTIVE STATISTICS

*Exploration of library statistics*

04.

## MACHINE LEARNING MODELS

*SVMR and Random Forest*

05.

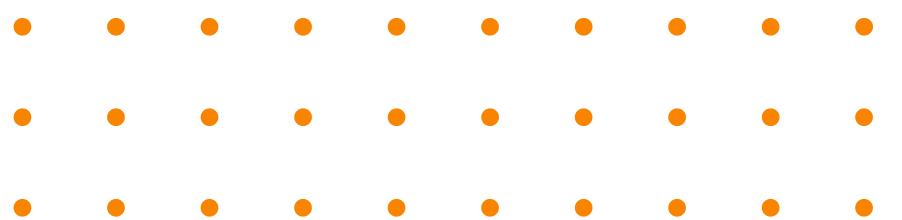
## MODEL RESULTS

*What did we learn from the models?*

06.

## RECOMMENDATIONS

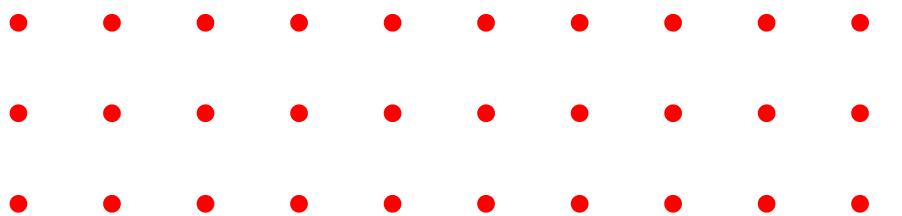
*Ideas for further research*



# TABLE OF CONTENTS

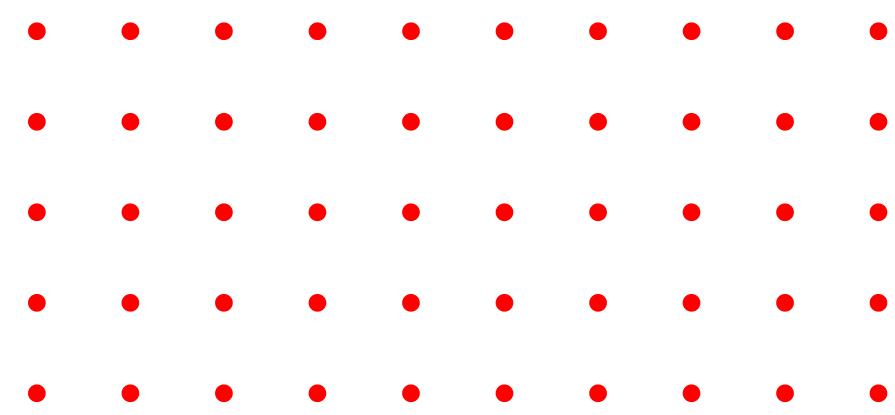
# RESEARCH PROBLEM

How do the characteristics  
of libraries in California  
contribute to variations in  
visitor attendance?



# California Library Statistics

<https://www.library.ca.gov/stats/>



# VARIABLES

## PROGRAMMING

Number of Programs,  
Program Attendance

## POPULATION OF SERVICE AREA

The population size of the legal  
service area

## COMPUTER USAGE

Number of computer use at the  
library

## CIRCULATION OF MATERIALS

Number of materials circulated  
(loaned and renewed)

## EXPENDITURE & INCOME

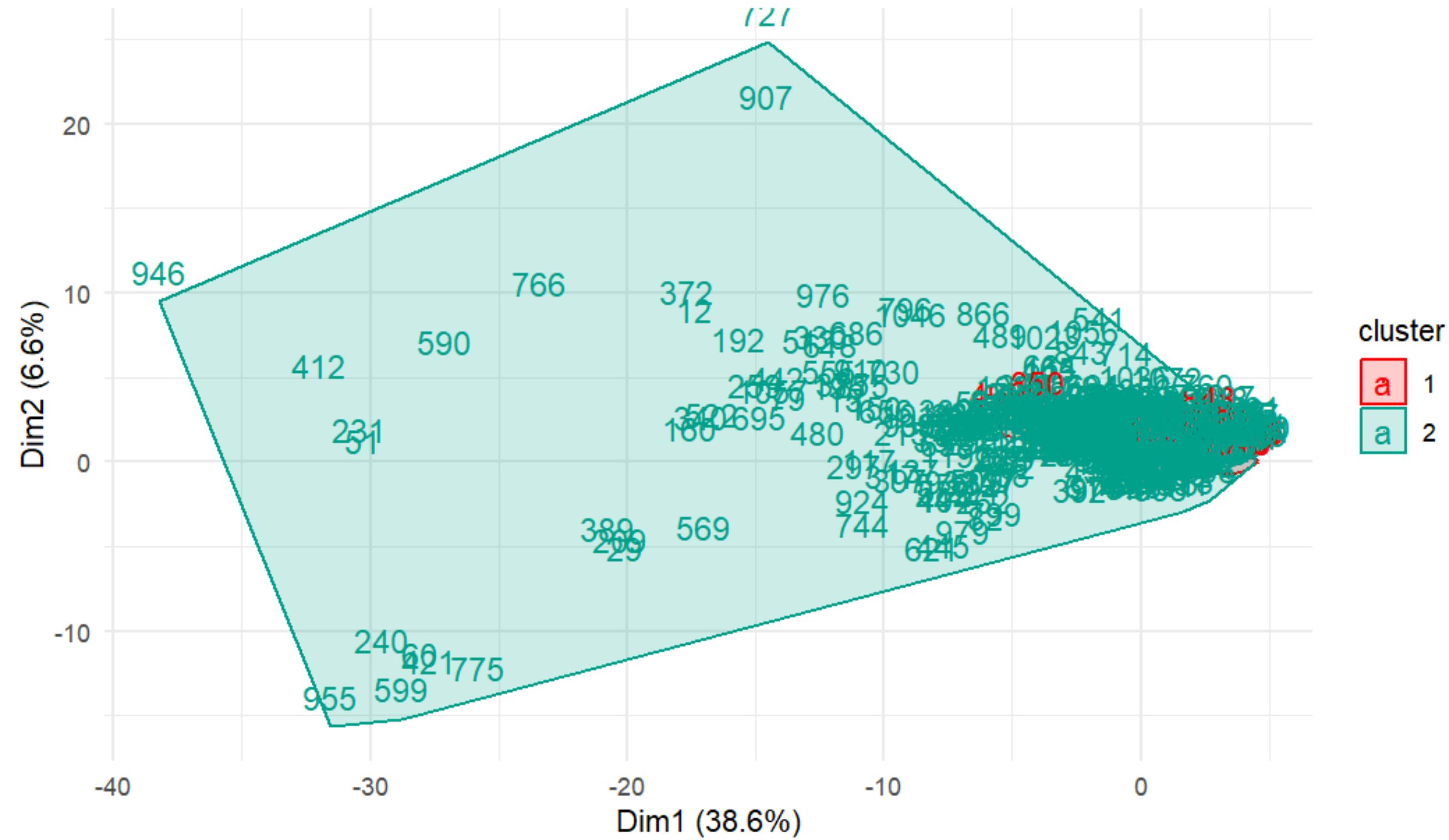
Income from state, federal, and  
other sources,  
Expenditures by type

