Data Visualization & Analysis on Statewide Standardized Test Results:

Modeling Average Scaled Scores Based on Four Major Demographics for the Massachusetts Comprehensive Assessment System (2022)

Timothy Hodgdon, Nolan Karsok, Broderick Kelly,
Adam Nelson, Shreyas Risbud, Souri Sasanfar

MATH 7343 – Applied Statistics

Professor Aidong Ding

Northeastern University

Monday, April 17th, 2023

Abstract:

This paper explores the applications of data visualization and analysis to the standardized test results achieved by elementary and middle school students across the state of Massachusetts as part of the 2022 Massachusetts Comprehensive Assessment System (MCAS). All throughout this investigation, the average scaled scores attained by every public school district in Massachusetts for the 2022 MCAS ELA and 2022 MCAS MATH examinations will be assessed against one another within four main demographic categories. Namely, these demographics are disability status, family income, gender, and race or ethnicity. The guiding purpose of evaluating the average scaled scores realized by all Massachusetts public school districts for the 2022 MCAS ELA and MATH tests on the basis of demographic is to determine if there exist any statistically significant differences in average scaled score between the individual groups in each category. First, the contextual information regarding the current condition of public education in Massachusetts, along with the functionality of MCAS and its yearly data collection, is reviewed. Next, the statistical computing executed in the R programming language on the initial data sets containing average scaled scores by demographic subgroup is discussed. In this section, the base programs that are employed to create side-by-side box plots and histograms, generate confidence intervals, and run comparative hypothesis tests for average scaled score according to demographic are presented. The outputs of the visualizations and hypothesis testing completed for every demographic for both the 2022 MCAS ELA and MATH exams are then displayed, and how these outcomes infer population average scaled scores is also deliberated upon. Finally, based on these results, it is reasoned whether or not there actually exist any statistically significant differences in average scaled score within each demographic.

Keywords: Average Scaled Score • Base R Commands • Bonferroni Pairwise Comparison Tests • Box Plot • Disability Status • Family Income • Gender • Histogram • Level of Significance • Massachusetts Comprehensive Assessment System (MCAS) • Massachusetts Department of Elementary and Secondary Education • Next Generation MCAS • 95% Confidence Interval • One-Way Analysis of Variance (ANOVA) • p-value • Race & Ethnicity • 2022 MCAS ELA (English Language Arts) • 2022 MCAS MATH (Mathematics) • Two-Sided t-Test

Background:

According to nationwide scholastic ranking data accumulated in early 2023, Massachusetts is currently rated as the fifth state in terms of overall educational quality, and first for student success (Scholaroo et el., 2023). Largely contributing to these leading positions are the public school systems that are present across the state of Massachusetts. Although nearly 12% of all precollegiate students in the state are educated in private schools, which is greater than the national mean of approximately 10% (Private School Review et al., 2023), Massachusetts is still home to 1,751 publicly accessible schools that are accommodated by 316 districts (Ballotpedia et al., 2022), all of which reside under a singular, overarching body within the state government.

The Massachusetts Department of Elementary and Secondary Education is a statewide organization that is responsible for overseeing the academic practices and curricular standards for all of the public and charter school systems that exist in the Commonwealth of

Massachusetts. While these guidelines are set for a range of fields and subject matters, the two that the department considers the most vital to monitor the performance of among elementary and middle school students, specifically from Grades 3 through 8, are English Language Arts (ELA) and Mathematics (MATH). This is done every year by way of the Massachusetts Comprehensive Assessment System (MCAS), which are state-proctored standardized tests in English Language Arts and Mathematics that are issued to the students within these particular grade levels.

In 2017, as part of an ongoing initiative to update and modernize standardized testing, the Massachusetts Department of Elementary and Secondary Education established the Next Generation MCAS, which is administered entirely in a digital format rather than with written assessments, and is aimed at providing students with a more practical approach to critical thinking and problem solving (Malden Public Schools et al., 2017). After the testing period, individual students receive their scores and get the opportunity to see if they are exceeding, meeting, partially meeting, or not meeting expectations in relation to the English Language Arts and Mathematics benchmarks set by the state for their respective age and grade. The table below charts the scoring criteria for these achievement levels:

Table 1: Next Generation MCAS achievement levels based on scoring range (Massachusetts Department of Elementary and Secondary Education)

Achievement Level	Scoring Range
Exceeding Expectations	530 – 560
Meeting Expectations	500 – 530
Partially Meeting Expectations	470 – 500
Not Meeting Expectations	440 – 470

On the state level, the department uses these scores to identify the best-performing school districts, and gauge the needs for potential funding initiatives and education reforms for the low-scoring districts. Furthermore, as every school district is expected to publish their test results, this allows taxpayers to know whether their mandated investments are being properly utilized toward their district's public education.

Over the last six years in which the Next Generation MCAS has been given to students between Grades 3 and 8, the Massachusetts Department of Elementary and Secondary Education has compiled and maintained an extensive repository of scoring data for every public and charter school district that took part in standardized testing. Moreover, this department has anonymously gathered key demographic information that is also used to weigh these results against. For example, some of these demographic categories include disability status, family income, gender, and race or ethnicity. Disability status, family income, and gender are each divided into two distinct groups, namely students with or without disabilities, low income or non-low income households, and male or female students, respectively. On the other hand, race or ethnicity comprises more than two groups, those being African American or Black, American Indian or Alaskan Native, Asian, Hispanic or Latino, Native Hawaiian or Pacific Islander, and White students, along with students from other races and nationalities.

Methods & Procedures:

As the primary entity that is in charge of upholding the academic standards of the public school systems in the state, aggregating the average scaled scores on the MCAS ELA and MATH tests based on these demographics has the potential to help the department know if their core curricula are in fact approachable by students from all types of backgrounds. Using the raw data sets which are made available by the Massachusetts Department of Elementary and Secondary Education, the validity of this statement can be determined. In particular, this paper will investigate the average scaled scores on the 2022 MCAS ELA and MATH tests by demographic in order to find out if there are indeed any statistically significant differences in these measurements between the groups that are in each category. Note that the R programming language will be applied for all of the analysis and visualization carried out in this study. In addition to displaying the base commands within this section, the actual R commands used will also be supplied as attachments to this document.

