Effect of Common Core on NYC Middle School Math Examination from 2013-2017

Thomas Kwok
M.A. Hunter College
Statistics Final Project

The implementation of Common Core to New York City Middle School math students from 2013-2017 and how effective has the shift been.

Section 1. Introduction

In 2013, New York State implemented a change to the New York State Math examination, aligning the test to the Common Core standards. The purpose of Common Core was to make a more rigorous exam that would prepare students for college and create higher thinking and problem solving. Looking at the National Assessment of Educational Progress (NAEP), the switch to Common Core made a lot of sense because the whole United States (not just New York) was performing in the middle of the pack and not excelling from the old assessments and benchmarks. According to the study by NAEP, in 2011, the last report card before the switch to Common Core, 8^{th} Graders in the United States received a score of 284 out of 400, whereas the NAEP criteria for Proficiency was 299 and Basic was 262, so the nation as a whole was performing below proficient but above Basic.

Common Core provided a lot of benefits; first it was internationally accepted with standards that are adopted from other countries, which allowed students to be better accessed in NAEP when compared to foreign countries. The standards also built off one another, so it was easier to progress-monitor student performance as they ascended from elementary school to middle school, and it culminated with more preparedness for high school and college. The idea behind implementing Common Core was smart, but how effective has it really been? This paper will look into the effects of New York State math exam scores from 2013-2017 for middle school students (Grade 6 to 8) and measure how groups of student performed over time with more exposure to the structure of Common Core.

Before we begin, let's first discuss what is Common Core and how to read the exam data. Common Core provides a set of standards that students are supposed to master by the time they finish high school, with the goal of being ready for college and the real world. These standards are broken up from pre-kindergarten to High School, with a progression schedule that would prepare the students to fully master the contents in college Math. In every grade, the students are supposed to take the foundational skills they learned from their prior grades and use it to solve more rigorous problem. One example is how in sixth grade, students take their multiplication and division skills from elementary school and utilize it to solve ratios and proportion questions. These same foundational skills are also used to learn expressions and equations, understanding the way to write equations and inequalities through the arithmetic. In seventh grade, they further their understanding of ratios and proportions and start to look at different units, for example turning inches to feet. They also take the equations skills they learned and start to solve one variable problems. This culminates in eighth grade, when they begin to utilize functions and linear equations, solving unknowns and building the skills necessary for High School algebra.

The state exam that they are tested annually contains multiple choice and short answer problems. Students are given mostly word problems and must use their comprehension skills to understand the question asked and come up with a formula to solve the problem. The skill learned here helps prepare them for college and the real world, where they are writing algorithms and taking a problem of interest and creating a method to solve it. The exam then contains a raw score and its scaled equivalence which is put into levels. Students are ranked on a scale of 1-4: one meaning the student is well below expected grade level, two meaning they are partially proficient, three meaning they are fully proficient for the grade, and four meaning they are well above expectation. Every student who takes the exam is given a rank of 1-4 and our data tracks the mean raw score and the percentage of students who received each rank.

Most research prior to this study only looks at the percentage of students who received a rank of 3 or 4, showing proficiency for the grade and utilizing that to measure the performance of students. This project will look at the raw score instead of the scaled score to look at performance, as the likelihood is that not all students will achieve proficiency by the time they graduate, but we should still be able to track their growth or lack thereof.

Section 2. Materials

2.1 Educational Data

The data used in this study comes from the New York State of Education (NYSED) and contains Borough and City data from 2013-2019 for Grades 3-8. The main focus of this research is Grades 6-8, to see how Middle School students in New York performed in the state exam. The time frame studied is 2013-2017 because in 2018, the examination changed as the test went from being three days to two days, thus 2018-2019 data can not be compared to 2013-2017 data since the exams are different.

The dataset contains approximately 400,000 students, aggregated into Boroughs and the entire New York City. The data is also split into Ethnicity, Gender, Economic Status, English as a Second Language (ELL), and students with disabilities (SWD) for both the Borough level and entire New York City. Our main focus will be on each Borough, as we measure how students performed longitudinally with these characteristic factors in mind.

2.2 Quality of Data

In this report the experimental unit we looked at are each group of students defined by grade, year, borough, and one of the five categories (ethnicity, gender, econ, ELL, SWD). For example, one group would consist of 6th Grade students from 2013, Asian, and living in the Bronx and we look at their longitudinal measurements of Mean Score at their 6th grade, 7th grade, and 8th grade. This group would be different than another that is 6th grade students from 2013, Hispanic, and living in the Bronx. Our goal is to have a longitudinal measurements of the same group of subjects to see if there are trends of growth. The majority of subjects studied will have three years of observations, but are also 8th Grade students from 2013 for each Borough and 6th Grade students from 2017 for each Borough that only have one year of observation. In addition there will be 2016 6th Graders and 2013 7th Graders for each Borough that will only have two years of observation. In total there are 34 groups measured in this research set.

2.3 Data Set

For each of the five data sets, there is a total of 17 variables, as shown below. For all of the data sets, 16 of the 17 variables are the same, with the only difference being the category studied which are ethnicity, gender, economic status, English Language Learner (ELL), and Students with Disabilities (SWD). Listed below is a table of the variables shown in the ethnicity data set. There will also be a slight change in the SWD data set as the classification of disability started being recorded in 2014 so the data will be from 2014-2017 instead of 2013-2017.

Table 2.3: Variables of data set.

Variables	Definition
Borough	Factor with 5 levels: Bronx, Brooklyn, Manhattan, Queens, Staten Island
Grade	Factor with 3 levels: 6^{th} Grade, 7^{th} Grade, 8^{th} Grade
Year	Factor with 5 levels: 2013, 2014, 2015, 2016, 2017
Ethnicity	Factor with 4 levels: Asian, Black, Hispanic, White
Number.Tested	Integer of Number of Students Tested
Mean.Scale.Score	Integer of Mean Score
X..Level.1	Number of Students with rank 1 (Well Below Grade Level)
X..Level.1.1	Percentage of students with rank 1
X..Level.2	Number of Students with rank 2 (Partially Proficient)
X..Level.2.1	Percentage of students with rank 2
X..Level.3	Number of students with rank 3 (Proficient)
X..Level.3.1	Percentage of students with rank 3
X..Level.4	Number of students with rank 4 (Advanced)
X..Level.4.1	Percentage of students with rank 4
X..Level.3.4	Number of students with rank 3 or 4 (Proficient and above)
X..Level.3.4.1	Percentage of students with rank 3 or 4
Group defined by Borough, Race, Year, and Grade:	Factor of our Longitudinal Analysis: Example Bronx, 6^{th} Grade, 2013, Asian would be one Group

Section 3: Preliminary graphical exploration

Exploratory analysis of longitudinal data seeks to discover patterns of systematic variation across groups of patients, as well as aspects of random variation that distinguish individual patients. In order to achieve this, we first created a few lattice plots that measure the student performance to see if there are any trends that can be found.

3.1 Measure of Proficiency (Rank 3 or 4)

The first exploration involved looking at the trends for proficiency during the period of 2013-2017. In doing so, we created a lattice plot that tracked the number of students from each Borough that achieved a rank of proficient in their state exams for grades 6-8. All of the Boroughs showed a similar trend, where 6th Grade students showed improvement in their 7th Grade exam score, but had a big decline by the time they took the exam in 8th Grade.

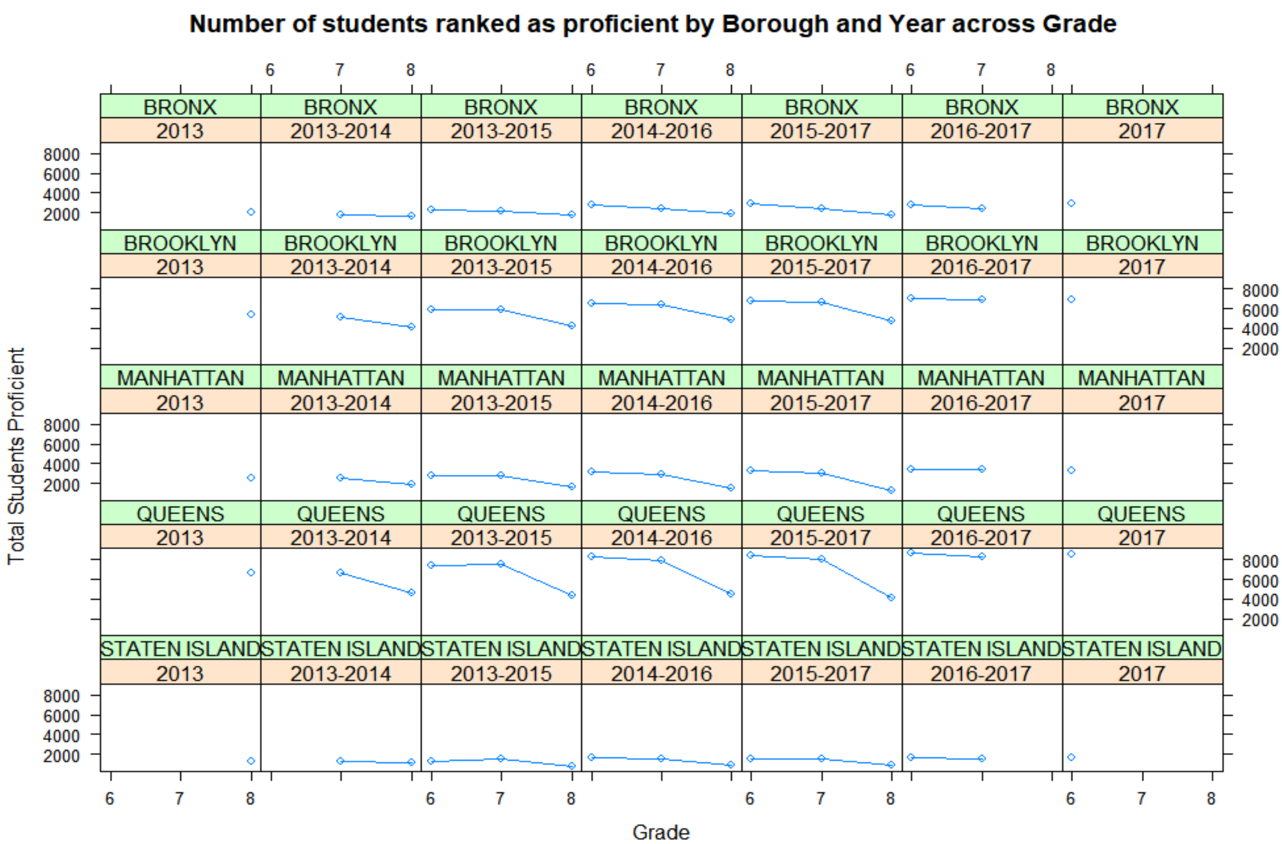

Figure 3.1.1: Number of students who achieved proficiency in the state exam from 2013-2017

The one issue with measuring the number of students who achieved proficiency in their state exam is that each Borough has a different population and thus it would be difficult to compare the Boroughs based off of total number of students alone.

