MATH 448 Project Progress Report II

Ulices Gonzalez

April 22nd, 2025

# Trends in Library Use Before, During, and After the COVID-19 Pandemic

## Introduction

In March of 2020 COVID-19 was declared a Pandemic Worldwide which disrupted infrastructure, business, travel, and the lifestyles of everyone. Given the way the virus is transmitted many libraries had to close their doors. The American Library Association (ALA) surveyed libraries nationwide which included public and academic libraries, in their findings the ALA reported that almost 99% of participating libraries had limited to no access to their branches. Between March of 2020 to April 2021 most libraries were closed to the public and offered alternative services such as Front Door Service and Virtual Library Catalogs. Having been hired at the time of libraries being reopened nationwide, I and other members of the public were not, and still not aware of the scale at which the Pandemic impacted the use of Library Branches. Libraries are not just used for checking out reading and visual materials but offer many resources to the public which range from computing, literacy programs, student meals, and many more necessities which many members of the public rely on. By looking at patron statistics from 2019 to 2024 it allows a better insight into the current state of library use while allowing us to understand the decline, recovery, and plateau libraries have had to face and still do and what factors have influenced this since.

## Description of Data and Variables

The Chicago Public Library is home to 8 branches across 77 different community areas throughout the city of Chicago and has provided statistics over their library operations through Data.gov for the last 12 years. Visitation by Location is a report of Patrons that visited a given library for services provided while excluding community room and program activities. The Holds by Location is a report of holds placed by patrons at a given library, does not include holds placed online. The Computer Sessions by location is a report of 1 hour and 15-minute computer sessions at a given library, does not exclude consecutive sessions from a given patron. Item Circulations by Location is a report of Items checked out and renewed at a given library branch, renewals are treated as new check outs and are not their own statistic. All four Data Sets present statistics by month with a year-to-date total for each library branch associated with the Chicago Public Library.

By using this data, we can examine trends in library usage but also the impact the pandemic had on the usage of library services overall. Specifically, by looking at the data from 2019 to understand the statistics prior to the pandemic we can get a good idea of where library usage peaked. Considering that the pandemic was officially declared in 2020 the best data to look at for the immediate impact would be 2021 as the data for 2022 would be after some safety guidelines were lifted. Lastly, to see the overall impact the pandemic has had a year after it was officially declared over the data from 2024 will be considered and compared along with that of 2019 and 2021.

## Data Processing

Due to the manner in which The Chicago Public Library data sets are provided, many predictors are separated into their own tables which require some additional work in order to combine them with other statistics collected in a give year. For example Item Circulation, Holds, and Computer Sessions are reported in their own sets so in order to look at their influence on Visitation. This took some work and adjustment but was a feasible task due to the size of each given data set being fixed. Some pieces of data were excluded from the final data set particularly online items, holds placed by patrons outside of a branch, and automatic renewals as these padded the statistics in a manner that treated them as their own library branches.

When processing the data from data.gov into r I began by reading each file and making the process of this easier by creating a function that could handle this faster. Once all the files have been read and I took a look at the data I realize that it would be tedious for each predictor to be output as their own data set. From here I began the process of compiling all the data for a given year into a table which led to the creation of the process\_year\_data function that would compile all the statistics into a single table with each predictor as its own column. This gave a really good output as the data can now be called however the caveat to this is that each library has a row every month of the year meaning that all 80 libraries now had 12 rows leading to each table for a given year having 960 different rows to examine. This was kept as this allows for an examination of extremes and other relevant data but in order to get a better overview the aggregate\_year\_data function was created to compile the data further by calculating the sum of all the statistics for a given library. With this now being completed it allowed for the data to finally be analyzed with greater ease and cleaner statistics.

## Descriptive Analysis

Above are the summaries of the data for 2019, 2021, and 2024 which display some relevant data regarding the use of public libraries within the city of Chicago. It’s important to state that the minimum values that are 0 correspond to the closure of a branch during a given month, these values were still included as the library branches would still be used during other times of the year. Considering that the data from 2021 is from during the pandemic, it goes without question that the lower number across all predictors corresponds to the impact that the pandemic had. However, it is notable that even a year after the end of the pandemic the average for all four predictors is much closer to that of the pandemic data. Specifically, if we look at the Mean visitation from 2021 it is under half of what the Median was in 2019, in 2024 while still higher by 25000 it shows that while library visitation is slowly coming back that the pandemic is having a long-lasting impact on the use of libraries throughout the city. However, circulation has remained consistent throughout 2021 – 2024 if we look at the means, Holds and Computer Sessions dropped during the pandemic but are showing signs of recovery with holds being slightly above what was reported in 2019 however the same cannot be said for Computer Sessions as they are still slowly increasing but are still below that of 2019. (Obvious outliers are not being considered in this analysis)

Over the course of the 5 years being examined there has been a clear change in the scale of visitation throughout the libraries of Chicago. If we excluded the obvious outliers of the Harold Washington and Sulzer Regional Branches, visitation throughout branches was between 100000 to 300000 in 2019. However, if we look at 2021 it’s not surprising that with safety guidelines visitation would go down between 20000 to 100000. Since then, it has steadily grown with it reaching between 50000 to 175000 in visitation but is a significant decrease from what it once was in 2019. (Obvious outliers are not being considered in this analysis)

In 2019 both visitation and circulation were consistent with them having a close relationship with one another. Moving into 2021 there is more variance in the relationship as it seems there was an increase in circulation but decrease in visitation, but it is important to note that this is also at a smaller scale compared to 2019. In 2024 there is an increase in both circulation and visitation, however the relationship is even more spread out than it was during the pandemic with a closer relationship being seen with smaller numbers in both ends. Although both have gone up it is still clear that it is at a smaller scale than in 2019 and only slightly larger than that of 2021. (Obvious outliers are not being considered in this analysis)

When looking at the relationship between holds and visitation it has the largest changes between the three being examined as in 2019 there was consistency in the number of holds increasing as there was less visitation which when compared to 2021 is very similar albeit at a smaller scale and more variability. However, the relationship has gone through a major change in 2024 with the increase in visitation leading to a much closer relationship between holds and visitation with it also being more consistent and even at a slightly larger scale than in 2019. (Obvious outliers are not being considered in this analysis)

Lastly Computer Sessions in 2019 had a clear relationship with visitation, with it being extremely consistent with one another with very few changes. In 2021, however, it is expected that this would become more varied throughout the pandemic along with it going down in scale. Of the three years being examined though, 2024 seems to show the largest change in variance as there is more inconsistency in the relationship between sessions and visitation. It also shows some signs of growth but is closer to that of what was reported in 2021. (Obvious outliers are not being considered in this analysis)

## Simple Linear Regression Model

Simple linear regression models were used to analyze the individual relationships between library visitation (Visitation) and each of the predictor variables: Circulation, Holds, and Computer Sessions (Sessions) across 2019, 2021, and 2024. For each year and each predictor, a separate linear model was fitted to observe the direction and strength of their relationship with Visitation. Specifically, the relationship between Circulation and Visitation appears positive and statistically significant across all three years. However it’s much more clear after applying the linear regression that there is a sharp fall off in visitation which shows the effect of low circulation on visitation numbers. Looking at the results from the regression for 2024 there is an increase both in circulation and visitation with the relationship becoming much more positive.

Looking at the Relationship between visitation and holds shows a similar Trend however after applying the linear regression the relationship is much more consistent between 2019 and 2021 but looking forward to 2024 there is a large increase in positivity between holds and visitation. While it is significant statistically it isn’t the most significant of the three predictors and if anything is the least significant. Computer sessions had the most varied relationship with visitation in 2019 it was clear that it was positively associated with visitation numbers however moving towards 2021 this was not the case as visitation dropped and so did computer sessions leading to the relationship less positive as a result. Lastly compared to what it is in 2024 it’s clear that computer sessions do have a positive relationship and impact on visitation but have not recovered to what they were in 2019 but have recovered and become more significant since 2021.

## Multiple Linear Regression Model

Similarly to the simple linear regression multiple linear regression models are applied to the data amongst all three years to take into account their collective relationships in influencing visitation. However compared to the simple linear regressions there is a much more obvious shift in variability in the between the data set and its regression models. The high R-squared values for all three years explain a substantial portion of the variance in library visitation while the rest can be associated to the inconsistency amongst each of the predictor variables data entries. The results are very much the same as the simple linear regression models but it is much clearer that the holds amongst all three years have the least influence on visitation as opposed to circulation and computer sessions. This also reiterates the fact that both circulation and computer sessions have a larger impact on visitation.

## Polynomial Regression Model

Lastly, polynomial regressions were used in order to take into account that there may be nonlinear relationships present amongst the data set. This was partly done due to the fact that the data set is looking at data from 12 months out of the years of 2019, 2021, and 2024. Using the polynomial regression allows to observe any kinds of nonlinear trends amongst these time windows as there’s the potential of them either positively and negatively affecting visitation due to library closures and specifically in the case of this project the covid-19 pandemic. For the regression squared terms were used after experimenting with different combinations of exponential values with the squared iteration being the most consistent amongst them all.

The higher R-squared values for the polynomial models compared to the linear models for each year suggest that incorporating non-linear terms improves the model’s ability to explain the variance in Visitation. The statistical significance of the linear and quadratic terms varies across predictors and years, indicating that the nature of the relationship may differ for each factor which could be associated with the pandemics effect on library services. The results of the polynomial regression were very similar to that of the multiple linear regressions albeit show more stable predictions if we look at the regression line on the plots. The main takeaway however is the fact that the relationships are shown to be slightly more consistent than the rest of the models.

