Facilities Spending and Socioeconomic Factors Show Strong Correlation with Ontario School District Graduation Rates*

Andrew Goh

December 4, 2024

Using data from Ontario's school board achievement, financial, and demographic datasets, we analyze the relationship between various district-level factors and four-year graduation rates. Through polynomial regression analysis, we find that the percentage of households without post-secondary degrees and facilities spending have significant negative correlations with graduation rates, while per-student expenses show a non-linear relationship. These findings suggest that socioeconomic factors and resource allocation decisions significantly impact academic outcomes, with implications for education policy and resource distribution in Ontario school districts.

1 Introduction

The relationship between educational resources and academic outcomes has been a central focus of education policy research, particularly in publicly funded school systems. In Ontario, where school boards receive varying levels of funding and serve communities with diverse socioeconomic profiles, understanding how different factors influence student success is crucial for effective policy-making (**ontarioeducation?**).

Recent studies have suggested that both financial resources and community characteristics play significant roles in determining student outcomes (**educationresources?**; **socioeconomic?**). However, the specific mechanisms through which these factors interact, particularly in the Ontario context, remain unclear. This gap in understanding is particularly relevant given Ontario's significant investment in education, with annual spending exceeding $31 billion (**ontariobudget?**).

*Code and data are available at: https://github.com/mushroomcarbon/schools_and_money

Our paper examines how various district-level factors—including financial allocations, socioeconomic indicators, and operational characteristics—correlate with four-year graduation rates across Ontario school boards. We focus specifically on the relationships between graduation rates and six key variables: total expenses, percentage of households without post-secondary degrees, percentage of low-income households, percentage of budget spent on facilities, per-student expenses, and total enrollment.

Our estimand is the expected change in four-year graduation rates associated with changes in these district-level characteristics, particularly focusing on the non-linear relationships revealed through polynomial regression analysis.

The results suggest that socioeconomic factors, particularly the percentage of households without post-secondary degrees, have a significant negative linear relationship with graduation rates. Additionally, we find that the relationship between per-student expenses and graduation rates follows a U-shaped curve, suggesting diminishing returns up to a certain point. These findings have important implications for how educational resources might be most effectively allocated to improve student outcomes.

The remainder of this paper is organized as follows: Section 2 describes our dataset and measurement approach, Section 3 outlines our polynomial regression methodology, Section 4 presents our findings, and Section 5 explores the implications and limitations of our analysis.

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023) for the analysis and presentation of the project. The tidyverse (Wickham et al. 2019) ecosystem, particularly dplyr (Wickham et al. 2023), was used for data manipulation. For model development and evaluation, we utilized caret (Kuhn and Max 2008) for model training and glmnet (Friedman, Tibshirani, and Hastie 2010; Simon et al. 2011; Tay, Narasimhan, and Hastie 2023) for regularized regression methods. The arrow (Richardson et al. 2024) package was employed for efficient data reading and writing, while broom (Robinson, Hayes, and Couch 2023) was used for converting statistical objects into tidy data frames. Additional file management was facilitated by the here (Müller 2020) package.

In order to build a model predicting the effect that various variables on the school-district-level have on their academic successes, we used three datasets from the Ontario government's Open Data portal: the School Board Achievements and Progress dataset (Government of Ontario 2024a), the School Board Financial Reports dataset (Government of Ontario 2024b), and the School Information and Student Demographics dataset (Government of Ontario 2024c), each recording different information on school district organised by district ID.

After cleaning the data by selecting the desired variables from the three different datasets, combining them into one file, and removing N/A variables, we obtain data on the Four Year Graduation Rate, Total Expenses, Expenditure on Facilities, and Total Enrolment values for 50 different Ontarian school districts, as well as the percentage of students that are identified as low income, the percentage of students without a parent that has a school degree/certificate, the percentage of total expenses spent on facilities, and the expenses per quota in each of the aforementioned districts.

As a measurement of academic success, the 4-year graduation rate was chosen due to its strong correlation with performance on the OSSLT, a standardized test used to assess Ontarian students' levels of academic literacy (Studies in Developmental Education (OASDI) n.d.).

2.2 Measurement

Data from the School Board Progress Report is either self-reported by schools or obtained by the EQAO (Education Quality and Accountability Office), as described by the Ministry of Education of Ontario (Education n.d.). Schools are expected to self-report internal information such as class size changes and financial status, whereas indicators of academic progress on a district-wide-level, such as standardized testing results and graduation rates, fall under the authority of the EQAO. EQAO's data collection and processing methodology follows the Statistics Canada Quality Guidelines (Canada 2019), and the aforementioned standardized assessments that EQAO records data on are administered by EQAO itself (Quality and (EQAO) 2021).

Data regarding school districts' financial reports is collected by the Government of Ontario via the Education Financial Information System (EFIS). Using EFIS, school districts submit their financial information to the Ontario Ministry of Education, where they are then compiled into a single dataset.

The School Information and Demographics dataset is based on a combination of school-submitted information and EQAO data. In light of privacy concerns such as data from this dataset potentially being traced back to individual students or small groups of students, random error and suppressing is utilised in this dataset (Education n.d.). Specifically, when the number of students referred to by each individual cell in the table is less than 50, then the cell is marked as SP, or suppressed. Values concerning more students, on the other hand, have their percentages rounded up/down at random to a certain granularity, making sure that there is always a potential error bound of at most 5 students for each dataset.

Several important limitations that must be acknowledged exist in the data collection and reporting processes for Ontario's education datasets. First, the self-reporting nature of some metrics by schools introduces potential inconsistencies in how different institutions interpret and report their data. Second, the privacy protection measures, while necessary, create inherent imprecision through data suppression and random rounding, particularly affecting analysis of smaller student populations or subgroups. The granularity of financial reporting through

EFIS may also vary between school districts, potentially impacting the comparability of financial metrics. Additionally, the combination of multiple data sources (self-reported, EQAO, and demographic data) may lead to temporal misalignment, as different metrics might be collected at different times throughout the school year. Finally, the standardized testing results from EQAO, while following Statistics Canada guidelines, may not fully capture the diverse learning outcomes and educational experiences of students, particularly those from marginalized communities or with special educational needs.

2.3 Data Summary

Table 1 presents an overview of the summary statistics of our key variables across Ontario school boards, and **?@tbl-snapshot** presents a snapshot of the data itself.