The next exploration involved the percentage of students from each Borough who achieved proficiency in the state exam from 2013-2017. From this data set we saw that students from the Bronx scored well less than the other Boroughs, but also that no Borough had more than 50% of their students scoring proficient in the exams.

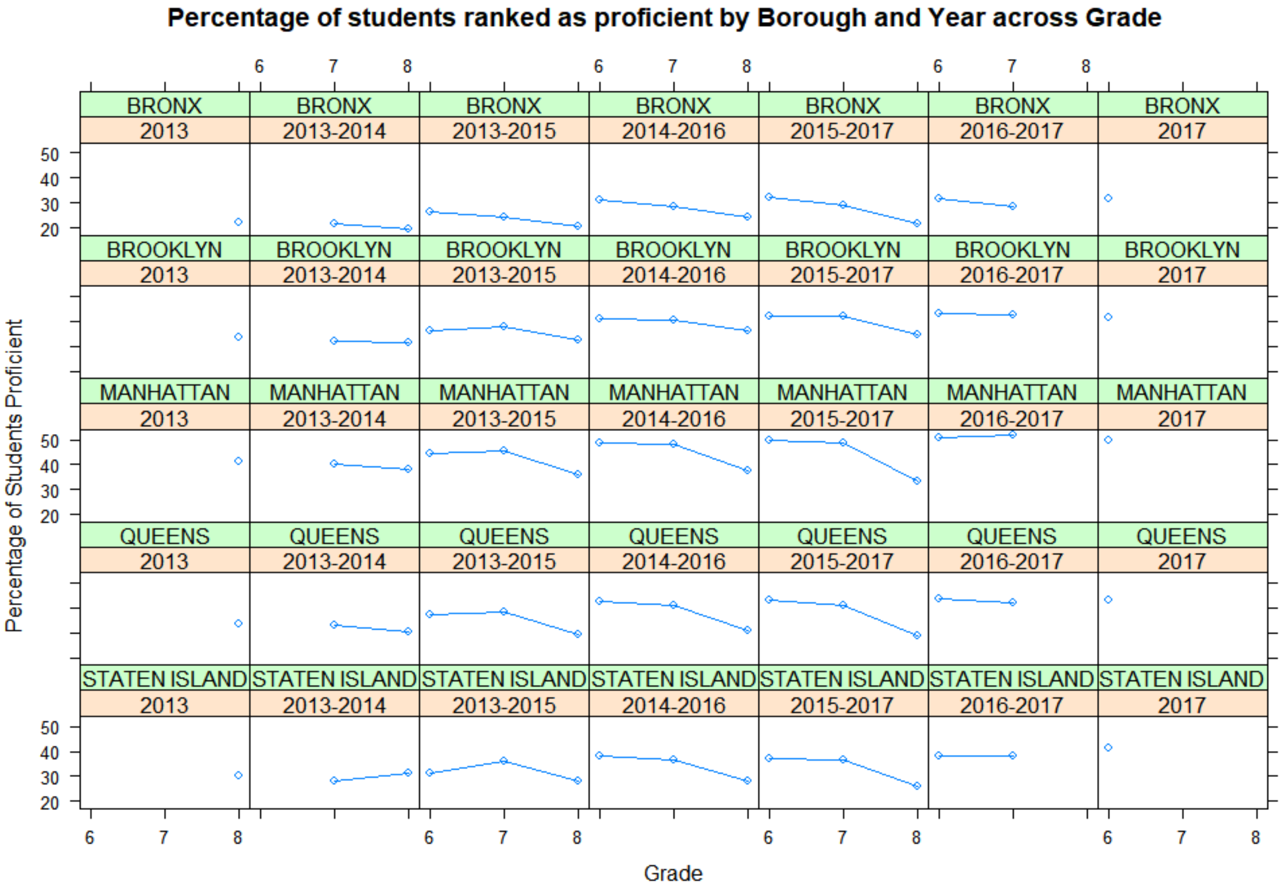

Figure 3.1.2: Percent of students who achieved proficiency in the state exam from 2013-2017

3.2 Number of students tested

The next exploration involved looking at the number of students tested for each Borough during 2013-2017 to see if there are any trends seen. The data shows that the number of 8th Graders who take the exam is much lower than their 7th Grade counter part. This will be revisited in the Discussion section, as there may be reasons for this.

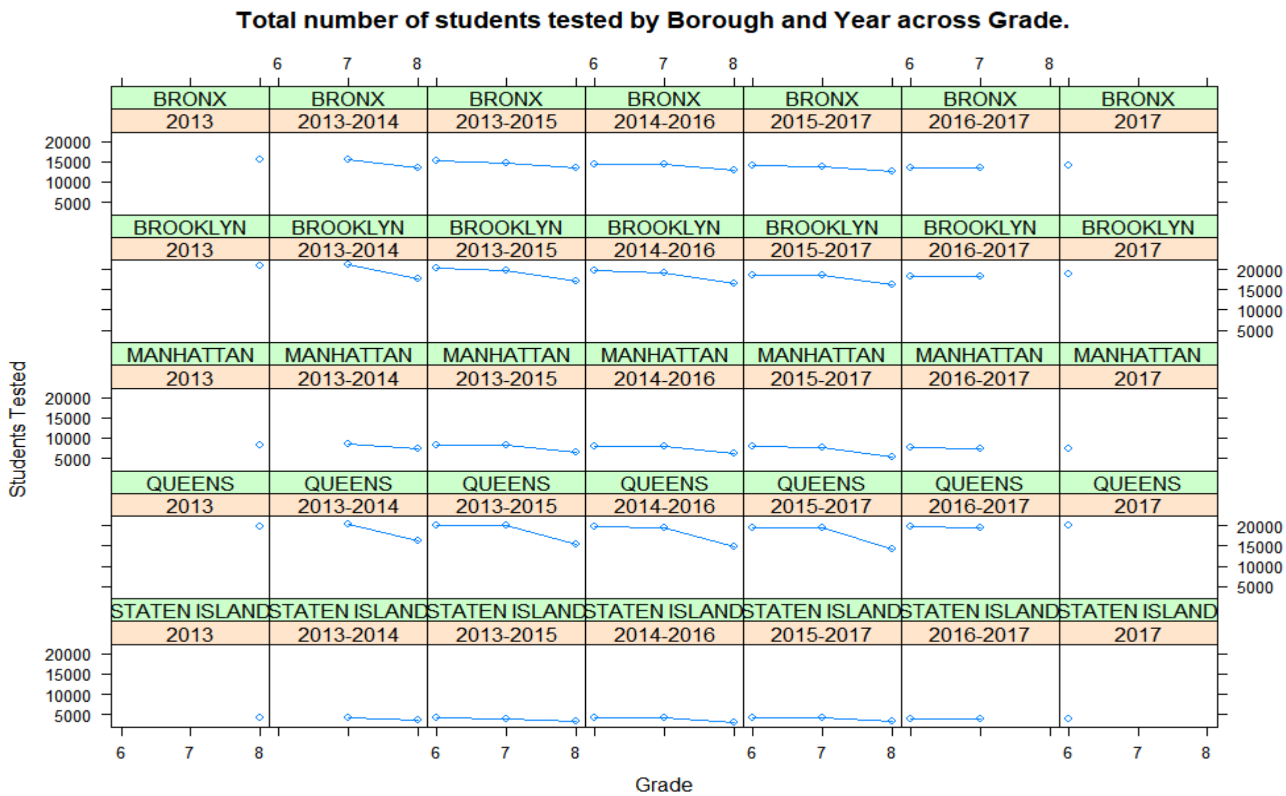

Figure 3.2.1: Number of Students Tested for the State Exam from 2013-2017

Below is a table that shows the total number of students in all five boroughs who took the exam in 7th Grade compared to 8th Grade. The decline is startling as there are approximately 10,000 students who take the 7th Grade Math state exam but do not take the 8th Grade state exam the following year.

Table 3.2.2: Number of students in all NYC who take the exams in 7th Grade and 8th Grade

	Grade 7	Grade 8
2013-2014	70305	59875
2014-2015	68091	56999
2015-2016	66259	54762
2016-2017	65371	53102

3.3 Raw Score Data across Grades and Boroughs

For the previous three examples, we looked at students based off the scaled ranking rubric that New York State created. Now, we will examine the raw scores that students receive to see if the trends remain the same. The reason for this is that a student can actually see an improvement in their raw score from one grade to the next, but the score itself is not high enough to move to the next ranking, so the data would make it seem like they stayed the same even though they could have improved. This is especially important if a student scored between a 1 and a 2 on one year and a 2 and 3 the next, but was measured at a 2 on both years.

The first figure that we decided to plot is a lattice plot that measures the mean score of middle school students in NYC from 2013-2017 for each grade and see if they improved. The data shows that there is a consistent improvement among the students from 6th Grade to 7th Grade, but a drastic drop from 7th Grade to 8th Grade.

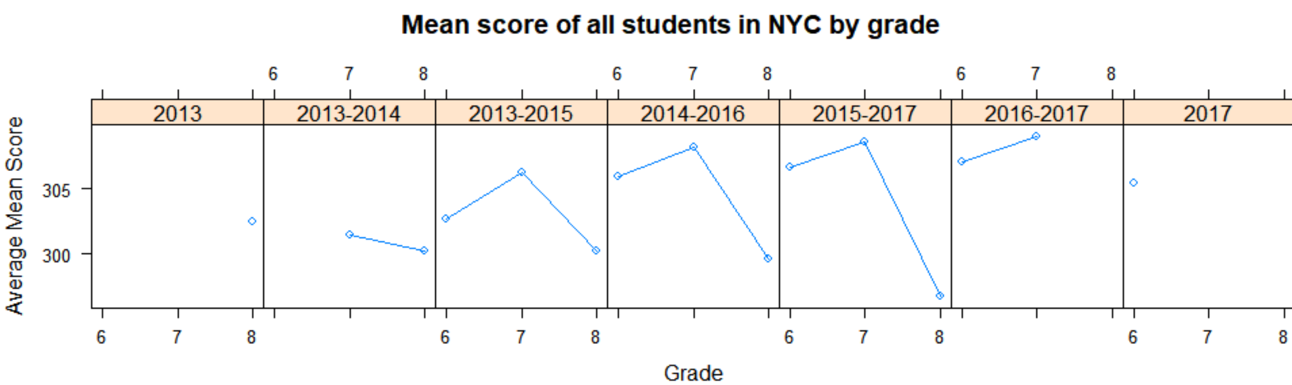

Figure 3.3.1 Mean score of all students in NYC from 2013-2017

Next we were created a correlation matrix that tracked the mean score of the group of students with at least two years of data to see if there is any correlation between them. There seems to be strong negative correlation between the two years of data students and a smaller negative correlation between the three years of data students.