## Model Comparisons

Due to the way that this data set is set up it was difficult to choose which regression methods to apply given that I’m looking at 3 years worth of data and working with a limited amount of predictor variables that influence visitation. While there are a lot of options in terms of the methods that we can apply the most suitable the linear regressions along with the polynomial regressions because of all the data entries being numerical and with them being somewhat of a small data set these regression methods were simple enough to apply to them. Of all the three the best performing one in my opinion had to be the polynomial regression because it was able to take into account not just the fact that there’s the potential for a nonlinear relationship amongst each of the predictors and visitation but also the fact that it showed more stability when it came to the predictions created with the regression. This made the data much easier to interpret and analyze compared to the two linear regression types that were performed. The simple in your regression while it was useful to look at the relationship of visitation with each predictor it didn’t perform as well as I thought given the fact that there are relationships between each of the predictors and not just with the visitation itself. This can be seen in the graphs that were plotted as strictly positive relationships and this isn’t realistic given the way that Trends in library use are. Similarly to the polynomial the multiple regression method was effective in considering the changes in the relationships amongst the visitation and the predictors showed accurate predictions when compared to the original data set with enough information being provided to explain the variability. However the main issue I saw when performing the multiple linear regression is the fact that there is a lot more variability compared to that of the polynomial making it more difficult to analyze. Overall, while all three of the methods I applied were useful in analyzing the data, the best performing and most suitable choice for the data set was the polynomial regression especially because it fit the data the best amongst the three.

## Appendix

### I. Data Read

data\_dir <- "LibraryData/"  
  
load\_year\_data <- function(year)   
{  
 visitors\_file <- paste0(data\_dir, year, "Visitors.csv")  
 circ\_file <- paste0(data\_dir, year, "Circ.csv")  
 holds\_file <- paste0(data\_dir, year, "Holds.csv")  
 computer\_file <- paste0(data\_dir, year, "Computer.csv")  
  
   
 visitors <- read\_csv(visitors\_file)  
 circ <- read\_csv(circ\_file)  
 holds <- read\_csv(holds\_file)  
 computer <- read\_csv(computer\_file)  
  
 return(list(visitors = visitors, circ = circ, holds = holds, computer = computer))  
}  
  
data\_2019 <- load\_year\_data(2019)

## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (12): JANUARY, FEBRUARY, MARCH, APRIL, MAY, JUNE, JULY, AUGUST, SEPTEMBE...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (12): JANUARY, FEBRUARY, MARCH, APRIL, MAY, JUNE, JULY, AUGUST, SEPTEMBE...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (11): JANUARY, MARCH, APRIL, MAY, JUNE, JULY, AUGUST, SEPTEMBER, OCTOBER...  
## num (1): FEBRUARY  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (12): JANUARY, FEBRUARY, MARCH, APRIL, MAY, JUNE, JULY, AUGUST, SEPTEMBE...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

data\_2021 <- load\_year\_data(2021)

## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (12): JANUARY, FEBRUARY, MARCH, APRIL, MAY, JUNE, JULY, AUGUST, SEPTEMBE...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (12): JANUARY, FEBRUARY, MARCH, APRIL, MAY, JUNE, JULY, AUGUST, SEPTEMBE...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (8): JANUARY, FEBRUARY, MARCH, JUNE, JULY, SEPTEMBER, OCTOBER, NOVEMBER  
## num (4): APRIL, MAY, AUGUST, DECEMBER  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (12): JANUARY, FEBRUARY, MARCH, APRIL, MAY, JUNE, JULY, AUGUST, SEPTEMBE...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

data\_2024 <- load\_year\_data(2024)

## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (12): JANUARY, FEBRUARY, MARCH, APRIL, MAY, JUNE, JULY, AUGUST, SEPTEMBE...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (12): JANUARY, FEBRUARY, MARCH, APRIL, MAY, JUNE, JULY, AUGUST, SEPTEMBE...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (12): JANUARY, FEBRUARY, MARCH, APRIL, MAY, JUNE, JULY, AUGUST, SEPTEMBE...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.  
## Rows: 80 Columns: 13  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (1): LOCATION  
## dbl (12): JANUARY, FEBRUARY, MARCH, APRIL, MAY, JUNE, JULY, AUGUST, SEPTEMBE...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

### II. Data Preparation

#### II.I. Data Processing

process\_year\_data <- function(year, data\_list)  
{  
 visitors <- data\_list$visitors  
 circ <- data\_list$circ  
 holds <- data\_list$holds  
 holds <- holds %>% mutate(across(c(-LOCATION), as.numeric))  
 computer <- data\_list$computer  
   
 visitors\_long <- visitors %>% pivot\_longer(cols = -LOCATION, names\_to = "Month", values\_to = "Visitation")  
 circ\_long <- circ %>% pivot\_longer(cols = -LOCATION, names\_to = "Month", values\_to = "Circulation")  
 holds\_long <- holds %>% pivot\_longer(cols = -LOCATION, names\_to = "Month", values\_to = "Holds")  
 computer\_long <- computer %>% pivot\_longer(cols = -LOCATION, names\_to = "Month", values\_to = "Sessions")  
   
 month\_order <- c("JANUARY", "FEBRUARY", "MARCH", "APRIL", "MAY", "JUNE", "JULY", "AUGUST", "SEPTEMBER", "OCTOBER", "NOVEMBER", "DECEMBER")  
   
 visitors\_long$Month <- factor(visitors\_long$Month, levels = month\_order)  
 circ\_long$Month <- factor(circ\_long$Month, levels = month\_order)  
 holds\_long$Month <- factor(holds\_long$Month, levels = month\_order)  
 computer\_long$Month <- factor(computer\_long$Month, levels = month\_order)  
   
 visitors\_long$LOCATION <- str\_trim(visitors\_long$LOCATION)  
 circ\_long$LOCATION <- str\_trim(circ\_long$LOCATION)  
 holds\_long$LOCATION <- str\_trim(holds\_long$LOCATION)  
 computer\_long$LOCATION <- str\_trim(computer\_long$LOCATION)  
   
 merged\_year\_data <- visitors\_long %>%   
 left\_join(circ\_long, by = c("LOCATION", "Month")) %>%  
 left\_join(holds\_long, by = c("LOCATION", "Month")) %>%  
 left\_join(computer\_long, by = c("LOCATION", "Month"))  
  
 merged\_year\_data <- merged\_year\_data %>%  
 mutate(Year = year) %>%  
 select(LOCATION, Visitation, Circulation, Holds, Sessions, Month, Year)  
  
 return(merged\_year\_data)  
}  
  
processed\_2019 <- process\_year\_data(2019, data\_2019)  
processed\_2021 <- process\_year\_data(2021, data\_2021)  
processed\_2024 <- process\_year\_data(2024, data\_2024)

#### II.II. Data Aggregation

aggregate\_year\_data <- function(year, data\_list)   
{  
 visitors <- data\_list$visitors  
 circ <- data\_list$circ  
 holds <- data\_list$holds  
 computer <- data\_list$computer  
   
 visitors\_yearly <- visitors %>%  
 mutate(across(c(-LOCATION), as.numeric)) %>%  
 mutate(LOCATION = str\_trim(LOCATION)) %>%  
 group\_by(LOCATION) %>%  
 summarise(Visitation = sum(across(where(is.numeric)), na.rm = TRUE))  
   
 circ\_yearly <- circ %>%  
 mutate(across(c(-LOCATION), as.numeric)) %>%  
 mutate(LOCATION = str\_trim(LOCATION)) %>%  
 group\_by(LOCATION) %>%  
 summarise(Circulation = sum(across(where(is.numeric)), na.rm = TRUE))  
   
 holds\_yearly <- holds %>%  
 mutate(across(c(-LOCATION), as.numeric)) %>%  
 mutate(LOCATION = str\_trim(LOCATION)) %>%  
 group\_by(LOCATION) %>%  
 summarise(Holds = sum(across(where(is.numeric)), na.rm = TRUE))  
   
 computer\_yearly <- computer %>%  
 mutate(across(c(-LOCATION), as.numeric)) %>%  
 mutate(LOCATION = str\_trim(LOCATION)) %>%  
 group\_by(LOCATION) %>%  
 summarise(Sessions = sum(across(where(is.numeric)), na.rm = TRUE))  
   
 merged\_year\_data <- visitors\_yearly %>%  
 left\_join(circ\_yearly, by = "LOCATION") %>%  
 left\_join(holds\_yearly, by = "LOCATION") %>%  
 left\_join(computer\_yearly, by = "LOCATION")  
   
 merged\_year\_data <- merged\_year\_data %>%  
 mutate(Year = year) %>%  
 select(LOCATION, Visitation, Circulation, Holds, Sessions, Year)  
   
 return(merged\_year\_data)  
}  
  
aggregated\_2019 <- aggregate\_year\_data(2019, data\_2019)  
aggregated\_2021 <- aggregate\_year\_data(2021, data\_2021)  
aggregated\_2024 <- aggregate\_year\_data(2024, data\_2024)

### III. Descriptive analysis

#### III.I. Processed/Aggregated Data Table

{{R}} combined\_data <- bind\_rows(aggregated\_2019, aggregated\_2021, aggregated\_2024) kable(combined\_data)

#### III.II. Data Summary

summary(aggregated\_2019 %>% select(-LOCATION, -Year))