Let i denote one of the four demographics, among disability status, family income, gender, and race or ethnicity, that are being considered. Before being able to show how the averaged scaled scores on the 2022 MCAS ELA and MATH assessments are distributed according to demographic group, they must first be arranged into separate lists. With that, let $n \geq 2$ be the number of groups that make up a given demographic i, such that $n \in \mathbb{Z}$. The basic R commands for converting the average scaled scores on the 2022 MCAS ELA and MATH of a group j in demographic i to lists, where $1 \leq j \leq n$ and $j \in \mathbb{Z}$, are in Figure 1 below:

```
Group_j_2022_Results <- read.csv("Group j - 2022 Results.csv")
Group_j_2022_MCAS_ELA_Results <- subset(Group_j_2022_Results, Subject == "ELA")
Group_j <- rep("Group j", nrow(Group_j_2022_MCAS_ELA_Results))
Group_j_2022_MCAS_ELA_Results$Group_j <- Group_j
Group_j_2022_MCAS_ELA_Scores <- Group_j_2022_MCAS_ELA_Results["Avg. Scaled Score"]
Group_j_2022_MCAS_ELA_Scores_List <- as.numeric(unlist(as.list(Group_j_2022_MCAS_ELA_Scores)))

Group_j_2022_Results <- read.csv("Group j - 2022 Results.csv")
Group_j_2022_MCAS_MATH_Results <- subset(Group_j_2022_Results, Subject == "MATH")
Group_j <- rep("Group j", nrow(Group_j_2022_MCAS_MATH_Results))
Group_j_2022_MCAS_MATH_Results$Group_j <- Group_j
Group_j_2022_MCAS_MATH_Scores <- Group_j_2022_MCAS_MATH_Results["Avg. Scaled Score"]
Group_j_2022_MCAS_MATH_Scores_List <- as.numeric(unlist(as.list(Group_j_2022_MCAS_MATH_Scores)))
```

Figure 1: Base R commands for generating lists of average scaled scores by demographic group

With the average scaled scores for every group j in demographic i organized into distinct lists, they can now be visualized. More precisely, for every demographic i, the average scaled scores in the 2022 MCAS ELA and MATH for all n groups that are within that category can be presented in both side-by-side box plots and histograms. The source code for creating these box plots is in the following figure:

```
Average_Scaled_Score <- c(Group_1_2022_MCAS_ELA_Scores_List, ..., Group_n_2022_MCAS_ELA_Scores_List)
Demographic <- c(Group_1, ..., Group_n)
Box_Plot <- boxplot(Average_Scaled_Score ~ Demographic,
                    main = "2022 MCAS ELA - Average Scaled Score By Demographic:")
Box_Plot

Average_Scaled_Score <- c(Group_1_2022_MCAS_MATH_Scores_List, ..., Group_n_2022_MCAS_MATH_Scores_List)
Demographic <- c(Group_1, ..., Group_n)
Box_Plot <- boxplot(Average_Scaled_Score ~ Demographic,
                    main = "2022 MCAS MATH - Average Scaled Score By Demographic:")
Box_Plot
```

Figure 2: Base R commands for constructing box plots of average scaled score by demographic group

while those for producing the side-by-side histograms of average scaled score for the n groups in demographic i are in Figure 3:

```
hist(Group_1_2022_MCAS_ELA_Scores_List, main = "2022 MCAS ELA - Group 1:")
.
.
.
hist(Group_n_2022_MCAS_ELA_Scores_List, main = "2022 MCAS ELA - Group n:")

hist(Group_1_2022_MCAS_MATH_Scores_List, main = "2022 MCAS MATH - Group 1:")
.
.
.
hist(Group_n_2022_MCAS_MATH_Scores_List, main = "2022 MCAS MATH - Group n:")
```

Figure 3: Base R commands for constructing histograms of average scaled score by demographic group

The purpose of running these codes and generating box plots and histograms of average scaled score is to demonstrate that this quantity adheres to an approximately normal distribution for every group j in a demographic i. This will be effective in confirming the use of the t-distribution to build 95% confidence intervals of the population means in average scaled score on the MCAS ELA and MATH tests for all n groups in demographic i. Figure 4 contains the foundational commands that are employed to construct these 95% confidence intervals for every group j in a particular demographic i:

```
Mean_Score <- mean(Group_j_2022_MCAS_ELA_Scores_List)
Standard_Deviation_Score <- sd(Group_j_2022_MCAS_ELA_Scores_List)
alpha <- 0.05
n <- nrow(Group_j_2022_MCAS_ELA_Scores)
df <- n - 1
Confidence_Interval <- Mean_Score + (c(-1, 1) * qt(1 - (alpha / 2), df) * (Standard_Deviation_Score / sqrt(n)))
Confidence_Interval

Mean_Score <- mean(Group_j_2022_MCAS_MATH_Scores_List)
Standard_Deviation_Score <- sd(Group_j_2022_MCAS_MATH_Scores_List)
alpha <- 0.05
n <- nrow(Group_j_2022_MCAS_ELA_Scores)
df <- n - 1
Confidence_Interval <- Mean_Score + (c(-1, 1) * qt(1 - (alpha / 2), df) * (Standard_Deviation_Score / sqrt(n)))
Confidence_Interval
```

Figure 4: Base R commands for constructing 95% confidence intervals of average scaled score by demographic group

The same line of reasoning for visualizing average scaled score also goes for being able to conduct hypothesis tests on the population means of this value for every demographic category i. As mentioned before, disability status, family income, and gender are three demographics that each consist of exactly $n = 2$ groups. Moreover, the total numbers of average scaled scores recorded for both groups in these each of these three demographics are unequal, thereby making these groups independent and unpaired. Because of this, two-sided t-tests, which evaluate the differences in mean average scaled score on the MCAS ELA and MATH for both groups, are carried out at the $\alpha = 0.05$ level of significance through the following outline in R:

```
Group_1_Data <- data.frame(score = Group_1_2022_MCAS_ELA_Scores_List, group = "Group 1")
Group_2_Data <- data.frame(score = Group_2_2022_MCAS_ELA_Scores_List, group = "Group 2")
Demographic_Data <- rbind(Group_1_Data, Group_2_Data)
t.test(score ~ group, data = Demographic_Data)

Group_1_Data <- data.frame(score = Group_1_2022_MCAS_MATH_Scores_List, group = "Group 1")
Group_2_Data <- data.frame(score = Group_2_2022_MCAS_MATH_Scores_List, group = "Group 2")
Demographic_Data <- rbind(Group_1_Data, Group_2_Data)
t.test(score ~ group, data = Demographic_Data)
```

Figure 5: Base R commands for conducting two-sided t-tests of average scaled score by demographic group

