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
```

```r
write.csv(data, "merged_data.csv")
```

Exporting the data to ExpertFit to fit distribution and test normality.

a. Data Summary

```r
library(nortest)
ad.test(data$All.Students.Math.Achievement)
```

```
##
##  Anderson-Darling normality test
##
## data:  data$All.Students.Math.Achievement
## A = 4.4728, p-value = 4.264e-11
```

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

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

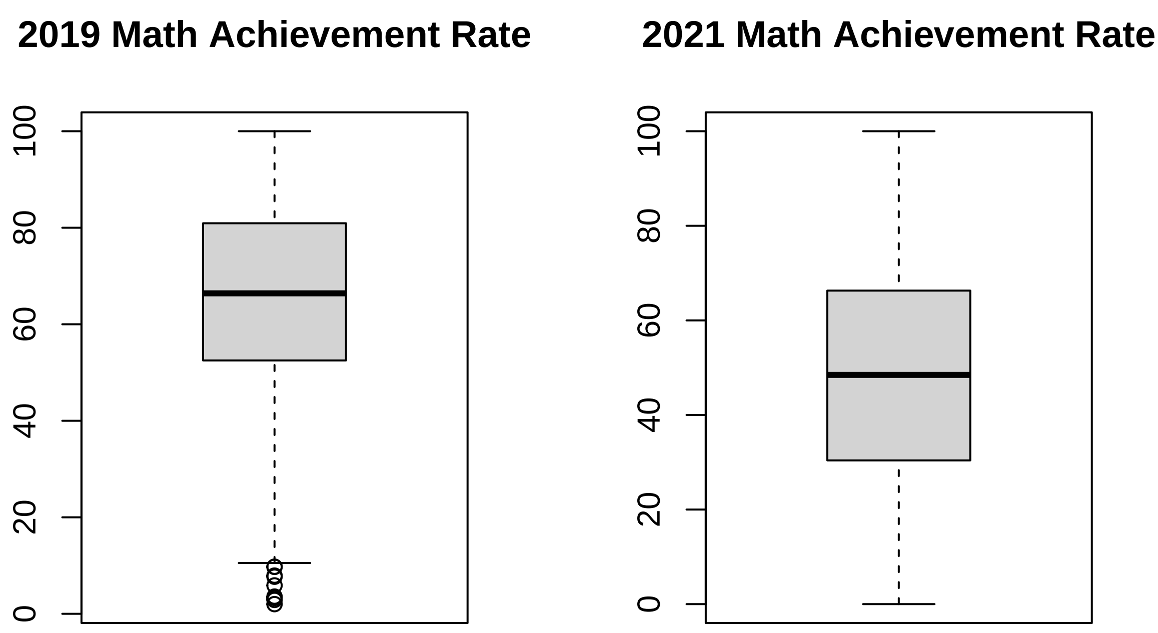
Mean and median Math test achievement rates are higher in 2019 than in 2021.

	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

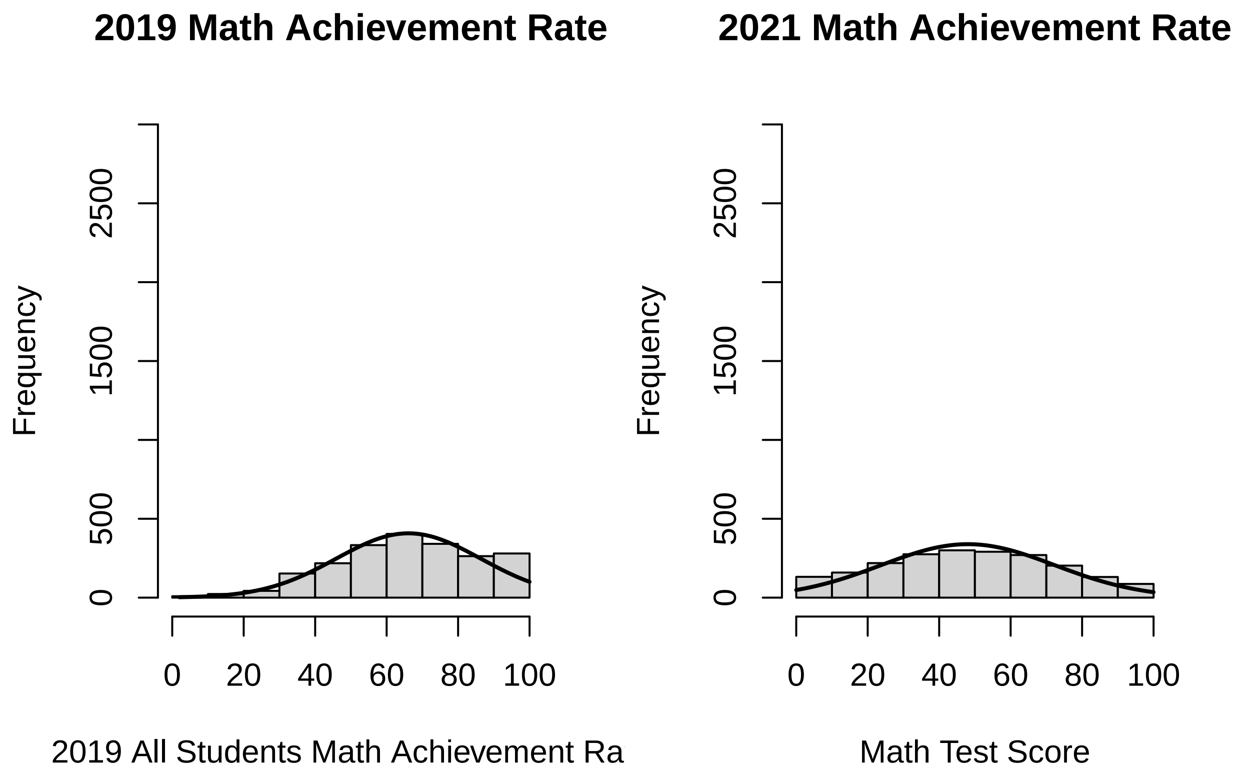
```
#average change in achievement rate
(52.23121-67.99686)/67.99686
```

