Testing, Testing…

Analyzing SAT Scores in NYC

By Ryan Fore

2/20/2018

Interactive versions of the visualizations contained in this report can be found at my Tableau page along with other graphs related to the project.

Executive Summary

This report examines education data provided by the city of New York regarding its public school system, with the aim of identifying patterns related to SAT scores in the city, and how scores relate with factors such as a school's racial makeup, the parents' incomes, and the perceived quality of the school. Using the relationships found, a model was built that can be used predict mean SAT scores at a school within 53 points on average (the SAT is out of 2400 points).

Out of the dozens of variables available, the single greatest predictor of SAT scores was the percentage of students at a school who receive a free or reduced price lunch (FRL). Since FRLs are only offered to students from low-income family, this shows that the parents' economic status plays a powerful effect in determining a student's academic success. It is also shown that schools in Brooklyn and the Bronx are overwhelmingly populated by black and Hispanic students, respectively, and that these schools that are mostly black and Hispanic do markedly worse than schools that have higher populations of whites and Asians.

One of the datasets used in the analysis contains the results of a survey administered to students, teachers, and parents, asking how they would rate their school in terms of safety, academic expectations, communication, and engagement. Surprisingly, safety scores are more highly correlated with SAT scores than any other category, including academics. Perhaps unsurprisingly, for every category, the ratings given by students were the most correlated with SAT scores, while parents' ratings were the lowest, often being negligible.

Possibly the most interesting finding of the study is that the SAT might be unfairly biased against students who are non-native English speakers. Schools with higher percentages of ELL students generally do much worse on the SAT, even on the Math section of the test. Since takers of the SAT are normally upperclassmen, who have likely gained a moderately high level of English fluency since they started high school, the strength of the correlation is somewhat surprising. Given the limited detail of the underlying data, it is impossible to definitively conclude whether the SAT is truly biased against non-native speakers or not, but the results of this report likely warrant a separate investigation focused solely on this issue.

1. Introduction

The city of New York posts numerous datasets regarding education in the city. In this report, we'll look specifically at SAT scores, and how different factors such as race, perceived safety, and other standardized test scores can be used to predict average SAT scores at a school.

If you are unfamiliar with the SAT, it is a standardized test offered to high school students who want to go to college. During the years the data covers, it comprised of 3 sections (Critical Reading, Math, Writing), each worth 800 points for a maximum possible score of 2400. Taking the SAT is optional, however many colleges require it as part of their admissions process.

Amongst other things, the analysis shows that:

- The percent of students receiving free or reduced price lunches is the single strongest predictor of SAT scores out of the data available
- The SAT might be unfairly hurting non-native English speakers' chances of college admissions
- What students think about a school is more highly correlated with a school's success than what parents think

2. The Data

All of the following datasets are available at https://data.cityofnewyork.us/

2012 SAT results- Contains information regarding how many students at each school took the SAT, and what their average score was, broken by section

DOE High School Directory 2013-2014- A directory with information such as extracurriculars, foreign languages offered, and information regarding the admissions process (if there is one)

2010-2011 Class Size - School-level detail – Information regarding class sizes at each school, broken down by program type and core subjects for each school (e.g. one row for general education math classes, another for gen ed English classes, etc.)

2010 AP (College Board) School Level Results- Data about how many students in each school took AP tests, how many exams were taken, and how many were passed. AP tests are exams that can be taken at the end of a school year (normally after taking the corresponding class) in order to obtain college credit for the subject.

2005-2010 Graduation Outcomes - School Level – Contains statistics such as graduation rate, dropout rate, and how many students obtained a "Regents" diploma (a specific distinction based on standardized tests that are offered to students in the state of NY). Data is broken down by cohort (the students' entry-year) for each school.

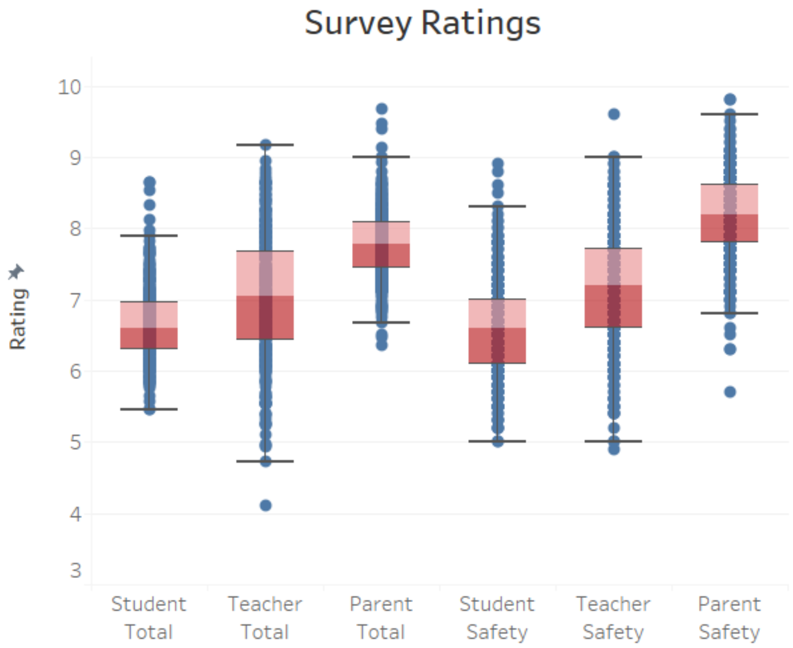
2006 - 2012 School Demographics and Accountability Snapshot- Offers information regarding the race and sex demographics, and how many students are in each grade. Information is broken down by school year for each school, and the dataset contains information for all NYC schools, not just high schools.

