

Final Project Fitting Models

Fall 2025

Define functions so we can remove outliers from training data.

```
# Function to remove outliers using the IQR rule
remove_outliers <- function(x) {
  if (is.numeric(x) && length(unique(na.omit(x))) > 2) { # skip binary
    q1 <- quantile(x, 0.25, na.rm = TRUE)
    q3 <- quantile(x, 0.75, na.rm = TRUE)
    iqr <- q3 - q1
    lower <- q1 - 1.5 * iqr
    upper <- q3 + 1.5 * iqr
    x[x < lower | x > upper] <- NA # mark outliers
  }
  return(x)
}

# Columns to ignore
ignore_cols <- c("Bookmobiles", "Branch Library", "state")

# Apply only to numeric, non-binary, and not ignored columns
libraries <- libraries %>%
  mutate(across(
    .cols = where(is.numeric) & !all_of(ignore_cols),
    .fns = remove_outliers
  )) %>%
  drop_na()

# Count occurrences per state
state_counts <- table(libraries$state)

# Find states with < 20 occurrences
rare_states <- names(state_counts[state_counts < 20])

# Combine them into "Other" (this is for CV purposes)
libraries$state <- as.character(libraries$state)
libraries$state[libraries$state %in% rare_states] <- "Other"
libraries$state <- factor(libraries$state) # re-factor to clean up
# Do the same for testing data:
library_testing <- read_csv("../data/libraries_testing.csv")

## Rows: 1367 Columns: 28
## -- Column specification -----
## Delimiter: ","
## chr (2): state, Locale
## dbl (26): Service Area Population, Total Circulation, Percentage of Children...
```

```

## 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.

library_testing$state <- as.character(library_testing$state)
library_testing$state[library_testing$state %in% rare_states] <- "Other"
library_testing$state[!(library_testing$state %in% libraries$state)] <- "Other"
library_testing$state <- factor(library_testing$state) # re-factor to clean up
libraries <- libraries %>%
  mutate(across(where(is.character), as.factor))

```

Libraries is already the training set.

Full Linear Model and Stepwise Selected Linear Model

We can start by using a full linear model:

```

full.library.model <- glm(`Total Circulation` ~ ., data = libraries, family = "gaussian")
summary(full.library.model)

```

```

##
## Call:
## glm(formula = 'Total Circulation' ~ ., family = "gaussian", data = libraries)
##
## Coefficients:
##                               Estimate Std. Error t value
## (Intercept)                 -8.894e+03  9.180e+03 -0.969
## stateAL                  -1.107e+04  3.428e+03 -3.231
## stateCO                  -1.610e+04  4.380e+03 -3.677
## stateIA                  -9.671e+03  3.114e+03 -3.105
## stateID                  -5.999e+03  3.807e+03 -1.576
## stateIL                  -1.580e+04  3.177e+03 -4.973
## stateIN                  -1.094e+04  3.556e+03 -3.076
## stateKS                  -1.296e+04  3.419e+03 -3.791
## stateKY                  -1.798e+04  4.787e+03 -3.755
## stateMA                  -5.387e+03  3.369e+03 -1.599
## stateME                  -1.265e+04  3.399e+03 -3.723
## stateMI                  -1.177e+04  3.241e+03 -3.630
## stateMN                  -1.111e+04  3.776e+03 -2.941
## stateMO                  -1.148e+04  3.514e+03 -3.267
## stateMT                  -1.304e+04  4.011e+03 -3.251
## stateND                  -1.084e+04  3.908e+03 -2.774
## stateNE                  -6.097e+03  4.322e+03 -1.411
## stateNH                  -8.222e+03  3.700e+03 -2.222
## stateNJ                  -1.861e+04  3.884e+03 -4.793
## stateNM                  -1.467e+04  3.957e+03 -3.706
## stateNY                  -1.572e+04  3.158e+03 -4.977
## stateOK                  -1.194e+04  3.611e+03 -3.306
## stateOther                -1.113e+04  3.353e+03 -3.321
## statePA                  -1.040e+04  3.423e+03 -3.039
## stateSD                  -5.370e+03  3.593e+03 -1.494

```

```

## stateTN          -1.154e+04  3.648e+03 -3.164
## stateTX          -1.664e+04  3.448e+03 -4.825
## stateVT          -1.014e+04  3.612e+03 -2.808
## stateWI          -1.526e+04  3.396e+03 -4.494
## LocaleRural      2.225e+04  6.288e+03 3.539
## LocaleSuburban   2.274e+04  6.282e+03 3.620
## LocaleTown        2.220e+04  6.268e+03 3.542
## 'Service Area Population' 4.689e-01 8.326e-02 5.632
## 'Percentage of Children's Material Circulation' 7.980e+01 2.274e+01 3.509
## 'Central Library' -5.139e+03 7.209e+03 -0.713
## 'Branch Library' -3.719e+03 1.440e+03 -2.582
## Bookmobiles       -1.938e+03 2.729e+03 -0.710
## 'Internet Computers' -2.280e+02 6.402e+01 -3.561
## 'Computer Uses'  -2.753e-01 2.323e-01 -1.185
## 'Wireless Sessions' 1.392e-01 5.509e-02 2.527
## 'Children's Programs' -3.537e+00 7.988e+00 -0.443
## 'Young Adult Programs' -8.948e+00 2.435e+01 -0.367
## 'Adult Programs'  -1.717e+01 1.107e+01 -1.551
## 'Children's Program Attendance' 1.997e+00 3.877e-01 5.152
## 'Adult Program Attendance' 1.191e+00 1.140e+00 1.044
## 'General Interest Program Attendance' 6.314e-01 1.120e+00 0.564
## 'General Interest Programs' -3.232e+01 3.715e+01 -0.870
## 'Local Revenue ($)' 1.953e-02 2.039e-03 9.579
## 'State Revenue ($)' -6.348e-02 2.834e-02 -2.240
## 'Federal Revenue ($)' 1.765e-01 1.132e-01 1.559
## 'Other Revenue ($)' 1.339e-02 1.434e-02 0.934
## 'Hours/Year'        1.300e+00 6.445e-01 2.017
## 'Physical Visits' 7.287e-01 3.321e-02 21.944
## 'Reference Transactions' 1.248e-01 1.278e-01 0.977
## 'Registered Users' 1.169e+00 1.475e-01 7.925
## 'Inter-library Loans to Other Library' -2.243e-01 1.833e-01 -1.223
## 'Inter-library Loans from Other Library' 2.381e+00 1.935e-01 12.301
## Pr(>|t|)
## (Intercept) 0.332686
## stateAL     0.001251 ***
## stateCO     0.000241 ***
## stateIA     0.001921 **
## stateID     0.115235
## stateIL     7.02e-07 ***
## stateIN     0.002119 **
## stateKS     0.000153 ***
## stateKY     0.000177 ***
## stateMA     0.109958
## stateME     0.000201 ***
## stateMI     0.000289 ***
## stateMN     0.003301 **
## stateMO     0.001099 **
## stateMT     0.001165 **
## stateND     0.005579 **
## stateNE     0.158425
## stateNH     0.026383 *
## stateNJ     1.73e-06 ***
## stateNM     0.000215 ***
## stateNY     6.86e-07 ***

