

# Reverse Engineering Segregation in North Carolina Public Schools

Jake Drew\*, Olivia Leeson†

\*DataScience@SMU

Southern Methodist University

jakemdrew@gmail.com, oleeson@smu.edu

**Abstract**—This research uses data science and classification techniques to identify and explore educational attainment disparities for highly minority segregated public school campuses in North Carolina. From elementary through high school levels, we find that the quality of education provided to students in highly minority segregated schools is notably different than that provided to students attending both non-highly and highly majority segregated public schools. We demonstrate that highly segregated public school campuses account for over 32% of all North Carolina public schools. Isolating highly majority and minority segregated public elementary, middle, and high school campuses, we build classification models to accurately distinguish between highly majority and minority segregated campuses. Removing data on student body race from the inputs, we are able to predict whether a school is highly minority segregated with an average of 96.8% accuracy. We consider trends in the quality of education for each segregated campus type across the 2014 – 17 academic school years showing that in many cases the educational attainment gap for highly minority segregated students is widening. We provide all code, models, and charts produced during this research in html formatted iPython Notebooks for further review and consideration [6]–[8].

## I. INTRODUCTION

The state of North Carolina is known for its efforts in leading the charge toward school integration during the 1980s following the 1964 Civil Rights Act [2]. By 1980, only 4.8% of black students and 0.8% of Hispanic students attended highly segregated schools with a greater than 90% white population [26]. North Carolina was able achieve integration through specific state and county initiatives born in the courts to integrate schools based on geographical attendance zones [2]. However, since the 2000s, these initiatives have waned. The use of race in student assignments policies has halted, leading to a resegregation of public schools in the state [2].

The Supreme Court of North Carolina ruled that every child has the constitutional “right to a sound basic education” in 1997 [1]. Currently, the segregation of public schools has not been proven in the courts to be linked to the ability of school districts to provide students with a sound basic education [2]. However, public school performance data collected in the school years between 2013 and 2017 show a clear disparity in educational attainment between majority and minority segregated campuses in North Carolina. On average, highly minority segregated campuses perform worse on standardized

tests and end-of-grade assessments compared with their non-highly minority and majority segregated counterparts.

The trend toward resegregation of public schools began following a 2001 Fourth Circuit Court of Appeals ruling which ordered districts to stop using race in pupil assignments [29]. This marked the end of mandatory busing programs employed in counties such as Charlotte-Mecklenburg and Wake [29]. Busing programs worked for integration measures as some neighborhoods are already segregated by race and poverty lines. In 2013, the UNC Center for Civil Rights conducted a report to explore the relationship between these “cluster residents” and their public schools [21]. The study defined cluster residents as those living in adjacent census blocks where at least 75% of residents are non-white. Their study suggested a relationship between these residents and the likelihood of their public school failing, having a racially identifiable student enrollment (segregated), and having a high-poverty student population. 48% of students in black cluster residencies attended an elementary school that was failing compared to the state average of 19% [21]. Additionally, 68% of students in black cluster residencies attended a high-poverty school compared to the state average of 33% [21].

Over 60 years of social science research finds that racially segregated schools remain decidedly unequal [2]. The inequality of quality in racially segregated schools is reflected in teachers, supplies, curriculum, and disciplinary policy. Studies have found that racially and socioeconomically segregated school campuses comprise a disproportionate share of teacher inexperience and turnover, lower peer group performance, substandard facilities, and limited supplies [2]. These underfunded schools also typically provide lower teacher salaries than their counterparts, and are more likely to employ fewer fully qualified teachers [18].

Segregated and low-income schools are also shown to use less challenging curricula [28] and offer less AP and early college credit courses [27]. Moreover, minority segregated school students are disciplined harsher, and expelled at higher rates than students of predominately white schools [2], facing higher dropout rates [3] and lower levels of attainment during post-secondary education [5]. The negative influence of segregation is also reflected by the fact that schools with higher concentrations of poverty influence student achievement greater than that of an individual student’s own socioeconomic status [4].

This research explores highly segregated campuses in North Carolina between 2014-2017 through the creation of classification models that predict whether a public school is highly majority or highly minority segregated based on school level factors excluding race. By reverse engineering the properties of a highly segregated school, we can identify the differences between the inputs that calculate whether a school is highly segregated. Our findings suggest that educational differences between highly minority and highly majority segregated campuses include educational attainment disparities and differences in overall quality of education.

We discuss the current trend in student body demographics of North Carolina and the presence of segregation in public schools in Section II. We discuss the data consolidated from the North Carolina Department of Public Instruction in Section III. In Sections IV and V, we present the classification models that accurately predict highly segregated schools at the elementary, middle, and high school levels. In Section VI we explore the differences in educational attainment between highly minority and highly majority segregated student populations for all three school types. Finally, in Section VII we utilize feature importance analysis to discuss how the quality of education varies between highly majority and highly minority segregated public elementary, middle, and high schools.

## II. HIGHLY SEGREGATED PUBLIC SCHOOL CAMPUSES IN NORTH CAROLINA

The percentage of North Carolina public schools that are racially isolated ( $>75\%$  students of color) has increased by 5% between the 2006-2007 and 2016-2017 school years according to a recent report by the North Carolina Justice Center [19]. The factors behind the increase in segregation of public schools may be attributed to a combination of pupil assignment policy changes and demographic changes of the overall student population. The cessation of using race in pupil assignments after 2001 disbanded busing programs aimed at increasing racial integration. Meanwhile, the share of minority race children in North Carolina's student population has climbed due to migration and birthrates [2]. As of 2013, racially identifiable elementary schools were located in closest proximity to 63% of all North Carolina residents, 79% for minority cluster residents, and 81% for African American cluster residents [17].

Our research shows that both the number of highly majority and highly minority segregated school campuses has remained relatively constant between 2014 and 2017 at the elementary, middle, and high school levels. Figure 1 shows a slight increase in the total number of middle and high school campuses  $< 80\%$  segregated over this four year period. During the 2016 – 17 academic school year, highly segregated public school campuses accounted for 32.29% of the 2264 public school campuses we reviewed. Table I shows that North Carolina's highly segregated public school campuses make up 35.41%, 30.42%, and 25.96% of all public elementary, middle, and high school campuses respectively.