Table 3.3.1: Correlation matrix of the mean score students received on their state exam in the year and grade they took exam

	2013 7th Grader	2013 6th Grader	2014 6th Grader	2015 6th Grader	2016 7th Grader
2013 7th Grader	1.000	-0.0419	0.1201	0.0989	-0.4092
2013 6th Grader	-0.0419	1.000	-0.2328	-0.2699	0.1804
2014 6th Grader	0.1201	-0.2328	1.000	-0.2744	0.0972
2015 6th Grader	0.0989	-0.2699	-0.2744	1.000	-0.0199
2016 7th Grader	-0.4092	0.1804	0.0972	-0.0199	1.000

In this figure, we break down the students by their Boroughs and see how the students perform across each Borough instead of looking at New York City as a whole. The data shows that students in the Bronx have a lower average score than students from all of the other Boroughs, and that students from Brooklyn see the smallest decrease from Grade 7 to Grade 8.

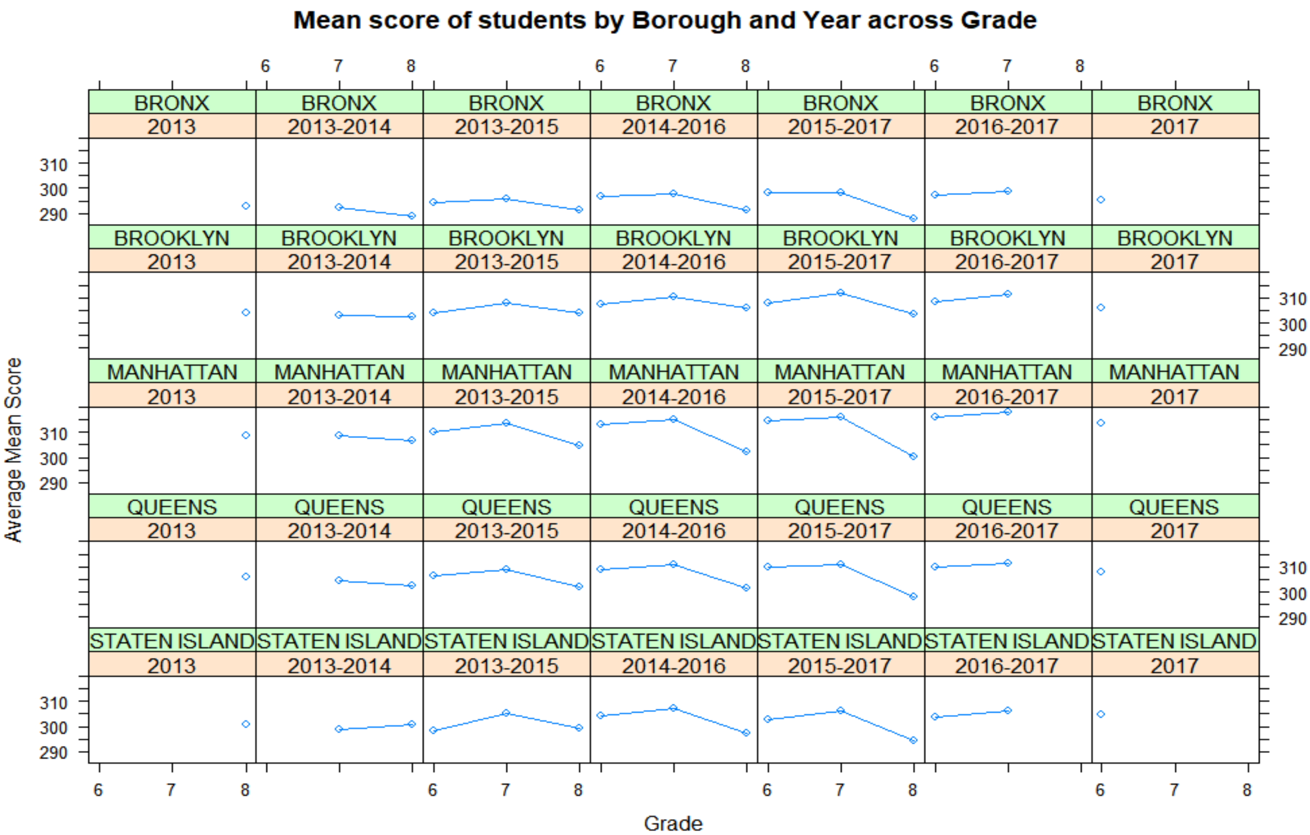

Figure 3.3.2 Average score of all students from 2013-2017 broken by their groups and Boroughs

Our last exploratory interest is to look at the data from each Borough from 2013-2017 from grades 6 to 8 without splitting the students into their groups. As most of the Boroughs show, there is a slight increase in performance from grade 6 to grade 7 but all of them have a sharp decrease in grade 8.