Unlike these three sets of demographics, race or ethnicity is made up of more than just two individual groups. Specifically, African American or Black, American Indian or Alaska Native, Asian, Hispanic or Latino, Native Hawaiian or Pacific Islander, White, and other students account for $n = 7$ groups in this category. Due to this, one-way ANOVA tests at the $\alpha = 0.05$ level of significance for the differences in mean average scaled scores across race or ethnicity are executed for both the 2022 MCAS ELA and MATH exams. Along with this, comparison tests using the Bonferroni method are run to figure out the exact pairs of mean average scaled scores with statistically significant differences. Figure 6 below shows the base R commands for the one-way ANOVA and Bonferroni pairwise comparison tests:

```
Group <- c(Group_1, ..., Group_n)
Score <- c(Group_1_2022_MCAS_ELA_Scores_List, ..., Group_n_2022_MCAS_ELA_Scores_List)
Demographic_Data <- data.frame(Group, Score)
Demographic_Data$Group <- as.factor(Demographic_Data$Group)
One_Way_ANOVA_Test <- aov(Score ~ Group, data = Demographic_Data)
summary(One_Way_ANOVA_Test)
pairwise.t.test(Demographic_Data$Score, Demographic_Data$Group, p.adjustment.method = "bonferroni")

Group <- c(Group_1, ..., Group_n)
Score <- c(Group_1_2022_MCAS_MATH_Scores_List, ..., Group_n_2022_MCAS_MATH_Scores_List)
Demographic_Data <- data.frame(Group, Score)
Demographic_Data$Group <- as.factor(Demographic_Data$Group)
One_Way_ANOVA_Test <- aov(Score ~ Group, data = Demographic_Data)
summary(One_Way_ANOVA_Test)
pairwise.t.test(Demographic_Data$Score, Demographic_Data$Group, p.adjustment.method = "bonferroni")
```

Figure 6: Base R commands for conducting one-way ANOVA tests and Bonferroni pairwise comparison tests of average scaled score by demographic group

Disability Status:

To reiterate, the demographic of disability status is one that is split between two groups, those being students with disabilities and students with no disabilities. From the initial data sets of average scaled scores for the 2022 MCAS ELA and 2022 MCAS MATH tests, the Massachusetts Department of Elementary and Secondary Education has not defined the disabilities that are encompassed by the students in the former grouping. Nevertheless, these distinctions have been made based on student enrollment data in every public and charter school district in Massachusetts. The results of the 2022 MCAS ELA show that the sample means in average scaled score for students with and without disabilities are ≈ 477.20 and ≈ 500.12, respectively, while the respective sample means for students with disabilities and students without disabilities on the 2022 MCAS MATH are ≈ 477.19 and ≈ 498.57.

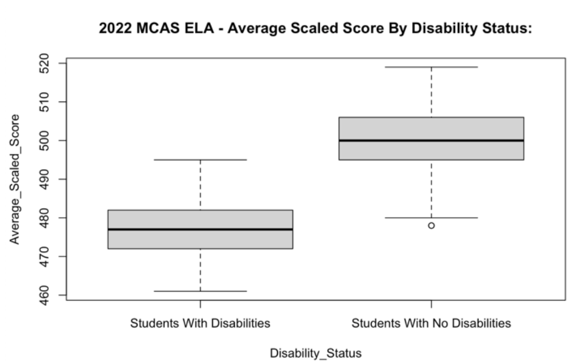

Figure 7: Box plots of average scaled score by disability status for the 2022 MCAS ELA

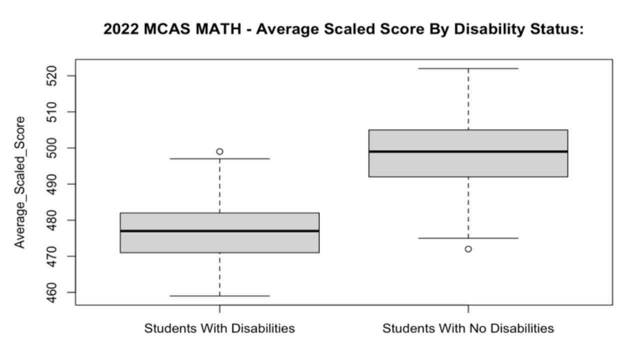

Figure 8: Box plots of average scaled score by disability status for the 2022 MCAS MATH

Figures 7 and 8 on the previous page display the boxplots of the average scaled scores for students with disabilities and students without disabilities on both the 2022 MCAS ELA and MATH. Just from these two sets of box plots, it is evident that the median average scaled scores for students with no disabilities are higher than those achieved by students with disabilities. In addition, these median scores appear to be similar to the corresponding sample means, and these box plots also seem to be symmetric, thus suggesting approximately normal distributions in average scaled score for students with and without disabilities in the MCAS ELA and MATH. As seen in the next two figures, presenting average scaled score by disability status for both the 2022 MCAS ELA and MATH exams in the format of histograms helps to corroborate this particular implication:

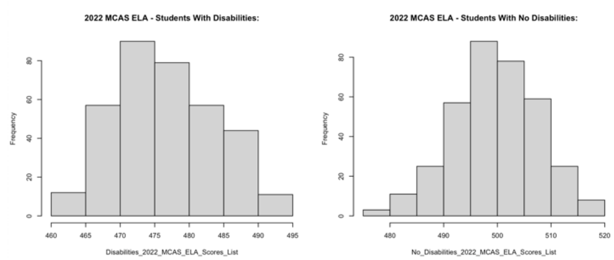

Figure 9: Histograms of average scaled score by disability status for the 2022 MCAS ELA

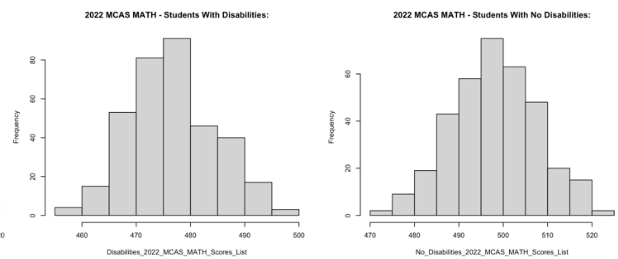

Figure 10: Histograms of average scaled score by disability status for the 2022 MCAS MATH

With the population distributions of average scaled score on the MCAS ELA and MATH approximated as normal for students with and without disabilities, the 95% confidence intervals of the mean average scaled scores on these tests for both these groups are calculated using the t-distribution as follows:

Table 2: 95% confidence intervals of mean average scaled score by disability status for the MCAS ELA and MCAS MATH

Disability Status	**95% Confidence Interval (MCAS ELA)**	**95% Confidence Interval (MCAS MATH)**
Students With Disabilities	≈ (476.43, 477.96)	≈ (476.36, 478.01)
Students With No Disabilities	≈ (499.30, 500.95)	≈ (497.57, 499.56)

While these 95% confidence intervals, along with the box plots generated earlier, reveal a clear difference in average scaled scores between students with disabilities and students with no disabilities, it now becomes necessary to establish if these differences are statistically significant. This is done with a two-sided t-test, in which the samples of average scaled score for students with and without disabilities are independent and unpaired for both assessments. At the $\alpha = 0.05$ level of significance, a two-sided t-test on the null hypothesis H_0 and alternative hypothesis H_1:

$\boldsymbol{H_0}$**:** The differences in mean average scaled score for students with and without disabilities are 0 for the MCAS ELA and MCAS MATH.