## STAFF & VOLUNTEERS

Number of staff and volunteers,  
Staff expenditures

## Library Visits

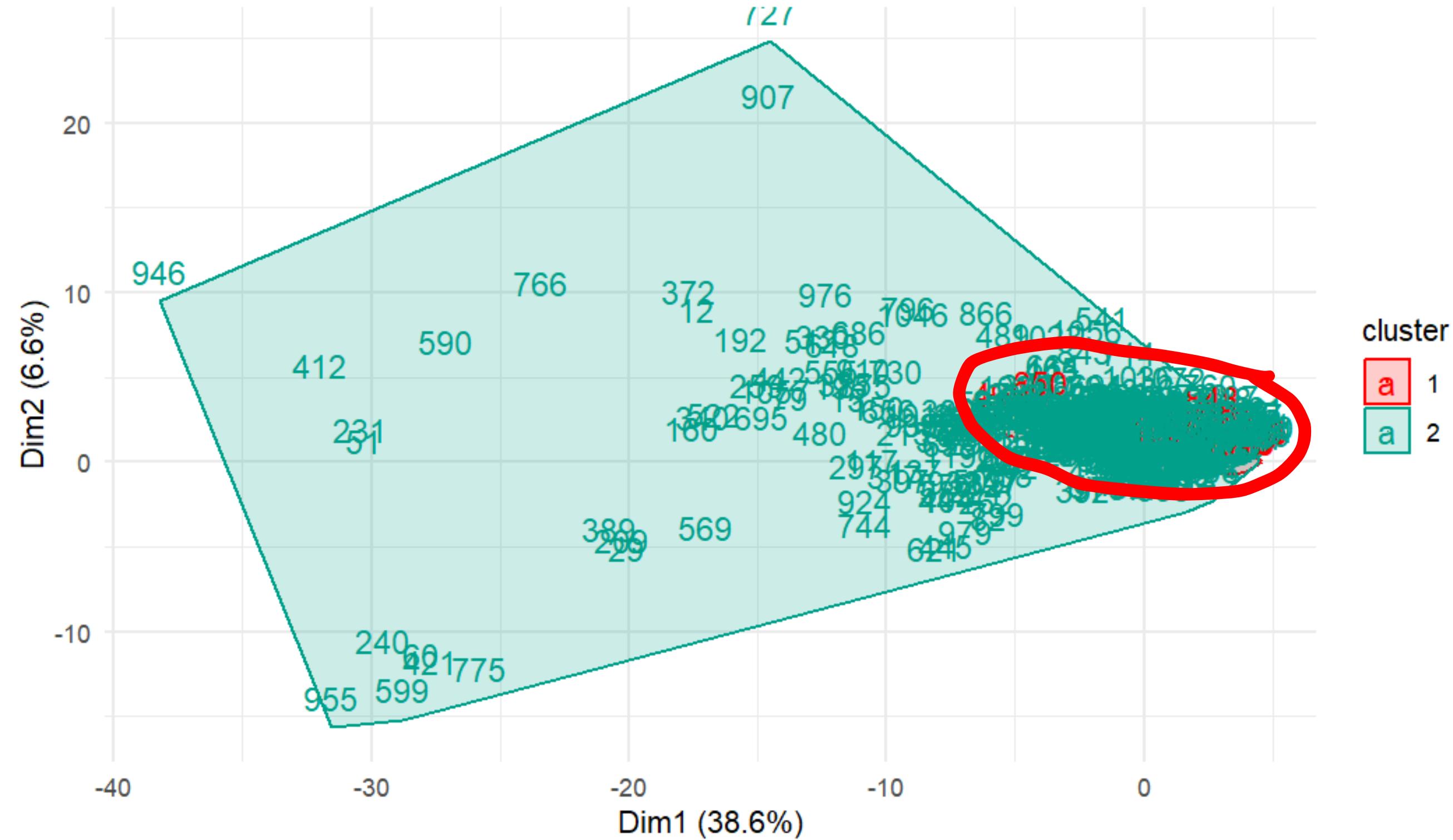
## Cluster plot



cluster

a	1
a	2

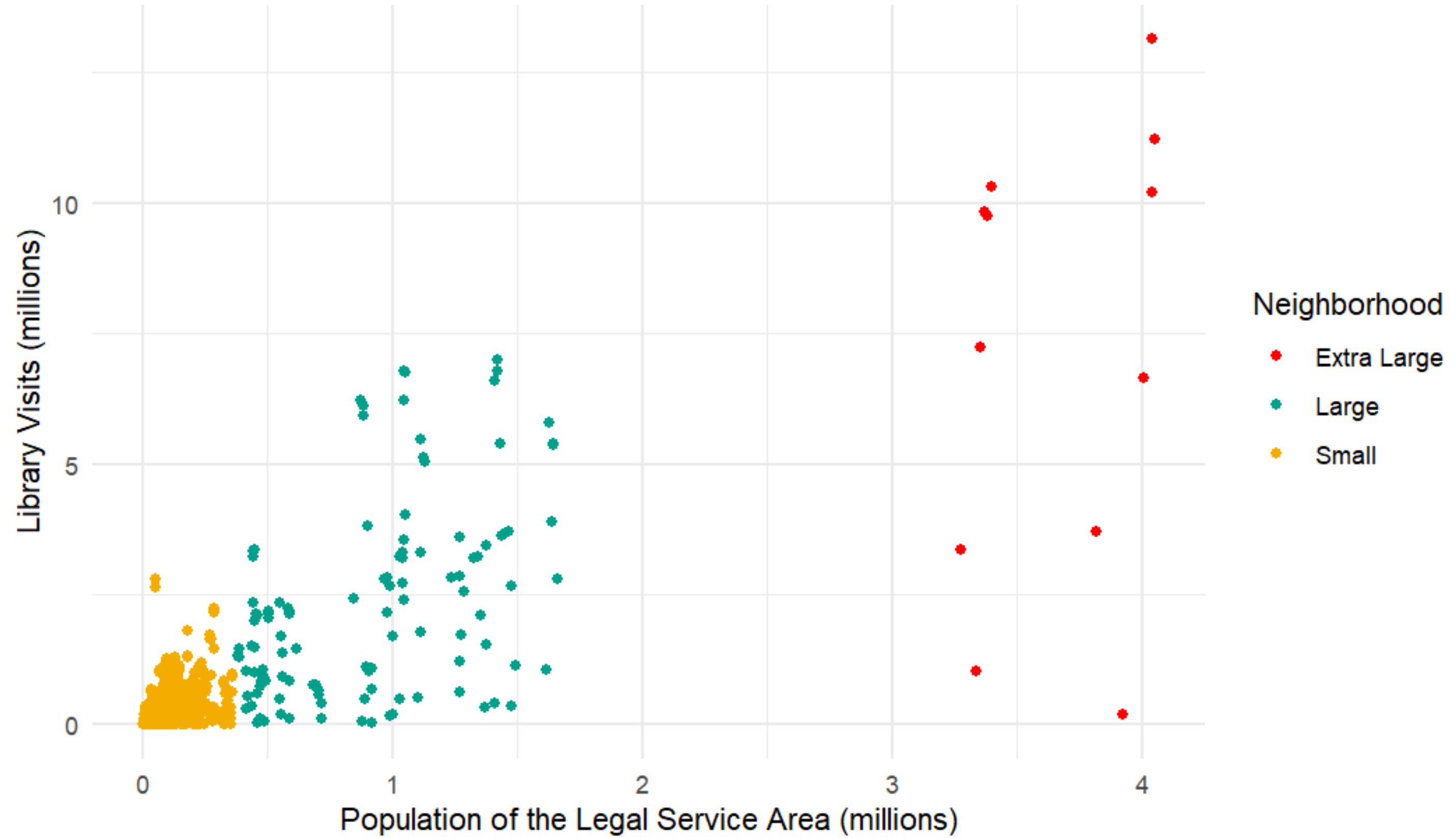
Cluster plot



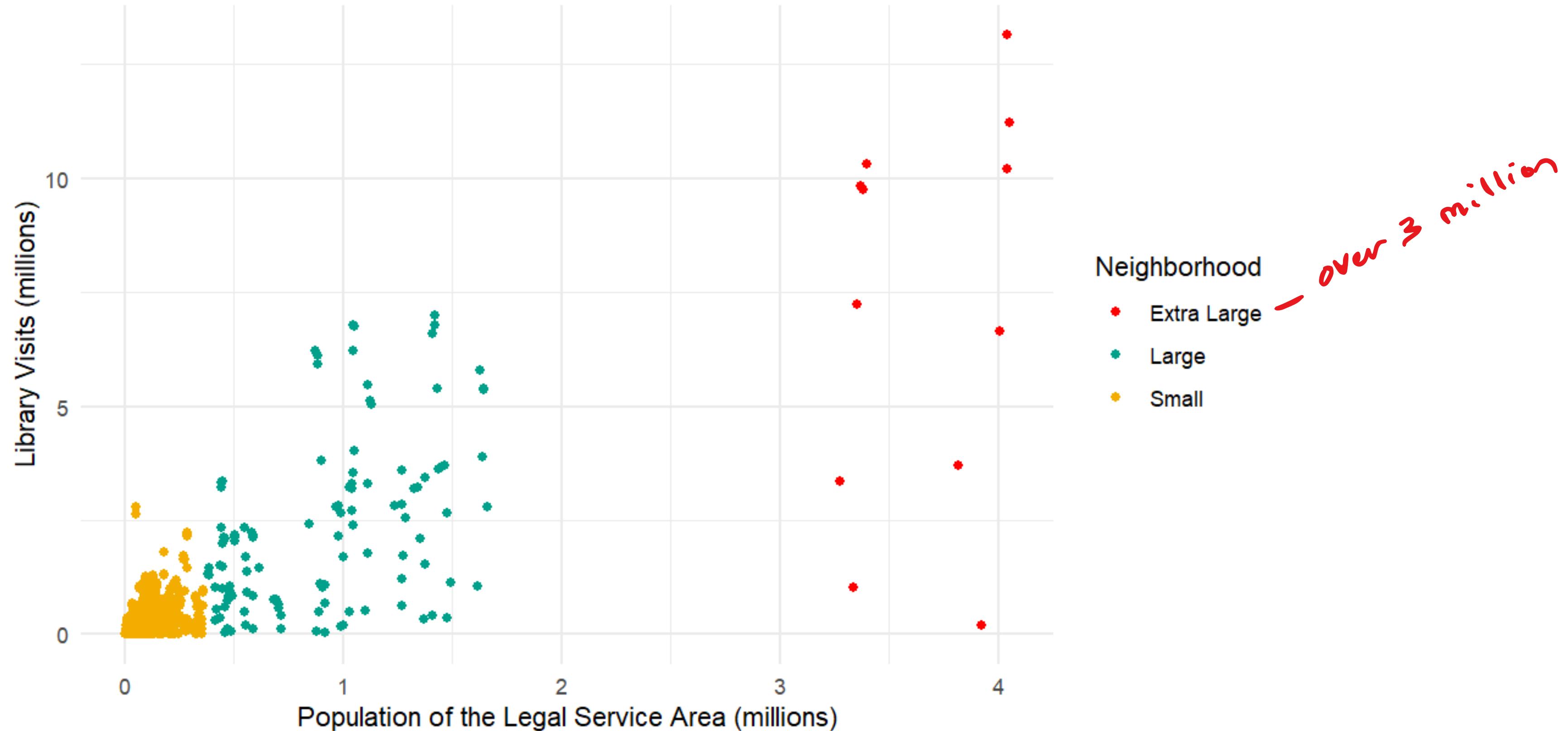
cluster

a	1
a	2

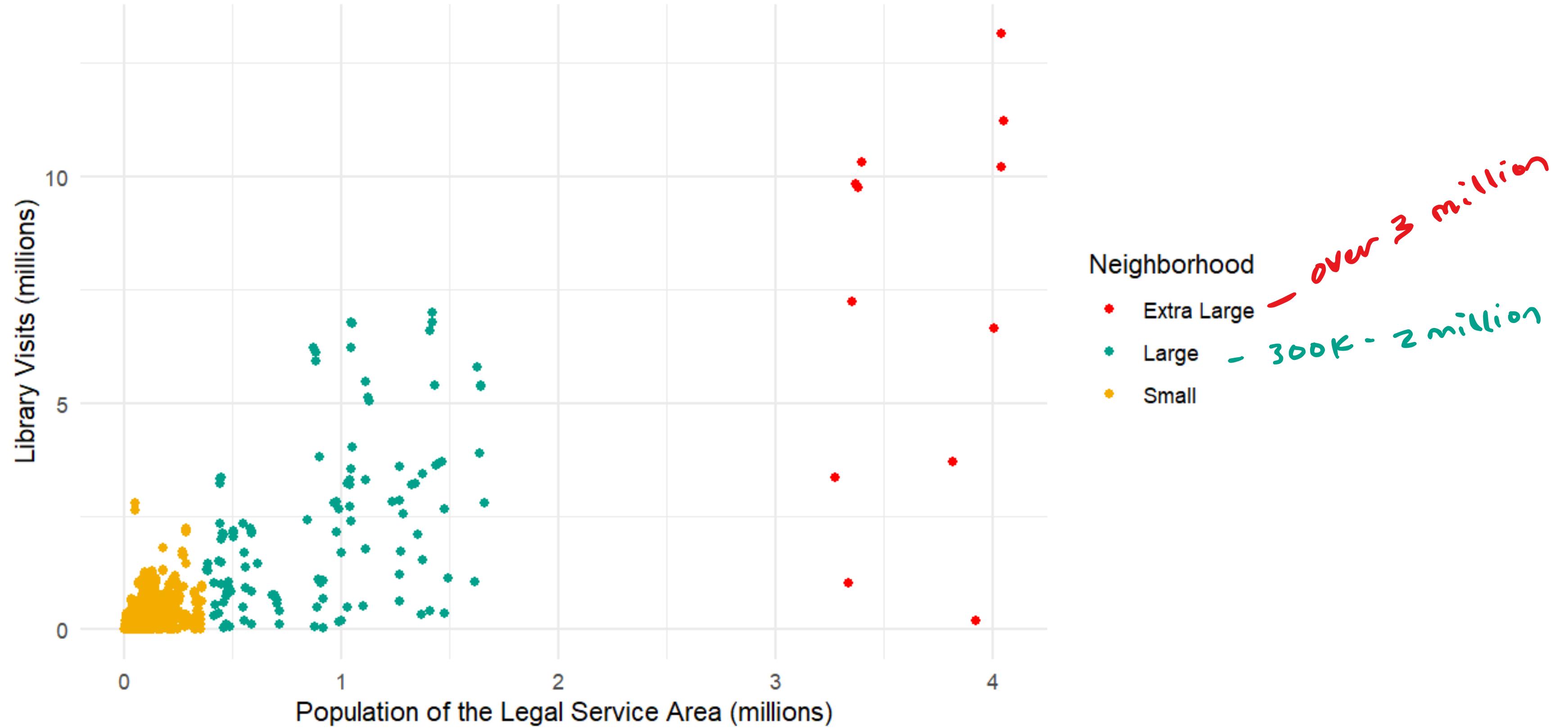
## Library Visits and Population Size



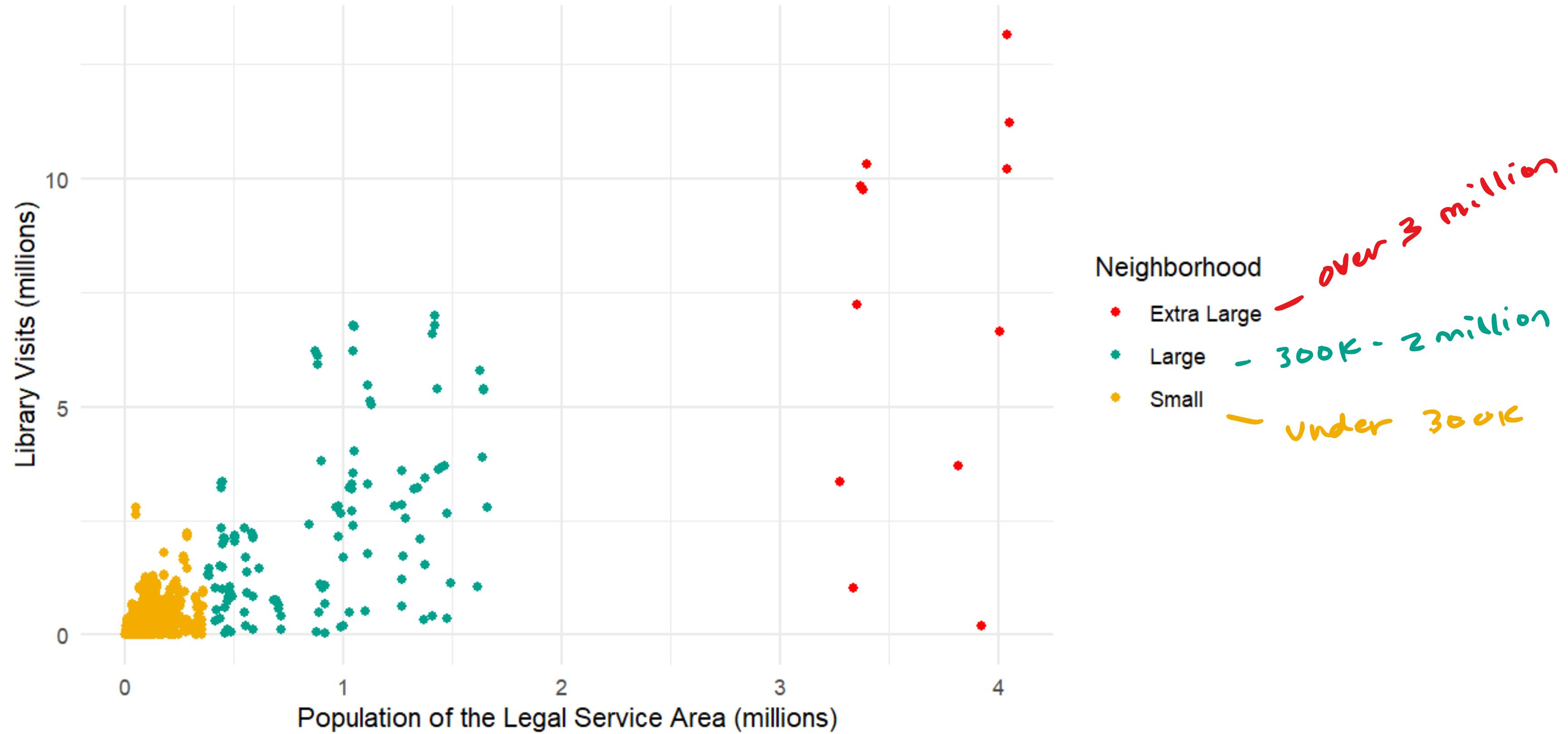