```

```

## stateOK          0.000960 ***
## stateOther       0.000908 ***
## statePA          0.002398 **
## stateSD          0.135171
## stateTN          0.001576 **
## stateTX          1.48e-06 ***
## stateVT          0.005024 **
## stateWI          7.31e-06 ***
## LocaleRural      0.000409 ***
## LocaleSuburban    0.000300 ***
## LocaleTown        0.000404 ***
## 'Service Area Population' 1.98e-08 ***
## 'Percentage of Children's Material Circulation' 0.000457 ***
## 'Central Library' 0.475989
## 'Branch Library' 0.009885 **
## Bookmobiles       0.477635
## 'Internet Computers' 0.000376 ***
## 'Computer Uses'   0.236130
## 'Wireless Sessions' 0.011553 *
## 'Children's Programs' 0.658015
## 'Young Adult Programs' 0.713302
## 'Adult Programs'   0.121020
## 'Children's Program Attendance' 2.77e-07 ***
## 'Adult Program Attendance' 0.296622
## 'General Interest Program Attendance' 0.573107
## 'General Interest Programs' 0.384432
## 'Local Revenue ($)' < 2e-16 ***
## 'State Revenue ($)' 0.025173 *
## 'Federal Revenue ($)' 0.119126
## 'Other Revenue ($)' 0.350252
## 'Hours/Year'        0.043794 *
## 'Physical Visits'   < 2e-16 ***
## 'Reference Transactions' 0.328869
## 'Registered Users' 3.34e-15 ***
## 'Inter-library Loans to Other Library' 0.221351
## 'Inter-library Loans from Other Library' < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1
##
## (Dispersion parameter for gaussian family taken to be 214086670)
##
## Null deviance: 2.1298e+12  on 2687  degrees of freedom
## Residual deviance: 5.6326e+11  on 2631  degrees of freedom
## AIC: 59248
##
## Number of Fisher Scoring iterations: 2

```

The full library model has lots of predictors from the dummy variables of state, but we can probably remove some of the predictors.

```
step(full.library.model, direction="both")
```

```
## Start:  AIC=59247.52
```

```

## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Central Library' +
##   'Branch Library' + Bookmobiles + 'Internet Computers' + 'Computer Uses' +
##   'Wireless Sessions' + 'Children's Programs' + 'Young Adult Programs' +
##   'Adult Programs' + 'Children's Program Attendance' + 'Adult Program Attendance' +
##   'General Interest Program Attendance' + 'General Interest Programs' +
##   'Local Revenue ($)' + 'State Revenue ($)' + 'Federal Revenue ($)' +
##   'Other Revenue ($)' + 'Hours/Year' + 'Physical Visits' +
##   'Reference Transactions' + 'Registered Users' + 'Inter-library Loans to Other Library' +
##   'Inter-library Loans from Other Library'

##
##                                     Df  Deviance    AIC
## - 'Young Adult Programs'           1 5.6329e+11 59246
## - 'Children's Programs'            1 5.6330e+11 59246
## - 'General Interest Program Attendance' 1 5.6333e+11 59246
## - Bookmobiles                      1 5.6337e+11 59246
## - 'Central Library'                1 5.6337e+11 59246
## - 'General Interest Programs'     1 5.6342e+11 59246
## - 'Other Revenue ($)'              1 5.6345e+11 59246
## - 'Reference Transactions'         1 5.6347e+11 59246
## - 'Adult Program Attendance'       1 5.6350e+11 59247
## - 'Computer Uses'                 1 5.6356e+11 59247
## - 'Inter-library Loans to Other Library' 1 5.6358e+11 59247
## <none>                            5.6326e+11 59248
## - 'Adult Programs'                 1 5.6378e+11 59248
## - 'Federal Revenue ($)'           1 5.6378e+11 59248
## - 'Hours/Year'                     1 5.6413e+11 59250
## - 'State Revenue ($)'              1 5.6434e+11 59251
## - 'Wireless Sessions'              1 5.6463e+11 59252
## - 'Branch Library'                 1 5.6469e+11 59252
## - Locale                           3 5.6607e+11 59255
## - 'Percentage of Children's Material Circulation' 1 5.6590e+11 59258
## - 'Internet Computers'             1 5.6598e+11 59258
## - 'Children's Program Attendance'  1 5.6894e+11 59273
## - 'Service Area Population'        1 5.7005e+11 59278
## - 'Registered Users'               1 5.7671e+11 59309
## - state                            28 5.9249e+11 59327
## - 'Local Revenue ($)'              1 5.8291e+11 59338
## - 'Inter-library Loans from Other Library' 1 5.9566e+11 59396
## - 'Physical Visits'                1 6.6636e+11 59697
##
## Step: AIC=59245.66
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Central Library' +
##   'Branch Library' + Bookmobiles + 'Internet Computers' + 'Computer Uses' +
##   'Wireless Sessions' + 'Children's Programs' + 'Adult Programs' +
##   'Children's Program Attendance' + 'Adult Program Attendance' +
##   'General Interest Program Attendance' + 'General Interest Programs' +
##   'Local Revenue ($)' + 'State Revenue ($)' + 'Federal Revenue ($)' +
##   'Other Revenue ($)' + 'Hours/Year' + 'Physical Visits' +
##   'Reference Transactions' + 'Registered Users' + 'Inter-library Loans to Other Library' +
##   'Inter-library Loans from Other Library'

##
##                                     Df  Deviance    AIC

