TERM PROJECT: ANALYSIS OF EQAO TEST SCORES

PRESENTED BY:

ASHOK MISTRY, RYAN SUAREZ AND MITHIL PATEL

Dataset from Government of Ontario's Open Data Directive website at:
https://www.ontario.ca/data/school-information-and-student-demographics

The Open Data Directive gives access to government data by requiring all data to be made public, unless it is of a commercially-sensitive or private nature.

This is data on all publicly funded schools in Ontario, including school board, school address, GPS location, provincial test results, and student population demographics.

This data is shown as reported by various sources:
▸ Board School Identification Database (BSID) / Ontario School Information System (OnSIS)
   as updated/maintained by boards December 2017
▸ OnSIS 2015-2016 (Student Population)
▸ Education Quality and Accountability Office (EQAO) (2016-2017)
▸ Class-size Tracker, 2015-2016
▸ 2011 National Household Survey (NHS)
▸ 2015 Tax Filer (T1 Family File)

Objective:

Investigate some community-based factors that may have possible influence on the student's standardized EQAO test performance.

Null Hypothesis H_0 :
The number of Ontario school students who pass the EQAO standardized tests is independent of the parents income and education level, the enrolment size, and number of students whose first language is not English.

Alternate Hypothesis H_a :
The number of Ontario school students who pass the EQAO standardized tests is dependent on the parents income and education level, the enrolment size, and number of students whose first language is not English.

We will test this hypothesis with an alpha level of 0.05.

‣ Original data size (4934 rows x 53 columns)

‣ Selected columns (4934 rows x 10 columns)

```
['School Number', 'Board Name', 'Grade Range', 'Enrolment', 'NotEnglish', 'Reading', 'Writing', 'Maths',
'LowIncomeHome', 'EducatedParents']
```

‣ Independent/explanatory variables used in analysis:

```
['LowIncomeHome', 'EducatedParents', 'Enrolment', 'NotEnglish']
```

‣ Dependent/response variables are the percentage of students that passed the EQAO tests at each grade:

    ‣ Grade3 & Grade6: `['Reading', 'Writing', 'Maths']`

    ‣ Grade9: `['MathsAcad', 'MathsAppl']`

    ‣ Grade10: `['OSSLT']`

‣ Other explanatory variables not considered: `Percentage of students identified as gifted students, and students with special education needs.`

▸ To reduce bias in the data, <u>Stratified Sampling</u> was used to sample the data:

For example:

- ▸ Large dataset of 4934 public schools, with 83 school boards.

- ▸ Data grouped by Grade 3 schools = 3652 rows

- ▸ Data cleanup and removal of NaN values = 2185 rows with 65 school boards.

- ▸ Use random sampling to select 10 schools from each school board.

- ▸ If a school board has less than 10 schools, pick all the schools for this school board.

- ▸ Left with 463 schools used for regression analysis.

