ISyE 4031 T09 - Georgia Achievement Gaps in K-12 Schools with Regression

Ruwei Ma, Yichen Ma, Yang Yang

1. Introduction

Covid-19 had brought a big impact to the education system across US. National test results for 2022 reveal the pandemic's devastating effects on American schoolchildren, with the performance of 9-year-olds in math and reading dropping to the lowest levels from two decades ago [1]. This lagging effect from the pandemic applies to all races and income levels and sparks a collective decline in academics for the generation that experienced school closures, frequent reliance on virtual and remote learning, and other pandemic effects. The setbacks will occupy the low-performing students for up to 9 months to catch up with the average, prompting an urgent need for the underlying solution to the achievement gap [2]. This setback further adds to, and likely aggravates, the pre-pandemic disparity in student achievement outcomes for vulnerable and at-risk student populations, especially in Georgia. Based on some of my preliminary analysis of the 2021 achievement data across 2,180 schools in Georgia, we found that there are 2 prominent factors that affect achievement rate: the student's economic status and race. The achievement rate in 2021 of economically disadvantaged students is 46.11%, compared to 52.32% across all students. A similar gap can be observed in the difference in achievement rate between white and black students in Georgia, the former as high as 66.99%, compared to the 39.88% of the latter. The gap within the economically-disadvantaged students' group is vast and depends on the county or school they attend. Further analysis at the school level shows strong correlation between achievement rate and the school's other demographics.

2. Problem Goal

We aim to adopt regression modeling to identify gaps in national test achievement rates between different demographic groups in Georgia, and recommend robust strategies to address such disparities. Specifically, the objectives are: (1) visualize the disparities in school resources, such as teacher certifications and FTE (Full-time Equivalent), and quantify its correlation with the student's achievement outcomes, especially among marginalized minority groups (e.g., White, Black, vs. Hispanic students, economically disadvantaged vs. affluent students, and rural vs. Urban schools) (2) quantify the achievement gap at the county level across Georgia's 159 counties at the school level to identify factors that predict student achievement and highlight intervention or resource allocation strategies, and (3) evaluate the impact and predict the trajectory of the policies and strategies produced from step 2 with adjustments.

3. Executive Summary

4. Data Description

```r
# Input Dataset
library(readxl)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

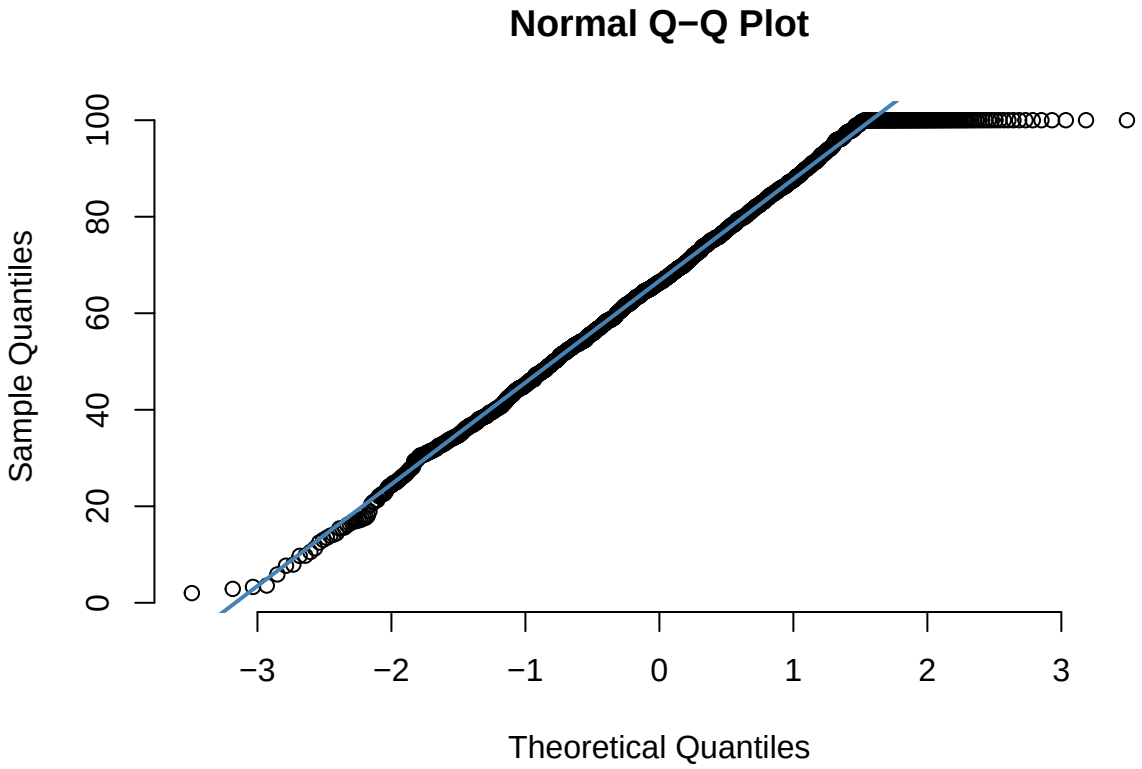
```r
Achievement_Rate = read.csv("2019 & 2021 Content Mastery Data.csv", header=TRUE)
Percentage = read.csv("Percentages & Certificates.csv", header=TRUE)
Salaries = read.csv("salaries.csv", header=TRUE)
Absent_Rate = read.csv("Absent Rate.csv", header=TRUE)
School_Expenditure = read_excel("2021_School-Level_PPE.xls")
```

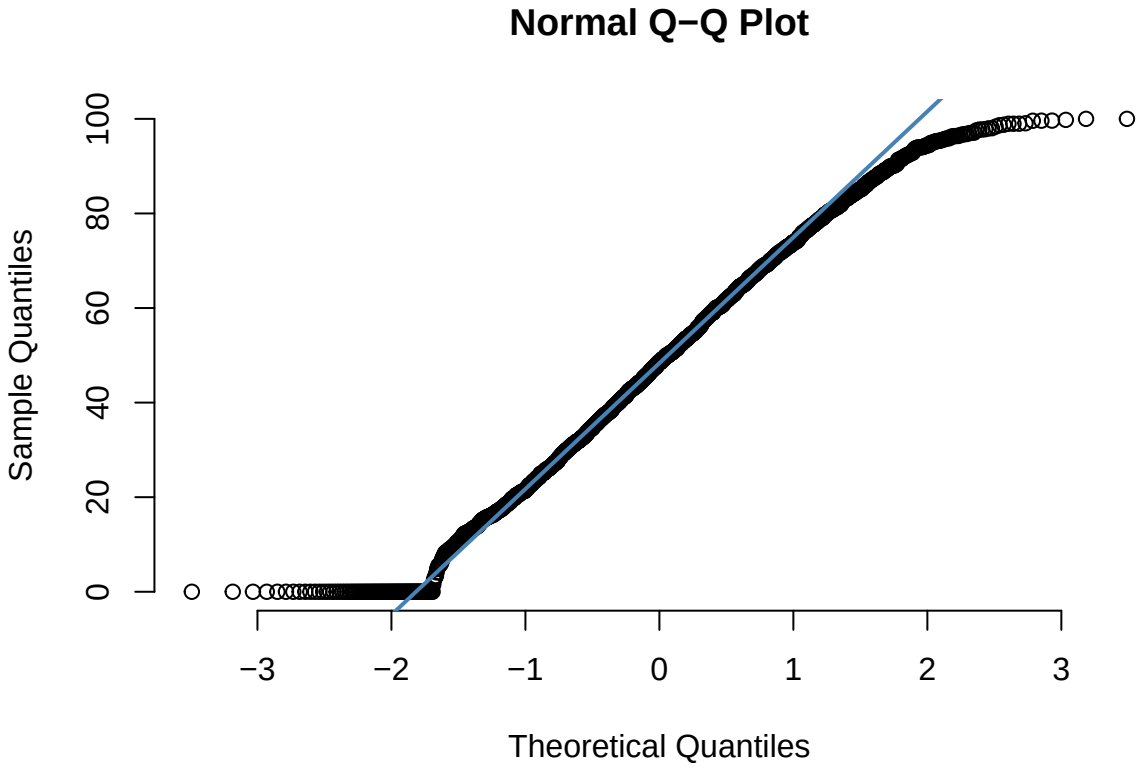
```
## Warning: Expecting numeric in Y2255 / R2255C25: got 'Non-Compliant'

## Warning: Expecting numeric in Z2255 / R2255C26: got 'Non-Compliant'

## Warning: Expecting numeric in AA2255 / R2255C27: got 'Non-Compliant'

## Warning: Expecting logical in AB2255 / R2255C28: got 'Note: This school did not
## report financial data for FY21.'

## New names:
## * `` -> `...28`
```

```r
School_Expenditure = select(School_Expenditure, schoolname, amount, school_ppe_21)
Poverty.Percentage = read_excel("2021_directly_certified_school.xls")
Poverty.Percentage = select(Poverty.Percentage, SCHOOL_NAME, direct_cert_perc)
Mobility = read_excel("2021_School_Mobility.xls")
Mobility = select(Mobility, school_name, mobility)
Enrollment = read.csv("Enrollment_by_Subgroups_Programs.csv", header=TRUE)
Enrollment = select(Enrollment, INSTN_NAME, ENROLL_PCT_GIFTED)

data = merge(Achievement_Rate, Percentage, by="School.Name")
data = merge(data, Salaries, by="School.Name")
data = merge(data, Absent_Rate, by="School.Name")

data = left_join(
          data %>% group_by(School.Name) %>% mutate(id = row_number()),
          School_Expenditure %>% group_by(schoolname) %>% mutate(id = row_number()),
          by = c("School.Name" = "schoolname", "id"))
```

```r
data = left_join(
          data %>% group_by(School.Name) %>% mutate(id = row_number()),
          Poverty.Percentage %>% group_by(SCHOOL_NAME) %>% mutate(id = row_number()),
          by = c("School.Name" = "SCHOOL_NAME", "id"))

data = left_join(
          data %>% group_by(School.Name) %>% mutate(id = row_number()),
          Mobility %>% group_by(school_name) %>% mutate(id = row_number()),
          by = c("School.Name" = "school_name", "id"))

data = left_join(
          data %>% group_by(School.Name) %>% mutate(id = row_number()),
          Enrollment %>% group_by(INSTN_NAME) %>% mutate(id = row_number()),
          by = c("School.Name" = "INSTN_NAME", "id"))
attach(data)

# Creating a Dummy Variable for Urban/Rural
data$u.r_dummy <- data$Urban.Rural
data$u.r_dummy <- as.character(data$u.r_dummy)
data$u.r_dummy[data$u.r_dummy == "Urban"] <- 1
data$u.r_dummy[data$u.r_dummy == "Rural"] <- 0
data$u.r_dummy <- as.numeric(data$u.r_dummy)
data$growth.rate.math <-data$X19.21.Difference.in.Math
```

a. Data Summary





```
##                2.5 %  97.5 %
## (Intercept) 65.23627 66.9811
```

```
##                2.5 %   97.5 %
## (Intercept) 46.99179 49.09119
```

We are 95% confident that the mean student achivement rate in Math in 2021 is higher than that in 2019.

```
##
## Attaching package: 'huxtable'
```

```
## The following object is masked from 'package:dplyr':
##
##     add_rownames
```

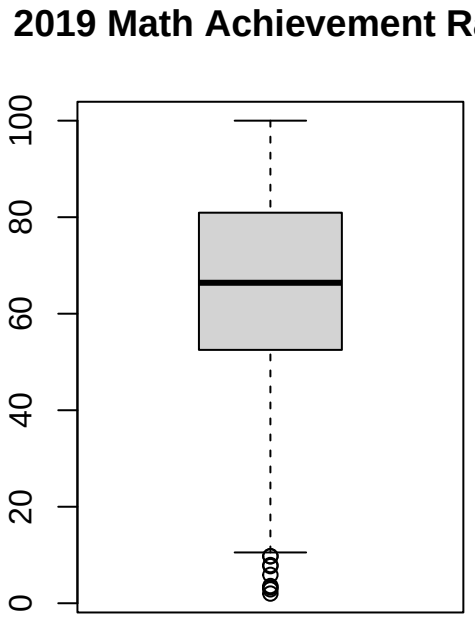
	2019	**2021**
Observations	2067.00	2067.00
Avg. Math achievement	66.1086840832124	48.0414900822448
Median Math achievement	66.41	48.47
Lower Bound of Math achievement	2.01	0
Upper Bound of Math achievement	100	100
Standard Deviation	20.2251098070928	24.3350230861084

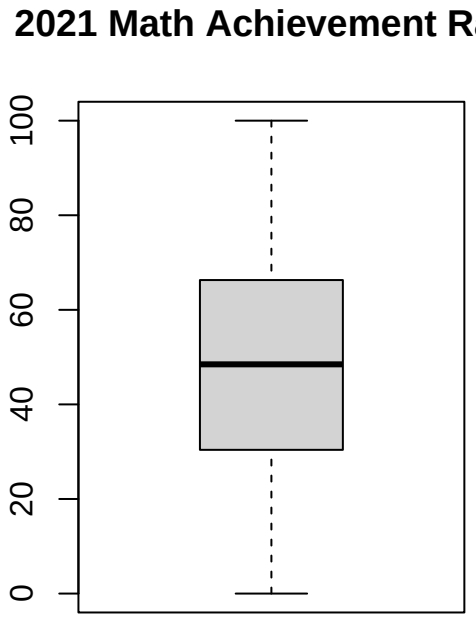
Mean and median Math test achievement rates are higher in 2019 than in 2021.

