Final Project

Alan Qiao, Jacqueline Dinh

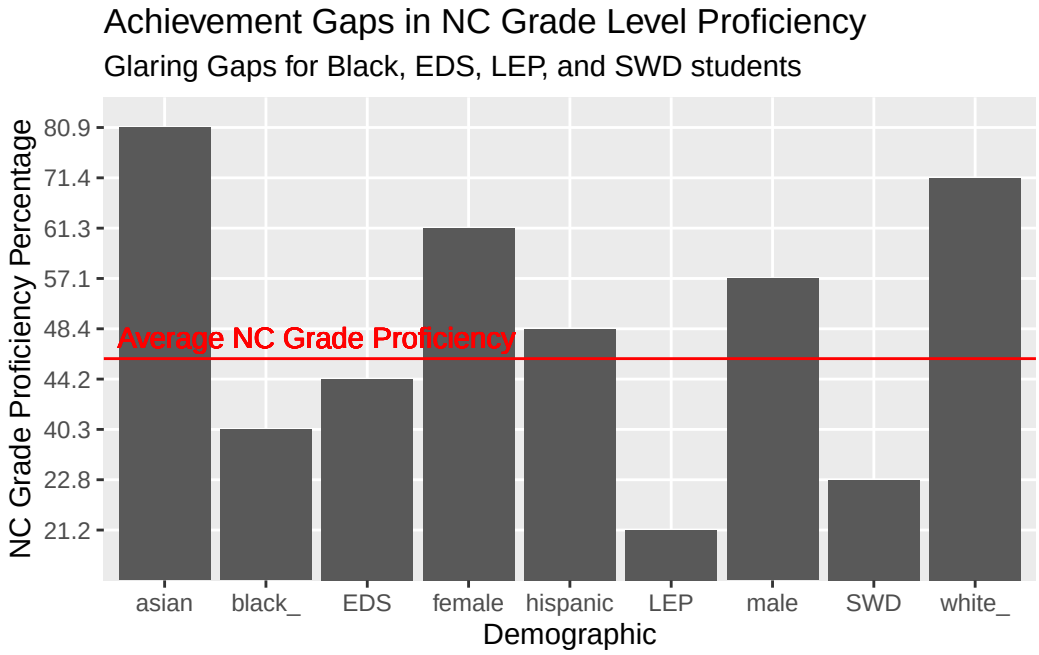
Read in the data

Introduction

As students from the south, where desegregation of public schools moved slowly, with some southern states actively resisting integration (cite), we were interested in analyzing how the legacy of segregation has impacted modern-day grade level proficiency in North Carolina public schools. North Carolina, specifically, is seen as a success story for public school integration; in 1971, the historic court case Swann v. Charlotte-Mecklenburg Board of Education established a busing system that transported students of color from historically redlined areas of Charlotte to schools located in predominantly white areas. However, in the 90s, the city started to see a pattern of resegregation, similar to that of pre-Swann (cite Mickelson). In this project, we expanded upon Dr. Roslyn Mickelson's work in The Cumulative Disadvantages of First- and Second-Generation Segregation for Middle School Achievement, which found that a student attending a school with a racial imbalance towards a Black population is more likely to be placed in a lower academic 'track', which Charlotte schools used to categorize students based on grade level proficiency.

We wanted to see if the conclusions and implications of Dr. Mickelson's paper rings true in modern times across the entire state of North Carolina. Using a dataset that recorded the percentage of students of different demographics that are grade level proficient and college/career ready in the 2016-2017 academic school year, the bar graph below plots the overall grade level proficiency of students that are Asian, Black, Hispanic, white, are economically disadvantaged, have limited English proficiency, and have disabilities across all North Carolina schools, it can be seen that there is a large gap between white students grade level proficiency (the plotted red line) and Black students, Hispanic students, economically disadvantaged students, students with limited English proficiency, and students with disabilities' grade level proficiency. The bar graph shows there are still existing achievement gaps between minority students and white students. This provided further motivation in analyzing how the percentage of white students and different minority students in a school, shaped by the legacy of segregation, may impact grade level proficiency. Our research question is: how do the white students, students of color, economically disadvantaged students, students with limited English proficiency, and

students with disabilities affect grade level proficiency of a North Carolina public school? We hypothesized that the higher the percentage of white students in a school's student body, the higher the overall grade level proficiency of the school. Conversely, the higher the percentage of students of color, economically disadvantaged students, students with limited English proficiency, and students with disabilities, the lower the grade level proficiency of the school.



Data

To investigate our research question, we used the dataset NC School Performance (2016 - 2017). This dataset is taken from Durham Open Data, a program managed jointly by the City of Durham's Technology Solutions Department and the County of Durham's Information Services & Technology Department to make data open and free to everyone. Durham Open Data collected this data from the North Carolina Department of Public Instruction and manipulated the annual testing data of public schools in North Carolina from the 2016-2017 school year. The main purpose of this dataset is to record the percentage of students of various races and other factors described that are grade level proficient and college/career ready. In this dataset, there are 32 variables and 85,766 observations separated by school, subject, and standard. The variable Standard was separated into Grade Level Proficient and College and Career Ready, determined by students' test scores of Level 4 & 5 on the CCR for college/career readiness and Level 3 & Above for grade level proficiency. The variable subject included all EOG/EOC Subjects and individual EOC Subjects along with the Math, Science, and Reading EOG and EOC assessments for each grade. To narrow the dataset down and compare schools, we chose

to filter the dataset to view scores of the observations that were determined by the subject of all EOG/EOC Subjects and the standard of Grade Level Proficient.