▸ Folium: MarkerClusters and HeatMaps
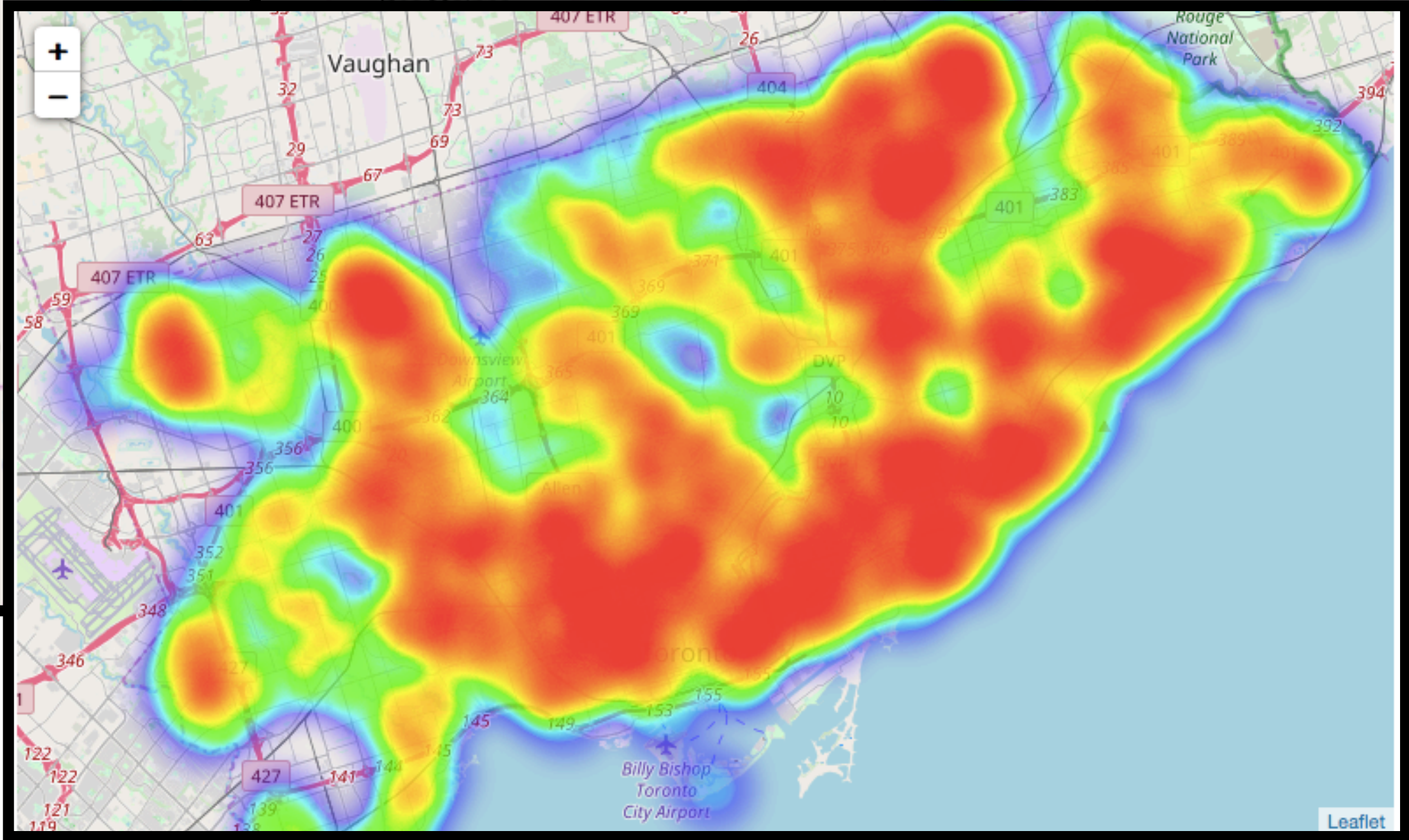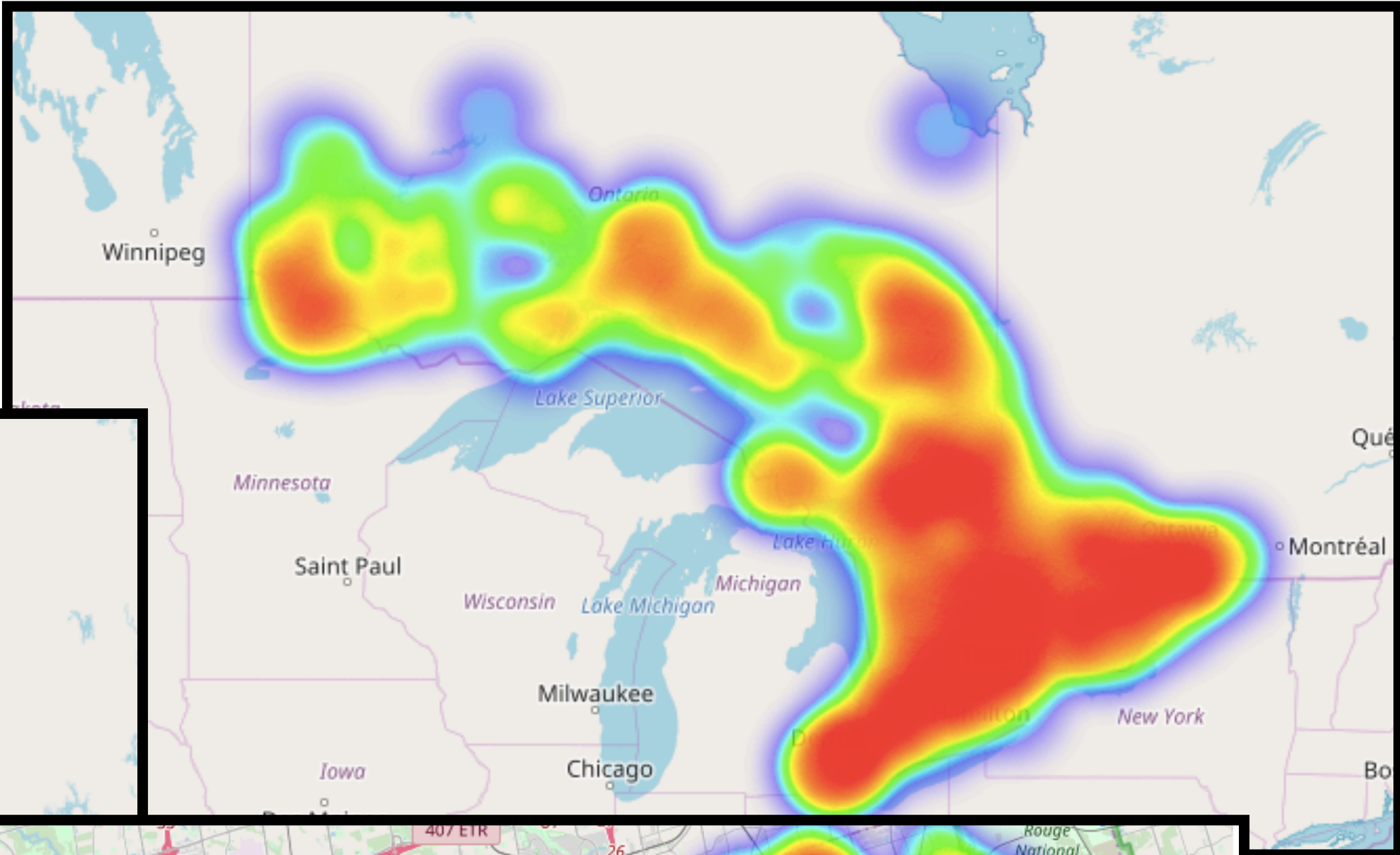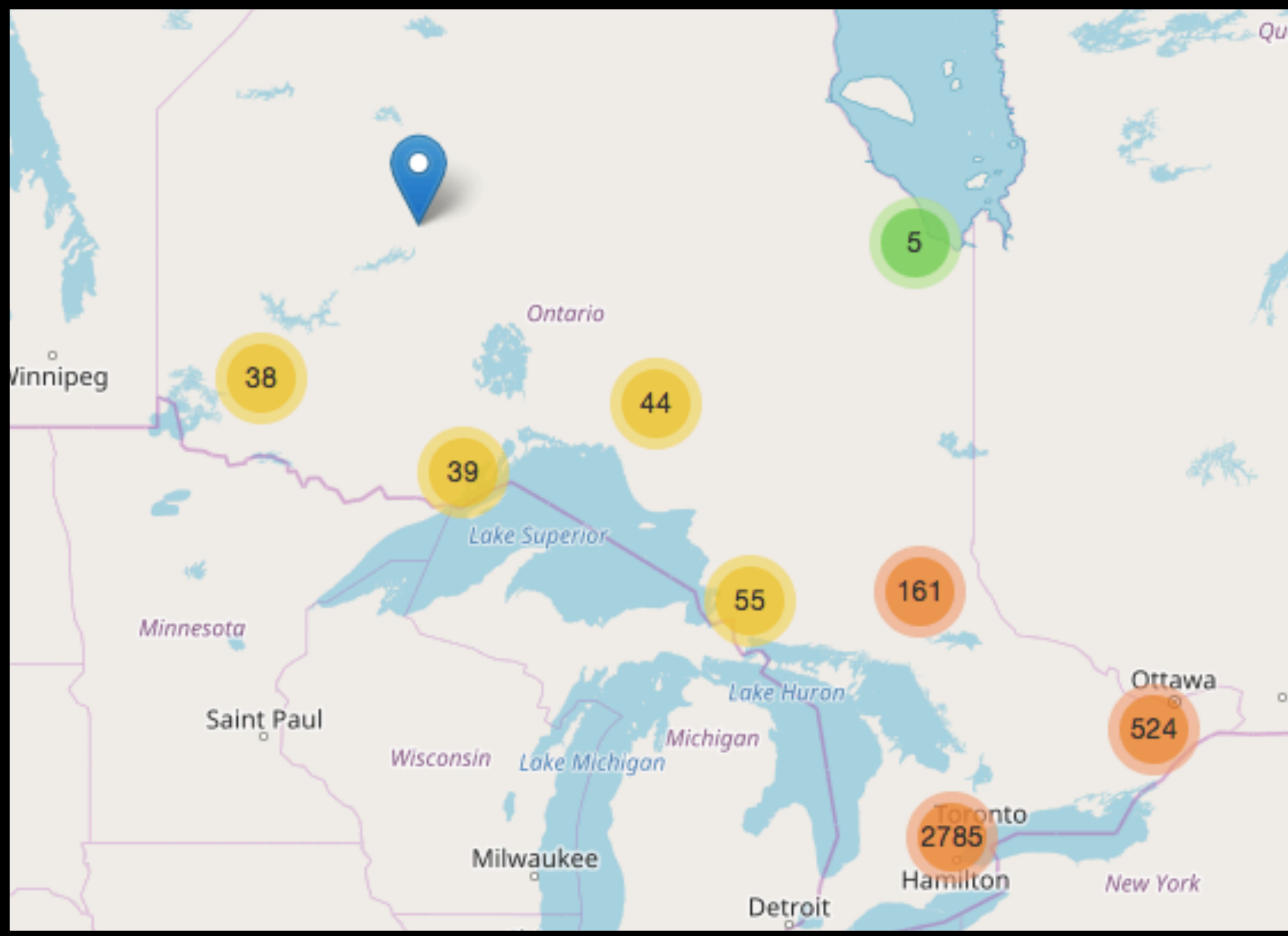
GRADE3: HISTOGRAMS OF NUMERIC VARIABLES