## Visitation Circulation Holds Sessions   
## Min. : 0 Min. : 1250 Min. : 28.0 Min. : 0   
## 1st Qu.: 50799 1st Qu.: 17074 1st Qu.: 509.5 1st Qu.: 10465   
## Median : 74725 Median : 37880 Median : 892.0 Median : 14292   
## Mean : 100495 Mean : 59089 Mean :1216.2 Mean : 19430   
## 3rd Qu.: 104360 3rd Qu.: 67621 3rd Qu.:1539.5 3rd Qu.: 21769   
## Max. :1405964 Max. :581151 Max. :4807.0 Max. :209583   
## NA's :1

summary(aggregated\_2021 %>% select(-LOCATION, -Year))

## Visitation Circulation Holds Sessions   
## Min. : 0 Min. : 0 Min. : 0.0 Min. : 0   
## 1st Qu.: 16756 1st Qu.: 7287 1st Qu.: 285.0 1st Qu.: 4898   
## Median : 33300 Median : 20184 Median : 600.0 Median : 6663   
## Mean : 44552 Mean : 37138 Mean : 855.9 Mean : 7385   
## 3rd Qu.: 53595 3rd Qu.: 55453 3rd Qu.:1209.0 3rd Qu.: 8963   
## Max. :543951 Max. :281183 Max. :3476.0 Max. :28566   
## NA's :3 NA's :3 NA's :3

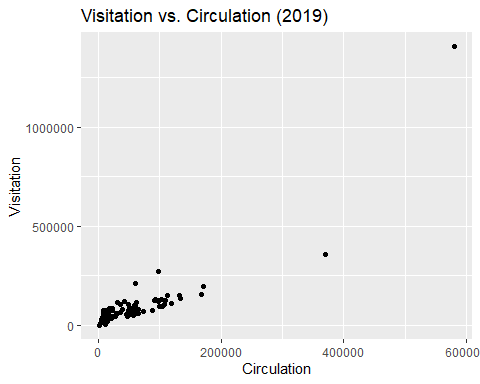
summary(aggregated\_2024 %>% select(-LOCATION, -Year))

## Visitation Circulation Holds Sessions   
## Min. : 0 Min. : 0 Min. : 0.0 Min. : 0   
## 1st Qu.: 37640 1st Qu.: 8863 1st Qu.: 366.0 1st Qu.: 7408   
## Median : 49725 Median : 20606 Median : 849.5 Median :10630   
## Mean : 69333 Mean : 38881 Mean : 1234.5 Mean :11432   
## 3rd Qu.: 76192 3rd Qu.: 56139 3rd Qu.: 1612.8 3rd Qu.:14220   
## Max. :653728 Max. :311851 Max. :15230.0 Max. :42065   
## NA's :3 NA's :3

#### III.III. Data Visualization

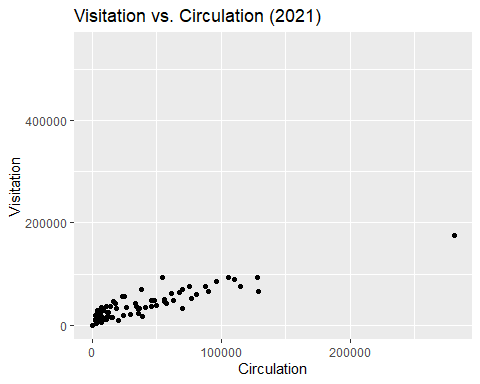
##### III.III.I. Scatter Plots

options(scipen = 999)  
  
ggplot(aggregated\_2019, aes(x= Circulation, y = Visitation)) + geom\_point() + labs(title = "Visitation vs. Circulation (2019)", x = "Circulation", y = "Visitation")



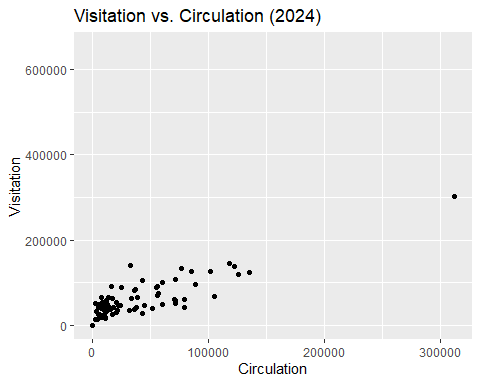
ggplot(aggregated\_2021, aes(x= Circulation, y = Visitation)) + geom\_point() + labs(title = "Visitation vs. Circulation (2021)", x = "Circulation", y = "Visitation")

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



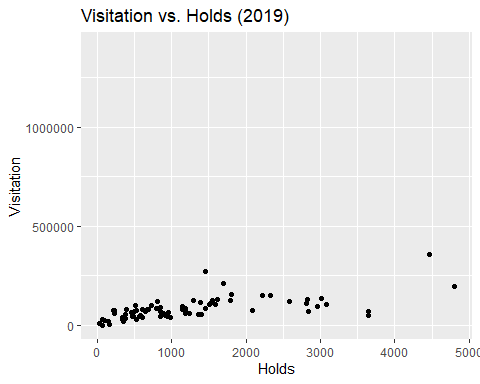
ggplot(aggregated\_2024, aes(x= Circulation, y = Visitation)) + geom\_point() + labs(title = "Visitation vs. Circulation (2024)", x = "Circulation", y = "Visitation")

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



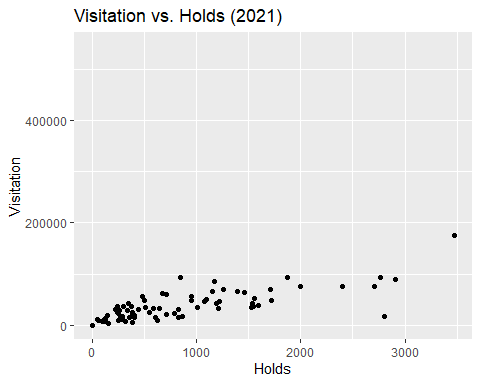
ggplot(aggregated\_2019, aes(x= Holds, y = Visitation)) + geom\_point() + labs(title = "Visitation vs. Holds (2019)", x = "Holds", y = "Visitation")

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_point()`).

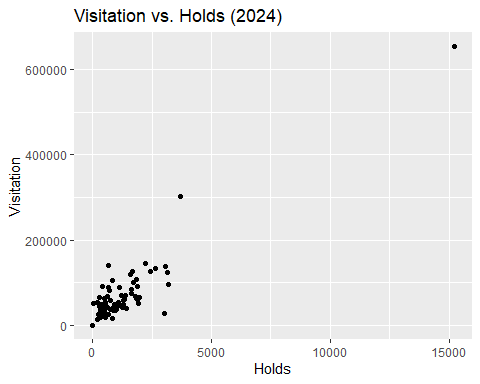


ggplot(aggregated\_2021, aes(x= Holds, y = Visitation)) + geom\_point() + labs(title = "Visitation vs. Holds (2021)", x = "Holds", y = "Visitation")

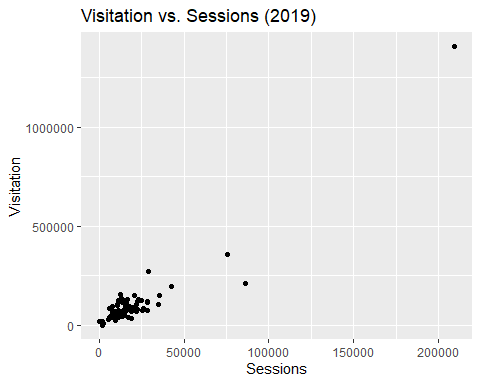
## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



ggplot(aggregated\_2024, aes(x= Holds, y = Visitation)) + geom\_point() + labs(title = "Visitation vs. Holds (2024)", x = "Holds", y = "Visitation")

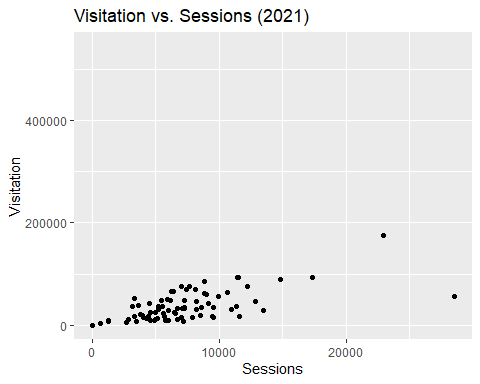


ggplot(aggregated\_2019, aes(x= Sessions, y = Visitation)) + geom\_point() + labs(title = "Visitation vs. Sessions (2019)", x = "Sessions", y = "Visitation")



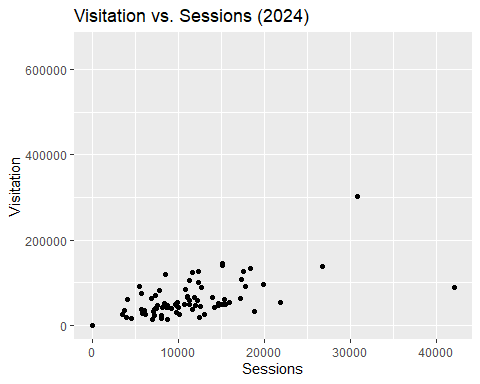
ggplot(aggregated\_2021, aes(x= Sessions, y = Visitation)) + geom\_point() + labs(title = "Visitation vs. Sessions (2021)", x = "Sessions", y = "Visitation")

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



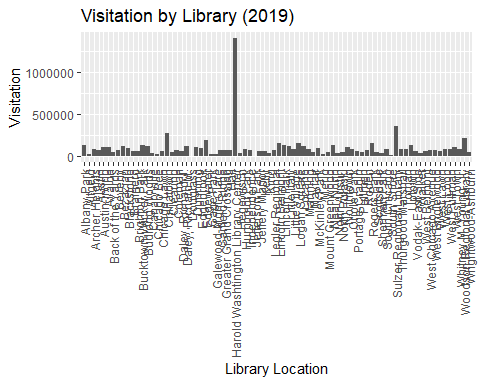
ggplot(aggregated\_2024, aes(x= Sessions, y = Visitation)) + geom\_point() + labs(title = "Visitation vs. Sessions (2024)", x = "Sessions", y = "Visitation")

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).