```
## [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

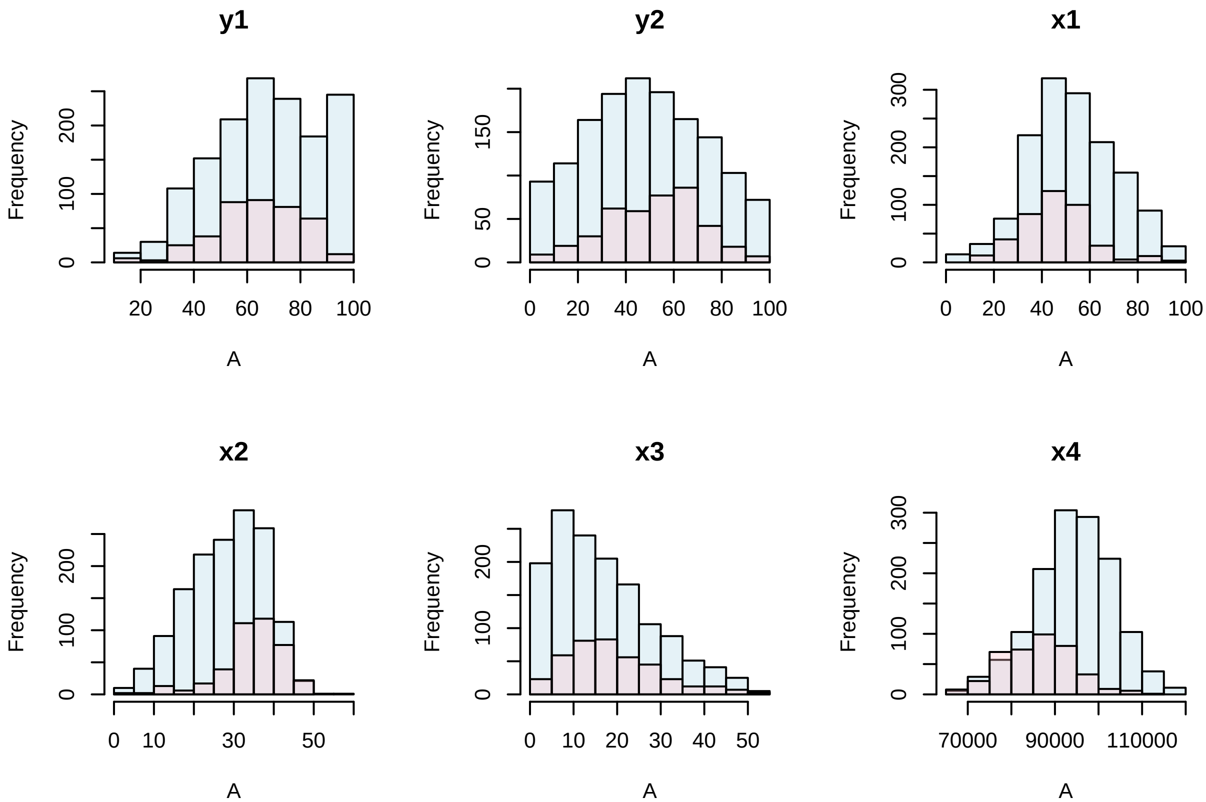
a. Iterations of the analysis process

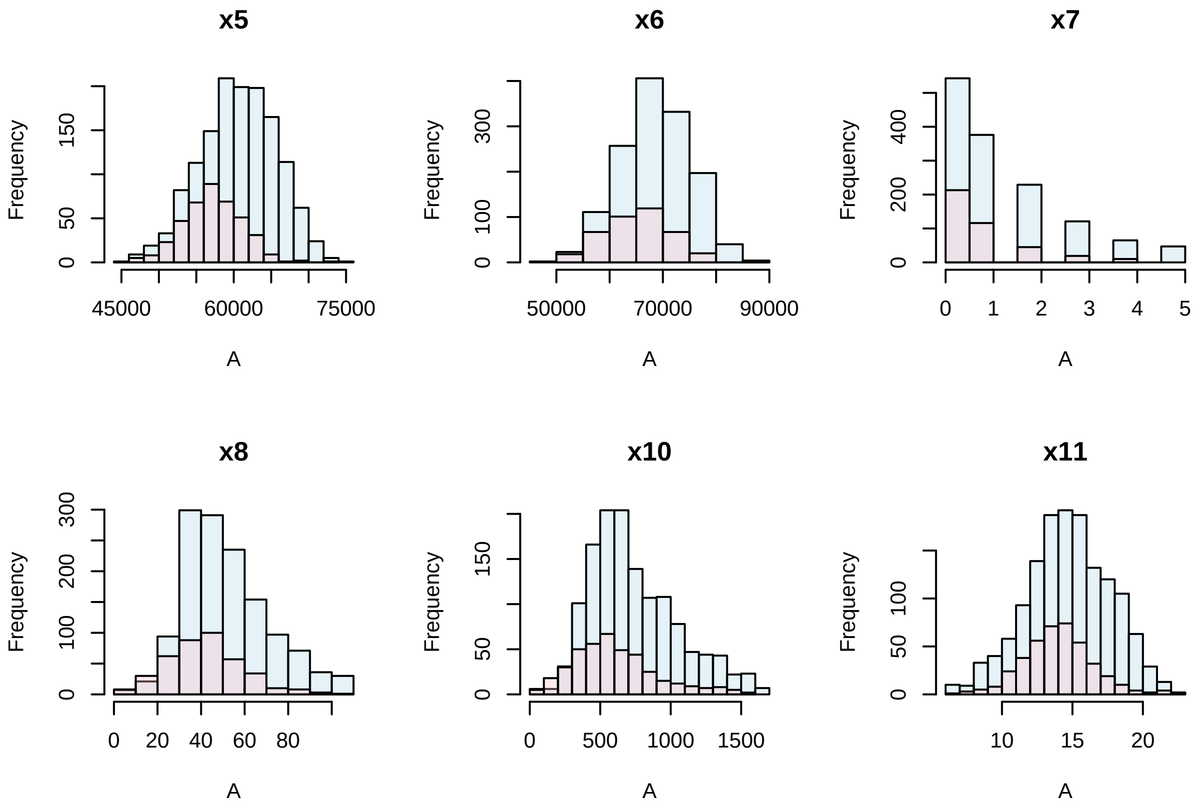
- paragraph description

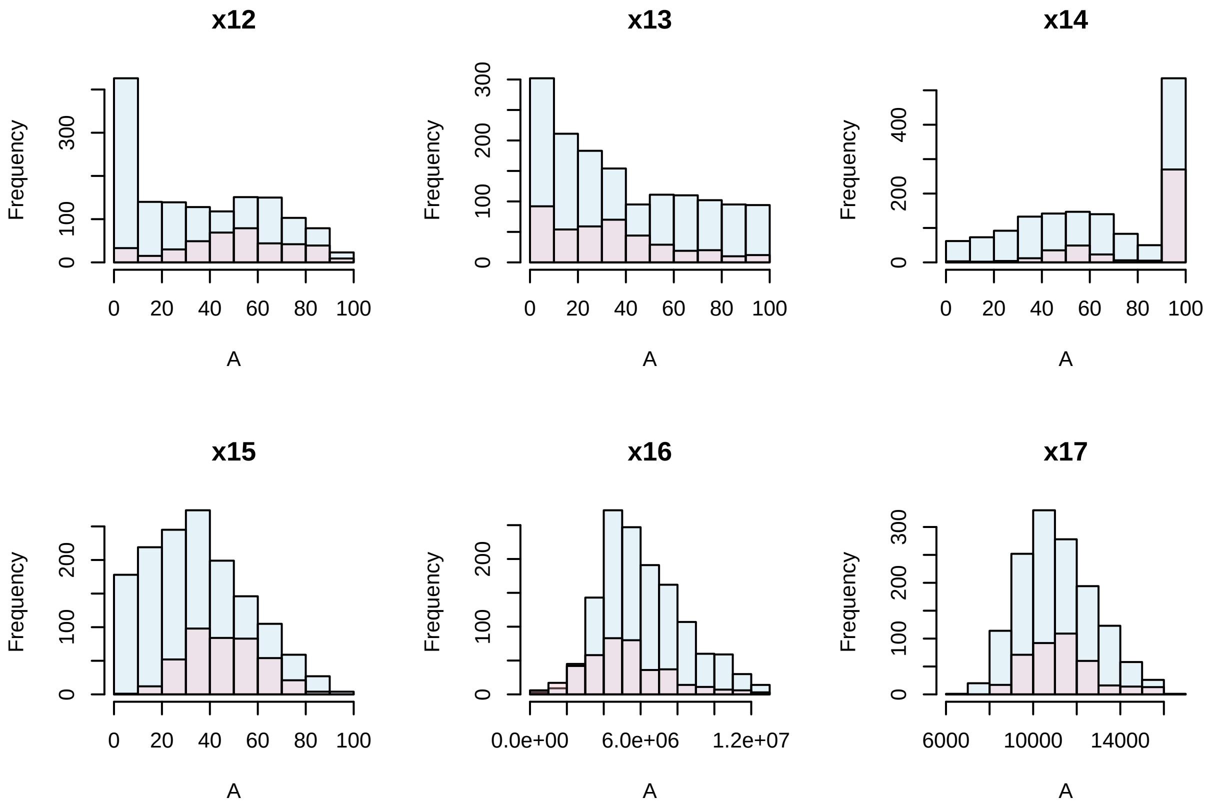