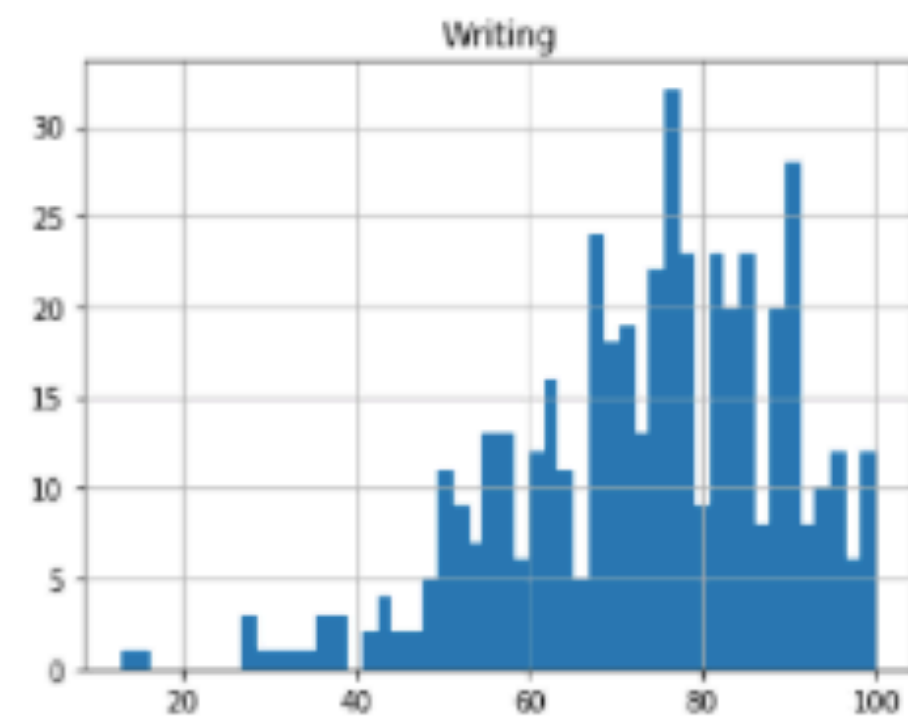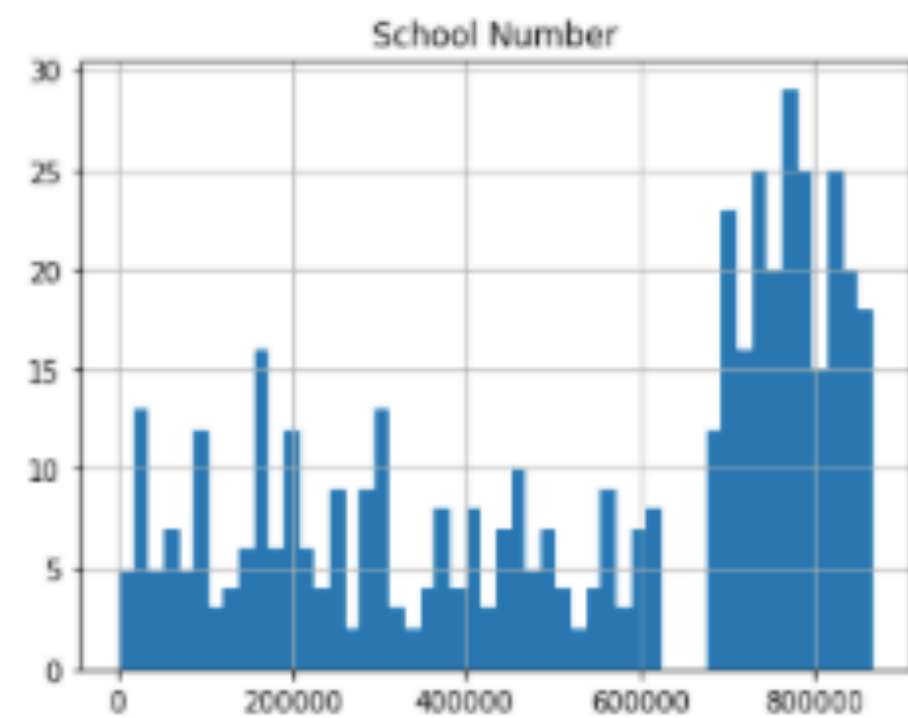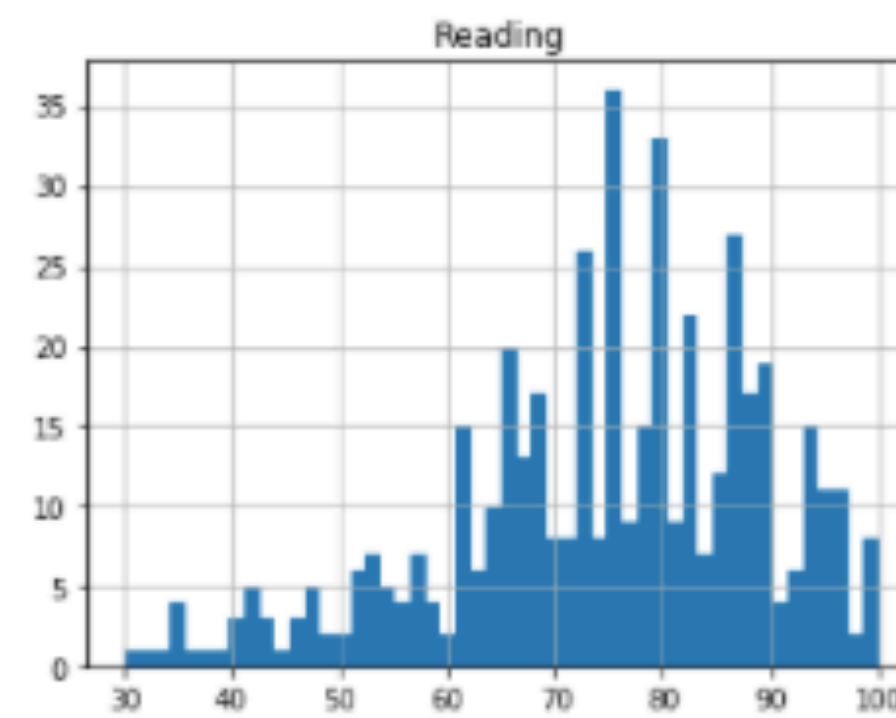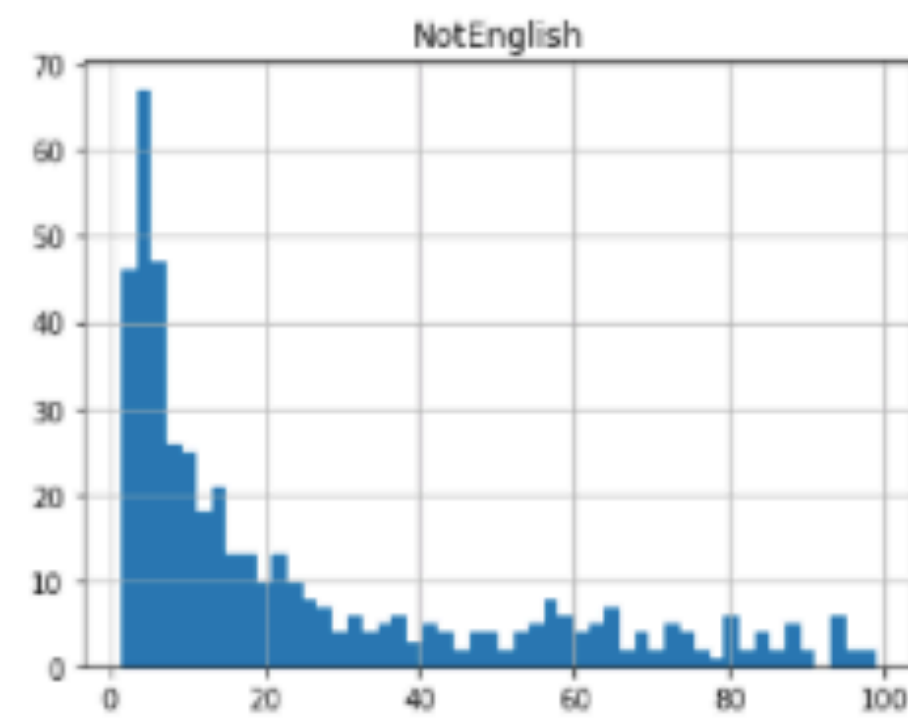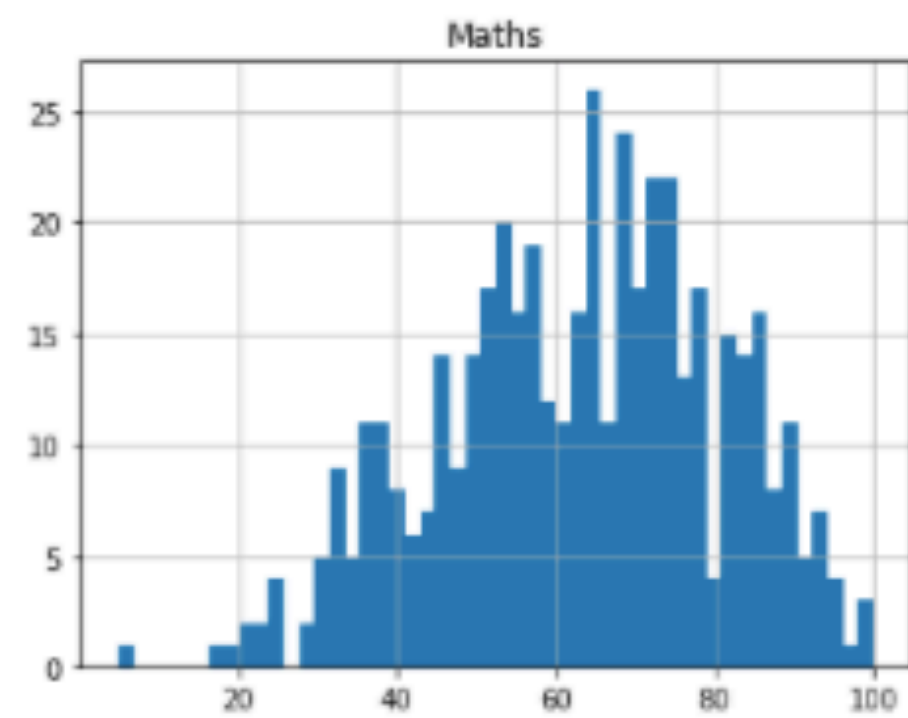