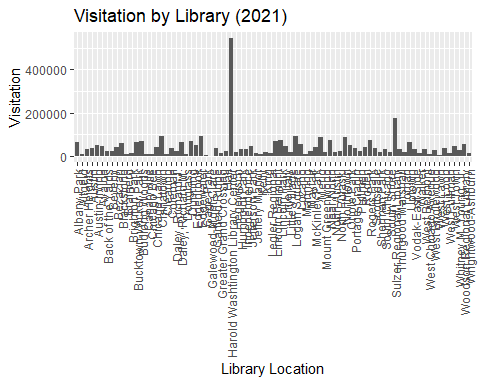


##### III.III.II. Bar Graphs

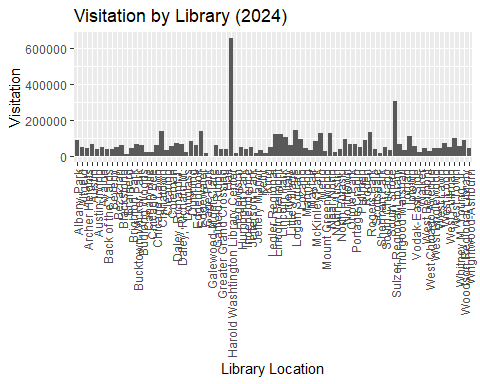
ggplot(aggregated\_2019, aes(x = LOCATION, y = Visitation)) + geom\_bar(stat = "identity") + labs(title = "Visitation by Library (2019)", x = "Library Location", y = "Visitation") +   
theme(axis.text.x = element\_text(angle = 90, hjust = 1, vjust = 0.5))



ggplot(aggregated\_2021, aes(x = LOCATION, y = Visitation)) + geom\_bar(stat = "identity") + labs(title = "Visitation by Library (2021)", x = "Library Location", y = "Visitation") +   
theme(axis.text.x = element\_text(angle = 90, hjust = 1, vjust = 0.5))



ggplot(aggregated\_2024, aes(x = LOCATION, y = Visitation)) + geom\_bar(stat = "identity") + labs(title = "Visitation by Library (2024)", x = "Library Location", y = "Visitation") +   
theme(axis.text.x = element\_text(angle = 90, hjust = 1, vjust = 0.5))



### IV. Regression Methods Applied

#### IV.I. Simple Linear Regression

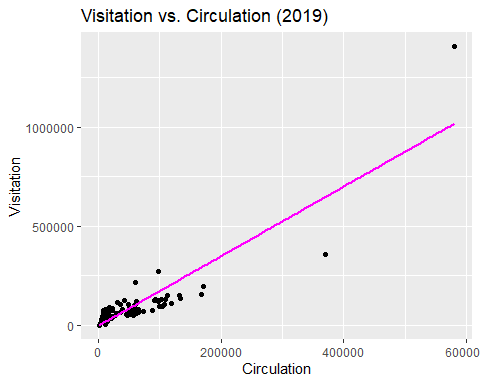
sLm2019Circ <- lm(Visitation ~ Circulation, data = aggregated\_2019)  
sLm2019Holds <- lm(Visitation ~ Holds, data = aggregated\_2019)  
sLm2019Sess <- lm(Visitation ~ Sessions, data = aggregated\_2019)  
  
sLm2021Circ <- lm(Visitation ~ Circulation, data = aggregated\_2021)  
sLm2021Holds <- lm(Visitation ~ Holds, data = aggregated\_2021)  
sLm2021Sess <- lm(Visitation ~ Sessions, data = aggregated\_2021)  
  
sLm2024Circ <- lm(Visitation ~ Circulation, data = aggregated\_2024)  
sLm2024Holds <- lm(Visitation ~ Holds, data = aggregated\_2024)  
sLm2024Sess <- lm(Visitation ~ Sessions, data = aggregated\_2024)

##### IV.I.I. 2019

options(scipen = 999)  
  
summary(sLm2019Circ)

##   
## Call:  
## lm(formula = Visitation ~ Circulation, data = aggregated\_2019)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -287085 -32546 5577 38380 388559   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -3285.3969 10170.7970 -0.323 0.748   
## Circulation 1.7563 0.1028 17.081 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 72960 on 78 degrees of freedom  
## Multiple R-squared: 0.7891, Adjusted R-squared: 0.7864   
## F-statistic: 291.8 on 1 and 78 DF, p-value: < 0.00000000000000022

ggplot(aggregated\_2019, aes(x = Circulation, y = Visitation)) + geom\_point() + geom\_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "magenta") + labs(title = "Visitation vs. Circulation (2019)", x = "Circulation", y = "Visitation")



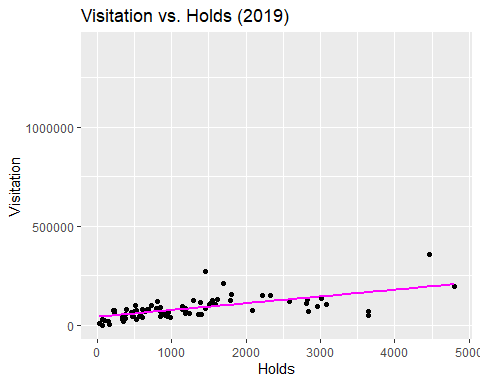
summary(sLm2019Holds)

##   
## Call:  
## lm(formula = Visitation ~ Holds, data = aggregated\_2019)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -115603 -23064 -6431 17409 179526   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 42436.324 7569.801 5.606 0.000000309516 \*\*\*  
## Holds 34.150 4.756 7.181 0.000000000377 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 43410 on 77 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.4011, Adjusted R-squared: 0.3933   
## F-statistic: 51.57 on 1 and 77 DF, p-value: 0.000000000377

ggplot(aggregated\_2019, aes(x = Holds, y = Visitation)) + geom\_point() + geom\_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "magenta") + labs(title = "Visitation vs. Holds (2019)", x = "Holds", y = "Visitation")

## Warning: Removed 1 row containing non-finite outside the scale range  
## (`stat\_smooth()`).

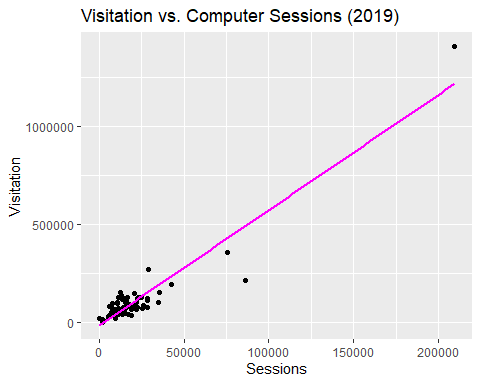
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_point()`).



summary(sLm2019Sess)

##   
## Call:  
## lm(formula = Visitation ~ Sessions, data = aggregated\_2019)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -277793 -25312 1448 22043 189819   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -13504.3430 7746.8944 -1.743 0.0852 .   
## Sessions 5.8671 0.2441 24.035 <0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 54790 on 78 degrees of freedom  
## Multiple R-squared: 0.881, Adjusted R-squared: 0.8795   
## F-statistic: 577.7 on 1 and 78 DF, p-value: < 0.00000000000000022

ggplot(aggregated\_2019, aes(x = Sessions, y = Visitation)) + geom\_point() + geom\_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "magenta") + labs(title = "Visitation vs. Computer Sessions (2019)", x = "Sessions", y = "Visitation")



##### IV.I.II. 2021

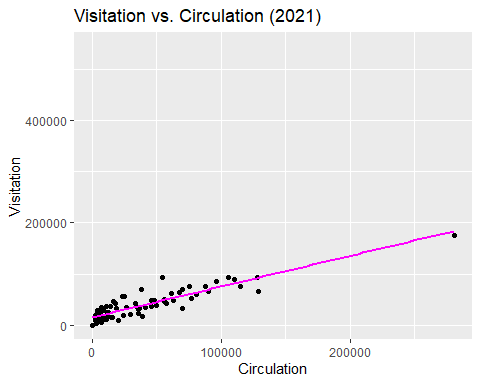
options(scipen = 999)  
  
summary(sLm2021Circ)

##   
## Call:  
## lm(formula = Visitation ~ Circulation, data = aggregated\_2021)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -25970 -8605 -3176 7713 45393   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 15872.93425 1905.69037 8.329 0.00000000000282 \*\*\*  
## Circulation 0.59825 0.03328 17.977 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 12730 on 75 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.8116, Adjusted R-squared: 0.8091   
## F-statistic: 323.2 on 1 and 75 DF, p-value: < 0.00000000000000022

ggplot(aggregated\_2021, aes(x = Circulation, y = Visitation)) + geom\_point() + geom\_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "magenta") + labs(title = "Visitation vs. Circulation (2021)", x = "Circulation", y = "Visitation")

## Warning: Removed 3 rows containing non-finite outside the scale range  
## (`stat\_smooth()`).

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).