2010 - 2011 NYC School Survey- The results of a survey offered to students, teachers, and parents that asked how they would rate their school in 4 different categories: Safety, Academic Expectations, Engagement, and Communication. For example, in the dataset saf_p_11 refers to the score parents gave regarding safety, whereas eng_tot_11 refers to the average of the engagement score given to a school by parents, teachers and students. In this report, these fields will generally be referred to by names like Parent Safety and Engagement Total.

---

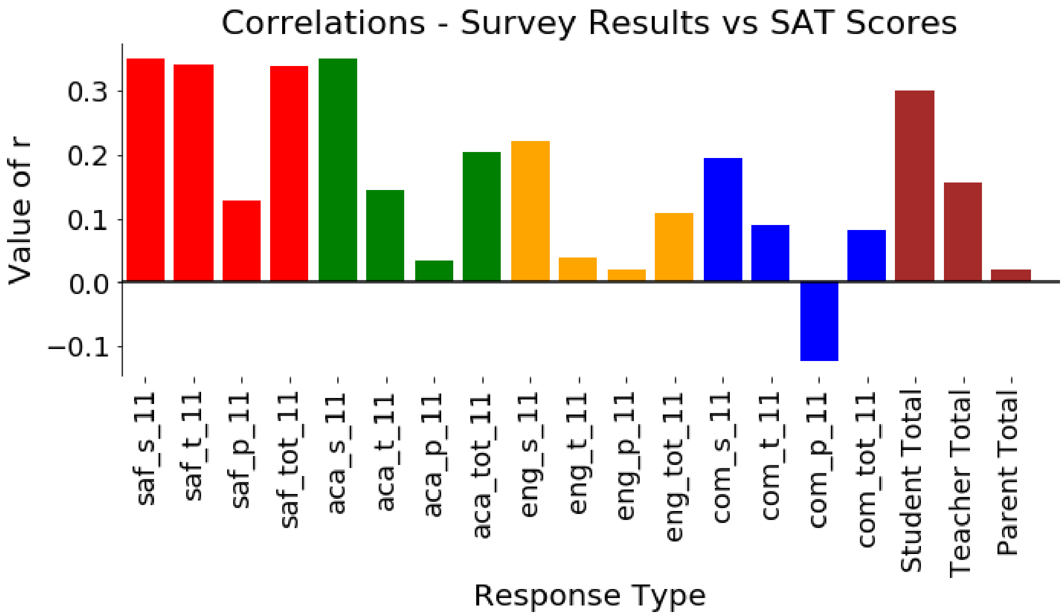
3. Survey Says…

As noted above, the survey asks parents, teachers and students what they think about the school in terms of 4 different aspects. The question is, what are the differences in how the respondent groups view the schools, and which, if any, response types are correlated with SAT scores. To aid in this analysis, a total score was calculated for each respondent/school combination, for example the average score students gave for safety, academics, engagement, and communication is represented as "Student Total". In the graph below, the three distributions on the left show the total scores for each group, and the three on the right show the safety scores. The red boxes indicate scores that fall in the 25th to 50th percentile, and the pink boxes indicate scores in the 50th to 75th percentile. Any score that falls outside of the horizontal black lines is a statistical outlier.



If we look first at the total scores, we see that parents by far give the highest scores, while students give the lowest scores, a trend that is repeated in the safety scores, and is true for any of the other categories as well. Parents and students clearly don't agree on how their schools perform on these measures, but whose responses are more relevant?

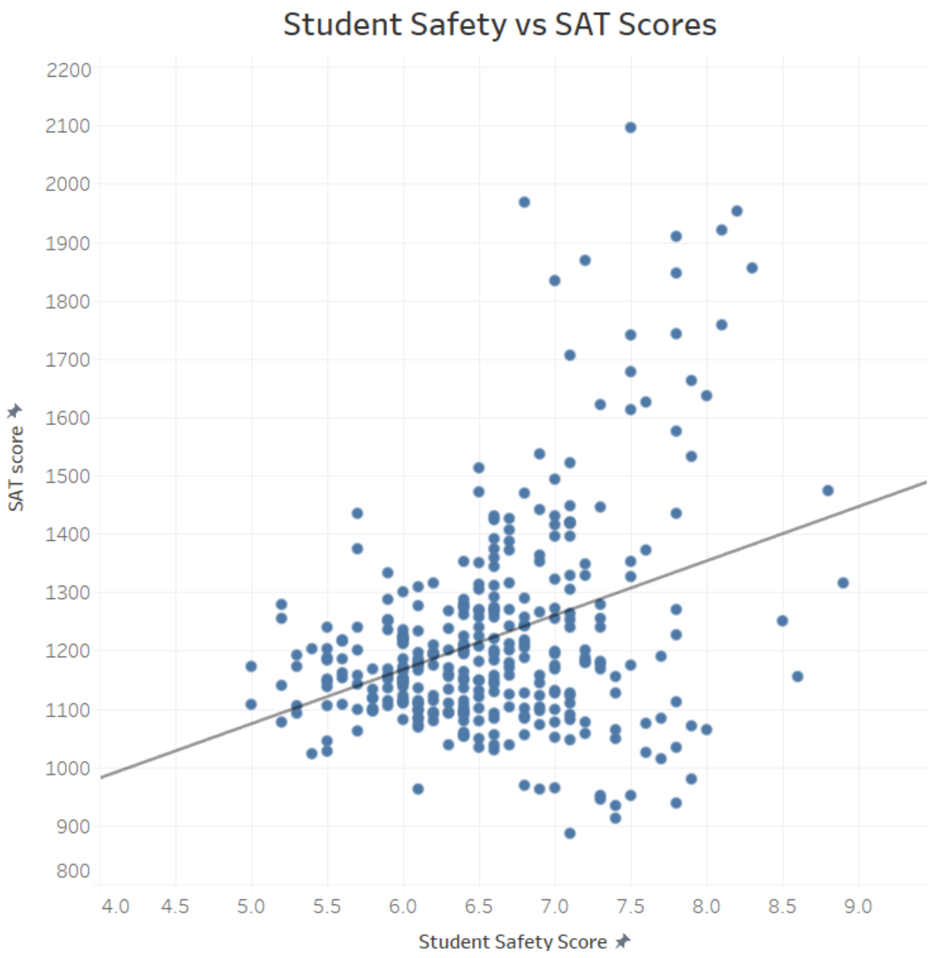
To see how these responses are related with SAT scores, let's look at the correlations between the different response types and SAT scores. In the graph below, each color corresponds to a different category, and within each color group the responses are in the order of student, teacher, parent, then total. The height of each bar is the r value for the response, which is an indicator of correlation strength.