```
#average change in achievement rate
(52.23121-67.99686)/67.99686
```

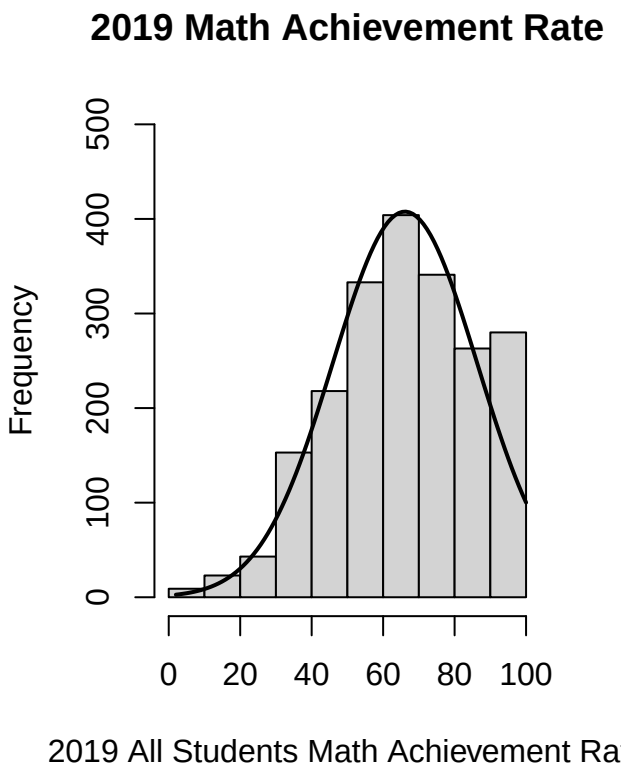
```
## [1] -0.2318585
```

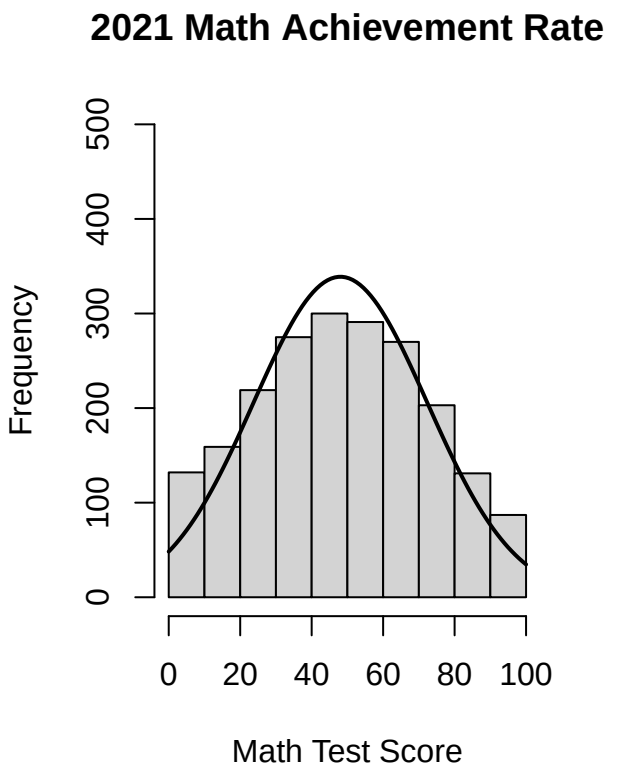
c. Data Visulization





The boxplot of both years' math achievement rate shows that in 2019, the data distribution is more compact, and all quartiles are significantly higher than those in 2021. A tremendous number of outliers are identified in both year's boxplots, suggesting many data points below the lower quartile by more than 1.5 interquartile range (IQR). Achievement rates are highly left skewed.





From both years' histogram, it can be confirmed that there is a very low frequency of math achievement rate between 0-30 for the 2019 data, as compared to the 2021 data. More outliers in the 2019 data could mean a higher . From plain sight, the 2019 data is better approximated by a normal distribution. The 2021 data seems skewed to the center.

d. Table of Variables

Variables	Description	Type
y1	2019 All Students Math Achievement Rate	Quantative
y2	2021 All Students Math Achievement Rate	Quantative
x1	Absent 0-5 Days Percentage	Quantative
x2	Absent 6-15 Days Percentage	Quantative
x3	Absent 15+ Days Percentage	Quantative
x4	Avg. Annual Salaries - Administrators	Quantative
x5	Avg. Annual Salaries - Teachers	Quantative
x6	Avg. Annual Salaries - Support.Personnel	Quantative
x7	Number of Teachers with a phd degree	Quantative
x8	Total Number of Certified Teachers	Quantative
x9	Post Grad Percentage	Quantative
x10	Total Students Enrolled	Quantative
x11	Teacher-Student Ratio	Quantative
x12	White Student Percentage	Quantative
x13	Black Student Percentage	Quantative
x14	Economically Disadvantaged Student Percentage	Quantative
x15	Directly Certified Students Percentage	Quantative
x16	Amount of Money Invested for Students	Quantative
x17	Per-Pupil Expenditure at School Level	Quantative
x18	Rate of Entries and Withdrawls to a School	Quantative
x19	Percentage of Gifted Students	Quantative
x20	Urban/Rural Area of the School	Qualitative

5. Regression Analysis

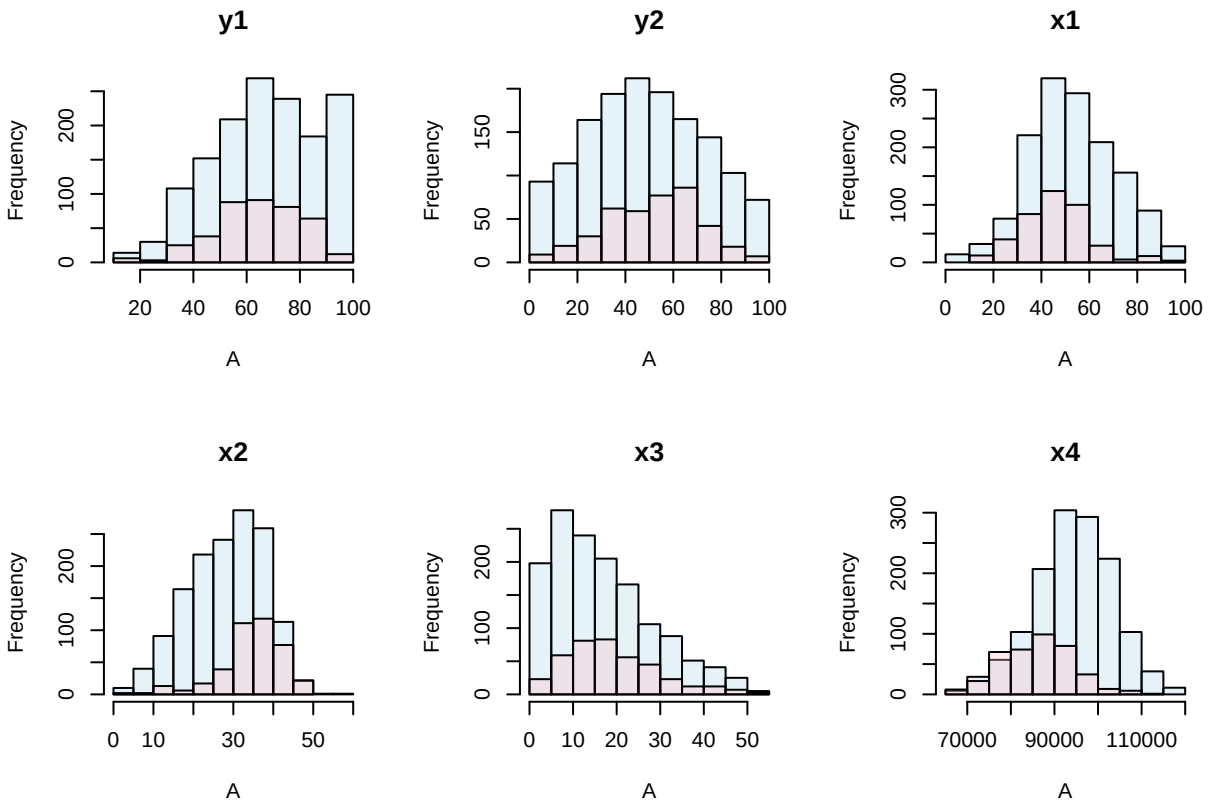
a. Iterations of the analysis process

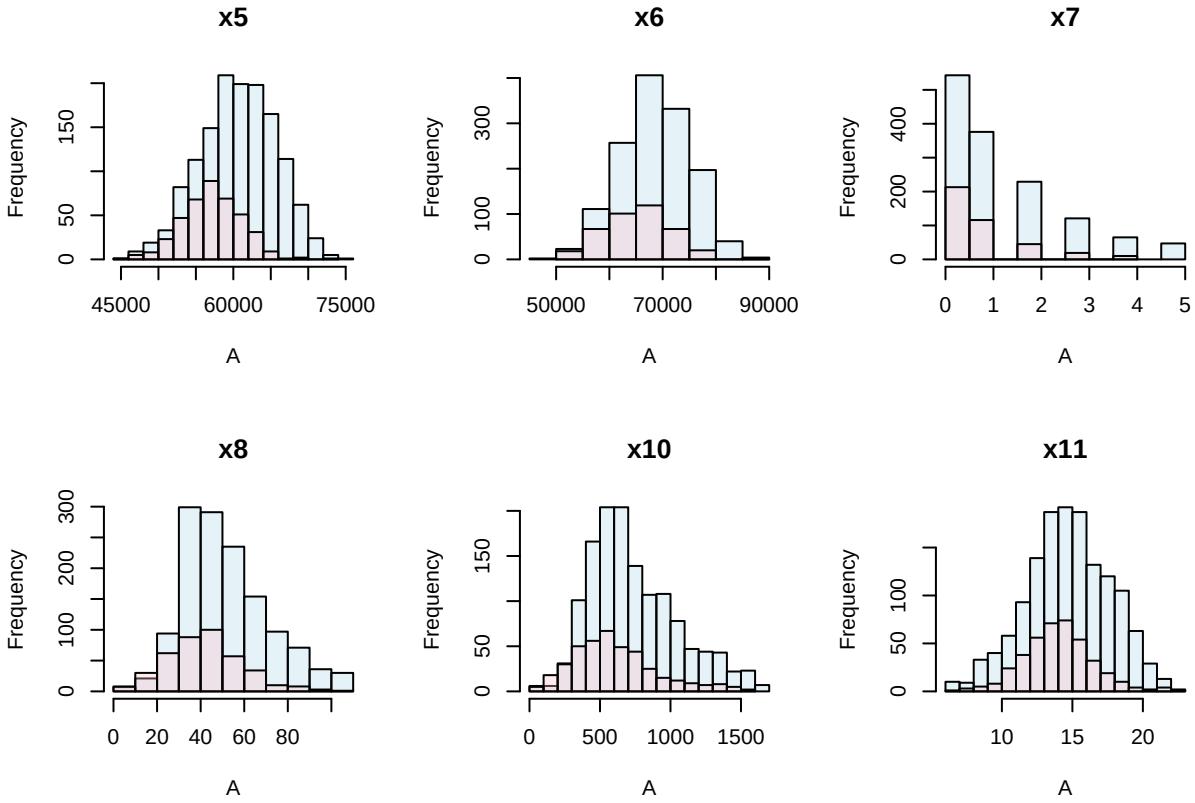
- paragraph description

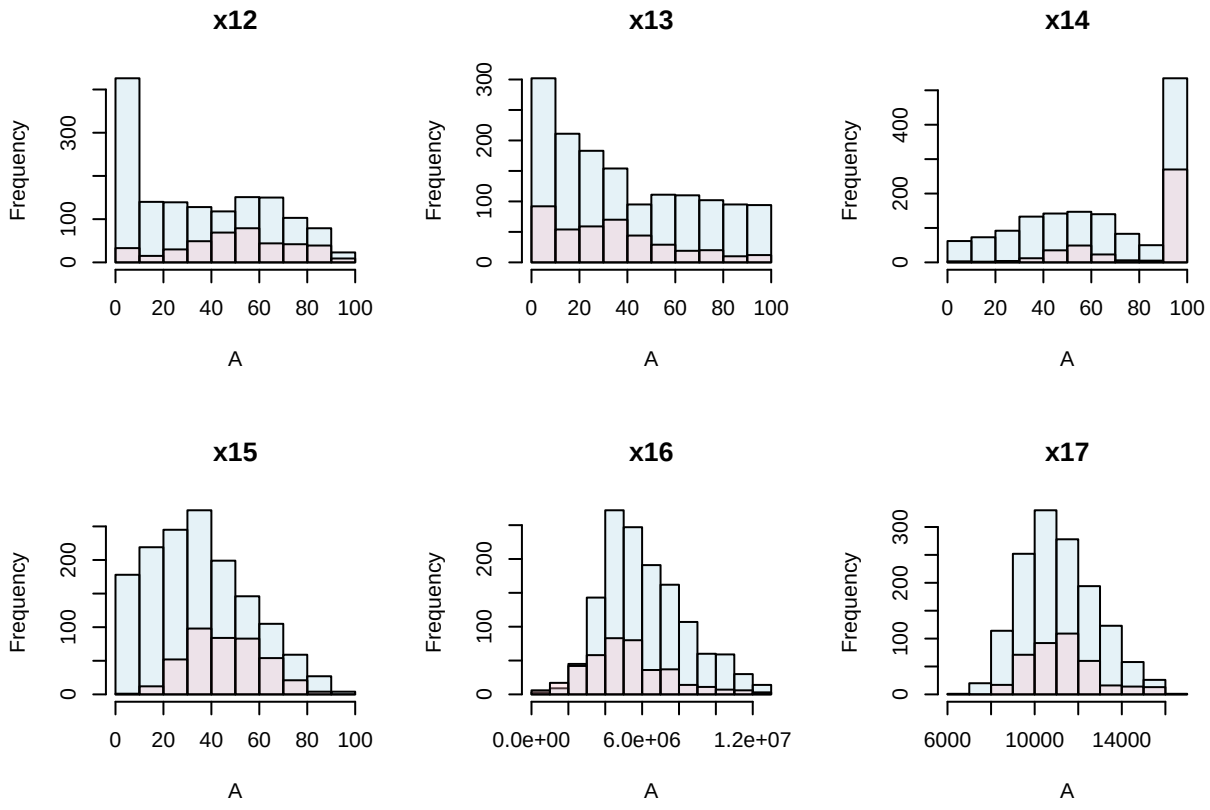
c. Plots of variables- Scatterplot

For the plots below, a light blue color indicates Urban Area and a light pink color indicates Rural Area.