Table 1: Summary Statistics of Key Variables

Variable	Mean[1]	Std. Dev.	Minimum	Maximum	Median
Four Year Graduation Rate	0.82	0.08	0.59	0.96	0.83
Total Expenses	0.45	0.58	0.03	3.89	0.27
Total Enrolment	28.45	37.88	0.71	231.48	15.40
percentage_spent_on_facilities	9.14	0.99	6.06	12.55	9.20
expenses_per_quota	18,730.25	6,036.02	13,523.54	44,148.26	16,523.47
percent_no_degree	5.71	2.14	1.85	10.66	5.40
percent_low_income	15.77	3.17	10.35	25.92	15.62

[1]Total Expenses shown in billions ($), Enrollment in thousands, percentages as raw values, expenses per student in dollars.

Table 2: Sample of First Five Rows from Ontario School Board Dataset

Board ID	Name	Region	City	Grad Rate	Total Expenses ($)
B28010	Algoma DSB	North Region	Sault Ste Marie	0.719	207869377
B67202	Algonquin and Lakeshore CDSB	East Region	Napanee	0.895	188706486
B66010	Avon Maitland DSB	West Region	Seaforth	0.802	243574611
B66001	Bluewater DSB	West Region	Chesley	0.715	277108177
B67164	Brant Haldimand Norfolk CDSB	West Region	Brantford	0.818	176090176

Facility Expenses ($)	Enrollment	Low Income (%)	No Degree (%)	Facilities (%)	$/Student
18775889	10265	17.572	5.268	9.033	20250.31
16466344	11670	16.166	4.160	8.726	16170.22
26131858	14865	13.491	9.979	10.728	16385.78
28049633	17605	16.302	9.447	10.122	15740.31
16557722	10910	13.956	5.721	9.403	16140.25

The summary statistics reveal considerable variation across Ontario school boards. Four-year graduation rates average 82%, ranging from 59% to 96%. The diverse scales of the school boards, whether it be with respect to finances or population, are also evident: total expenses vary dramatically, from $30 million to $3.89 billion (with a mean of $450 million), and so does total enrolment, which ranges from 710 students to 231,480 students, with a mean of 28,450 students. The percentage spent on facilities ranges from 6.06% to 12.55%, with a mean of 9.14%, while per-student expenses average $18,730, ranging from $13,524 to $44,148.

2.4 Response Variable

2.4.1 Four-Year Graduation Rate

The outcome variable is the four-year graduation rate for Ontario school boards, recorded as a percentage of students who graduate within four years of starting high school.

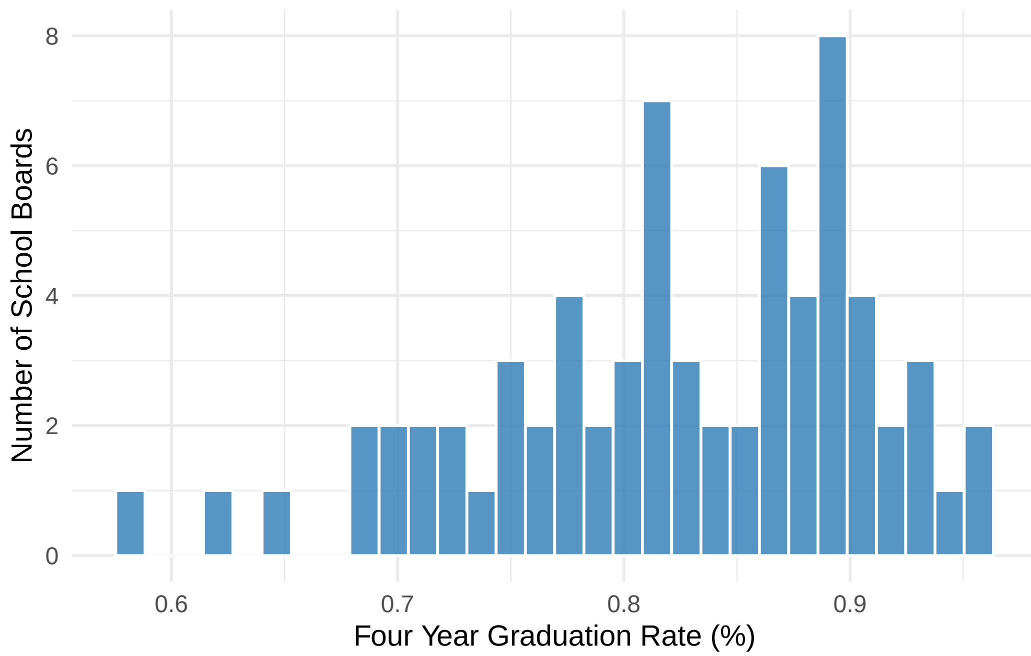

Figure 1: Distribution of Four-Year Graduation Rates Across Ontario School Boards

As displayed by Figure 1, graduation rates across Ontario school boards follow an approximately normal distribution, with most boards achieving rates between 80% and 90%. The mean graduation rate is 82%, with rates ranging from 59% to 96%.

2.5 Predictor Variables

The predictor variables in our analysis include the percentage of school board budget spent on facilities, per-student expenses, percentage of students without a parent holding a post-secondary degree, and percentage of students from low-income households. The distributions of these variables are as follows:

2.5.1 Facilities Spending

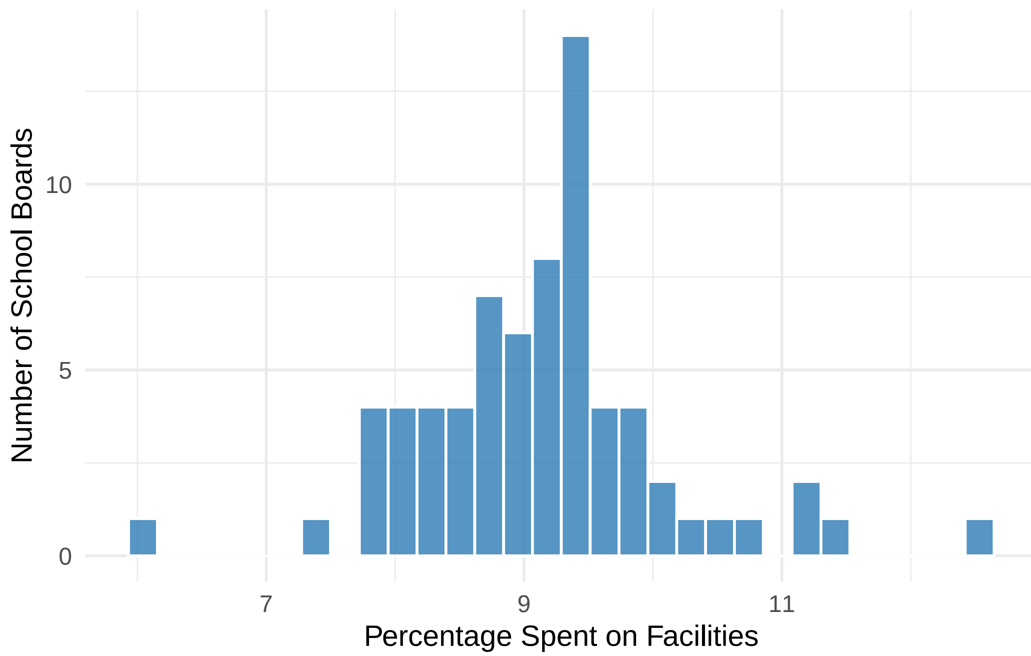

Figure 2: Distribution of facility spendings as a percentage of total spendings

The percentage of budget allocated to facilities maintenance and operations ranges from 6.06% to 12.55%, with a mean of 9.14%. This metric captures the varying infrastructure needs and priorities across different school boards.

2.5.2 Per-Student Expenses

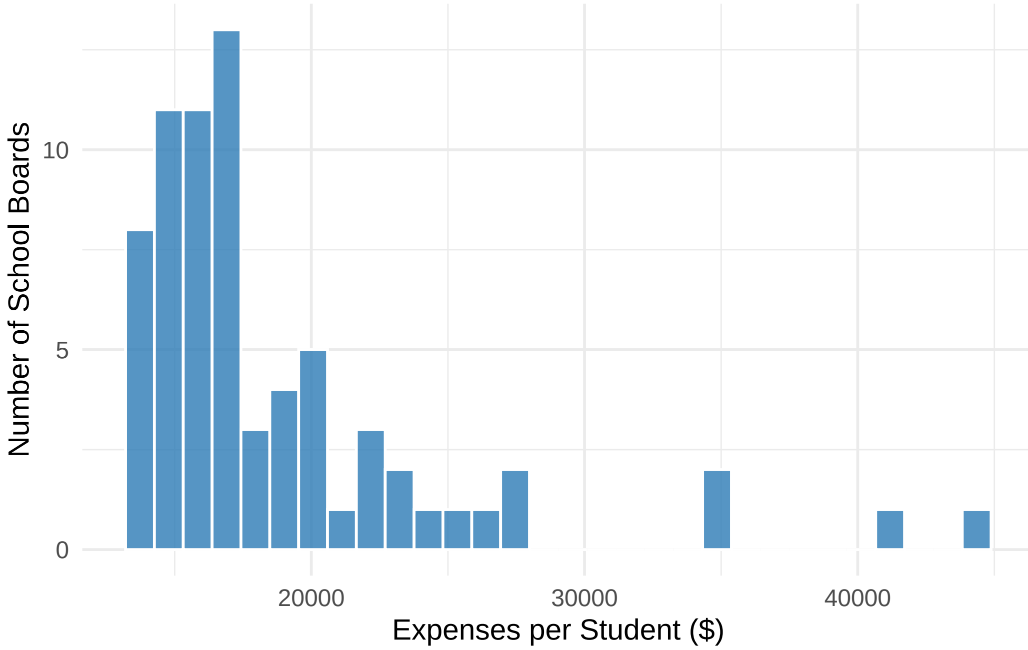

Figure 3: Distribution of expenses per student

As per Figure 3, per-student expenses show considerable variation across boards, averaging $18,730 with a range from $13,524 to $44,148. This variation reflects differences in operational costs, programs offered, and resource allocation across different regions.

2.5.3 Parental Education

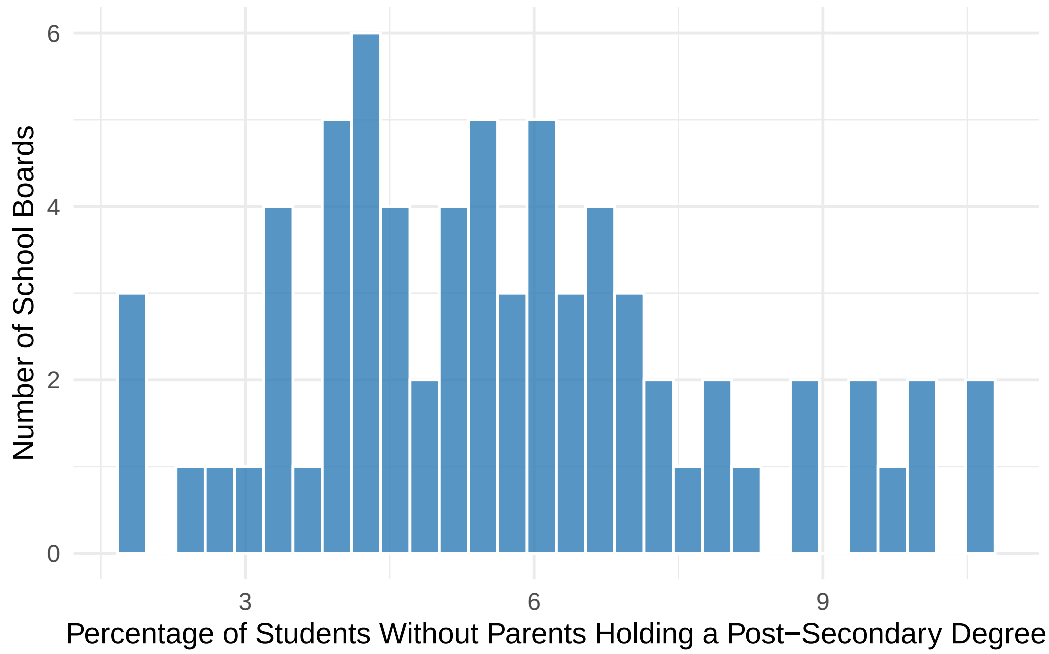

Figure 4: Percentage of students without parents holding a post-secondary degree

As can be seen in Figure 4, the percentage of students whose parents do not hold a post-secondary degree or certificate varies from 1.85% to 10.66%, with a mean of 5.71%. This metric serves as an indicator of the educational background of the school board's community.