c. Plots of variables- Scatterplot

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
```

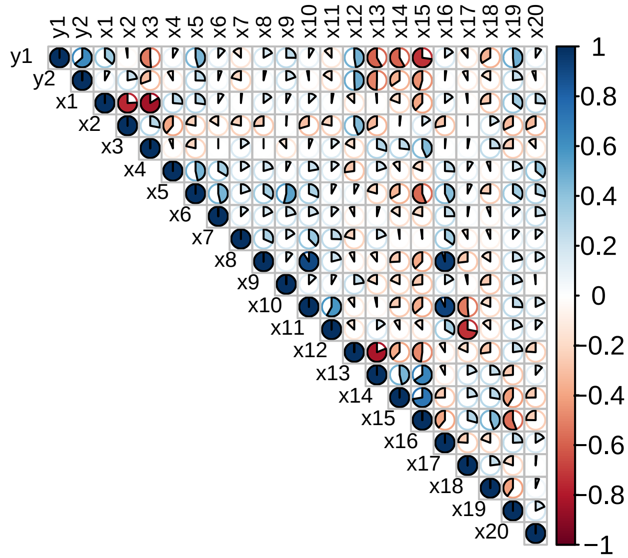

y1
y2
x1
x2
x3
x4
Frequency
A
0
100
200
50
150
300
20
40
60
80
100
10
30
50
70000
90000
110000

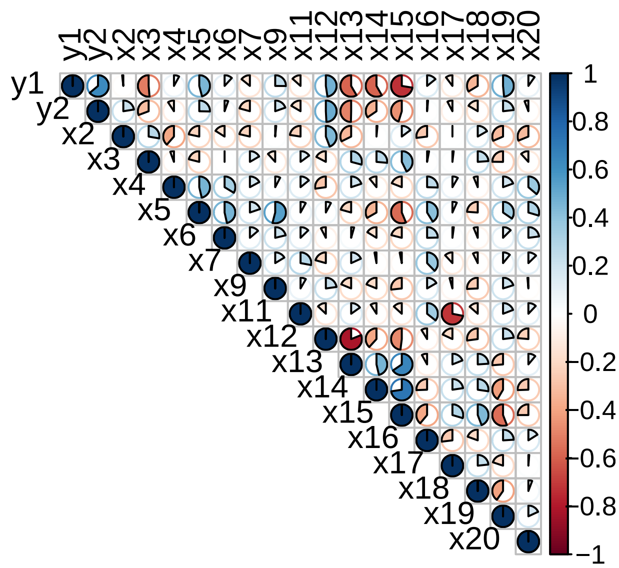

x5
Frequency
0
50
150
45000
60000
75000
A
x6
Frequency
0
100
300
50000
70000
90000
A
x7
Frequency
0
200
400
0
1
2
3
4
5
A
x8
Frequency
0
100
200
300
0
20
40
60
80
A
x10
Frequency
0
50
150
0
500
1000
1500
A
x11
Frequency
0
50
100
10
15
20
A



b. Multicollinearity