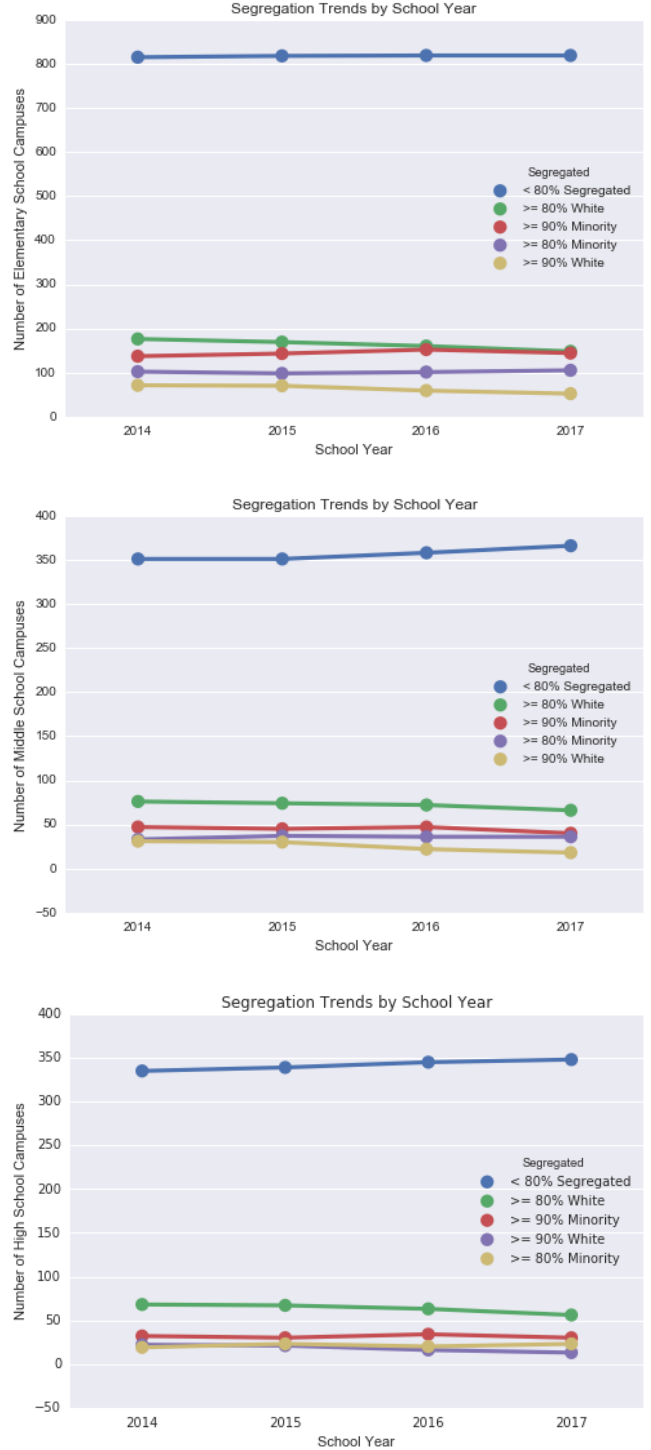


Fig. 1. Changes in Highly Minority and Majority Segregated Schools 2014 - 2017

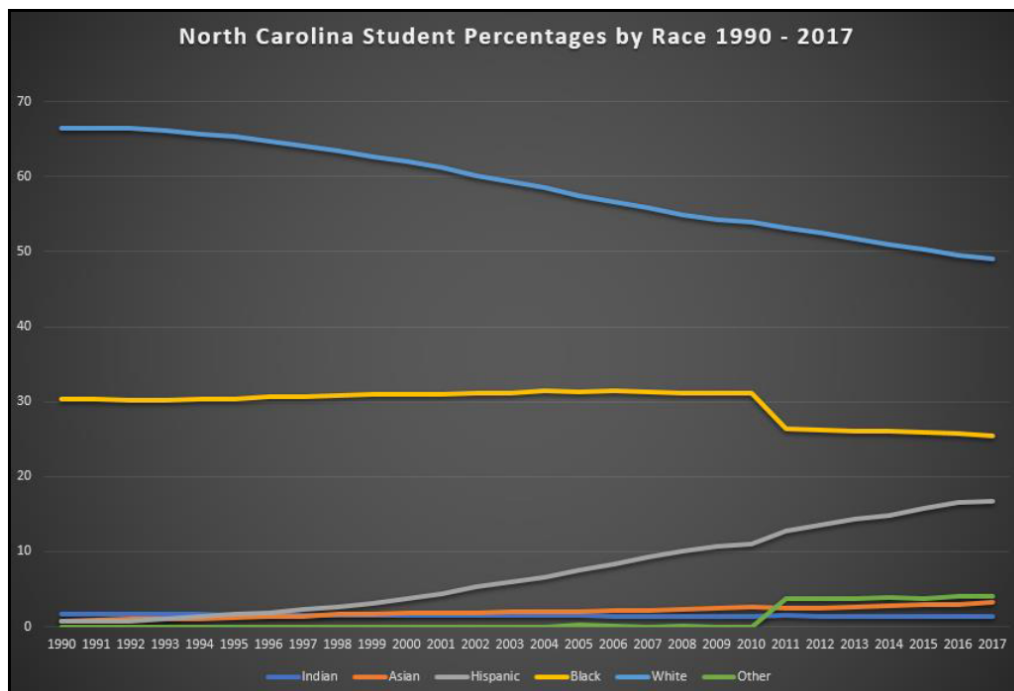


Fig. 2. North Carolina Student Percentages by Race 1990-2017 [25]

#### A. Changes in Student Body Demographics

Data contained in the North Carolina Public School's Statistical Profile Database [24], [25] provides detailed insights on the history of student body racial compositions at the school level. Overall, student body populations in the State of North Carolina have increased 142% over the past 28 years. Statewide, the student body population has increased from 1.078 million in 1990 to 1.531 million in 2017.

While all race memberships have increased during this period, the Hispanic and Asian populations have seen substantial growth at 3622% and 546% respectively. In 1990, there were 7,100 Hispanic children attending North Carolina Public schools. This number has grown to 257,147 during the 2016-2017 school year. The Asian student body membership has grown from 8,938 to 48,781 during the same period.

Figure 2 illustrates North Carolina student body membership percentages by race. Since 1990, membership percentages in white (-17.4%), black (-4.9%), and Indian (-0.3%) subgroups have decreased or remained flat while Hispanic (16.1%), Asian (2.4%), and Other (4.1%) student memberships have increased. Other race memberships currently include the Pacific Islanders and Mixed-Race subgroups. Statistics for these two memberships were not available until 2010. Since 2010, other student memberships have increased by 3.9%. Both the black and Other racial subgroups show similar positive and negative changes during 2010. This may indicate that a majority of the Mixed-Race subgroup was previously defined as black prior to 2010.

While North Carolina public school student populations and all racial subgroups have increased in size since 1990,

Figure 2 shows that the white subgroup has decreased from 66.5% of the statewide North Carolina student population to 49.1% during 2017. This is predominately due to the growth in the Hispanic (0.7% to 16.8%) and Asian (0.8% to 3.2%) subgroups during the same period. In addition, two new racial subgroups were added during 2010 (Pacific Islander, and Two or More Races) which now account for 4.1% of the student population during 2017.

Changes in the racial composition of North Carolina's student body are primarily driven by migration patterns and a change in birthrates for both white and minority subgroups [2]. During 2010, live births per 1000 people for whites (10.9) was lower than blacks (13.9) and less than half that of Hispanics (23.4) [2]. This trend is also reflected in the Figure 3's 2011 - 2017 student body racial composition with notable minority increases at the elementary and middle school campuses. Selected Vital Statistics for North Carolina [20] show that the 2016 birth rate per 1000 people for whites, blacks, and Hispanics are 10.3, 12.8, and 20.3 respectively.

#### B. A New Perspective on Majority-Minority

A majority-minority school is typically defined as any public school in the state of North Carolina which comprises less than 50% membership in the white racial subgroup. This definition of majority-minority schools no longer represents public school campuses with a highly segregated student body racial composition. The changes in student demographics between 2011 and 2017 suggest that North Carolina should be seeing more school campuses becoming majority-minority.

The white majority student body subgroup in North Carolina has consistently decreased on a percentage basis (53.1% to

North Carolina Highly Segregated Schools 2016-17				
Campus	Total	< 80% Segregated	>= 80% Segregated	Percentage
Elementary	1268	819	449	35.41%
Middle	526	366	160	30.42%
High School	470	348	122	25.96%
Totals	2264	1533	731	32.29%

TABLE I  
NORTH CAROLINA PUBLIC SCHOOL CAMPUSES >80% SEGREGATED DURING THE 2016-17 ACADEMIC SCHOOL YEAR.

49.1%) between 2011 and 2017. During this same period, statewide minority subgroups have grown from 46.9% of the student body population to 50.9%. It is significant to note that in 2017, Caucasians still represent the largest racial subgroup by a substantial majority (49.1%) with black children representing the second largest student body racial subgroup at 25.5%.

The statewide percentage of majority-minority schools in North Carolina has increased from 44.3% in 2011 to 49.1% in 2017. As of 2017, 1,200 schools qualified as majority-minority out of 2,446 North Carolina public schools reporting student body memberships by race [24], [25]. Out of the 2,579 active school profiles identified in the North Carolina report card database [23], 166 or 6.44% of these profiles were missing student body membership statistics by race in the North Carolina Public Schools Statistical Profile Database [24], [25]. This included 160 charter schools and 6 vocational schools. While the total students attending these 6 vocational schools is unknown, the 160 charter schools enrolled a total of 80,546 or 5.26% of NC students during 2016-2017. The majority-minority school percentages shown in Figure 3 do not reflect the impact of North Carolina charter schools.

According to Figure 3, the North Carolina majority-minority school percentage is around 4.3% higher at the elementary and middle school levels. Minority student subgroups take a majority share of both the elementary and middle school student populations. There were 1,284 out of 1320 active North Carolina elementary school profiles identified providing racial subgroup data during 2017. Figure 7 shows that 53.3% of these students are minorities and 46.7% are white. If all 1,284 of these elementary schools had student body populations which exactly matched the statewide racial subgroup distributions, then all elementary schools in the State of North Carolina would be majority-minority schools. This goes to show that the entire student population is not evenly distributed by race. For 35.41% of public schools, the distribution of students is highly segregated.