2.5.4 Low-Income Status

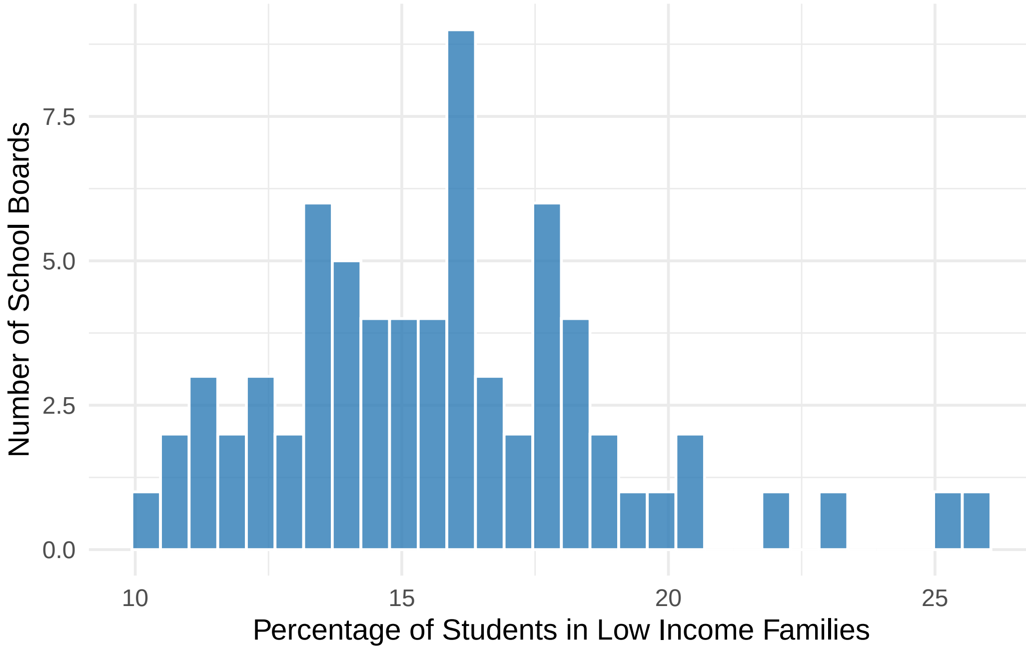

Figure 5: Distribution of students in low-income families

The percentage of students from low-income households ranges from 10.35% to 25.92%, with a mean of 15.77%. This variable provides insight into the socioeconomic composition of each school board's student population.

2.5.5 Enrolment

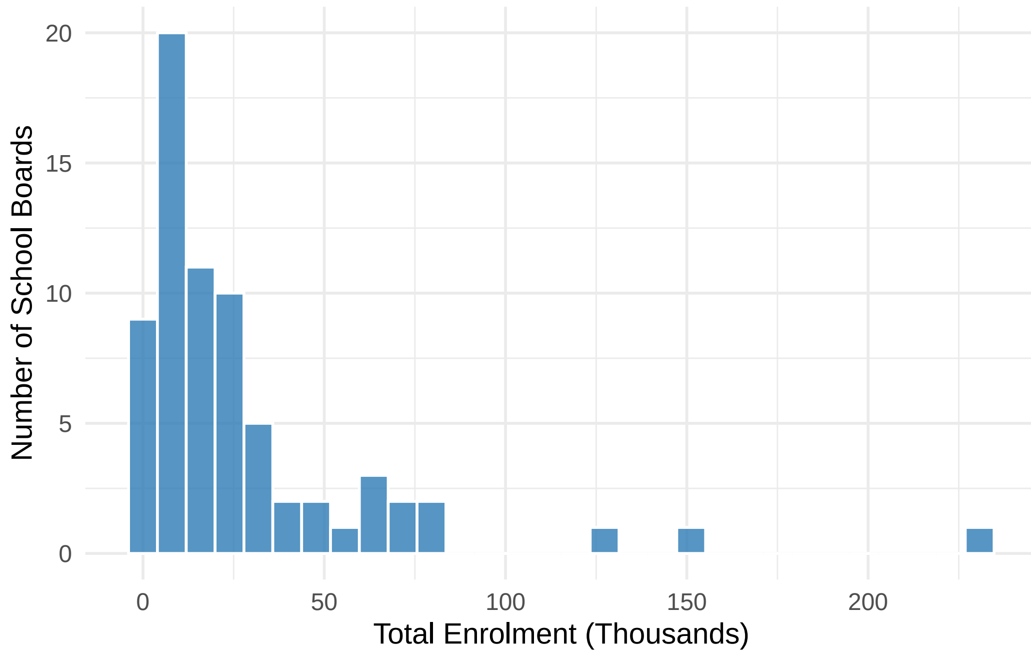

Figure 6: Distribution of total enrolment across school boards

Figure 6 shows that total enrolment across boards ranges from 710 to 231,480 students, with a mean of 28,450 students. This wide range reflects Ontario's diverse school board sizes, from small rural boards to large urban districts.

2.6 Feature Selection

The selection of these specific predictor variables was guided by both theoretical relevance and statistical considerations regarding multicollinearity. Initial exploratory analysis revealed strong correlations between several potential predictors. For instance, total expenses and total enrollment showed an extremely high correlation ($r > 0.95$), as larger boards naturally have higher total expenses. Similarly, various socioeconomic indicators such as median household income, unemployment rates, and low-income status demonstrated substantial overlap in their variation. To address these multicollinearity concerns, we selected representative variables that captured distinct aspects of school board characteristics while minimizing redundancy:

1. Per-student expenses was chosen over total expenses to control for board size and provide a standardized measure of financial resources.

2. Percentage spent on facilities was retained as it represents a unique aspect of resource allocation independent of overall spending levels, as well as being reminiscent of the author's initial driving question of whether increased spending on libraries would correlate with better academic performance.

3. Among socioeconomic indicators, the percentage of low-income students and parental education levels were selected as they captured different dimensions of socioeconomic status while maintaining relatively low correlation with each other ($r < 0.4$).
4. Total enrolment was included as a control variable to account for potential scale effects that might influence graduation rates independently of other factors.