$\boldsymbol{H_1}$**:** The differences in mean average scaled score for students with and without disabilities are not 0 for the MCAS ELA and MCAS MATH.

yields a p-value that is less than $2.2 \cdot 10^{-16}$ for both the MCAS ELA and MCAS MATH. Since $\alpha = 0.05$, and $2.2 \cdot 10^{-16} \ll 0.05$, this undoubtedly means that p-value $< \alpha$ for these two assessments. As a result, the stated null hypothesis stands to be rejected, meaning that there exists sufficient evidence to indicate that the mean average scaled scores for students with disabilities and for students with no disabilities are significantly difference from one another with respect to both the MCAS ELA and MCAS MATH.

Family Income:

Family income is another demographic category observed by the Massachusetts Department of Elementary and Secondary Education that is made up of two different groups, namely students from low income households and students from non-low income households. As is the case with disability status, the department does not clearly delineate the salary ranges which constitute low income as opposed to non-low income households. Rather, these differences have also been made on the district level. After the 2022 testing period, the sample mean average scaled scores for the MCAS ELA among students from low income households is found to be ≈ 488.85, and ≈ 499.92 for students from non-low income households. For the 2022 MCAS MATH, the respective sample mean average scaled scores for students from low income households and students from non-low income households are ≈ 486.80 and ≈ 499.06.

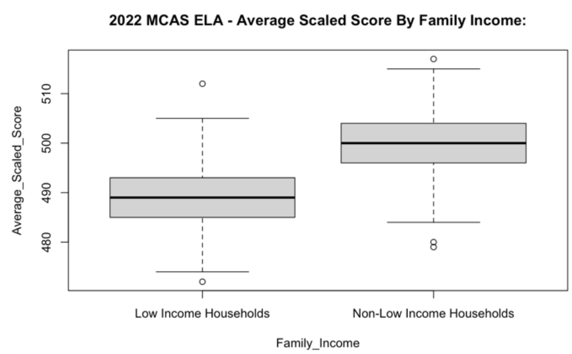

Figure 11: Box plots of average scaled score by family income for the 2022 MCAS ELA

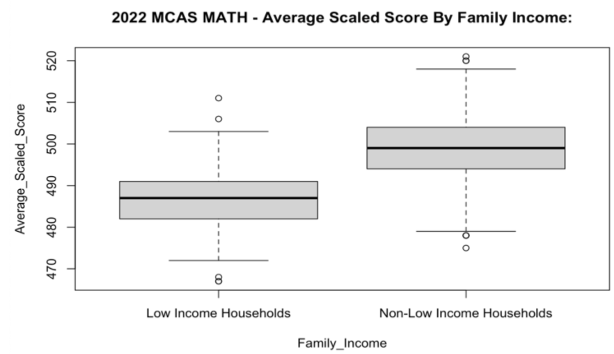

Figure 12: Box plots of average scaled score by family income for the 2022 MCAS MATH

The two sets of side-by-side box plots above show the distributions of averaged scaled scores for students from low income households and non-low income households. The pair of side-by-side histograms below also display how these values are distributed by family income for the 2022 MCAS ELA and MATH exams:

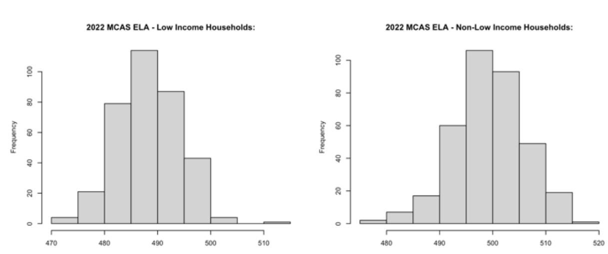

Figure 13: Histograms of average scaled score by family income for the 2022 MCAS ELA

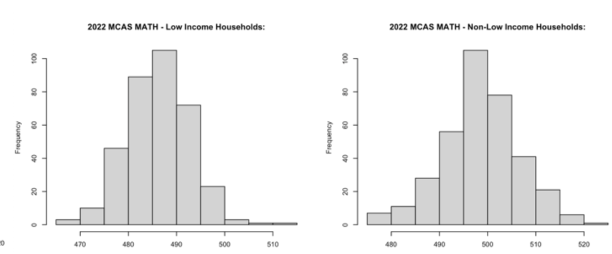

Figure 14: Histograms of average scaled score by family income for the 2022 MCAS MATH

From the box plots provided in Figures 11 and 12, there is a discrepancy in the median average scaled scores for students from low income households and students from non-low income households. Additionally, on the 2022 MCAS ELA, the median average scaled scores for students from low income households and for students from non-low income households look like they are close to their respective sample mean average scaled scores. This statement holds true for the 2022 MCAS MATH as well. The near symmetry illustrated in the box plots and the histograms in Figures 13 and 14 further demonstrates that average scaled scores on the MCAS ELA and MATH follow approximately normal population distributions for students from both low income and non-low income households. This means that the t-distribution can be utilized to construct the 95% confidence intervals of the mean average scaled scores for students from low and non-low income households on the MCAS ELA and MATH. These 95% confidence intervals are cataloged in Table 3:

Table 3: 95% confidence intervals of mean average scaled score by family income for the MCAS ELA and MCAS MATH

Family Income	95% Confidence Interval (MCAS ELA)	95% Confidence Interval (MCAS MATH)
Low Income Households	≈ (488.24, 489.46)	≈ (486.11, 487.49)
Non-Low Income Households	≈ (499.24, 500.61)	≈ (498.22, 499.89)

The sample mean and median average scaled scores for the 2022 MCAS ELA and 2022 MCAS MATH for students from non-low income households are noticeably higher than those for students from low income households, as seen in both the data visualizations and the 95% confidence intervals. With that, the subsequent null hypothesis H_0 and the alternative hypothesis H_1, which are defined as:

$\boldsymbol{H_0}$: The differences in mean average scaled score for students from low and non-low income households are 0 for the MCAS ELA and MCAS MATH.