Looking at the r values for each response type, we see that:

- Amongst the 4 categories, safety has the highest correlation, then academic expectations, followed by engagement score and then finally communication
- For each category, the student responses have the highest correlation, followed by the teachers, and then the parents.

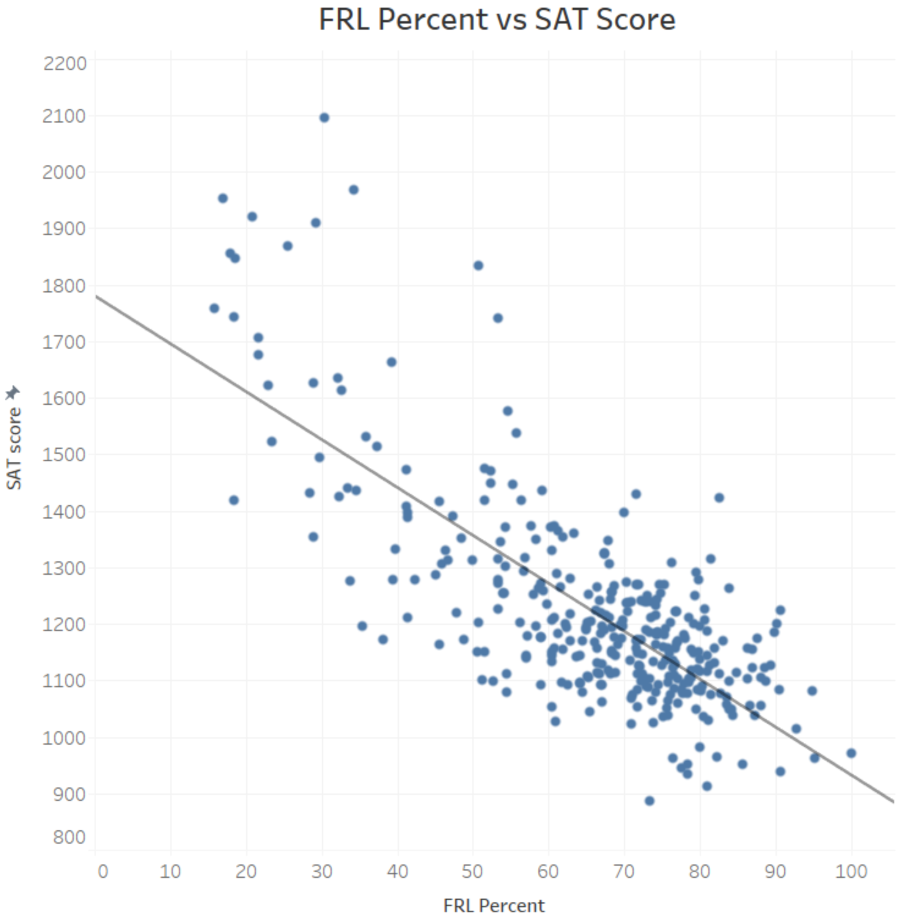
These correlation values however should be taken with a grain of salt. If for example we look the following scatter plot of the student safety score, we'll see a large part of this correlation is due to a cluster of schools with a high safety score and very high SAT scores.



Without the cluster at the top right, the correlation would be a lot less apparent, however this is not to say there's no relation at all. For example, only one school that received a safety grade lower than 6.5 has an SAT score higher than 1400, while there are dozens of schools who have safety scores higher than 6.5 with SAT scores over 1400. Clearly if a student wants to excel academically, it helps to be in an environment they feel safe in.