```

```

## - 'General Interest Program Attendance'          1 5.6336e+11 59244
## - 'Children's Programs'                         1 5.6336e+11 59244
## - Bookmobiles                                  1 5.6340e+11 59244
## - 'Central Library'                            1 5.6340e+11 59244
## - 'General Interest Programs'                 1 5.6346e+11 59244
## - 'Other Revenue ($)'                          1 5.6348e+11 59245
## - 'Reference Transactions'                   1 5.6349e+11 59245
## - 'Adult Program Attendance'                  1 5.6351e+11 59245
## - 'Computer Uses'                            1 5.6360e+11 59245
## - 'Inter-library Loans to Other Library'    1 5.6361e+11 59245
## <none>                                       5.6329e+11 59246
## - 'Federal Revenue ($)'                      1 5.6381e+11 59246
## - 'Adult Programs'                           1 5.6382e+11 59246
## + 'Young Adult Programs'                     1 5.6326e+11 59248
## - 'Hours/Year'                                1 5.6417e+11 59248
## - 'State Revenue ($)'                        1 5.6435e+11 59249
## - 'Wireless Sessions'                        1 5.6466e+11 59250
## - 'Branch Library'                           1 5.6470e+11 59250
## - Locale                                     3 5.6610e+11 59253
## - 'Percentage of Children's Material Circulation' 1 5.6593e+11 59256
## - 'Internet Computers'                      1 5.6606e+11 59257
## - 'Children's Program Attendance'            1 5.6904e+11 59271
## - 'Service Area Population'                 1 5.7006e+11 59276
## - 'Registered Users'                        1 5.7671e+11 59307
## - state                                      28 5.9249e+11 59326
## - 'Local Revenue ($)'                       1 5.8291e+11 59336
## - 'Inter-library Loans from Other Library'  1 5.9568e+11 59394
## - 'Physical Visits'                          1 6.6706e+11 59698
##
## Step: AIC=59243.99
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Central Library' +
##   'Branch Library' + Bookmobiles + 'Internet Computers' + 'Computer Uses' +
##   'Wireless Sessions' + 'Children's Programs' + 'Adult Programs' +
##   'Children's Program Attendance' + 'Adult Program Attendance' +
##   'General Interest Programs' + 'Local Revenue ($)' + 'State Revenue ($)' +
##   'Federal Revenue ($)' + 'Other Revenue ($)' + 'Hours/Year' +
##   'Physical Visits' + 'Reference Transactions' + 'Registered Users' +
##   'Inter-library Loans to Other Library' + 'Inter-library Loans from Other Library'
##
##                                         Df  Deviance  AIC
## - 'Children's Programs'                      1 5.6343e+11 59242
## - Bookmobiles                               1 5.6346e+11 59242
## - 'General Interest Programs'               1 5.6346e+11 59242
## - 'Central Library'                         1 5.6347e+11 59243
## - 'Other Revenue ($)'                      1 5.6355e+11 59243
## - 'Reference Transactions'                 1 5.6356e+11 59243
## - 'Adult Program Attendance'                1 5.6361e+11 59243
## - 'Computer Uses'                          1 5.6366e+11 59243
## - 'Inter-library Loans to Other Library'  1 5.6367e+11 59243
## <none>                                       5.6336e+11 59244
## - 'Federal Revenue ($)'                   1 5.6386e+11 59244
## - 'Adult Programs'                        1 5.6391e+11 59245
## + 'General Interest Program Attendance'  1 5.6329e+11 59246

```

```

## + 'Young Adult Programs'          1 5.6333e+11 59246
## - 'Hours/Year'                   1 5.6426e+11 59246
## - 'State Revenue ($)'           1 5.6443e+11 59247
## - 'Wireless Sessions'            1 5.6473e+11 59249
## - 'Branch Library'               1 5.6479e+11 59249
## - Locale                         3 5.6616e+11 59251
## - 'Percentage of Children's Material Circulation' 1 5.6606e+11 59255
## - 'Internet Computers'           1 5.6612e+11 59255
## - 'Children's Program Attendance' 1 5.6923e+11 59270
## - 'Service Area Population'      1 5.7010e+11 59274
## - 'Registered Users'             1 5.7678e+11 59305
## - state                          28 5.9262e+11 59324
## - 'Local Revenue ($)'            1 5.8292e+11 59334
## - 'Inter-library Loans from Other Library'        1 5.9571e+11 59392
## - 'Physical Visits'              1 6.6754e+11 59698
##
## Step: AIC=59242.33
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Central Library' +
##   'Branch Library' + Bookmobiles + 'Internet Computers' + 'Computer Uses' +
##   'Wireless Sessions' + 'Adult Programs' + 'Children's Program Attendance' +
##   'Adult Program Attendance' + 'General Interest Programs' +
##   'Local Revenue ($)' + 'State Revenue ($)' + 'Federal Revenue ($)' +
##   'Other Revenue ($)' + 'Hours/Year' + 'Physical Visits' +
##   'Reference Transactions' + 'Registered Users' + 'Inter-library Loans to Other Library' +
##   'Inter-library Loans from Other Library'
##
##                                         Df  Deviance   AIC
## - Bookmobiles                      1 5.6353e+11 59241
## - 'Central Library'                 1 5.6354e+11 59241
## - 'General Interest Programs'      1 5.6355e+11 59241
## - 'Other Revenue ($)'              1 5.6362e+11 59241
## - 'Reference Transactions'         1 5.6363e+11 59241
## - 'Adult Program Attendance'       1 5.6371e+11 59242
## - 'Computer Uses'                  1 5.6372e+11 59242
## - 'Inter-library Loans to Other Library' 1 5.6375e+11 59242
## <none>                            5.6343e+11 59242
## - 'Federal Revenue ($)'           1 5.6394e+11 59243
## - 'Adult Programs'                 1 5.6407e+11 59243
## + 'Children's Programs'            1 5.6336e+11 59244
## + 'General Interest Program Attendance' 1 5.6336e+11 59244
## + 'Young Adult Programs'           1 5.6337e+11 59244
## - 'Hours/Year'                     1 5.6429e+11 59244
## - 'State Revenue ($)'              1 5.6451e+11 59245
## - 'Wireless Sessions'              1 5.6479e+11 59247
## - 'Branch Library'                 1 5.6485e+11 59247
## - Locale                          3 5.6623e+11 59250
## - 'Percentage of Children's Material Circulation' 1 5.6608e+11 59253
## - 'Internet Computers'             1 5.6622e+11 59254
## - 'Service Area Population'        1 5.7021e+11 59272
## - 'Children's Program Attendance'   1 5.7179e+11 59280
## - 'Registered Users'               1 5.7689e+11 59304
## - state                           28 5.9265e+11 59322
## - 'Local Revenue ($)'              1 5.8298e+11 59332