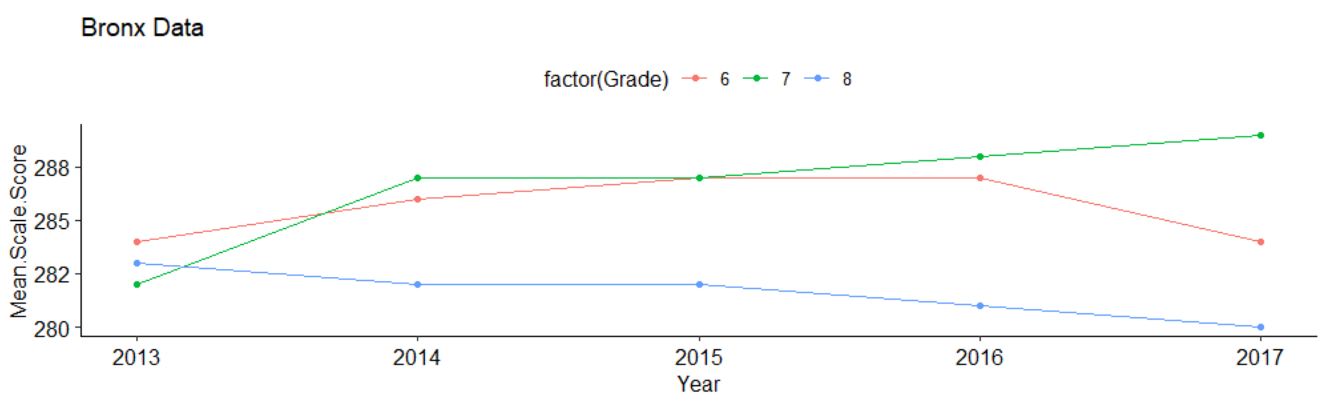

Figure 3.3.3 Mean Score of Bronx students from 2013-2017 grouped by Grade

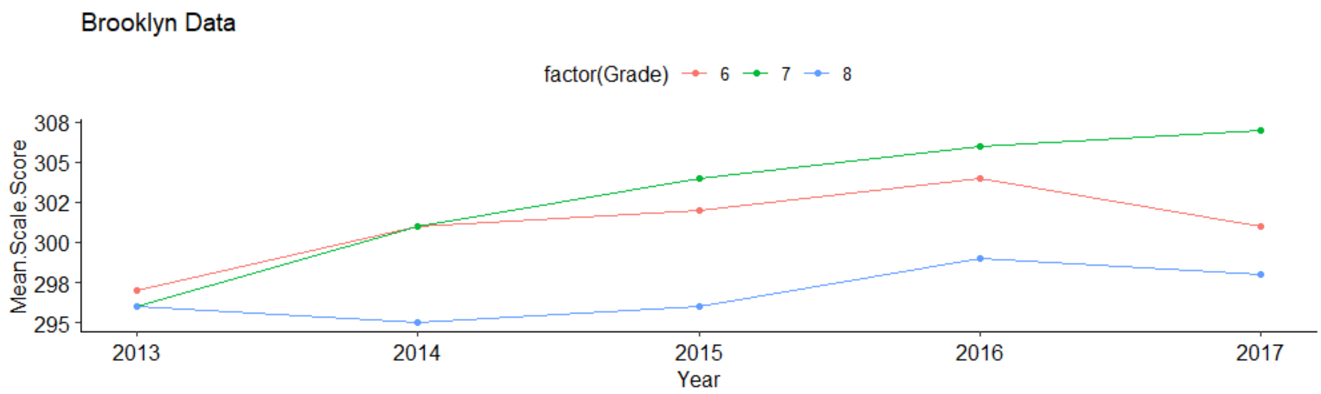

Figure 3.3.4 Mean Score of Brooklyn students from 2013-2017 grouped by Grade

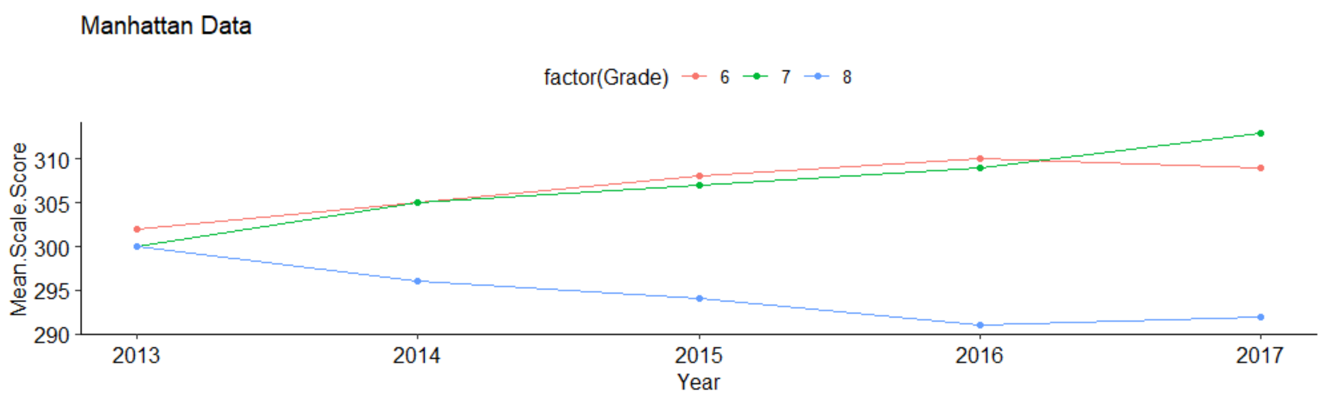

Figure 3.3.5 Mean Score of Manhattan students from 2013-2017 grouped by Grade

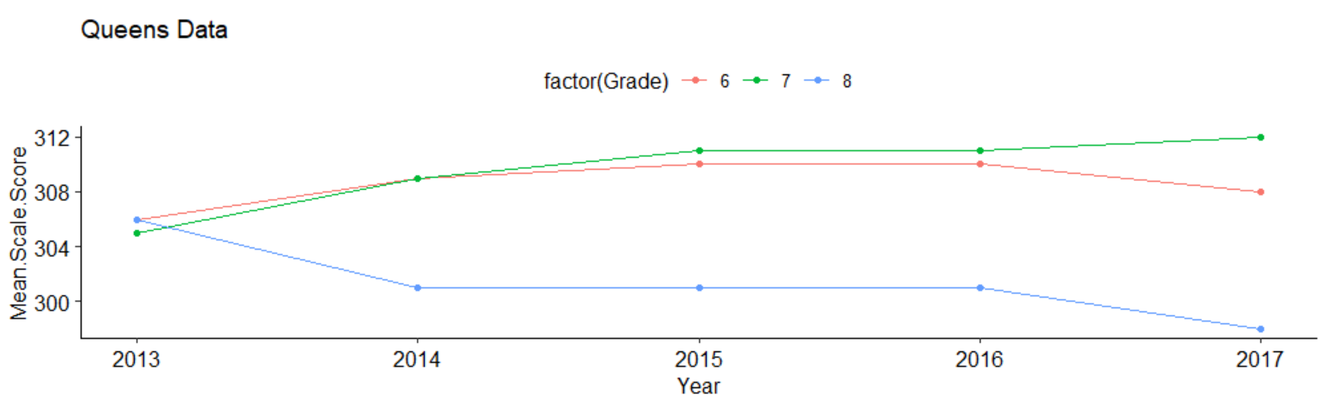

Figure 3.3.6 Mean Score of Queens students from 2013-2017 grouped by Grade

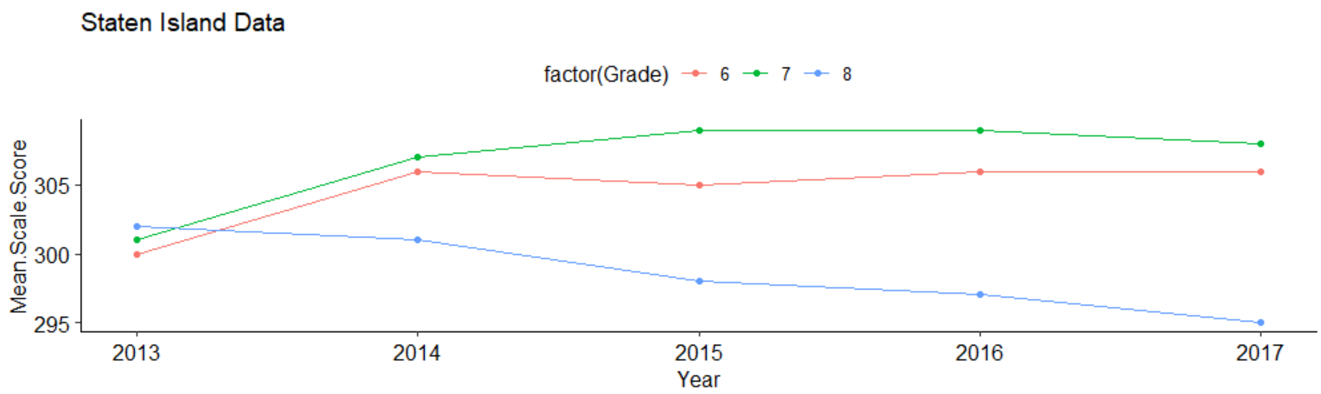

Figure 3.3.7 Mean Score of Staten Island students from 2013-2017 grouped by Grade

Section 4. Statistical Modeling

In this project, we are looking at Diagnostic Analysis instead of Predictive Analysis. This is an important distinction to bring up as Predictive Analysis usually involves training and test data to see how our model can be used to predict patterns. Diagnostic Analysis involves using only training data and interpreting it, explaining trends and correlations. We are choosing models that best explain the data and trend from 2013-2017 Grades 6 to 8 Math exam data, instead of further predicting how unknown students will perform.

Since our data set involves tracking the mean score of each Group of students (as defined in Section 2.3) from grades 6 to 8, we will be doing Longitudinal analysis, using a Mixed Effects Model that tracks Fixed Effects and Random Effects. For the Ethnicity data set, based on the variables available in this data (Table 2.3), we choose Borough, Ethnicity, Grade, and Year as our fixed effects. We use a random intercept to account for the within-group correlation among students from the same group. The models for the other data sets will be analogous, with the ethnicity variable changing to the study of interest variable.

For each study, we will fit a mixed effects model with an unstructured correlation structure, a mixed effects model with compound symmetry and autoregressive correlation structures, and a GEE model. The main R packages that we will use for modeling are lme4 which fits linear mixed effect modeling, nlme which measures linear and nonlinear mixed effects (used for autoregressive and Compound Symmetry) and geepack for the Generalized Estimating equations model.

Our interest will be in AIC, p-value, Chi-Square, ANOVA, and looking at the coefficients of each fixed effect.

We will use tidyverse and dplyr to break up our variables, as some of the variables are categorical and others may be misinterpreted for rank. We split Borough, Ethnicity, Grade, and Years into factors with Bronx, Asian, Grade 6, and 2013 as the reference levels.

Table 4.0.1 Reference Levels for each Category

Category	Reference Level	Other levels
Borough	Bronx	Brooklyn, Manhattan, Queens, Staten Island
Ethnicity	Asian	Black, White, Hispanic
Grade	6^{th} Grade	7^{th} Grade, 8^{th} Grade
Year	2013	2014, 2015, 2016, 2017

4.1 Linear Mixed Effect Models

The first approach to study longitudinal data is to use a linear mixed effects model. For this research we will focus on three correlation structures for this model, the uncorrelated structure (also known as the default structure), the compound symmetry correlation structure, and the autoregressive correlation structure. Listed below is the equation for a linear mixed effects model.

$$Y_{ij} = \underbrace{\beta_0 + \beta_1 \cdot X_{i1} + \ldots + \beta_p \cdot X_{ip}}_{\text{fixed effects}} + \underbrace{b_{i,0} + b_{i,1} \cdot X_{i1} + \ldots + b_{i,q} \cdot X_{iq}}_{\text{random effects}} + \epsilon_{ij}$$

This model deals with fixed effects being variables whose effects are constant across groups and random effects being variables that impact individuals differently. Random effects are then broken into two categories, the between subjects and within subjects.

$$Y_{ij} = \beta_0 + \beta_1 \cdot X_{ij} + \underbrace{b_{i,0} + b_{i,1} \cdot X_{ij}}_{\text{between-subject}} + \underbrace{\epsilon_{ij}}_{\text{within-subject}}$$

For our model, the equation is represented as:

Mean.Scale.Score = Borough + Ethnicity + Grade + Year + (1 | Group) + error where Borough, Ethnicity, Grade, and Year are the fixed effects and the groups are the random effects used to predict the Mean Score.

The default model had uncorrelated structure, so then we tested a linear mixed effects model with two correlation structures: Compound Symmetry and Autoregressive. The assumption of a compound symmetry structure is that the variances pooled within the subjects and the covariances pooled across the subjects would be identical. This means that there would be no differences between the within and across subjects, which is a more strict assumption. Listed below is a figure of the compound symmetry structure, which shows the correlation matrix for compound symmetry. As the figure shows, the correlation between any two measurements on a given subject is assumed to be equal.

$$\mathrm{corr}(Y_i) = \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & 1 & \dots & \rho \\ \vdots & & & \ddots & \vdots \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix}$$

Figure 4.1.1 Correlation Matrix for Compound Symmetry Structure

An autoregressive structure is often used for time series and forecasting, as it forecasts the variable of interest through a linear combination of the past values of the variable. The assumes that the time separation between measurements determines their correlation. The figure below shows an autoregressive correlation structure, which shows that the correlation is dependent on a time variable known as t.

$$\mathrm{corr}(Y_i) = \begin{bmatrix} 1 & \rho^{|t_1-t_2|} & \rho^{|t_1-t_3|} & \dots & \rho^{|t_1-t_n|} \\ \rho^{|t_2-t_1|} & 1 & \rho^{|t_2-t_3|} & \dots & \rho^{|t_2-t_n|} \\ \rho^{|t_3-t_1|} & \rho^{|t_3-t_2|} & 1 & \dots & \rho^{|t_3-t_n|} \\ \vdots & & & \ddots & \vdots \\ \rho^{|t_n-t_1|} & \rho^{|t_n-t_2|} & \rho^{|t_n-t_3|} & \dots & 1 \end{bmatrix} .$$

Figure 4.1.2 Correlation Matrix for Autoregressive Structure

4.2 Model Analysis – Mixed Effects Model Default

For our first models, we decided to run a default mixed effects model that has an unstructured correlation and we ran year as a continuous variable and a categorical variable. The reason why we performed two models for this is to see the impact of year as a whole on the mean score and also then the effect of each year separately on the mean score. The model output is listed below.

Output 4.2.1 Linear Mixed Model with Year as Continuous Variable

---

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: Mean.Scale.Score ~ Borough + Ethnicity + Grade + Year + (1 |      Group)
   Data: eth_model
     AIC      BIC   logLik deviance df.resid
  1800.0   1848.2   -887.0   1774.0      287

Scaled residuals:
     Min      1Q  Median      3Q     Max
-3.3604 -0.4511  0.0589  0.4578  2.4568

Random effects:
 Groups   Name        Variance Std.Dev.
 Group    (Intercept) 19.54    4.421
 Residual             10.50    3.241
Number of obs: 300, groups:  Group, 140

Fixed effects:
                     Estimate Std. Error      df  t value Pr(>|t|)
(Intercept)          -59.8594   473.6866 194.3131  -0.126  0.89957
BoroughBROOKLYN       11.7368     1.3425 135.5641   8.742 7.75e-15 ***
BoroughMANHATTAN      16.4695     1.3425 135.5641  12.268  < 2e-16 ***
BoroughQUEENS         12.2444     1.3425 135.5641   9.121 9.01e-16 ***
BoroughSTATEN ISLAND   7.6677     1.3425 135.5641   5.711 6.79e-08 ***
EthnicityBlack       -44.1262     1.2008 135.5641 -36.748  < 2e-16 ***
EthnicityHispanic    -37.6823     1.2008 135.5641 -31.382  < 2e-16 ***
EthnicityWhite       -10.8065     1.2008 135.5641  -9.000 1.80e-15 ***
Grade7                 1.4686     0.5076 230.2676   2.893  0.00418 **
Grade8                -5.5248     0.5939 297.3480  -9.303  < 2e-16 ***
Year                   0.1881     0.2352 194.2039   0.800  0.42483
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results showed that that year as a continuous random variable has a positive impact on the mean score, which means that there is positive correlation between year and test score result. The data also shows students from Bronx perform worst than students from the other Boroughs while Manhattan performed the best, and students of Asian ethnicity perform the best while students of Black ethnicity perform the worst. The results also show that students improve from grade 6 to grade 7 but then have a massive decline in grade 8. The AIC for this model was 1800 and according to the t-tests p-values, it seems that Year is not significant as it fails an alpha of 0.05 test.

We then decided to run the model again, but this time splitting year into factors to see how each year would compare to one another. The results are listed below.