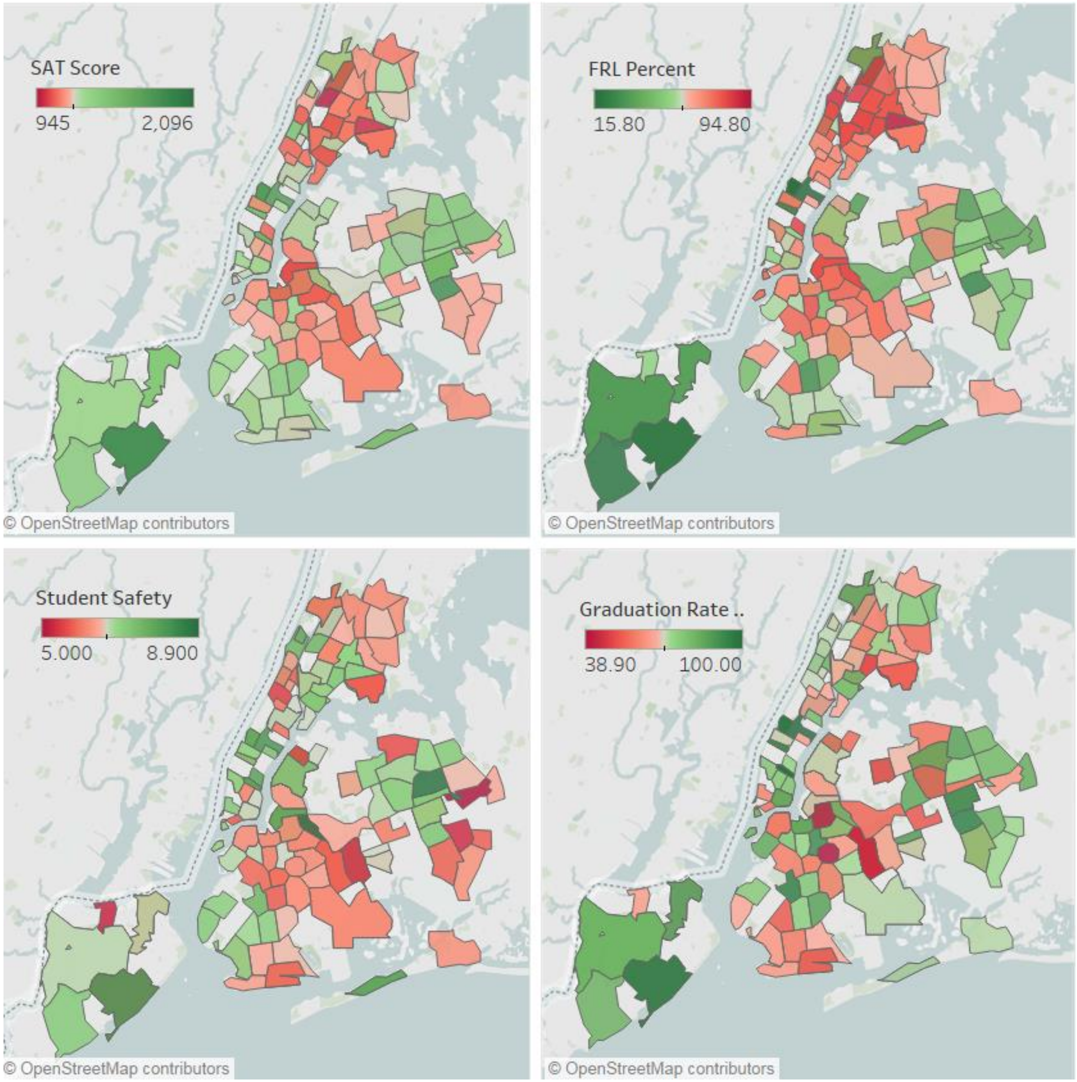
4. The Price of a Free Lunch

Out of all the factors found in the data, the one that is most correlated with SAT scores is the percent of students who receive a free or reduced price lunch (FRL), with an r value of -0.75. The graph below clearly shows that as FRL percent goes up, SAT scores go down.



FRL students come from families with low incomes, so in a sense FRL percent serves as a proxy for the average household income in a school. In other words, out of all the variables we have access to, the single most powerful predictor for academic success is how much money the parents have. The playing field is clearly not at all level for students of different economic backgrounds.

To further illustrate the effects student safety and FRL percent have on academic success, it's important to look at how these factors vary over different areas of the city. The dashboard below contains four maps: The top left map shows the average SAT scores in zip codes around the city, the top right shows FRL by zip code, the bottom left the safety scores given by students, and the bottom right the graduation rates, another indicator of the academic success of a school.