```

```

## - 'Inter-library Loans from Other Library'      1 5.9584e+11 59391
## - 'Physical Visits'                           1 6.6758e+11 59696
##
## Step: AIC=59240.8
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Central Library' +
##   'Branch Library' + 'Internet Computers' + 'Computer Uses' +
##   'Wireless Sessions' + 'Adult Programs' + 'Children's Program Attendance' +
##   'Adult Program Attendance' + 'General Interest Programs' +
##   'Local Revenue ($)' + 'State Revenue ($)' + 'Federal Revenue ($)' +
##   'Other Revenue ($)' + 'Hours/Year' + 'Physical Visits' +
##   'Reference Transactions' + 'Registered Users' + 'Inter-library Loans to Other Library' +
##   'Inter-library Loans from Other Library'
##
##                                         Df  Deviance    AIC
## - 'Central Library'                      1 5.6363e+11 59239
## - 'General Interest Programs'           1 5.6365e+11 59239
## - 'Reference Transactions'              1 5.6371e+11 59240
## - 'Other Revenue ($)'                  1 5.6372e+11 59240
## - 'Adult Program Attendance'           1 5.6381e+11 59240
## - 'Computer Uses'                     1 5.6382e+11 59240
## - 'Inter-library Loans to Other Library' 1 5.6385e+11 59240
## <none>                                5.6353e+11 59241
## - 'Federal Revenue ($)'               1 5.6404e+11 59241
## - 'Adult Programs'                   1 5.6417e+11 59242
## + Bookmobiles                         1 5.6343e+11 59242
## + 'Children's Programs'              1 5.6346e+11 59242
## + 'General Interest Program Attendance' 1 5.6346e+11 59242
## + 'Young Adult Programs'            1 5.6348e+11 59243
## - 'Hours/Year'                       1 5.6439e+11 59243
## - 'State Revenue ($)'                1 5.6460e+11 59244
## - 'Wireless Sessions'                1 5.6490e+11 59245
## - 'Branch Library'                  1 5.6506e+11 59246
## - Locale                            3 5.6631e+11 59248
## - 'Percentage of Children's Material Circulation' 1 5.6619e+11 59251
## - 'Internet Computers'             1 5.6629e+11 59252
## - 'Service Area Population'        1 5.7033e+11 59271
## - 'Children's Program Attendance'   1 5.7197e+11 59279
## - 'Registered Users'              1 5.7692e+11 59302
## - state                             28 5.9289e+11 59321
## - 'Local Revenue ($)'              1 5.8299e+11 59330
## - 'Inter-library Loans from Other Library' 1 5.9595e+11 59389
## - 'Physical Visits'                1 6.6770e+11 59695
##
## Step: AIC=59239.27
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Branch Library' +
##   'Internet Computers' + 'Computer Uses' + 'Wireless Sessions' +
##   'Adult Programs' + 'Children's Program Attendance' + 'Adult Program Attendance' +
##   'General Interest Programs' + 'Local Revenue ($)' + 'State Revenue ($)' +
##   'Federal Revenue ($)' + 'Other Revenue ($)' + 'Hours/Year' +
##   'Physical Visits' + 'Reference Transactions' + 'Registered Users' +
##   'Inter-library Loans to Other Library' + 'Inter-library Loans from Other Library'
##

```

```

##                                     Df Deviance AIC
## - 'General Interest Programs'      1 5.6374e+11 59238
## - 'Other Revenue ($)'             1 5.6380e+11 59238
## - 'Reference Transactions'        1 5.6381e+11 59238
## - 'Adult Program Attendance'      1 5.6390e+11 59239
## - 'Computer Uses'                1 5.6392e+11 59239
## - 'Inter-library Loans to Other Library' 1 5.6394e+11 59239
## <none>                           5.6363e+11 59239
## - 'Federal Revenue ($)'          1 5.6413e+11 59240
## - 'Adult Programs'               1 5.6426e+11 59240
## + 'Central Library'              1 5.6353e+11 59241
## + Bookmobiles                    1 5.6354e+11 59241
## + 'General Interest Program Attendance' 1 5.6356e+11 59241
## + 'Children's Programs'          1 5.6357e+11 59241
## + 'Young Adult Programs'         1 5.6357e+11 59241
## - 'Hours/Year'                  1 5.6445e+11 59241
## - 'State Revenue ($)'           1 5.6471e+11 59242
## - 'Wireless Sessions'           1 5.6504e+11 59244
## - 'Branch Library'              1 5.6506e+11 59244
## - Locale                         3 5.6631e+11 59246
## - 'Percentage of Children's Material Circulation' 1 5.6629e+11 59250
## - 'Internet Computers'          1 5.6637e+11 59250
## - 'Service Area Population'     1 5.7042e+11 59269
## - 'Children's Program Attendance' 1 5.7215e+11 59278
## - 'Registered Users'            1 5.7701e+11 59300
## - state                          28 5.9292e+11 59319
## - 'Local Revenue ($)'           1 5.8299e+11 59328
## - 'Inter-library Loans from Other Library' 1 5.9606e+11 59388
## - 'Physical Visits'              1 6.6803e+11 59694
##
## Step: AIC=59237.81
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Branch Library' +
##   'Internet Computers' + 'Computer Uses' + 'Wireless Sessions' +
##   'Adult Programs' + 'Children's Program Attendance' + 'Adult Program Attendance' +
##   'Local Revenue ($)' + 'State Revenue ($)' + 'Federal Revenue ($)' +
##   'Other Revenue ($)' + 'Hours/Year' + 'Physical Visits' +
##   'Reference Transactions' + 'Registered Users' + 'Inter-library Loans to Other Library' +
##   'Inter-library Loans from Other Library'
##
##                                     Df Deviance AIC
## - 'Other Revenue ($)'             1 5.6390e+11 59237
## - 'Reference Transactions'        1 5.6391e+11 59237
## - 'Adult Program Attendance'      1 5.6402e+11 59237
## - 'Computer Uses'                1 5.6404e+11 59237
## - 'Inter-library Loans to Other Library' 1 5.6405e+11 59237
## <none>                           5.6374e+11 59238
## - 'Federal Revenue ($)'          1 5.6423e+11 59238
## - 'Adult Programs'               1 5.6443e+11 59239
## + 'General Interest Programs'    1 5.6363e+11 59239
## + 'Central Library'              1 5.6365e+11 59239
## + Bookmobiles                    1 5.6365e+11 59239
## + 'Children's Programs'          1 5.6366e+11 59239
## + 'Young Adult Programs'         1 5.6367e+11 59239