Output 4.2.2 Linear Mixed Model with Year as Categorical Variables

```
Linear mixed model fit by maximum likelihood . t-tests use Satterthwaite's method ['lmerModLmerTest']
Formula: Mean.Scale.Score ~ Borough + Ethnicity + Grade + Year_ + (1 |      Group)
   Data: eth_model

     AIC      BIC   logLik deviance df.resid
  1756.2   1815.4   -862.1   1724.2      284

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.5617 -0.4594  0.0467  0.5226  2.4447

Random effects:
 Groups   Name        Variance Std.Dev.
 Group    (Intercept) 23.651   4.863
 Residual              7.134   2.671
Number of obs: 300, groups:  Group, 140

Fixed effects:
                     Estimate Std. Error      df t value Pr(>|t|)
(Intercept)          317.1194     1.3298 157.4248 238.475  < 2e-16 ***
BoroughBROOKLYN       11.7139     1.4038 133.2613   8.344 7.97e-14 ***
BoroughMANHATTAN      16.4946     1.4038 133.2613  11.750  < 2e-16 ***
BoroughQUEENS         12.2595     1.4038 133.2613   8.733 9.08e-15 ***
BoroughSTATEN ISLAND   7.6931     1.4038 133.2613   5.480 2.05e-07 ***
EthnicityBlack       -44.1688     1.2556 133.2613 -35.176  < 2e-16 ***
EthnicityHispanic    -37.6930     1.2556 133.2613 -30.019  < 2e-16 ***
EthnicityWhite       -10.7852     1.2556 133.2613  -8.589 2.03e-14 ***
Grade7                 1.7421     0.4428 255.8738   3.934 0.000107 ***
Grade8                -5.4987     0.5603 269.3670  -9.815  < 2e-16 ***
Year_2014              3.4222     0.5698 224.9873   6.006 7.58e-09 ***
Year_2015              3.8156     0.6863 299.6723   5.560 5.98e-08 ***
Year_2016              2.8830     0.8320 255.8485   3.465 0.000621 ***
Year_2017              0.9315     0.9641 183.0235   0.966 0.335242
```

In this result, we see that all subsequent years after 2013 performed better than 2013, with a steady rise from 2014-2015 and then a drop that is especially significant in 2017. The p-value though shows that 2017 may be more of an aberration than anything significant. There is also a sizable improvement in AIC from 1800 to 1756. From this point on, we decided to measure year as a categorical variable as we felt that it provided more information.

4.3 Model Analysis – Mixed Effects Model Correlation Structures

Next we decided to look at the compound symmetry correlation structure of a mixed effects model to see the impact if we assumed no difference between the subjects and within the subjects. Listed below is the model results and a variance-covariance matrix of the results from the compound symmetry structure.

Output 4.3.1 Linear Mixed Effects Model with Compound Symmetry Correlation Structure

```
Generalized least squares fit by REML
  Model: Mean.Scale.Score ~ Ethnicity + Borough + Grade + Year_
  Data: eth_model
       AIC      BIC    logLik
  1738.077 1796.573 -853.0385

Correlation Structure: Compound symmetry
 Formula: ~1 | Group
 Parameter estimate(s):
      Rho 0.7762614

Coefficients:
                         Value Std.Error   t-value p-value
(Intercept)           317.1168 1.3743429 230.74070  0.0000
EthnicityBlack        -44.1716 1.2997219 -33.98545  0.0000
EthnicityHispanic     -37.6937 1.2997219 -29.00136  0.0000
EthnicityWhite        -10.7838 1.2997219  -8.29703  0.0000
BoroughBROOKLYN        11.7124 1.4531333   8.06012  0.0000
BoroughMANHATTAN       16.4962 1.4531333  11.35216  0.0000
BoroughQUEENS          12.2605 1.4531333   8.43727  0.0000
BoroughSTATEN ISLAND    7.6948 1.4531333   5.29531  0.0000
Grade7                  1.7491 0.4517880   3.87150  0.0001
Grade8                 -5.4972 0.5757313  -9.54824  0.0000
Year_2014               3.4389 0.5805060   5.92397  0.0000
Year_2015               3.8272 0.7027641   5.44596  0.0000
Year_2016               2.8868 0.8555864   3.37403  0.0008
Year_2017               0.9263 0.9950293   0.93088  0.3527
```

The first thing that stood out from the compound symmetry correlation structure is that the AIC increased as opposed to an unstructured correlation model. The next thing of interest is that the p-value of year 2017 was again not significant, so it shows that it may be more of an aberration than anything definitive, since all three models so far had shown a non-significant p-value for 2017 data.

We also display the output of the variance-covariance matrix for compound symmetry as listed below. This specific variance covariance matrix focused on an Bronx Asian student from 2013-2015 and produced a 3 by 3 matrix for each year.

Table 4.3.2 Compound Symmetry Variance-Covariance Matrix

	2013	2014	2015
2013	32.847	25.498	25.498
2014	25.498	32.847	25.498
2015	25.498	25.498	32.847

Table 4.3.3 Compound Symmetry Variance-Covariance Matrix

	2013	2014	2015
2013	1.00000	0.77626	0.77626
2014	0.77626	1.00000	0.77626
2015	0.77626	0.77626	1.00000

Lastly, we ran a linear mixed model with an autoregressive structure to see the result if we did not assume the correlation to be interchangeable, but instead dependent on time.

Output 4.3.4 Linear Effects model with Autoregressive Correlation Structure

---

```
Generalized least squares fit by REML
  Model: Mean.Scale.Score ~ Ethnicity + Borough + Grade + Year_
  Data: eth_model
       AIC      BIC   logLik
  1724.028 1782.524 -846.014

Correlation Structure: AR(1)
 Formula: ~1 | Group
 Parameter estimate(s):
      Phi 0.8150158

Coefficients:
                          Value Std.Error   t-value p-value
(Intercept)            317.0478 1.3734023 230.84843  0.0000
EthnicityBlack         -44.2471 1.2957760 -34.14718  0.0000
EthnicityHispanic      -37.6768 1.2957760 -29.07661  0.0000
EthnicityWhite         -11.0867 1.2957760  -8.55604  0.0000
BoroughBROOKLYN         11.6501 1.4487217   8.04162  0.0000
BoroughMANHATTAN        16.0998 1.4487217  11.11313  0.0000
BoroughQUEENS           12.0659 1.4487217   8.32866  0.0000
BoroughSTATEN ISLAND     7.3929 1.4487217   5.10307  0.0000
Grade7                   2.1174 0.4150444   5.10173  0.0000
Grade8                  -5.0379 0.6082504  -8.28257  0.0000
Year_2014                3.3370 0.5414724   6.16290  0.0000
Year_2015                3.8354 0.7257406   5.28475  0.0000
Year_2016                3.1788 0.8580398   3.70474  0.0003
Year_2017                0.9893 0.9979666   0.99130  0.3224
```

The results gave us the best AIC of all of the linear mixed models with different correlated structures, although it also still showed that year 2017 data was not significant. The autoregressive correlated structure model also showed less variability between 2014-2016 than the other models, showing that there may be some forecasting that can be done.

Lastly, we showed the variance-covariance matrix for autoregressive strucutre, which again focused on the Bronx 2013-2015 Asian subjects.

Table 4.3.5 Compound Symmetry Variance-Covariance Matrix

	2013	2014	2015
2013	32.806	26.737	21.791
2014	26.737	32.806	26.737
2015	21.791	26.737	32.806

Table 4.3.6 Compound Symmetry Correlation Matrix

	2013	2014	2015
2013	1.00000	0.81502	0.66425
2014	0.81502	1.00000	0.81502
2015	0.66425	0.81502	1.00000

4.4 Generalized Estimating Equations (GEE) approach and Analysis

A second regression approach for inference with longitudinal data is known as the generalized estimating equations approach of GEE approach. A GEE approach requires a regression model and a working correlation model. The GEE approach simply models the mean response instead of attempting to model the within-subject covariance structure. For this approach, we applied a GEE approach with Compound symmetry correlation to observe the impact of each variable on the mean score.

Output 4.4.1 GEE Approach with Compound Symmetry Correlation Structure

```
GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
  gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
  Link:                      Identity
  Variance to Mean Relation: Gaussian
  Correlation Structure:     Exchangeable

Call:
```

```
gee(formula = Mean.Scale.Score ~ Ethnicity + Borough + Grade +
    Year_, id = Group, data = eth_model, corstr = "exchangeable")

Summary of Residuals:
    Min     1Q  Median     3Q    Max
-13.884  -4.064  -0.840   2.706  17.520

Coefficients:
                        Estimate Naive S.E. Naive z Robust S.E. Robust z
(Intercept)            317.1442     1.3175 240.718      1.3994 226.6310
EthnicityBlack         -44.1395     1.2241 -36.059      1.2131 -36.3853
EthnicityHispanic      -37.6856     1.2241 -30.787      1.1105 -33.9370
EthnicityWhite         -10.7997     1.2241  -8.823      1.2864  -8.3955
BoroughBROOKLYN         11.7296     1.3686   8.571      1.3016   9.0120
BoroughMANHATTAN        16.4776     1.3686  12.040      1.8816   8.7573
BoroughQUEENS           12.2492     1.3686   8.950      1.4408   8.5019
BoroughSTATEN ISLAND     7.6757     1.3686   5.609      1.3477   5.6953
Grade7                   1.6722     0.4924   3.396      0.3512   4.7619
Grade8                  -5.5141     0.5869  -9.395      0.6239  -8.8387
Year_2014                3.2539     0.6406   5.079      0.5544   5.8697
Year_2015                3.6994     0.7399   5.000      0.7122   5.1941
Year_2016                2.8489     0.8643   3.296      0.8397   3.3929
Year_2017                0.9857     0.9691   1.017      1.0469   0.9416

Estimated Scale Parameter:  30.58
Number of Iterations:  4

Working Correlation
       [,1]   [,2]   [,3]
[1,] 1.0000 0.6868 0.6868
[2,] 0.6868 1.0000 0.6868
[3,] 0.6868 0.6868 1.0000
```

The results of the GEE Approach was similar to the results of the linear effects model with compound symmetry correlation structure. The coefficient estimates for all of the models were around the same, showing that ethnicity, borough, grade, and year do play an impact on the student performance in the state exam. They also all show that there is growth among the students from 6^{th} to 7^{th} grade, but also a drop when the students take the 8^{th} grade exam. The growth in exam scores from 2013-2016 seemed steady, but then dropped substantially in 2017.

4.5 Results from other data sets

Since the other data sets used the same models and approaches, we decided to simply show the models with the best AIC for each of the data set and speak of them specifically. We will list the code in the reference section for each of the models in case someone wants to look into a specific model that was not referenced in this section.

For the gender data set, we had female as the reference level and compared male and female performance in addition to borough, grade, and year. The results of our linear mixed effects model with compound symmetry correlation is listed below.