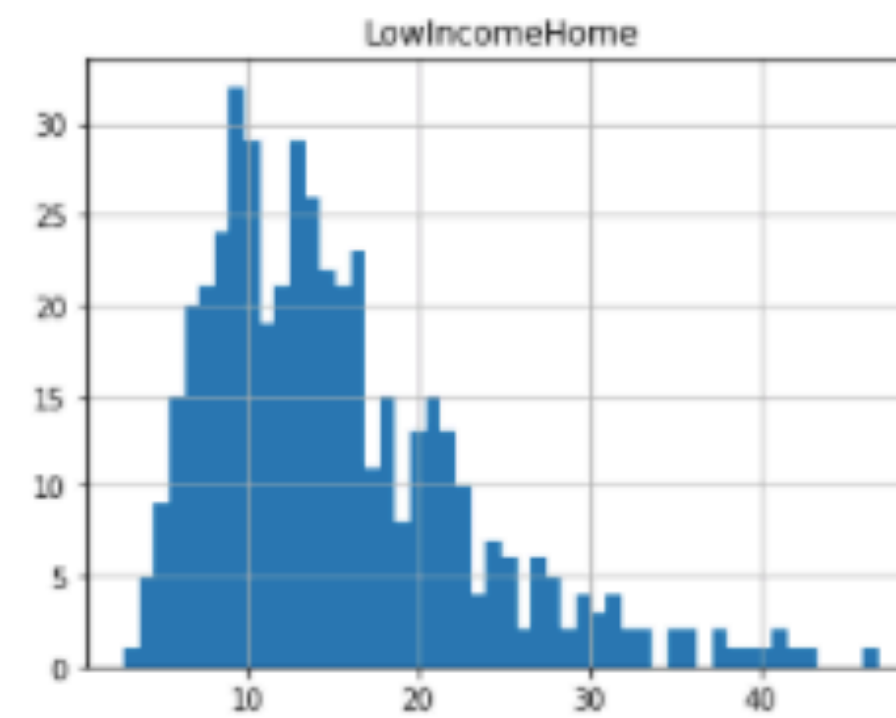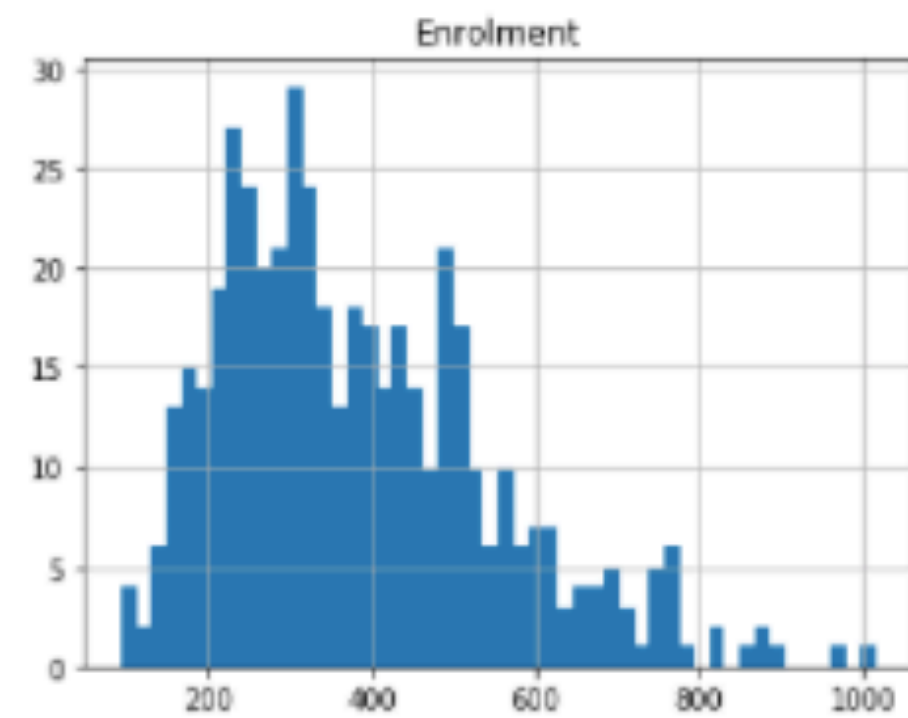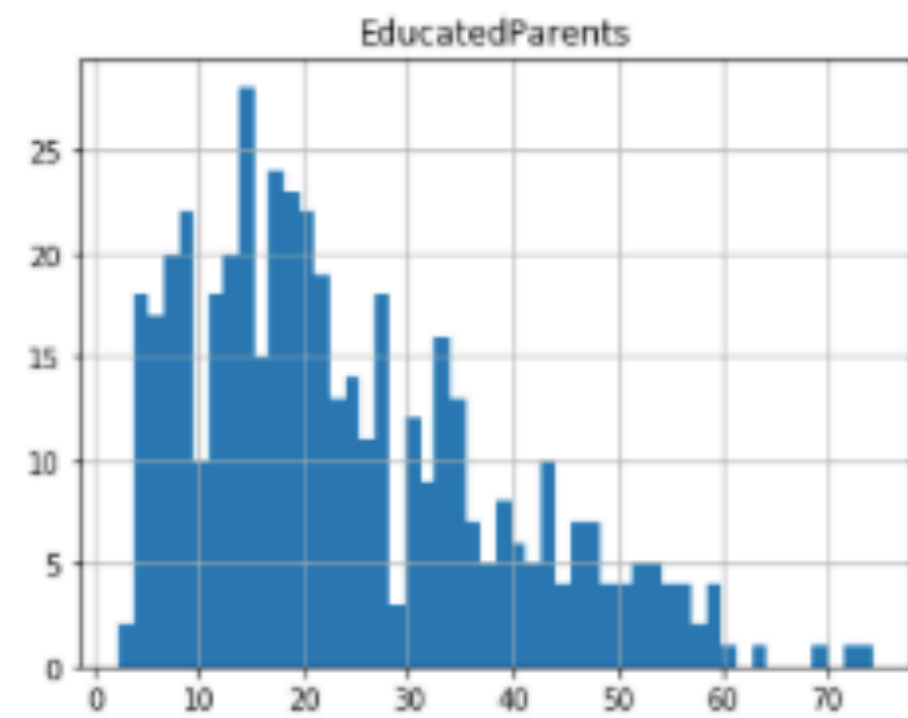
▸ Right skewed distribution for LowIncomeHome and NotEnglish