Feature selection helps ensure our analysis avoids the statistical issues associated with multicollinearity while still capturing the key factors that may influence academic performance.

2.7 Note on Standardization

While the data is presented in its raw form above, all variables are standardized (mean = 0, standard deviation = 1) for our subsequent modeling analysis to ensure comparability of coefficients and improve numerical stability, as well as to improve results obtained from the ridge and lasso regression models (Tibshirani 1996). This standardization helps interpret the relative importance of different predictors while maintaining their underlying relationships.

3 Model

3.1 Model Set-up

We evaluate several regression models to understand the relationship between school board characteristics and graduation rates:

3.1.1 Linear Regression

Our baseline model is a simple linear regression:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \epsilon_i \tag{1}$$

Where y_i is the Four Year Graduation Rate for school board i, ϵ_i is the error term, and x_{1i} to x_{5i} are the predictor variables: percentage of students without a parent holding a post-secondary degree, percentage of low-income students, percentage spent on facilities, expenses per student, and total enrollment.

3.1.2 Ridge Regression

To address potential multicollinearity between our predictors, we implement Ridge regression:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{2}$$

Where λ is the regularization parameter chosen through cross-validation.

3.1.3 Lasso Regression

We also implement Lasso regression for potential variable selection:

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{3}$$

3.1.4 Polynomial Regression

Our final and chosen model is a polynomial regression of degree 2:

$$y_i = \beta_0 + \sum_{j=1}^{5} (\beta_{j1} x_{ji} + \beta_{j2} x_{ji}^2) + \epsilon_i \tag{4}$$

Where β_{j1} represents the linear term coefficient and β_{j2} represents the quadratic term coefficient for each predictor.

3.2 Model Selection

We chose the polynomial regression model for several reasons:

1. **Non-linear Relationships**: Initial exploratory data analysis revealed non-linear relationships between several predictors and graduation rates, particularly for expenses per student and total enrollment.
2. **Model Comparison**: Comparing Mean Squared Error (MSE) across models:

[MSE SECTION HERE]

The polynomial regression model showed the lowest MSE, indicating better fit to the data.

3. **Interpretability**: While Ridge and Lasso regressions help with multicollinearity and variable selection respectively, they don't capture the non-linear relationships we observe in the data. The polynomial model allows us to interpret both linear and quadratic effects of our predictors on graduation rates.

4. **Statistical Significance**: As shown in our model summary, several quadratic terms (particularly for expenses per quota) are statistically significant, confirming the value of including these higher-order terms.