$\boldsymbol{H_1}$: The differences in mean average scaled score for students from low and non-low income households are not 0 for the MCAS ELA and MCAS MATH.

are evaluated in an unpaired, independent two-sided t-test at the $\alpha = 0.05$ level of significance. For both the MCAS ELA and the MCAS MATH, the resulting p-value is less than $2.2 \cdot 10^{-16}$, which means that p-value $< \alpha$, as $2.2 \cdot 10^{-16} \ll 0.05$. Because of this, the null hypothesis that the differences in mean average scaled score for students from low and non-low income households are 0 for the MCAS ELA and the MCAS MATH must be rejected. Therefore, there is enough evidence which signifies that the differences in mean average scaled scores for students from low income households and for students from non-low income households on the MCAS ELA and MATH assessments are statistically significant at the $\alpha = 0.05$ level of significance.

Gender:

The 2022 MCAS ELA posts sample mean average scaled scores of ≈ 492.46 for male students and ≈ 498.78 for female students. Meanwhile the sample mean average scaled scores for male and female students on the 2022 MCAS MATH tests are ≈ 495.01 and ≈ 493.71, respectively. For the 2022 MCAS ELA, Figure 15 presents box plots of the average scaled scores recorded for male and female students, and Figure 16 does the same for the 2022 MCAS MATH. In these two sets of side-by-side box plots, the median average scaled scores for male and female students appear to be similar to their coincident sample means. There seems to be an observable distinction between the median average scaled scores in the 2022 MCAS ELA among male and female students, such that the median score for female students is higher than that for male students. Meanwhile, for the 2022 MCAS MATH, the median scores for male and female students look nearly equivalent to one another, with the male scores being only slightly bigger than the female scores:

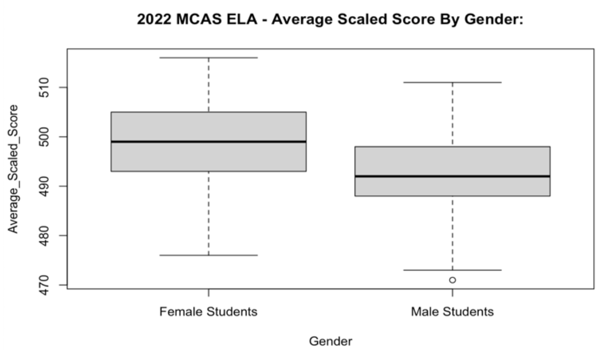

Figure 15: Box plots of average scaled score by gender for the 2022 MCAS ELA

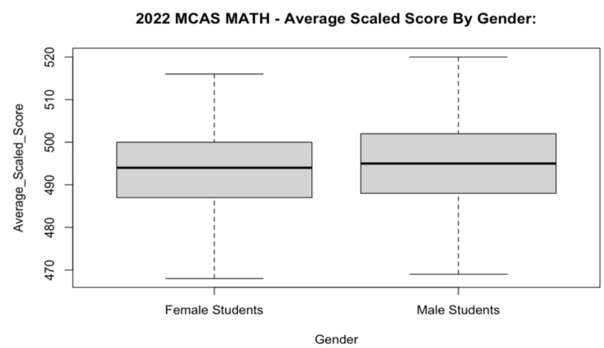

Figure 16: Box plots of average scaled score by gender for the 2022 MCAS MATH

Based on these sets of side-by-side box plots, there is an approximate symmetry to the distributions of average scaled scores for male and female students in the 2022 MCAS ELA and 2022 MCAS MATH tests, thereby implying that the population distributions of average scaled score on the MCAS ELA and MATH for male and female students are roughly normal. Putting the scoring data by gender for the 2022 MCAS ELA and MATH into the form of side-by-side histograms in the subsequent figures supports this supposition:

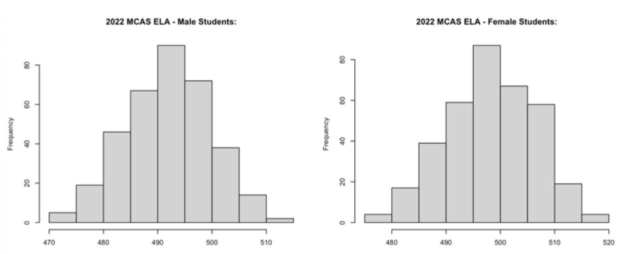

Figure 17: Histograms of average scaled score by gender for the 2022 MCAS ELA

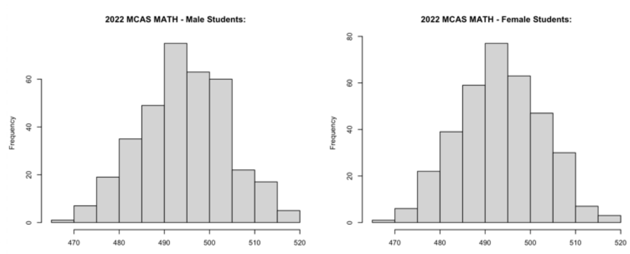

Figure 18: Histograms of average scaled score by gender for the 2022 MCAS MATH

Because the average scaled scores achieved by male students and female students on the MCAS ELA and MCAS MATH assessments are expected to follow approximately normal distributions, this means that the t-distribution can be used to compute the 95% confidence intervals of mean average scaled score by gender for both of these standardized tests. These 95% confidence intervals of mean average scaled score for the MCAS ELA and the MCAS MATH are tabulated on the next page:

Table 4: 95% confidence intervals of mean average scaled score by gender for the MCAS ELA and MCAS MATH

Gender	95% Confidence Interval (MCAS ELA)	95% Confidence Interval (MCAS MATH)
Male Students	≈ (491.64, 493.27)	≈ (494.00, 496.02)
Female Students	≈ (497.93, 499.62)	≈ (492.75, 494.67)

As previously described, the differences in the sample mean and median average scaled scores between male and female students on the 2022 MCAS ELA are discernible, while the differences in sample mean and median scores on the 2022 MCAS MATH are more subtle. This is reflected in the side-by-side box plots of average scaled score for male and female students for the 2022 MCAS ELA and MATH, along with the 95% confidence intervals for population mean scores within the two groups of this particular demographic. By running two-sided t-tests at the $\alpha = 0.05$ level of significance on the null H_0 and alternative H_1 hypotheses below:

$\boldsymbol{H_0}$**:** The differences in mean average scaled score for male and female students are 0 for the MCAS ELA and MCAS MATH.