### C. Highly Minority Segregated Schools and Economically Disadvantaged Student Populations

Highly minority segregated public schools and economic disadvantaged student populations are linked. Looking at majority-minority defined schools, Figure 4 reflects levels of segregation in under performing public school campuses across North Carolina. Specifically, campuses identified as Consistently Low Performing (101) and those eligible for

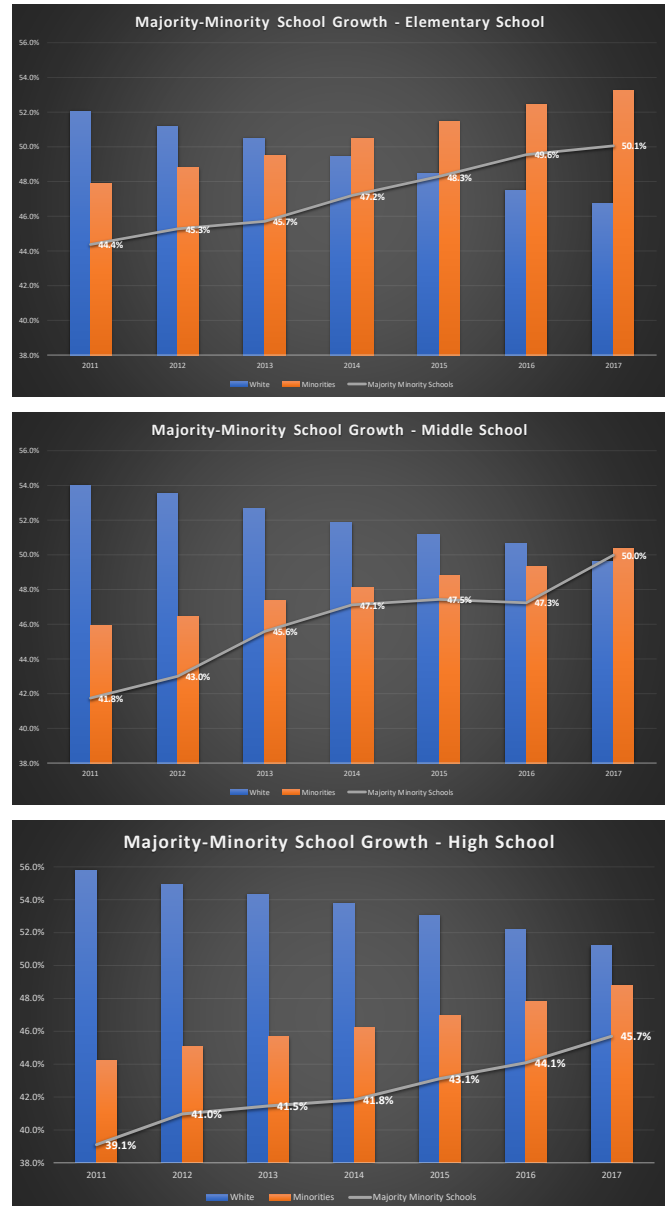


Fig. 3. Majority-Minority School Growth by Campus 2011-2017 [25]

flexibility under the ESEA - Elementary and Secondary Education Act (195) have racial compositions which include underrepresented white subgroup populations.

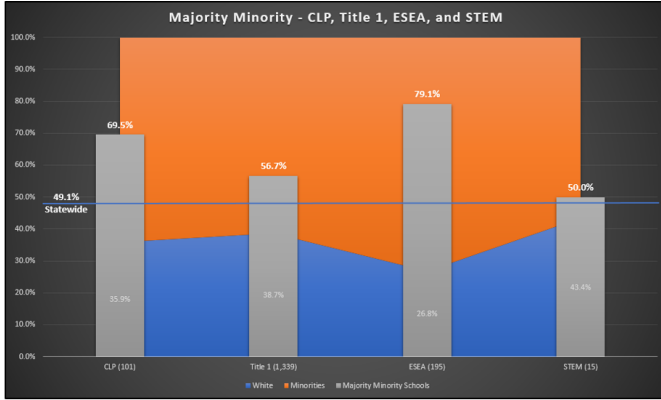


Fig. 4. Racial Compositions for CLP, Title 1, ESEA, and STEM Campuses during 2017

The white subgroup for CLP, Title 1, ESEA, and STEM campuses make up 35.9%, 38.7%, 26.8%, and 43.4% of the student body respectively. The 2017 statewide majority-minority school percentage reflected in Figure 4 is currently 49.1%. However, public schools classified as consistently low performing (70) and ESEA campuses (201) have drastically higher majority-minority school percentages which are 69.5% and 79.1% respectively. The link between segregation and socioeconomic status discovered in our findings are described in section VII.

### III. SCHOOL PERFORMANCE DATA: EXPLORING HIGHLY SEGREGATED PUBLIC SCHOOL CAMPUSES

We explore key differences in the quality of education for students attending highly minority and majority segregated public school campuses using data from the North Carolina Public School Report Cards [23] and Statistical Profiles databases [22]. The *Belk Endowment Educational Attainment Data Repository for North Carolina Public Schools* [9] currently consolidates data from 25 different School Report Card and Statistical Profile data structures for the 2014 – 17 academic school years.

The repository consolidates all data collected into one record per public school campus at the elementary, middle, and high school levels. The repository is open source and discloses all data transformations from each data structure’s origination URL to the final consolidated output. Data roughly falls into the following categories of information: school performance, educator experience and effectiveness, personnel experience and licensure, student demographics, classroom environment, school environment, and funding. Educator experience data contains information such as the percentage of teachers with 0 – 3 years of experience. Environment data comprises information on suspensions and crime ratios as well as data about wireless access points and student-to-internet-connected-computer ratios.

Statistical Profiles [22] data reports student body racial compositions for each public school campus in North Carolina. For the 2016 – 17 academic school year, student body racial compositions by gender included statistics for: Indian, Asian, Hispanic, black, white, Pacific Island, and students of two or more races. In addition, public high school campuses in North Carolina maintained 96 different student body racial composition features for certain metrics. This data includes the percentage of students of each racial subgroup who achieved End of Course benchmarks, College and Career Ready benchmarks, ACT scores, SAT scores, and Graduation Rates.

Data contained within the Belk repository’s consolidated data structures allow us to isolate highly segregated public school campuses. We define *highly minority segregated* campuses as any public school campus with a student body racial compositions including  $\geq 80\%$  minorities, or non-white students. We define *highly majority segregated* campuses as any public school campus with a student body racial composition including  $\geq 80\%$  majority, or white students. In North Carolina, our research shows that highly majority and minority segregated campuses make up 35.4%, 30.4% and 25.9% of all public elementary, middle, and high campuses respectively.

### IV. REVERSE ENGINEERING HIGHLY SEGREGATED CAMPUSES

We isolate highly majority and minority segregated public school campuses to determine how accurately each population may be identified using Logistic Regression classification. Reviewing data sets from the 2014 - 2017 academic school years, separate models are created for public school campuses at the elementary, middle, and high school levels. This is necessary to accommodate for the different data features provided by each campus type. For example, data for elementary campuses would not include the average size of a 12th grade class.

Prior to building the models, all student body racial composition data elements for each public school campus are removed. A model which accurately distinguishes between highly majority and minority segregated public school campuses is not particularly useful for making predictions, but it can indicate that significant educational disparities exist between the two populations. Furthermore, using Logistic Regression in combination with other machine learning techniques provides us with a set of model coefficients from which we can determine how the individual models accurately distinguish between these two highly segregated campus types. Furthermore, we can identify significant differences in the quality of education received by students attending highly majority segregated, highly minority segregated, and non-highly segregated public school campuses by performing feature importance analysis for each model created.