```
## corrplot 0.92 loaded
```



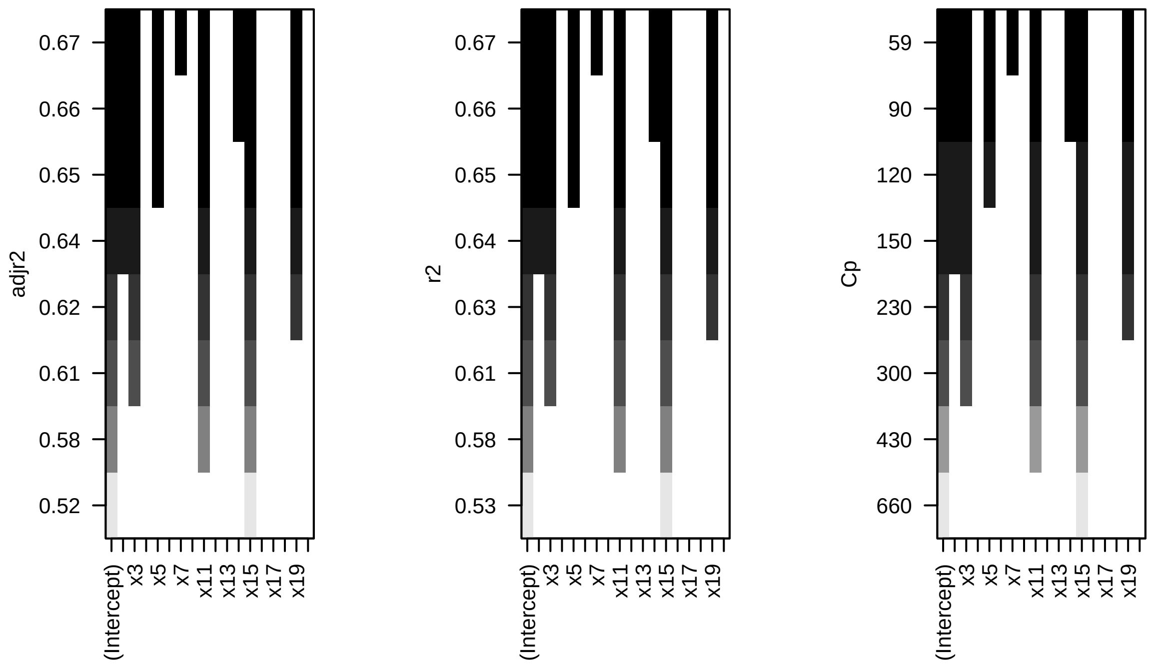


Rewrite!

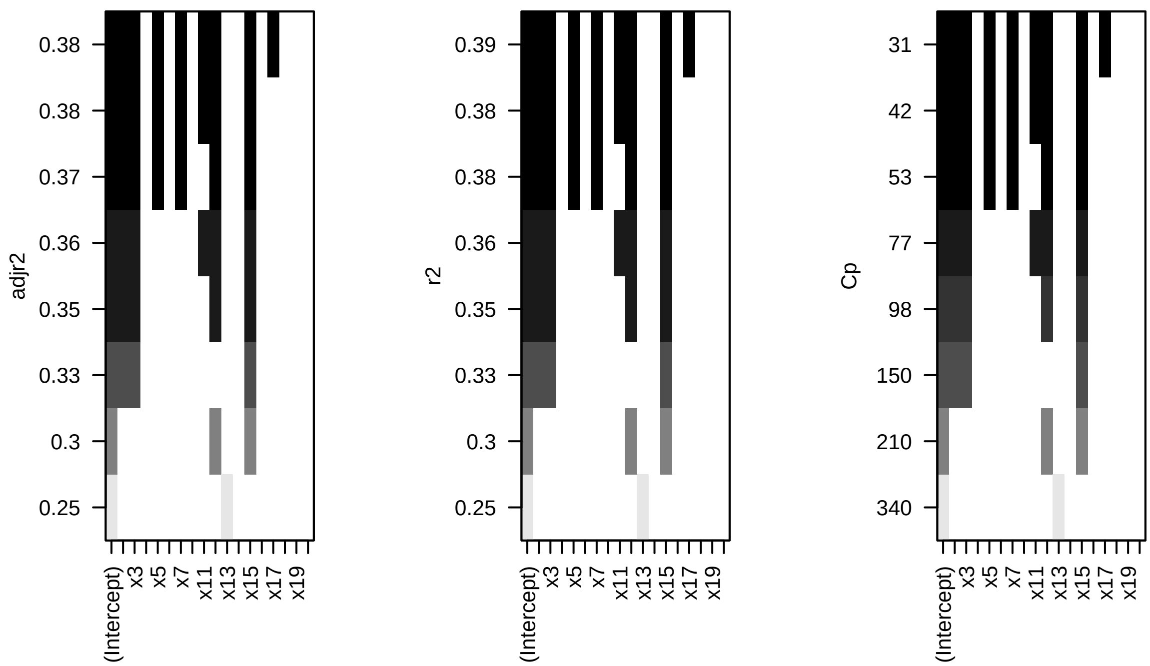
Before doing the model selection process, a Multicollinearity check produces high correlation of (x1:x2,x3), (x8:x10,x16), (x5:x7), (x9:x4,x5,x6), and (x12:x4,x5,x6,x8). And another set of variables that have a high correlation is y1 and y2, since we are modeling them separately as response variables, we do not need to drop any of them. The renewed plot is on the right.

d. Model Selection

2019 Model Selection



2021 Model Selection



d. Best Model

```
##
## 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'.