## Library Visits and Population Size



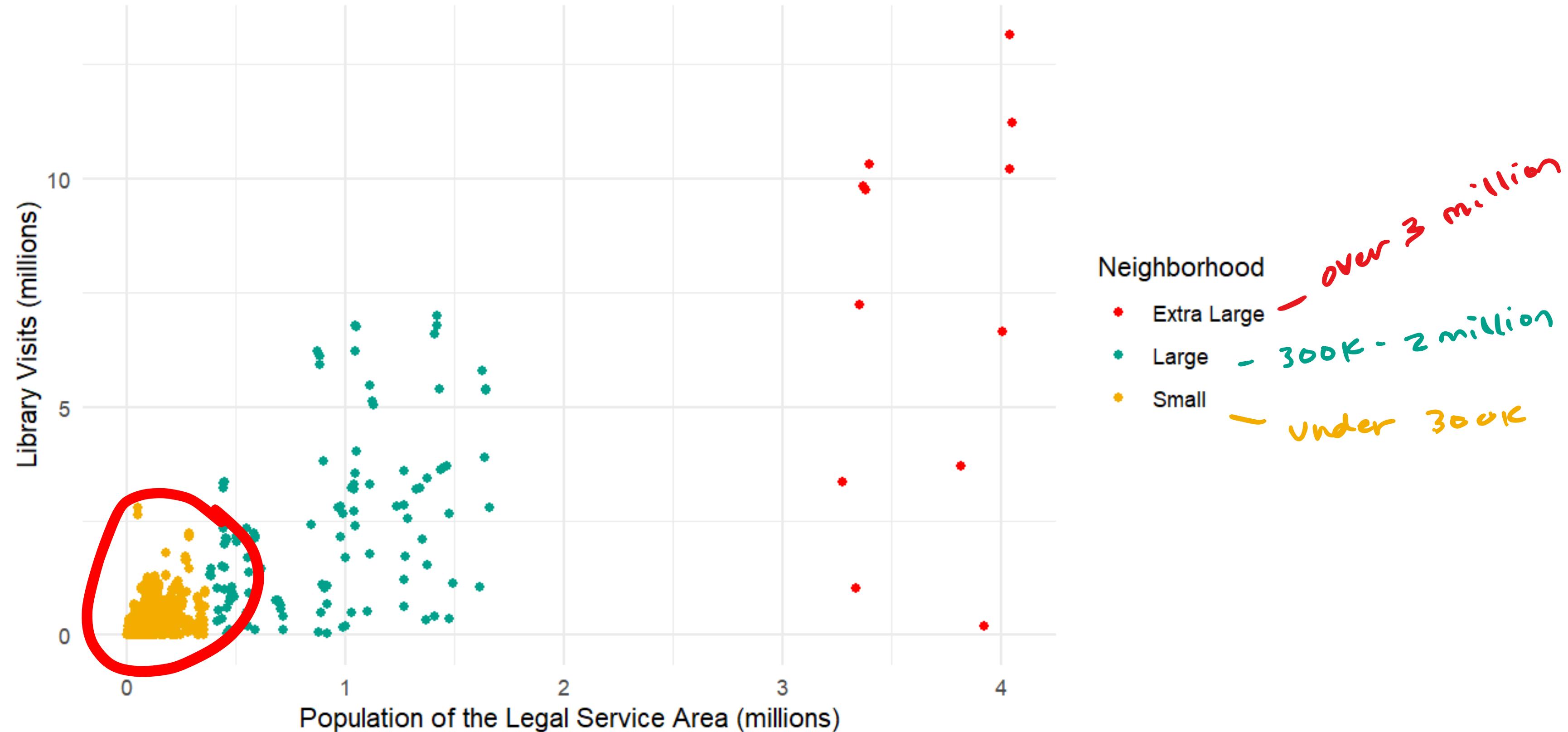
## Library Visits and Population Size



## Library Visits and Population Size

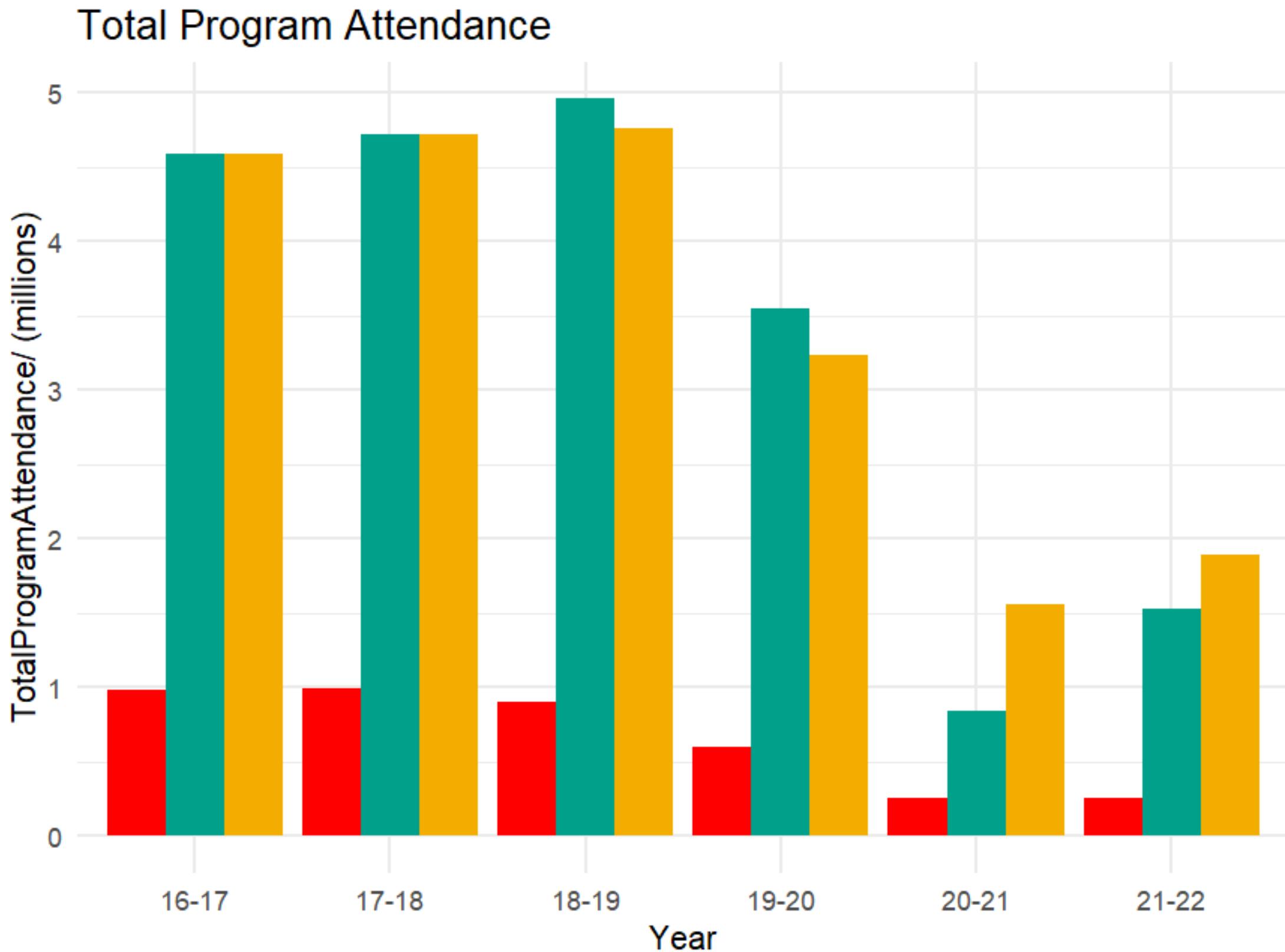


## Library Visits and Population Size



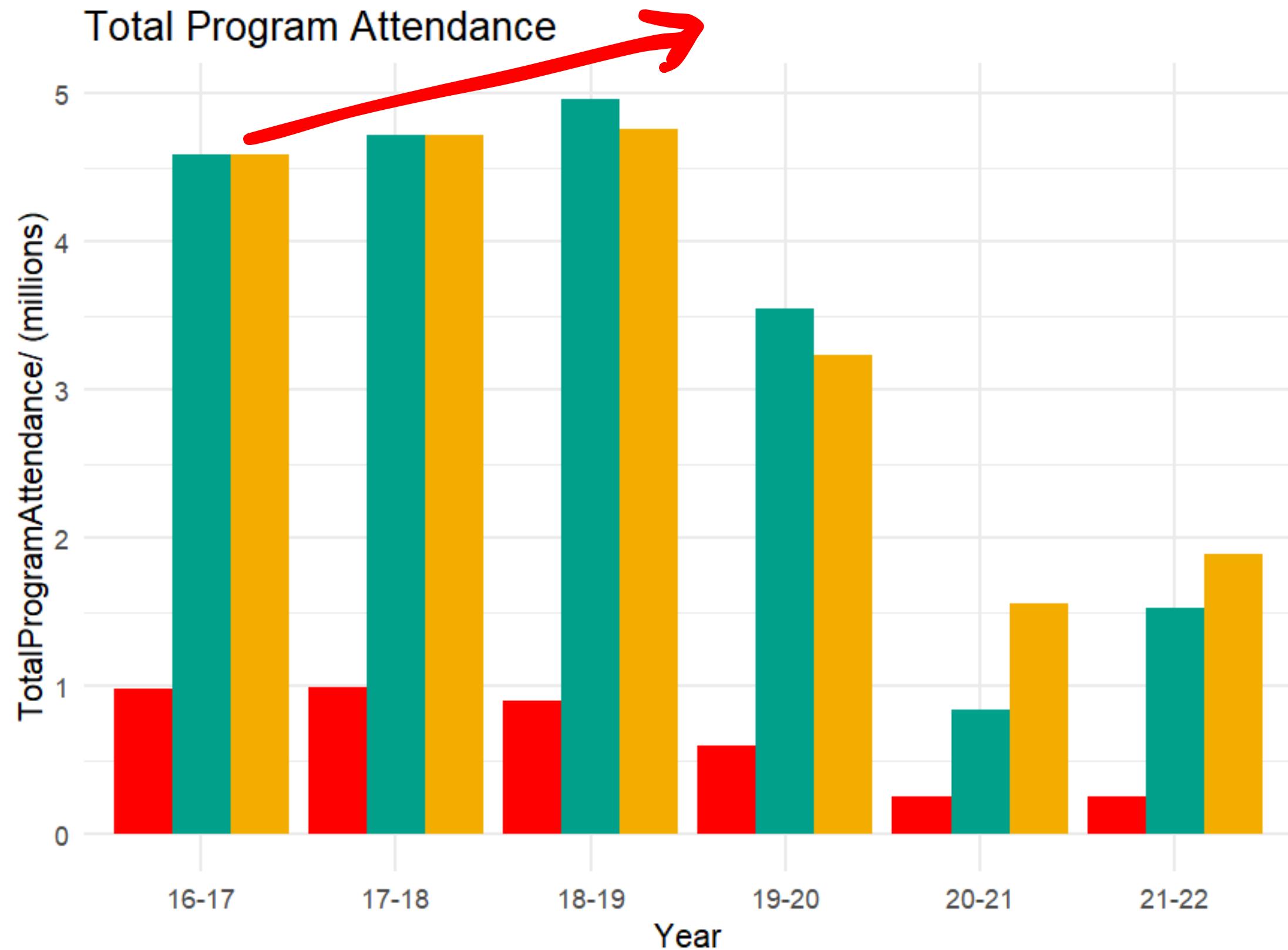
## Library Visits and Population Size





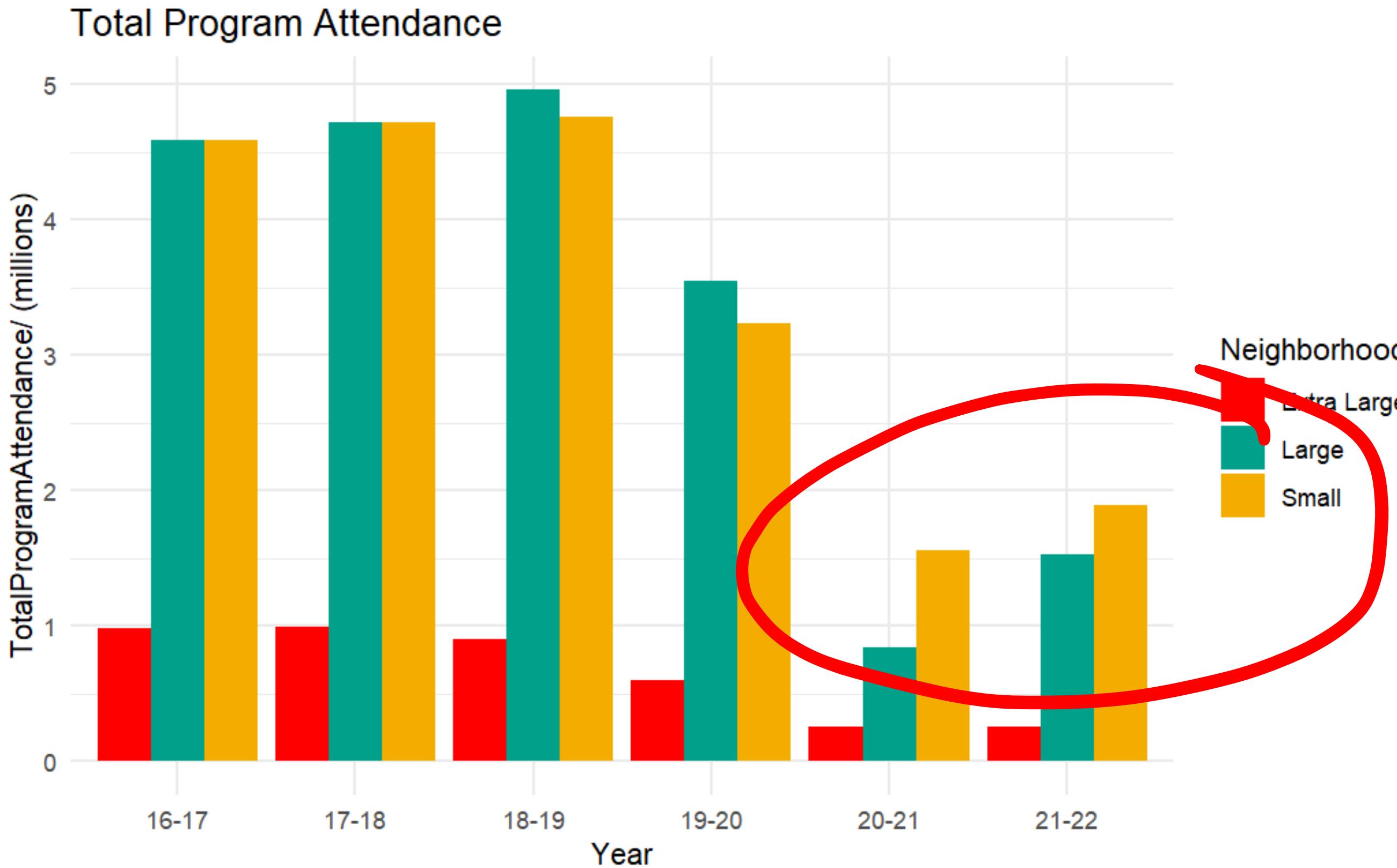
Neighborhood

- Extra Large
- Large
- Small

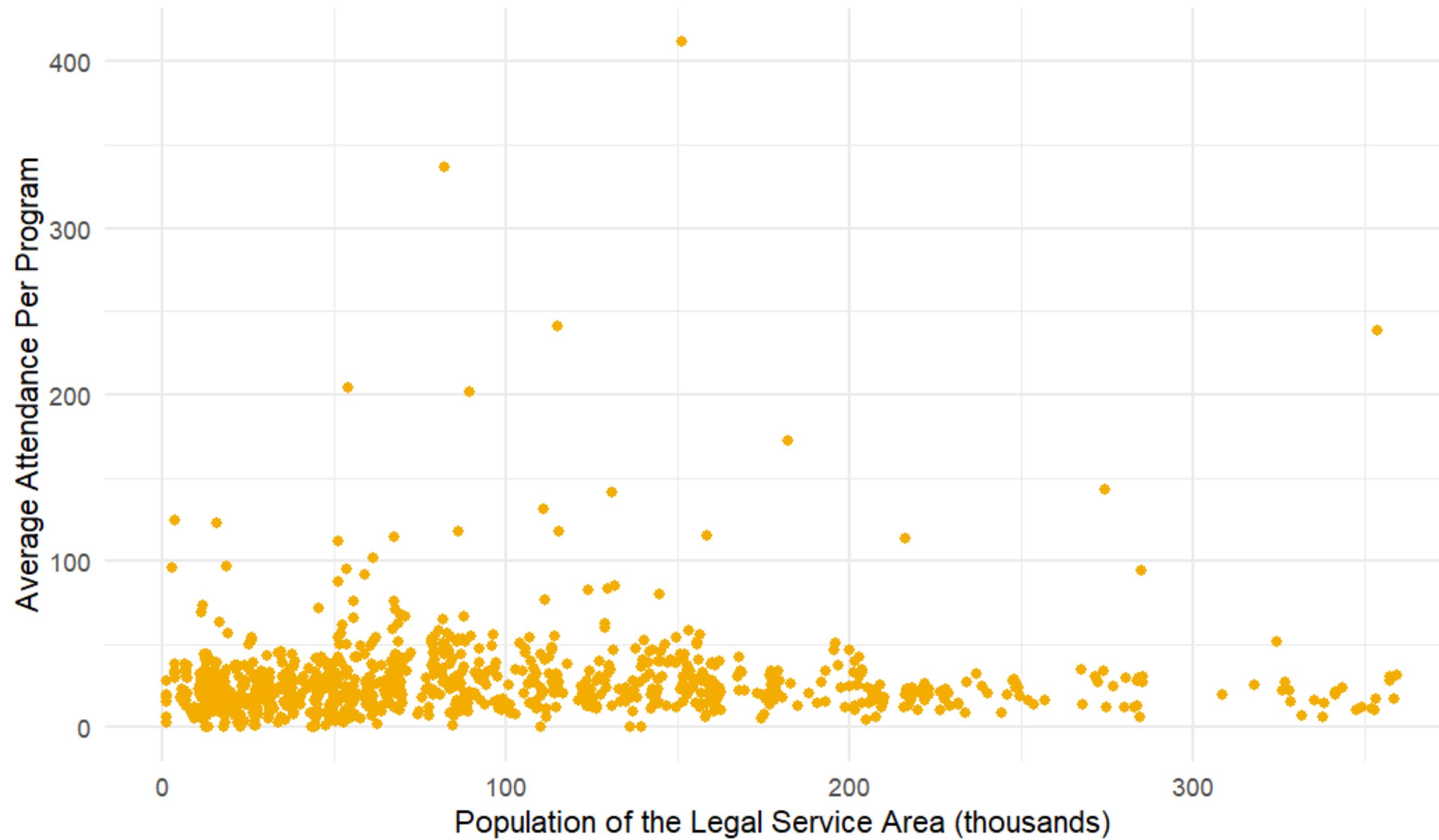