▸ Nearly normal distribution for EducatedParents

▸ Normal distributions for Reading, Writing and Maths

▸ Distributions for Reading and Writing follow very closely with each other.

DESCRIPTIVE STATISTICS

Percentages of Grade 3 EQAO Passes in Reading, Writing and Maths

Percentages of Grade 6 EQAO Passes in Reading, Writing and Maths

Percentages of Grade 9 EQAO Passes in Maths Academic and Applied

Percentage of Students That Passed the Grade 10 OSSLT on Their First Attempt
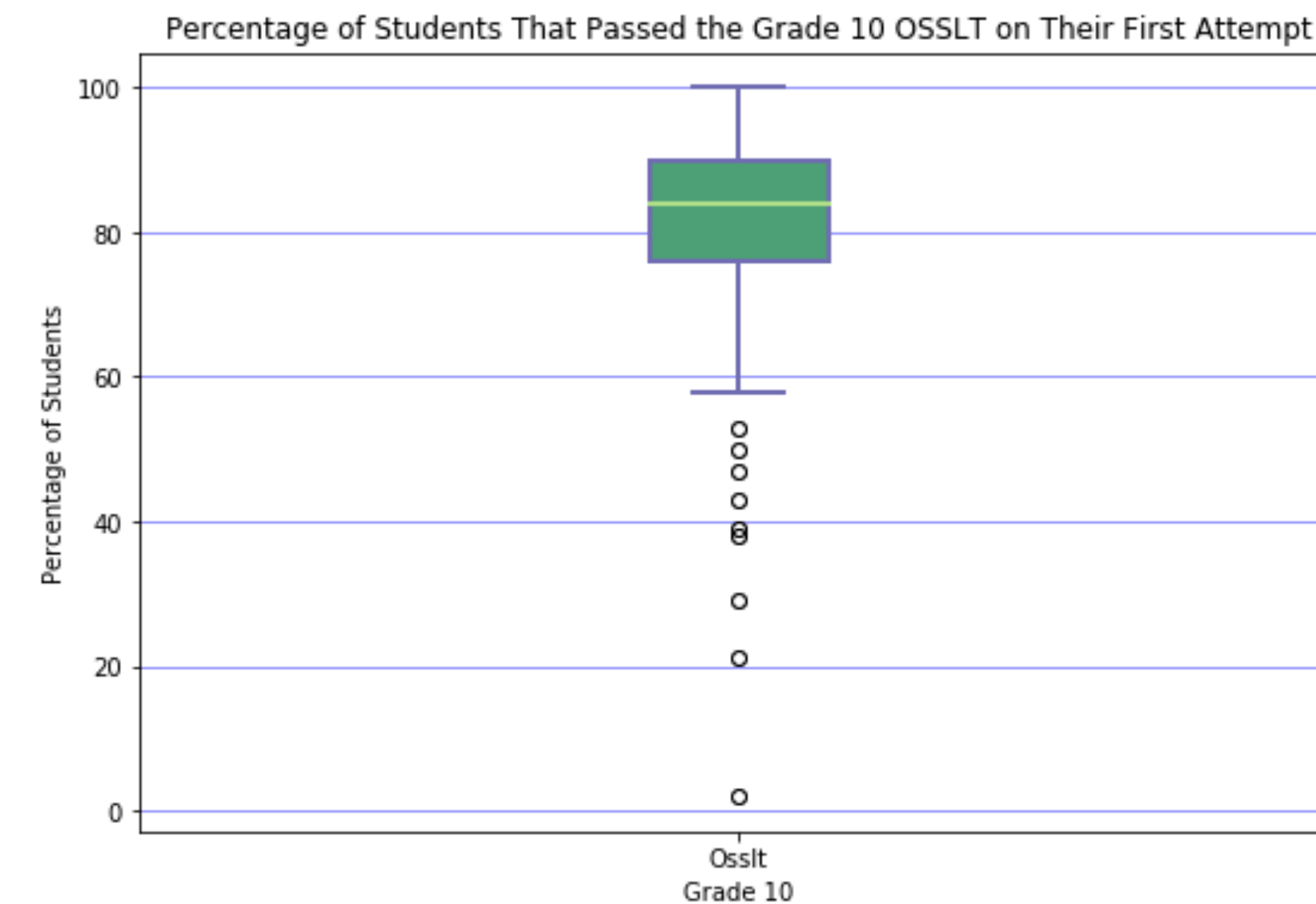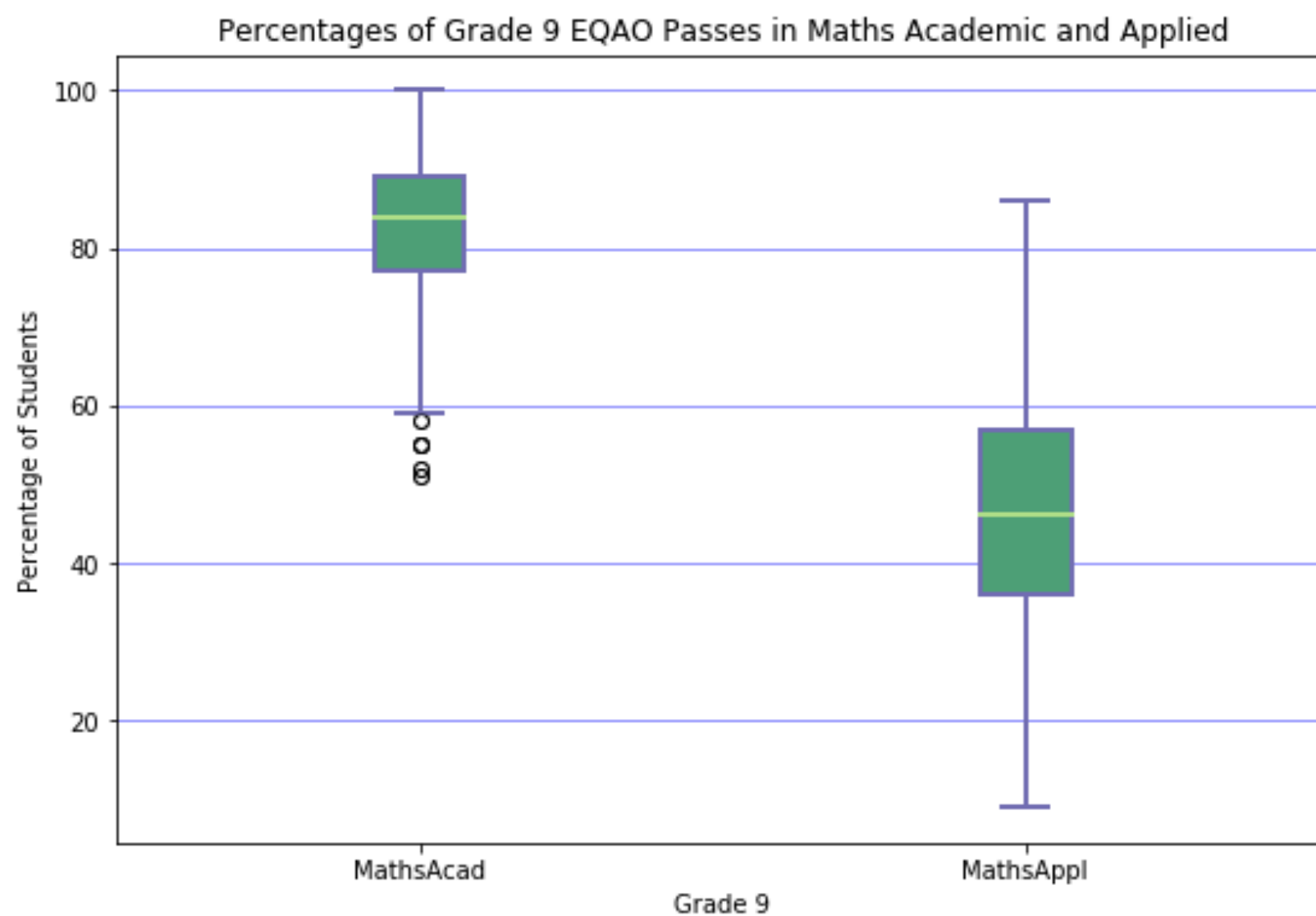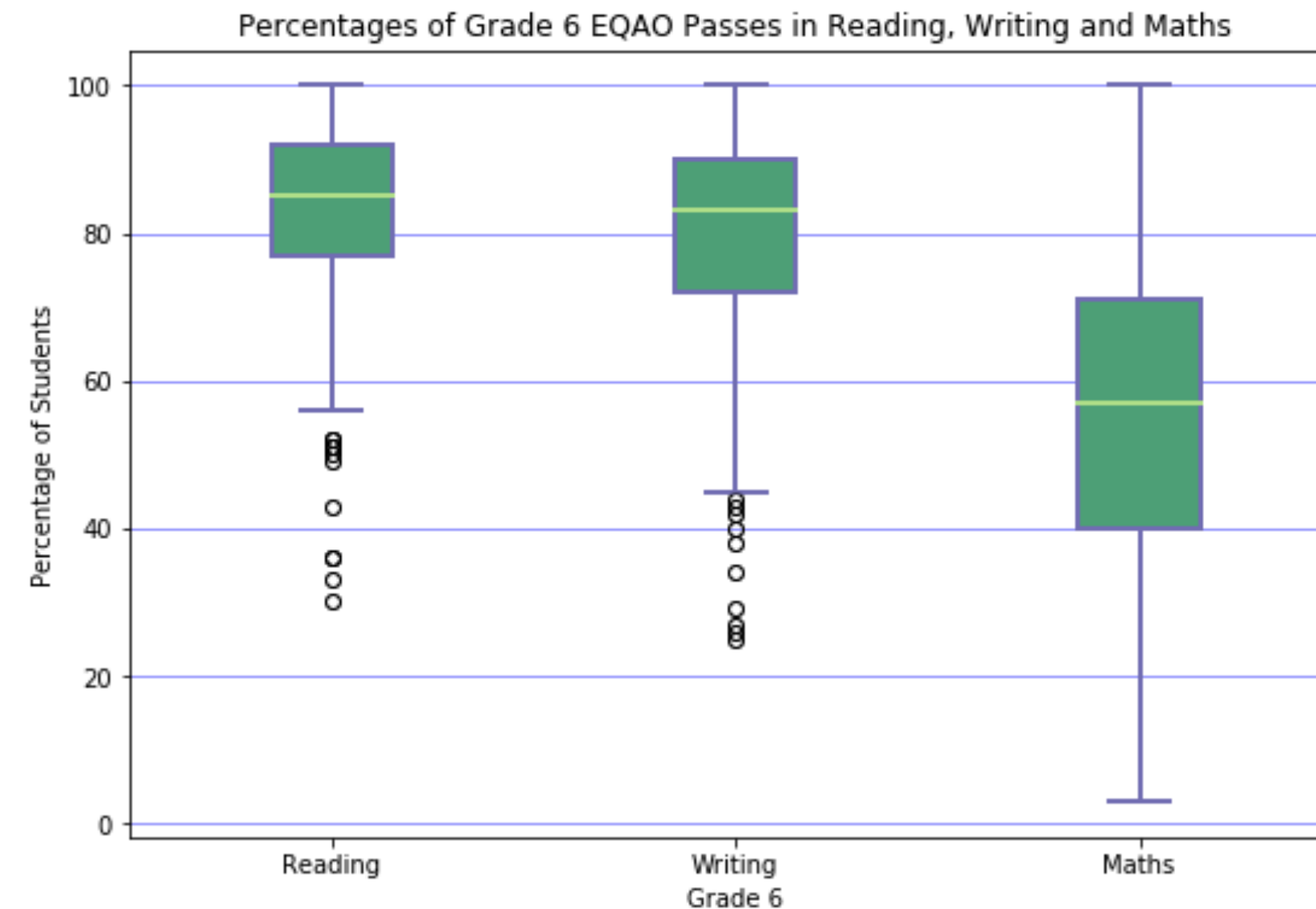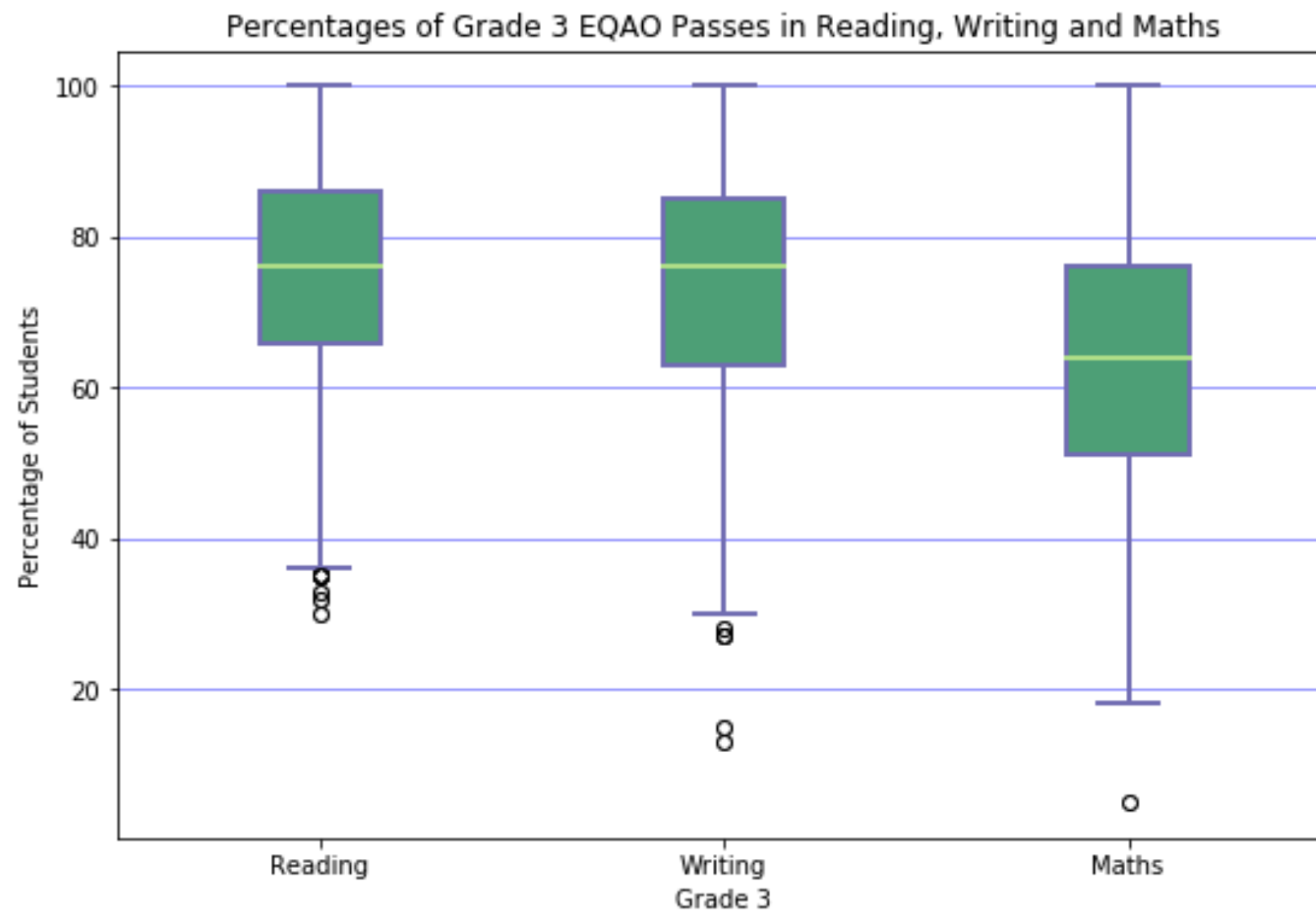
▶ 75th percentiles are similar for Grades 3 and 6 in Reading, Writing and Maths