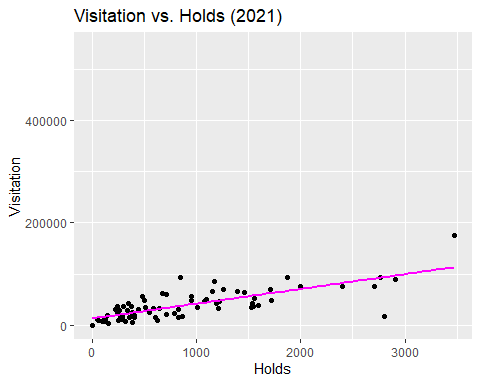


summary(sLm2021Holds)

##   
## Call:  
## lm(formula = Visitation ~ Holds, data = aggregated\_2021)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -75546 -9846 -4150 8077 62009   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 13501.987 3214.939 4.20 0.000072768215060234 \*\*\*  
## Holds 28.730 2.791 10.29 0.000000000000000538 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 18880 on 75 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.5855, Adjusted R-squared: 0.58   
## F-statistic: 106 on 1 and 75 DF, p-value: 0.000000000000000538

ggplot(aggregated\_2021, aes(x = Holds, y = Visitation)) + geom\_point() + geom\_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "magenta") + labs(title = "Visitation vs. Holds (2021)", x = "Holds", y = "Visitation")

## Warning: Removed 3 rows containing non-finite outside the scale range (`stat\_smooth()`).  
## Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).

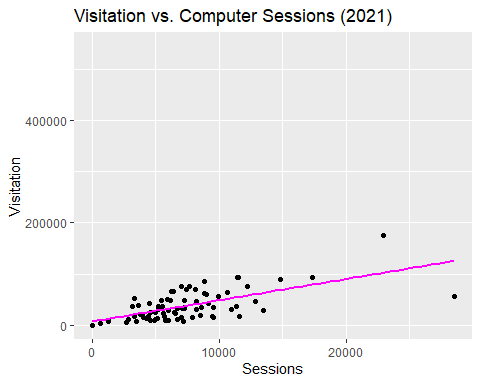


summary(sLm2021Sess)

##   
## Call:  
## lm(formula = Visitation ~ Sessions, data = aggregated\_2021)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -69698 -14273 -3451 15695 72966   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 7551.2721 5047.4092 1.496 0.139   
## Sessions 4.1355 0.5865 7.051 0.000000000745 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 22740 on 75 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.3986, Adjusted R-squared: 0.3906   
## F-statistic: 49.72 on 1 and 75 DF, p-value: 0.0000000007445

ggplot(aggregated\_2021, aes(x = Sessions, y = Visitation)) + geom\_point() + geom\_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "magenta") + labs(title = "Visitation vs. Computer Sessions (2021)", x = "Sessions", y = "Visitation")

## Warning: Removed 3 rows containing non-finite outside the scale range (`stat\_smooth()`).  
## Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



##### IV.I.III. 2024

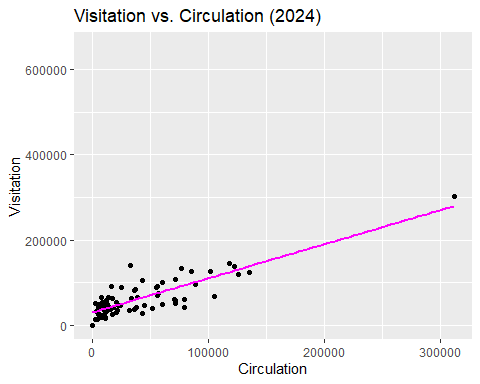
options(scipen = 999)  
  
summary(sLm2024Circ)

##   
## Call:  
## lm(formula = Visitation ~ Circulation, data = aggregated\_2024)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -52122 -17749 -183 15198 82884   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 30672.75484 3495.60742 8.775 0.0000000000004 \*\*\*  
## Circulation 0.79881 0.05794 13.787 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 23460 on 75 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.7171, Adjusted R-squared: 0.7133   
## F-statistic: 190.1 on 1 and 75 DF, p-value: < 0.00000000000000022

ggplot(aggregated\_2024, aes(x = Circulation, y = Visitation)) + geom\_point() + geom\_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "magenta") + labs(title = "Visitation vs. Circulation (2024)", x = "Circulation", y = "Visitation")

## Warning: Removed 3 rows containing non-finite outside the scale range  
## (`stat\_smooth()`).

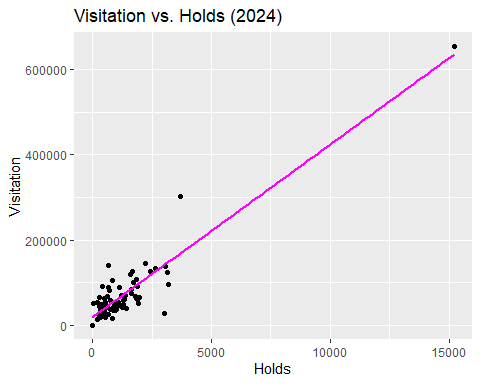
## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



summary(sLm2024Holds)

##   
## Call:  
## lm(formula = Visitation ~ Holds, data = aggregated\_2024)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -114055 -17904 -5670 14242 133678   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 19393.560 4341.152 4.467 0.0000264 \*\*\*  
## Holds 40.453 2.008 20.149 < 0.0000000000000002 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 31880 on 78 degrees of freedom  
## Multiple R-squared: 0.8388, Adjusted R-squared: 0.8368   
## F-statistic: 406 on 1 and 78 DF, p-value: < 0.00000000000000022

ggplot(aggregated\_2024, aes(x = Holds, y = Visitation)) + geom\_point() + geom\_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "magenta") + labs(title = "Visitation vs. Holds (2024)", x = "Holds", y = "Visitation")

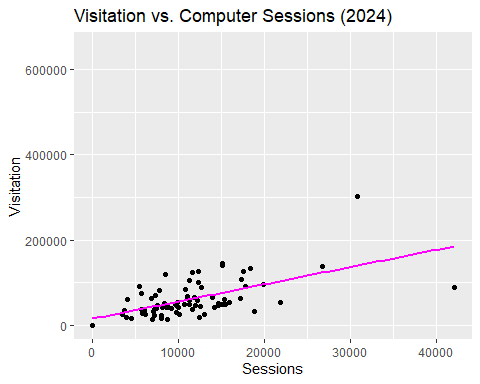


summary(sLm2024Sess)

##   
## Call:  
## lm(formula = Visitation ~ Sessions, data = aggregated\_2024)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -97200 -21011 -5986 15267 162573   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 15544.5158 8574.4340 1.813 0.0738 .   
## Sessions 4.0401 0.6587 6.133 0.0000000371 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 35990 on 75 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.334, Adjusted R-squared: 0.3252   
## F-statistic: 37.62 on 1 and 75 DF, p-value: 0.00000003713

ggplot(aggregated\_2024, aes(x = Sessions, y = Visitation)) + geom\_point() + geom\_smooth(method = "lm", formula = y ~ x, se = FALSE, color = "magenta") + labs(title = "Visitation vs. Computer Sessions (2024)", x = "Sessions", y = "Visitation")

## Warning: Removed 3 rows containing non-finite outside the scale range (`stat\_smooth()`).  
## Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).



#### IV.II. Multiple Linear Regression

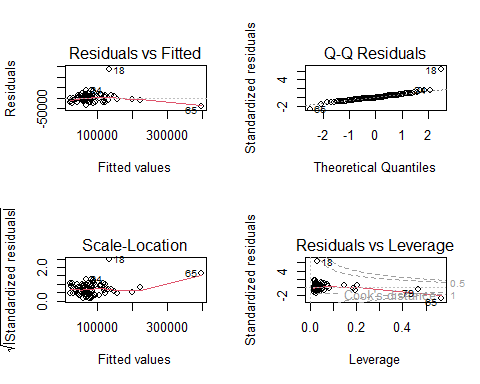
mLm2019 <- lm(Visitation ~ Circulation + Holds + Sessions, data = aggregated\_2019)  
mLm2021 <- lm(Visitation ~ Circulation + Holds + Sessions, data = aggregated\_2021)  
mLm2024 <- lm(Visitation ~ Circulation + Holds + Sessions, data = aggregated\_2024)

##### IV.II.I. 2019

summary(mLm2019)

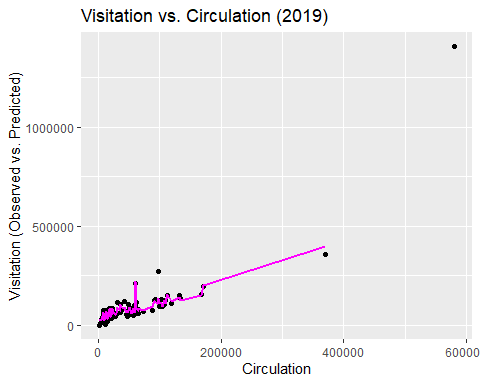
##   
## Call:  
## lm(formula = Visitation ~ Circulation + Holds + Sessions, data = aggregated\_2019)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -39005 -10563 -3648 8522 135512   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 19439.29378 4463.93455 4.355 0.0000416123095094 \*\*\*  
## Circulation 0.65147 0.07889 8.258 0.0000000000038549 \*\*\*  
## Holds -2.66614 3.89539 -0.684 0.496   
## Sessions 1.97279 0.21638 9.117 0.0000000000000892 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 22000 on 75 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.8501, Adjusted R-squared: 0.8441   
## F-statistic: 141.8 on 3 and 75 DF, p-value: < 0.00000000000000022

par(mfrow = c(2, 2))  
plot(mLm2019)



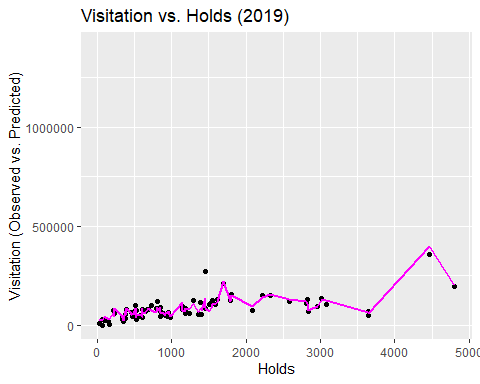
par(mfrow = c(1, 1))  
  
options(scipen = 999)  
  
predicted2019 <- predict(mLm2019, newdata = aggregated\_2019)  
  
plotted2019Circ <- data.frame(Circulation = aggregated\_2019$Circulation, Predicted\_Visitation = predicted2019)  
plotted2019Circ <- plotted2019Circ %>% arrange(Circulation)  
  
ggplot(aggregated\_2019, aes(x = Circulation, y = Visitation)) + geom\_point() + geom\_line(data = plotted2019Circ, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Circulation (2019)", x = "Circulation", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_line()`).