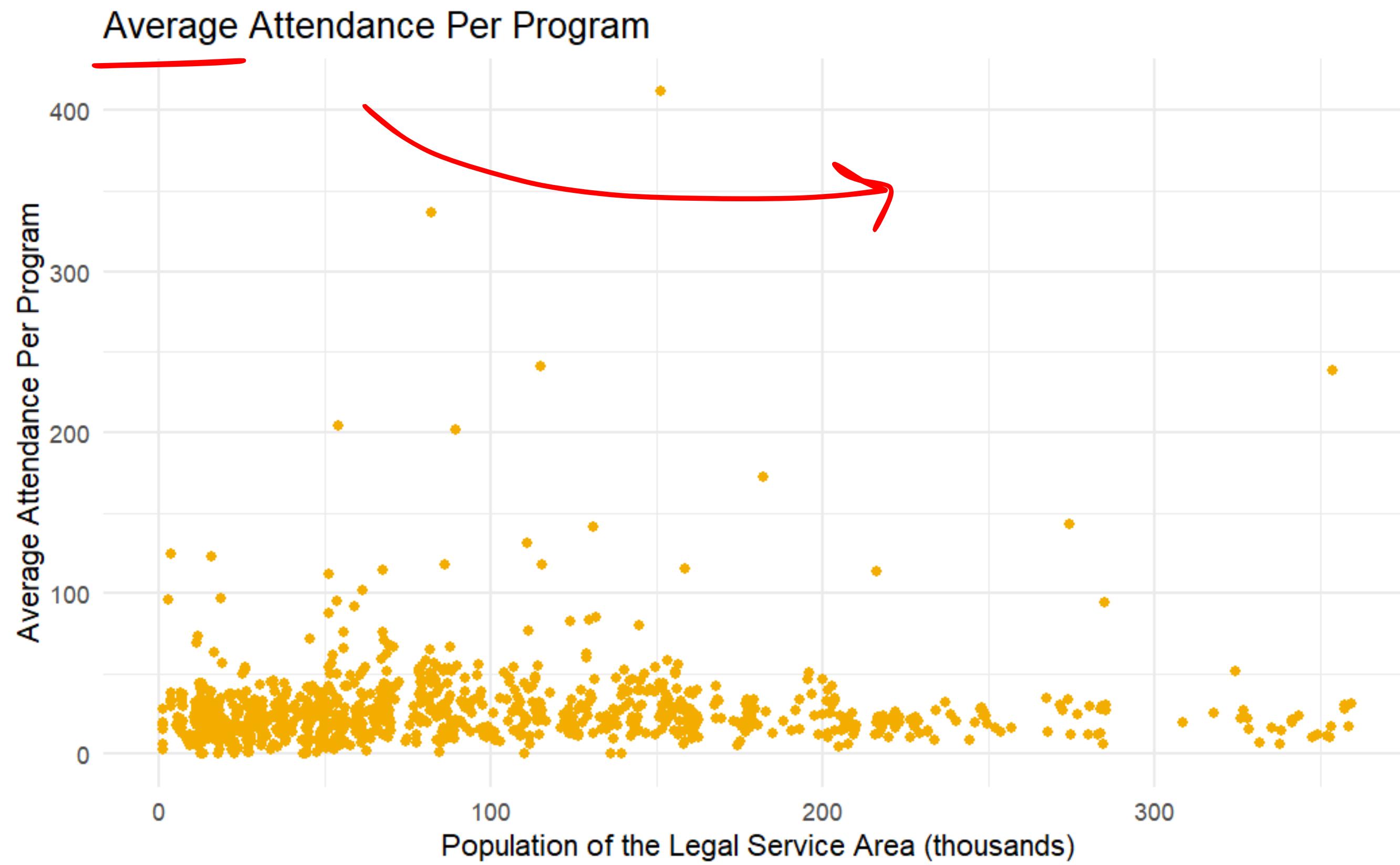
Neighborhood

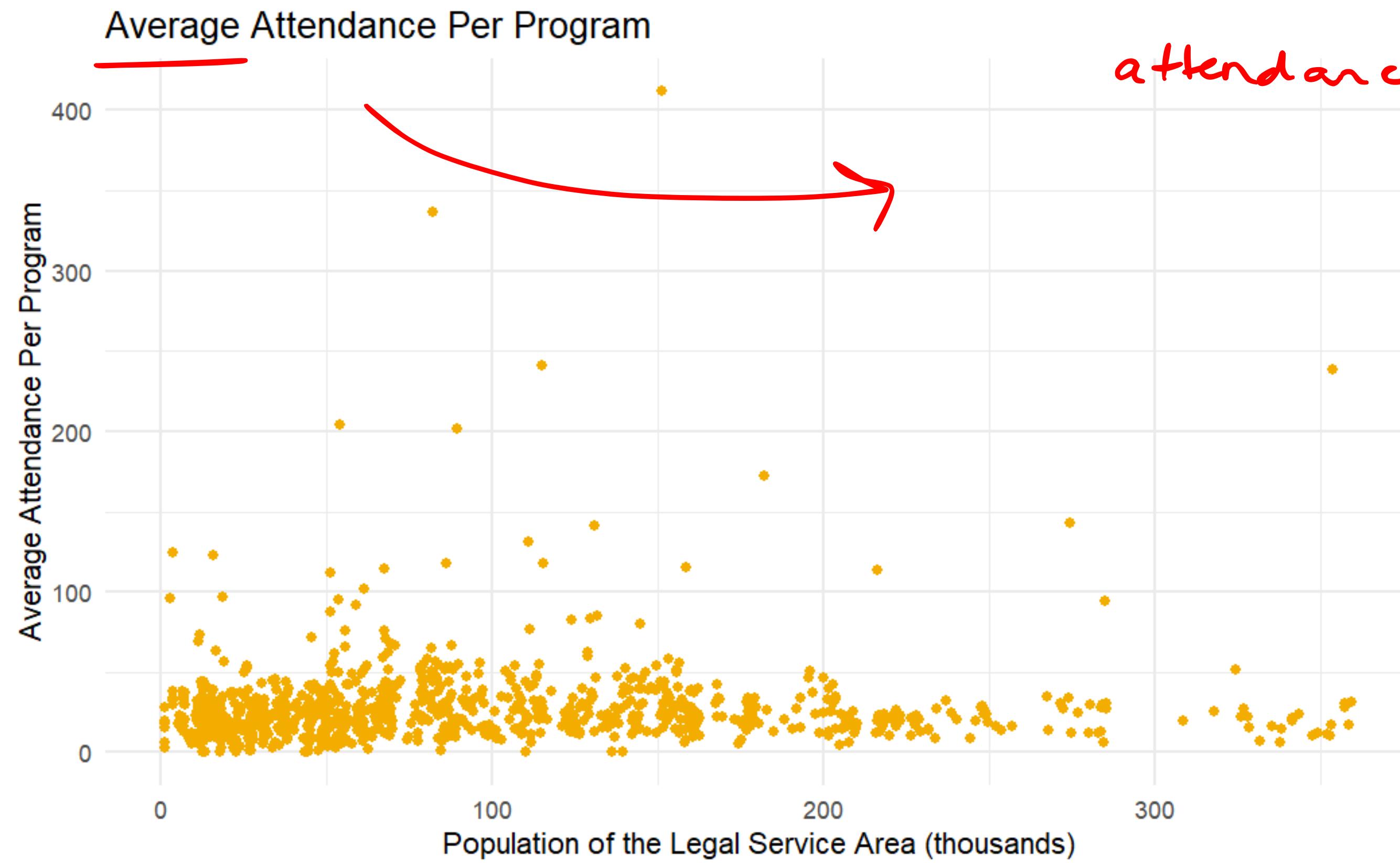
- Extra Large
- Large
- Small



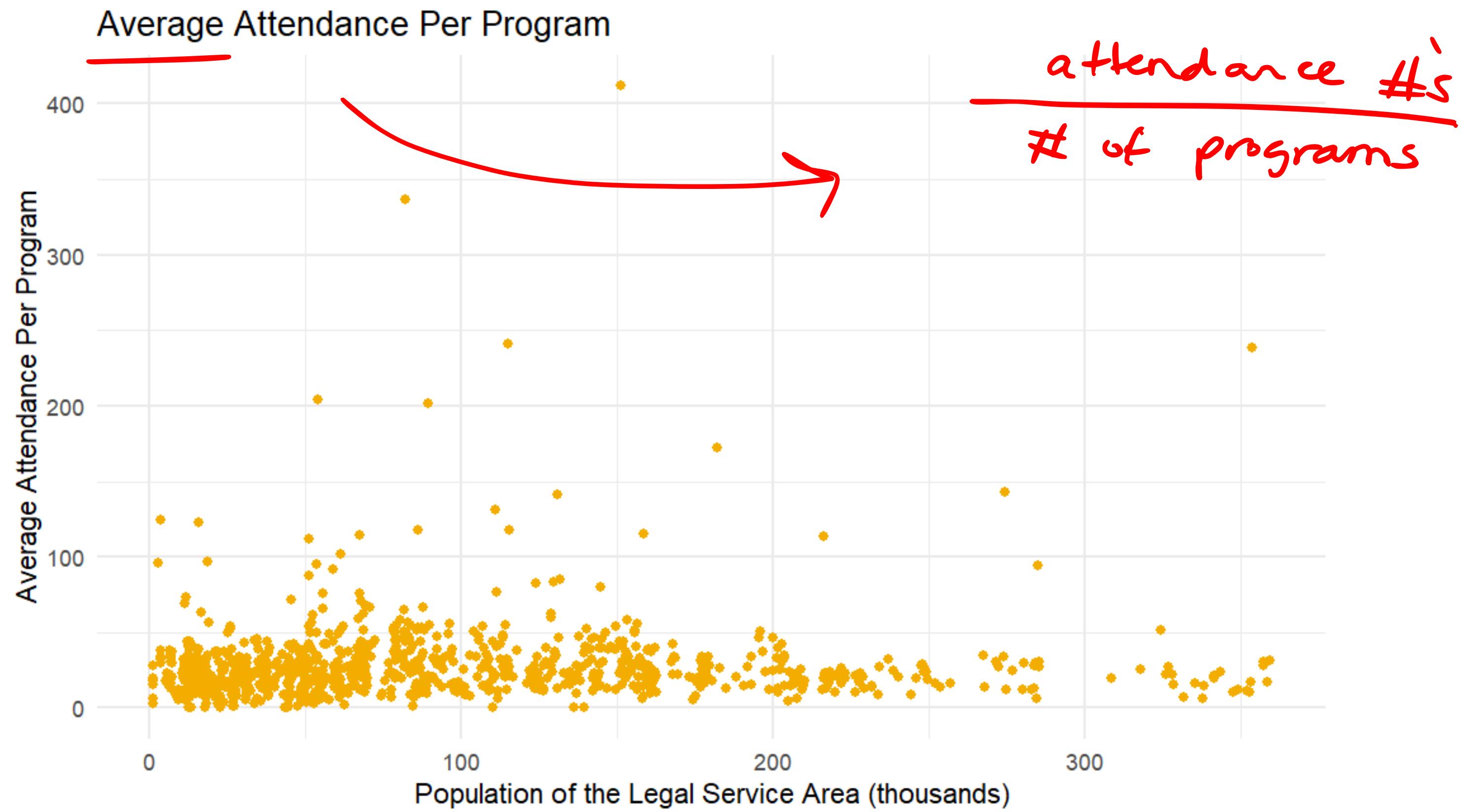
## Average Attendance Per Program

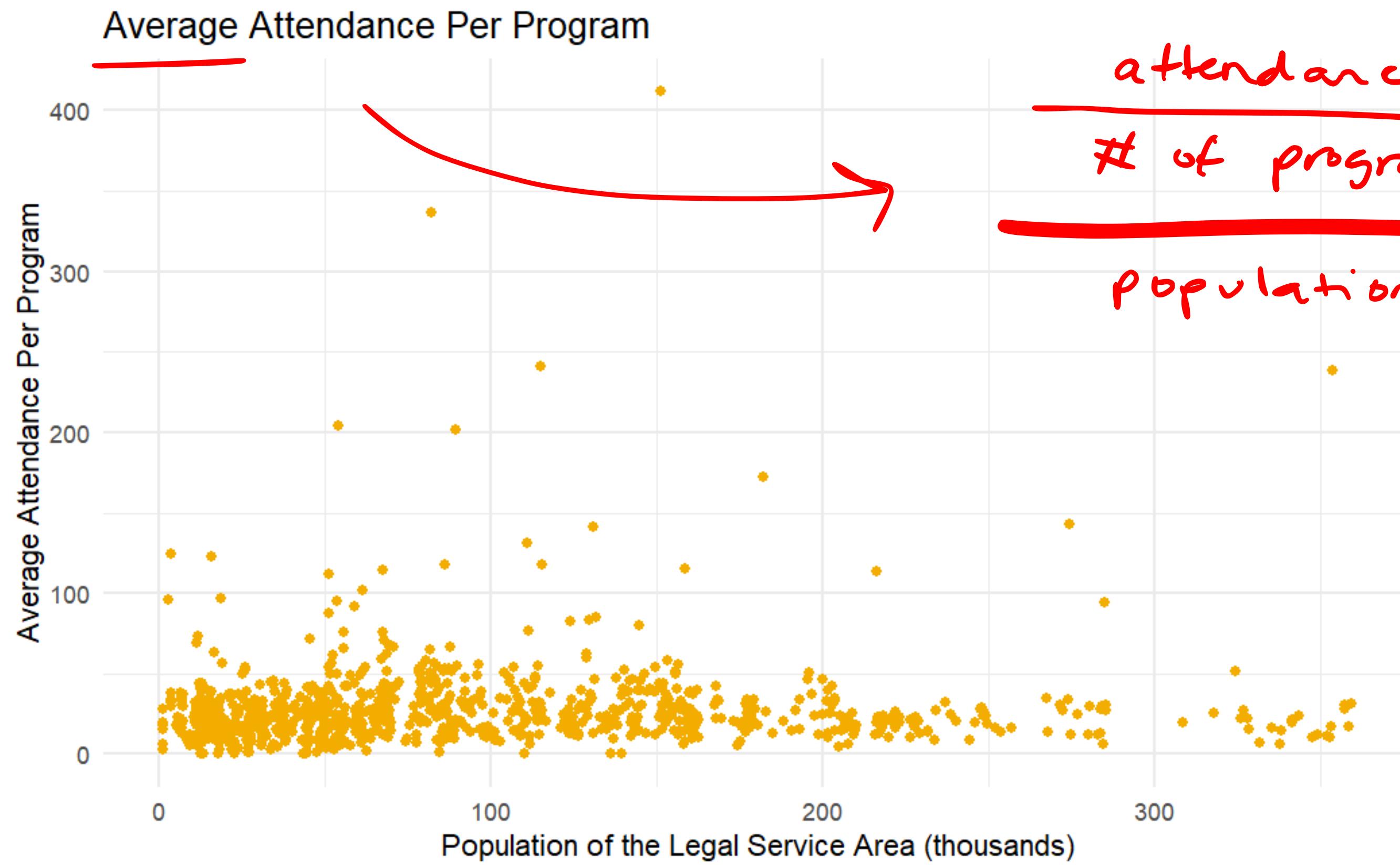






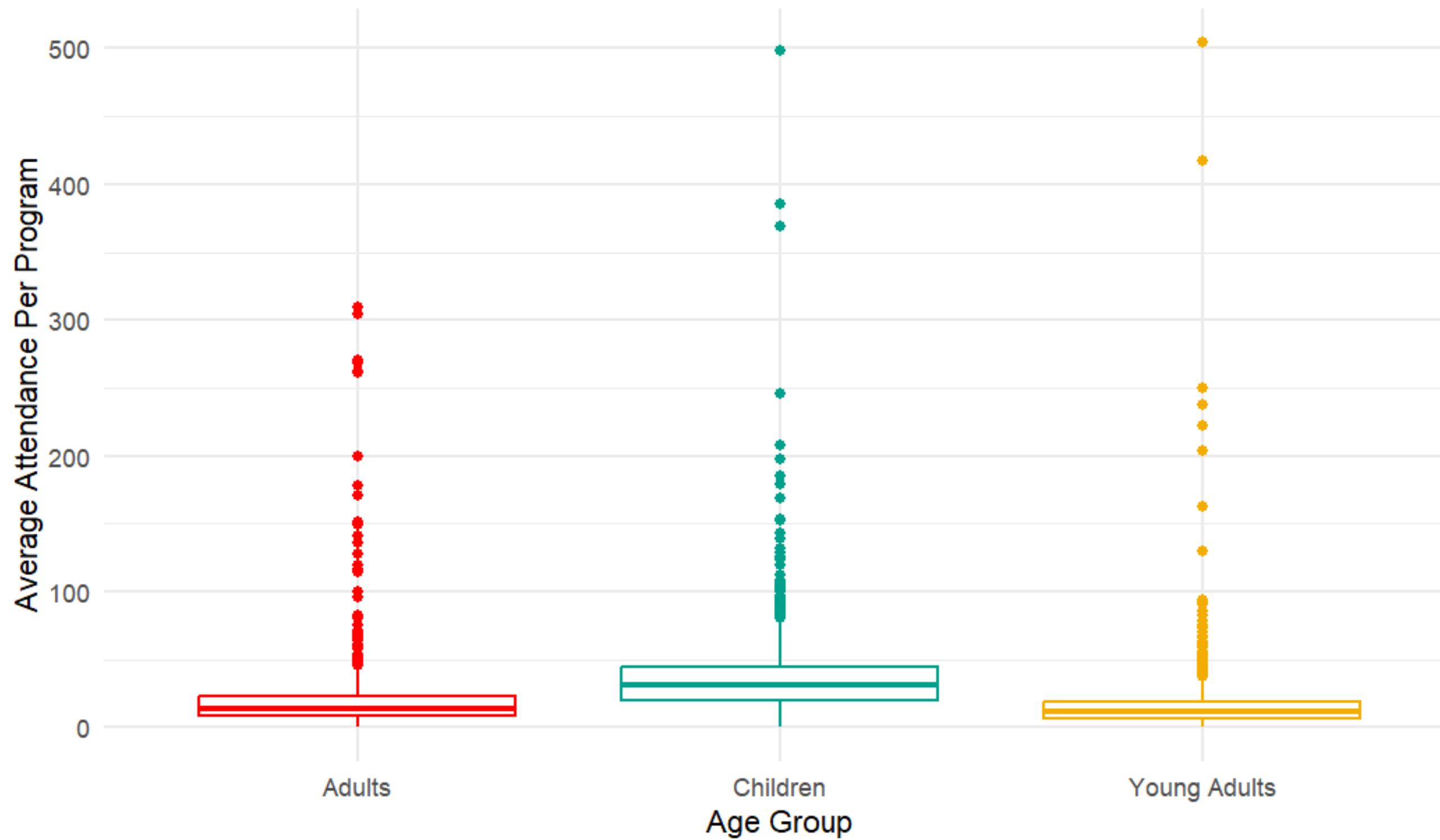
attendance #'s

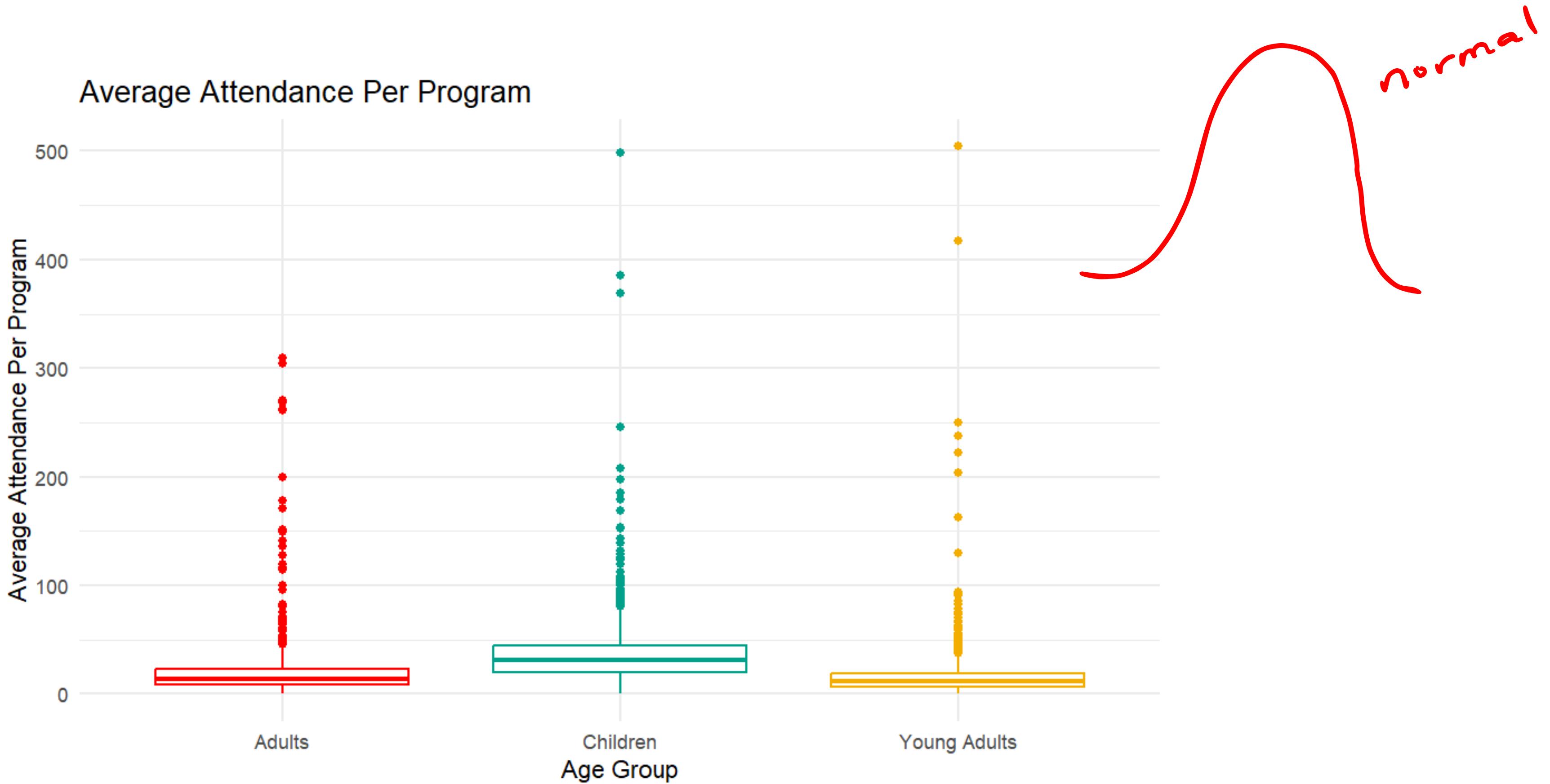




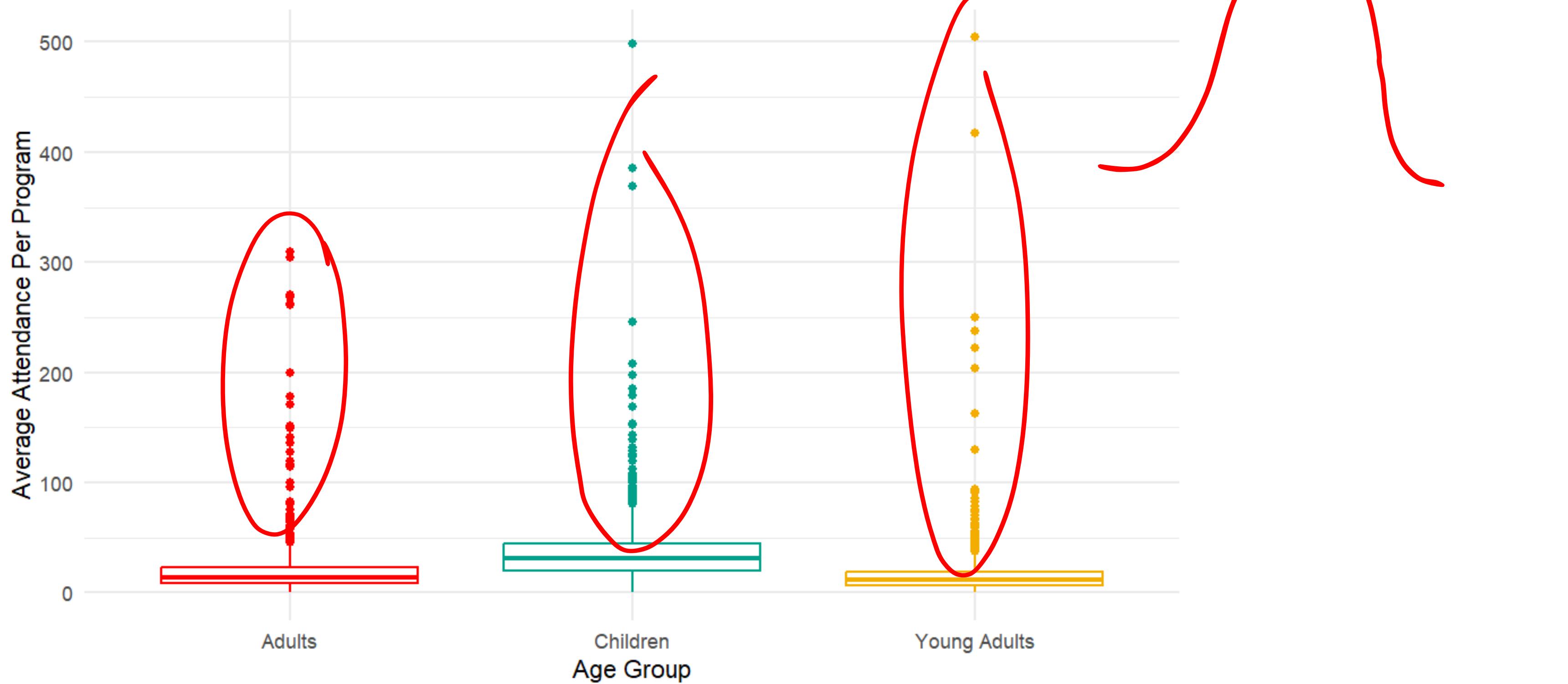
attendance #'s  
# of programs  
population size