```
## Warning in hist.default(A, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of

## Warning in hist.default(B, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of

## Warning in hist.default(A, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of

## Warning in hist.default(B, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of

## Warning in hist.default(A, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of

## Warning in hist.default(B, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of

## Warning in hist.default(A, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of

## Warning in hist.default(B, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of

## Warning in hist.default(A, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of

## Warning in hist.default(B, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of

## Warning in hist.default(A, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of

## Warning in hist.default(B, breaks = 12, plot = FALSE, na.rm = TRUE): argument
## '...' is not made use of
```

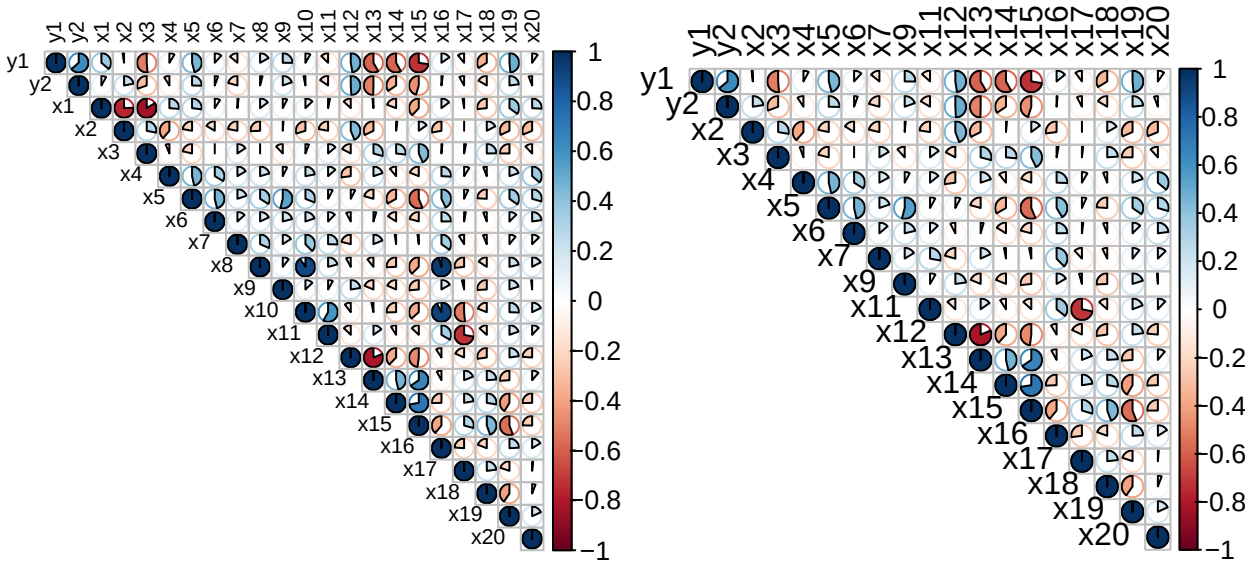

y1
Frequency
0
100
200
20
40
60
80
100
A
y2
Frequency
0
50
150
0
20
40
60
80
100
A
x1
Frequency
0
100
200
300
0
20
40
60
80
100
A
x2
Frequency
0
100
200
0
10
30
50
A
x3
Frequency
0
100
200
0
10
20
30
40
50
A
x4
Frequency
0
100
200
300
70000
90000
110000
A


x5
Frequency
0
50
150
45000
60000
75000
A
x6
Frequency
0
100
300
50000
70000
90000
A
x7
Frequency
0
200
400
0
1
2
3
4
5
A
x8
Frequency
0
100
200
300
0
20
40
60
80
A
x10
Frequency
0
50
150
0
500
1000
1500
A
x11
Frequency
0
50
100
10
15
20
A



b. Multicollinearity

```
## corrplot 0.92 loaded
```



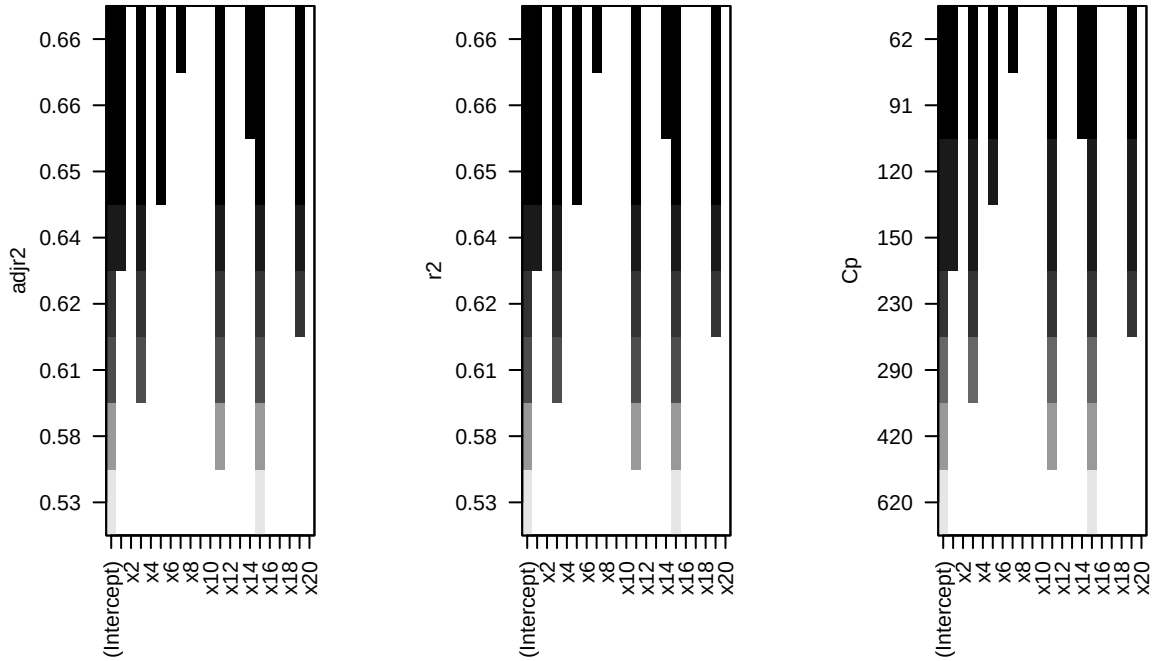

```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'ggplot2'

## The following object is masked from 'package:huxtable':
##
##     theme_grey

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:huxtable':
##
##     contents, label, label<-

## The following objects are masked from 'package:dplyr':
##
##     src, summarize
```

```
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

Before doing the model selection process, a Multicollinearity check produces high correlation of (x1:x3), (x8:x10,x16), (x10:x16), (x12: x13), (x11: x17), (x15: x13, x14). And another set of variables that have a high correlation is y1 and y2, since we are modeling them separately as response variables, we do not need to drop any of them. The renewed plot is on the right.