#### A. Overview of Methodology

We create and test independent models for elementary, middle, and high school campus types using the *Machine*



*Learning* data sets from the Belk Endowment Educational Attainment Data Repository for North Carolina Public Schools [9]. The repository performs the following transformations to prepare each data set for creating classification models:

- Columns that have the same value in every single row are deleted.
- Columns that have a unique value in every single row (all values are different) are deleted.
- Empty columns (all values are NA or NULL) are deleted.
- Numeric columns with more than 60% of missing values are deleted.
- The remaining numeric, non-race columns with missing values are imputed / populated with a value of zero. In many cases, schools are not reporting values when they are missing. However, mean imputation or some other more sophisticated strategy might be considered here.
- Categorical / text based columns with > 25 unique values are deleted.
- All remaining categorical / text based columns are one-hot encoded. In categorical columns, one-hot encoding creates one new binary field per unique value in the target column, converting all categorical columns to numeric data types.
- Duplicated or highly linear dependent columns with > 95% correlation are deleted.
- Any raw data features with masking applied were mapped to the following values: "\*" : 0, ">95" : 100, "<5" : 0, "<10" : 5.

Non-highly segregated campuses are removed from each data set prior to model creation and validation. The remaining data is then divided into ten testing and training cross-validation folds using a 90% random selection of data for training and the remaining 10% for testing. Next, we isolate and remove any student body racial composition features from each data set. For example, there were 96 out of 405 total features which were removed from the public high school data set. These features included factors such as average SAT, ACT, College and Career Ready Benchmarks, and Graduation Rates tracked by student body racial compositions. A binary response variable is also created to indicate *TRUE* when a public school campus is  $\geq 80\%$  minority segregated and *FALSE* when a public school campus is  $\geq 80\%$  majority segregated.

For each public elementary, middle, and high school segregation data set, we use the following steps to develop an optimized model for distinguishing between highly majority and minority segregated schools:

- 1) *Recursive Feature Elimination* [15] is performed to recursively select a smaller set of features for most accurately distinguishing between highly majority and minority segregated schools.
- 2) The remaining features are transformed [16] by removing the mean and scaling to unit variance. This allows each model's coefficients to be compared without concern for large variations in each individual feature's

dynamic range. Comparing model coefficient's for unscaled years of experience and average SAT scores could produce misleading results when performing feature importance analysis.

- 3) A grid search [13] is performed to identify the optimal parameters for each Logistic Regression [14] model created. The grid search performs cross-validation using various parameter values for the Cost, Class Weight, and Maximum Iterations to identify the most accurate parameter combination.
- 4) Ten-fold Cross-Validation [12] is performed using the scaled features and top performing model to determine a final model accuracy and review important features.

A very accurate model would indicate that the differences in school level factors such as test scores and teacher experience are clear enough that we can tell whether a school is highly minority segregated without even considering the race of the student body. Given accurate models, we can then consider which factors are most important in predicting whether or not a school is classified as highly majority or highly minority segregated by reviewing the models' coefficients.

## V. RESULTS: CLASSIFYING HIGHLY MINORITY AND MAJORITY SEGREGATED CAMPUSES

Recursive Feature Elimination is used to select a subset of features producing the highest classification accuracy for distinguishing between highly minority and majority segregated public elementary, middle, and high school campuses. Once the grid search has identified the optimal model parameters, a final model is produced for each public school campus type. Each model's features are ranked by the absolute value of the model's coefficients, and the top 75 features with the largest absolute coefficient values are reviewed.

Table II shows the average ten-fold cross-validation results for models distinguishing between highly minority and majority segregated public school campuses during the 2016 – 17 academic school year. On average, we were able to create Logistic Regression classification models with 96.8% accuracy. Elementary school campuses showed the highest degree of accuracy at 97.9% with Middle and High School campuses at 97.5% and 96.8% respectively.

Such a high degree of classification accuracy indicates that there are distinct differences within the model's training data features for these two populations of schools. We utilize each model's coefficients to perform feature importance analysis and uncover substantial inequalities in the education that is disproportionately provided to highly minority segregated public school students in the state of North Carolina.

To visualize the differences between the features used to create the models, Kernel Density Estimation is performed to approximate the probability density function for each significant continuous feature given each population: majority segregated vs. minority segregated campuses. We also compare the population means for highly segregated and non-highly segregated public school campuses. Our complete research

Classifying Highly Segregated Schools				
Campus	Total	$\geq 80\%$ Minority	$\geq 80\%$ Majority	Classification Accuracy
Elementary	449	249	200	97.9%
Middle	160	76	84	97.5%
High School	122	53	69	95.0%
Totals	731	378	353	96.8%

TABLE II

NORTH CAROLINA HIGHLY MINORITY AND MAJORITY SEGREGATED CAMPUSES ARE CLASSIFIED WITH 96.8% ACCURACY DURING THE 2016-17 ACADEMIC SCHOOL YEAR

considers each significant feature across the 2014 – 17 academic school years for elementary, middle, and high school campuses [6]–[8].

## VI. FINDINGS: KEY PERFORMANCE DIFFERENCES IN HIGHLY MINORITY AND MAJORITY SEGREGATED CAMPUSES

### A. Highly Segregated Elementary Schools and Educational Attainment

We reviewed a total of 1268 elementary school campuses for the 2016 – 17 academic school year finding that 449 or 35.41% had student body racial compositions greater than or equal to 80% segregated. This included 249 highly minority segregated campuses and 200 highly majority segregated campuses respectively representing 19.6% and 15.7% of the public elementary school campuses reviewed. There were a total of 315 data features specific to public elementary school campuses. We identified and removed 83 fields containing data related to student body racial compositions. After performing Recursive Feature Elimination, 221 features were used in the final model which distinguished between highly minority and majority segregated elementary school campuses with 97.9% classification accuracy.

The model [11] identifies *lea\_wap\_num* (number of wireless access points in school district), *pct\_ed*s (percentage of students of economically disadvantaged status), *SBE District\_Sandhills* (Sandhills district schools), *lea\_sat\_avg\_score\_num* (average SAT score of the school district), and *lea\_licensed\_teach\_pct* (percentage of fully licensed teachers in the School District) as among the top features that influence whether an elementary school is classified as highly majority or highly minority segregated. It is interesting to note that the average SAT score of the school district is associated with elementary schools. It is possible that scores achieved at the high school level are an indicator of education at the elementary school level.

We find that highly minority segregated elementary schools under perform highly majority and non-highly segregated elementary schools. Figure 5 shows the count of highly segregated public elementary school campuses that received an overall school performance grade of D. The same disparity in school performance grades is reflected specifically in the Reading and Math subjects. In all three instances, less than 3.5% (7) of highly majority segregated schools received school performance grades of D. However, over 70, or 28.1%, of

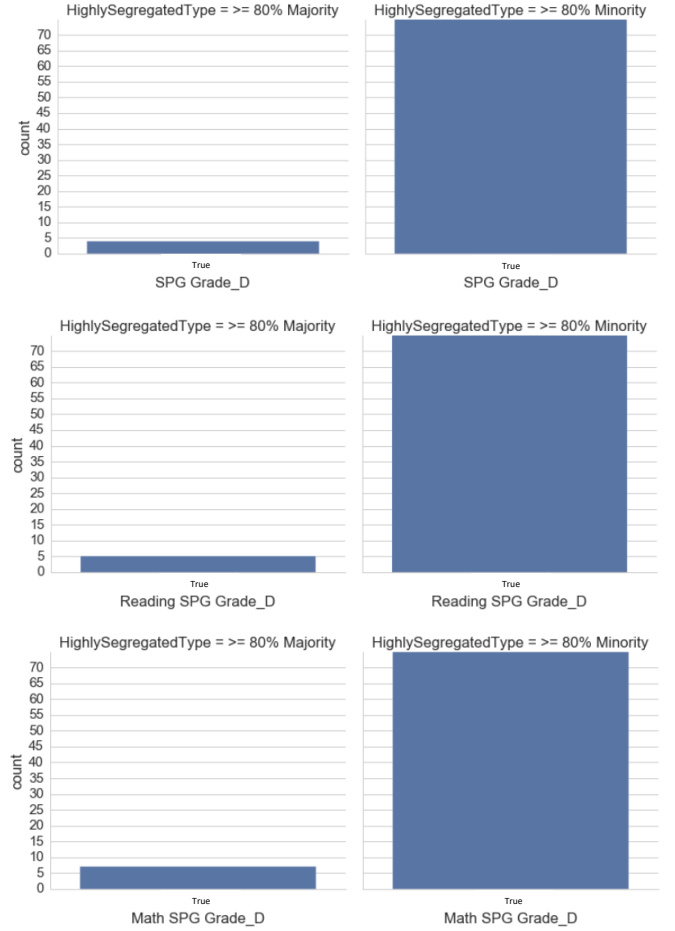


Fig. 5. Highly Segregated Elementary School Performance Grades 2016 - 2017

highly minority segregated elementary schools received school performance grades of D.