## Average Attendance Per Program

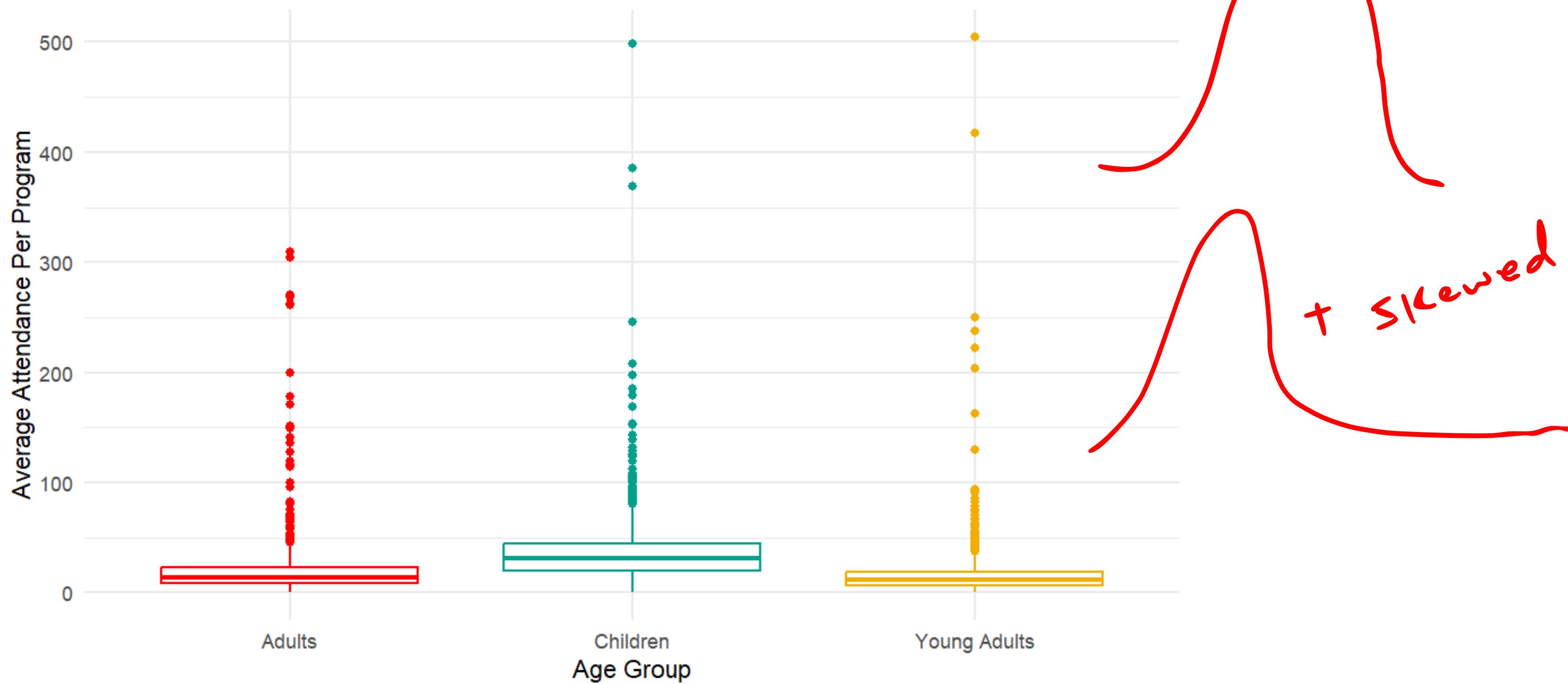




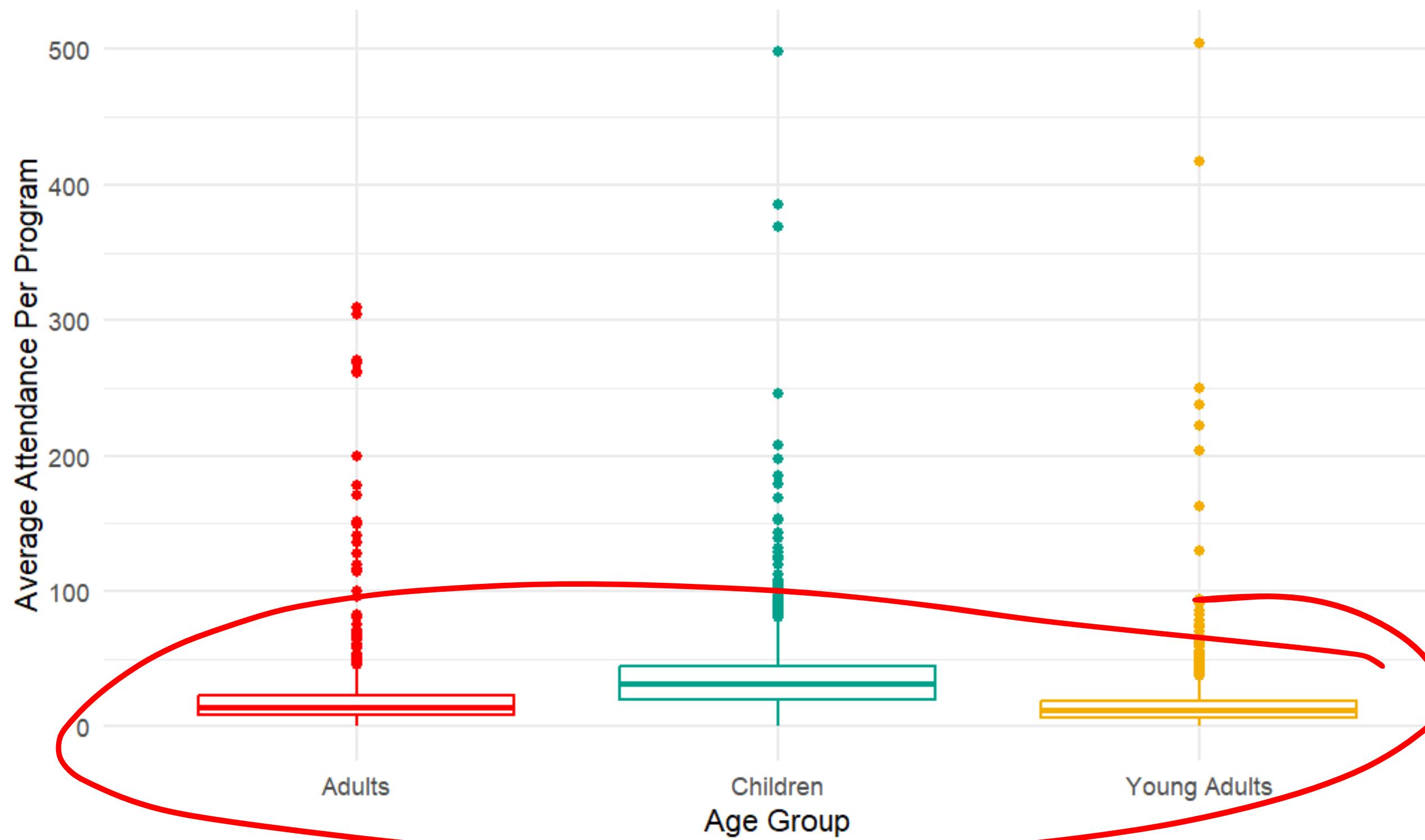
## Average Attendance Per Program



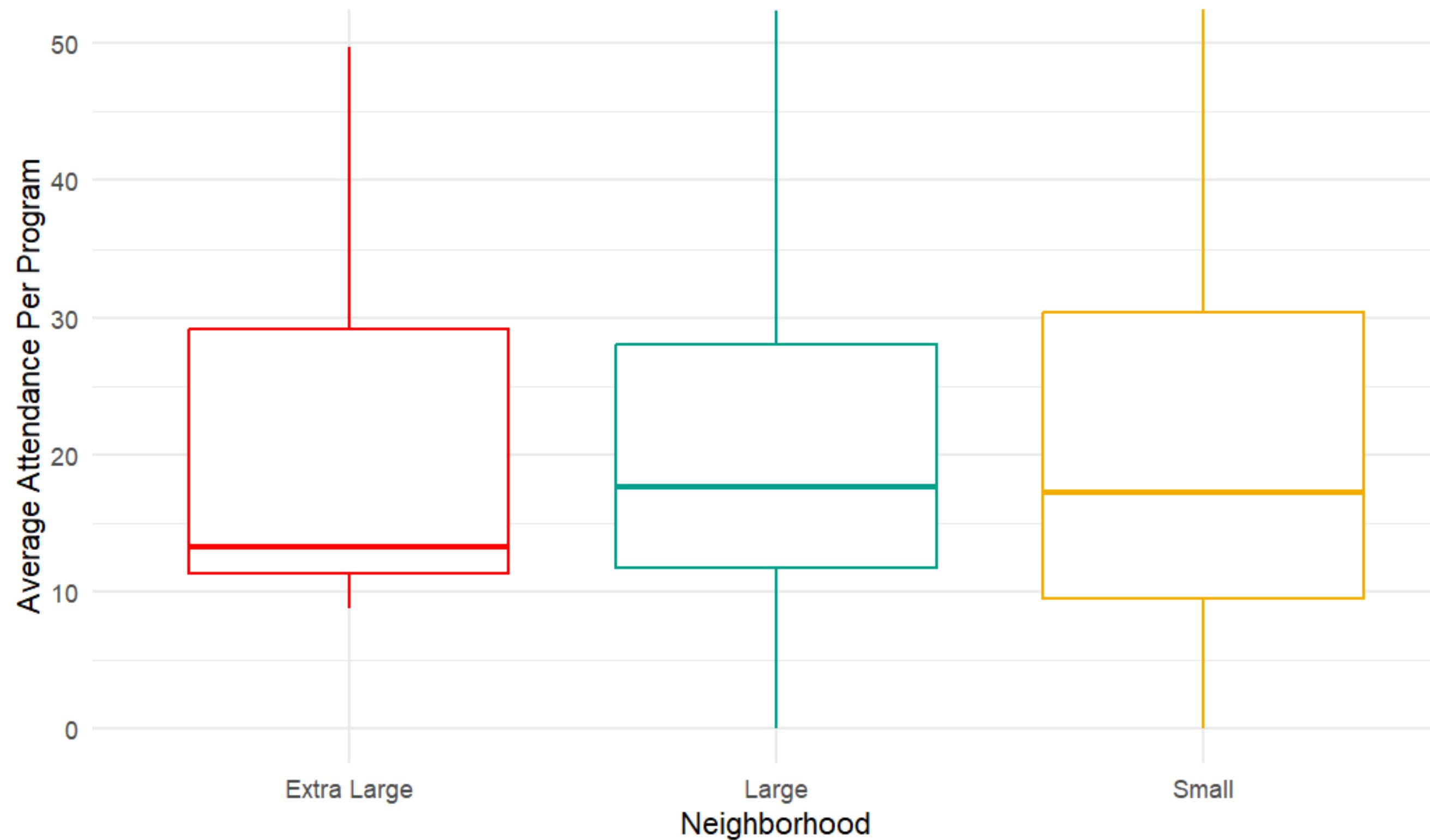
## Average Attendance Per Program



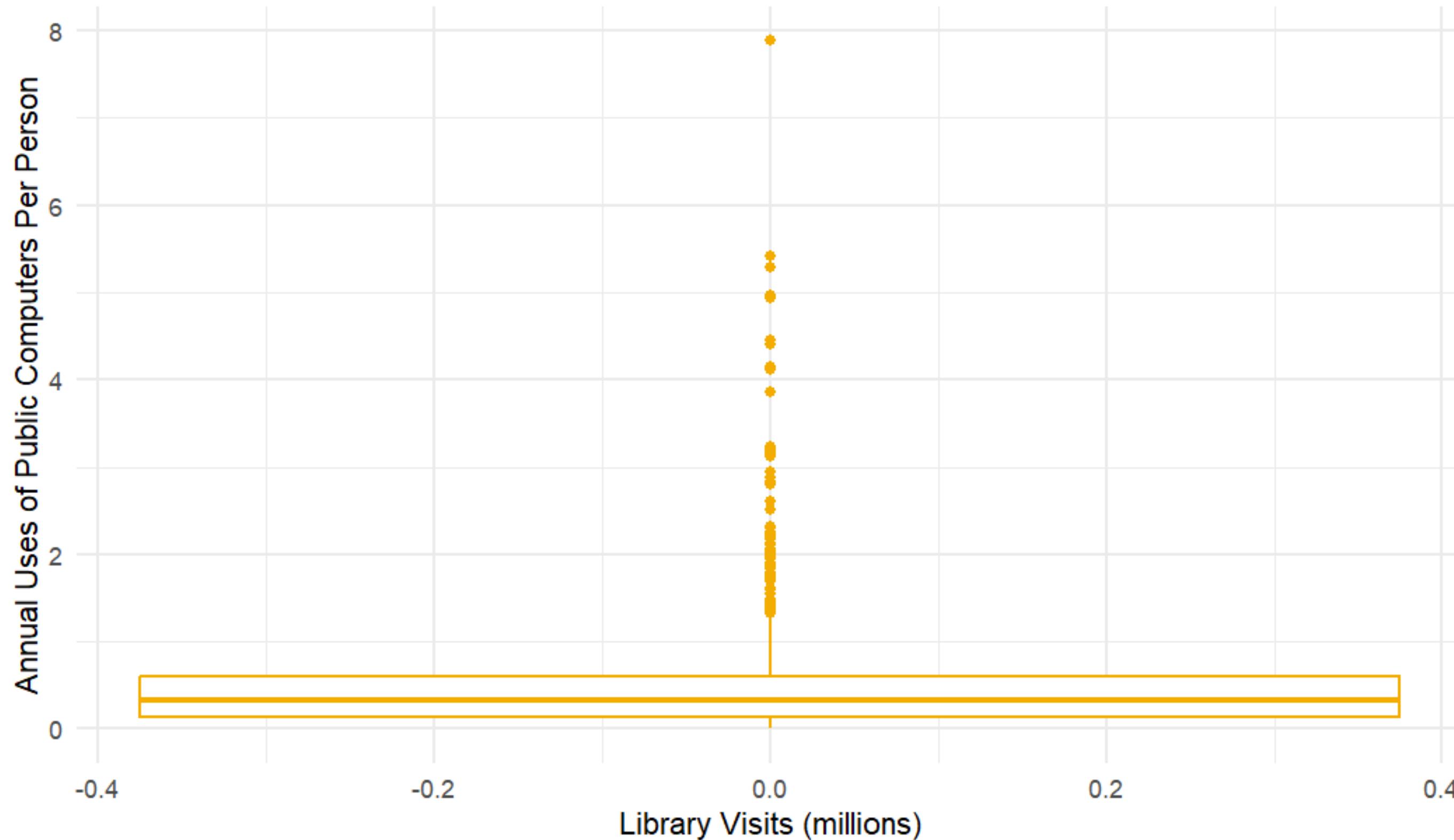
## Average Attendance Per Program



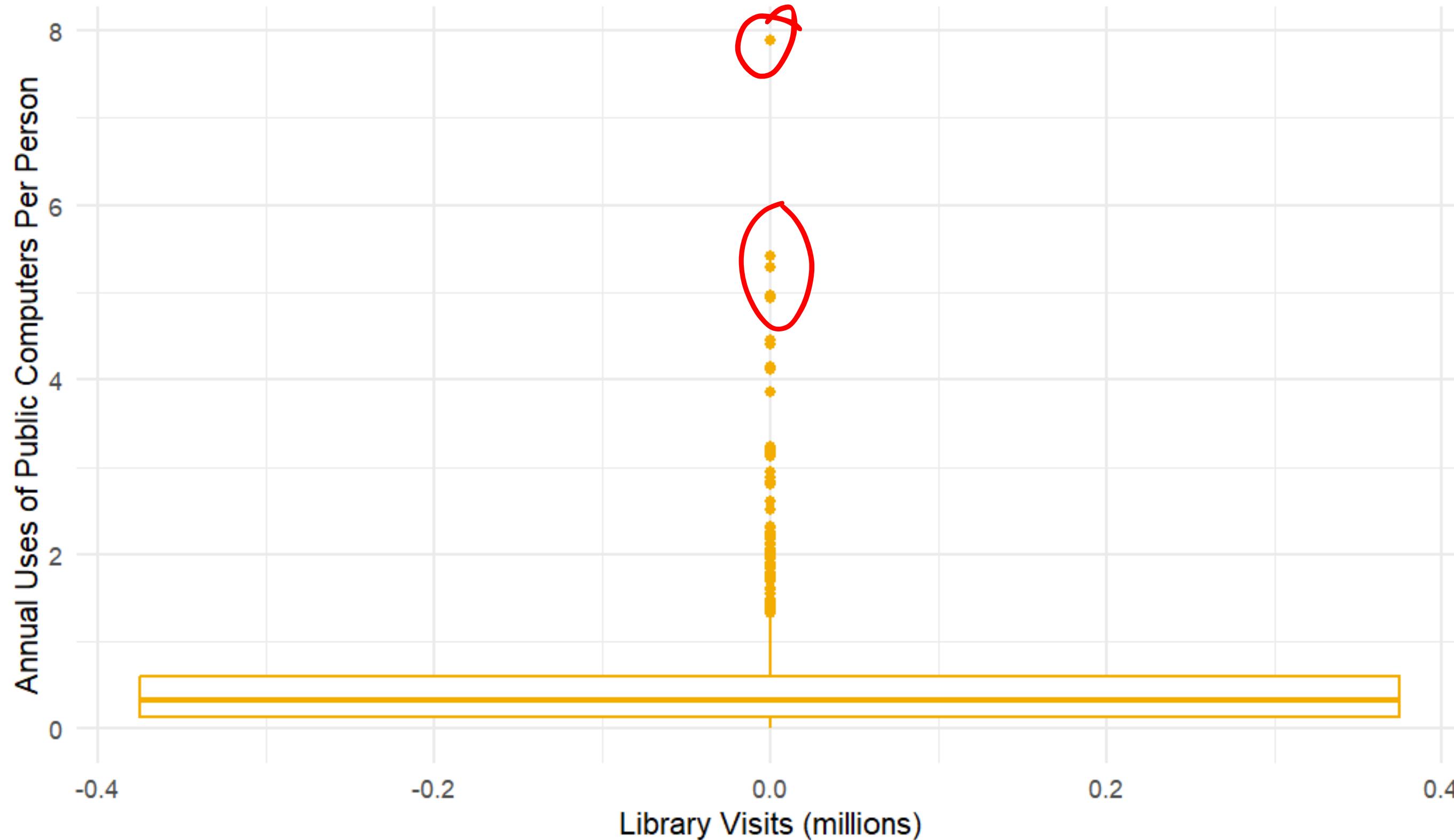
## Average Attendance Per Program (Zoomed Into the Median)



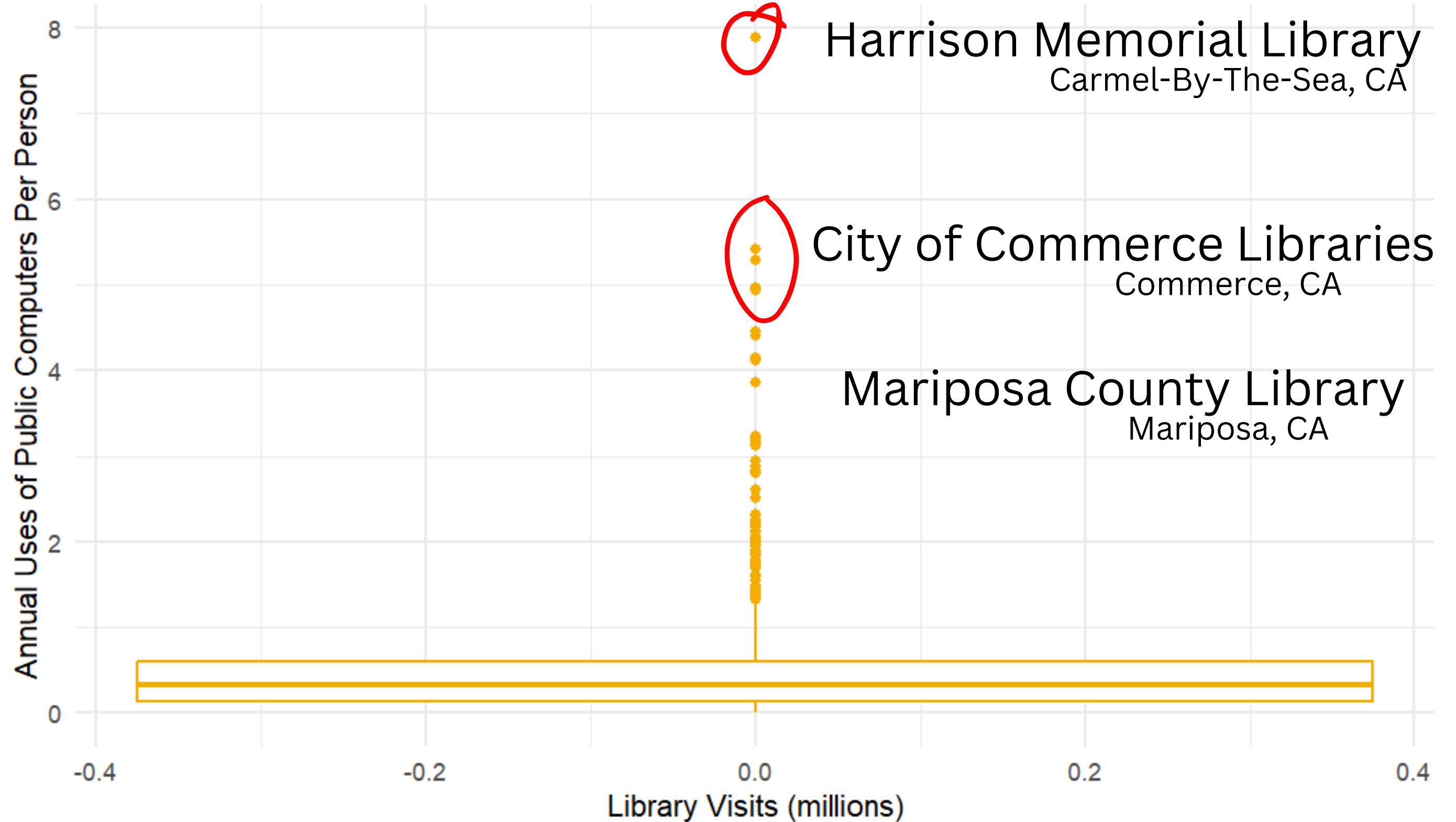
## Computer Use Per Person and Library Visits



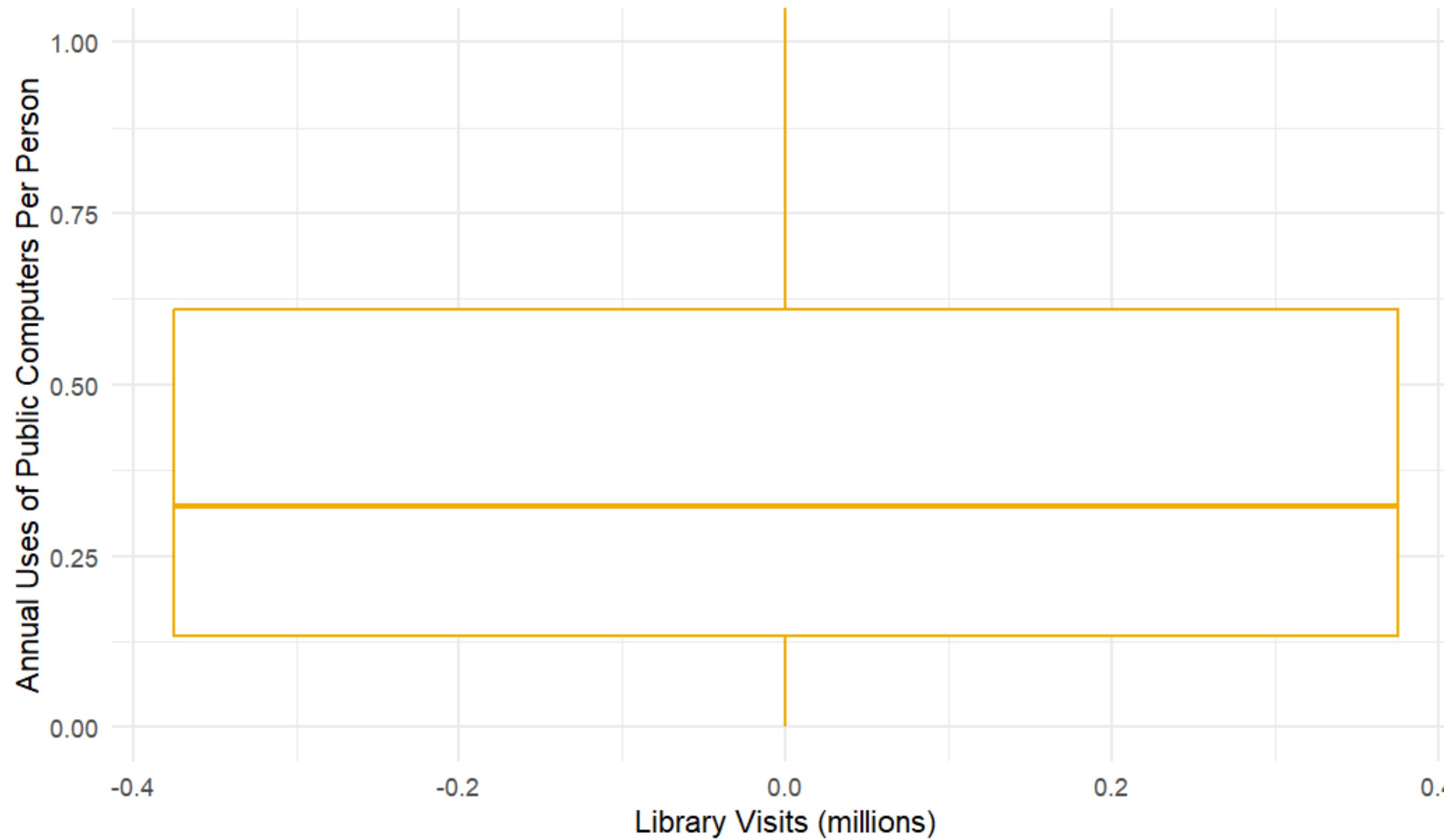
## Computer Use Per Person and Library Visits



## Computer Use Per Person and Library Visits



## Computer Use Per Person and Library Visits





01.

## RESEARCH PROBLEM

*Define research problem*

02.

## DATA DESCRIPTION

*Dataset and variables used*

03.

## DESCRIPTIVE STATISTICS

*Exploration of library statistics*

04.

## MACHINE LEARNING MODELS

*SVMR and Random Forest*

05.

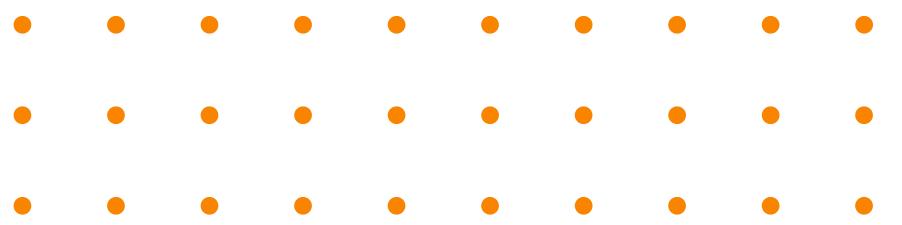
## MODEL RESULTS

*What did we learn from the models?*

06.

## RECOMMENDATIONS

*Ideas for further research*



# TABLE OF CONTENTS



01.

## RESEARCH PROBLEM

*Define research problem*

02.

## DATA DESCRIPTION

*Dataset and variables used*

03.

## DESCRIPTIVE STATISTICS

*Exploration of library statistics*

04.  
 

## MACHINE LEARNING MODELS

*SVMR and Random Forest*

05.

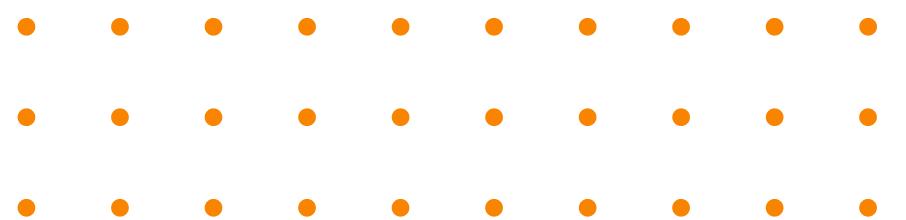
## MODEL RESULTS

*What did we learn from the models?*

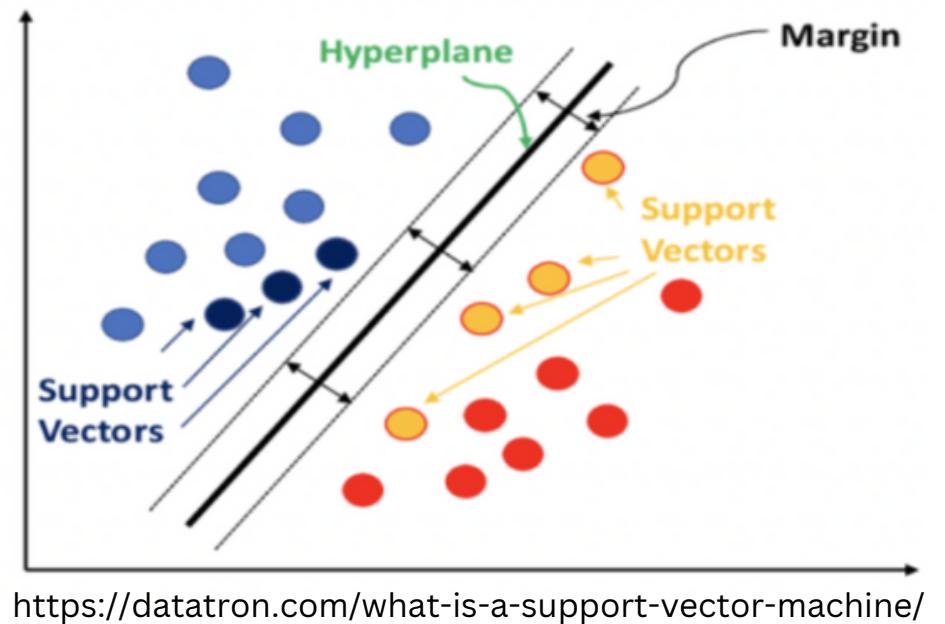
06.

## RECOMMENDATIONS

*Ideas for further research*

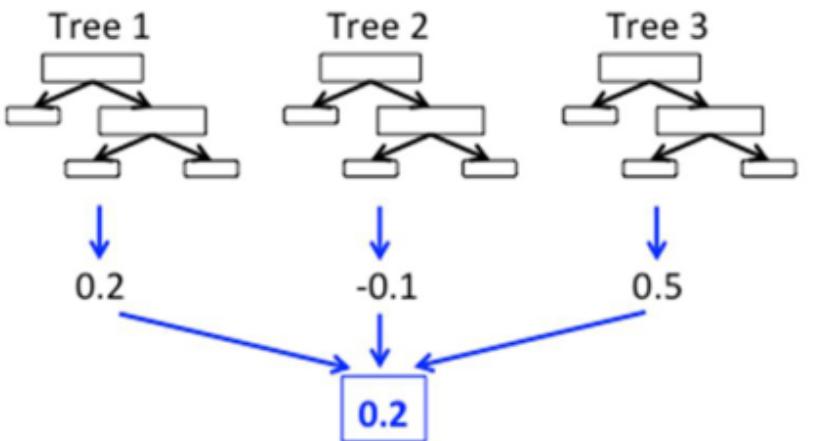
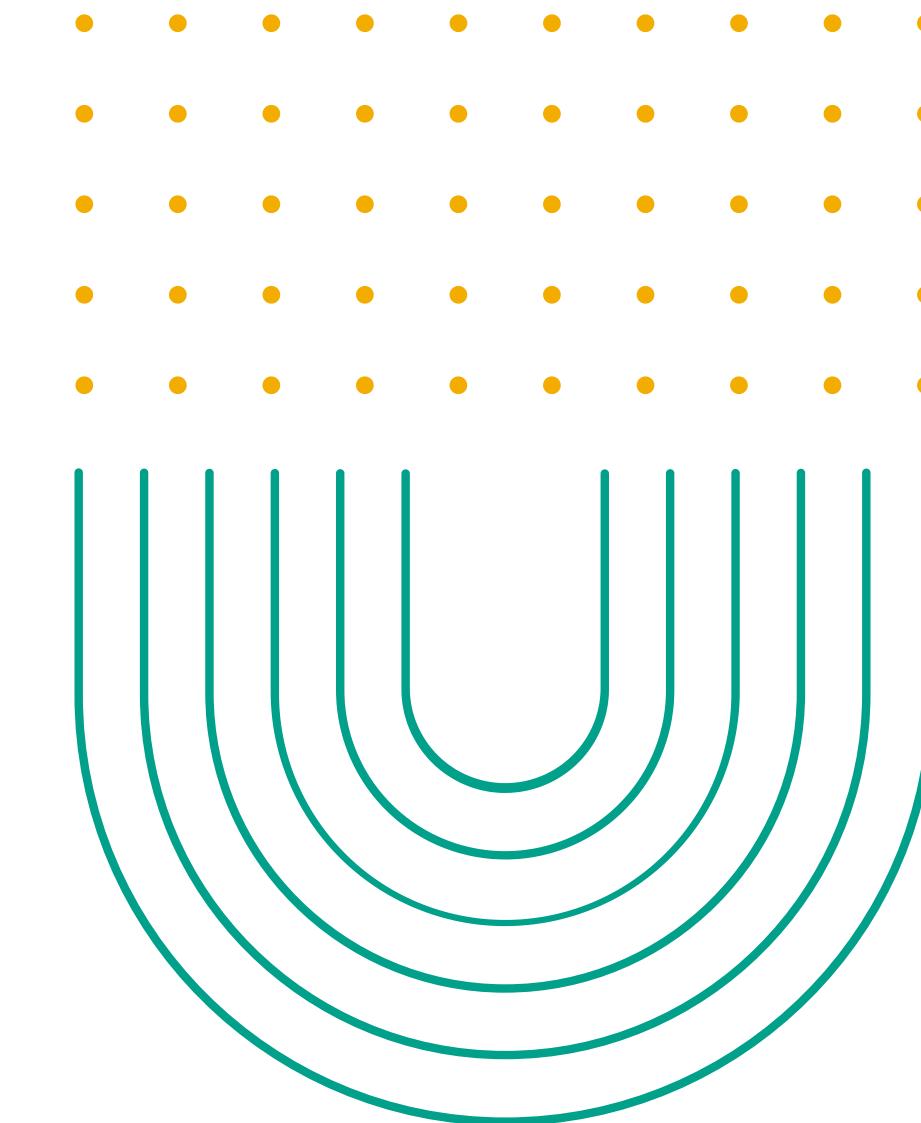


# TABLE OF CONTENTS



## Support Vector Machine Regression

models the relationship between variables and target values by finding the hyperplane that maximizes the margin while minimizing the error between predicted and actual values.