d. Model Selection

2019 Model Selection




```
## [1] "(Intercept)" "x1"          "x3"          "x5"          "x7"
## [6] "x11"         "x14"         "x15"         "x19"
```

```
##
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:datasets':
##
##     rivers
```

```
## Forward Selection Method
## --------------------------
##
## Candidate Terms:
##
## 1. x1
## 2. x2
## 3. x3
## 4. x4
## 5. x5
## 6. x6
## 7. x7
## 8. x8
## 9. x9
## 10. x10
## 11. x11
## 12. x12
## 13. x13
## 14. x14
## 15. x15
## 16. x16
## 17. x17
## 18. x18
## 19. x19
## 20. x20
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - x15
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.702       RMSE                14.159
## R-Squared               0.492       Coef. Var           21.333
## Adj. R-Squared          0.492       MSE                200.489
## Pred R-Squared          0.492       MAE                 10.974
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F          Sig.
## --------------------------------------------------------------------------
## Regression    400015.583           1     400015.583    1995.198    0.0000
## Residual      412205.640        2056        200.489
## Total         812221.223        2057
## --------------------------------------------------------------------------
##
```

```
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    91.915         0.651                141.087    0.000    90.637    93.193
##         x15    -0.671         0.015       -0.702   -44.668    0.000    -0.701    -0.642
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 2
##
## - x3
##
##                          Model Summary
## ---------------------------------------------------------------
## R                       0.728       RMSE                13.240
## R-Squared               0.531       Coef. Var           19.638
## Adj. R-Squared          0.530       MSE                175.304
## Pred R-Squared          0.529       MAE                 10.317
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                 Squares          DF    Mean Square       F          Sig.
## --------------------------------------------------------------------------
## Regression    390660.910           2     195330.455    1114.238    0.0000
## Residual      345524.186        1971        175.304
## Total         736185.095        1973
## --------------------------------------------------------------------------
##
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    95.902         0.677                141.639    0.000    94.574    97.230
##         x15    -0.576         0.016       -0.599   -35.642    0.000    -0.608    -0.545
##          x3    -0.406         0.028       -0.241   -14.379    0.000    -0.461    -0.351
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 3
##
## - x12
##
##                          Model Summary
## ---------------------------------------------------------------
## R                       0.755       RMSE                12.674
## R-Squared               0.570       Coef. Var           18.798
```

```
## Adj. R-Squared          0.569       MSE                160.639
## Pred R-Squared          0.568       MAE                  9.909
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                 Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression    419726.867           3     139908.956    870.954    0.0000
## Residual      316458.229        1970        160.639
## Total         736185.095        1973
## --------------------------------------------------------------------------
##
##                                Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    85.429         1.013                84.326    0.000    83.442    87.415
##         x15    -0.465         0.018       -0.483   -26.527    0.000    -0.500    -0.431
##          x3    -0.398         0.027       -0.236   -14.703    0.000    -0.451    -0.345
##         x12     0.161         0.012        0.231    13.451    0.000     0.137     0.184
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 4
##
## - x11
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.767       RMSE               12.325
## R-Squared               0.588       Coef. Var          18.216
## Adj. R-Squared          0.587       MSE               151.893
## Pred R-Squared          0.586       MAE                 9.625
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                 Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression    417367.108           4     104341.777    686.941    0.0000
## Residual      292090.985        1923        151.893
## Total         709458.092        1927
## --------------------------------------------------------------------------
##
```

```
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta       t        Sig      lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)    100.655         1.956                51.457    0.000    96.819    104.491
##         x15     -0.524         0.018       -0.546   -28.651    0.000    -0.560     -0.488
##          x3     -0.351         0.027       -0.210   -12.909    0.000    -0.404     -0.297
##         x12      0.133         0.012        0.192    11.138    0.000     0.110      0.157
##         x11     -0.874         0.103       -0.131    -8.495    0.000    -1.076     -0.672
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 5
##
## - x19
##
##                          Model Summary
## ------------------------------------------------------------------
## R                       0.786       RMSE                11.518
## R-Squared               0.618       Coef. Var           16.800
## Adj. R-Squared          0.617       MSE                132.675
## Pred R-Squared          0.615       MAE                  9.097
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression    395927.440           5      79185.488    596.838    0.0000
## Residual      244785.347        1845        132.675
## Total         640712.787        1850
## --------------------------------------------------------------------------
##
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower      upper
## ----------------------------------------------------------------------------------------
## (Intercept)    96.339         1.989                48.436    0.000    92.438    100.240
##         x15    -0.445         0.020       -0.465   -21.943    0.000    -0.484     -0.405
##          x3    -0.342         0.026       -0.210   -13.009    0.000    -0.394     -0.291
##         x12     0.130         0.011        0.193    11.496    0.000     0.108      0.153
##         x11    -1.016         0.100       -0.156   -10.153    0.000    -1.212     -0.820
##         x19     0.357         0.039        0.164     9.246    0.000     0.281      0.433
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 6
##
```

```
## - x5
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.805       RMSE                11.052
## R-Squared               0.648       Coef. Var           16.111
## Adj. R-Squared          0.647       MSE                122.152
## Pred R-Squared          0.645       MAE                  8.725
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression    402886.927           6      67147.821    549.706    0.0000
## Residual      219141.011        1794        122.152
## Total         622027.938        1800
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    66.492         4.895                 13.583    0.000    56.890    76.093
##         x15    -0.402         0.023       -0.417    -17.266    0.000    -0.448    -0.356
##          x3    -0.340         0.026       -0.209    -13.019    0.000    -0.391    -0.289
##         x12     0.133         0.012        0.198     11.545    0.000     0.111     0.156
##         x11    -1.123         0.104       -0.165    -10.820    0.000    -1.327    -0.919
##         x19     0.328         0.037        0.151      8.752    0.000     0.255     0.402
##          x5     0.001         0.000        0.134      7.770    0.000     0.000     0.001
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 7
##
## - x2
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.810       RMSE                10.881
## R-Squared               0.657       Coef. Var           15.846
## Adj. R-Squared          0.655       MSE                118.388
## Pred R-Squared          0.653       MAE                  8.606
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
```

```
## ----------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F         Sig.
## ----------------------------------------------------------------------
## Regression    404005.684           7      57715.098    487.506    0.0000
## Residual      211086.586        1783        118.388
## Total         615092.270        1790
## ----------------------------------------------------------------------
##
##                                  Parameter Estimates
## ------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ------------------------------------------------------------------------------------------
## (Intercept)    59.192         4.945                  11.969    0.000    49.492    68.891
##         x15    -0.416         0.023       -0.432    -17.854    0.000    -0.462    -0.371
##          x3    -0.399         0.027       -0.246    -14.802    0.000    -0.452    -0.347
##         x12     0.077         0.014        0.115      5.617    0.000     0.050     0.104
##         x11    -1.064         0.103       -0.157    -10.312    0.000    -1.266    -0.861
##         x19     0.435         0.040        0.194     10.784    0.000     0.356     0.515
##          x5     0.001         0.000        0.140      8.149    0.000     0.000     0.001
##          x2     0.267         0.037        0.134      7.125    0.000     0.194     0.341
## ------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 8
##
## - x14
##
##                           Model Summary
## ----------------------------------------------------------------
## R                        0.816       RMSE                10.748
## R-Squared                0.665       Coef. Var           15.653
## Adj. R-Squared           0.664       MSE                115.521
## Pred R-Squared           0.662       MAE                  8.495
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## ----------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F         Sig.
## ----------------------------------------------------------------------
## Regression    409234.027           8      51154.253    442.814    0.0000
## Residual      205858.243        1782        115.521
## Total         615092.270        1790
## ----------------------------------------------------------------------
##
##                                  Parameter Estimates
## ------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ------------------------------------------------------------------------------------------
```

```
## (Intercept)    58.770         4.885                12.030    0.000    49.188    68.351
##         x15    -0.314         0.028        -0.326  -11.404    0.000    -0.369    -0.260
##          x3    -0.413         0.027        -0.254  -15.446    0.000    -0.465    -0.361
##         x12     0.079         0.014         0.117    5.802    0.000     0.052     0.106
##         x11    -1.042         0.102        -0.153  -10.216    0.000    -1.242    -0.842
##         x19     0.421         0.040         0.188   10.546    0.000     0.343     0.500
##          x5     0.001         0.000         0.151    8.892    0.000     0.000     0.001
##          x2     0.253         0.037         0.126    6.814    0.000     0.180     0.325
##         x14    -0.080         0.012        -0.134   -6.727    0.000    -0.103    -0.057
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 9
##
## - x7
##
##                          Model Summary
## ------------------------------------------------------------------
## R                       0.818       RMSE                10.661
## R-Squared               0.669       Coef. Var           15.534
## Adj. R-Squared          0.667       MSE                113.655
## Pred R-Squared          0.665       MAE                  8.396
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression    392891.404          9      43654.600    384.099    0.0000
## Residual      194463.162        1711        113.655
## Total         587354.566        1720
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    52.786         4.996                10.566    0.000    42.988    62.585
##         x15    -0.307         0.028        -0.320  -11.029    0.000    -0.362    -0.253
##          x3    -0.375         0.028        -0.230  -13.533    0.000    -0.429    -0.321
##         x12     0.066         0.014         0.099    4.831    0.000     0.039     0.093
##         x11    -0.889         0.106        -0.129   -8.406    0.000    -1.096    -0.681
##         x19     0.437         0.042         0.186   10.384    0.000     0.355     0.520
##          x5     0.001         0.000         0.178   10.214    0.000     0.001     0.001
##          x2     0.221         0.037         0.110    5.886    0.000     0.147     0.294
##         x14    -0.086         0.012        -0.144   -7.047    0.000    -0.110    -0.062
##          x7    -1.383         0.227        -0.096   -6.079    0.000    -1.829    -0.937
## ----------------------------------------------------------------------------------------
##
```

```
##
##
## Forward Selection: Step 10
##
## - x17
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.819       RMSE                10.552
## R-Squared               0.671       Coef. Var           15.294
## Adj. R-Squared          0.669       MSE                111.348
## Pred R-Squared          0.666       MAE                  8.320
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F          Sig.
## --------------------------------------------------------------------------
## Regression    366571.180          10      36657.118    329.213    0.0000
## Residual      179603.621        1613        111.348
## Total         546174.801        1623
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    73.054         6.198                 11.786    0.000    60.896    85.211
##         x15    -0.284         0.030       -0.291     -9.409    0.000    -0.343    -0.225
##          x3    -0.355         0.029       -0.213    -12.209    0.000    -0.412    -0.298
##         x12     0.056         0.014        0.084      3.952    0.000     0.028     0.084
##         x11    -1.632         0.174       -0.226     -9.358    0.000    -1.974    -1.290
##         x19     0.443         0.043        0.192     10.381    0.000     0.359     0.527
##          x5     0.001         0.000        0.207     10.541    0.000     0.001     0.001
##          x2     0.181         0.040        0.090      4.534    0.000     0.103     0.260
##         x14    -0.084         0.013       -0.141     -6.597    0.000    -0.109    -0.059
##          x7    -1.332         0.232       -0.092     -5.735    0.000    -1.787    -0.876
##         x17    -0.001         0.000       -0.125     -5.156    0.000    -0.002    -0.001
## ----------------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 11
##
## - x20
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.825       RMSE                10.440
## R-Squared               0.681       Coef. Var           15.138
```

```
## Adj. R-Squared          0.678       MSE                108.988
## Pred R-Squared          0.675       MAE                  8.199
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                 Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression    353940.990          11      32176.454    295.231    0.0000
## Residual      166097.010        1524        108.988
## Total         520038.000        1535
## --------------------------------------------------------------------------
##
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    78.226         6.525                11.989    0.000    65.428    91.024
##         x15    -0.298         0.032       -0.303     -9.460    0.000    -0.360    -0.236
##          x3    -0.375         0.030       -0.224    -12.638    0.000    -0.433    -0.316
##         x12     0.037         0.015        0.055      2.414    0.016     0.007     0.067
##         x11    -1.721         0.180       -0.237     -9.532    0.000    -2.075    -1.367
##         x19     0.460         0.044        0.197     10.511    0.000     0.374     0.546
##          x5     0.001         0.000        0.203     10.062    0.000     0.001     0.001
##          x2     0.189         0.041        0.094      4.615    0.000     0.109     0.269
##         x14    -0.093         0.013       -0.154     -7.077    0.000    -0.119    -0.067
##          x7    -1.342         0.237       -0.093     -5.663    0.000    -1.806    -0.877
##         x17    -0.001         0.000       -0.125     -5.003    0.000    -0.002    -0.001
##         x20    -2.431         0.734       -0.056     -3.313    0.001    -3.870    -0.991
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 12
##
## - x13
##
##                           Model Summary
## ----------------------------------------------------------------
## R                       0.825       RMSE               10.430
## R-Squared               0.681       Coef. Var          15.125
## Adj. R-Squared          0.679       MSE               108.789
## Pred R-Squared          0.675       MAE                 8.203
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
```

```
##                    Sum of
##                   Squares          DF    Mean Square       F         Sig.
## ------------------------------------------------------------------------
## Regression    354352.532          12      29529.378    271.437    0.0000
## Residual      165685.468        1523        108.789
## Total         520038.000        1535
## ------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    77.676         6.525                  11.905    0.000    64.877    90.474
##         x15    -0.265         0.036       -0.270     -7.389    0.000    -0.335    -0.194
##          x3    -0.372         0.030       -0.222    -12.533    0.000    -0.430    -0.314
##         x12     0.016         0.019        0.024      0.866    0.386    -0.020     0.052
##         x11    -1.651         0.184       -0.228     -8.985    0.000    -2.012    -1.291
##         x19     0.467         0.044        0.200     10.635    0.000     0.380     0.553
##          x5     0.001         0.000        0.205     10.173    0.000     0.001     0.001
##          x2     0.174         0.042        0.087      4.176    0.000     0.092     0.256
##         x14    -0.096         0.013       -0.160     -7.269    0.000    -0.122    -0.070
##          x7    -1.337         0.237       -0.092     -5.647    0.000    -1.801    -0.873
##         x17    -0.001         0.000       -0.121     -4.806    0.000    -0.002    -0.001
##         x20    -2.348         0.734       -0.054     -3.198    0.001    -3.788    -0.908
##         x13    -0.042         0.022       -0.062     -1.945    0.052    -0.085     0.000
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 13
##
## - x10
##
##                           Model Summary
## ------------------------------------------------------------------
## R                       0.824       RMSE                10.452
## R-Squared               0.680       Coef. Var           15.208
## Adj. R-Squared          0.677       MSE                109.250
## Pred R-Squared          0.672       MAE                  8.204
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## ------------------------------------------------------------------------
##                    Sum of
##                   Squares          DF    Mean Square       F         Sig.
## ------------------------------------------------------------------------
## Regression    336549.123          13      25888.394    236.965    0.0000
## Residual      158739.931        1453        109.250
## Total         495289.054        1466
## ------------------------------------------------------------------------
##
```