$\boldsymbol{H_1}$**:** The differences in mean average scaled score for male and female students are not 0 for the MCAS ELA and MCAS MATH.

the resultant p-value for the MCAS ELA is less than $2.2 \cdot 10^{-16}$ and that for the MCAS MATH is 0.06683. Considering the MCAS ELA first, with a p-value that is less than $2.2 \cdot 10^{-16}$, it is clear that p-value $< \alpha$ due to $2.2 \cdot 10^{-16} \ll 0.05$. According to this result, the null hypothesis that there is no difference between the mean average scaled scores for male and female students on the MCAS ELA must be rejected. On the other hand, with p-value $= 0.06683$ for the MCAS MATH, because p-value $> \alpha$ as $0.06683 > 0.05$, then the null hypothesis that there is no difference between the mean average scaled scores for male and female students on the MCAS MATH fails to be rejected. As a result, these two conclusions substantiate the observations made for this demographic category. Specifically, there exists evidence to suggest that the difference in average scaled scores on the MCAS ELA standardized assessments between male and female students is statistically significant, while there does not exist enough evidence to suggest that the difference in scores between male and female students is statistically significant for the MCAS MATH examinations.

Race & Ethnicity:

Disability status, family income, and gender are demographics that are each been broken down into two separate groups by the Massachusetts Department of Elementary and Secondary Education. Different from these three is race or ethnicity, which is a category that the department has partitioned into seven main groups. Currently, the Massachusetts Department of Elementary and Secondary Education recognizes African American or Black, American Indian or Alaska Native, Asian, Hispanic or Latino, Native Hawaiian or Pacific Islander, and White students, as well as those from other backgrounds. For the 2022 MCAS ELA and 2022 MCAS MATH, the sample means in average scaled score by race or ethnicity are given in Table 5:

Table 5: Sample means of average scaled score by race or ethnicity for the 2022 MCAS ELA and 2022 MCAS MATH

Race & Ethnicity	Sample Mean (2022 MCAS ELA)	Sample Mean (2022 MCAS MATH)
African American or Black Students	≈ 488.86	≈ 485.40
American Indian or Alaska Native Students	≈ 487.12	≈ 486.88
Asian Students	≈ 504.20	≈ 508.08
Hispanic or Latino Students	≈ 489.28	≈ 487.24
Native Hawaiian or Pacific Islander Students	≈ 500.60	≈ 495.80
White Students	≈ 497.69	≈ 496.55
Other Students	≈ 497.65	≈ 496.31

The side-by-side box plots of average scaled scores by race or ethnicity on the 2022 MCAS ELA and the 2022 MCAS MATH are shown in Figures 19 and 20 below. From these distributions, it appears as though the median average scaled scores for African American or Black, Asian, Hispanic or Latino, White, and other students are roughly equal to their corresponding sample mean scores on both the 2022 MCAS MATH and ELA. Because of this, and the fact that the distributions for these five groups look fairly symmetric, it stands to reason that the population distributions of average scaled score on the MCAS ELA and the MCAS MATH for these five groups are approximately normal:

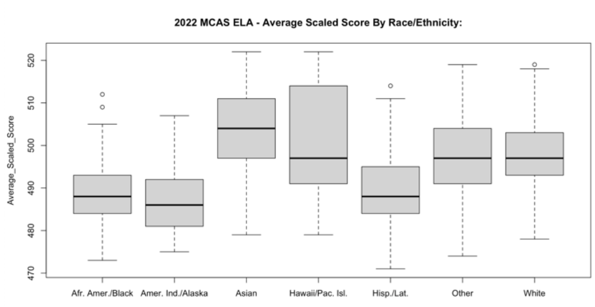

Figure 19: Box plots of average scaled score by race or ethnicity for the 2022 MCAS ELA

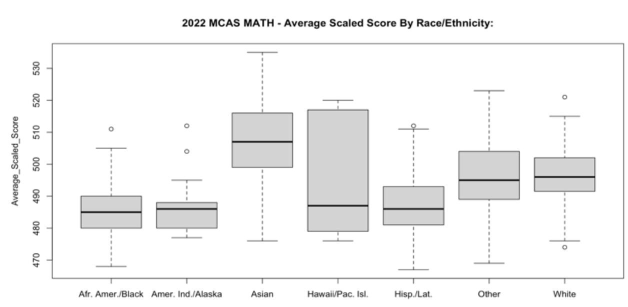

Figure 20: Box plots of average scaled score by race or ethnicity for the 2022 MCAS MATH

The sample mean average scaled scores for American Indian or Alaska Native students on the 2022 MCAS ELA and MATH are ≈ 487.12 and ≈ 486.88, respectively. While the median average scaled scores on these assessments for American Indian or Alaska Native students are close to the sample mean average scaled scores on the coincident tests, the distributions of average scaled score for this group do not seem to be symmetric. The same can evidently be said for students who are Native Hawaiians or Pacific Islanders. However, the sample means for

Native Hawaiians or Pacific Islanders are ≈ 500.60 for the 2022 MCAS ELA and ≈ 495.80 for the 2022 MCAS MATH, and the median scores denoted in the box plots for these students do not look like they are in the neighborhood of these values.

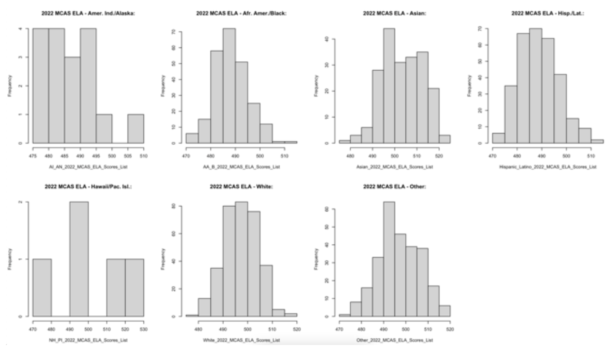

Figure 21: Histograms of average scaled score by gender for the 2022 MCAS ELA

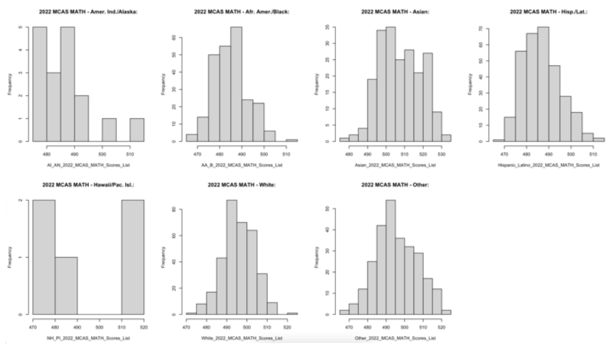

Figure 22: Histograms of average scaled score by gender for the 2022 MCAS MATH

The histograms above display relatively normal distributions for African American or Black, Asian, Hispanic or Latino, White, and other students. In contrast, the histograms for American Indians or Alaska Natives and for Native Hawaiians or Pacific Islanders, which are the leftmost histograms in the top and bottom rows of both figures, do not look normally distributed. Note that the sample sizes for these two groups are much smaller than the sample sizes for the other groups. If these sample sizes were akin to those of the other groups, then it is likely that their distributions of average scaled scores would be close to normal. Nevertheless, assuming that average scaled scores in the MCAS ELA and MATH are approximately normal for all seven of these racial and ethnic backgrounds, the t-distribution is applied to generate the 95% confidence intervals of mean average scaled score for all of these groups:

Table 6: 95% confidence intervals of mean average scaled score by race or ethnicity for the MCAS ELA and MCAS MATH

Race & Ethnicity	**95% Confidence Interval (MCAS ELA)**	**95% Confidence Interval (MCAS MATH)**
African American or Black Students	≈ (488.01, 489.72)	≈ (484.43, 486.36)
American Indian or Alaska Native Students	≈ (482.95, 491.29)	≈ (482.01, 491.76)
Asian Students	≈ (502.96, 505.44)	≈ (506.56, 509.60)
Hispanic or Latino Students	≈ (488.41, 490.16)	≈ (486.31, 488.16)
Native Hawaiian or Pacific Islander Students	≈ (479.01, 522.19)	≈ (469.56, 522.04)
White Students	≈ (496.95, 498.42)	≈ (495.71, 497.39)
Other Students	≈ (496.53, 498.77)	≈ (494.97, 497.64)

These 95% confidence intervals, along with the data visualizations in Figures 19 through 22, indicate that the mean average scaled scores among all groups within the race or ethnicity demographic are not all equal for the MCAS ELA and MATH. With a one-way analysis of variance (ANOVA) test where the single categorical variable is race or ethnicity, the statistical significance of these differences can be authenticated. Testing the following null hypothesis H_0 and alternative hypothesis H_1:

$\boldsymbol{H_0}$**:** The mean average scaled scores for students across all races or ethnicities are equal for the MCAS ELA and MCAS MATH.

$\boldsymbol{H_1}$**:** The mean average scaled scores for students across all races or ethnicities are not equal for the MCAS ELA and MCAS MATH.

yields a p-value that is less than $2 \cdot 10^{-16}$ for both exams. At the $\alpha = 0.05$ level of significance, it is unmistakable that p-value $< \alpha$ as $2 \cdot 10^{-16} \ll 0.05$. This means that the null hypothesis, stating that the mean average scaled scores across all races or ethnicities are equal for the MCAS ELA and MATH, must be rejected. As a result, there exists at least one pair of races or ethnicities with mean average scaled scores that are significantly different from one another. Due to the null hypothesis of a one-way ANOVA test being rejected, it now becomes necessary to determine the exact pairs of races or ethnicities with statistically significant differences in mean average scaled score. This is done through pairwise comparison tests under the Bonferroni method. In general, if there are n groups that make up a single categorical variable, then the level of significance α' for which every group pair must be tested at by the Bonferroni method is:

$$\alpha' = \frac{\alpha}{\binom{n}{2}} \tag{1}$$

Since there are $n = 7$ groups in the race or ethnicity demographic, then substituting this value and $\alpha = 0.05$ into the equality above returns $\alpha' \approx 0.0024$. The statistically significant results of the Bonferroni pairwise comparison tests for the MCAS ELA and MATH are in Tables 7 and 8:

Table 7: Statistically significant Bonferroni pairwise comparison test results by race or ethnicity for the MCAS ELA

Race & Ethnicity Pair	**p-value**	**Race & Ethnicity Pair**	**p-value**
African American or Black; Asian	$< 2 \cdot 10^{-16}$	Asian; Hispanic or Latino	$< 2 \cdot 10^{-16}$
African American or Black; White	$< 2 \cdot 10^{-16}$	Asian; White	$< 2 \cdot 10^{-16}$
African American or Black; Other	$< 2 \cdot 10^{-16}$	Asian; Other	$< 2 \cdot 10^{-16}$
American Indian or Alaska Native; Asian	$6.5 \cdot 10^{-16}$	Hispanic or Latino; White	$< 2 \cdot 10^{-16}$
American Indian or Alaska Native; White	$1.3 \cdot 10^{-6}$	Hispanic or Latino; Other	$< 2 \cdot 10^{-16}$
American Indian or Alaska Native; Other	$1.6 \cdot 10^{-6}$		

Table 8: Statistically significant Bonferroni pairwise comparison test results by race or ethnicity for the MCAS MATH

Race & Ethnicity Pair	p-value	Race & Ethnicity Pair	p-value
African American or Black; Asian	$< 2 \cdot 10^{-16}$	Asian; Hispanic or Latino	$< 2 \cdot 10^{-16}$
African American or Black; White	$< 2 \cdot 10^{-16}$	Asian; White	$< 2 \cdot 10^{-16}$
African American or Black; Other	$< 2 \cdot 10^{-16}$	Asian; Other	$< 2 \cdot 10^{-16}$
American Indian or Alaska Native; Asian	$< 2 \cdot 10^{-16}$	Hispanic or Latino; White	$< 2 \cdot 10^{-16}$
American Indian or Alaska Native; White	0.00032	Hispanic or Latino; Other	$< 2 \cdot 10^{-16}$
American Indian or Alaska Native; Other	0.00051		

Conclusion:

The preceding paper investigated the scoring data of standardized testing in English Language Arts (ELA) and Mathematics (MATH) for students from Grades 3 through 8 who participated in the 2022 Massachusetts Comprehensive Assessment System (MCAS). This scoring information was extracted from raw data sets of average scaled scores that have been made openly available by the Massachusetts Department of Elementary and Secondary Education. The department is accountable for overseeing the statewide educational guidelines and standards that are implemented in every public and charter school district in Massachusetts, and for ensuring that public education is approachable for students across different demographics. Using the R programming language, lists of average scaled scores on the 2022 MCAS ELA and the 2022 MCAS MATH were created according to four major demographics categories, namely disability status, family income, gender, and race or ethnicity. Side-by-side box plots and histograms of the average scaled scores recorded for every group in each category were then produced. Since these distributions were found to be approximately normal, the t-distribution was applied to construct 95% confidence intervals of mean average scaled score on the MCAS ELA and MATH for every demographic group being explored. Finally, hypothesis testing at the $\alpha = 0.05$ level of significance was completed for every demographic in order to assess if there are any statistically significant differences in average scaled scores between groups for these four demographics.