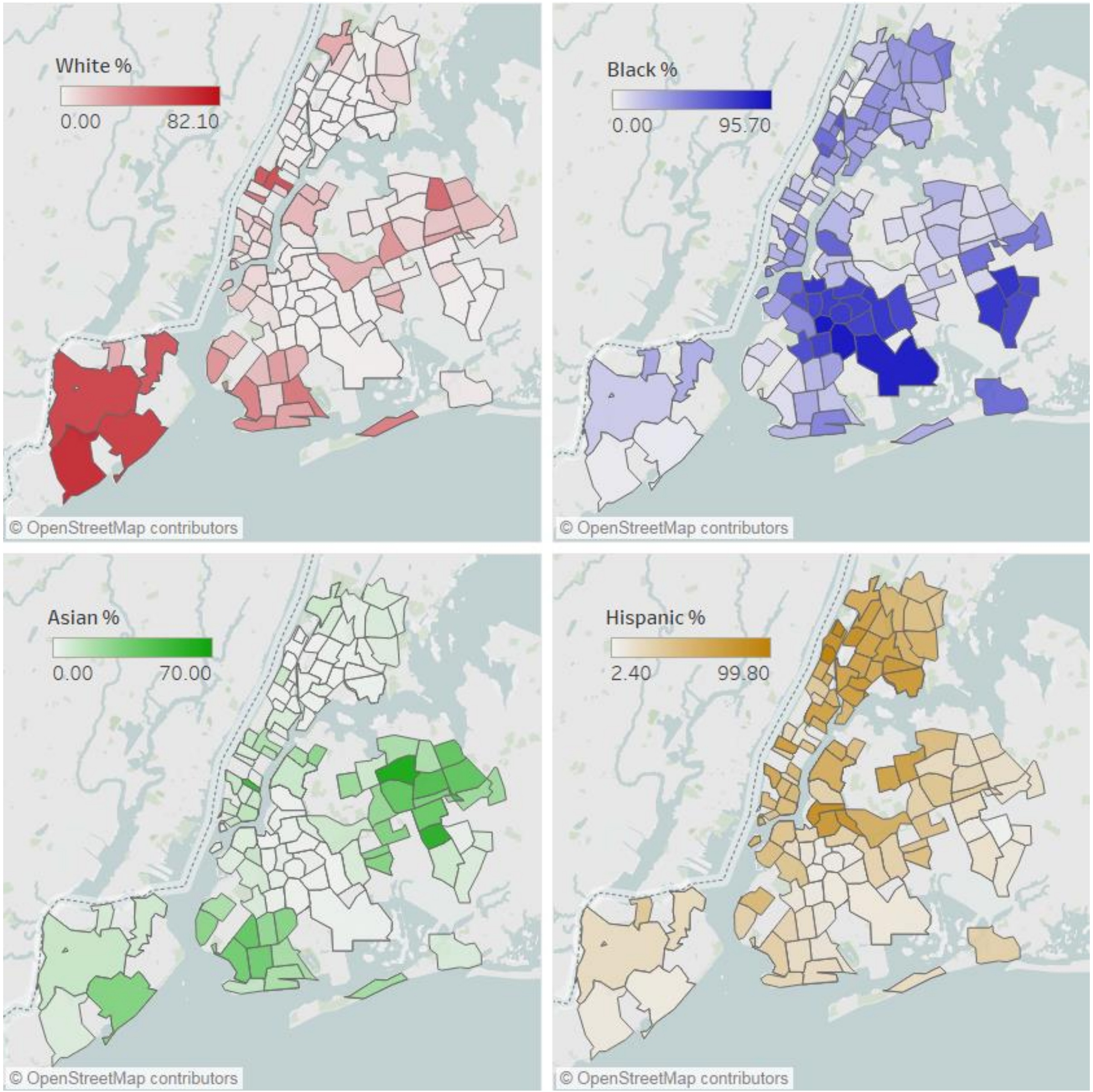

SAT Score
945
2,096
© OpenStreetMap contributors
FRL Percent
15.80
94.80
© OpenStreetMap contributors
Student Safety
5.000
8.900
© OpenStreetMap contributors
Graduation Rate ..
38.90
100.00
© OpenStreetMap contributors

Note: The color schemes are centered at or near the median for clarity purposes

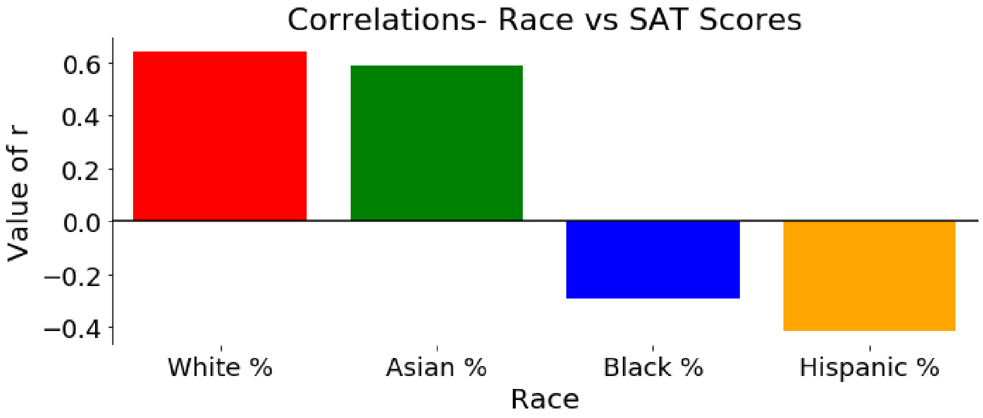
Notice how all of the maps look fairly similar? This shows that these 4 dimensions generally move together and are all fairly correlated. For example, if we look at the top or middle sections of the maps (The Bronx and Brooklyn), we see that SAT scores, graduation rates and student safety scores are all relatively low, while FRL percent is high. Conversely, in the western and northeastern portions of the maps (Staten Island and Queens), FRL percents are low, while the other dimensions are high. There are clearly two very distinct subsets of learning environments to be found in the city. This leads one to ask, what more do we know about these different zones, namely, who lives in them?

5. Diversity, or Rather the Lack Thereof

To see how different segments of the city's population fit in to the model we've built so far, let's look at the population density of the four races with the largest populations in the city.

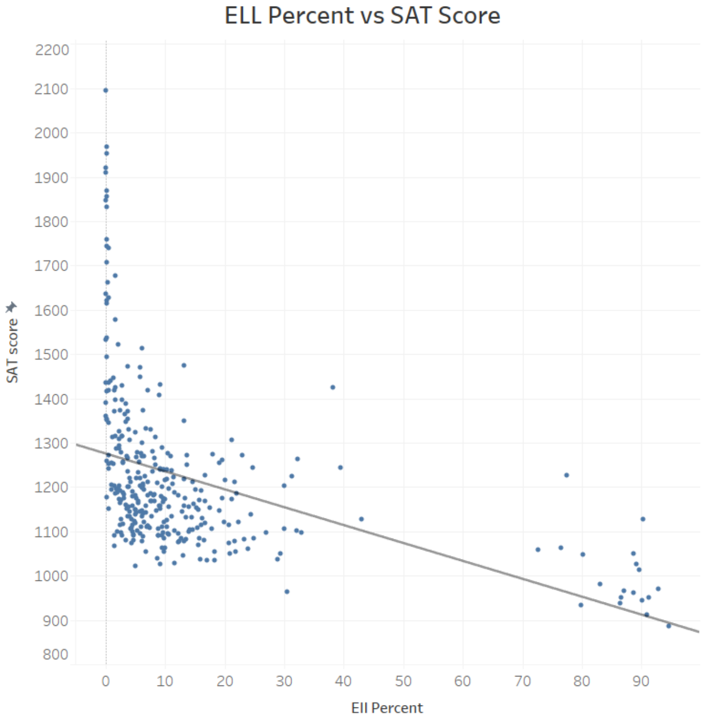

White %
0.00
82.10
© OpenStreetMap contributors
Black %
0.00
95.70
© OpenStreetMap contributors
Asian %
0.00
70.00
© OpenStreetMap contributors
Hispanic %
2.40
99.80
© OpenStreetMap contributors