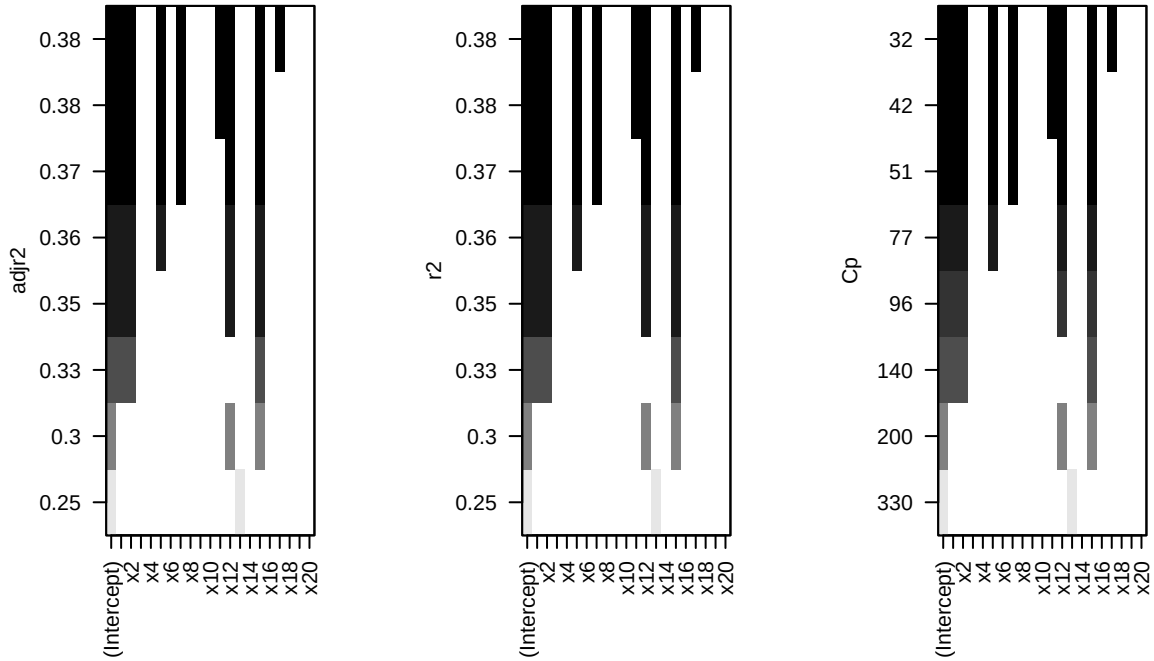
```
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    80.325         6.805                 11.805    0.000    66.978    93.673
##         x15    -0.272         0.037       -0.275     -7.400    0.000    -0.345    -0.200
##          x3    -0.368         0.031       -0.220    -11.806    0.000    -0.429    -0.307
##         x12     0.013         0.020        0.020      0.681    0.496    -0.025     0.052
##         x11    -1.601         0.192       -0.211     -8.338    0.000    -1.978    -1.225
##         x19     0.430         0.046        0.180      9.385    0.000     0.340     0.520
##          x5     0.001         0.000        0.212     10.179    0.000     0.001     0.001
##          x2     0.146         0.043        0.073      3.403    0.001     0.062     0.230
##         x14    -0.096         0.014       -0.157     -7.100    0.000    -0.123    -0.070
##          x7    -1.311         0.266       -0.083     -4.921    0.000    -1.834    -0.789
##         x17    -0.002         0.000       -0.133     -5.085    0.000    -0.002    -0.001
##         x20    -2.355         0.742       -0.055     -3.175    0.002    -3.810    -0.900
##         x13    -0.042         0.022       -0.062     -1.893    0.059    -0.086     0.002
##         x10    -0.002         0.001       -0.040     -1.802    0.072    -0.005     0.000
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 14
##
## - x8
##
##                           Model Summary
## ------------------------------------------------------------------
## R                       0.825       RMSE                10.446
## R-Squared               0.680       Coef. Var           15.197
## Adj. R-Squared          0.677       MSE                109.124
## Pred R-Squared          0.673       MAE                  8.193
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                   ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression    336259.392          14      24018.528    220.104    0.0000
## Residual      158011.243        1448        109.124
## Total         494270.634        1462
## --------------------------------------------------------------------------
##
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    69.044         8.666                  7.967    0.000    52.045    86.043
##         x15    -0.272         0.037       -0.275     -7.391    0.000    -0.344    -0.200
##          x3    -0.367         0.031       -0.219    -11.750    0.000    -0.428    -0.305
```

```
##         x12     0.019         0.020        0.029      0.985    0.325    -0.019     0.058
##         x11    -0.861         0.395       -0.113     -2.180    0.029    -1.635    -0.086
##         x19     0.430         0.046        0.180      9.384    0.000     0.340     0.520
##          x5     0.001         0.000        0.210     10.076    0.000     0.001     0.001
##          x2     0.148         0.043        0.074      3.450    0.001     0.064     0.233
##         x14    -0.096         0.014       -0.157     -7.108    0.000    -0.123    -0.070
##          x7    -1.260         0.268       -0.079     -4.707    0.000    -1.785    -0.735
##         x17    -0.002         0.000       -0.133     -5.050    0.000    -0.002    -0.001
##         x20    -2.467         0.744       -0.057     -3.316    0.001    -3.927    -1.008
##         x13    -0.038         0.022       -0.056     -1.705    0.088    -0.082     0.006
##         x10    -0.018         0.007       -0.292     -2.559    0.011    -0.032    -0.004
##          x8     0.238         0.105        0.219      2.279    0.023     0.033     0.444
## ----------------------------------------------------------------------------------------
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + x15
## + x3
## + x12
## + x11
## + x19
## + x5
## + x2
## + x14
## + x7
## + x17
## + x20
## + x13
## + x10
## + x8
##
##
## Final Model Output
## ------------------
##
##                            Model Summary
## ----------------------------------------------------------------
## R                       0.825       RMSE                10.446
## R-Squared               0.680       Coef. Var           15.197
## Adj. R-Squared          0.677       MSE                109.124
## Pred R-Squared          0.673       MAE                  8.193
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F         Sig.
```

```
## ------------------------------------------------------------------------
## Regression    336259.392          14      24018.528    220.104    0.0000
## Residual      158011.243        1448        109.124
## Total         494270.634        1462
## ------------------------------------------------------------------------
##
##                                   Parameter Estimates
## ---------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ---------------------------------------------------------------------------------------
## (Intercept)    69.044         8.666                   7.967    0.000    52.045    86.043
##         x15    -0.272         0.037       -0.275     -7.391    0.000    -0.344    -0.200
##          x3    -0.367         0.031       -0.219    -11.750    0.000    -0.428    -0.305
##         x12     0.019         0.020        0.029      0.985    0.325    -0.019     0.058
##         x11    -0.861         0.395       -0.113     -2.180    0.029    -1.635    -0.086
##         x19     0.430         0.046        0.180      9.384    0.000     0.340     0.520
##          x5     0.001         0.000        0.210     10.076    0.000     0.001     0.001
##          x2     0.148         0.043        0.074      3.450    0.001     0.064     0.233
##         x14    -0.096         0.014       -0.157     -7.108    0.000    -0.123    -0.070
##          x7    -1.260         0.268       -0.079     -4.707    0.000    -1.785    -0.735
##         x17    -0.002         0.000       -0.133     -5.050    0.000    -0.002    -0.001
##         x20    -2.467         0.744       -0.057     -3.316    0.001    -3.927    -1.008
##         x13    -0.038         0.022       -0.056     -1.705    0.088    -0.082     0.006
##         x10    -0.018         0.007       -0.292     -2.559    0.011    -0.032    -0.004
##          x8     0.238         0.105        0.219      2.279    0.023     0.033     0.444
## ---------------------------------------------------------------------------------------
```

```
##
##                                Selection Summary
## ------------------------------------------------------------------------------
##         Variable                  Adj.
## Step    Entered     R-Square    R-Square      C(p)          AIC         RMSE
## ------------------------------------------------------------------------------
##    1    x15           0.4925      0.4922    1740.7492    16753.3141    14.1594
##    2    x3            0.5307      0.5302    1212.8823    15805.6807    13.2402
##    3    x12           0.5701      0.5695     947.3022    15634.2226    12.6743
##    4    x11           0.5883      0.5874     770.9783    15163.1094    12.3245
##    5    x19           0.6179      0.6169     414.4844    14308.4081    11.5185
##    6    x5            0.6477      0.6465     230.4037    13774.2900    11.0522
##    7    x2            0.6568      0.6555     168.2550    13642.8021    10.8806
##    8    x14           0.6653      0.6638     122.1230    13599.8827    10.7481
##    9    x7            0.6689      0.6672      89.2204    13041.7338    10.6609
##   10    x17           0.6712      0.6691      51.4240    12275.0291    10.5521
##   11    x20           0.6806      0.6783      17.0827    11578.6668    10.4397
##   12    x13           0.6814      0.6789      15.2940    11576.8563    10.4302
##   13    x10           0.6795      0.6766      22.3537    11064.6636    10.4523
##   14    x8            0.6803      0.6772      21.6454    11033.8392    10.4462
## ------------------------------------------------------------------------------
```