## Random Forest

constructs multiple decision trees during training and outputs the average prediction of the individual trees

**ML  
MODELS**

# PREDICTION

SVMR

Mean Squared  
Error (MSE)

0.20

Root Mean Squared  
Error (RMSE)

0.44

Mean Absolute  
Error (MAE)

0.30

R-Squared

0.79 OR 79%

RF

3.22

1.79

1.18

0.82 OR 82%

# PREDICTION

SVMR

RF

Mean Absolute  
Error (MAE)

0.30

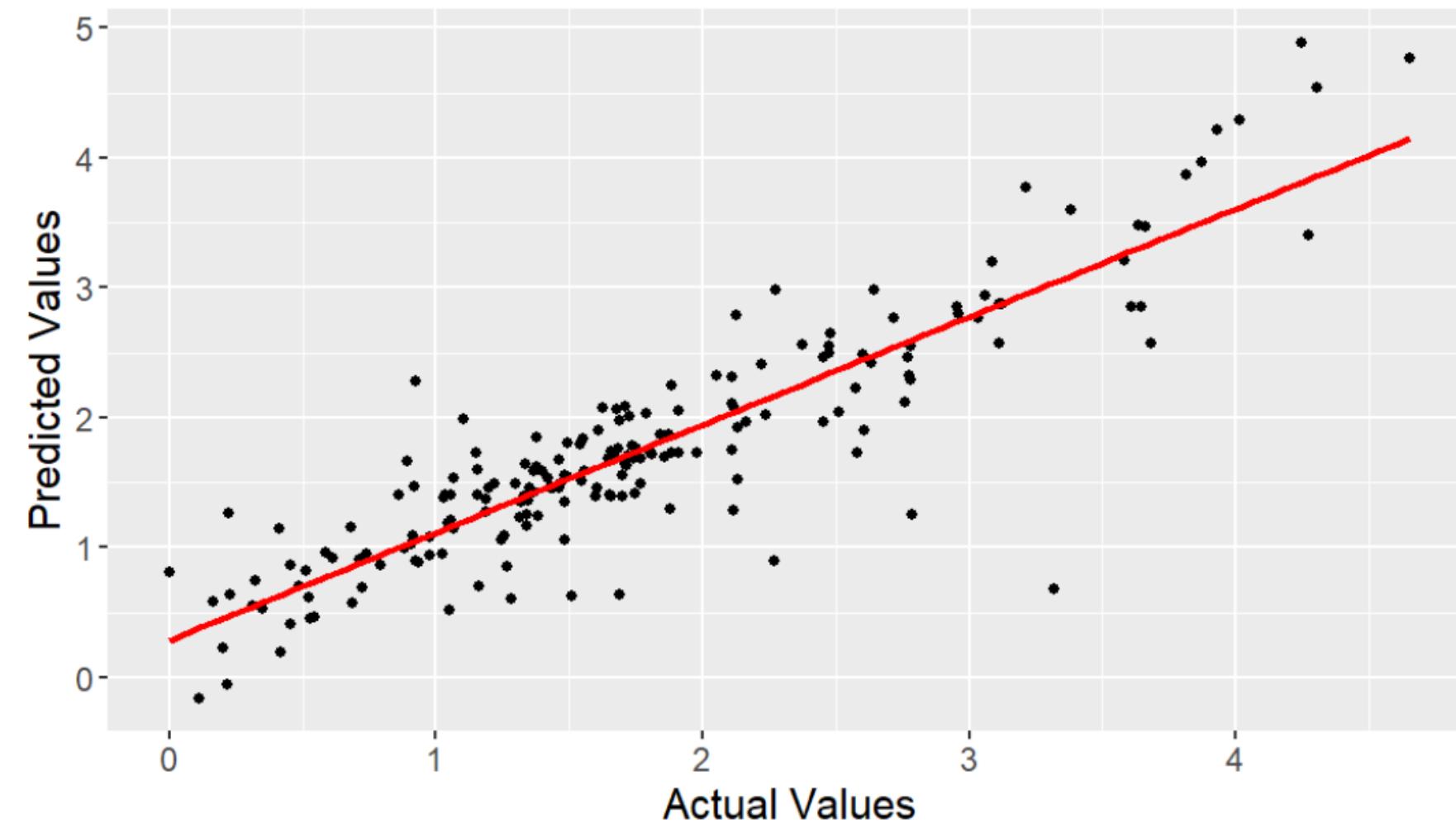
1.18

R-Squared

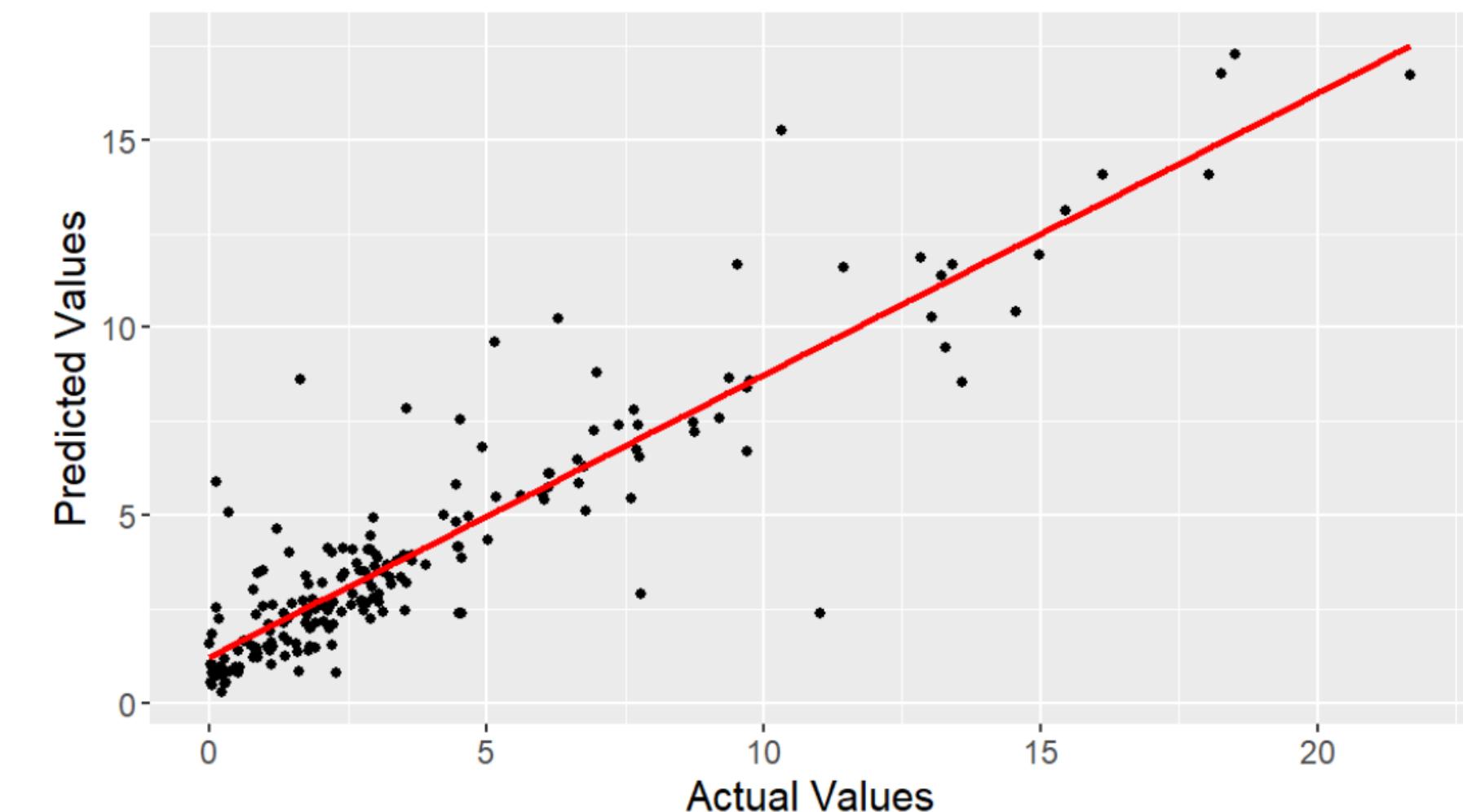
0.79 OR 79%

0.82 OR 82%

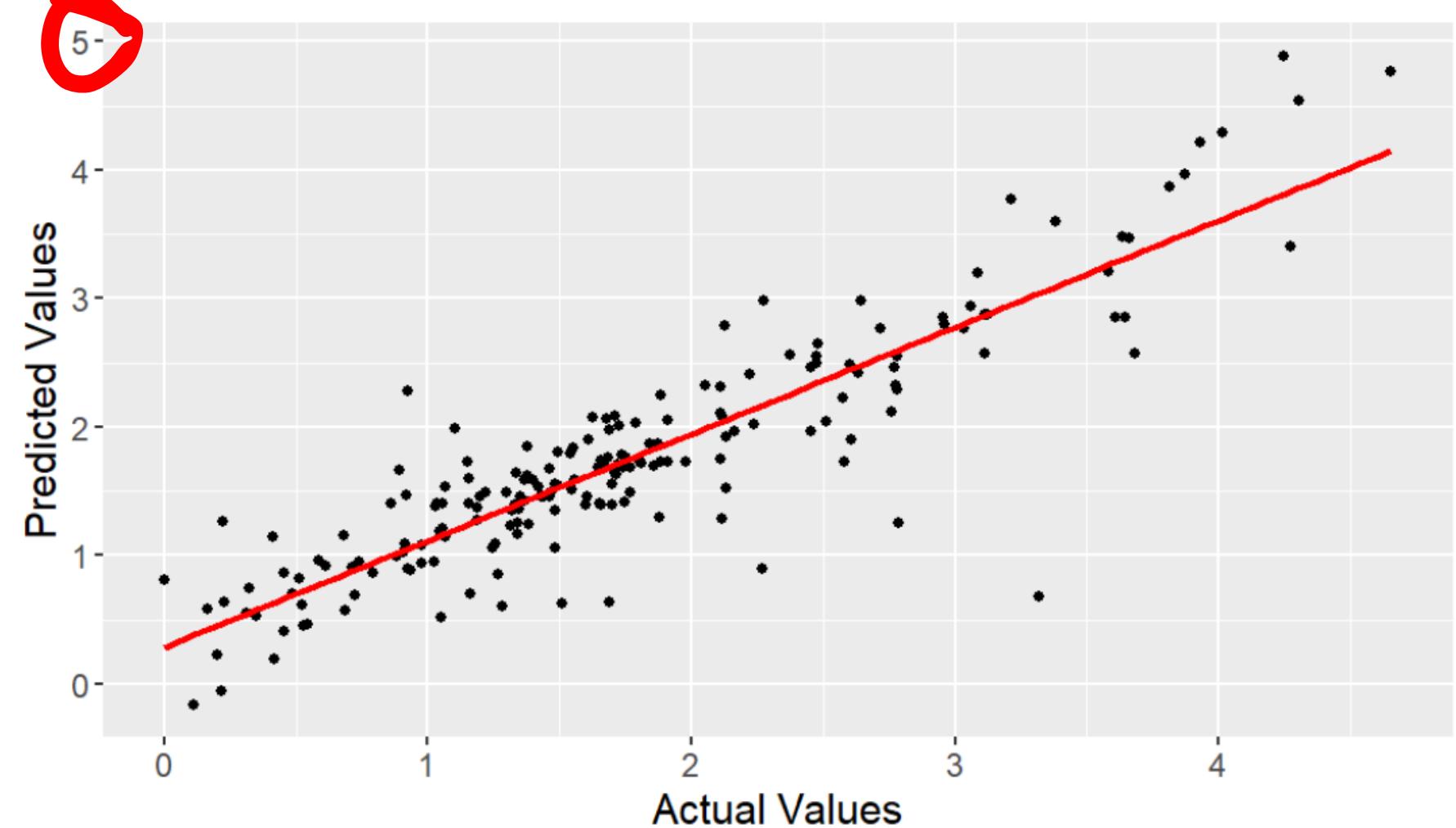
SVMR: Actual vs Predicted Values



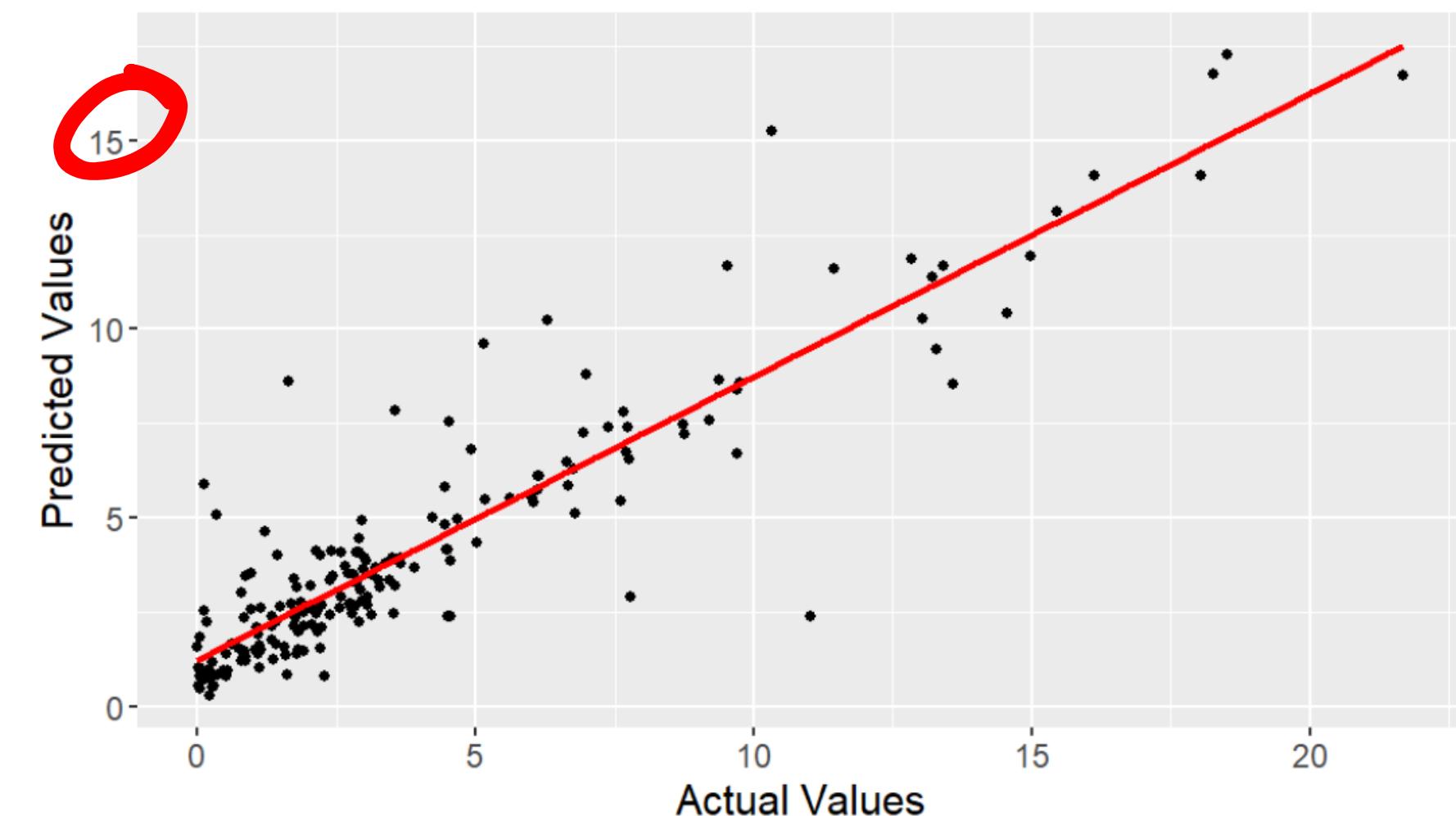
Random Forest: Actual vs Predicted Values



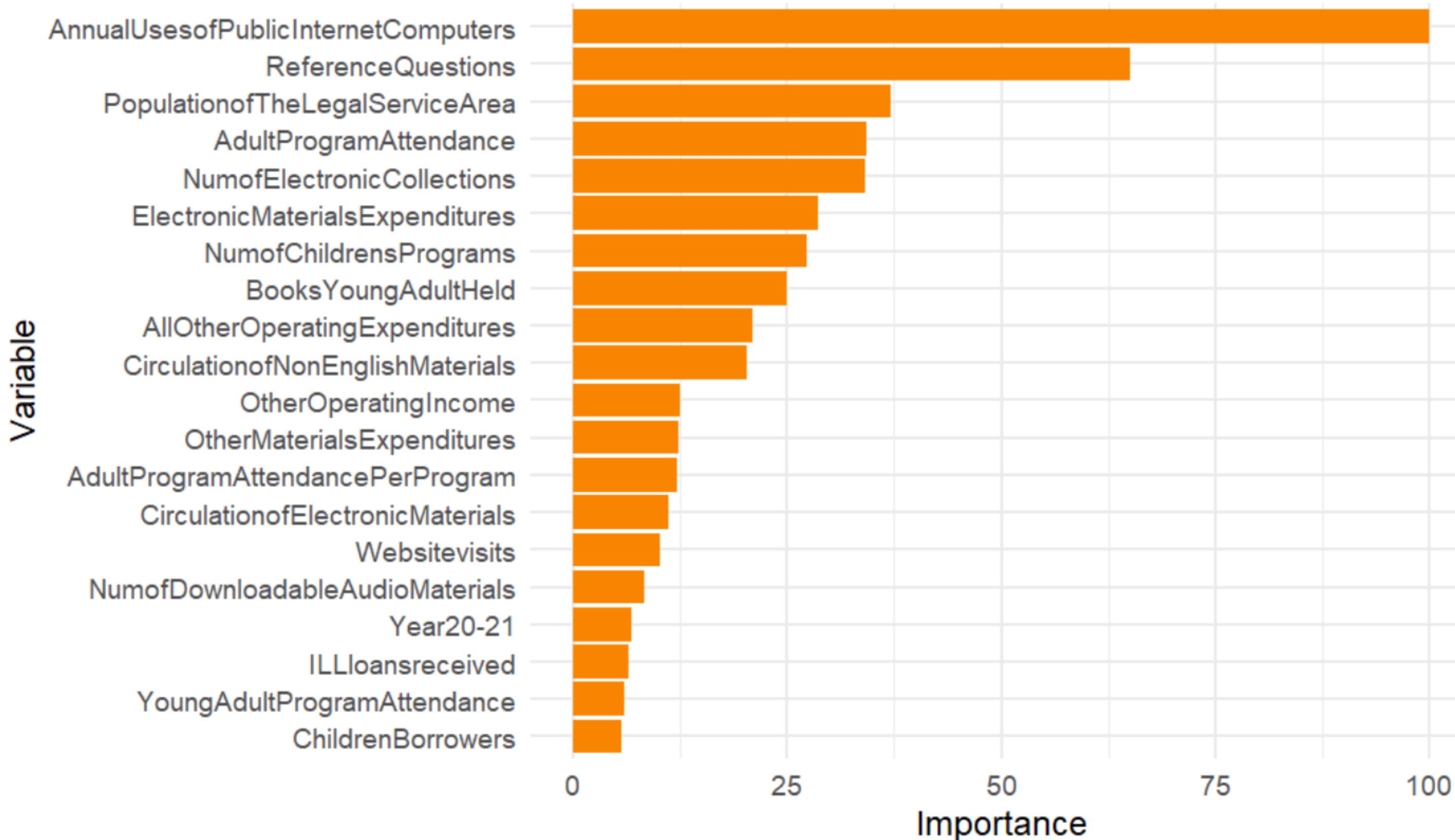
SVMR: Actual vs Predicted Values



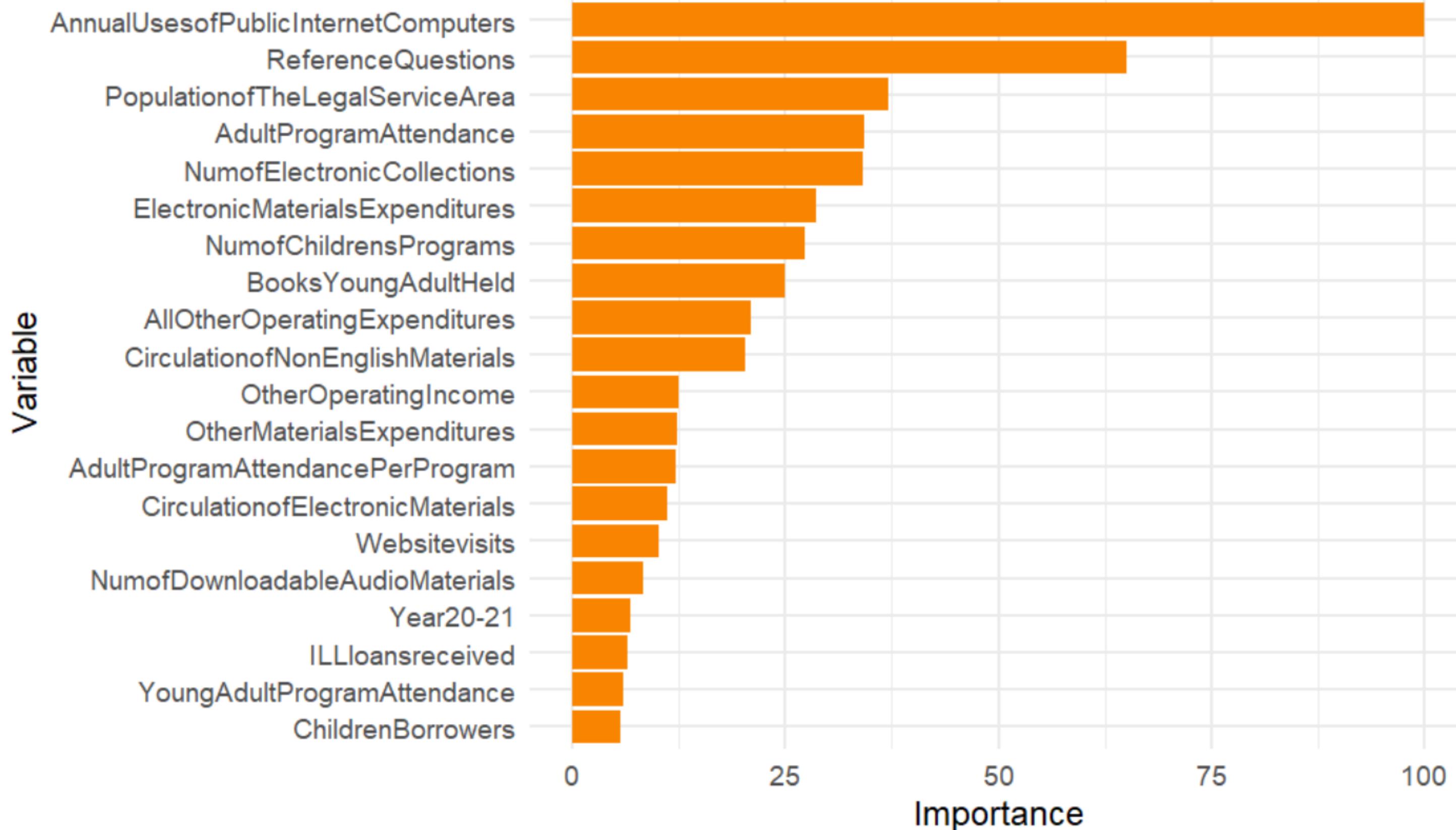
Random Forest: Actual vs Predicted Values