These maps show that NYC public schools are still very segregated, with most schools either being composed of either a mix of whites and Asians, or a mix of blacks and Hispanics. If we compare this dashboard with the dashboard from earlier, a clear pattern emerges. The high achieving zones are the areas that are mostly populated by whites and Asians, while the low achieving areas are mostly populated by blacks and Hispanics. The r values between the 4 different races and SAT scores backs this up, showing positive correlations for white and Asian percentages, and negative for black and Hispanic ones:



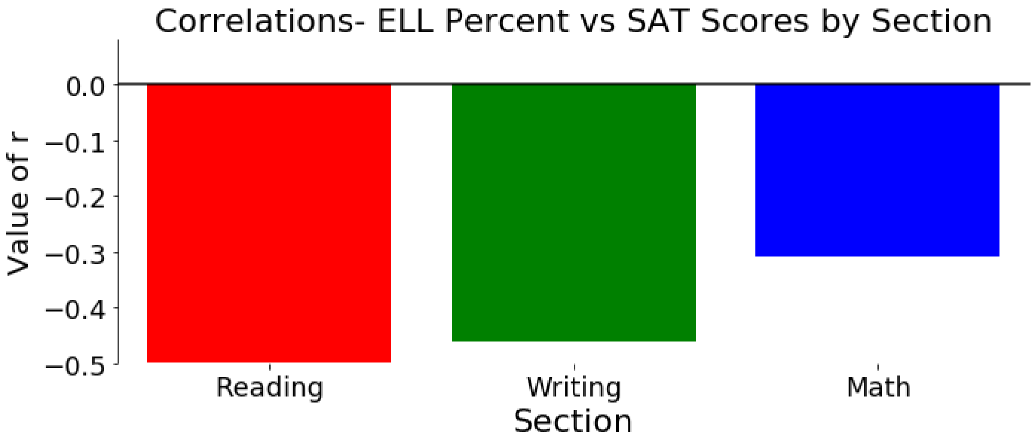
6. English Language Learners

The percent of students at a school who are English Language Learners (students who still aren't fluent in English) is also a good predictor of SAT scores. Testing to determine ELL status is administered every Spring, and in general ELL enrollment decreases as student age increases. For example, there are over twice as many ELL students in 9th grade than 12th grade, showing that as they progress through the system, ELL students can quickly gain English proficiency. Since most students take the SAT their Senior year or at the end of their Junior year, this means that as long as they started high school in America, they likely have a fairly high English proficiency level when they take the test. Despite this though, a highly negative correlation was found between the ELL percent and SAT scores, as shown the graph below:



The most obvious part of the graph is the fact that there's a group of schools that have an ELL percent higher than 70%, all of whom do relatively poorly on the SAT. In fact, out of the 12 schools in NYC that have average SAT scores lower than 1000, 11 of them belong to this high ELL category. Even without this cluster though, there is still a fairly strong correlation between the measures. As proof of this, it can be seen that out of the dozens of schools that have SAT scores higher than 1400, only two have ELL percents higher than 10%, and none of the schools with scores over 1600 have ELL percents higher than 2%.

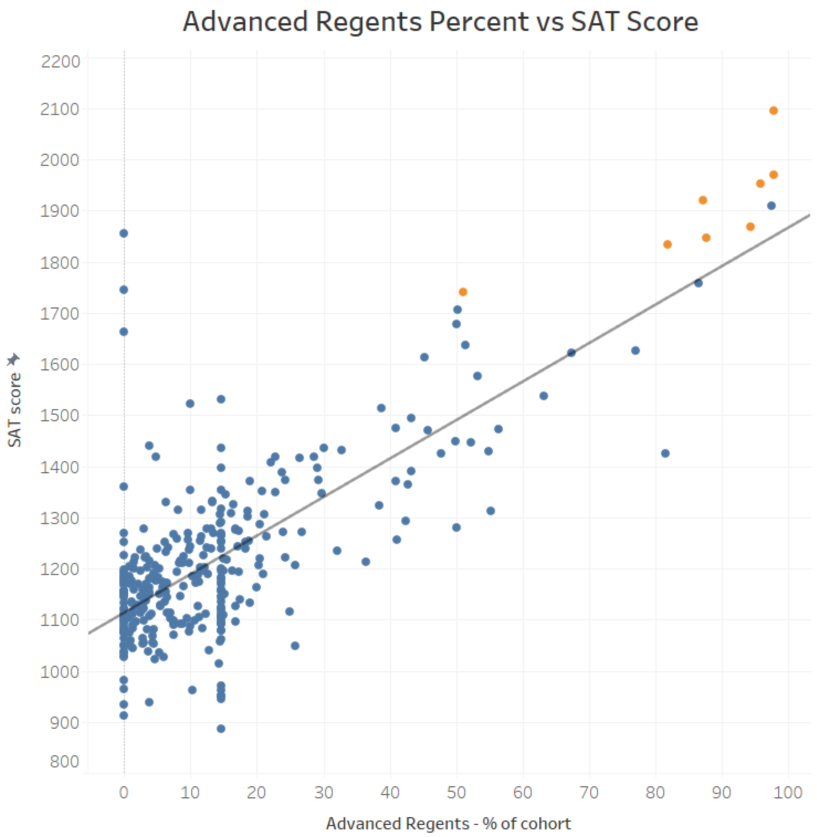
Predictably, the percent of ELL students has a higher effect on Critical Reading and Writing scores than on Math scores. The correlation value for the Math section however is still sizable, which is likely due in part to the fact that many of the problems in the section are word problems, thus testing the student's language skills instead of just qualitative skills.