Output 4.5.1 Linear Mixed-Effects model with Compound Symmetry Correlation for Gender

```
Generalized least squares fit by REML
   Model: Mean.Scale.Score ~ Gender + Borough + Grade + Year
   Data: gend_model
   AIC BIC logLik
   756 797   -364

Correlation Structure: Compound symmetry
  Formula: ~1 | Group
  Parameter estimate(s):
    Rho 0.1984

Coefficients:
                        Value Std.Error t-value p-value
(Intercept)           286.21    0.8662   330.4  0.0000
GenderMale             -3.97    0.5498    -7.2  0.0000
BoroughBROOKLYN        15.67    0.8694    18.0  0.0000
BoroughMANHATTAN       18.36    0.8694    21.1  0.0000
BoroughQUEENS          21.89    0.8694    25.2  0.0000
BoroughSTATEN ISLAND   18.90    0.8694    21.7  0.0000
Grade7                  1.33    0.5457     2.4  0.0164
Grade8                 -6.58    0.5591   -11.8  0.0000
Year2014                2.35    0.7118     3.3  0.0012
Year2015                2.93    0.7347     4.0  0.0001
Year2016                3.22    0.7595     4.2  0.0000
Year2017                2.50    0.7669     3.3  0.0014
```

The results of this model show that males perform worst than females on the state exam for middle school and Bronx is again the worst performing borough. This data also shows that the p-value for year 2017 is significant. In the gender data set compared to the ethnicity data set, the coefficients of the boroughs are more closely related with Queens performing the best as opposed to Manhattan for ethnicity. Also there is an even more substantial drop off in Grade 8 performance. Though it seems that for gender, there is a larger improvement from 2013 to 2014 before the impact flattens out before a sharp decline in 2017.

Next we looked at the ELL data set to see how students performed from 2014-2017. The variable of interest in this case, ELL, is split into three categories: students who are currently considered ELL (English as their second language), students who were ELL but not anymore, and students that were never ELL. In this study, students that are currently still considered ELL is the reference level. The results of the data is shown below using a GEE approach with autoregressive correlation structure, since the linear mixed effects model approach showed that the autoregressive correlation structure had the best AIC.

Output 4.5.2 GEE Approach with AR Correlation structure for ELL data set

```
geeglm(formula = Mean.Scale.Score ~ ELL + Borough + Grade + Year,
        data = ell_model, id = Group, corstr = "ar1")

  Coefficients:

                          Estimate  Std.err   Wald Pr(>|W|)
(Intercept)        262.3385   0.9087 83338.32  < 2e-16 ***
ELLEver ELL         37.3746   0.8157  2099.26  < 2e-16 ***
ELLNever ELL        29.2121   1.0198   820.59  < 2e-16 ***
BoroughBROOKLYN     16.2803   1.2305   175.04  < 2e-16 ***
BoroughMANHATTAN    13.8420   1.0945   159.94  < 2e-16 ***
BoroughQUEENS       18.4415   0.7600   588.80  < 2e-16 ***
BoroughSTATEN ISLAND 10.6576  1.1395    87.48  < 2e-16 ***
Grade7               1.4742   0.3755    15.42  8.6e-05 ***
Grade8              -5.8639   0.6864    72.98  < 2e-16 ***
Year2015             0.1724   0.6336     0.07     0.79
Year2016            -0.0294   0.7391     0.00     0.97
Year2017            -0.3126   0.8602     0.13     0.72
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The first thing that stood out from this analysis is that the years 2015-2017 are all not significant when compared to the reference level of year 2014, this is the first data set which showed that the years were not a significant indicator. The other thing of interest is that students who were once classified as ELL had a higher coefficient estimate compared to the reference level of current ELL students than the students who were never classified as ELL. This would be an area where more research can be done to see the influence of former ELL students to never ELL students, to see if there is a significant correlation between the two.

Next we looked at students with disability (SWD) data set, where students without disability was used as the reference level and compared to students with disabilities. The model chosen was a linear mixed effects model with autoregressive correlation structure.

Output 4.5.3 Linear Mixed Effects model with AR correlation structure for SWD data

```
Generalized least squares fit by REML
    Model: Mean.Scale.Score ~ SWD + Borough + Grade + Year
    Data: swd_model
    AIC BIC logLik
```

```
    760 801   -366

Correlation Structure: AR(1)
 Formula: ~1 | Group
 Parameter estimate(s):
   Phi 0.666

Coefficients:
                          Value Std.Error t-value p-value
(Intercept)         293.9     1.049   280.3  0.0000
SWDSWD              -37.6     0.758   -49.5  0.0000
BoroughBROOKLYN      11.4     1.199     9.5  0.0000
BoroughMANHATTAN     15.3     1.199    12.7  0.0000
BoroughQUEENS        16.8     1.199    14.0  0.0000
BoroughSTATEN ISLAND 16.4     1.199    13.7  0.0000
Grade7                2.0     0.446     4.5  0.0000
Grade8               -5.0     0.590    -8.5  0.0000
Year2014              4.0     0.594     6.7  0.0000
Year2015              4.0     0.736     5.5  0.0000
Year2016              3.8     0.809     4.6  0.0000
Year2017              2.2     0.881     2.5  0.0141
```

The one thing of interest from this data set is that all variables were seen as significant according to their p-values, which no other model from the other data sets had shown yet. Besides that, the trends remain the same for Borough, Grade, and Year. Also as expected, students without disabilities perform much better than students with disabilities.

Lastly, we looked at the econ data set which looked at students who were economically disadvantaged and students that were not economically disadvantaged. The way this was measured is by looking at students who are eligible for free or reduced lunch and students not eligible. The reference level in this case are students that are economically disadvantaged, so they are eligible for free or reduced lunch and their parents income is likely on the poverty line or below. In this data set, we looked at a GEE approach with compound symmetry correlation.

Output 4.5.4 GEE Approach with compound symmetry correlation structure for econ data set

---

```
Call:
geeglm(formula = Mean.Scale.Score ~ ECON + Borough + Grade +
      Year, data = econ_model, id = Group, corstr = "exchangeable")

 Coefficients:
                       Estimate  Std.err     Wald Pr(>|W|)
(Intercept)          280.5054   1.6339 29473.23  < 2e-16 ***
ECONNot Econ Disadv   18.3328   1.0664   295.53  < 2e-16 ***
BoroughBROOKLYN       17.4984   1.4836   139.11  < 2e-16 ***
BoroughMANHATTAN      20.3486   2.2806    79.61  < 2e-16 ***
BoroughQUEENS         21.7369   1.6793   167.54  < 2e-16 ***
BoroughSTATEN ISLAND  17.2328   1.5328   126.40  < 2e-16 ***
Grade7                 0.6714   0.5382     1.56   0.2122
```

```
Grade8            -8.2756  1.0334   64.13  1.1e-15 ***
Year2014           2.7445  0.9346    8.62  0.0033 **
Year2015           1.0011  1.2917    0.60  0.4383
Year2016           0.9701  1.3139    0.55  0.4603
Year2017          -0.0359  1.4748    0.00  0.9806
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

In this data set there is a much larger drop off for grade 8 student performance as opposed to grade 6 and grade 7 student performance compared to the other data sets, and that 2015-2017 coefficient estimate are not significant because of the p-value. Also students who are economically disadvantaged perform worst than students who are not economically disadvantaged.

Section 5. Discussion of Results

The Common Core exam was meant to prepare students for college and the real world. It increased the rigor of the Math exams and the mission was to give students the fundamental skills to solve problems. The results show that these desired goals have not yet been reached yet. The improvement from Grade 6 to Grade 7 showed that if the standards remain the same but get more complex, students saw a slight improvement. But in Grade 8 when the standards changed the students were supposed to take what they learned and use it to solve different problems with these foundational skills, they struggled mightily to achieve this. The goal was to build a foundation to solve different kinds of problems, not to use what is learned to solve the same problems and that is where the Common Core seemed to have failed.

One thing to keep in mind, is that in 8th Grade, students are allowed to take the Algebra regents over the 8th Grade Common Core exam, which many of the top performing students opt to do to skip the beginning level of High School Math. This hurts the overall performance of 8th Graders for the Math Common Core state exam, as the top students who performed in 7th Grade probably did not take the exam in 8th Grade. Since we can not track an individual student and instead look at the group of subjects, we do not know for sure how many students opted out of the 8^{th} Grade exam to take the Algebra regents, as it is possible for students to take both.

The problem with this thinking is that the evidence shown, even with this in mind, still shows that the Common Core has not met its goals. According to NAEP, the United States have not grown substantially in 8th Grade since the implementation of Common Core. Since the implementation of Common Core in 2013, the United States have fluctuated between 282-285 as an average score with the Proficiency at 299, so the nation as a whole has not reached proficiency in the six years. In addition, the United States scored highest in 2013 and has steadily declining since, so one can not even use the argument of continual growth.

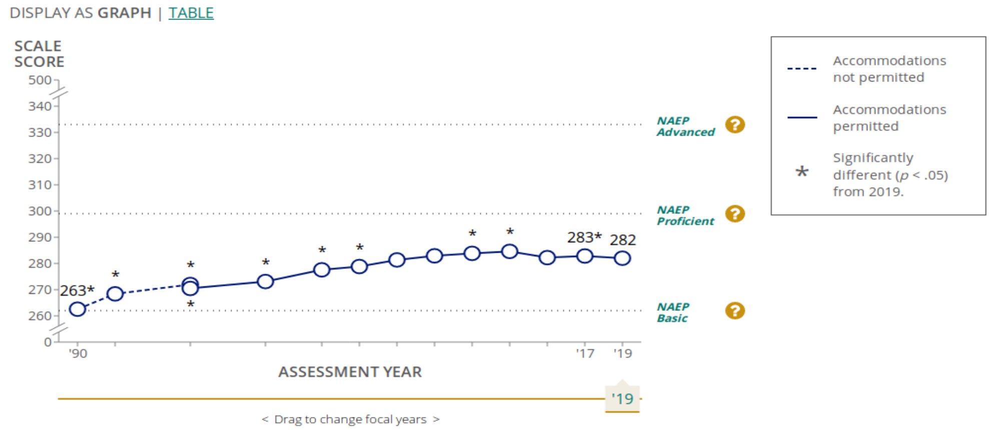

Figure 5.1.1 United States NAEP score from 2000 to 2019

In addition to the NAEP report of the United States as a nation, the data for New York City specifically hasn't shown much improvement either. Since the implementation of Common Core in 2013, New York City students have generally scored below the National average, which is a shift from when they used to show above the nation before the Common Core was added.