▶ Maths achievements are lower in Grade 6

▶ Academic Maths shows a significant improvement in Grade 9

▶ Applied Maths has a mean score below 50%

▶ Grade 10 OSSLT pass-rate is high overall, with mean of 82%. There are some outliers in the lower range worth investigating
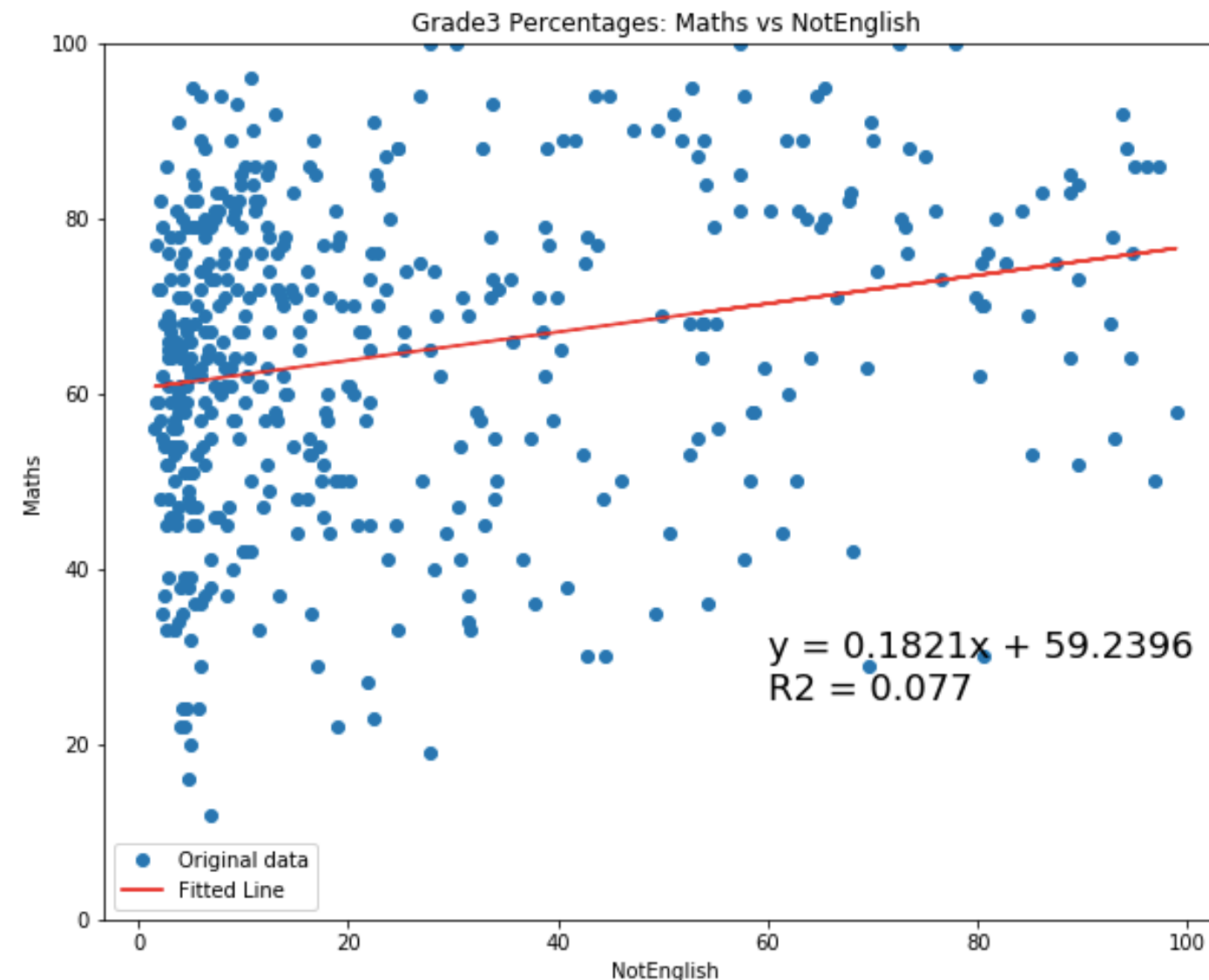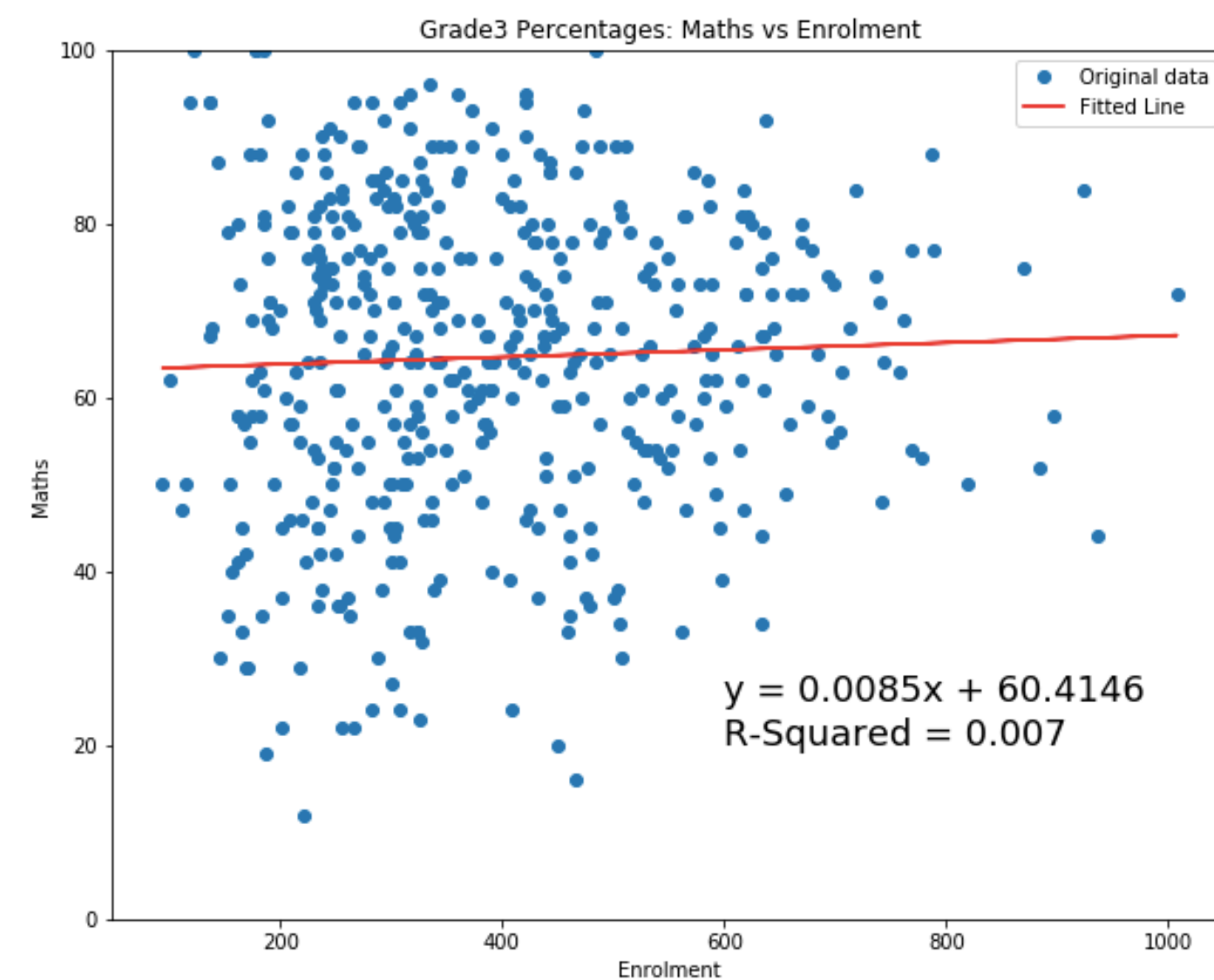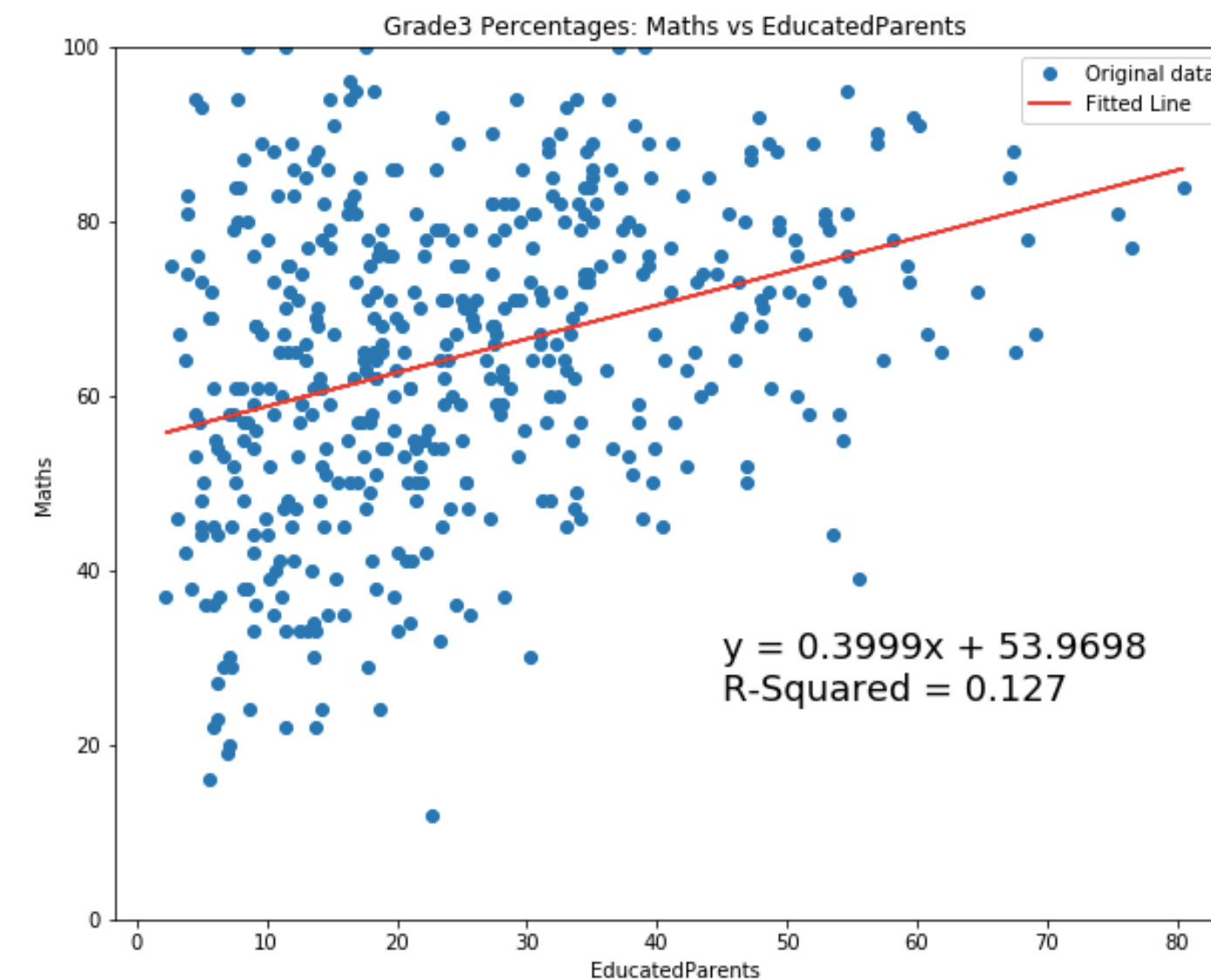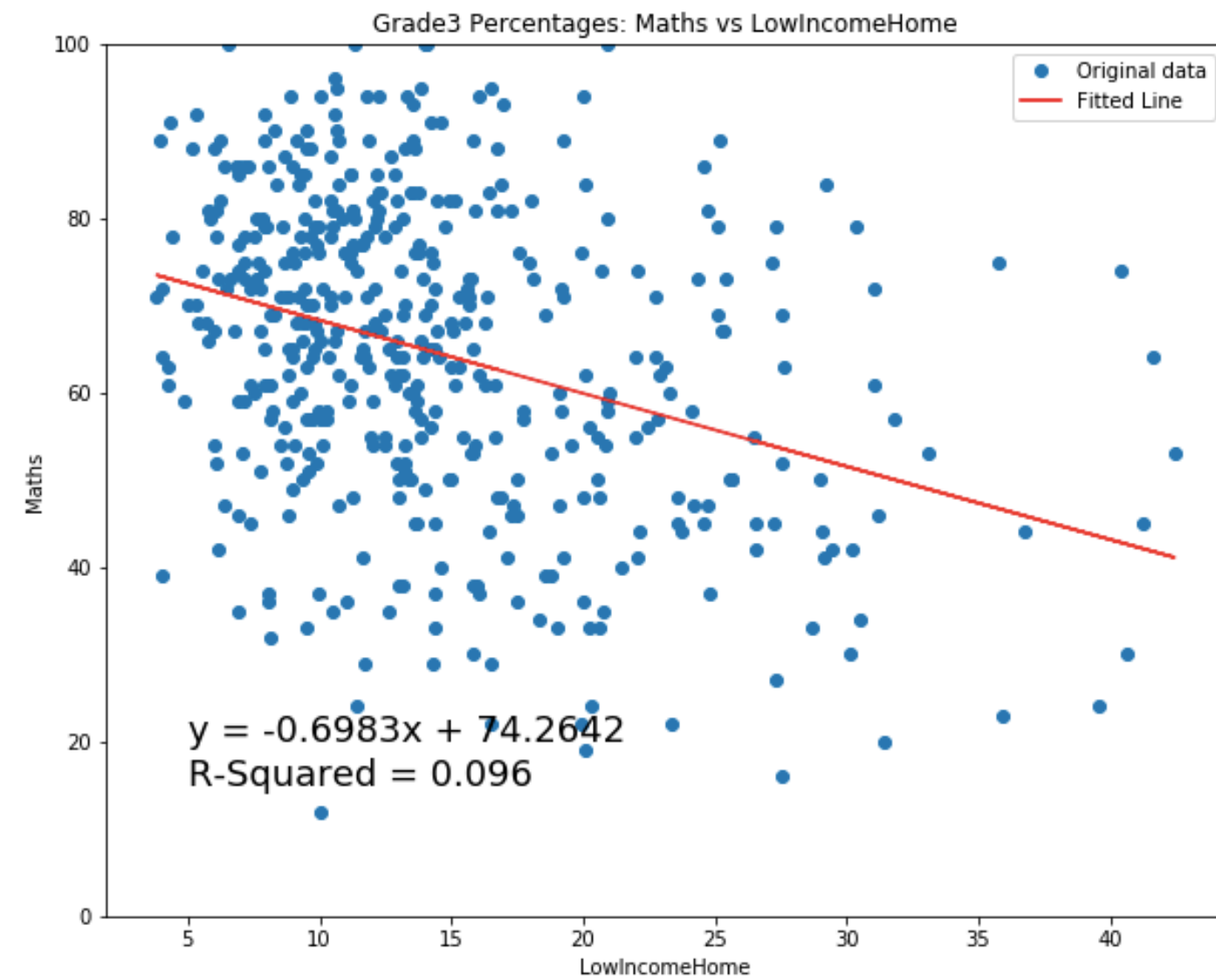
Regression showing the percentages of students who passed the EQAO Maths test with each of the explanatory variables