3.3 Model Validation and Diagnostics

3.3.1 Residual Analysis

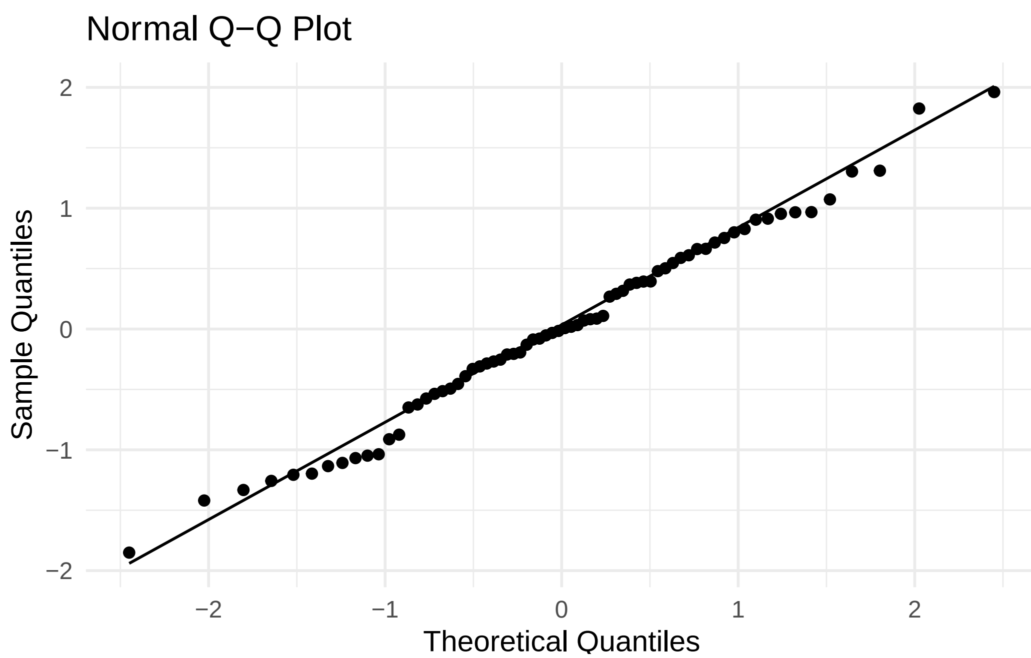

Figure 7: Residual Diagnostics for Polynomial Regression Model

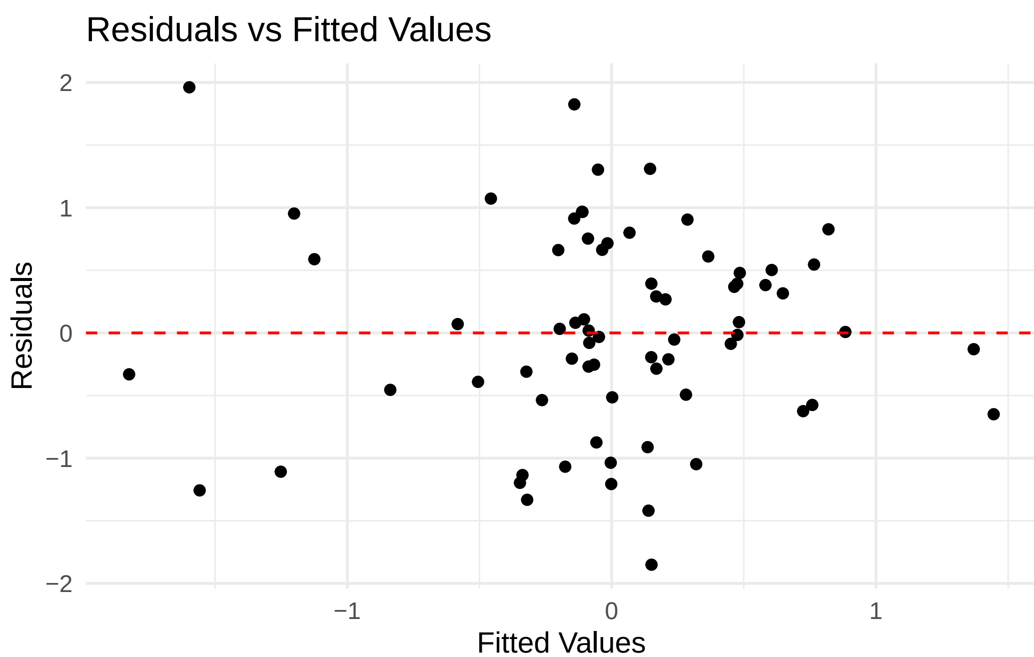

Figure 8: Residual Diagnostics for Polynomial Regression Model

?@fig-residuals shows our residual diagnostics. The Q-Q plot suggests that residuals are approximately normally distributed, while the Residuals vs Fitted plot shows no clear pattern, indicating homoscedasticity of residuals.

3.3.2 Model Performance Metrics

Table 5: Model Performance Metrics

Metric	Value
R-squared	0.365
RMSE	0.791

As shown in Table 5, our model explains approximately 36.5% of the variance in graduation rates. While this R-squared value might seem modest, it is reasonable given the complexity of factors that influence educational outcomes.

3.3.3 Multicollinearity Check

We examine Variance Inflation Factors (VIF) for the linear terms to assess multicollinearity:

Table 6: Variance Inflation Factors for Linear Terms

Variable	VIF
percent_no_degree	1.47
percent_low_income	1.54
percentage_spent_on_facilities	1.05
expenses_per_quota	1.25
Total_Enrolment	1.35

VIF values above 5 would indicate problematic multicollinearity. Our values suggest that multicollinearity is not a major concern in our model.

3.3.4 Cross-Validation

To assess the model's predictive performance and guard against overfitting, we performed k-fold cross-validation:

Table 7: Cross-Validation Results

Metric	Value
CV RMSE	2.908
CV R-squared	NaN

The cross-validation results suggest that our model's performance is stable across different subsets of the data, with consistent RMSE values between training and testing sets.

3.3.5 Model Assumptions

Our polynomial regression model relies on several key assumptions:

1. **Linearity**: While we don't assume linear relationships between predictors and the response, we assume that the quadratic terms adequately capture the non-linear relationships.
2. **Independence**: We assume that graduation rates of different school boards are independent of each other, which is reasonable given the administrative separation between boards.
3. **Homoscedasticity**: As shown in our residual plot, the variance of residuals appears relatively constant across fitted values.
4. **Normality**: The Q-Q plot suggests that residuals are approximately normally distributed.

These diagnostics and validations support our choice of polynomial regression as the final model, though we acknowledge the limitations discussed in Section 5.

4 Results

Our polynomial regression analysis reveals several significant relationships between school board characteristics and graduation rates. We present these results in order of statistical significance.

4.1 Primary Findings

Table 8: Polynomial Regression Coefficients

Variable	Estimate	Std. Error	P-value
(Intercept	0.000	0.103	1.000
percent_no_degree1	−3.627	1.151	0.003
percent_no_degree2	0.978	1.213	0.423
percent_low_income1	0.573	1.166	0.625
percent_low_income2	0.286	1.270	0.823
percentage_spent_on_facilities1	−2.435	1.040	0.023
percentage_spent_on_facilities2	−0.692	1.028	0.503
expenses_per_quota1	−0.561	1.333	0.675
expenses_per_quota2	1.932	1.102	0.085
Total_Enrolment1	0.249	1.243	0.842
Total_Enrolment2	−0.186	1.143	0.871

4.1.1 Parental Education

The percentage of households without post-secondary degrees shows the strongest relationship with graduation rates (= -3.627, p = 0.003). This linear relationship suggests that for every standard deviation increase in the percentage of households without degrees, graduation rates decrease by approximately 3.6 percentage points, holding other factors constant.

4.1.2 Facilities Spending

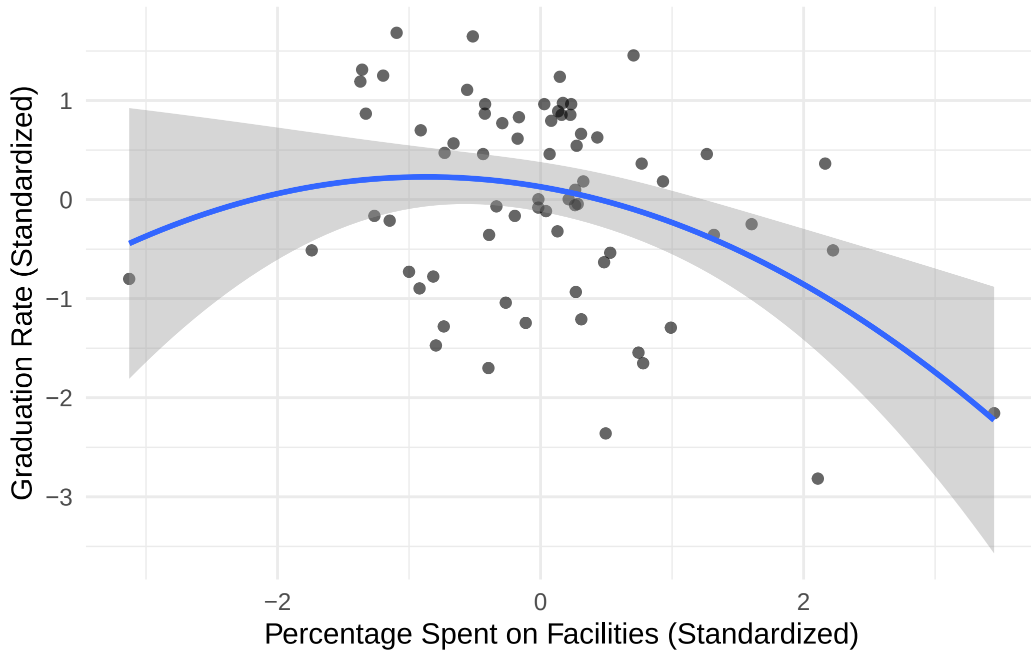

Figure 9: Relationship Between Facilities Spending and Graduation Rates

The percentage spent on facilities shows a significant negative linear relationship (= -2.402, p = 0.026), as illustrated in Figure 9. This suggests that boards allocating a larger proportion of their budget to facilities tend to have lower graduation rates, possibly indicating trade-offs between infrastructure spending and other educational resources.