There is also evidence that highly minority segregated elementary school campuses are less capable of passing students to the next grade. During the 2016 – 17 academic school year, 14.8% of students attending highly minority segregated elementary school campuses were held back because they did not meet benchmarks for passing to the next grade. The mean percentage of students retained at highly minority segregated campuses is 1.92 times greater than other non-

highly segregated public elementary school campuses and 9.25 times greater than other highly majority segregated public elementary school campuses.

Figure 6 illustrates that the 2016 – 17 academic school year student retention numbers for some types of highly minority segregated campuses are an improvement from past years. However, the trend across the 2014 – 17 academic school years remains consistent with highly minority segregated campuses holding back students at much higher rates than other non-highly highly segregated and highly majority segregated elementary schools. In addition, highly minority segregated campuses with  $\geq 90\%$  minority student body racial compositions have increased the percentage of students held back because they did not meet benchmarks for passing to the next grade.

Highly minority segregated elementary school campuses also show distinct differences in attainment for numerous End of Grade (EOG) and College and Career Ready (CACR) benchmarks. The End of Grade 3 College and Career Ready All benchmark for the 2016 – 17 academic school year shows that highly minority segregated campuses perform 14.2% lower than non-highly segregated campuses and 22.5% lower than highly majority segregated campuses on average. During the fourth grade, this attainment gap widens to 14.6% lower than non-highly segregated campuses and 23.2% lower than highly majority segregated campuses on average.

Males attending highly minority segregated elementary schools perform 14.8% lower than non-highly segregated campuses and 20.3% lower than highly majority segregated campuses on average for 2016 – 17 EOG-CACR fourth grade Reading. Math EOG-CACR fourth grade benchmarks show similar results with males performing 15.8% lower than non-highly segregated campuses and 21.6% lower than highly majority segregated campuses on average for 2016 – 17.

#### B. Highly Segregated Middle Schools and Educational Attainment

Over the 2014 – 17 academic school years the number of highly segregated middle school campuses remains rather consistent. Table I shows that out of 526 middle school campuses reviewed during 2016 – 17, we identified 30.42% or (106) campuses with student body racial compositions which were  $\geq 80\%$  segregated. According to Table II there are 76 highly minority segregated and 84 highly majority segregated middle school campuses in North Carolina.

We considered a total of 406 data features for middle school campuses removing 99 fields containing data related to student body racial compositions. After performing Recursive Feature Elimination, 78 features were used in the final model which distinguished between highly minority and majority segregated middle school campuses with 97.5% classification accuracy.

The model [11] identifies *pct\_eds*, *lea\_nbpts\_num* (number of nationally board certified staff in the school district), *lea\_short\_susp\_per\_c\_num* (number of suspensions per 100 students in the school district), *lea\_sat\_avg\_score\_num*, and *SPG Score* (school performance grade score) as among the top

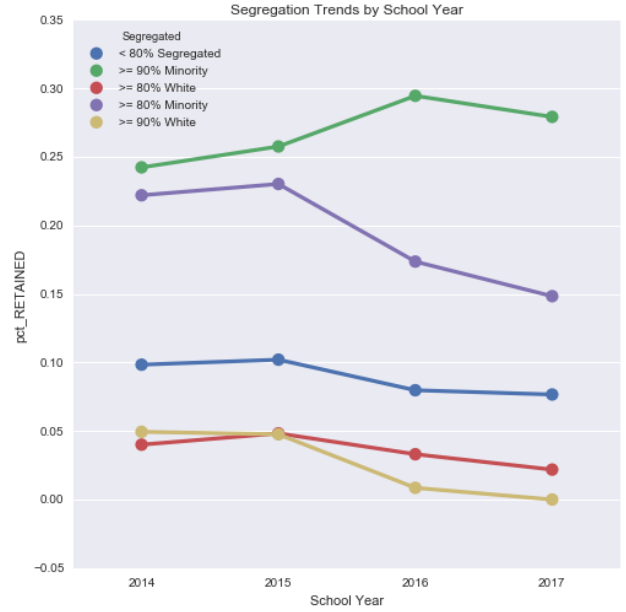


Fig. 6. Mean Percentage of Elementary School Students Retained by Segregated Campus Type 2014 - 2017

features that influence whether a middle school is classified as highly majority or highly minority segregated. We find similar results as with the elementary school model: the percentage of economically disadvantaged students and the average SAT score of the school district have a strong influence in predicting highly segregated schools.

Subsequent feature and kernel density analysis show that highly segregated middle school campuses have consistently lower school performance grades than both non-highly segregated and highly majority segregated middle schools. During 2016 – 17, non-highly segregated campuses had school performance grades that were 15.3% higher on average than highly minority segregated schools. For highly majority segregated campuses, this number was even higher at 21.3%. Figure 7 illustrates that school performance grades for highly segregated middle school campuses have consistently remained lower than both non-highly segregated and highly majority segregated middle schools across the 2014 – 17 academic school years. When considering Figure 7, it is clear that higher concentrations of minority segregation further degrade school performance. For example, Figure 7 shows that middle school campuses with  $\geq 90\%$  minority segregation have the lowest average school performance grades reviewed during the 2014 – 17 period.

As students are consolidated into highly minority segregated middle school campuses, the educational attainment gap also widens. At the middle school level, highly minority segregated campuses belong to school districts with an average SAT score that is 43.5 points less than non-highly segregated campuses and 79.3 points less than highly majority segregated campuses. Respectively, this is a 9.5 and 15.3 point decrease when



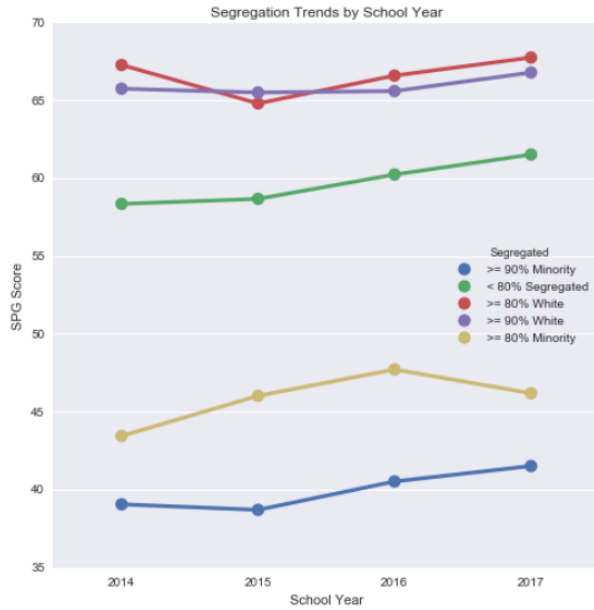


Fig. 7. Middle School SPG Grade by Segregated Campus Type 2014 - 2017

compared to highly segregated elementary schools.

Mean school performance grades for highly segregated middle school campuses are 15.3% lower than non-highly segregated campuses and 21.3% lower than highly majority segregated campuses. Since 2014, this gap has remained rather consistent, increasing by 0.4% for non-highly segregated campuses and decreasing by 2% when compared to highly majority segregated schools.

We also find notable differences in attainment levels for the economically disadvantaged with end of grade college and career ready benchmarks for all subjects being 1.67 times higher for highly majority segregated and 1.37 times higher for non-highly segregated middle school students. By 7th grade, highly minority segregated middle school students have end of grade reading benchmarks that are 1.62 times lower than highly majority segregated and 1.88 times lower than non-highly segregated middle school students.

Educational attainment gaps for middle school students remain between the 6th and 8th grades when considering end of grade college and career ready benchmarks for all subjects. During 6th grade, students attending highly segregated middle school campuses score 16.6% lower than non-highly segregated campuses and 20.4% lower than highly majority segregated campuses. By 8th grade, average attainment gaps for this benchmark increase to 22.4% when compared to highly majority segregated campuses and decrease by 0.9% for non-highly segregated campuses.