The visualizations and 95% confidence intervals for disability status clearly displayed higher average scaled scores for students without disabilities. For the 2022 MCAS ELA, the sample means calculated show that students with disabilities partially met expectations, and that students without disabilities were barely above the lower bound necessary to be meeting expectations. In addition, for the 2022 MCAS MATH, students with disabilities were also partially meeting expectations, but students with no disabilities were just under the lower bound to be meeting expectations. After analyzing the scores on the two exams based on disability status, the two-sided t-test on this data for students with and without disabilities revealed statistically significant differences in average scaled scores between these groups.

Students from low income households posted sample mean average scaled scores on the 2022 MCAS ELA and the 2022 MCAS MATH which were in the middle of the range needed to partially meet expectations. On the other hand, students from non-low income households on average just fell short of meeting expectations on these tests. These discrepancies were noticeable in the box plots, histograms, and 95% confidence intervals that were built for this demographic, and the two-sided t-tests confirmed statistically significant mean differences in average scaled score between students from low income and non-low income households.

For both standardized tests, male and female students achieved sample mean average scaled scores that were on the higher end of partially meeting expectations. However, the sample mean for female students on the 2022 MCAS ELA was discernibly higher than that for male students, while the sample mean for males on the 2022 MCAS MATH was only narrowly bigger than that for females. The scoring analysis and hypothesis tests conducted on the basis of gender verified these visual observations, as the difference in mean average scaled score between male and female students on the MCAS ELA was found to be statistically significant, while this difference was determined to have no statistical significance on the MCAS MATH.

Due to the race or ethnicity category containing more than two distinct groups, one-way ANOVA tests with this demographic as the single categorical variable were carried out in order to see if there is at least one pair of racial or ethnic groups with statistically significant mean differences in average scaled score on the MCAS ELA and MATH. Both of these one-way ANOVA tests outputted p-values much less than the $\alpha = 0.05$ level of significance, meaning that there is sufficient evidence to suggest that at least one pair of differences in mean average scaled score between racial or ethnic groups for the MCAS ELA and MATH is statistically significant. Bonferroni pairwise comparison tests were then conducted to find out the exact pairs of groups that are significantly different from one another. Tables 7 and 8 list the same 11 pairs of races or ethnicities that have statistically significant differences in mean average scaled score.

As mentioned before, the Massachusetts Department of Elementary and Secondary Education and the individual school districts share responsibility in offering evenly attainable education across demographics. Under ideal circumstances, the mean score for every group within each demographic would fall within meeting expectations or exceeding expectations. Since this is not always the case, it is at least important to make sure that every group in a given demographic is equally receptive to the same quality and standards of education. The hypothesis tests which returned statistically significant results could be indicators of potential biases or skews in how the groups in the corresponding demographics are instructed. Specifically, some groups may have advantages in terms finding material more approachable than other groups. For example, students with no physical or learning disabilities inherently have the advantage when it comes to education, and by extension standardized testing, as opposed to students who possess disabilities. While a bias like this does make sense and can therefore be difficult to eliminate, it is still undoubtedly essential for schools to be equal opportunist when it comes to family income, gender, and race or ethnicity.

Further Study & Improvements:

The scope of this analysis was restricted to the most recent MCAS score distributions from 2022. However, the Massachusetts Department of Elementary and Secondary Education has available data from every year since 2017. Therefore, similar analysis could be done over the past five years to see if there are any meaningful changes in trends.

Additionally, the mean difference between some categories could be compared across years. For example, the difference between the average score of students in low income families and non-low income families could be charted against an axis of years since 2017. Linear regression, or some other model, could be used to indicate if there is a trend in the difference over the past five years, and if the gap in scores is widening or decreasing. This analysis could be done on a number of different pairs examined in the initial study, such as disability status and gender. It is likely that the five years of data is not enough to demonstrate any meaningful trend. However, perhaps this could be done in the future when more data is available, or archived data from further in the past could be found. The MCAS was first given in 1993, and it is likely that this data exists somewhere, and it would be very interesting to see if initiatives to give support to marginalized students or other influences have had a positive impact.

Finally, while it may be difficult to standardize between states, it would be interesting to look at differences across state lines. Being native to Massachusetts, MCAS would not be able to provide this type of data. However, there are some federal standardized tests which could be examined. In particular, the PSATs and SATs could give interesting federal data to compare the U.S. by state or region. ANOVA testing could be used to see if there are particular areas which perform statistically better than others. Furthermore, particular subsets of the student population could be compared to evaluate how well their needs are being met. Perhaps there are outlier states which do a particularly good job supporting low income, disabled, minority or otherwise disadvantaged students. Perhaps there are some which do a particularly bad job. ANOVA testing on PSAT or SAT scores could reveal which states need to change their approaches, and which states can be used as models for others.

Acknowledgments:

We would like to thank Professor Aidong Ding for reviewing and approving the initial proposal submitted for this paper, and for providing feedback on the ideas put forth within it. In addition, we would like to acknowledge Professor Ding for his instruction throughout the semester, especially in regard to the visualization and analysis techniques utilized in this project.

Team C Members & Contributions:

The following list details the six members of Team C, as well as each individual's contributions to the overall project and its goals:

- Timothy Hodgdon
  – Original Data Sets; Abstract
- Nolan Karsok
  – Written Analysis; Compilation
- Broderick Kelly
  – Further Study & Improvements
- Adam Nelson
  – PowerPoint Presentation
- Shreyas Risbud
  – R Programming; Report Write-Up
- Souri Sasanfar
  – Methodology; Report Outline

References and Original Data Sets:

Ballotpedia. (2022). Public Education in Massachusetts. [Research Article]. Available at https://ballotpedia.org/Public_education_in_Massachusetts#:~:text=The%20Massachusetts%20public%20school%20system,schools%20in%20316%20school%20districts.

Malden Public Schools. (2017). Understanding the Next-Generation MCAS and 2017 Accountability Results. [Informational Article]. Available at https://maldenps.org/wp-content/uploads/2017/11/MCAS-FY17-Slides-School-Commt..pdf.

Massachusetts Department of Elementary and Secondary Education. (2017 – 2022). Next Generation MCAS Achievement Results. [Original Data Sets]. Available at https://profiles.doe.mass.edu/statereport/nextgenmcas.aspx.

Massachusetts Department of Elementary and Secondary Education. (2022). Testing Matters. [Informational Article]. Available at https://www.doe.mass.edu/mcas/testingmatters.html.

Private School Review. (2023). Best Massachusetts Private Schools. [Informational Article]. Available at https://www.privateschoolreview.com/massachusetts#:~:text=Best%20Massachusetts%20Private%20Schools%20(2023,national%20average%20of%2010%25).

Scholaroo. (2023). 2023 State Education Rankings – Best to Worst. [Research Article]. Available at https://scholaroo.com/report/state-education-rankings/.