▸ Strong negative relationship with LowIncomeHome

▸ Strong positive relationship with EducatedParents

▸ No discernible relationship with Enrolment

▸ Moderately strong relationship with NotEnglish

	R-Squared	Adj R-Squared	Coefficient	P-Value	SE	Observations
GRADE 3						
Maths, LowIncomeHome	0.125	0.125	-0.8116	0.000	0.100	483
Maths, EducatedParents	0.103	0.101	0.3876	0.000	0.053	483
Maths, Enrolment	**0.004**	0.002	0.0069	**0.160**	0.005	483
Maths, NotEnglish	0.035	0.033	0.031	0.000	0.031	483
Maths, LowIncomeHome,EducatedParents	0.174	0.170	-0.6437, 0.2818	0.000, 0.000	0.103, 0.054	483
Maths, LowIncomeHome, EducatedParents, Enrolment	0.181	0.176	—0.6793, 0.3104, -0.0096	0.000, 0.000, **0.046**	0.104, 0.056, 0.005	483
Maths, LowIncomeHome, EducatedParents, Enrolment, NotEnglish	**0.237**	0.230	-0.7913, 0.2628, -0.0025, 0.1697	0.000, 0.000, **0.596**, 0.000	0.102, 0.054, 0.005, 0.029	483
Reading, LowIncomeHome	0.119	0.117	-0.6544	0.000	0.083	483
Reading, EducatedParents	0.124	0.122	0.3507	0.000	0.043	483
Reading, Enrolment	0.016	0.014	0.0112	0.006	0.004	483
Reading, NotEnglish	0.018	0.016	0.0751	0.004	0.026	483
Reading, LowIncomeHome,EducatedParents	0.185	0.181	-0.4938, 0.2696	0.000, 0.000	0.084, 0.044	483
Reading, LowIncomeHome, EducatedParents, Enrolment	0.186	0.180	-0.5024, 0.2765, -0.0023	0.000, 0.000, **0.557**	0.085, 0.046, 0.004	483
Reading, LowIncomeHome, EducatedParents, Enrolment, NotEnglish	**0.222**	0.215	-0.5772, 0.2448, 0.0024, 0.1134	0.000, 0.000, **0.550**, 0.000	0.085, 0.045, 0.004, 0.024	483

- R-squared of 0.004 indicates there is very little that explains the variations in the Maths score and Enrolment

- R-squared values of 0.237 and 0.222 in multiple regression indicates there is an association with Maths and Reading and the exploratory variables, more specifically with LowIncomeHome and EducatedParents

- P-value for Enrolment remain high in all regression tests.

GRADE 6: REGRESSION RESULTS

	R-Squared	Adj R-Squared	Coefficient	P-Value	SE	Observations
GRADE 6						
Maths, LowIncomeHome	0.105	0.103	-0.9015	0.000	0.122	463
Maths, EducatedParents	0.113	0.111	0.4686	0.000	0.061	463
Maths, Enrolment	0.000	-0.002	0.0010	0.871	0.006	463
Maths, NotEnglish	0.171	0.169	0.3429	0.000	0.035	463
Maths, LowIncomeHome,EducatedParents	0.161	0.158	-0.6524, 0.3527	0.000, 0.000	0.127, 0.064	463
Maths, LowIncomeHome, EducatedParents, Enrolment	0.188	0.183	-0.7144, 0.4377, -0.0230	0.000, 0.000, 0.000	0.126, 0.066, 0.006	463
Maths, LowIncomeHome, EducatedParents, Enrolment, NotEnglish	0.383	0.377	-0.9772, 0.3320, -0.0130, 0.3798	0.000, 0.000, 0.014, 0.000	0.122, 0.059, 0.005, 0.032	463
Reading, LowIncomeHome	0.145	0.143	-0.6001	0.000	0.068	463
Reading, EducatedParents	0.112	0.110	0.2638	0.000	0.035	463
Reading, Enrolment	0.008	0.006	0.0066	0.054	0.003	463
Reading, NotEnglish	0.032	0.030	0.0839	0.000	0.022	463
Reading, LowIncomeHome,EducatedParents	0.191	0.187	-0.4731, 0.1798	0.000, 0.000	0.071, 0.035	463
Reading, LowIncomeHome, EducatedParents, Enrolment	0.197	0.192	-0.4904, 0.2034, -0.0064	0.000, 0.000, 0.057	0.071, 0.037, 0.003	463
Reading, LowIncomeHome, EducatedParents, Enrolment, NotEnglish	0.247	0.241	-0.5662, 0.1729, -0.0035, 0.1096	0.000, 0.000, 0.285, 0.000	0.070, 0.037, 0.003, 0.020	463

- R-squared values of 0.000 for Maths and 0.008 for Reading indicates no discernible association with Enrolment

- R-squared values of 0.383 and 0.247 for multi-variants moderately explains the variations in the EQAO scores for Maths and Reading

	R-Squared	Adj R-Squared	Coefficient	P-Value	SE	Observations
GRADE 9						
MathsAcad, LowIncomeHome	0.028	0.024	-0.2503	0.004	0.087	297
MathsAcad, EducatedParents	0.159	0.156	0.3696	0.000	0.050	297
MathsAcad, Enrolment	0.023	0.020	0.0037	0.008	0.001	297
MathsAcad, NotEnglish	0.026	0.022	0.0547	0.006	0.020	297
MathsAcad, LowIncomeHome,EducatedParents	0.170	0.165	-01648, 0.3545	0.043, 0.000	0.081, 0.050	297
MathsAcad, LowIncomeHome, EducatedParents, Enrolment	0.171	0.162	-0.1636, 0.3472, 0.0006	0.044, 0.000, 0.656	0.081, 0.53, 0.001	297
MathsAcad, LowIncomeHome, EducatedParents, Enrolment, NotEnglish	**0.183**	0.172	-0.2407, 0.2924, 0.0015, 0.0458	0.007, 0.000, 0.277, 0.036	0.089, 0.058, 0.001, 0.022	297
GRADE 10						
Osslt, LowIncomeHome	0.071	0.068	—0.5281	0.000	0.108	312
Osslt, EducatedParents	0.174	0.171	0.5278	0.000	0.065	312
Osslt, Enrolment	**0.000**	-0.003	0.0001	**0.953**	0.002	312
Osslt, NotEnglish	0.112	0.109	0.1499	0.000	0.024	312
Osslt, LowIncomeHome,EducatedParents	0.210	0.204	-0.3810, 0.4800	0.000, 0.000	0.102, 0.065	312
Osslt, LowIncomeHome, EducatedParents, Enrolment	0.224	0.217	-0.3862, 0.5241, -0.0038	0.000,0.000, 0.016	0.101, 0.067, 0.002	312
Osslt, LowIncomeHome, EducatedParents, Enrolment, NotEnglish	**0.312**	0.303	-0.6456, 0.3188, -0.0002, 0.1608	0.000, 0.0.000, 0.888, 0.000	0.104, 0.071, 0.002, 0.026	312

▸ R-squared=0.000 for Enrolment , and the p-value of 0.953 for Grade 10, indicates there is no discernible association with the number of students passing the EQAO OSSLT

▸ R-squared=0.312 for Grade10 multi-variants regression indicates some variations of EQAO OSSLT scores can be explained by the explanatory variables

▸ The parents income and education level has a linear relationship with the student's EQAO test score.

▸ For students where English is not their first language there is a positive and moderately linear association with the EQAO test scores; which is a result of most children having an early start at learning English in kindergarten.

▸ There is very little association between Enrolment and the EQAO test scores.

In conclusion, we reject the null hypothesis in favour of the alternate hypothesis. The students success in EQAO tests is dependent on the parents income and education levels. Schools where there are more children of rich and well educated parents get better results. School enrolment size has no bearing on the percentage of students passing. The growing percentage of school students who do not speak English as their first language does not negatively impact the percentage of students who pass the EQAO tests.

ANY QUESTIONS?