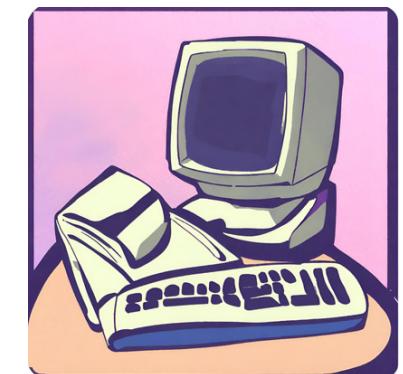
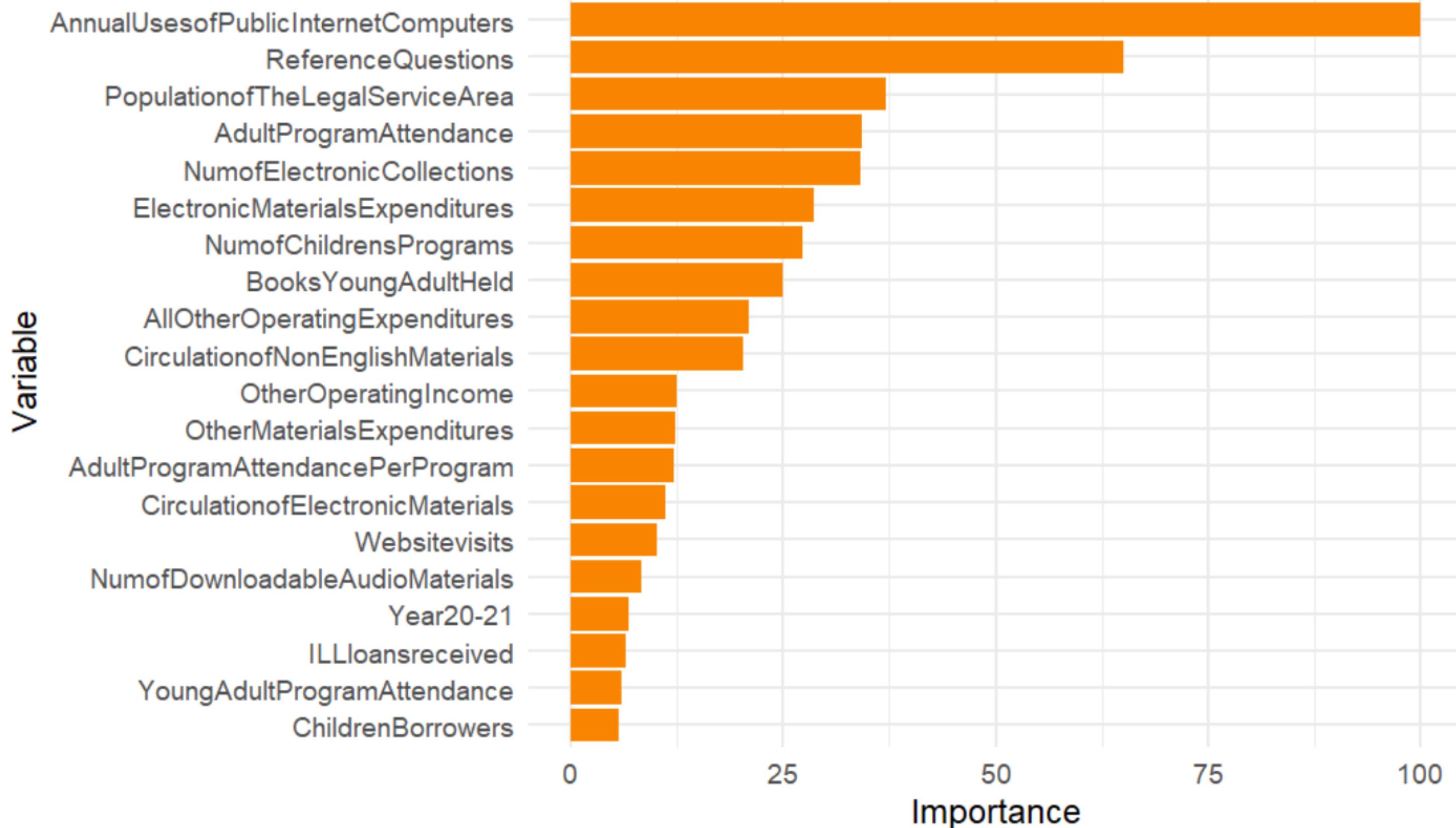
## Random Forest: Variable Importance Plot



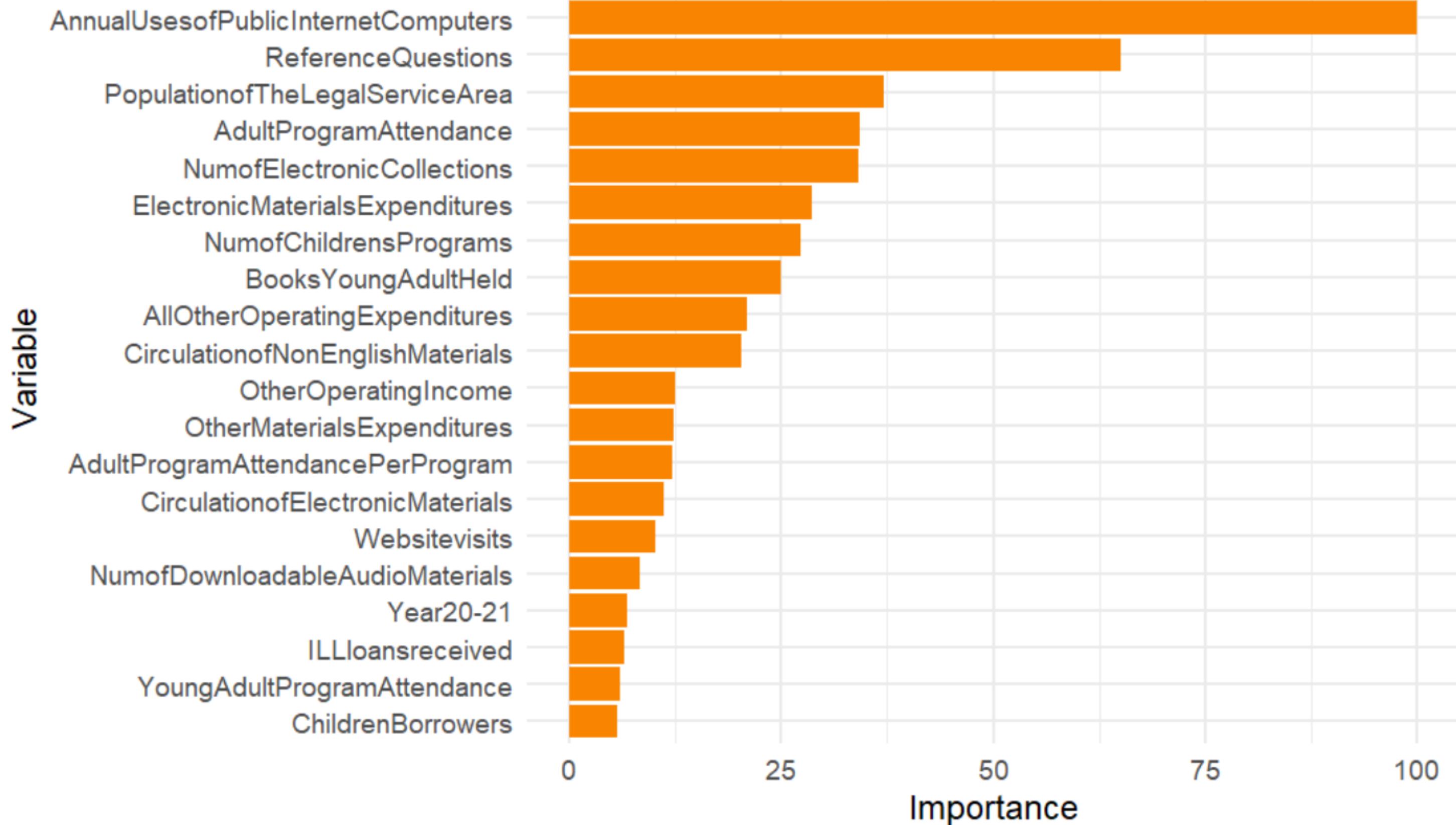
## Random Forest: Variable Importance Plot



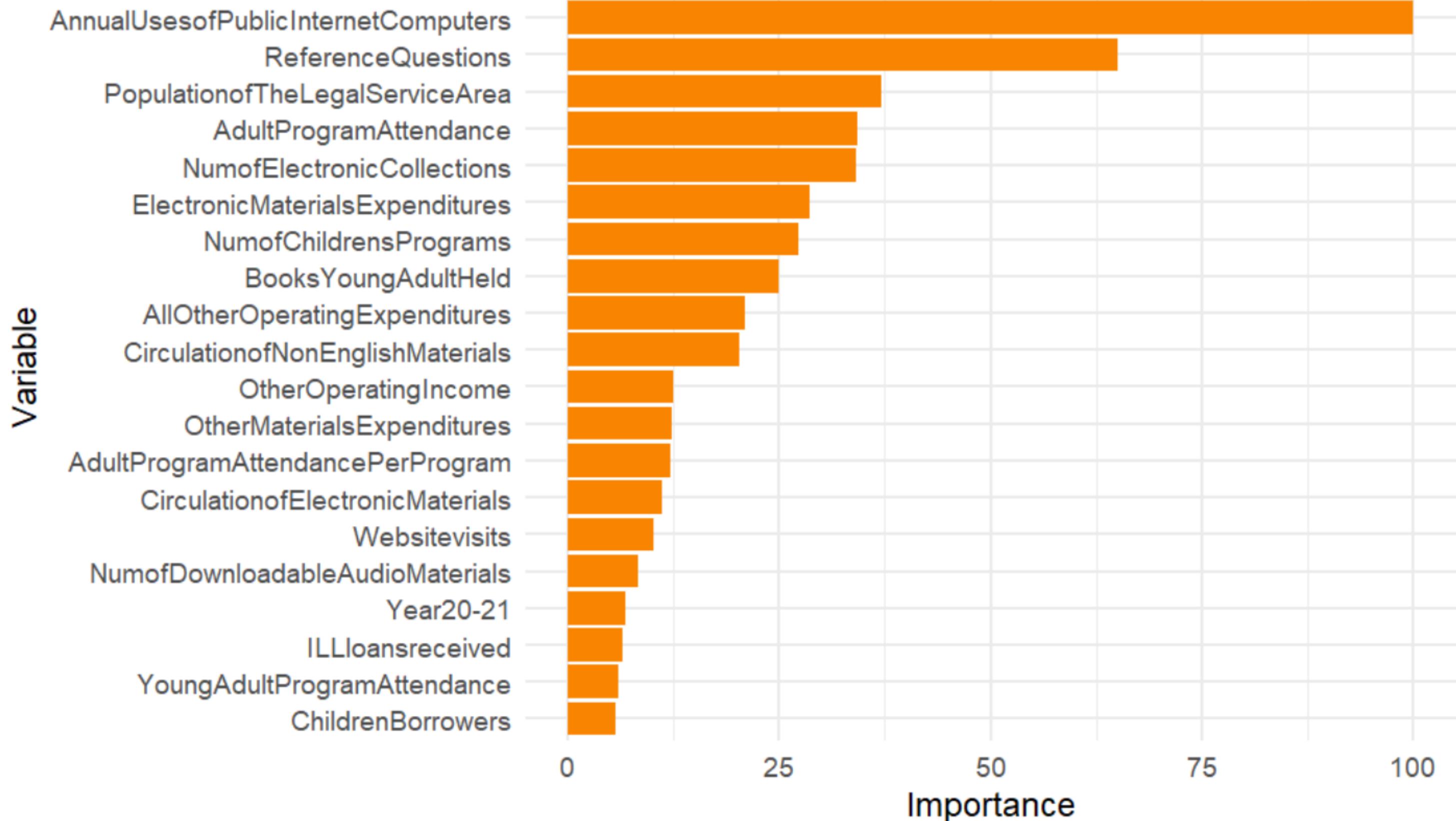
## Random Forest: Variable Importance Plot