```

```

## - 'Hours/Year'                                1 5.6453e+11 59240
## + 'General Interest Program Attendance'      1 5.6374e+11 59240
## - 'State Revenue ($)'                        1 5.6480e+11 59241
## - 'Branch Library'                           1 5.6513e+11 59242
## - 'Wireless Sessions'                        1 5.6515e+11 59243
## - 'Locale'                                  3 5.6643e+11 59245
## - 'Percentage of Children's Material Circulation' 1 5.6635e+11 59248
## - 'Internet Computers'                      1 5.6653e+11 59249
## - 'Service Area Population'                 1 5.7054e+11 59268
## - 'Children's Program Attendance'            1 5.7226e+11 59276
## - 'Registered Users'                        1 5.7726e+11 59300
## - state                                    28 5.9300e+11 59318
## - 'Local Revenue ($)'                      1 5.8315e+11 59327
## - 'Inter-library Loans from Other Library' 1 5.9625e+11 59387
## - 'Physical Visits'                         1 6.6818e+11 59693
##
## Step: AIC=59236.57
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Branch Library' +
##   'Internet Computers' + 'Computer Uses' + 'Wireless Sessions' +
##   'Adult Programs' + 'Children's Program Attendance' + 'Adult Program Attendance' +
##   'Local Revenue ($)' + 'State Revenue ($)' + 'Federal Revenue ($)' +
##   'Hours/Year' + 'Physical Visits' + 'Reference Transactions' +
##   'Registered Users' + 'Inter-library Loans to Other Library' +
##   'Inter-library Loans from Other Library'
##
##                                         Df  Deviance  AIC
## - 'Reference Transactions'                  1 5.6407e+11 59235
## - 'Computer Uses'                          1 5.6420e+11 59236
## - 'Adult Program Attendance'               1 5.6421e+11 59236
## - 'Inter-library Loans to Other Library'  1 5.6421e+11 59236
## <none>                                    5.6390e+11 59237
## - 'Federal Revenue ($)'                  1 5.6440e+11 59237
## - 'Adult Programs'                        1 5.6456e+11 59238
## + 'Other Revenue ($)'                    1 5.6374e+11 59238
## + 'General Interest Programs'            1 5.6380e+11 59238
## + Bookmobiles                            1 5.6381e+11 59238
## + 'Central Library'                      1 5.6382e+11 59238
## + 'Children's Programs'                  1 5.6383e+11 59238
## + 'Young Adult Programs'                1 5.6383e+11 59238
## - 'Hours/Year'                            1 5.6474e+11 59239
## + 'General Interest Program Attendance'  1 5.6390e+11 59239
## - 'State Revenue ($)'                   1 5.6493e+11 59239
## - 'Branch Library'                       1 5.6532e+11 59241
## - 'Wireless Sessions'                   1 5.6534e+11 59241
## - 'Locale'                               3 5.6649e+11 59243
## - 'Percentage of Children's Material Circulation' 1 5.6648e+11 59247
## - 'Internet Computers'                  1 5.6663e+11 59248
## - 'Service Area Population'             1 5.7072e+11 59267
## - 'Children's Program Attendance'       1 5.7245e+11 59275
## - 'Registered Users'                   1 5.7762e+11 59299
## - state                                28 5.9301e+11 59316
## - 'Local Revenue ($)'                  1 5.8315e+11 59325
## - 'Inter-library Loans from Other Library' 1 5.9670e+11 59387

```

```

## - 'Physical Visits' 1 6.6934e+11 59695
##
## Step: AIC=59235.39
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Branch Library' +
##   'Internet Computers' + 'Computer Uses' + 'Wireless Sessions' +
##   'Adult Programs' + 'Children's Program Attendance' + 'Adult Program Attendance' +
##   'Local Revenue ($)' + 'State Revenue ($)' + 'Federal Revenue ($)' +
##   'Hours/Year' + 'Physical Visits' + 'Registered Users' + 'Inter-library Loans to Other Library' +
##   'Inter-library Loans from Other Library'
##
##                                     Df Deviance AIC
## - 'Computer Uses'           1 5.6433e+11 59235
## - 'Inter-library Loans to Other Library' 1 5.6436e+11 59235
## - 'Adult Program Attendance' 1 5.6440e+11 59235
## <none>                      5.6407e+11 59235
## - 'Federal Revenue ($)'    1 5.6456e+11 59236
## - 'Adult Programs'          1 5.6474e+11 59237
## + 'Reference Transactions' 1 5.6390e+11 59237
## + 'Other Revenue ($)'      1 5.6391e+11 59237
## + 'Central Library'         1 5.6399e+11 59237
## + 'General Interest Programs' 1 5.6399e+11 59237
## + Bookmobiles               1 5.6400e+11 59237
## + 'Children's Programs'     1 5.6401e+11 59237
## + 'Young Adult Programs'   1 5.6402e+11 59237
## + 'General Interest Program Attendance' 1 5.6407e+11 59237
## - 'Hours/Year'              1 5.6492e+11 59237
## - 'State Revenue ($)'      1 5.6511e+11 59238
## - 'Branch Library'          1 5.6549e+11 59240
## - 'Wireless Sessions'       1 5.6556e+11 59240
## - Locale                    3 5.6664e+11 59242
## - 'Percentage of Children's Material Circulation' 1 5.6669e+11 59246
## - 'Internet Computers'     1 5.6677e+11 59246
## - 'Service Area Population' 1 5.7103e+11 59266
## - 'Children's Program Attendance' 1 5.7270e+11 59274
## - 'Registered Users'        1 5.7823e+11 59300
## - state                     28 5.9322e+11 59315
## - 'Local Revenue ($)'      1 5.8358e+11 59325
## - 'Inter-library Loans from Other Library' 1 5.9693e+11 59386
## - 'Physical Visits'         1 6.7104e+11 59700
##
## Step: AIC=59234.62
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Branch Library' +
##   'Internet Computers' + 'Wireless Sessions' + 'Adult Programs' +
##   'Children's Program Attendance' + 'Adult Program Attendance' +
##   'Local Revenue ($)' + 'State Revenue ($)' + 'Federal Revenue ($)' +
##   'Hours/Year' + 'Physical Visits' + 'Registered Users' + 'Inter-library Loans to Other Library' +
##   'Inter-library Loans from Other Library'
##
##                                     Df Deviance AIC
## - 'Adult Program Attendance' 1 5.6461e+11 59234
## - 'Inter-library Loans to Other Library' 1 5.6464e+11 59234
## <none>                      5.6433e+11 59235

```

```

## - 'Federal Revenue ($)'
## + 'Computer Uses'
## - 'Adult Programs'
## + 'Other Revenue ($)'
## + 'Reference Transactions'
## + 'Central Library'
## + 'General Interest Programs'
## + Bookmobiles
## + 'Young Adult Programs'
## + 'Children's Programs'
## - 'Hours/Year'
## + 'General Interest Program Attendance'
## - 'State Revenue ($)'
## - 'Wireless Sessions'
## - 'Branch Library'
## - Locale
## - 'Percentage of Children's Material Circulation'
## - 'Internet Computers'
## - 'Service Area Population'
## - 'Children's Program Attendance'
## - 'Registered Users'
## - state
## - 'Local Revenue ($)'
## - 'Inter-library Loans from Other Library'
## - 'Physical Visits'

##
## Step: AIC=59233.97
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##      'Percentage of Children's Material Circulation' + 'Branch Library' +
##      'Internet Computers' + 'Wireless Sessions' + 'Adult Programs' +
##      'Children's Program Attendance' + 'Local Revenue ($)' + 'State Revenue ($)' +
##      'Federal Revenue ($)' + 'Hours/Year' + 'Physical Visits' +
##      'Registered Users' + 'Inter-library Loans to Other Library' +
##      'Inter-library Loans from Other Library'
##
##                                     Df  Deviance   AIC
## - 'Adult Programs'           1 5.6494e+11 59234
## - 'Inter-library Loans to Other Library' 1 5.6494e+11 59234
## <none>                      5.6461e+11 59234
## - 'Federal Revenue ($)'
## + 'Adult Program Attendance' 1 5.6508e+11 59234
## + 'Computer Uses'           1 5.6433e+11 59235
## + 'Other Revenue ($)'
## + 'Reference Transactions'  1 5.6442e+11 59235
## + 'General Interest Programs' 1 5.6447e+11 59235
## + 'Children's Programs'      1 5.6452e+11 59236
## + 'Central Library'          1 5.6454e+11 59236
## + Bookmobiles                1 5.6454e+11 59236
## - 'Hours/Year'               1 5.6540e+11 59236
## + 'Young Adult Programs'    1 5.6457e+11 59236
## + 'General Interest Program Attendance' 1 5.6460e+11 59236
## - 'State Revenue ($)'
## - 'Branch Library'          1 5.6569e+11 59237
## - 'Wireless Sessions'        1 5.6599e+11 59239
##                                     1 5.6606e+11 59239