2021 Model Selection




```
##   (Intercept)    x1    x2    x3    x4    x5    x6    x7    x8    x9   x10   x11
## 1        TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 2        TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 3        TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 4        TRUE  TRUE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 5        TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE
## 6        TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE FALSE
## 7        TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE
## 8        TRUE  TRUE  TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE  TRUE
##     x12   x13   x14   x15   x16   x17   x18   x19   x20
## 1 FALSE  TRUE FALSE FALSE FALSE FALSE FALSE FALSE FALSE
## 2  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
## 3 FALSE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
## 4  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
## 5  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
## 6  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
## 7  TRUE FALSE FALSE  TRUE FALSE FALSE FALSE FALSE FALSE
## 8  TRUE FALSE FALSE  TRUE FALSE  TRUE FALSE FALSE FALSE

## [1] "(Intercept)" "x1"          "x2"          "x5"          "x7"
## [6] "x11"         "x12"         "x15"         "x17"

## Forward Selection Method
```

```
## ---------------------------
##
## Candidate Terms:
##
## 1. x1
## 2. x2
## 3. x3
## 4. x4
## 5. x5
## 6. x6
## 7. x7
## 8. x8
## 9. x9
## 10. x10
## 11. x11
## 12. x12
## 13. x13
## 14. x14
## 15. x15
## 16. x16
## 17. x17
## 18. x18
## 19. x19
## 20. x20
##
## We are selecting variables based on p value...
##
##
## Forward Selection: Step 1
##
## - x13
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.553       RMSE                20.284
## R-Squared               0.306       Coef. Var           42.222
## Adj. R-Squared          0.305       MSE                411.437
## Pred R-Squared          0.304       MAE                 15.238
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                    Sum of
##                   Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression     373855.078           1     373855.078    908.658    0.0000
## Residual       849616.380        2065        411.437
## Total         1223471.458        2066
## --------------------------------------------------------------------------
##
##                                 Parameter Estimates
```

```
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    65.707         0.737                89.210    0.000    64.263    67.152
##         x13    -0.454         0.015       -0.553   -30.144    0.000    -0.484    -0.425
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 2
##
## - x12
##
##                          Model Summary
## ---------------------------------------------------------------
## R                       0.575       RMSE                19.927
## R-Squared               0.330       Coef. Var           41.478
## Adj. R-Squared          0.329       MSE                397.077
## Pred R-Squared          0.328       MAE                 14.814
## ---------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F          Sig.
## --------------------------------------------------------------------------
## Regression     403904.492           2     201952.246    508.597    0.0000
## Residual       819566.967        2064        397.077
## Total         1223471.458        2066
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ----------------------------------------------------------------------------------------
## (Intercept)    49.856         1.961                25.429    0.000    46.011    53.701
##         x13    -0.273         0.026       -0.333   -10.710    0.000    -0.323    -0.223
##         x12     0.235         0.027        0.270     8.699    0.000     0.182     0.288
## ----------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 3
##
## - x3
##
##                          Model Summary
## ---------------------------------------------------------------
## R                       0.584       RMSE                19.461
## R-Squared               0.341       Coef. Var           39.572
## Adj. R-Squared          0.340       MSE                378.746
```

```
## Pred R-Squared          0.338       MAE                  14.192
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                    Sum of
##                   Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression     387701.851           3     129233.950    341.216    0.0000
## Residual       748401.261        1976        378.746
## Total         1136103.113        1979
## --------------------------------------------------------------------------
##
##                                 Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta      t        Sig      lower     upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    53.699         1.975                27.186    0.000    49.825    57.573
##         x13    -0.200         0.027       -0.240    -7.519    0.000    -0.253    -0.148
##         x12     0.251         0.027        0.291     9.412    0.000     0.199     0.304
##          x3    -0.386         0.040       -0.186    -9.585    0.000    -0.465    -0.307
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 4
##
## - x1
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.599       RMSE                19.145
## R-Squared               0.358       Coef. Var           38.830
## Adj. R-Squared          0.357       MSE                366.516
## Pred R-Squared          0.355       MAE                 13.994
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                    Sum of
##                   Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression     401833.936           4     100458.484     274.09    0.0000
## Residual       719471.733        1963        366.516
## Total         1121305.669        1967
## --------------------------------------------------------------------------
##
##                                 Parameter Estimates
```

```
## ---------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower     upper
## ---------------------------------------------------------------------------------------
## (Intercept)    83.106         4.753                17.483    0.000    73.784    92.429
##         x13    -0.181         0.026       -0.217    -6.865    0.000    -0.233    -0.130
##         x12     0.201         0.027        0.234     7.348    0.000     0.148     0.255
##          x3    -0.884         0.081       -0.425   -10.894    0.000    -1.043    -0.725
##          x1    -0.366         0.054       -0.253    -6.755    0.000    -0.473    -0.260
## ---------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 5
##
## - x15
##
##                          Model Summary
## ------------------------------------------------------------------
## R                       0.620       RMSE                18.761
## R-Squared               0.384       Coef. Var           38.052
## Adj. R-Squared          0.383       MSE                351.969
## Pred R-Squared          0.380       MAE                 13.393
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression     430742.642          5      86148.528    244.762    0.0000
## Residual       690563.027        1962        351.969
## Total         1121305.669        1967
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ---------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta       t        Sig      lower      upper
## ---------------------------------------------------------------------------------------
## (Intercept)    103.343         5.166                20.005    0.000    93.212    113.473
##         x13     -0.046         0.030       -0.055    -1.539    0.124    -0.105      0.013
##         x12      0.184         0.027        0.213     6.819    0.000     0.131      0.236
##          x3     -1.053         0.082       -0.506   -12.894    0.000    -1.213     -0.893
##          x1     -0.573         0.058       -0.397    -9.913    0.000    -0.687     -0.460
##         x15     -0.289         0.032       -0.242    -9.063    0.000    -0.351     -0.226
## ---------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 6
##
## - x11
```

```
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.620       RMSE                18.660
## R-Squared               0.384       Coef. Var           37.677
## Adj. R-Squared          0.382       MSE                348.192
## Pred R-Squared          0.379       MAE                 13.221
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                    Sum of
##                   Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression     415745.187           6      69290.865    199.002    0.0000
## Residual       665743.731        1912        348.192
## Total         1081488.918        1918
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta       t        Sig       lower      upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    114.353         5.649                 20.244    0.000    103.275    125.431
##         x13     -0.006         0.031       -0.008     -0.204    0.838     -0.067      0.054
##         x12      0.182         0.027        0.213      6.717    0.000      0.129      0.236
##          x3     -1.028         0.084       -0.495    -12.289    0.000     -1.192     -0.864
##          x1     -0.588         0.059       -0.408     -9.968    0.000     -0.704     -0.472
##         x15     -0.359         0.035       -0.301    -10.393    0.000     -0.427     -0.291
##         x11     -0.648         0.160       -0.078     -4.056    0.000     -0.961     -0.334
## ----------------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 7
##
## - x19
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.607       RMSE                18.592
## R-Squared               0.369       Coef. Var           36.795
## Adj. R-Squared          0.366       MSE                345.680
## Pred R-Squared          0.361       MAE                 13.120
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
```

```
##                    Sum of
##                   Squares          DF    Mean Square       F         Sig.
## ------------------------------------------------------------------------
## Regression     370267.283           7      52895.326    153.018    0.0000
## Residual       633977.744        1834        345.680
## Total         1004245.027        1841
## ------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta       t        Sig       lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    112.604         5.836                19.294    0.000    101.158    124.050
##         x13     -0.012         0.032       -0.014     -0.362    0.717     -0.074      0.051
##         x12      0.172         0.028        0.203      6.161    0.000      0.118      0.227
##          x3     -1.043         0.088       -0.507    -11.851    0.000     -1.215     -0.870
##          x1     -0.609         0.063       -0.428     -9.675    0.000     -0.732     -0.485
##         x15     -0.288         0.039       -0.239     -7.416    0.000     -0.364     -0.212
##         x11     -0.740         0.165       -0.091     -4.499    0.000     -1.063     -0.418
##         x19      0.267         0.068        0.094      3.911    0.000      0.133      0.401
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 8
##
## - x7
##
##                          Model Summary
## ----------------------------------------------------------------
## R                        0.608       RMSE                18.653
## R-Squared                0.370       Coef. Var           36.649
## Adj. R-Squared           0.367       MSE                347.923
## Pred R-Squared           0.361       MAE                 13.045
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## --------------------------------------------------------------------------
##                    Sum of
##                   Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression     357879.989           8      44734.999    128.577    0.0000
## Residual       610604.445        1755        347.923
## Total          968484.433        1763
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta       t        Sig       lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    112.006         5.958                18.799    0.000    100.320    123.692
```

```
##         x13     -0.025         0.034       -0.030     -0.749    0.454     -0.092       0.041
##         x12      0.143         0.029        0.168      4.930    0.000      0.086       0.200
##          x3     -0.964         0.092       -0.465    -10.461    0.000     -1.145      -0.784
##          x1     -0.574         0.065       -0.401     -8.870    0.000     -0.701      -0.447
##         x15     -0.298         0.041       -0.246     -7.293    0.000     -0.378      -0.218
##         x11     -0.717         0.172       -0.086     -4.160    0.000     -1.055      -0.379
##         x19      0.340         0.073        0.114      4.666    0.000      0.197       0.483
##          x7     -1.295         0.375       -0.071     -3.454    0.001     -2.030      -0.560
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 9
##
## - x5
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.616       RMSE                18.484
## R-Squared               0.380       Coef. Var           36.431
## Adj. R-Squared          0.377       MSE                341.660
## Pred R-Squared          0.371       MAE                 12.815
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square       F         Sig.
## --------------------------------------------------------------------------
## Regression    358304.965           9      39811.663    116.524    0.0000
## Residual      584921.429        1712        341.660
## Total         943226.394        1721
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model      Beta    Std. Error    Std. Beta       t        Sig      lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    87.743         9.887                   8.874    0.000    68.351    107.135
##         x13     0.005         0.035        0.005      0.132    0.895    -0.064      0.073
##         x12     0.148         0.029        0.174      5.065    0.000     0.091      0.205
##          x3    -1.048         0.093       -0.508    -11.304    0.000    -1.230     -0.866
##          x1    -0.657         0.066       -0.457     -9.910    0.000    -0.787     -0.527
##         x15    -0.279         0.047       -0.229     -5.872    0.000    -0.372     -0.186
##         x11    -0.560         0.186       -0.064     -3.010    0.003    -0.926     -0.195
##         x19     0.335         0.073        0.112      4.579    0.000     0.192      0.479
##          x7    -1.805         0.395       -0.098     -4.568    0.000    -2.580     -1.030
##          x5     0.000         0.000        0.092      3.877    0.000     0.000      0.001
## ----------------------------------------------------------------------------------------------
##
##
```

```
##
## Forward Selection: Step 10
##
## - x17
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.614       RMSE                18.341
## R-Squared               0.377       Coef. Var           35.572
## Adj. R-Squared          0.373       MSE                336.403
## Pred R-Squared          0.367       MAE                 12.640
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                 ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square      F          Sig.
## --------------------------------------------------------------------------
## Regression    328616.612          10      32861.661    97.685    0.0000
## Residual      542955.129        1614        336.403
## Total         871571.741        1624
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig      lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    102.696        11.465                 8.958    0.000    80.209    125.183
##         x13      0.008         0.036        0.009     0.215    0.830    -0.063      0.079
##         x12      0.139         0.030        0.164     4.589    0.000     0.080      0.199
##          x3     -0.940         0.099       -0.446    -9.511    0.000    -1.134     -0.746
##          x1     -0.587         0.070       -0.408    -8.400    0.000    -0.724     -0.450
##         x15     -0.232         0.050       -0.188    -4.642    0.000    -0.329     -0.134
##         x11     -1.595         0.307       -0.175    -5.201    0.000    -2.196     -0.993
##         x19      0.330         0.074        0.113     4.441    0.000     0.184      0.475
##          x7     -2.018         0.405       -0.110    -4.989    0.000    -2.812     -1.225
##          x5      0.001         0.000        0.139     5.192    0.000     0.000      0.001
##         x17     -0.002         0.000       -0.122    -3.668    0.000    -0.003     -0.001
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 11
##
## - x4
##
##                          Model Summary
## ----------------------------------------------------------------
## R                       0.616       RMSE                18.342
## R-Squared               0.379       Coef. Var           35.649
## Adj. R-Squared          0.375       MSE                336.413
```

```
## Pred R-Squared          0.368       MAE                  12.628
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                  Sum of
##                 Squares          DF    Mean Square      F         Sig.
## --------------------------------------------------------------------------
## Regression    328483.691          11      29862.154    88.766    0.0000
## Residual      537588.150        1598        336.413
## Total         866071.841        1609
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig      lower      upper
## ----------------------------------------------------------------------------------------------------
## (Intercept)    106.987        11.645                  9.187    0.000    84.145    129.829
##         x13      0.019         0.036        0.022     0.531    0.596    -0.052      0.091
##         x12      0.133         0.031        0.157     4.350    0.000     0.073      0.193
##          x3     -0.895         0.100       -0.424    -8.954    0.000    -1.091     -0.699
##          x1     -0.556         0.071       -0.385    -7.871    0.000    -0.695     -0.418
##         x15     -0.247         0.051       -0.200    -4.888    0.000    -0.346     -0.148
##         x11     -1.586         0.309       -0.174    -5.132    0.000    -2.192     -0.980
##         x19      0.339         0.075        0.116     4.545    0.000     0.193      0.485
##          x7     -2.030         0.406       -0.111    -5.005    0.000    -2.825     -1.234
##          x5      0.001         0.000        0.166     5.777    0.000     0.001      0.001
##         x17     -0.002         0.000       -0.117    -3.477    0.001    -0.003     -0.001
##          x4      0.000         0.000       -0.066    -2.677    0.008     0.000      0.000
## ----------------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 12
##
## - x20
##
##                           Model Summary
## ----------------------------------------------------------------
## R                       0.630       RMSE                18.054
## R-Squared               0.397       Coef. Var           35.108
## Adj. R-Squared          0.392       MSE                325.937
## Pred R-Squared          0.385       MAE                 12.407
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                ANOVA
## --------------------------------------------------------------------------
##                  Sum of
```

```
##                   Squares          DF    Mean Square      F         Sig.
## ----------------------------------------------------------------------------
## Regression    324182.762          12      27015.230    82.885    0.0000
## Residual      492817.223        1512        325.937
## Total         816999.984        1524
## ----------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig      lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    105.487        12.108                 8.712    0.000    81.737    129.238
##         x13      0.048         0.038        0.056     1.276    0.202    -0.026      0.122
##         x12      0.142         0.032        0.168     4.431    0.000     0.079      0.205
##          x3     -0.926         0.101       -0.441    -9.150    0.000    -1.124     -0.727
##          x1     -0.576         0.071       -0.403    -8.067    0.000    -0.716     -0.436
##         x15     -0.280         0.053       -0.227    -5.267    0.000    -0.384     -0.176
##         x11     -1.644         0.319       -0.180    -5.150    0.000    -2.270     -1.018
##         x19      0.330         0.076        0.112     4.345    0.000     0.181      0.479
##          x7     -2.031         0.410       -0.112    -4.953    0.000    -2.835     -1.227
##          x5      0.001         0.000        0.180     6.152    0.000     0.001      0.001
##         x17     -0.002         0.001       -0.120    -3.464    0.001    -0.003     -0.001
##          x4      0.000         0.000       -0.058    -2.320    0.020     0.000      0.000
##         x20     -2.475         1.275       -0.045    -1.941    0.052    -4.975      0.026
## ----------------------------------------------------------------------------------------------
##
##
##
## Forward Selection: Step 13
##
## - x18
##
##                           Model Summary
## ------------------------------------------------------------------
## R                       0.624       RMSE                18.198
## R-Squared               0.390       Coef. Var           35.190
## Adj. R-Squared          0.384       MSE                331.175
## Pred R-Squared          0.377       MAE                 12.538
## ------------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
##                                  ANOVA
## ----------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square      F         Sig.
## ----------------------------------------------------------------------------
## Regression    311427.484          13      23955.960    72.336    0.0000
## Residual      487488.930        1472        331.175
## Total         798916.413        1485
## ----------------------------------------------------------------------------
##
##                                  Parameter Estimates
```

```
## ---------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig      lower      upper
## ---------------------------------------------------------------------------------------
## (Intercept)    102.640        12.517                 8.200    0.000    78.087    127.193
##         x13      0.068         0.039         0.078     1.749    0.081    -0.008      0.145
##         x12      0.158         0.033         0.185     4.781    0.000     0.093      0.223
##          x3     -0.905         0.105        -0.429    -8.648    0.000    -1.110     -0.700
##          x1     -0.556         0.074        -0.389    -7.533    0.000    -0.701     -0.412
##         x15     -0.299         0.056        -0.238    -5.318    0.000    -0.409     -0.188
##         x11     -1.746         0.331        -0.191    -5.281    0.000    -2.395     -1.098
##         x19      0.366         0.078         0.125     4.680    0.000     0.212      0.519
##          x7     -1.951         0.419        -0.107    -4.652    0.000    -2.774     -1.128
##          x5      0.001         0.000         0.184     6.172    0.000     0.001      0.001
##         x17     -0.002         0.001        -0.132    -3.693    0.000    -0.003     -0.001
##          x4      0.000         0.000        -0.060    -2.325    0.020     0.000      0.000
##         x20     -2.768         1.332        -0.051    -2.078    0.038    -5.381     -0.155
##         x18      0.248         0.114         0.054     2.167    0.030     0.024      0.472
## ---------------------------------------------------------------------------------------
##
##
##
## No more variables to be added.
##
## Variables Entered:
##
## + x13
## + x12
## + x3
## + x1
## + x15
## + x11
## + x19
## + x7
## + x5
## + x17
## + x4
## + x20
## + x18
##
##
## Final Model Output
## ------------------
##
##                            Model Summary
## ----------------------------------------------------------------
## R                        0.624       RMSE                18.198
## R-Squared                0.390       Coef. Var           35.190
## Adj. R-Squared           0.384       MSE                331.175
## Pred R-Squared           0.377       MAE                 12.538
## ----------------------------------------------------------------
##  RMSE: Root Mean Square Error
##  MSE: Mean Square Error
##  MAE: Mean Absolute Error
##
```

```
##                                 ANOVA
## --------------------------------------------------------------------------
##                   Sum of
##                  Squares          DF    Mean Square      F         Sig.
## --------------------------------------------------------------------------
## Regression    311427.484          13      23955.960    72.336    0.0000
## Residual      487488.930        1472        331.175
## Total         798916.413        1485
## --------------------------------------------------------------------------
##
##                                  Parameter Estimates
## ----------------------------------------------------------------------------------------------
##       model       Beta    Std. Error    Std. Beta      t        Sig      lower      upper
## ----------------------------------------------------------------------------------------------
## (Intercept)    102.640        12.517                 8.200    0.000    78.087    127.193
##         x13      0.068         0.039        0.078     1.749    0.081    -0.008      0.145
##         x12      0.158         0.033        0.185     4.781    0.000     0.093      0.223
##          x3     -0.905         0.105       -0.429    -8.648    0.000    -1.110     -0.700
##          x1     -0.556         0.074       -0.389    -7.533    0.000    -0.701     -0.412
##         x15     -0.299         0.056       -0.238    -5.318    0.000    -0.409     -0.188
##         x11     -1.746         0.331       -0.191    -5.281    0.000    -2.395     -1.098
##         x19      0.366         0.078        0.125     4.680    0.000     0.212      0.519
##          x7     -1.951         0.419       -0.107    -4.652    0.000    -2.774     -1.128
##          x5      0.001         0.000        0.184     6.172    0.000     0.001      0.001
##         x17     -0.002         0.001       -0.132    -3.693    0.000    -0.003     -0.001
##          x4      0.000         0.000       -0.060    -2.325    0.020     0.000      0.000
##         x20     -2.768         1.332       -0.051    -2.078    0.038    -5.381     -0.155
##         x18      0.248         0.114        0.054     2.167    0.030     0.024      0.472
## ----------------------------------------------------------------------------------------------

##
##                               Selection Summary
## ------------------------------------------------------------------------------
##         Variable                 Adj.
## Step    Entered     R-Square    R-Square      C(p)         AIC         RMSE
## ------------------------------------------------------------------------------
##    1    x13           0.3056      0.3052    504.6346    18312.5172    20.2839
##    2    x12           0.3301      0.3295    415.8220    18240.0870    19.9268
##    3    x3            0.3413      0.3403    289.7513    17379.9846    19.4614
##    4    x1            0.3584      0.3571    216.3231    17211.0930    19.1446
##    5    x15           0.3841      0.3826    130.9578    17132.3853    18.7608
##    6    x11           0.3844      0.3825    106.9512    16686.3101    18.6599
##    7    x19           0.3687      0.3663     89.9508    16004.7898    18.5925
##    8    x7            0.3695      0.3667     99.3141    15339.8857    18.6527
##    9    x5            0.3799      0.3766     65.6972    14944.6250    18.4840
##   10    x17           0.3770      0.3732     37.8704    14079.2684    18.3413
##   11    x4            0.3793      0.3750     38.6508    13950.4644    18.3416
##   12    x20           0.3968      0.3920     -9.6519    13167.4321    18.0537
##   13    x18           0.3898      0.3844     15.2454    12855.7503    18.1982
## ------------------------------------------------------------------------------
```