	2019 Best Model	2021 Best Model
(Intercept)	16.436	−12.483
	(4.505)	(7.417)
x1	0.221	0.106
	(0.019)	(0.031)
x5	0.0009	0.001
	(0.000 07)	(0.0001)
x7	−2.141	−3.135
	(0.237)	(0.458)
x12	0.058	0.349
	(0.019)	(0.022)
x13	−0.197	
	(0.019)	
x14	−0.150	−0.091
	(0.012)	(0.020)
x19	0.409	
	(0.043)	
x20	−2.143	−2.880
	(0.764)	(1.283)
x16		0.000 000 3
		(0.000 000 3)
x17		−0.0004
		(0.0003)
Num.Obs.	1628	1561
R2	0.638	0.355
R2 Adj.	0.636	0.352
AIC	12 515.8	13 646.9
BIC	12 569.7	13 700.4
Log.Lik.	−6247.892	−6813.438
F		106.702
RMSE	11.23	19.03

e. Best Model (Outlier Excluded)

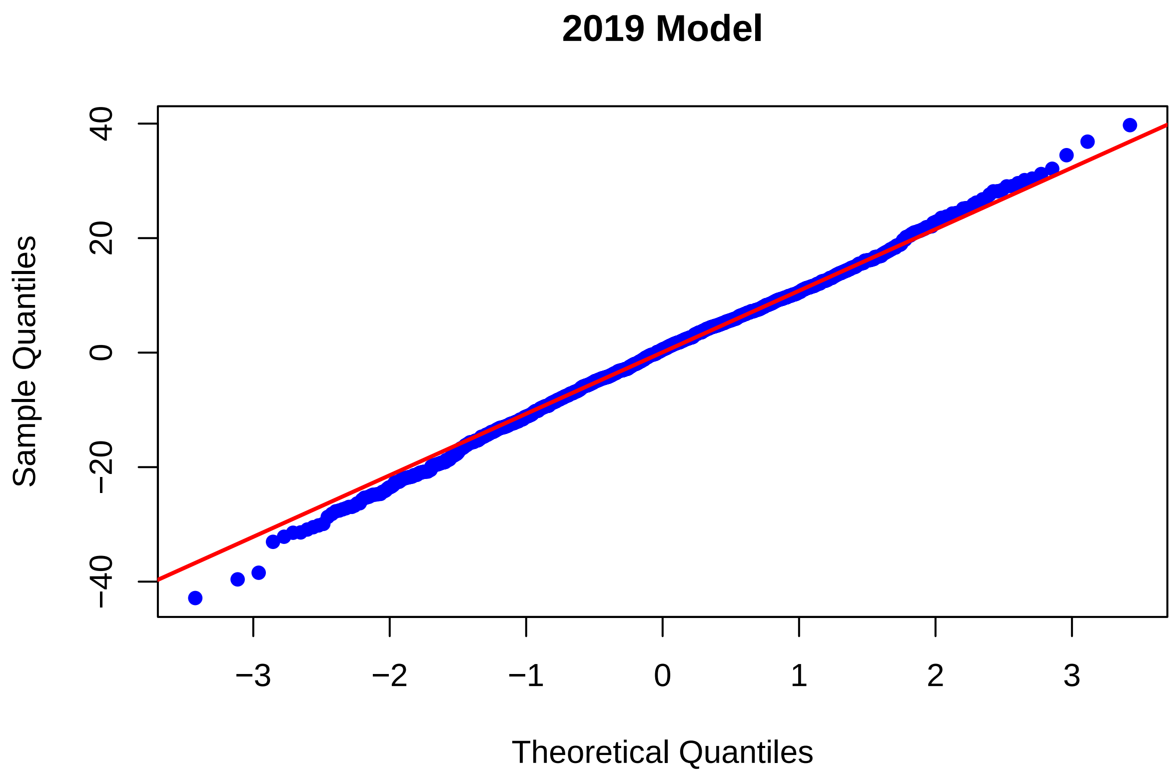
2019

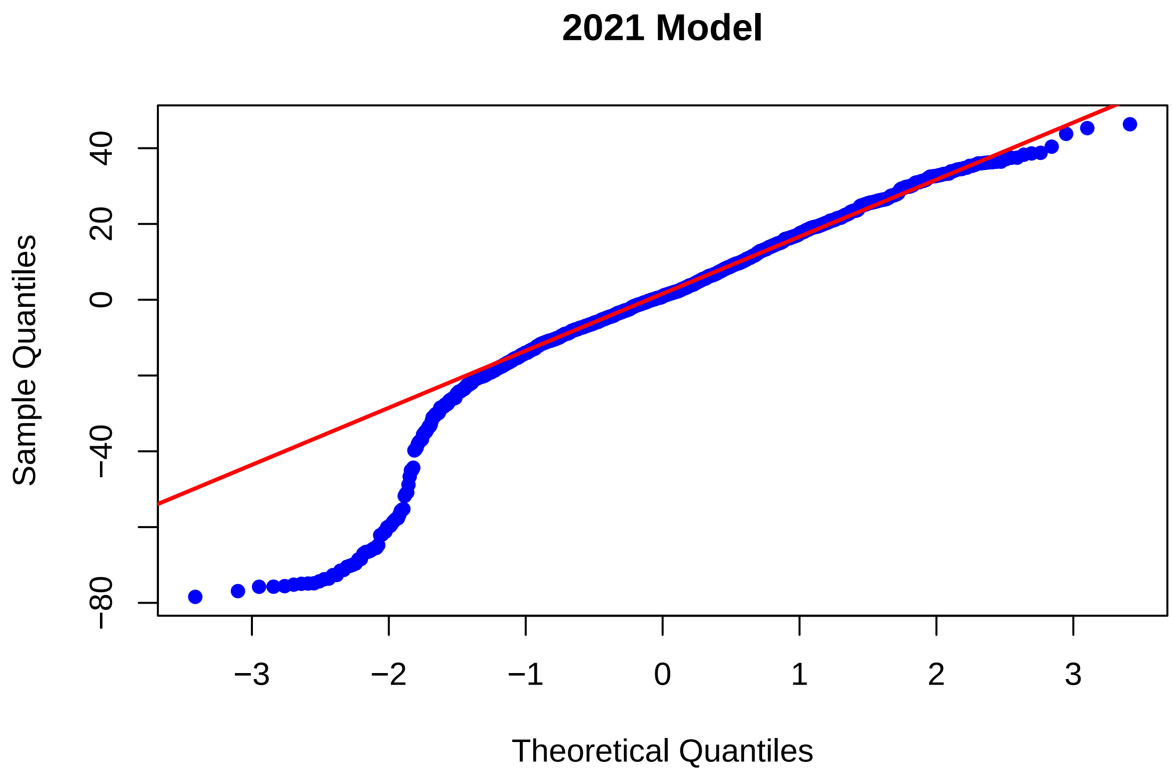
2021

f. Normality Check

```
##
##  Anderson-Darling normality test
##
## data:  resid(best_model_2019)
## A = 1.2077, p-value = 0.003783
```

```
##
##  Anderson-Darling normality test
##
## data:  resid(best_model_2021)
## A = 20.141, p-value < 2.2e-16
```





g. Transformation

2019

```
##
## Call:
## lm(formula = y1 ~ x1 + x2 + x5 + x7 + x12 + x14 + x17 + x19,
##     data = data_numeric)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -46.313  -7.365   0.463   7.765  36.663
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.112e+01  4.865e+00  -6.397 2.06e-10 ***
## x1           4.374e-01  2.554e-02  17.125  < 2e-16 ***
## x2           5.872e-01  5.295e-02  11.088  < 2e-16 ***
## x5           1.004e-03  6.577e-05  15.261  < 2e-16 ***
## x7          -1.630e+00  2.429e-01  -6.710 2.66e-11 ***
## x12          1.403e-01  1.362e-02  10.302  < 2e-16 ***
## x14         -1.561e-01  1.108e-02 -14.087  < 2e-16 ***
## x17          5.903e-05  1.846e-04   0.320    0.749
## x19          5.076e-01  4.392e-02  11.559  < 2e-16 ***
## ---
```