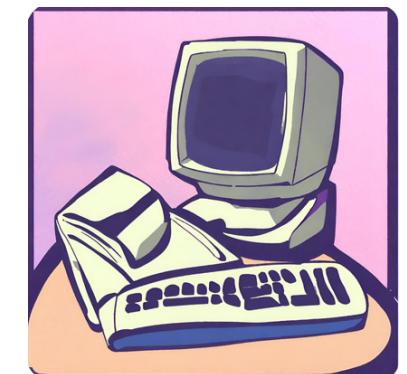


## Random Forest: Variable Importance Plot

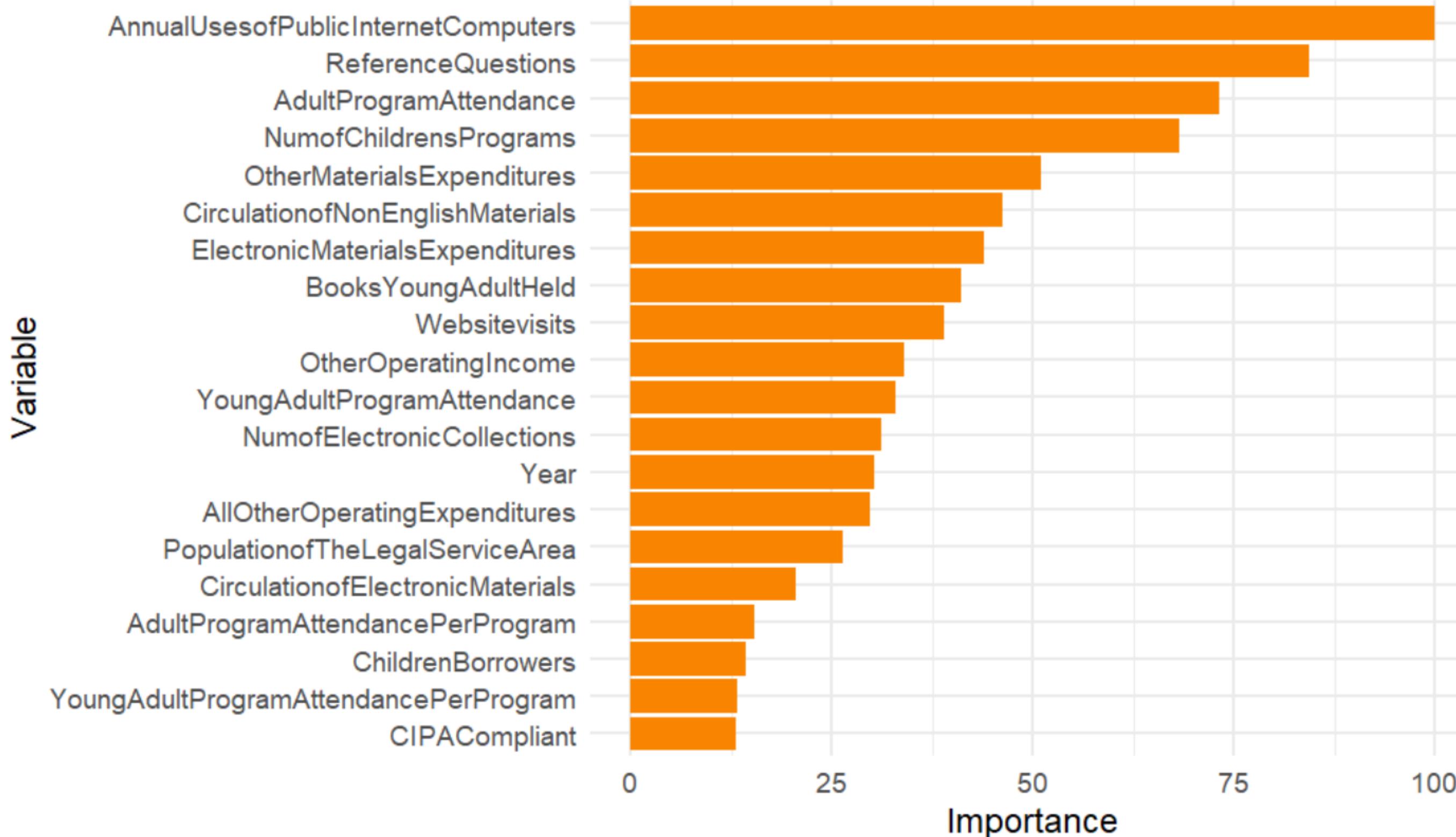


## Random Forest: Variable Importance Plot

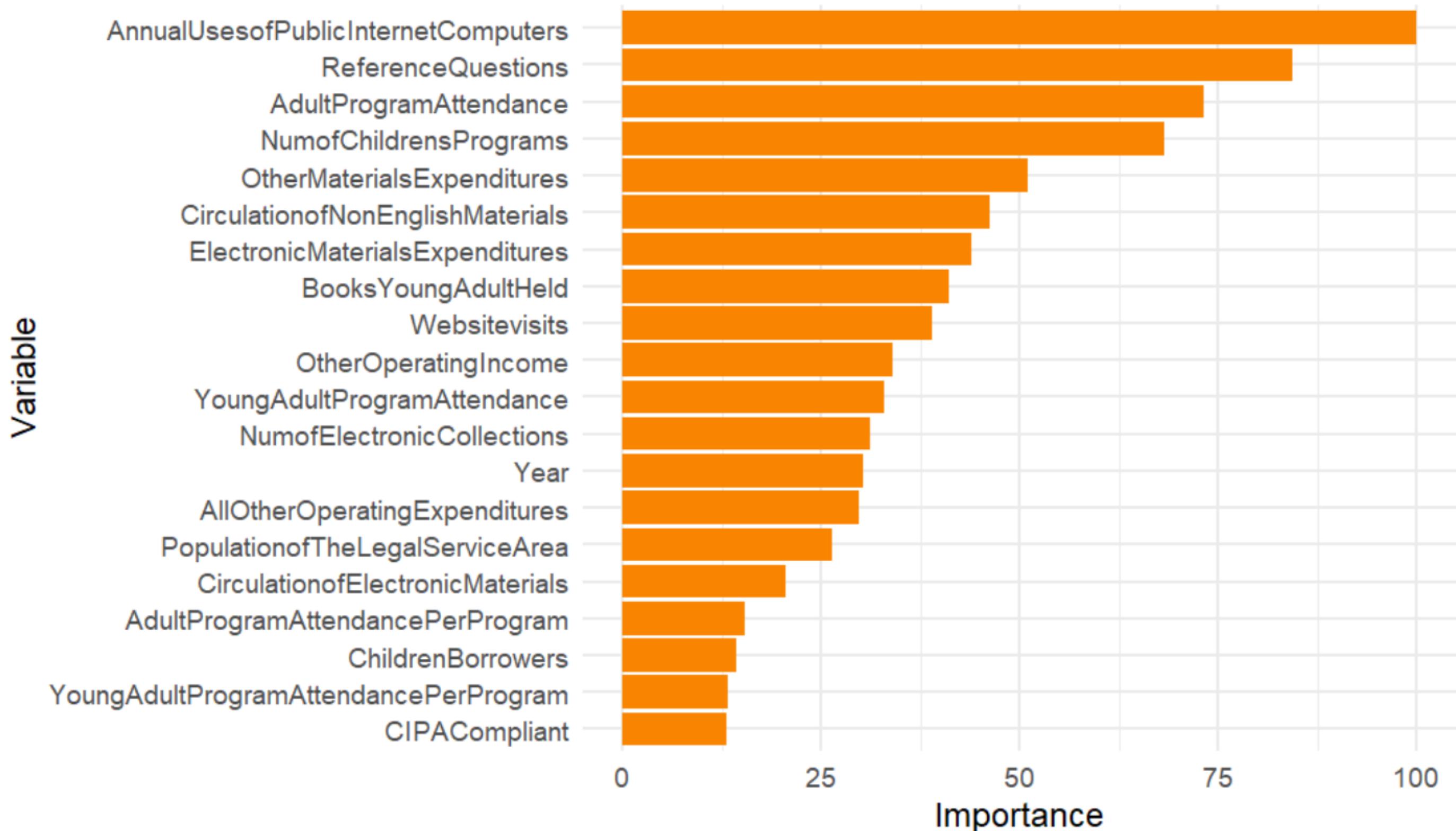




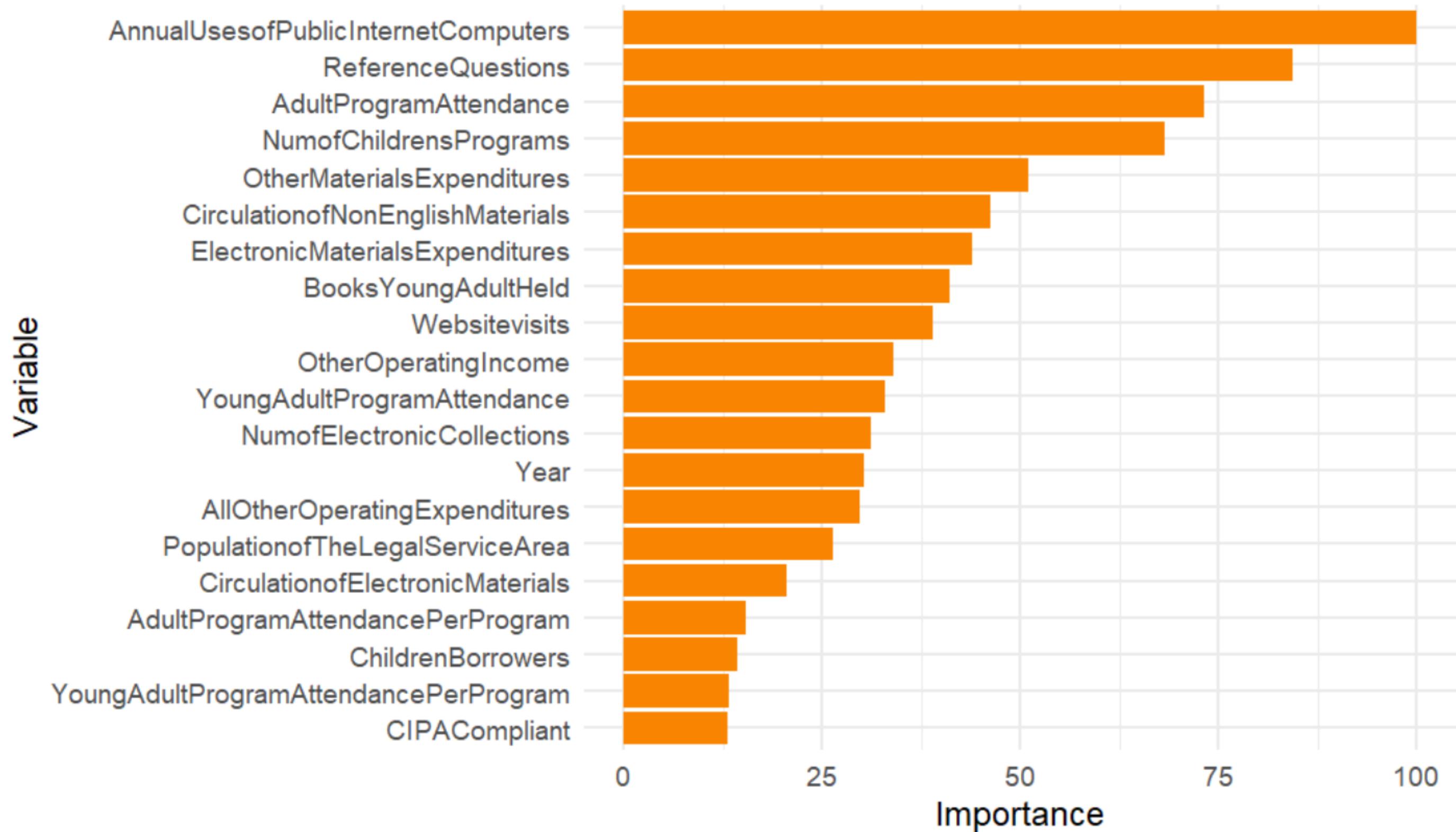
## SVM: Variable Importance Plot



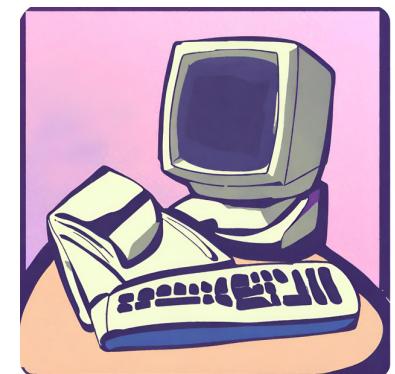
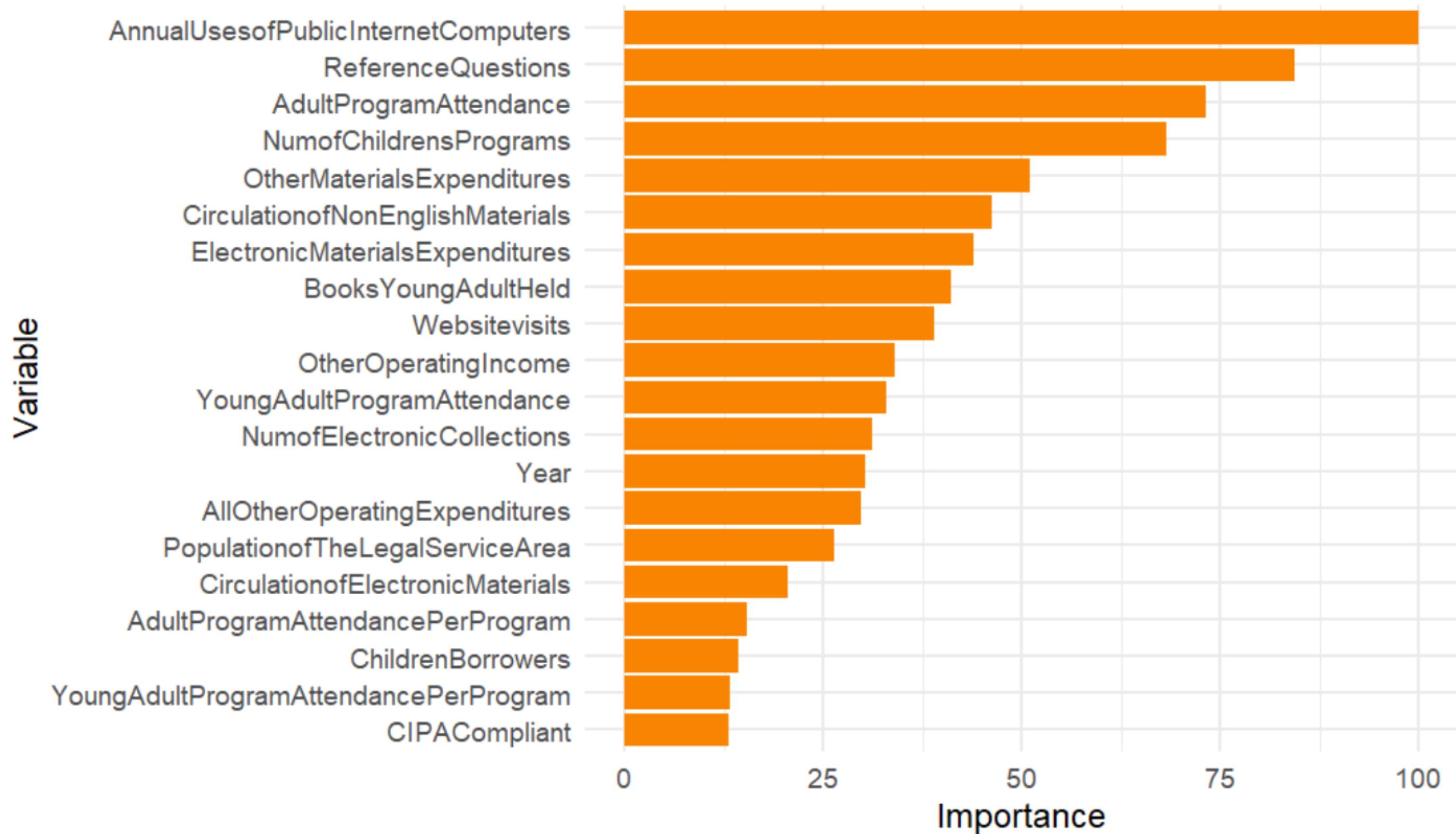
## SVM: Variable Importance Plot

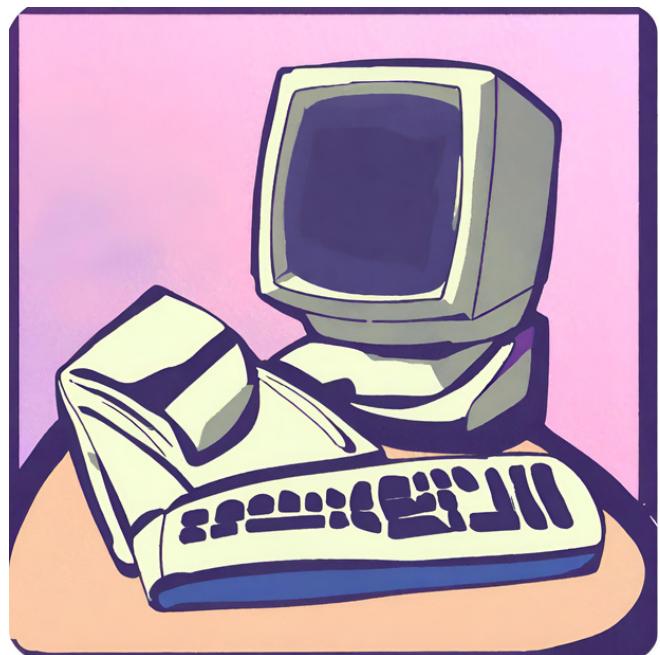


## SVM: Variable Importance Plot



## SVM: Variable Importance Plot





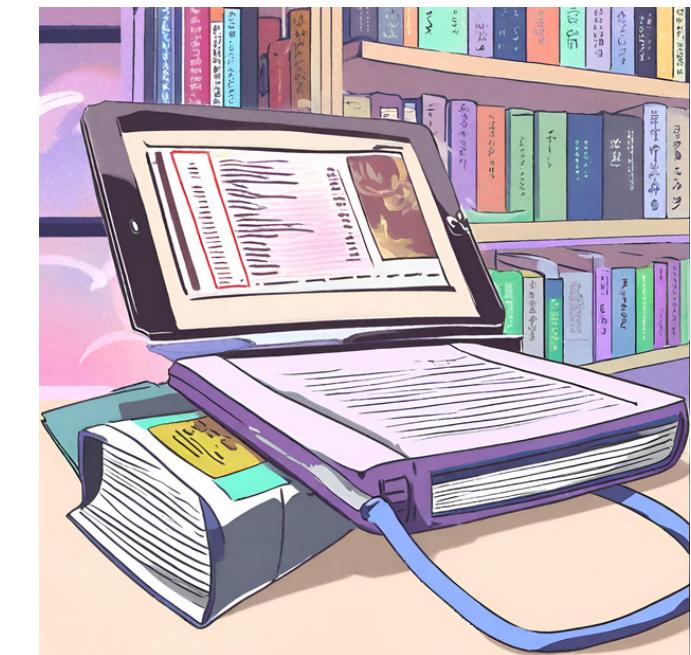
Computer Usage



Ref Questions



Programs for Adults



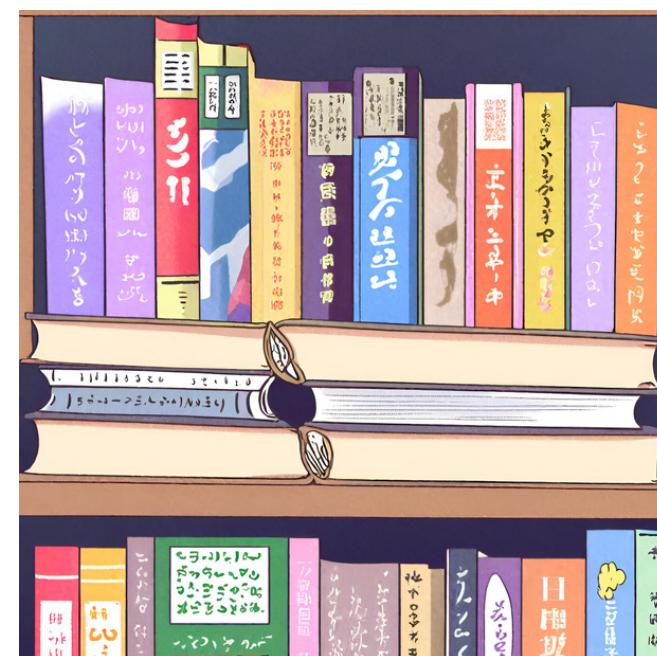
Electronic Materials



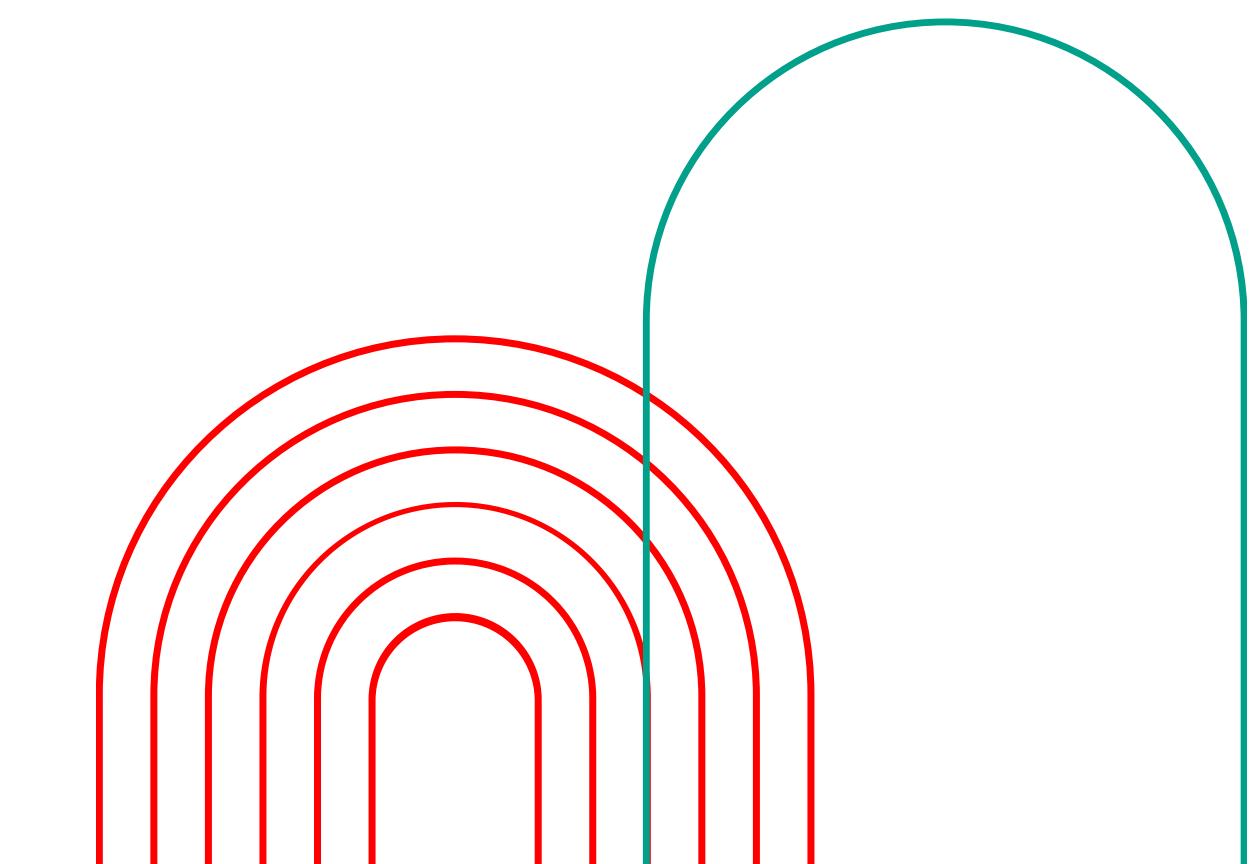
Programs for Children



Expenditure



Non-English Materials

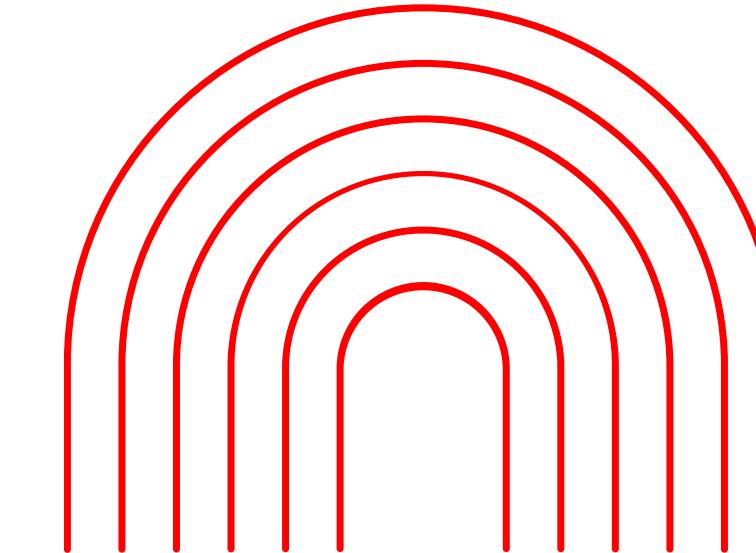
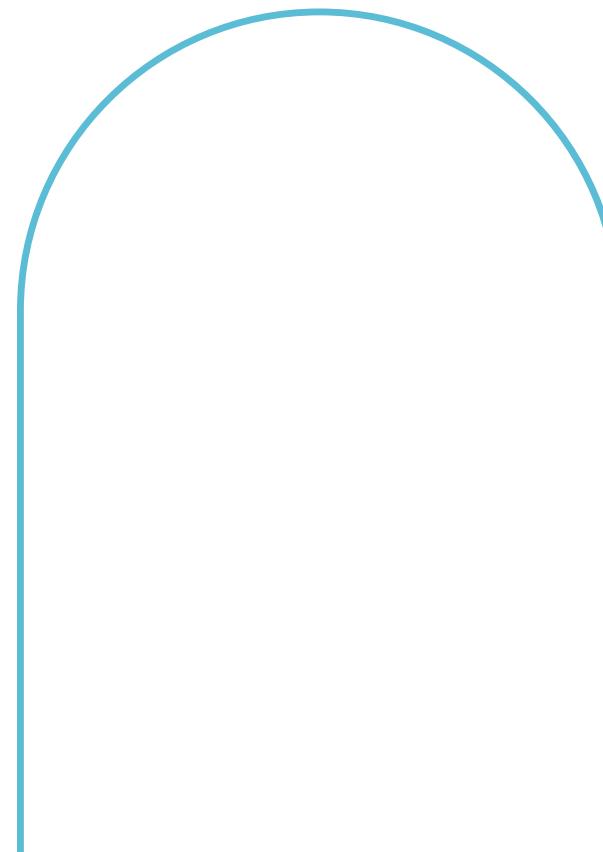
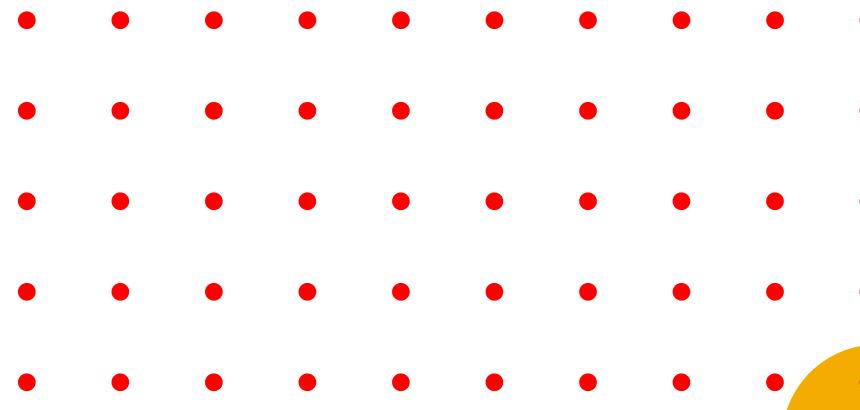




Computer Usage

# RECOMMENDATIONS

1. Conduct qualitative studies  
(formal or informal)
2. A/B Experimentation
3. Complete this study again  
with census data



**THANK YOU**