When considering gender specific benchmarks, we find differences in end of 7th grade reading benchmarks for females with highly minority segregated campuses scoring 30.7, non-highly segregated campuses scoring 47.4, and highly majority segregated campuses scoring 53.5, on average. Males also

show differences in 7th grade reading benchmarks with highly minority segregated students attaining scores which are 16.6 and 21.1 points lower on average respectively. End of 6th grade math benchmarks also show disparities for females with highly minority segregated campuses scoring 27.2, non-highly segregated campuses scoring 42.2, and highly majority segregated campuses scoring 43.6, on average.

While non-highly segregated and highly majority segregated 7th grade students with disabilities have mean scores are within less than 1% on end of grade math benchmarks, disabled students attending highly minority segregated campuses have math benchmark scores which are less than 57% of both the scores achieved by other campus types. Likewise, scores on end of 7th grade reading benchmarks for students with disabilities are 67% of the scores at non-highly segregated campuses and only 60% of the scores at highly majority segregated campuses.

Only 24.7% of students attending highly minority segregated campuses are ranked as *proficient* by the end of 6th grade, while students attending other campus types have a far greater opportunities for proficiency. Over 43% of non-highly segregated and 46.4% of highly majority segregated students are proficient by 6th grade. It is clear that students of highly minority segregated public schools are at a disadvantage compared to the states' average and their highly majority segregated public school counterparts.

### C. Highly Segregated High Schools and Educational Attainment

We reviewed a total of 470 public high school campuses for the 2014 – 17 academic school years. Table I illustrates that 25.96% of the high school campuses considered had student body racial compositions which were  $\geq 80\%$  segregated. In Table II, we see that 53 or 11.28% of public high school campuses have student body racial compositions with  $\geq 80\%$  minorities and 69 or 14.68% of campuses have student body racial compositions which are  $\geq 80\%$  white.

We reviewed 406 high school data features removing 96 features containing data specific to student body racial compositions. After performing Recursive Feature Elimination, 75 features were used in the final model which distinguished between highly minority and majority segregated high school campuses with 95.0% classification accuracy.

The model [11] identifies *lea\_short\_susp\_per\_c\_num*, *lea\_avg\_student\_num* (average number of students in the school district), *lea\_sat\_avg\_score\_num*, *pct\_eds*, and *short\_susp\_per\_c\_num* (number of suspensions per 100 students in the school) as among the top features that influence whether a high school is classified as highly majority or highly minority segregated. These results are congruous with the elementary and middle school models. In at least two of the three models, the number of short term suspensions per 100 students and average SAT score of the school district are identified as top predictors for segregation. In all three models, the percentage of economically disadvantaged students is a top important feature.

Females attending highly minority segregated high schools enroll in an institute of higher education within 16 months of receiving a regular high school diploma at similar rates when compared to both non-segregated and highly majority segregated high schools. However, we find that during 2017 only 17.8% of males attending highly minority segregated high schools enrolled in an institute of higher education within 16 months of receiving a regular high school diploma. This was 3.9% lower than non-highly segregated high schools and 4.6% lower than highly majority segregated high schools.

When considering average ACT Scores, highly minority segregated high schools showed substantial differences during both the 2016 – 17 and 2015 – 16 academic school years. During 2016 – 17, ACT performance gaps widened for both non-highly segregated and highly majority segregated high schools. ACT scores at non-highly segregated high schools were 1.45 times higher during 2016–17 and 1.26 times higher during 2015–16 than highly minority segregated high schools. These numbers are even greater for highly majority segregated schools which performed 1.6 times higher on average during 2016 – 17. This was also an increase from 1.33 times higher during 2015 – 16.

Students attending highly minority segregated high schools also show deficits in average SAT scores at the district level scoring 29.7 points lower than non-highly segregated high schools and 71.2 points lower than highly majority segregated schools, on average. Similar district level score deficits exist for the 2014 – 17 academic school years.

The average high school performance scores reflected in Figure 8 illustrate that as minority segregation levels increase at the campus level school performance scores decrease. Figure 8 shows that highly minority segregated campuses with  $\geq 90\%$  segregation levels have the lowest average school performance scores of all populations considered during each of the 2014 – 17 academic school years. During 2016 – 17, highly minority segregated high schools scored 14% lower than highly majority segregated and 10.6% lower than non-highly segregated high schools. On average, the school performance score gaps for both highly majority and non-highly segregated high schools have increased from 10% and 7.3% during 2014.

We find that both 4 and 5 year graduation rates are lower for many types of students that attend highly minority segregated high school campuses. Our models utilized graduation rates for students with limited English proficiency, students with disabilities, academically or intellectually gifted students, and male students to distinguish between highly minority and majority segregated campuses. When considering the 5 year graduation rate for all students, highly minority segregated high school campuses have graduation rates which were 10.3% lower than highly majority segregated campuses and 8.8% lower than non-highly segregated campuses. Most alarmingly, these gaps were much lower during previous years at 2.4% and 1.6% respectively during 2014.

Likewise, 4 year cohort graduation rate scores show attainment gaps for highly minority segregated high school students.

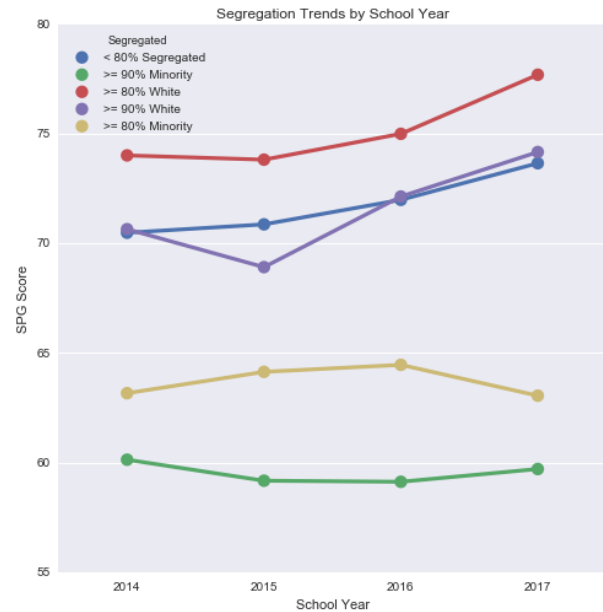


Fig. 8. Mean School Performance Scores for Public High Schools by Segregated Campus Type 2014 - 2017

During 2016 – 17, highly majority segregated student graduation rates were 8.4% higher and non-highly segregated high school student graduation rates were 6% higher than highly minority segregated schools. This attainment gap doubled from the prior academic school year for highly majority segregated students and increased 27.6% for non-highly segregated students. These findings suggest that educational disadvantages of highly majority segregated schools leads into post-secondary education.

## VII. FINDINGS: KEY EDUCATION QUALITY DIFFERENCES IN HIGHLY MINORITY AND MAJORITY SEGREGATED CAMPUSES

The feature importance results of our three models identify school level factors outside of average academic performance that point to differences in education quality between highly minority and highly majority segregated school campuses. Highly minority segregated school campuses are comprised of a much higher proportion of economically disadvantaged students than highly majority segregated schools. Highly minority segregated schools also have higher rates of short-term suspensions, fewer fully licensed teachers and nationally board certified staff, and have more teachers with less experience than their highly majority segregated counterparts. These findings suggest that students in highly minority segregated schools are receiving an unequal quality of education.