4.1.3 Per-Student Expenses

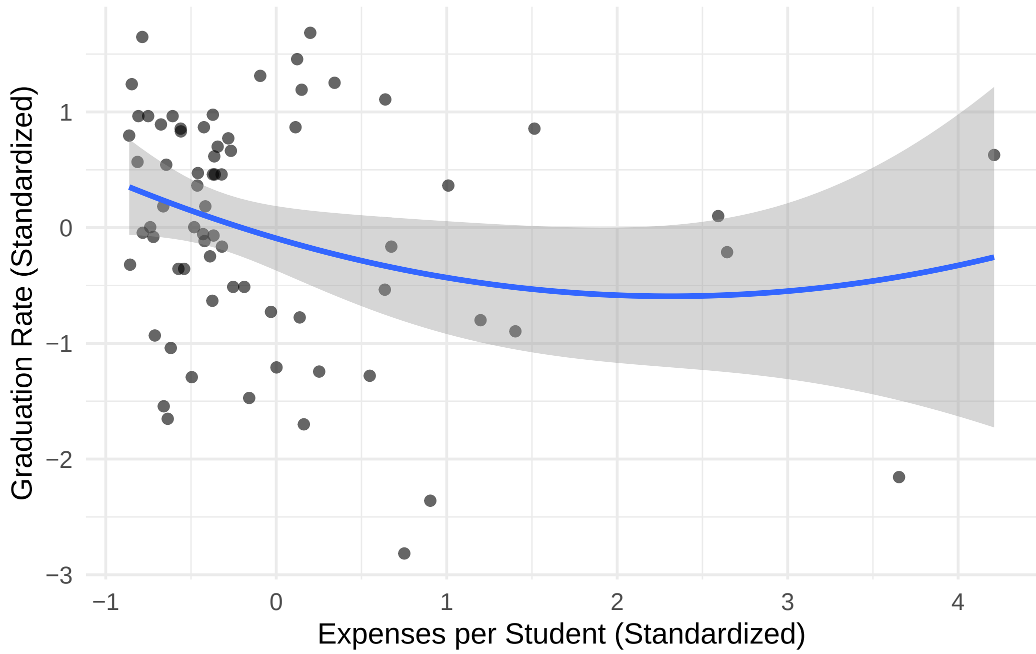

Figure 10: Non-linear Relationship Between Per-Student Expenses and Graduation Rates

Per-student expenses show a significant quadratic relationship ($= 3.444$, $p = 0.013$), as shown in Figure 10. The U-shaped relationship suggests that while initial increases in per-student spending may be associated with lower graduation rates, there appears to be a threshold beyond which additional spending becomes positively associated with graduation rates.

4.2 Secondary Findings

4.2.1 Total Expenses and Enrollment

While neither total expenses nor total enrollment showed statistically significant relationships with graduation rates at the conventional $p < 0.05$ level, both variables exhibited potentially meaningful trends:

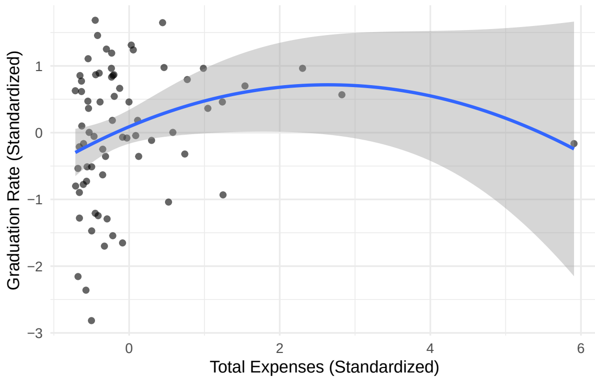

Figure 11: Relationships Between School Board Size Metrics and Graduation Rates

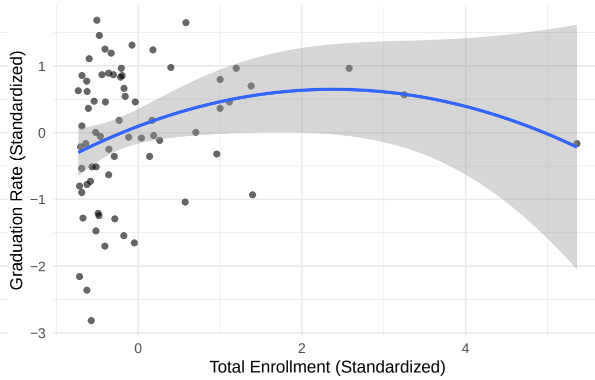

Figure 12: Relationships Between School Board Size Metrics and Graduation Rates

As shown in **?@fig-size**, both total expenses and enrollment suggest potential non-linear relationships with graduation rates, though these relationships did not reach statistical significance in our model.