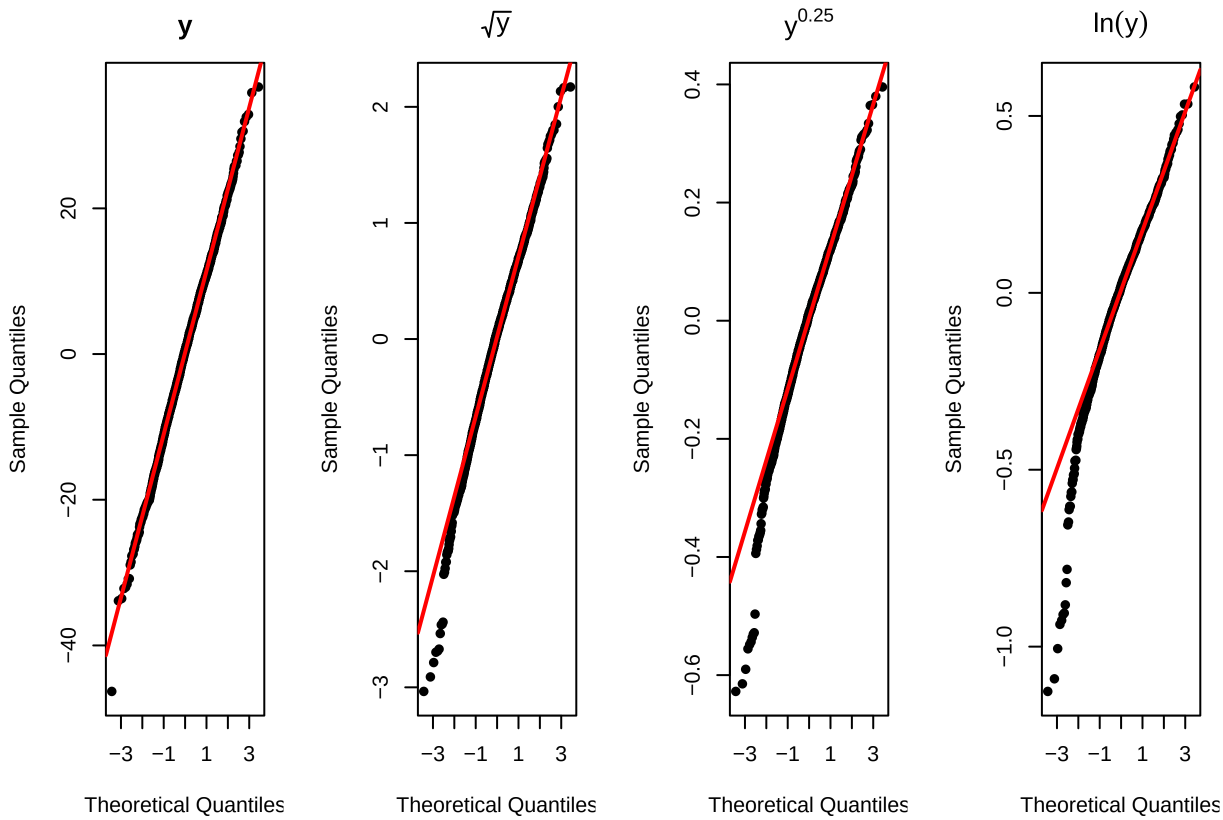
```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.22 on 1647 degrees of freedom
##   (411 observations deleted due to missingness)
## Multiple R-squared:  0.6354, Adjusted R-squared:  0.6336
## F-statistic: 358.7 on 8 and 1647 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = trans_y1 ~ x1 + x2 + x5 + x7 + x12 + x14 + x17 +
##     x19, data = data_numeric)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.03527 -0.44353  0.05277  0.48535  2.17113
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.750e+00  3.109e-01   5.629 2.13e-08 ***
## x1           2.840e-02  1.632e-03  17.395  < 2e-16 ***
## x2           4.048e-02  3.384e-03  11.963  < 2e-16 ***
## x5           6.324e-05  4.203e-06  15.044  < 2e-16 ***
## x7          -1.006e-01  1.553e-02  -6.482 1.20e-10 ***
## x12          9.002e-03  8.704e-04  10.342  < 2e-16 ***
## x14         -9.254e-03  7.083e-04 -13.065  < 2e-16 ***
## x17          2.116e-06  1.179e-05   0.179    0.858
## x19          3.064e-02  2.807e-03  10.915  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7173 on 1647 degrees of freedom
##   (411 observations deleted due to missingness)
## Multiple R-squared:  0.6237, Adjusted R-squared:  0.6219
## F-statistic: 341.3 on 8 and 1647 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = trans_y2 ~ x1 + x2 + x5 + x7 + x12 + x14 + x17 +
##     x19, data = data_numeric)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.62758 -0.07664  0.01175  0.08596  0.39569
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.685e+00  5.664e-02  29.758  < 2e-16 ***
## x1           5.158e-03  2.974e-04  17.343  < 2e-16 ***
## x2           7.575e-03  6.165e-04  12.286  < 2e-16 ***
## x5           1.132e-05  7.658e-07  14.783  < 2e-16 ***
## x7          -1.777e-02  2.829e-03  -6.283 4.24e-10 ***
## x12          1.621e-03  1.586e-04  10.220  < 2e-16 ***
## x14         -1.601e-03  1.290e-04 -12.405  < 2e-16 ***
## x17          2.480e-07  2.149e-06   0.115    0.908
```

```
## x19          5.373e-03  5.113e-04  10.508  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1307 on 1647 degrees of freedom
##   (411 observations deleted due to missingness)
## Multiple R-squared:  0.6128, Adjusted R-squared:  0.6109
## F-statistic: 325.8 on 8 and 1647 DF,  p-value: < 2.2e-16
```

```
##
## Call:
## lm(formula = trans_y3 ~ x1 + x2 + x5 + x7 + x12 + x14 + x17 +
##     x19, data = data_numeric)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.12661 -0.10550  0.01824  0.12123  0.58191
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.476e+00  8.377e-02  29.558  < 2e-16 ***
## x1           7.538e-03  4.399e-04  17.137  < 2e-16 ***
## x2           1.140e-02  9.119e-04  12.503  < 2e-16 ***
## x5           1.632e-05  1.133e-06  14.406  < 2e-16 ***
## x7          -2.519e-02  4.184e-03  -6.022 2.12e-09 ***
## x12          2.343e-03  2.345e-04   9.988  < 2e-16 ***
## x14         -2.223e-03  1.909e-04 -11.646  < 2e-16 ***
## x17          1.825e-07  3.178e-06   0.057    0.954
## x19          7.592e-03  7.563e-04  10.039  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1933 on 1647 degrees of freedom
##   (411 observations deleted due to missingness)
## Multiple R-squared:  0.5974, Adjusted R-squared:  0.5955
## F-statistic: 305.5 on 8 and 1647 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       x5       x7      x12      x13      x14      x19      x20
## 1.251926 1.349700 1.160524 3.499350 3.446830 1.613986 1.470418 1.320757
```

```
##       x1       x5       x7      x12      x14      x16      x17      x20
## 1.141084 1.541745 1.234421 1.564701 1.592821 1.575792 1.269641 1.297179
```

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
```

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

```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