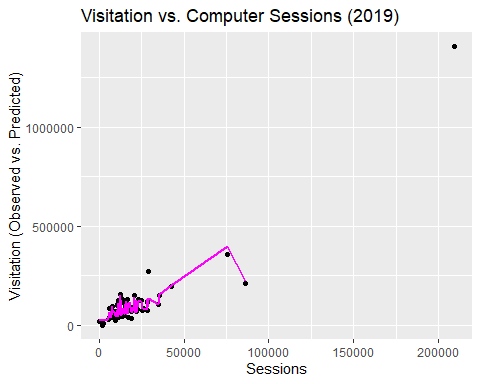
plotted2019Holds <- data.frame(Holds = aggregated\_2019$Holds, Predicted\_Visitation = predicted2019)  
plotted2019Holds <- plotted2019Holds %>% arrange(Holds)  
  
ggplot(aggregated\_2019, aes(x = Holds, y = Visitation)) + geom\_point() + geom\_line(data = plotted2019Holds, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Holds (2019)", x = "Holds", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_point()`).  
## Removed 1 row containing missing values or values outside the scale range  
## (`geom\_line()`).



plotted2019Sess <- data.frame(Sessions = aggregated\_2019$Sessions, Predicted\_Visitation = predicted2019)  
plotted2019Sess <- plotted2019Sess %>% arrange(Sessions)  
  
ggplot(aggregated\_2019, aes(x = Sessions, y = Visitation)) + geom\_point() + geom\_line(data = plotted2019Sess, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Computer Sessions (2019)", x = "Sessions", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_line()`).

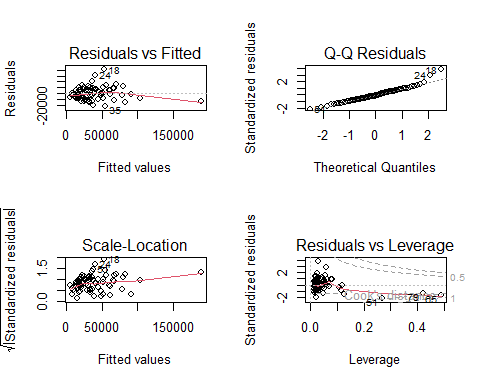


##### IV.II.II. 2021

summary(mLm2021)

##   
## Call:  
## lm(formula = Visitation ~ Circulation + Holds + Sessions, data = aggregated\_2021)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -21013 -7689 -1058 5978 40375   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5637.32338 2498.98885 2.256 0.0271 \*   
## Circulation 0.48634 0.04942 9.842 0.00000000000000501 \*\*\*  
## Holds 2.11269 2.75021 0.768 0.4449   
## Sessions 1.70399 0.31946 5.334 0.00000103255419934 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10840 on 73 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.8671, Adjusted R-squared: 0.8617   
## F-statistic: 158.8 on 3 and 73 DF, p-value: < 0.00000000000000022

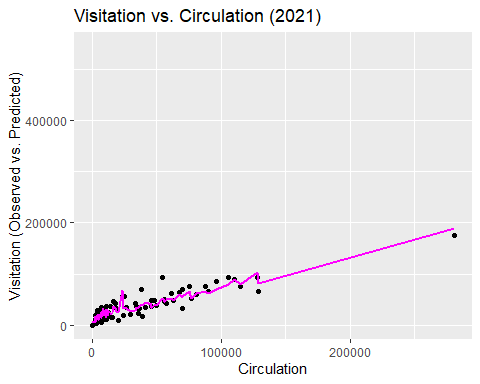
par(mfrow = c(2, 2))  
plot(mLm2021)



par(mfrow = c(1, 1))  
  
options(scipen = 999)  
  
predicted2021 <- predict(mLm2021, newdata = aggregated\_2021)  
  
plotted2021Circ <- data.frame(Circulation = aggregated\_2021$Circulation, Predicted\_Visitation = predicted2021)  
plotted2021Circ <- plotted2021Circ %>% arrange(Circulation)  
  
ggplot(aggregated\_2021, aes(x = Circulation, y = Visitation)) + geom\_point() + geom\_line(data = plotted2021Circ, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Circulation (2021)", x = "Circulation", y = "Visitation (Observed vs. Predicted)")

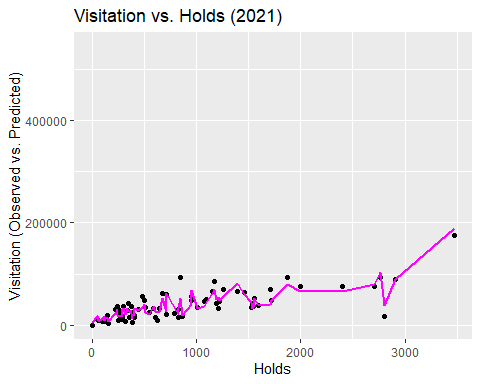
## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_line()`).



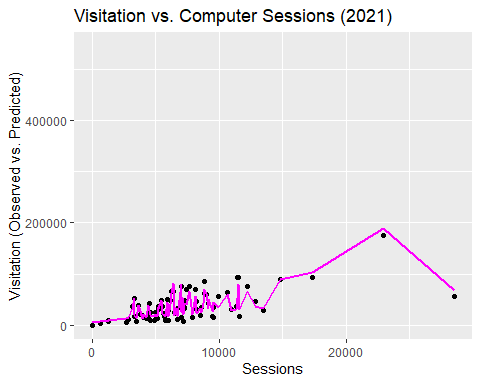
plotted2021Holds <- data.frame(Holds = aggregated\_2021$Holds, Predicted\_Visitation = predicted2021)  
plotted2021Holds <- plotted2021Holds %>% arrange(Holds)  
  
ggplot(aggregated\_2021, aes(x = Holds, y = Visitation)) + geom\_point() + geom\_line(data = plotted2021Holds, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Holds (2021)", x = "Holds", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).  
## Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_line()`).



plotted2021Sess <- data.frame(Sessions = aggregated\_2021$Sessions, Predicted\_Visitation = predicted2021)  
plotted2021Sess <- plotted2021Sess %>% arrange(Sessions)  
  
ggplot(aggregated\_2021, aes(x = Sessions, y = Visitation)) + geom\_point() + geom\_line(data = plotted2021Sess, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Computer Sessions (2021)", x = "Sessions", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).  
## Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_line()`).

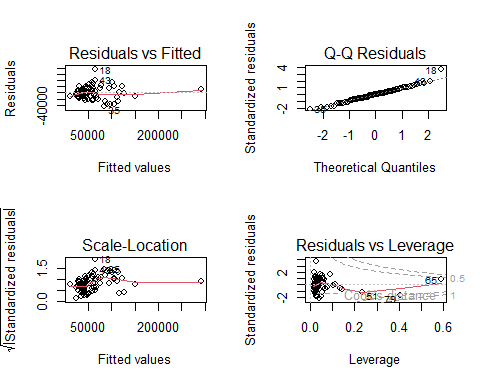


##### IV.II.III. 2024

summary(mLm2024)

##   
## Call:  
## lm(formula = Visitation ~ Circulation + Holds + Sessions, data = aggregated\_2024)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -42826 -12912 1117 10572 74371   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 11920.07846 5382.74039 2.214 0.0299 \*   
## Circulation 0.71036 0.08946 7.940 0.0000000000182 \*\*\*  
## Holds -1.29647 4.82986 -0.268 0.7891   
## Sessions 2.06153 0.40559 5.083 0.0000027682528 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20430 on 73 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.7911, Adjusted R-squared: 0.7825   
## F-statistic: 92.14 on 3 and 73 DF, p-value: < 0.00000000000000022

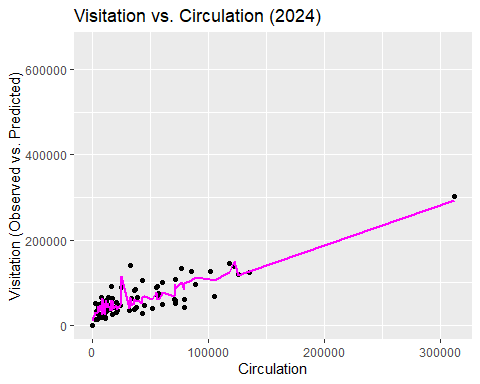
par(mfrow = c(2, 2))  
plot(mLm2024)



par(mfrow = c(1, 1))  
  
options(scipen = 999)  
  
predicted2024 <- predict(mLm2024, newdata = aggregated\_2024)  
  
plotted2024Circ <- data.frame(Circulation = aggregated\_2024$Circulation, Predicted\_Visitation = predicted2024)  
plotted2024Circ <- plotted2024Circ %>% arrange(Circulation)  
  
ggplot(aggregated\_2024, aes(x = Circulation, y = Visitation)) + geom\_point() + geom\_line(data = plotted2024Circ, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Circulation (2024)", x = "Circulation", y = "Visitation (Observed vs. Predicted)")

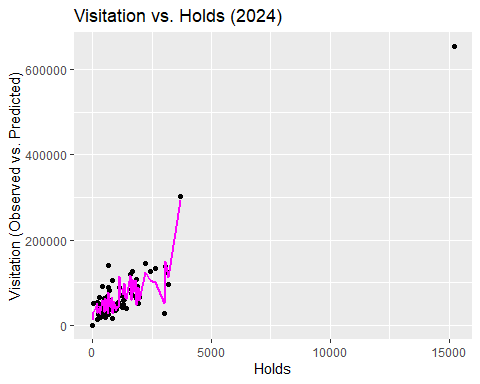
## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_line()`).