```

```

## - Locale                                     3 5.6713e+11 59240
## - 'Percentage of Children's Material Circulation' 1 5.6726e+11 59245
## - 'Internet Computers'                      1 5.6817e+11 59249
## - 'Service Area Population'                 1 5.7138e+11 59264
## - 'Children's Program Attendance'           1 5.7411e+11 59277
## - 'Registered Users'                       1 5.7885e+11 59299
## - state                                      28 5.9348e+11 59312
## - 'Local Revenue ($)'                      1 5.8433e+11 59324
## - 'Inter-library Loans from Other Library' 1 5.9806e+11 59387
## - 'Physical Visits'                        1 6.7825e+11 59725
##
## Step: AIC=59233.51
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Branch Library' +
##   'Internet Computers' + 'Wireless Sessions' + 'Children's Program Attendance' +
##   'Local Revenue ($)' + 'State Revenue ($)' + 'Federal Revenue ($)' +
##   'Hours/Year' + 'Physical Visits' + 'Registered Users' + 'Inter-library Loans to Other Library' +
##   'Inter-library Loans from Other Library'
##
##                                         Df  Deviance   AIC
## - 'Inter-library Loans to Other Library'    1 5.6526e+11 59233
## <none>                                    5.6494e+11 59234
## - 'Federal Revenue ($)'                   1 5.6537e+11 59234
## + 'Adult Programs'                      1 5.6461e+11 59234
## + 'Computer Uses'                      1 5.6474e+11 59235
## - 'Hours/Year'                          1 5.6563e+11 59235
## + 'Children's Programs'                 1 5.6479e+11 59235
## + 'Other Revenue ($)'                  1 5.6480e+11 59235
## + 'Reference Transactions'            1 5.6480e+11 59235
## + 'General Interest Programs'        1 5.6481e+11 59235
## + 'Young Adult Programs'             1 5.6483e+11 59235
## + 'Central Library'                  1 5.6486e+11 59235
## + Bookmobiles                         1 5.6486e+11 59235
## + 'Adult Program Attendance'         1 5.6493e+11 59235
## + 'General Interest Program Attendance' 1 5.6494e+11 59235
## - 'State Revenue ($)'                1 5.6598e+11 59236
## - 'Branch Library'                  1 5.6624e+11 59238
## - 'Wireless Sessions'               1 5.6635e+11 59238
## - Locale                           3 5.6741e+11 59239
## - 'Percentage of Children's Material Circulation' 1 5.6762e+11 59244
## - 'Internet Computers'              1 5.6866e+11 59249
## - 'Service Area Population'        1 5.7201e+11 59265
## - 'Children's Program Attendance'   1 5.7411e+11 59275
## - 'Registered Users'              1 5.7917e+11 59298
## - state                           28 5.9408e+11 59313
## - 'Local Revenue ($)'              1 5.8434e+11 59322
## - 'Inter-library Loans from Other Library' 1 5.9808e+11 59385
## - 'Physical Visits'               1 6.7826e+11 59723
##
## Step: AIC=59233.06
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##   'Percentage of Children's Material Circulation' + 'Branch Library' +
##   'Internet Computers' + 'Wireless Sessions' + 'Children's Program Attendance' +
##   'Local Revenue ($)' + 'State Revenue ($)' + 'Federal Revenue ($)' +

```

```

##      'Hours/Year' + 'Physical Visits' + 'Registered Users' + 'Inter-library Loans from Other Library'
##
##      Df    Deviance   AIC
## - 'Federal Revenue ($)'          1 5.6568e+11 59233
## <none>                          5.6526e+11 59233
## + 'Inter-library Loans to Other Library' 1 5.6494e+11 59234
## + 'Adult Programs'              1 5.6494e+11 59234
## + 'Computer Uses'              1 5.6505e+11 59234
## - 'Hours/Year'                  1 5.6590e+11 59234
## + 'Children's Programs'         1 5.6510e+11 59234
## + 'Other Revenue ($)'          1 5.6512e+11 59234
## + 'General Interest Programs' 1 5.6513e+11 59234
## + 'Reference Transactions'     1 5.6515e+11 59235
## + 'Young Adult Programs'       1 5.6516e+11 59235
## + 'Central Library'            1 5.6519e+11 59235
## + Bookmobiles                  1 5.6519e+11 59235
## + 'Adult Program Attendance'   1 5.6526e+11 59235
## + 'General Interest Program Attendance' 1 5.6526e+11 59235
## - 'State Revenue ($)'          1 5.6641e+11 59237
## - 'Branch Library'             1 5.6651e+11 59237
## - 'Wireless Sessions'          1 5.6671e+11 59238
## - Locale                       3 5.6772e+11 59239
## - 'Percentage of Children's Material Circulation' 1 5.6799e+11 59244
## - 'Internet Computers'        1 5.6905e+11 59249
## - 'Service Area Population'   1 5.7237e+11 59265
## - 'Children's Program Attendance' 1 5.7446e+11 59274
## - 'Registered Users'          1 5.7989e+11 59300
## - state                        28 5.9454e+11 59313
## - 'Local Revenue ($)'          1 5.8434e+11 59320
## - 'Inter-library Loans from Other Library' 1 6.2019e+11 59480
## - 'Physical Visits'            1 6.7847e+11 59722
##
## Step: AIC=59233.04
## 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##      'Percentage of Children's Material Circulation' + 'Branch Library' +
##      'Internet Computers' + 'Wireless Sessions' + 'Children's Program Attendance' +
##      'Local Revenue ($)' + 'State Revenue ($)' + 'Hours/Year' +
##      'Physical Visits' + 'Registered Users' + 'Inter-library Loans from Other Library'
##
##      Df    Deviance   AIC
## <none>                          5.6568e+11 59233
## + 'Federal Revenue ($)'          1 5.6526e+11 59233
## + 'Inter-library Loans to Other Library' 1 5.6537e+11 59234
## + 'Adult Programs'              1 5.6539e+11 59234
## + 'Computer Uses'              1 5.6548e+11 59234
## + 'Children's Programs'         1 5.6552e+11 59234
## + 'Other Revenue ($)'          1 5.6553e+11 59234
## - 'Hours/Year'                  1 5.6637e+11 59234
## + 'General Interest Programs' 1 5.6556e+11 59234
## + 'Reference Transactions'     1 5.6558e+11 59235
## + 'Young Adult Programs'       1 5.6559e+11 59235
## + Bookmobiles                  1 5.6561e+11 59235
## + 'Central Library'            1 5.6561e+11 59235
## + 'General Interest Program Attendance' 1 5.6568e+11 59235