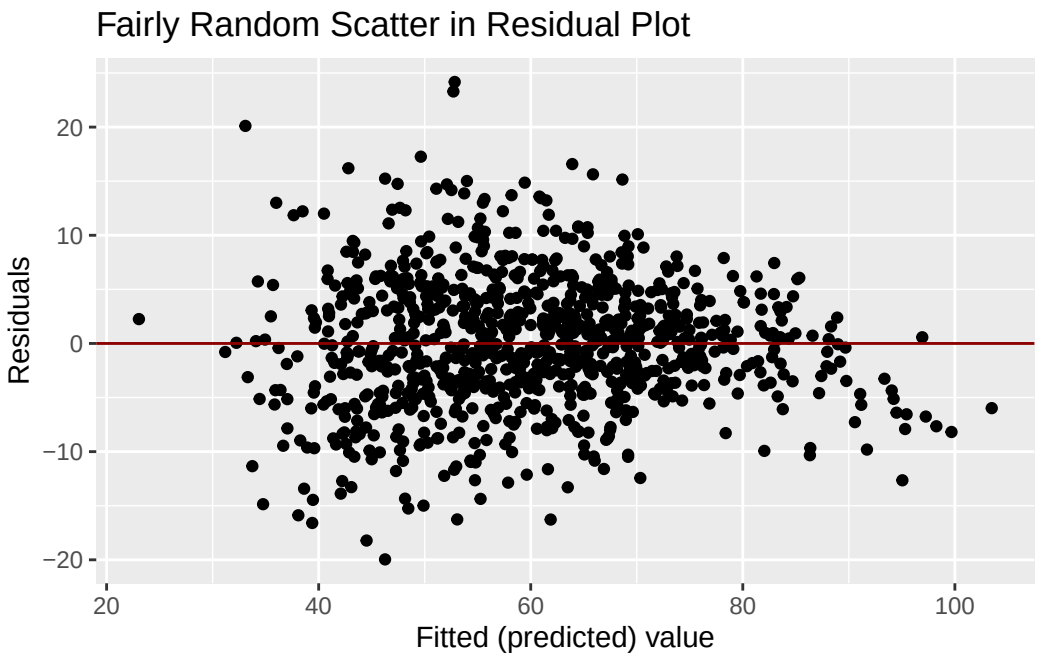
The dataset presented percentages that are greater than 95 as ">95" and percentages less than 5 as "<5." To make sure our model does not skip over these observations, we manipulated the dataset to change these percentages to 97.5 and 2.5, respectively, as they are the averages of 95-100 and 0-5. Also, in the dataset, if the number of students that met a condition were less than 10, it would just display "<10." We then manipulated the dataset again to change these counts to 5, the average of 0 and 10. In the dataset, a * indicates a school does not have tested grades/sufficient data for reporting, and these were changed to NA's in our model. Rows that did not have sufficient data for the 'All Students' subgroup are excluded from this file. In addition, we didn't include the demographic of American Indian in our model, as there was too little data present in the file because of how small their population is. Finally, we filtered out the State of North Carolina, which was listed as a school, as this was the average of all the schools present and we wanted to investigate individual school demographics.

In our manipulated dataset, to look at and compare an overall assessment for all public NC schools, we included the percentage and number of students that are grade level proficient in all EOG/EOC subjects for all, female, male, American Indian, Asian, Black, two or more races, white, economically disadvantaged (EDS), limited English proficient, disabled, and academically or intellectually gifted students for each school.

```
Call:
lm(formula = all ~ female_dem + male_dem + asian_dem + black_dem +
    hispanic_dem + white_dem + EDS_dem + LEP_dem + SWD_dem +
    AIG_dem + white_comp, data = nc_school)

Residuals:
     Min       1Q   Median       3Q      Max
-19.9618  -3.8129  -0.0862   3.8170  24.1717

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)
(Intercept)             251.201098  47.756557   5.260 1.80e-07 ***
female_dem               -1.797082   0.489232  -3.673 0.000253 ***
male_dem                 -1.963939   0.463676  -4.236 2.51e-05 ***
asian_dem                 0.192906   0.054991   3.508 0.000474 ***
black_dem                -0.000600   0.033876  -0.018 0.985874
hispanic_dem              0.120169   0.040113   2.996 0.002812 **
white_dem                 0.169690   0.041335   4.105 4.40e-05 ***
EDS_dem                  -0.244715   0.015923 -15.369  < 2e-16 ***
LEP_dem                  -0.053874   0.010088  -5.340 1.17e-07 ***
```

```
SWD_dem                    -0.080386   0.008947  -8.985  < 2e-16 ***
AIG_dem                     0.506832   0.033728  15.027  < 2e-16 ***
white_compbmid white        2.372891   0.809075   2.933 0.003443 **
white_compminority white    0.185998   1.450784   0.128 0.898015
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.081 on 912 degrees of freedom
  (1753 observations deleted due to missingness)
Multiple R-squared:  0.8208,    Adjusted R-squared:  0.8184
F-statistic: 348.1 on 12 and 912 DF,  p-value: < 2.2e-16
```

```r
schools_aug <- augment(schools)
schools_aug |>
  ggplot(
    aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, color = "darkred") +
  labs(x = "Fitted (predicted) value",
       y = "Residuals",
       title = "Fairly Random Scatter in Residual Plot")
```



```
schools_aug |>
  ggplot(
    aes(sample= .resid)) +
  stat_qq() +
  stat_qq_line() +
  labs(x = "Theoretical Quantiles",
       y = "Sample quantiles",
       title = "Low Deviation for QQ Plot")
```