### A. Highly Segregated Elementary Schools and Quality of Education

1) *Highly Segregated Elementary Schools and Economically Disadvantaged Students:* When compared to both non-highly segregated and highly majority segregated elementary

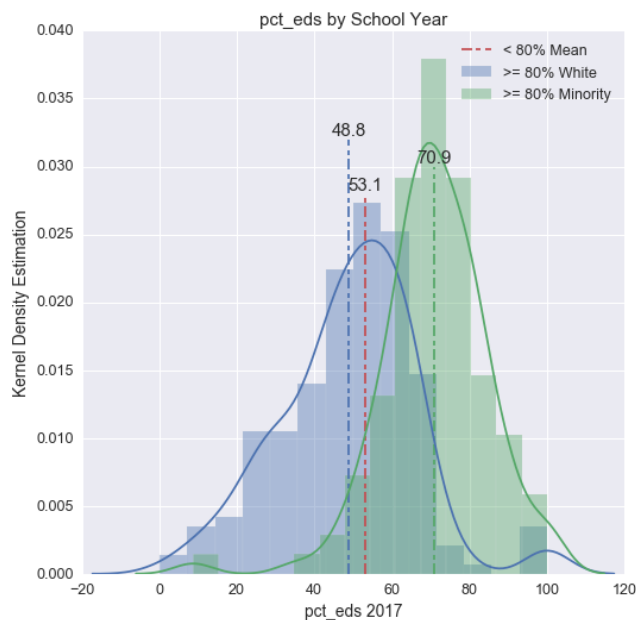


Fig. 9. Campus Percentage of Economically Disadvantaged Elementary School Students 2016 - 2017

schools, we find that the percentage of economically disadvantaged students is substantially greater at highly minority segregated campuses. The mean percentage of economically disadvantaged elementary school students shown in Figure 9 illustrates that white students attending highly majority segregated schools, on average, experience an educational environment that contains 22.1% less economically disadvantaged students than highly minority segregated schools. Highly majority segregated students also experience 4.3% less economically disadvantaged students than other non-highly segregated elementary school campuses. Highly minority segregated students attend campuses which are made up of 70.9% economically disadvantaged students on average. This could be thought of as a disproportionate share of North Carolina minority students receiving an approximately 22.1% less privileged educational environment than students attending highly majority segregated schools and an approximately 17.8% less privileged educational environment than students attending non-highly segregated elementary schools.

2) *Highly Segregated Elementary Schools and School Districts:* Our findings also show that highly segregated elementary school campus types are highly concentrated within SBE Education Districts. For example, we identified 53 highly minority segregated elementary school campuses and only 5 highly majority segregated campuses in the Sandhills district. In stark contrast, the Northwest district contained 54 highly majority segregated elementary schools and less than 3 highly minority segregated schools. Similar concentrations of highly majority segregated campuses were also identified in the Western District, while the Northeastern district contains much higher concentrations of highly minority segregated

campuses. The Piedmont-Triad District contained rather equal amounts of highly segregated school campuses including 42 highly majority segregated and 45 highly minority segregated elementary school campuses.

We find that 11 out of the top 20 features used by our model to distinguish between highly minority and majority segregated elementary schools are specific to each school's local education agency or school district. Students attending highly minority segregated schools also belong local education agencies which provide a lower quality of education when compared to both non-segregated and highly majority segregated elementary school campuses. For example, highly segregated elementary schools reside in school districts which have a mean SAT score that is 34 points lower than non-highly segregated schools and 64 points lower than highly majority segregated schools.

3) *Highly Segregated Elementary Schools and Staffing:* Highly segregated elementary schools reside in school districts which have less fully licensed teachers, greater expense percentages for services, more teachers with 0 – 3 years experience, higher federal funding per pupil, higher 1 year teacher turnover, less teachers with 11 plus years experience, more average students than highly majority segregated schools, three times as many short term suspensions than highly majority segregated schools, and higher crime rates per 100 students than both non-highly segregated and highly majority segregated elementary school campuses.

We uncover a number of differences in elementary school teachers employed at highly minority segregated campuses when compared to both non-segregated and highly majority segregated campuses. Highly minority segregated campuses belong to school districts which employ less fully licensed teachers. There are also less fully licensed teachers employed at highly minority segregated elementary schools. Highly minority segregated elementary schools belong to districts that also employ a higher percentage of teachers with 0 – 3 experience and less teachers with 11 or more years experience. One year teacher turnover in these districts is also higher.

On average, teachers in highly minority segregated elementary schools have the highest teacher to student ratios educating 35.3 students per classroom. While district average student numbers per campus are over 95 students higher than highly majority segregated campuses, district average salary expenses are also the lowest for highly minority segregated schools. Highly minority segregated elementary campuses also have the highest percentage of teachers not meeting expected teacher student growth.

We find that the quality of education received by students attending highly minority segregated public elementary school campuses is sub-standard when compared to both non-highly segregated and highly majority segregated elementary schools in North Carolina. Educational deficits at these campuses are compounded when students move on to attend highly minority segregated middle schools.

## B. Middle Schools and Quality of Education

1) *Highly Segregated Middle Schools and Economically Disadvantaged Students:* The *privilege segregation* gap widens for highly segregated middle school students. The total number of middle school campuses (526) reflected in Table I shows a consolidation of 742 campuses when compared to the elementary school campus type (1268). As the total number of campuses decreases highly minority segregated middle school students are placed into campuses with even higher concentrations of economically disadvantaged students. Over 72.3% of students attending highly minority segregated middle schools are economically disadvantaged. This is 19.4% greater than non-highly segregated campuses at 52.9% and 24.9% greater than highly majority segregated campuses at 47.4%.

2) *Highly Segregated Middle Schools and Staffing:* Students attending highly minority segregated campuses also belong to school districts with a lower average percentage of Nationally Board Certified staff. During 2016 – 17, highly minority segregated middle school campuses belonged to school districts that have 1.66 times less Nationally Board Certified staff than non-highly segregated campuses and 2.03 times less Nationally Board Certified staff than highly majority segregated campuses. This trend is consistent across the 2014 – 17 period.

We also find that the percentage of Nationally Board Certified staff at individual highly minority segregated middle school campuses is 2.22 times lower than non-highly segregated schools and 2.45 times lower than highly majority segregated schools. This staff certification gap has also increased each year since 2014 from 1.73 and 1.85 times respectively. Alarming, the percentage of Nationally Board Certified staff at all North Carolina middle school campuses has steadily decreased each year since 2014. Highly minority segregated middle schools also tend to be larger in terms of student body than the state average or highly majority segregated schools. In 2017, highly minority segregated middle schools enrolled 613 students on average compared with 579 students in highly majority segregated schools. This difference of 34 students compounded with fewer nationally board certified staff means less resources are spread across a larger student body for highly minority segregated schools.

3) *Highly Segregated Middle Schools and Environment:* Highly minority segregated schools have higher rates of short term suspensions than non-highly and highly majority segregated public schools. In 2017, the average number of short term suspensions per 100 students was 38.2. Meanwhile, highly segregated majority schools gave out short term suspensions to only 7.5 out of 100 students on average, while non-highly segregated schools gave out 19.7 short term suspensions. The implications of the ratio of short term suspensions are that highly minority segregated students are more likely to miss class time due to suspensions, and thus miss out on the opportunity to learn or attain education at the same levels as highly majority segregated school students.

### C. High Schools and Quality of Education

1) *Highly Segregated High Schools and Economically Disadvantaged Students:* The ratios of economically disadvantaged students and *privilege segregation* remain steady throughout a highly minority segregated student's educational career. In 2017, there was a +22.1 difference in the percentage points of economically disadvantaged students in highly minority versus highly majority segregated elementary schools. In this same year, the difference on average for highly segregated middle schools was +25.1 percentage points. For high schools, highly minority segregated schools had +22.6 percentage points more economically disadvantaged students than highly majority segregated schools. By the time a highly minority segregated student reaches high school in North Carolina, there are 1.59 times more economically disadvantaged students at their high school than highly majority segregated campuses and 1.17 times more economically disadvantaged students than other non-highly segregated public school campuses.

2) *Highly Segregated High Schools and Staffing:* The impacts of *privilege segregation* are clear when considering the average number of students attending highly minority segregated campuses at the district level. On average, highly minority segregated students belong to school districts with 199.5 additional students per campus when compared to non-highly segregated campuses and 211.6 additional students per campus when compared to highly majority segregated campuses. Highly minority segregated high school campuses have far less nationally board certified staff with non-highly segregated campuses having 1.79 times more certified staff per campus during 2014 and 2.25 times more during 2017. This trend is similar for highly majority segregated campuses having twice as many certified staff per campus during 2014 and 2.62 times more during 2017.