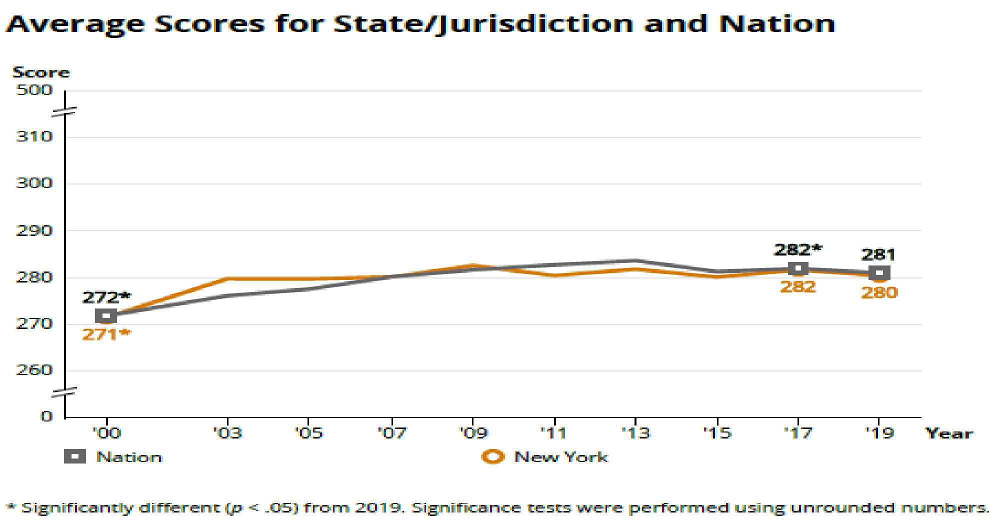

Figure 5.1.2 New York City NAEP score from 2000 to 2019

In addition to that, an article shows that the first students who started High School with the implementation of Common Core are said to be the worst prepared for college in 15 years. The goal of the Common Core was to prepare students for College and it seems that the Common Core has not done that, there has been little growth in student performance in middle school, little growth in New York City (and United States) performance in NAEP, and the students who started High School with Common Core are the least prepared students for college.

Also startling to see is that the gap between different ethnicity and Boroughs remain large throughout the years of Common Core. Asian and White students still perform substantially better than their Black and Hispanic counterparts. Also the Bronx still performs substantially worst than all of the other Boroughs. Part of the initiative in education was to close the gap and it seems that has not been achieved either. This is even more alarming because it isn't just the school test result data that shows this, but also the NAEP data breaking down Ethnicity shows the same startling gap.

Results for Student Groups in 2019

Reporting Groups	Percentage of students	Avg. score	Percentage at or above *NAEP* *Basic*	*Proficient*	Percentage at *NAEP* *Advanced*
Race/Ethnicity					
White	46	292	77	44	15
Black	17	258	45	15	3
Hispanic	26	263	51	17	3
Asian	9	309	85	60	31
American Indian/Alaska Native	1	‡	‡	‡	‡
Native Hawaiian/Pacific Islander	#	‡	‡	‡	‡
Two or more races	2	‡	‡	‡	‡
Gender					
Male	51	281	65	34	12
Female	49	280	67	33	10
National School Lunch Program					
Eligible	50	266	53	21	5
Not eligible	48	294	79	46	17

Figure 5.1.3 NAEP rank split up by Ethnicity, Gender, and Economic Status

It isn't all bad though and there is still hope for the future as we continue using the Common Core standards. The increase in student performance from 6th Grade to 7th Grade shows that if the standards remain similar but rigor increases, students are able to utilize their foundational skills and solve more challenging problems. They are not yet able to take this skill to understand that these foundational skills can be used to solve problems that may look different but are actually very similar to the problems they have solved (as shown from the drop in 8th Grade scores). Another positive is that there seems to be an upward trend for student performance after each year. The 2014 6th Graders performed better than the 2013 6th Graders and the 2015 6th Graders performed better than the 2014 6th Graders. This is further exemplified when we took year as a continuous and categorical variable and saw a positive growth in both results. Although students did not continue to grow after each year, they were all better than the control year of 2013 when Common Core was first implemented. This shows that there is some benefit here. The NAEP data also shows this as less students in New York City are performing basic or below and more are performing in a Proficient or Advanced level since 2000, so that is one positive.

NAEP Achievement-Level Percentages and Average Score Results

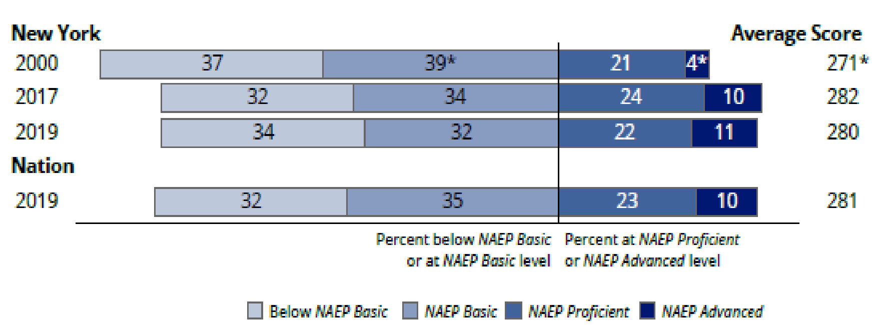

Figure 5.1.4 New York City NAEP ranked score broken up from 2000, 2017, and 2019

One last thing to keep in mind, which none of the data would show is the importance of the tests that are administered. In New York City, most students apply for High School during their 8th Grade year and find out the decision of their High School a few months prior to taking the Common Core state exam. For many High School admissions, the schools focus predominantly on the 7th Grade Math and Reading exam and also take a look at the 6th Grade exams too. They do not look at the 8th Grade exams at all as they had made their decision before it was even administered.

This is important, as the 7th Grade exam is seen as the student's last chance to showcase their testing abilities to their prospective High Schools, whereas the 8th Grade exam is not seen by any admissions office at all. The fact that there is a growth from 6th Grade to 7th Grade in the Math State exam shows that there is some impact of Common Core, at least in terms of High School admissions. And one thing to keep in mind with the lack of participation and overall drop of performance in the 8th Grade Math exam could also deal with lack of interest from the students, as they know it will not likely effect their admissions.

In the end, has the growth been what we hoped or expected in the six years since Common Core was implemented? No. But has there been growth shown from the implementation of Common Core? That answer is yes. Now as we continue to a new decade, we will see how much students can improve with even more years of experience of Common Core. There is an old saying that Rome was not built in one day and that is one of the points that need to be taken into consideration when measuring the success of Common Core. The actual results may not be fully realized until the end of the 2020s, but the data has also shown that the actual results may not be what we hoped for either as some studies show that students are even less prepared now than they were before the switch to Common Core.

Section 6.1 References:

New York State Education Department. [2019]. “Timeline & History of New York State Assessments.” <http://www.p12.nysed.gov/assessment/timeline-historyrev.pdf>

National Assessment Governing Borough. [2019]. "NAEP Report Card Mathematics". <https://www.nationsreportcard.gov/mathematics/nation/scores/?grade=8>

Meador, Derrick. [2019]. "What Are Some Pros and Cons of the Common Core State Standards?" <https://www.thoughtco.com/common-core-state-standards-3194603>

State Department of New York. [2019]. "New York State P-12 Common Core Learning Standards for Mathematics." <https://www.engageny.org/resource/new-york-state-p-12-common-core-learning-standards-for-mathematics/file/741>

Brunton, Jeremy A. [2019]. "New York State Common Core Test Results: How to read your child's test results." <https://www.lumoslearning.com/llwp/teachers-speak/new-york-state-common-core-test-results-how-to-read-your-childs-test-results-by-jeremy-a-brunton.html>

O'Hare, JP and Jeanne Beattie. [2019]. "State Education Department Releases Spring 2019 Grades 3-8 ELA & Math Assessment Results." <http://www.nysed.gov/news/2019/state-education-department-releases-spring-2019-grades-3-8-ela-math-assessment-results>

Campbell, Jon. [2017]. "Education: Opt-out movement remains strong." <https://www.lohud.com/story/news/politics/albany/2017/03/02/opt-out-movement-new-york/98608956/>

McMahon, Julie. [2019]. "Why are our kids doing worse in math as they get older? (Look up state test scores)." <https://www.syracuse.com/schools/2016/10/why_are_our_kids_doing_worse_in_math_as_they_get_older_look_up_state_test_scores.html>

Tibco Software Inc. [2019]. “Assumptions and Effects of Violating Assumptions – Sphericity and Compound Symmetry.” <https://documentation.statsoft.com/STATISTICAHelp.aspx?path=Gxx/ANOVAMANOVA/Overview/AssumptionsSphericityandCompoundSymmetry>

Hyndman, R.J., & Athanasopoulos, G. (2018) *Forecasting: principles and practice*, 2nd edition, OTexts: Melbourne, Australia. OTexts.com/fpp2.

Pullmann, Joy. [2019]. "First common Core High School Grads Worst-Prepared For College in 15 Years." <https://thefederalist.com/2019/10/31/first-common-core-high-school-grads-worst-prepared-for-college-in-15-years/>

Bates, Douglas M. [2010]. *lme4: Mixed-effects modeling with R*.

Belle, Van et al. [2002]. “Chapter 1: Longitudinal Data Analysis.” <https://faculty.washington.edu/heagerty/Courses/VA-longitudinal/private/LDAchapter.pdf>