d. Best Model

```
##
## Attaching package: 'modelsummary'

## The following object is masked from 'package:Hmisc':
##
##     Mean

##
## Attaching package: 'kableExtra'

## The following object is masked from 'package:huxtable':
##
##     add_footnote

## The following object is masked from 'package:dplyr':
##
##     group_rows
```

Based on the model selection, the best model for the 2019 Math Achievement Rate consists of independent variables of 'Absent 0-5 Days Percentage', 'Avg. Annual Salaries for Teachers', 'Number of Teachers with a phd degree', 'White Student Percentage', 'Black Student Percentage', 'Economically Disadvantaged Student Percentage', 'Percentage of Gifted Students', and 'Urban/Rural Area of the School'. The best model for the 2021 Math Achievement Rate consists of independent variables of 'Absent 0-5 Days Percentage', 'Avg. Annual Salaries for Teachers', 'Number of Teachers with a phd degree', 'White Student Percentage', 'Economically Disadvantaged Student Percentage', 'Amount of Money Invested for Students', 'Per-Pupil Expenditure at School Level', and 'Urban/Rural Area of the School'.

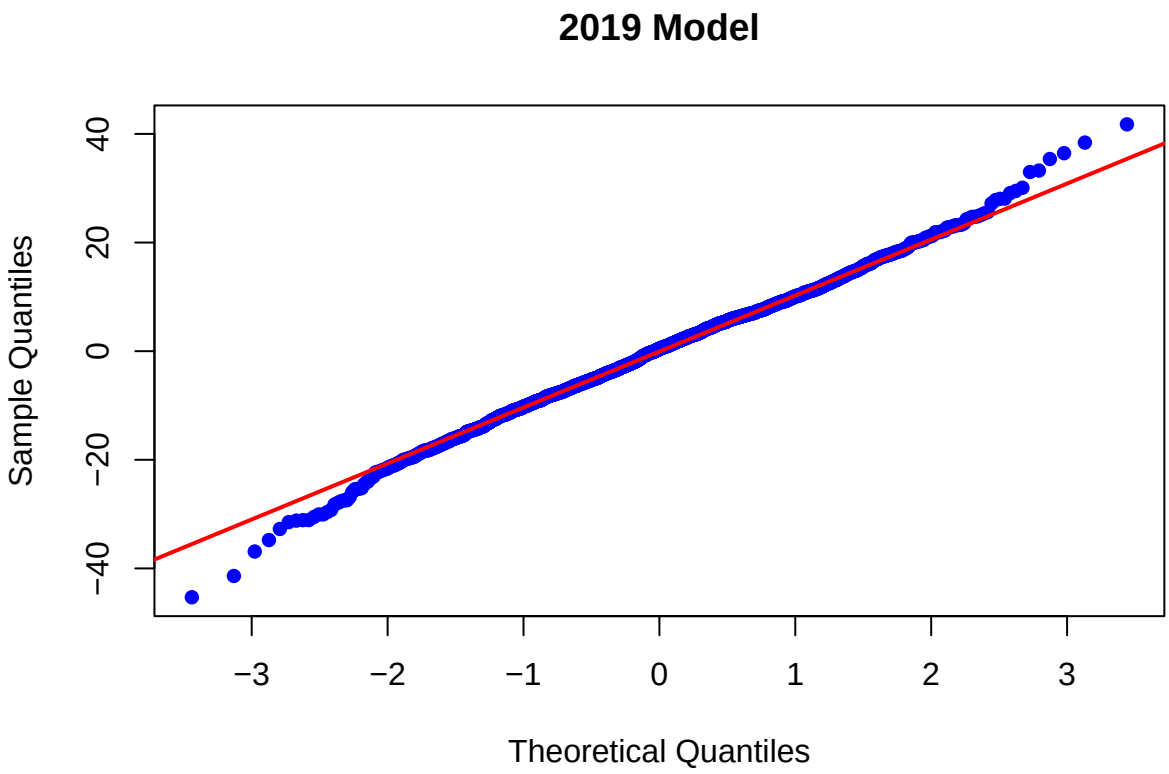
e. Best Model (Outlier Excluded)

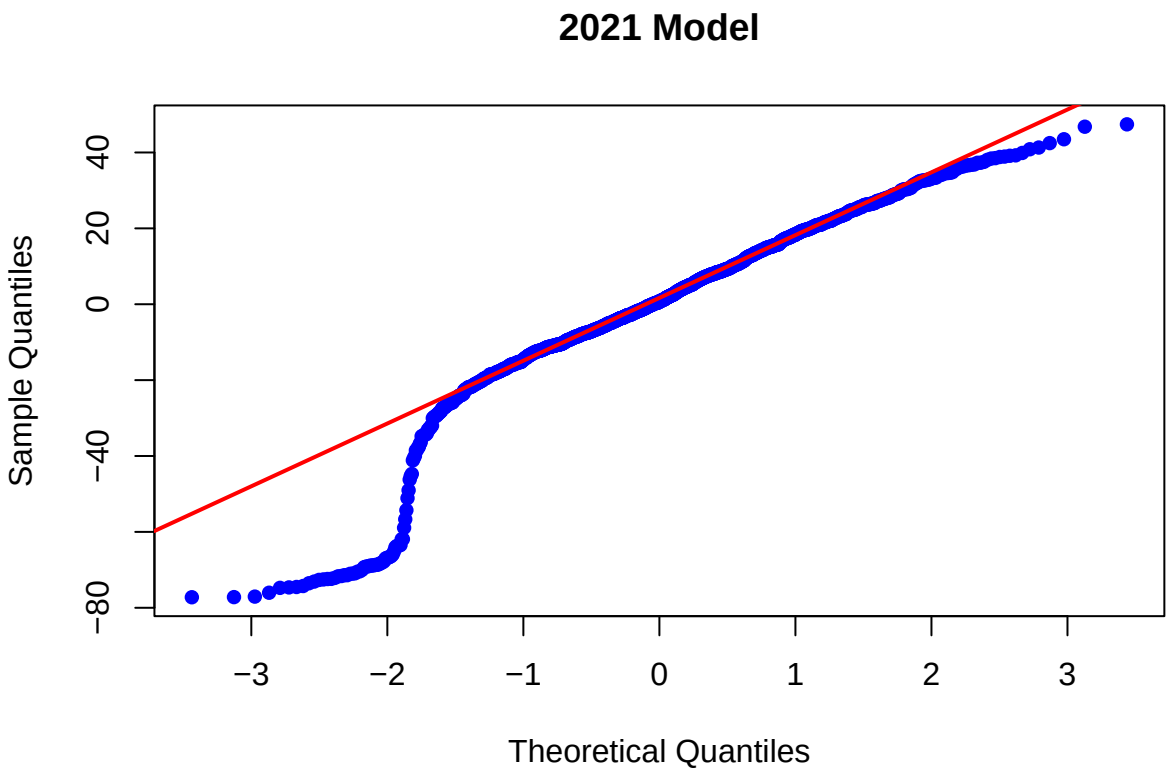
2019

2021

f. Normality Check

```
##
##  Anderson-Darling normality test
##
## data:  resid(best_model_2019)
## A = 1.0953, p-value = 0.007156

##
##  Anderson-Darling normality test
##
## data:  resid(best_model_2021)
## A = 21.181, p-value < 2.2e-16
```