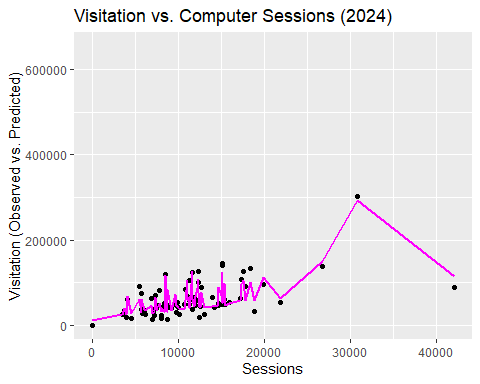
plotted2024Holds <- data.frame(Holds = aggregated\_2024$Holds, Predicted\_Visitation = predicted2024)  
plotted2024Holds <- plotted2024Holds %>% arrange(Holds)  
  
ggplot(aggregated\_2024, aes(x = Holds, y = Visitation)) + geom\_point() + geom\_line(data = plotted2024Holds, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Holds (2024)", x = "Holds", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_line()`).



plotted2024Sess <- data.frame(Sessions = aggregated\_2024$Sessions, Predicted\_Visitation = predicted2024)  
plotted2024Sess <- plotted2024Sess %>% arrange(Sessions)  
  
ggplot(aggregated\_2024, aes(x = Sessions, y = Visitation)) + geom\_point() + geom\_line(data = plotted2024Sess, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Computer Sessions (2024)", x = "Sessions", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).  
## Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_line()`).



#### IV.III. Polynomial Regression

pm2019 <- lm(Visitation ~ Circulation + I(Circulation^2) + Holds + I(Holds^2) + Sessions + I(Sessions^2), data = aggregated\_2019)  
pm2021 <- lm(Visitation ~ Circulation + I(Circulation^2) + Holds + I(Holds^2) + Sessions + I(Sessions^2), data = aggregated\_2021)  
pm2024 <- lm(Visitation ~ Circulation + I(Circulation^2) + Holds + I(Holds^2) + Sessions + I(Sessions^2), data = aggregated\_2024)

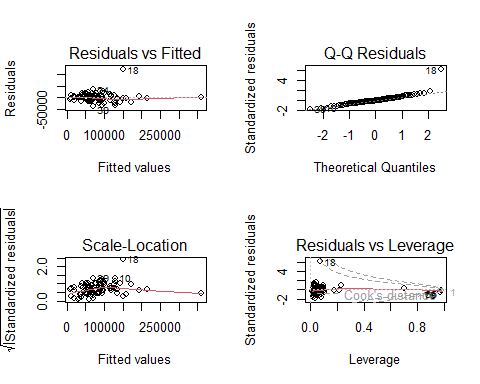
##### IV.III.I. 2019

summary(pm2019)

##   
## Call:  
## lm(formula = Visitation ~ Circulation + I(Circulation^2) + Holds +   
## I(Holds^2) + Sessions + I(Sessions^2), data = aggregated\_2019)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -35743 -11164 -1403 8002 120556   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2183.6897090187 6877.9529373172 0.317 0.752   
## Circulation 0.8476277934 0.1597167557 5.307 0.000001179 \*\*\*  
## I(Circulation^2) -0.0000005737 0.0000004704 -1.220 0.227   
## Holds 2.1155382437 10.6052532080 0.199 0.842   
## I(Holds^2) -0.0021275207 0.0022189943 -0.959 0.341   
## Sessions 2.8262942737 0.4814756755 5.870 0.000000123 \*\*\*  
## I(Sessions^2) -0.0000102636 0.0000062309 -1.647 0.104   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20690 on 72 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.8728, Adjusted R-squared: 0.8622   
## F-statistic: 82.31 on 6 and 72 DF, p-value: < 0.00000000000000022

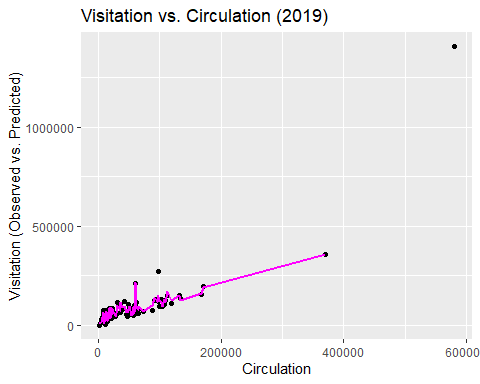
par(mfrow = c(2, 2))  
plot(pm2019)

## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced  
## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced



par(mfrow = c(1, 1))  
  
options(scipen = 999)  
  
predicted2019 <- predict(pm2019, newdata = aggregated\_2019)  
  
plotted2019Circ <- data.frame(Circulation = aggregated\_2019$Circulation, Predicted\_Visitation = predicted2019)  
plotted2019Circ <- plotted2019Circ %>% arrange(Circulation)  
  
ggplot(aggregated\_2019, aes(x = Circulation, y = Visitation)) + geom\_point() + geom\_line(data = plotted2019Circ, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Circulation (2019)", x = "Circulation", y = "Visitation (Observed vs. Predicted)")

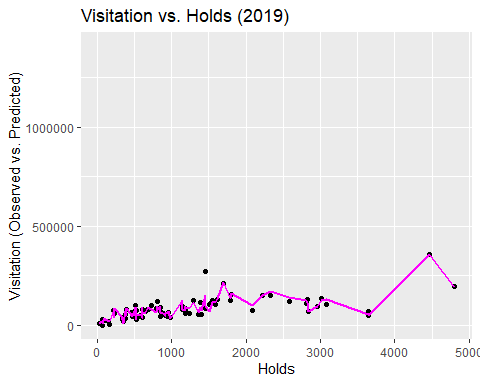
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_line()`).



plotted2019Holds <- data.frame(Holds = aggregated\_2019$Holds, Predicted\_Visitation = predicted2019)  
plotted2019Holds <- plotted2019Holds %>% arrange(Holds)  
  
ggplot(aggregated\_2019, aes(x = Holds, y = Visitation)) + geom\_point() + geom\_line(data = plotted2019Holds, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Holds (2019)", x = "Holds", y = "Visitation (Observed vs. Predicted)")

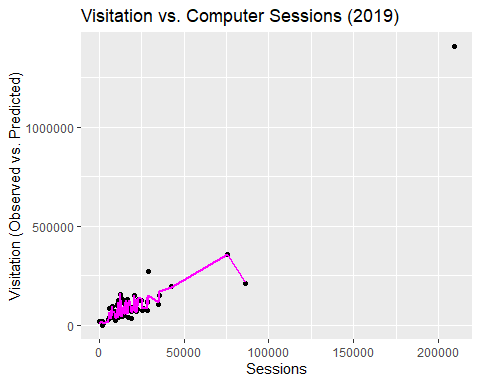
## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_point()`).

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_line()`).



plotted2019Sess <- data.frame(Sessions = aggregated\_2019$Sessions, Predicted\_Visitation = predicted2019)  
plotted2019Sess <- plotted2019Sess %>% arrange(Sessions)  
  
ggplot(aggregated\_2019, aes(x = Sessions, y = Visitation)) + geom\_point() + geom\_line(data = plotted2019Sess, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Computer Sessions (2019)", x = "Sessions", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_line()`).