3) *Highly Segregated High Schools and Environment:* The trend of short term suspensions shown in middle schools continues at the high school level. Highly minority segregated high schools belong to school districts that have substantially higher average short term suspensions. Students attending highly minority segregated high schools are suspended 2.44 times more often than highly minority segregated campuses and 1.72 times more often than non-highly segregated campuses at the district level. For minorities, this trend becomes even more pronounced at the individual school level when compared to the more privileged highly majority segregated campuses. During 2017, short term suspensions at highly minority segregated high schools were 1.71 times higher than non-highly segregated campuses and 3.2 times higher than highly majority segregated campuses. Such differences between highly majority and minority segregated campuses for short term suspensions have increased at both the district and campus levels since 2014.

## VIII. CONCLUSION

From the elementary through high school levels, we find that the quality of education provided to highly minority segregated students in North Carolina is substantially different than that of students attending non-highly and highly majority segregated

schools. Highly segregated public school campuses account for over 32% of all North Carolina public schools with the highest levels of segregation occurring at the elementary and middle school campuses. *Privilege segregation* seems to define the highly majority segregated public school campuses in North Carolina which outperform both non-highly and highly minority segregated schools in almost all areas of educational attainment.

Our machine learning models accurately distinguish between highly majority and minority segregated campuses with an average of 96.8% accuracy, even when removing all data features which contain any details of student body racial compositions. Performing feature importance analysis, each model's coefficients uncover where the primary educational differences exist between highly minority segregated, highly majority segregated, and non-highly segregated public school campuses. These educational differences include attainment measured by End of Grade/End of Course and standardized test scores along with quality differences measured by staffing and environment.

The trends in the quality of education for each segregated campus type across the 2014 – 17 academic school years show that in many cases, the educational attainment gap for highly minority segregated students is widening. We provide all code, models, and charts produced during this research in html formatted iPython Notebooks for further review and consideration [6]–[8]. We also provide an appendix [11] with additional model coefficient charts. Our hope is that this work will assist with data driven decisions in education and philanthropy to close the educational attainment gaps identified for highly minority segregated students in North Carolina.

## IX. ACKNOWLEDGEMENTS, SOURCE CODE, AND SUPPORTING DATA

This work was made possible by a grant from the Belk Endowment [10]. All data used in this research is open source and available in The Belk Endowment Educational Attainment Data Repository for North Carolina Public Schools [9]. The models, charts, and source code for this research are also available within the Belk NCEA Repository [6]–[8], [11].

## REFERENCES

- [1] Leandro v. state. 346 N.C. 336 (1997).
- [2] J. B. Ayscue and B. Woodward. Segregation again: North carolina's transition from leading desegregation then to accepting segregation now. *Civil Rights Project-Proyecto Derechos Civiles*, 2014.
- [3] R. Balfanz and N. Legters. Locating the dropout crisis, which high schools produce the nation's dropouts? where are they located? who attends them? report 70. *Center for Research on the Education of Students Placed at Risk CRESPAR*, 2004.
- [4] G. Borman and M. Dowling. Schools and inequality: A multilevel analysis of coleman's equality of educational opportunity data. *Teachers College Record*, 112(5):1201–1246, 2010.
- [5] E. M. Camburn. College completion among students from high schools located in large metropolitan areas. *American Journal of Education*, 98(4):551–569, 1990.
- [6] J. Drew. Elementary school segregation trends 2014 - 2017. <http://nbviewer.jupyter.org/github/jakemdraw/EducationDataNC/blob/master/2017/Models/2017ElementarySchoolSegregationTrends.ipynb>. Last retrieved 2018-09-10.
- [7] J. Drew. High school segregation trends 2014 - 2017. <http://nbviewer.jupyter.org/github/jakemdraw/EducationDataNC/blob/master/2017/Models/2017HighSchoolSegregationTrends.ipynb>. Last retrieved 2018-09-10.
- [8] J. Drew. Middle school segregation trends 2014 - 2017. <http://nbviewer.jupyter.org/github/jakemdraw/EducationDataNC/blob/master/2017/Models/2017MiddleSchoolSegregationTrends.ipynb>. Last retrieved 2018-09-10.
- [9] J. Drew. The belk endowment educational attainment data repository for north carolina public schools. <https://github.com/jakemdraw/EducationDataNC>, 2018.
- [10] B. Endowment. Belk endowment. <http://jmbendowment.org/>. Last retrieved 2018-10-01.
- [11] D. et al. Appendix - reverse engineering segregation in north carolina public schools. <https://github.com/jakemdraw/EducationDataNC/blob/master/Reports/ReverseEngineeringSegregation/APPENDIX%20-%20Reverse%20Engineering%20Segregation%20in%20North%20Carolina%20Public%20Schools.pdf>. Last retrieved 2018-10-01.
- [12] <http://scikit-learn.org>. Cross val predict. [http://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.cross\\_val\\_predict.html](http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.cross_val_predict.html). Last retrieved 2018-09-10.
- [13] <http://scikit-learn.org>. Grid search cv. [http://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html). Last retrieved 2018-09-10.
- [14] <http://scikit-learn.org>. Logistic regression. [http://scikit-learn.org/stable/modules/generated/sklearn.linear\\_model.LogisticRegression.html](http://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html). Last retrieved 2018-09-10.
- [15] <http://scikit-learn.org>. Recursive feature elimination. [http://scikit-learn.org/stable/modules/generated/sklearn.feature\\_selection.RFE.html](http://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html). Last retrieved 2018-09-10.
- [16] <http://scikit-learn.org>. Standard scaler. <http://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>. Last retrieved 2018-09-10.
- [17] R. A. Mickelson. Subverting swann: First-and second-generation segregation in the charlotte-mecklenburg schools. *American Educational Research Journal*, 38(2):215–252, 2001.
- [18] R. T. Miller. *Comparable schomparable: Evidence of inequity in the allocation of funds for teacher salary within California's public school districts*. Center for American Progress, 2010.
- [19] K. Nordstrom. Stymied by segregation: How integration can transform north carolina schools and the lives of its students. North Carolina Justice Center, 2018.
- [20] N. C. D. o. H. North Carolina State Center for Health Statistics and H. S. (2016). North carolina vital statistics, volume 1: Population, births, deaths, marriages, divorces (p. 37 / 149). [http://www.schs.state.nc.us/data/vital/volume1/2016/Vol1\\_2016\\_PRT.pdf](http://www.schs.state.nc.us/data/vital/volume1/2016/Vol1_2016_PRT.pdf). Last retrieved 2017-12-27.
- [21] U. of North Carolina at Chapel Hill. Center for Civil Rights. *The State of Exclusion: An Empirical Analysis of the Legacy of Segregated Communities in North Carolina*. UNC Center for Civil Rights, 2013.
- [22] N. C. D. of Public Instruction. Financial and business services statistical profiles. <http://www.ncpublicschools.org/fbs/resources/data/#statistical-profile>. Last retrieved 2018-08-09.
- [23] N. C. D. of Public Instruction. Nc school report cards. <http://www.ncpublicschools.org/src/researchers/>. Last retrieved 2018-08-09.
- [24] N. C. D. of Public Instruction. Statistical profile table 10-1 pupils in membership by race and sex at school level ec-pupils.csv. <http://apps.schools.nc.gov/ords/f?p=145:2:::NO:::>. Last retrieved 2017-12-18.
- [25] N. C. D. of Public Instruction. Statistical profile, table 11, a history of pupil membership by race. <http://apps.schools.nc.gov/ords/f?p=145:2:::NO:::>. Last retrieved 2017-12-20.
- [26] G. Orfield. Public school desegregation in the united states, 1968-1980. *UCLA: The Civil Rights Project / Proyecto Derechos Civiles*, pages 1–62, 1983.
- [27] G. Orfield and S. E. Eaton. *Dismantling Desegregation. The Quiet Reversal of Brown v. Board of Education*. ERIC, 1996.
- [28] R. W. Rumberger and G. J. Palardy. Does segregation still matter? the impact of student composition on academic achievement in high school. *Teachers college record*, 107(9):1999, 2005.
- [29] C. Smith. The desegregation and resegregation of charlotte's schools. The New Yorker, October 2016. <https://www.newyorker.com/news/news-desk/the-desegregation-and-resegregation-of-charlottes-schools>. Last retrieved 2018-08-18.