2019 Model
Sample Quantiles
40
20
0
−20
−40
−3
−2
−1
0
1
2
3
Theoretical Quantiles



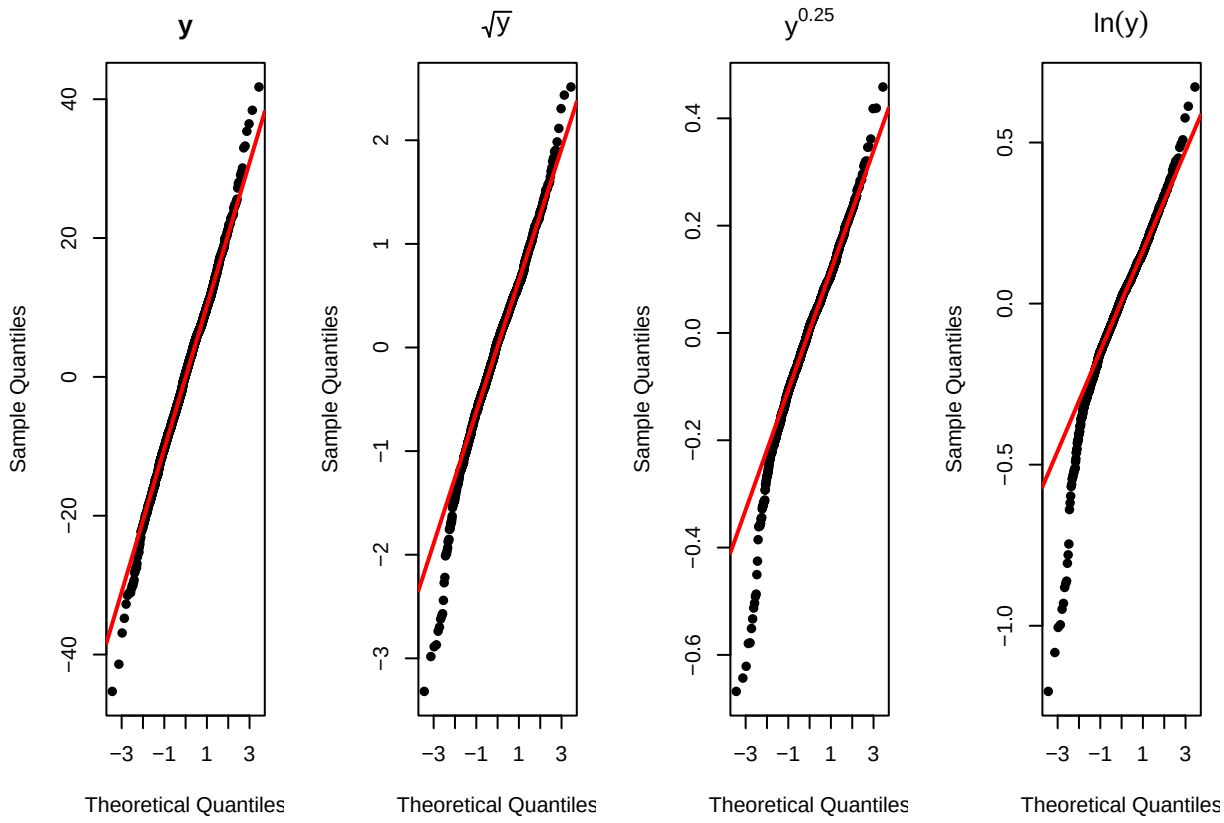
g. Transformation

2019

```
##
## Call:
## lm(formula = y1 ~ x1 + x3 + x5 + x7 + x11 + x14 + x15 + x19,
##     data = data_numeric)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -45.308  -7.014   0.417   6.890  41.763
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.969e+01  4.782e+00  18.755  < 2e-16 ***
## x1          -3.166e-01  3.147e-02 -10.058  < 2e-16 ***
## x3          -6.992e-01  4.776e-02 -14.639  < 2e-16 ***
## x5           6.207e-04  6.515e-05   9.527  < 2e-16 ***
## x7          -1.495e+00  2.281e-01  -6.556 7.31e-11 ***
## x11         -9.312e-01  1.060e-01  -8.784  < 2e-16 ***
## x14         -8.498e-02  1.226e-02  -6.930 5.94e-12 ***
## x15         -3.631e-01  2.557e-02 -14.200  < 2e-16 ***
## x19          4.665e-01  4.193e-02  11.127  < 2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.73 on 1713 degrees of freedom
##   (345 observations deleted due to missingness)
## Multiple R-squared:  0.6648, Adjusted R-squared:  0.6632
## F-statistic: 424.6 on 8 and 1713 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = trans_y1 ~ x1 + x3 + x5 + x7 + x11 + x14 + x15 +
##     x19, data = data_numeric)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.3193 -0.4164  0.0531  0.4400  2.5143
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  9.767e+00  3.060e-01  31.912  < 2e-16 ***
## x1          -2.257e-02  2.014e-03 -11.208  < 2e-16 ***
## x3          -4.749e-02  3.057e-03 -15.537  < 2e-16 ***
## x5           3.861e-05  4.170e-06   9.261  < 2e-16 ***
## x7          -9.309e-02  1.460e-02  -6.377 2.32e-10 ***
## x11         -6.075e-02  6.784e-03  -8.955  < 2e-16 ***
## x14         -4.778e-03  7.847e-04  -6.089 1.40e-09 ***
## x15         -2.309e-02  1.636e-03 -14.110  < 2e-16 ***
## x19          2.812e-02  2.683e-03  10.480  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6866 on 1713 degrees of freedom
##   (345 observations deleted due to missingness)
## Multiple R-squared:  0.6528, Adjusted R-squared:  0.6511
## F-statistic: 402.5 on 8 and 1713 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = trans_y2 ~ x1 + x3 + x5 + x7 + x11 + x14 + x15 +
##     x19, data = data_numeric)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.66808 -0.07039  0.01083  0.08012  0.45844
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.158e+00  5.587e-02  56.521  < 2e-16 ***
## x1          -4.284e-03  3.677e-04 -11.652  < 2e-16 ***
## x3          -8.822e-03  5.580e-04 -15.810  < 2e-16 ***
## x5           6.870e-06  7.612e-07   9.025  < 2e-16 ***
## x7          -1.653e-02  2.665e-03  -6.203 6.93e-10 ***
## x11         -1.108e-02  1.238e-03  -8.948  < 2e-16 ***
## x14         -8.023e-04  1.433e-04  -5.600 2.49e-08 ***
## x15         -4.142e-03  2.987e-04 -13.865  < 2e-16 ***
```

```
## x19          4.931e-03  4.898e-04  10.067  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1253 on 1713 degrees of freedom
##   (345 observations deleted due to missingness)
## Multiple R-squared:  0.6412, Adjusted R-squared:  0.6395
## F-statistic: 382.6 on 8 and 1713 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = trans_y3 ~ x1 + x3 + x5 + x7 + x11 + x14 + x15 +
##     x19, data = data_numeric)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.20386 -0.09666  0.01693  0.11288  0.67243
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.652e+00  8.290e-02  56.115  < 2e-16 ***
## x1          -6.532e-03  5.456e-04 -11.972  < 2e-16 ***
## x3          -1.319e-02  8.280e-04 -15.927  < 2e-16 ***
## x5           9.842e-06  1.129e-06   8.714  < 2e-16 ***
## x7          -2.359e-02  3.954e-03  -5.966 2.95e-09 ***
## x11         -1.628e-02  1.838e-03  -8.860  < 2e-16 ***
## x14         -1.079e-03  2.126e-04  -5.076 4.28e-07 ***
## x15         -5.968e-03  4.433e-04 -13.463  < 2e-16 ***
## x19          6.972e-03  7.269e-04   9.592  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.186 on 1713 degrees of freedom
##   (345 observations deleted due to missingness)
## Multiple R-squared:  0.6247, Adjusted R-squared:  0.623
## F-statistic: 356.5 on 8 and 1713 DF,  p-value: < 2.2e-16
```



2021

h. Influential Points

```
## named numeric(0)
```

```
## named numeric(0)
```

VIF

```
## Loading required package: carData
```

```
##
## Attaching package: 'carData'
```

```
## The following object is masked _by_ '.GlobalEnv':
##
##     Salaries
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
##       x1       x3       x5       x7      x11      x14      x15      x19
## 3.926058 4.385703 1.533826 1.261083 1.203689 2.165951 3.608543 1.622482
```

```
##       x3       x5      x11      x14      x15      x17      x19
## 1.272579 1.835423 2.455273 2.155140 4.054052 2.658713 1.537905
```

7. Residual Plot

```
# Residual Plot
# plot(data_numeric$y2, resid(best_model_2021), pch=16, col="blue")
# abline(0, 0, col = "red", lwd = 3)
# plot(fitted(best_model_2021), resid(best_model_2021), pch=16, col="blue", ylab=bquote(paste("e")))
# abline(0, 0, col = "red", lwd = 3)
```

Category

1. Urban & Rural

```
urban = data[data$Urban.Rural == "Urban", ]
rural = data[data$Urban.Rural == "Rural", ]
```

Testing if mean of Urban and Rural Math Achievement Rates are equal

$$H_0 : \mu_{Urban} - \mu_{Rural} = 0$$
$$H_0 : \mu_{Urban} - \mu_{Rural} > 1$$
$$p - value = 0.006737 < \alpha = 0.05 \rightarrow Reject\ H_0$$

```
mean(urban$All.Students.Math.Achievement)
```

```
## [1] 67.43092
```

```
mean(rural$All.Students.Math.Achievement)
```

```
## [1] 63.96621
```

```
t.test(urban$All.Students.Math.Achievement, rural$All.Students.Math.Achievement,
       mu=1, alternative='greater')
```

```
##
##  Welch Two Sample t-test
##
## data:  urban$All.Students.Math.Achievement and rural$All.Students.Math.Achievement
```

```
## t = 2.5042, df = 809.73, p-value = 0.006234
## alternative hypothesis: true difference in means is greater than 1
## 95 percent confidence interval:
##  1.843955      Inf
## sample estimates:
## mean of x mean of y
##  67.43092  63.96621
```

2. Race

Testing if the difference in mean of White and Black Math Achievement Rates is greater than 13

$$H_0 : \mu_{White} - \mu_{Black} = 0$$
$$H_0 : \mu_{White} - \mu_{Black} > 13$$
$$p - value = 0.004886 < \alpha = 0.05 \rightarrow Reject\ H_0$$

```
mean(data$White.Math.Achievement)
```

```
## [1] 63.89831
```

```
mean(data$Black.Math.Achievement)
```

```
## [1] 48.41171
```

```
t.test(data$White.Math.Achievement, data$Black.Math.Achievement,
       mu=13, alternative='greater')
```

```
##
##  Welch Two Sample t-test
##
## data:  data$White.Math.Achievement and data$Black.Math.Achievement
## t = 2.6881, df = 3705.6, p-value = 0.003609
## alternative hypothesis: true difference in means is greater than 13
## 95 percent confidence interval:
##  13.96467      Inf
## sample estimates:
## mean of x mean of y
##  63.89831  48.41171
```

```
mean(urban$White.Percentage)
```

```
## [1] 35.32944
```

```
mean(rural$White.Percentage)
```

```
## [1] 50.33252
```

```r
mean(urban$Black.Percentage)
```

```
## [1] 38.83198
```

```r
mean(rural$Black.Percentage)
```

```
## [1] 32.6066
```

3. Economy

```r
# 100% Econ Disadv Percentage
Econ_Dia_100 = data[data$Econ.Disadvantaged.Percentage == '100', ]
Econ_Dia_100_urban = Econ_Dia_100[Econ_Dia_100$Urban.Rural == "Urban",]
Econ_Dia_100_rural = Econ_Dia_100[Econ_Dia_100$Urban.Rural == "Rural",]
# 2019
c(mean(Econ_Dia_100_urban$All.Students.Math.Achievement),
  mean(Econ_Dia_100_rural$All.Students.Math.Achievement))
```

```
## [1] 54.89206 58.82841
```

```r
# 2021
c(mean(Econ_Dia_100_urban$X2021.All.Students.Math.Achievement),
  mean(Econ_Dia_100_rural$X2021.All.Students.Math.Achievement))
```

```
## [1] 36.54889 47.08022
```

$$H_0 : \mu_{Rural\ EconDis} - \mu_{Urban\ EconDis} = 0$$
$$H_0 : \mu_{Rural\ EconDis} - \mu_{Urban\ EconDis} > 15$$
$$p{-}value = 0.04061 < \alpha = 0.05 \rightarrow Reject\ H_0$$

```r
mean(urban$Econ.Disadvantaged.Percentage)
```

```
## [1] 65.87646
```

```r
mean(rural$Econ.Disadvantaged.Percentage)
```

```
## [1] 83.43863
```

```r
t.test(rural$Econ.Disadvantaged.Percentage, urban$Econ.Disadvantaged.Percentage ,
       mu=15, alternative='greater')
```

```
##
##  Welch Two Sample t-test
##
## data:  rural$Econ.Disadvantaged.Percentage and urban$Econ.Disadvantaged.Percentage
```

```
## t = 1.7462, df = 809.96, p-value = 0.04058
## alternative hypothesis: true difference in means is greater than 15
## 95 percent confidence interval:
##  15.14594      Inf
## sample estimates:
## mean of x mean of y
##  83.43863  65.87646
```

4. Teacher Certificates

$$H_0 : \mu_{Urban\ Certificates} - \mu_{Rural\ Certificates} = 0$$
$$H_0 : \mu_{Urban\ Certificates} - \mu_{Rural\ Certificates} > 10$$
$$p{-}value = 0.001039 < \alpha = 0.05 \rightarrow Reject\ H_0$$

```
# Number of total certificates at school level
mean(urban$Total)
```

```
## [1] 59.34454
```

```
mean(rural$Total)
```

```
## [1] 44.75061
```

```
t.test(urban$Total, rural$Total,
       mu=10, alternative='greater')
```

```
##
##  Welch Two Sample t-test
##
## data:  urban$Total and rural$Total
## t = 3.1267, df = 806.62, p-value = 0.0009156
## alternative hypothesis: true difference in means is greater than 10
## 95 percent confidence interval:
##  12.17444      Inf
## sample estimates:
## mean of x mean of y
##  59.34454  44.75061
```

Reference

[1] Mervosh, Sarah. "The Pandemic Erased Two Decades of Progress in Math and Reading." The New York Times, The New York Times, 1 Sept. 2022, https://www.nytimes.com/2022/09/01/us/national-test-scores-math-reading- pandemic.html?smid=nytcore-ios-share&referringSource=articleShare.

[2] Stern, Paul. "The Pandemic Worsened Racial Achievement Gaps. Making up the Difference Won't Be Easy." CT Mirror, 23 May 2022, https://ctmirror.org/2022/05/22/the-pandemic-worsened-racial-achievement-gaps-making-up-the-difference-wont-be-easy/.

[3] Georgia Department of Education. CCRPI Reports. Retrieved from https://www.gadoe.org/CCRPI/

Pages/default.aspx

[4] The Governor's Office of Student Achievement. Downloadable Dataset. Retrieved from https://gosa.georgia.gov/dashboards-data-report-card/downloadable-data

row	column	cor	p
x10	x16	0.896	0
x8	x16	0.894	0
x8	x10	0.844	0
x1	x3	-0.833	0
x12	x13	-0.815	0
y1	x15	-0.702	0
x14	x15	0.699	0
x13	x15	0.688	0
y1	y2	0.679	0
x11	x17	-0.64	0
y1	x13	-0.639	0
x1	x2	-0.592	0
y1	x14	-0.588	0
x15	x19	-0.583	0
x5	x9	0.575	0
y2	x13	-0.553	0
y1	x12	0.55	0
y2	x12	0.541	0
x12	x15	-0.532	0
x5	x15	-0.504	0
x10	x11	0.503	0
y1	x19	0.501	0
x5	x6	0.494	0
x13	x14	0.489	0
x4	x5	0.488	0
y1	x3	-0.478	0
x10	x17	-0.472	0
y2	x15	-0.472	0
x12	x14	-0.431	0
x15	x18	0.43	0
x14	x19	-0.43	0
x18	x19	-0.425	0
x2	x12	-0.424	0

	2019 Best Model	2021 Best Model
(Intercept)	89.695	140.757
	(4.782)	(9.562)
x1	−0.317	
	(0.031)	
x3	−0.699	−0.182
	(0.048)	(0.049)
x5	0.0006	0.0004
	(0.000 07)	(0.0001)
x7	−1.495	
	(0.228)	
x11	−0.931	−3.604
	(0.106)	(0.290)
x14	−0.085	−0.036
	(0.012)	(0.023)
x15	−0.363	−0.366
	(0.026)	(0.051)
x19	0.467	0.047
	(0.042)	(0.069)
x17		−0.004
		(0.0005)
Num.Obs.	1722	1702
R2	0.665	0.272
R2 Adj.	0.663	0.269
AIC	13 070.1	15 002.4
BIC	13 124.6	15 051.3
Log.Lik.	−6525.067	−7492.190
F		90.288
RMSE	10.70	19.75