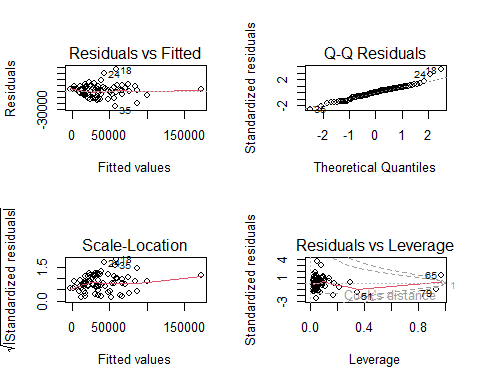
##### IV.III.II. 2021

summary(pm2021)

##   
## Call:  
## lm(formula = Visitation ~ Circulation + I(Circulation^2) + Holds +   
## I(Holds^2) + Sessions + I(Sessions^2), data = aggregated\_2021)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -25063 -5771 386 5940 35435   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -2804.2192310407 3855.1817217629 -0.727 0.469413   
## Circulation 0.5459048319 0.0952013961 5.734 0.000000229 \*\*\*  
## I(Circulation^2) -0.0000001733 0.0000004031 -0.430 0.668453   
## Holds 7.6387095264 7.1618196150 1.067 0.289823   
## I(Holds^2) -0.0026711179 0.0022925471 -1.165 0.247921   
## Sessions 2.9646502710 0.7709233205 3.846 0.000262 \*\*\*  
## I(Sessions^2) -0.0000499194 0.0000298696 -1.671 0.099138 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 10350 on 70 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.8838, Adjusted R-squared: 0.8738   
## F-statistic: 88.71 on 6 and 70 DF, p-value: < 0.00000000000000022

par(mfrow = c(2, 2))  
plot(pm2021)

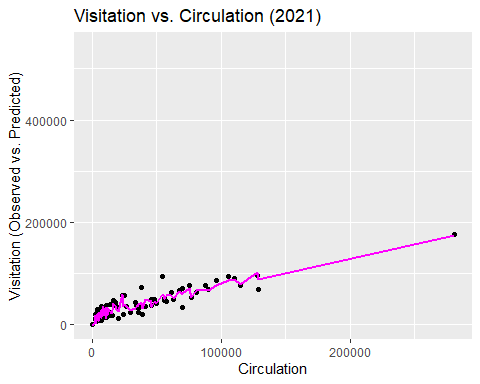
## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced  
## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced



par(mfrow = c(1, 1))  
  
options(scipen = 999)  
  
predicted2021 <- predict(pm2021, newdata = aggregated\_2021)  
  
plotted2021Circ <- data.frame(Circulation = aggregated\_2021$Circulation, Predicted\_Visitation = predicted2021)  
plotted2021Circ <- plotted2021Circ %>% arrange(Circulation)  
  
ggplot(aggregated\_2021, aes(x = Circulation, y = Visitation)) + geom\_point() + geom\_line(data = plotted2021Circ, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Circulation (2021)", x = "Circulation", y = "Visitation (Observed vs. Predicted)")

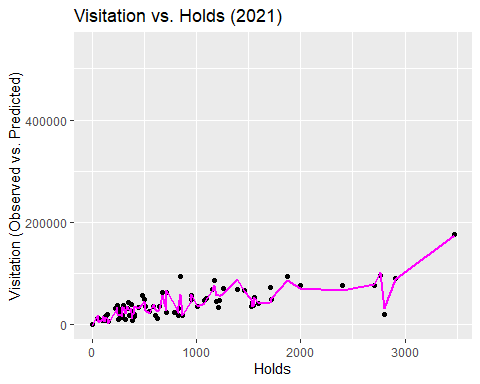
## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_line()`).



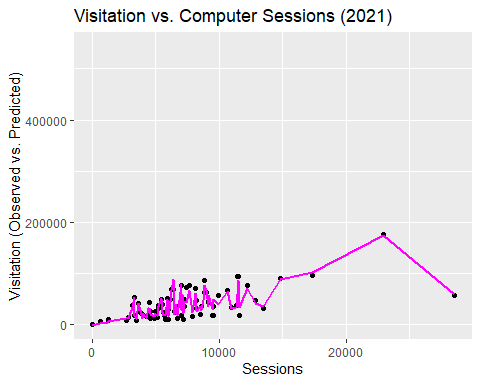
plotted2021Holds <- data.frame(Holds = aggregated\_2021$Holds, Predicted\_Visitation = predicted2021)  
plotted2021Holds <- plotted2021Holds %>% arrange(Holds)  
  
ggplot(aggregated\_2021, aes(x = Holds, y = Visitation)) + geom\_point() + geom\_line(data = plotted2021Holds, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Holds (2021)", x = "Holds", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).  
## Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_line()`).



plotted2021Sess <- data.frame(Sessions = aggregated\_2021$Sessions, Predicted\_Visitation = predicted2021)  
plotted2021Sess <- plotted2021Sess %>% arrange(Sessions)  
  
ggplot(aggregated\_2021, aes(x = Sessions, y = Visitation)) + geom\_point() + geom\_line(data = plotted2021Sess, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Computer Sessions (2021)", x = "Sessions", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).  
## Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_line()`).



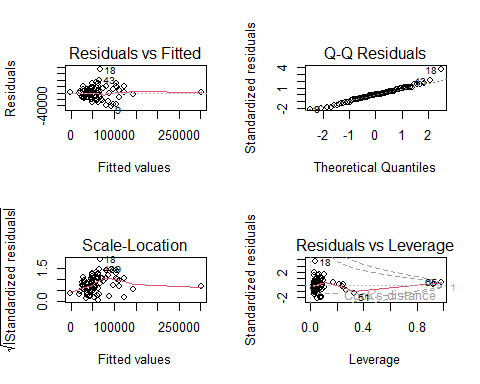
##### IV.III.III. 2024

summary(pm2024)

##   
## Call:  
## lm(formula = Visitation ~ Circulation + I(Circulation^2) + Holds +   
## I(Holds^2) + Sessions + I(Sessions^2), data = aggregated\_2024)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -40514 -12022 225 10476 72406   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) -1885.4795282580 10005.1961957638 -0.188 0.851069   
## Circulation 0.4906139843 0.1982417344 2.475 0.015755 \*   
## I(Circulation^2) 0.0000007862 0.0000006534 1.203 0.232941   
## Holds 9.3652833807 13.7567577533 0.681 0.498258   
## I(Holds^2) -0.0022653734 0.0037424277 -0.605 0.546923   
## Sessions 3.9475666562 1.0706579402 3.687 0.000444 \*\*\*  
## I(Sessions^2) -0.0000546476 0.0000280856 -1.946 0.055701 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 20240 on 70 degrees of freedom  
## (3 observations deleted due to missingness)  
## Multiple R-squared: 0.8033, Adjusted R-squared: 0.7864   
## F-statistic: 47.64 on 6 and 70 DF, p-value: < 0.00000000000000022

par(mfrow = c(2, 2))  
plot(pm2024)

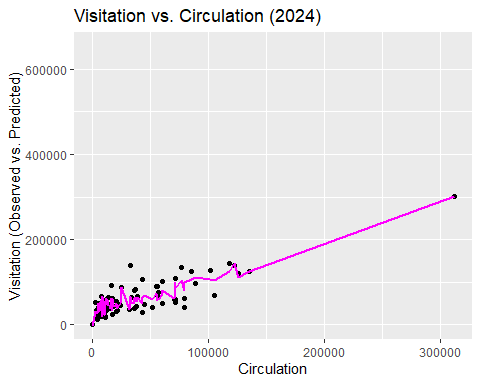
## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced  
## Warning in sqrt(crit \* p \* (1 - hh)/hh): NaNs produced



par(mfrow = c(1, 1))  
  
options(scipen = 999)  
  
predicted2024 <- predict(pm2024, newdata = aggregated\_2024)  
  
plotted2024Circ <- data.frame(Circulation = aggregated\_2024$Circulation, Predicted\_Visitation = predicted2024)  
plotted2024Circ <- plotted2024Circ %>% arrange(Circulation)  
  
ggplot(aggregated\_2024, aes(x = Circulation, y = Visitation)) + geom\_point() + geom\_line(data = plotted2024Circ, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Circulation (2024)", x = "Circulation", y = "Visitation (Observed vs. Predicted)")

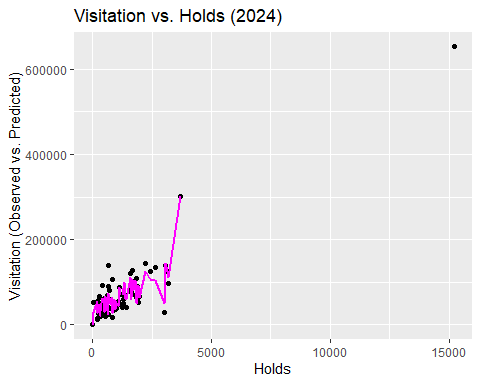
## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_line()`).



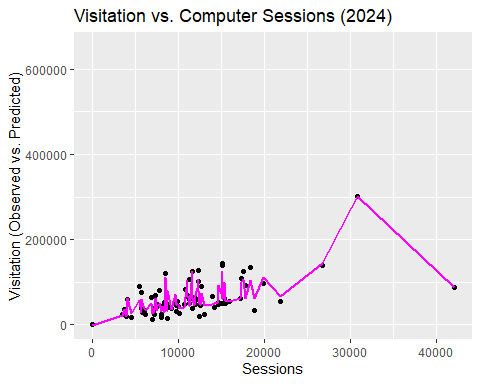
plotted2024Holds <- data.frame(Holds = aggregated\_2024$Holds, Predicted\_Visitation = predicted2024)  
plotted2024Holds <- plotted2024Holds %>% arrange(Holds)  
  
ggplot(aggregated\_2024, aes(x = Holds, y = Visitation)) + geom\_point() + geom\_line(data = plotted2024Holds, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Holds (2024)", x = "Holds", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 1 row containing missing values or values outside the scale range  
## (`geom\_line()`).



plotted2024Sess <- data.frame(Sessions = aggregated\_2024$Sessions, Predicted\_Visitation = predicted2024)  
plotted2024Sess <- plotted2024Sess %>% arrange(Sessions)  
  
ggplot(aggregated\_2024, aes(x = Sessions, y = Visitation)) + geom\_point() + geom\_line(data = plotted2024Sess, aes(y = Predicted\_Visitation), color = "magenta", linewidth = 1) + labs(title = "Visitation vs. Computer Sessions (2024)", x = "Sessions", y = "Visitation (Observed vs. Predicted)")

## Warning: Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_point()`).  
## Removed 3 rows containing missing values or values outside the scale range  
## (`geom\_line()`).



#### IV.IV. Cubic Spline

##### IV.IV.I. 2019

##### IV.IV.II. 2021

##### IV.IV.III. 2024

#### IV.V. Natural Spline

##### IV.V.I. 2019

##### IV.V.II. 2021

##### IV.V.III. 2024