```

```

## + 'Adult Program Attendance'          1 5.6568e+11 59235
## - 'State Revenue ($)'              1 5.6678e+11 59236
## - 'Branch Library'                 1 5.6701e+11 59237
## - 'Wireless Sessions'              1 5.6723e+11 59238
## - Locale                           3 5.6817e+11 59239
## - 'Percentage of Children's Material Circulation' 1 5.6847e+11 59244
## - 'Internet Computers'            1 5.6938e+11 59249
## - 'Service Area Population'       1 5.7277e+11 59265
## - 'Children's Program Attendance' 1 5.7490e+11 59275
## - 'Registered Users'              1 5.8021e+11 59299
## - state                            28 5.9615e+11 59318
## - 'Local Revenue ($)'             1 5.8483e+11 59321
## - 'Inter-library Loans from Other Library' 1 6.2060e+11 59480
## - 'Physical Visits'               1 6.7879e+11 59721

##
## Call: glm(formula = 'Total Circulation' ~ state + Locale + 'Service Area Population' +
##           'Percentage of Children's Material Circulation' + 'Branch Library' +
##           'Internet Computers' + 'Wireless Sessions' + 'Children's Program Attendance' +
##           'Local Revenue ($)' + 'State Revenue ($)' + 'Hours/Year' +
##           'Physical Visits' + 'Registered Users' + 'Inter-library Loans from Other Library',
##           family = "gaussian", data = libraries)
##
## Coefficients:
##                               (Intercept)
##                               -1.162e+04
## stateAL                         -1.090e+04
## stateCO                         -1.635e+04
## stateIA                         -1.040e+04
## stateID                         -6.715e+03
## stateIL                         -1.626e+04
## stateIN                         -1.157e+04
## stateKS                         -1.361e+04
## stateKY                         -1.826e+04
## stateMA                         -6.365e+03
## stateME                         -1.304e+04
## stateMI                         -1.210e+04
## stateMN                         -1.171e+04
## stateMO                         -1.174e+04
## stateMT                         -1.412e+04

```

```

##                                     stateND
##                                     -1.155e+04
##                                     stateNE
##                                     -6.055e+03
##                                     stateNH
##                                     -8.633e+03
##                                     stateNJ
##                                     -1.972e+04
##                                     stateNM
##                                     -1.503e+04
##                                     stateNY
##                                     -1.656e+04
##                                     stateOK
##                                     -1.199e+04
##                                     stateOther
##                                     -1.143e+04
##                                     statePA
##                                     -1.081e+04
##                                     stateSD
##                                     -6.088e+03
##                                     stateTN
##                                     -1.143e+04
##                                     stateTX
##                                     -1.698e+04
##                                     stateVT
##                                     -9.788e+03
##                                     stateWI
##                                     -1.611e+04
##                                     LocaleRural
##                                     2.055e+04
##                                     LocaleSuburban
##                                     2.097e+04
##                                     LocaleTown
##                                     2.064e+04
##                                     'Service Area Population'
##                                     4.753e-01
##                                     'Percentage of Children's Material Circulation'
##                                     8.119e+01
##                                     'Branch Library'
##                                     -3.377e+03
##                                     'Internet Computers'
##                                     -2.530e+02
##                                     'Wireless Sessions'
##                                     1.458e-01
##                                     'Children's Program Attendance'
##                                     1.911e+00
##                                     'Local Revenue ($)'
##                                     1.879e-02
##                                     'State Revenue ($)'
##                                     -6.373e-02
##                                     'Hours/Year'
##                                     1.133e+00
##                                     'Physical Visits'
##                                     7.192e-01

```

```

##                               'Registered Users'
##                               1.199e+00
## 'Inter-library Loans from Other Library'
##                               2.230e+00
##
## Degrees of Freedom: 2687 Total (i.e. Null);  2644 Residual
## Null Deviance:      2.13e+12
## Residual Deviance: 5.657e+11    AIC: 59230

```

The best model uses less predictors, and we can save it:

```

best.stepwise.model <- glm(formula = `Total Circulation` ~ state + Locale + `Service Area Population` +
  `Percentage of Children's Material Circulation` + `Branch Library` +
  `Internet Computers` + `Wireless Sessions` + `Children's Program Attendance` +
  `Local Revenue ($)` + `State Revenue ($)` + `Hours/Year` +
  `Physical Visits` + `Registered Users` + `Inter-library Loans from Other Library`,
  data = libraries, family = "gaussian")

```

We can compare the prediction errors for both models using CV:

```

set.seed(13)
cv.glm(libraries, best.stepwise.model, K=100)$delta[2]

```

```
## [1] 222741182
```

```

set.seed(13)
cv.glm(libraries, full.library.model, K=100)$delta[2]

```

```
## [1] 225166463
```

Here, we see that the stepwise model has a lower delta than the full selection model.

We can test both models against the testing data.

```

full.predictions <- predict(full.library.model, newdata = library_testing, type = "response")
full.residuals <- library_testing$`Total Circulation` - full.predictions
stepwise.predictions <- predict(best.stepwise.model, newdata = library_testing, type = "response")
stepwise.residuals <- library_testing$`Total Circulation` - stepwise.predictions
full.MSE <- mean((full.residuals)^2)
stepwise.MSE <- mean((stepwise.residuals)^2)
print(c(full.MSE, stepwise.MSE))

```

```
## [1] 436277589977 437612100407
```

On the testing data, the full model actually performed better than the stepwise model, going against the CV results.

Ridge and LASSO Shrinkage Models

Start by getting the design and response matrices

```

libraries.X <- model.matrix(full.library.model) [, -1]
libraries.Y <- full.library.model$data$`Total Circulation` 

#lets use cross validation and get a ridge and LASSO model
set.seed(4545)
libraries.ridge <- cv.glmnet(x = libraries.X, y=libraries.Y, alpha=0)
set.seed(4545)
libraries.lasso <- cv.glmnet(x = libraries.X, y=libraries.Y, alpha=1)

```

Tree Methods

I don't know why this isn't working

```

set.seed(4545)
libraries.tree <- select(libraries, - state, -Locale, -`Central Library`, -`Branch Library`, -Bookmobile)

# tree.libraries <- tree(`Total Circulation` ~ ., data = libraries.tree)

```

PLSR

```

libraries.plsr <- plsr(`Total Circulation` ~ ., data=libraries, scale=T, validation="CV")
summary(libraries.plsr)