Section 6.2 R Code

```
setwd("C:/Users/thoma/Desktop/Thomas Research Set/")
ethnicity <- read.csv("Ethnicity_Borough.csv")
library(tidyverse)
library(dplyr)
library(car)
library(lattice)
library(lme4)
library(nlme)
library(MASS)
library(lmerTest)
library(geepack)
library(ggthemes)
library(ggplot2)
library(gridExtra)
library(gee)

##########################################    Introduction
#Number of Proficient
number_proficient_borough <- ethnicity %>%
  group_by(Borough, Group, Grade) %>%
  summarise(Sum=sum(X..Level.3.4, na.rm = TRUE))

#lattice plot Number
xyplot(Sum~Grade | factor(Group)+factor(Borough), data=number_proficient_borough,
       scales=list(x=list(at=c(6,7,8))),
       xlab = "Grade",
       ylab = "Total Students Proficient",
       main = "Number of students ranked as proficient by Borough and Year across Grade",
       type="b",
       as.table=TRUE)


#Percent of Proficient
percent_proficient_borough <- ethnicity %>%
  group_by(Borough, Group, Grade) %>%
  summarise(Avg=mean(X..Level.3.4.1, na.rm = TRUE))

#lattice plot Percent
xyplot(Avg~Grade | factor(Group)+factor(Borough), data=percent_proficient_borough,
       scale=list(x=list(at=c(6,7,8))),
       xlab = "Grade",
       ylab = "Percentage of Students Proficient",
       main = "Percentage of students ranked as proficient by Borough and Year across Grade",
       type="b",
       as.table=TRUE)

################################################   Materials

#Number Students Tested
number_tested_borough <- ethnicity %>%
  group_by(Borough, Group, Grade) %>%
```

```
  summarise(Sum=sum(Number.Tested, na.rm = TRUE))

#Lattice Plot number Tested
xyplot(Sum~Grade | factor(Group)+factor(Borough), data=number_tested_borough,
       scale=list(x=list(at=c(6,7,8))),
       xlab = "Grade",
       ylab = "Students Tested",
       main = "Total number of students tested by Borough and Year across Grade. ",
       type="b",
       as.table=TRUE)

#Testers
df <- matrix(c(70305, 59875, 68091, 56999, 66259, 54762, 65371, 53102), 4, 2, byrow=TRUE)
colnames(df) <- c("Grade 7", "Grade 8")
rownames(df) <- c("2013-2014", "2014-2015", "2015-2016", "2016-2017")
df

######################################################  Modeling

#Average Score Overall per Group
ethnicity_lattice <- ethnicity %>%
  group_by(Group, Grade) %>%
  summarise(Avg=mean(Mean.Scale.Score, na.rm = TRUE))
#Average Score Per Borough per Group
ethnicity_lat_borough <- ethnicity %>%
  group_by(Borough, Group, Grade) %>%
  summarise(Avg=mean(Mean.Scale.Score, na.rm = TRUE))

#Lattice Plot Average Score Overall per Group
xyplot(Avg~Grade | factor(Group), data=ethnicity_lattice,
       scale=list(x=list(at=c(6,7,8))),
       xlab = "Grade",
       ylab = "Average Mean Score",
       main = "Mean score of all students in NYC by grade",
       type="b",
       layout=c(7,1))

#Lattice Plot Average Score per Borough per Group
xyplot(Avg~Grade | factor(Group)+factor(Borough), data=ethnicity_lat_borough,
       scale=list(x=list(at=c(6,7,8))),
       xlab = "Grade",
       ylab = "Average Mean Score",
       main = "Mean score of students by Borough and Year across Grade",
       type="b",
       as.table=TRUE)

#Borough trend
Bronx <- read.csv("Bronx.csv")
Brooklyn <- read.csv("Brooklyn.csv")
Manhattan <- read.csv("Manhattan.csv")
Queens <- read.csv("Queens.csv")
Staten_Island <- read.csv("Staten Island.csv")

ggplot(data=Bronx, aes(x=Year, y=Mean.Scale.Score, color=factor(Grade)))+geom_line()+geom_point()+ggtitle("Bronx Data")
ggplot(data=Brooklyn, aes(x=Year, y=Mean.Scale.Score, color=factor(Grade)))+geom_point()+geom_line()+ggtitle("Brooklyn Data")
ggplot(data=Manhattan, aes(x=Year, y=Mean.Scale.Score, color=factor(Grade)))+geom_point()+geom_line()+ggtitle("Manhattan Data")
```

```
ggplot(data=Queens, aes(x=Year, y=Mean.Scale.Score, color=factor(Grade)))+geom_point()+geom_line()+ggtitle("Queens Data")
ggplot(data=Staten_Island, aes(x=Year, y=Mean.Scale.Score, color=factor(Grade)))+geom_point()+geom_line()+ggtitle("Staten Island Data")


#corr matrix
mat <- with(eth_model, matrix(c(Mean.Scale.Score[Group==0], Mean.Scale.Score[Group==1], Mean.Scale.Score[Group==2],
Mean.Scale.Score[Group==3], Mean.Scale.Score[Group==4]), ncol=5))
dr <- cor(mat)
colnames(dr) <- c("Group 0", "Group 1", "Group 2", "Group 3", "Group 4")
rownames(dr) <- c("Group 0", "Group 1", "Group 2", "Group 3", "Group 4")
dr

####################### Ethnicity Model

#convert into factors
ethnicity <- read.csv("Ethnicity_Borough2.csv")

ethnicity <- ethnicity %>%
  mutate(
    Borough = as.factor(Borough),
    Ethnicity = as.factor(Ethnicity),
    Grade = as.factor(Grade),
    Year_ = as.factor(Year))
eth_model <- ethnicity %>%
  dplyr::select(Group, Borough, Grade, Year, Year_, Ethnicity, Mean.Scale.Score)

#With Year
mixed_eth_model1 <- lmerTest::lmer(Mean.Scale.Score~Borough+Ethnicity+Grade+Year+(1|Group), data=eth_model, REML=FALSE)
mixed_eth_model2 <- lmerTest::lmer(Mean.Scale.Score~Borough+Ethnicity+Grade+Year_+(1|Group), data=eth_model, REML=FALSE)

#eth_model1
summary(mixed_eth_model1)
print(mixed_eth_model1,correlation=TRUE)
VarCorr(mixed_eth_model1)

#eth_model2
summary(mixed_eth_model2)
VarCorr(mixed_eth_model2)

#################################### Model #2

#model 2 - Compound Symmetry
eth_cs_model2 <- gls(Mean.Scale.Score~Ethnicity+Borough+Grade+Year_,
                     correlation = corCompSymm(form= ~1|Group),
                     data=eth_model)

#model 2 results
summary(eth_cs_model2)

#correlation matrix
VarCov <- getVarCov(eth_cs_model2, individual="Asian_Bronx_2013-2015")
Corr<-VarCov/VarCov[1,1]
VarCov
Corr
```

```
##################################### Model #3

#model 3 - Autoregressive model
eth_ar_model2 <- gls(Mean.Scale.Score~Ethnicity+Borough+Grade+Year_, data=eth_model,
                   correlation = corAR1(form= ~1|Group))

#model 3 results
summary(eth_ar_model2)

#correlation matrix
VarCov <- getVarCov(eth_ar_model2, individual="Asian_Bronx_2013-2015")
Corr<-VarCov/VarCov[1,1]
VarCov
Corr
##################################### GEE model

#gee
eth_mgee <- gee(Mean.Scale.Score~Ethnicity+Borough+Grade+Year_, data=eth_model, id=Group, corstr = "exchangeable")

#gee results
summary(eth_mgee)

#gee
eth_mgee2 <- geese(Mean.Scale.Score~Ethnicity+Borough+Grade+Year_, data=eth_model, id=Group, corstr="ar1")

#gee results
summary(eth_mgee2)

######################## Gender Model

gender <- read.csv("Gender_Borough.csv")

#convert into factors
gender <- gender %>%
  mutate(
    Borough = as.factor(Borough),
    Gender = as.factor(Gender),
    Grade = as.factor(Grade),
    Year = as.factor(Year))
gend_model <- gender %>%
  dplyr::select(Borough, Grade, Year, Gender, Group, Mean.Scale.Score)

#model 1 - Mixed Model

#With Year
mixed_gend_model2 <- lmerTest::lmer(Mean.Scale.Score~Borough+Gender+Grade+Year+(1|Group), data=gend_model, REML=FALSE)


#model 1 results
summary(mixed_gend_model2)

##################################### Model #2

#model 2 - Compound Symmetry
gend_cs_model <- gls(Mean.Scale.Score~Gender+Borough+Grade+Year, data=gend_model,
               corr = corCompSymm(0, form= ~1|Group))
```

```
#model 2 results
summary(gend_cs_model)

##################################### Model #3

#model 3 - Autoregressive model
gend_ar_model <- gls(Mean.Scale.Score~Gender+Borough+Grade+Year, data=gend_model,
        corr = corAR1(0, form= ~1|Group))

#model 3 results
summary(gend_ar_model)

##################################### GEE model

#gee
gend_mgee <- geeglm(Mean.Scale.Score~Gender+Borough+Grade+Year, data=gend_model, id=Group, corstr="ar1")

summary(gend_mgee)



######################## ELL Model

ell <- read.csv("ELL_Borough.csv")

#convert into factors
ell <- ell %>%
  mutate(
    Borough = as.factor(Borough),
    ELL = as.factor(ELL),
    Grade = as.factor(Grade),
    Year=as.factor(Year))
ell_model <- ell %>%
  dplyr::select(Borough, Grade, Year, ELL, Group, Mean.Scale.Score)

#model 1 - Mixed Model

#With Year
mixed_ell_model2 <- lmerTest::lmer(Mean.Scale.Score~Borough+ELL+Grade+Year+(1|Group), data=ell_model, REML=FALSE)


#model 1 results
summary(mixed_ell_model2)

##################################### Model #2

#model 2 - Compound Symmetry
ell_cs_model <- gls(Mean.Scale.Score~ELL+Borough+Grade+Year, data=ell_model,
          corr = corCompSymm(0, form= ~1|Group))

#model 2 results
summary(ell_cs_model)

##################################### Model #3
```

```
#model 3 - Autoregressive model
ell_ar_model <- gls(Mean.Scale.Score~ELL+Borough+Grade+Year, data=ell_model,
                    corr = corAR1(0, form= ~1|Group))

#model 3 results
summary(ell_ar_model)

##################################### GEE model

#gee
ell_mgee <- geeglm(Mean.Scale.Score~ELL+Borough+Grade+Year, data=ell_model, id=Group, corstr="ar1")

summary(ell_mgee)

######################## SWD Model

swd <- read.csv("SWD_Borough.csv")

#convert into factors
swd <- swd %>%
  mutate(
    Borough = as.factor(Borough),
    SWD = as.factor(SWD),
    Grade = as.factor(Grade),
    Year = as.factor(Year))
swd_model <- swd %>%
  dplyr::select(Borough, Grade, Year, SWD, Group, Mean.Scale.Score)

#model 1 - Mixed Model

#With Year
mixed_swd_model2 <- lmerTest::lmer(Mean.Scale.Score~Borough+SWD+Grade+Year+(1|Group), data=swd_model, REML=FALSE)


#model 1 results
summary(mixed_swd_model2)

##################################### Model #2

#model 2 - Compound Symmetry
swd_cs_model <- gls(Mean.Scale.Score~SWD+Borough+Grade+Year, data=swd_model,
                    corr = corCompSymm(0, form= ~1|Group))

#model 2 results
summary(swd_cs_model)

##################################### Model #3

#model 3 - Autoregressive model
swd_ar_model <- gls(Mean.Scale.Score~SWD+Borough+Grade+Year, data=swd_model,
                    corr = corAR1(0, form= ~1|Group))

#model 3 results
summary(swd_ar_model)

##################################### GEE model
```

```
#gee
swd_mgee <- geeglm(Mean.Scale.Score~SWD+Borough+Grade+Year, data=swd_model, id=Group, corstr="ar1")

summary(swd_mgee)

####################### ECON Model

econ <- read.csv("Econ_Borough.csv")

#convert into factors
econ <- econ %>%
  mutate(
    Borough = as.factor(Borough),
    ECON = as.factor(ECON),
    Grade = as.factor(Grade),
    Year = as.factor(Year))
econ_model <- econ %>%
  dplyr::select(Borough, Grade, Year, ECON, Group, Mean.Scale.Score)

#model 1 - Mixed Model

#With Year
mixed_econ_model2 <- lmerTest::lmer(Mean.Scale.Score~Borough+ECON+Grade+Year+(1|Group), data=econ_model, REML=FALSE)


#model 1 results
summary(mixed_econ_model2)

##################################### Model #2

#model 2 - Compound Symmetry
econ_cs_model <- gls(Mean.Scale.Score~ECON+Borough+Grade+Year, data=econ_model,
                corr = corCompSymm(0, form= ~1|Group))

#model 2 results
summary(econ_cs_model)

##################################### Model #3

#model 3 - Autoregressive model
econ_ar_model <- gls(Mean.Scale.Score~ECON+Borough+Grade+Year, data=econ_model,
                corr = corAR1(0, form= ~1|Group))

#model 3 results
summary(econ_ar_model)

##################################### GEE model

#gee
econ_mgee <- geeglm(Mean.Scale.Score~ECON+Borough+Grade+Year, data=econ_model, id=Group, corstr="exchangeable")

summary(econ_mgee)
```