The SAT is supposed to gauge a student's college readiness, but while the SAT is administered under strictly timed conditions, causing even native speakers to often feel rushed, most college coursework is under no such time constraint. It's conceivable then there are non-native students who have a high enough level of English for it to not negatively affect their college coursework, but who still do poorly on the SAT, as they likely feel the time rush even more than native students. It could be argued then that the SAT is unfairly biased against ELL students, hurting their chances to succeed as (presumably) immigrants in America. As mentioned before, there are a lot of confounding variables regarding this effect, and a lack of detail in the data used in this analysis means that a definitive conclusion is out of the scope of this report, however it is definitely a matter worth further investigation

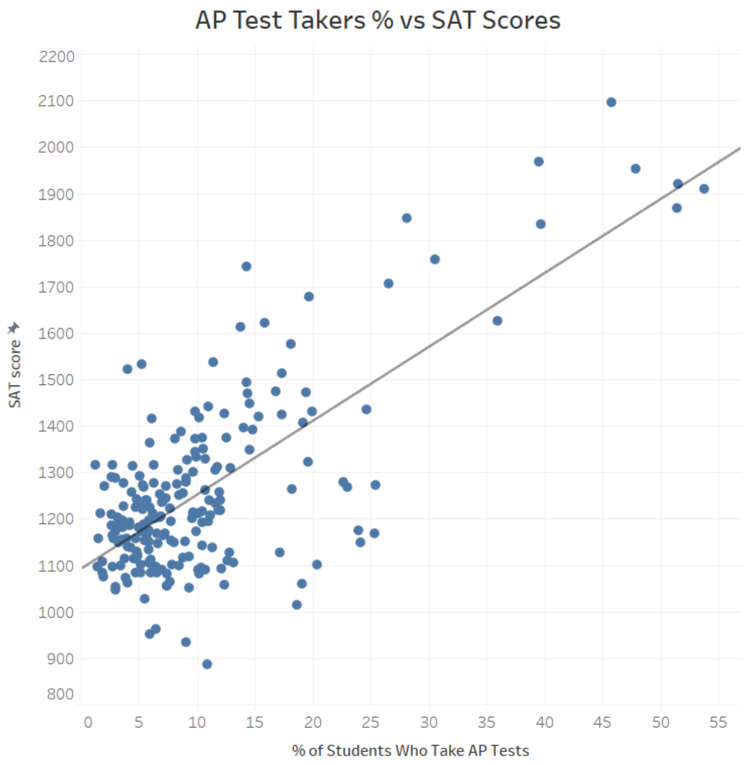
7. SAT vs Other Test Scores

The SAT is only one of many standardized tests NYC students might take growing up. Another example is the SHSAT, a sort of mock-SAT offered to 8th and 9th graders that is used to determine entrance to 8 of the most prestigious schools in the city. There's also the Regents Examinations, a group of 10 exams in a variety of fields such as History, Math, Science, and Foreign Languages, offered to all students in the state of New York. If a student passes 5 exams, they earn a Regents distinction on their diploma, and if they pass 9 of them, including a foreign language exam, they earn an Advanced Regents distinction. Below is a chart showing the percent of students in a school who earn Advanced Regents distinctions versus average SAT scores, with the schools that use SHSAT scores for admissions highlighted in orange.



It turns out, next to the FRL percent, the Advanced Regents percent has the highest correlation with SAT scores out of all of the fields in the dataset. And notice how we finally have an explanation for why some schools are so incredibly high-achieving: all 8 SHSAT schools are in the top 12 highest scoring schools on the SAT. This isn't too surprising though, as it makes sense that schools that accept students based on mock-SAT scores would then have high SAT scores.

As mentioned before, AP tests are another type of standardized exam that high school students can take. These tests are normally only taken after taking the corresponding course in school, and since they offer the chance for college credit if they're passed, it's usually only the higher achieving students who take the classes/exams. As such, we'd expect that as the percent of students who take AP tests goes up, SAT scores would also go up, and we find that this is indeed the case:

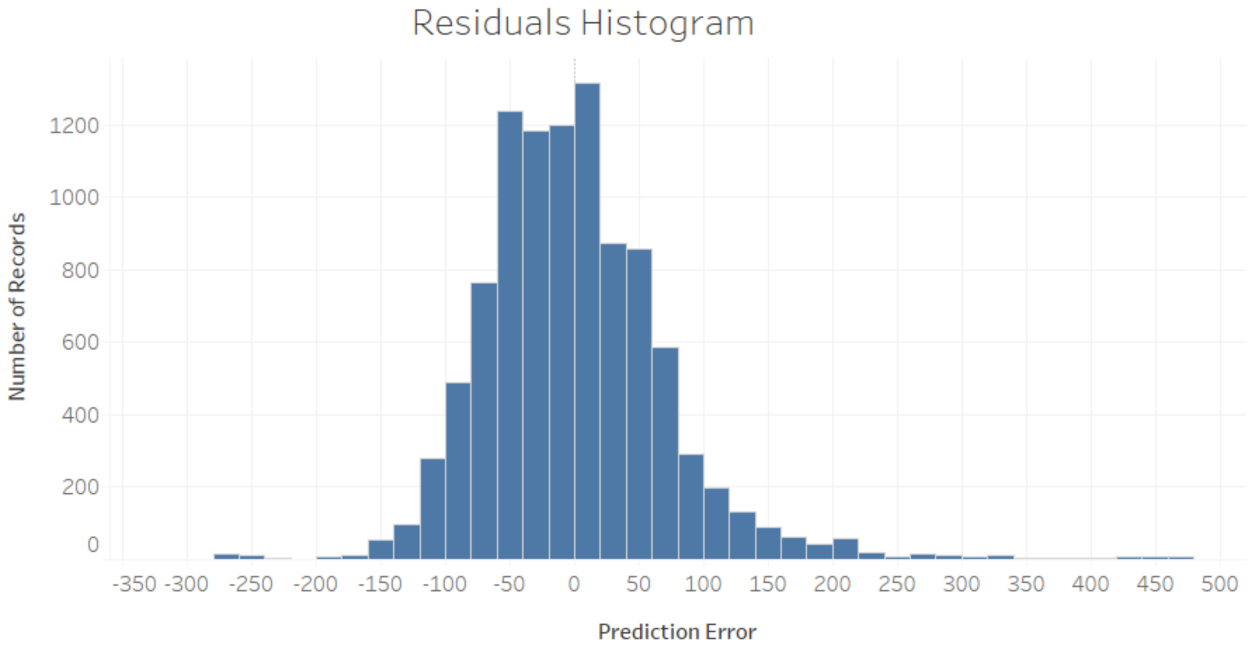


8. Predictive Modelling

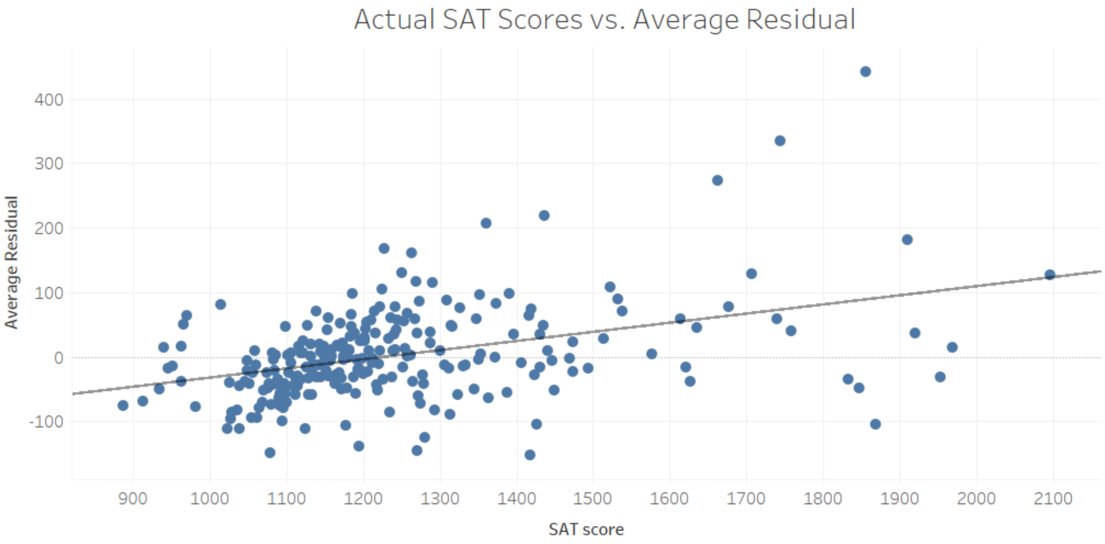
To create a predictive model for SAT scores, the cleaned dataset was randomly broken into two smaller sets: a learning set, comprised of 70% of the cleaned set, and a test set comprised of the remaining 30%, used to test the accuracy of the model. The following fields were used as predictors in a multiple linear regression model:

- 'frl_percent'
- 'asian_per'
- 'black_per'
- 'hispanic_per'
- 'ell_percent'
- 'Advanced Regents - % of cohort'
- 'SHSAT School' (1 if the school is one of the 8 that uses the SHSAT for admissions, else 0)
- 'High Ell Per' (1 if the school has an ELL percent higher than 70, else 0)

The process of assigning sets, fitting a model, and calculating errors was repeated 100 times. Over the course of 100 trials, the model's predictions were on average 53.3 points off of the actual values in the test sets, while the median size of the error was only 43.8. Given that the range of scores in the underlying dataset is 1209 points, this means the model was able to reliably predict the SAT scores with an average error of only 4.4% relative to the range.



Looking at the histogram of the residuals above, we see they follow a fairly normal distribution, which is a good sign. What is concerning though is how long the tails are, specifically towards the positive end. If we look below at the relationship between SAT Scores and residuals, we see that there is a slight upward trend:



A large reason for this trend though is due to only a few points that have very large residuals. For example, the data point in the top right that has an average residual of over 400 correspond to Bard High School Early College, a school where students move at a highly accelerated pace in order to fulfill the requirements for a high school diploma and a 2-year college Associate's Degree, all in 4 years. The students are clearly highly gifted, but the school doesn't use the SHSAT for admissions, nor are the students allowed to pursue Advanced Regents distinctions, which is why the model is not able to accurately predict its SAT score. The other data points with overly large residuals are likely due to similar reasons, but without individually going through each of them and finding a way to group them, it would be hard to improve the model's ability to predict their SAT scores.

9. Closing Thoughts and Areas for Further Analysis

In this report, it has been shown the income levels and the racial make-up of a school generally go hand in hand, and are both strong indicators of a school's SAT scores. NYC schools are still fairly segregated, with the schools that are mostly black and Hispanic generally being poorer and less successful academically. One potential avenue for continuing this analysis would be to add a data related to housing costs throughout the city. This could be used to find pockets of the city that give the most educational value for what is paid in rent, potentially allowing some of these lower income minorities a better chance to succeed academically.

Another possible area for further analysis is looking at the effects being a non-native English speaker has on SAT scores and thus college admissions. It is very possible that a non-native student who can speak English fluently in their day-to-day life does disproportionately worse on the SAT due to time-constraints and the pressure of the test. Since a lot of work in college is done under no such time constraint, the SAT could be a poor reflection of a non-native speaker's college readiness, hurting their chances of getting into good schools. Further analysis could include how SAT scores compare to scores on foreign language tests meant to judge someone's English skills such as the TOEFL, and how these foreign language scores correlate to success in high school and college.

The accuracy of the analyses in this report could be improved upon by going through the data at a more granular level, looking for information that could separate certain kinds of schools into different subsets, and researching individual schools if necessary, like what was done with the SHSAT admissions based schools. Similarly, using more advanced predictive modeling techniques such as a Lasso regression could yield better results than a multiple linear regression model.