```

```

## Data:      X dimension: 2688 56
##   Y dimension: 2688 1
## Fit method: kernelpls
## Number of components considered: 56
##
## VALIDATION: RMSEP
## Cross-validated using 10 random segments.
##             (Intercept) 1 comps  2 comps  3 comps  4 comps  5 comps  6 comps
## CV          28159    17212    15671    15309    15136    15071    15053
## adjCV       28159    17209    15647    15283    15110    15046    15028
##             7 comps  8 comps  9 comps  10 comps 11 comps 12 comps 13 comps
## CV          15045    15049    15046    15049    15046    15043    15038
## adjCV       15019    15023    15020    15022    15019    15015    15009
##             14 comps 15 comps 16 comps 17 comps 18 comps 19 comps 20 comps
## CV          15018    15011    15007    15009    15012    15015    15019
## adjCV       14989    14983    14980    14982    14985    14987    14990
##             21 comps 22 comps 23 comps 24 comps 25 comps 26 comps 27 comps
## CV          15025    15026    15030    15028    15026    15024    15024
## adjCV       14996    14997    15000    14998    14996    14995    14995
##             28 comps 29 comps 30 comps 31 comps 32 comps 33 comps 34 comps
## CV          15024    15024    15024    15024    15024    15024    15024
## adjCV       14995    14995    14995    14995    14995    14995    14995
##             35 comps 36 comps 37 comps 38 comps 39 comps 40 comps 41 comps
## CV          15024    15024    15024    15024    15024    15024    15024
## adjCV       14995    14995    14995    14995    14995    14995    14995

```

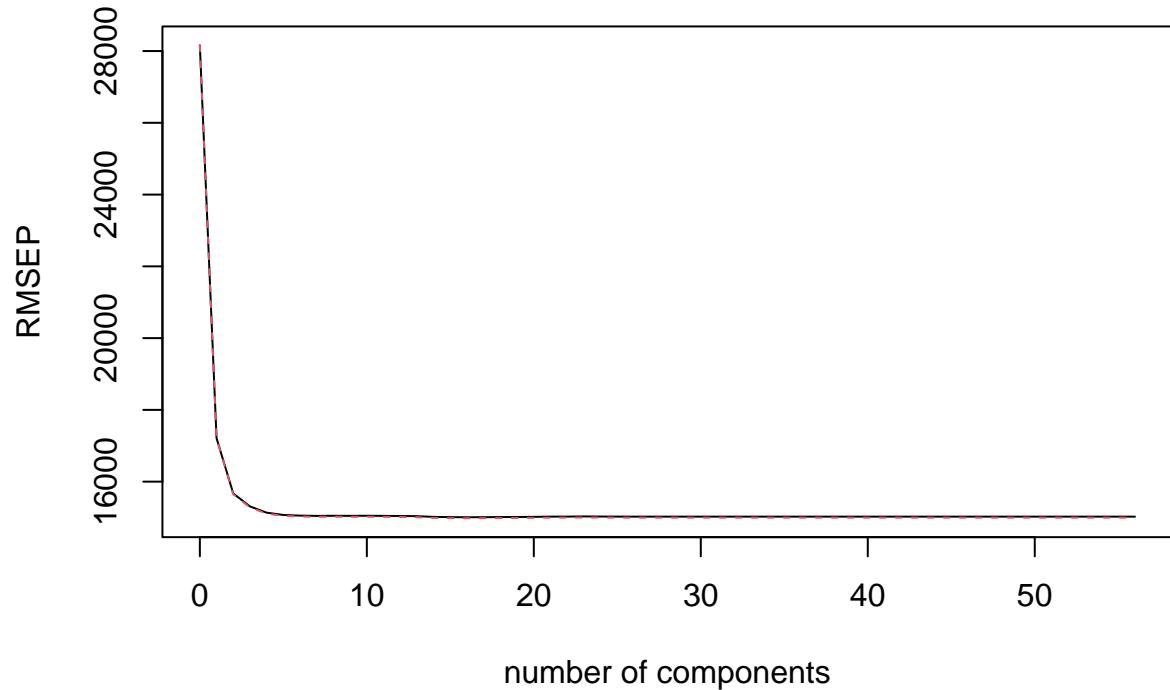
```

##          42 comps 43 comps 44 comps 45 comps 46 comps 47 comps 48 comps
## CV        15024   15024   15024   15024   15024   15024   15024
## adjCV    14995   14995   14995   14995   14995   14995   14995
##          49 comps 50 comps 51 comps 52 comps 53 comps 54 comps 55 comps
## CV        15024   15024   15024   15024   15024   15024   15024
## adjCV    14995   14995   14995   14995   14995   14995   14995
##          56 comps
## CV        15024
## adjCV    14995
##
## TRAINING: % variance explained
##          1 comps 2 comps 3 comps 4 comps 5 comps 6 comps
## X         14.64   16.93   19.67   22.61   24.85   27.30
## Total Circulation 62.98   70.59   72.21   72.89   73.14   73.22
##          7 comps 8 comps 9 comps 10 comps 11 comps 12 comps
## X         29.13   30.81   32.32   34.26   35.81   37.01
## Total Circulation 73.27   73.29   73.32   73.33   73.35   73.39
##          13 comps 14 comps 15 comps 16 comps 17 comps 18 comps
## X         38.34   39.83   41.39   43.15   44.64   46.06
## Total Circulation 73.45   73.50   73.52   73.52   73.52   73.52
##          19 comps 20 comps 21 comps 22 comps 23 comps 24 comps
## X         47.60   49.06   50.28   51.28   52.53   53.92
## Total Circulation 73.53   73.53   73.54   73.55   73.55   73.55
##          25 comps 26 comps 27 comps 28 comps 29 comps 30 comps
## X         55.30   56.55   57.85   59.16   60.22   61.65
## Total Circulation 73.55   73.55   73.55   73.55   73.55   73.55
##          31 comps 32 comps 33 comps 34 comps 35 comps 36 comps
## X         62.65   64.05   65.26   66.62   67.85   69.19
## Total Circulation 73.55   73.55   73.55   73.55   73.55   73.55
##          37 comps 38 comps 39 comps 40 comps 41 comps 42 comps
## X         70.58   71.89   73.13   74.51   76.02   77.46
## Total Circulation 73.55   73.55   73.55   73.55   73.55   73.55
##          43 comps 44 comps 45 comps 46 comps 47 comps 48 comps
## X         78.60   79.90   81.33   82.76   84.19   85.93
## Total Circulation 73.55   73.55   73.55   73.55   73.55   73.55
##          49 comps 50 comps 51 comps 52 comps 53 comps 54 comps
## X         87.13   88.97   90.82   92.65   94.49   96.33
## Total Circulation 73.55   73.55   73.55   73.55   73.55   73.55
##          55 comps 56 comps
## X         98.17   100.01
## Total Circulation 73.55   73.55

```

```
validationplot(libraries.plsr)
```

Total Circulation



```
# 14 comps for minimum MSE, but I think that we could less
```