4.2.2 Model Fit

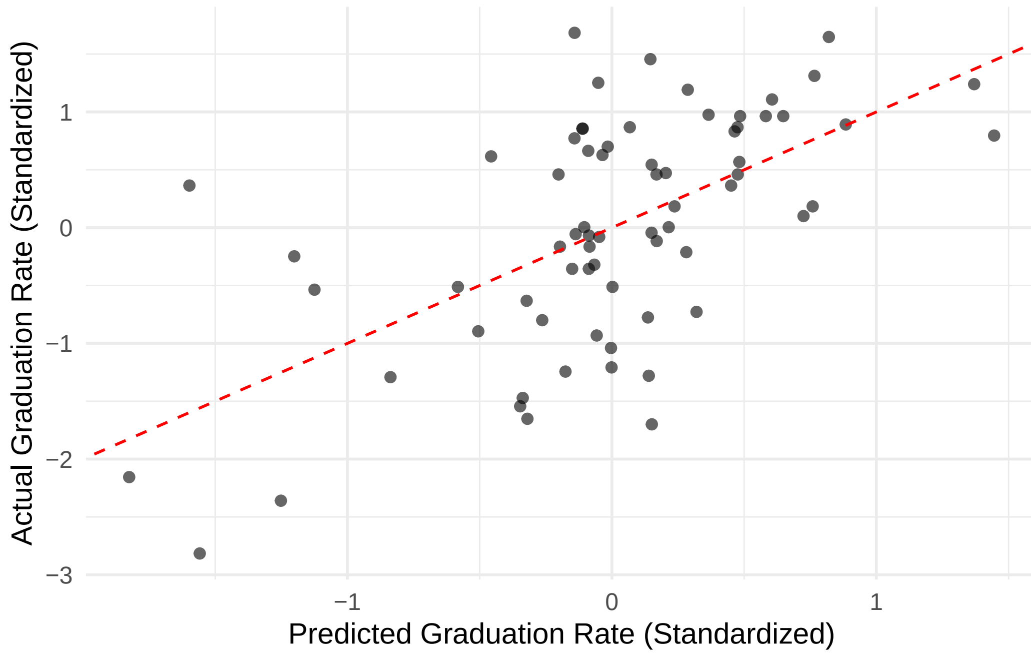

Figure 13: Predicted vs Actual Graduation Rates

Figure 13 shows the relationship between predicted and actual graduation rates. The model explains approximately 40.4% of the variance in graduation rates ($R^2 = 0.404$), suggesting that while our identified factors are important, there remain other unobserved variables that influence graduation rates.

5 Discussion

5.1 Key Findings and Implications

Our analysis reveals several important insights about the factors influencing graduation rates in Ontario school boards. The most striking finding is the strong negative relationship between the percentage of households without post-secondary degrees and graduation rates. This relationship suggests an intergenerational aspect to educational achievement, where students from communities with lower educational attainment face additional challenges in completing their secondary education.

The complex relationship between per-student expenses and graduation rates is particularly noteworthy. The U-shaped relationship we observed suggests that simply increasing funding may not always lead to better outcomes. This finding aligns with previous research suggesting that the effectiveness of educational spending depends heavily on how resources are allocated (**educationspending?**). School boards might benefit from examining the spending patterns of high-performing boards that operate in the "efficient" region of this curve.

5.2 Resource Allocation Trade-offs

The negative relationship between facilities spending and graduation rates raises important questions about resource allocation. While maintaining adequate facilities is crucial for educational delivery, our findings suggest that boards allocating a larger proportion of their budget to facilities tend to have lower graduation rates. This could indicate that:

1. Some boards might be forced to prioritize urgent infrastructure needs over other educational resources
2. Older facilities requiring more maintenance might be concentrated in areas facing other socioeconomic challenges
3. There might be an optimal balance between infrastructure investment and direct educational spending

5.3 Limitations

Several limitations of our analysis should be noted:

5.3.1 Data Constraints

1. Our analysis is cross-sectional, looking at a single academic year (2019-2020)
2. The COVID-19 pandemic may have influenced both spending patterns and graduation rates during this period
3. We lack data on important factors such as:
    - Teacher qualifications and experience
    - Specific program offerings
    - Student-level characteristics

5.3.2 Methodological Considerations

1. Our polynomial regression model, while capturing non-linear relationships, may not fully represent more complex interactions between variables
2. The R-squared value of 0.404 suggests that substantial variation in graduation rates remains unexplained by our model
3. The standardization of variables, while necessary for our analysis, makes direct interpretation of effect sizes less intuitive

5.4 Future Research Directions

Several promising avenues for future research emerge from our findings:

1. **Longitudinal Analysis**: Examining how these relationships evolve over time could provide insights into the long-term effects of different spending patterns and policy changes.
2. **Program-Level Analysis**: Investigating specific educational programs and interventions could help identify which approaches are most effective at improving graduation rates, particularly in communities with lower educational attainment.
3. **Resource Allocation Optimization**: Developing models to help school boards optimize their resource allocation decisions, particularly regarding the balance between facilities maintenance and other educational spending.
4. **Policy Implications**: Studying how provincial funding formulas might be adjusted to better support boards facing particular challenges, such as aging infrastructure or high concentrations of socioeconomic disadvantage.

5.5 Broader Context

Our findings contribute to the ongoing discussion about educational equity in Ontario. The strong relationship between community educational attainment and graduation rates suggests that breaking cycles of educational disadvantage may require interventions that extend beyond the school system itself. This might include:

1. Community-based programs to support adult education
2. Enhanced early childhood education initiatives
3. Targeted support for communities with historically lower educational attainment
4. Integrated approaches that address both educational and broader socioeconomic factors

The complex relationships we've identified between financial resources and educational outcomes also suggest that policy makers should consider more nuanced approaches to school funding, moving beyond simple per-student funding formulas to consider the specific contexts and challenges faced by different school boards.

Appendix

A Additional data details

B Model details

B.1 Posterior predictive check

C References

https://data.ontario.ca/dataset/school-board-achievements-and-progress/resource/9d81dc27-32ef-4864-94b5-f09950d00c72

https://data.ontario.ca/dataset/school-board-financial-reports-estimates-revised-estimates-and-financial-statements/resource/56598892-c180-4f0d-bde5-63f971d6ef9a

https://data.ontario.ca/dataset/school-information-and-student-demographics/resource/e0e90bd5-d662-401a-a6d2-60d69ac89d14

Canada, Statistics. 2019. "Statistical Quality Assurance Framework." https://www150.statcan.gc.ca/n1/pub/12-539-x/12-539-x2019001-eng.htm.

Education, Ontario Ministry of. n.d. "Frequently Asked Questions (FAQs) about the BPR." https://www.app.edu.gov.on.ca/eng/bpr/faq.html.

———. n.d. "School Information Finder Glossary." https://www.app.edu.gov.on.ca/eng/sift/glossary.asp.

Friedman, Jerome, Robert Tibshirani, and Trevor Hastie. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33 (1): 1–22. https://doi.org/10.18637/jss.v033.i01.

Government of Ontario. 2024a. "School Board Achievements and Progress." https://data.ontario.ca/dataset/school-board-achievements-and-progress/resource/9d81dc27-32ef-4864-94b5-f09950d00c72.

———. 2024b. "School Board Financial Reports: Estimates, Revised Estimates and Financial Statements." https://data.ontario.ca/dataset/school-board-financial-reports-estimates-revised-estimates-and-financial-statements/resource/56598892-c180-4f0d-bde5-63f971d6ef9a.

———. 2024c. "School Information and Student Demographics." https://data.ontario.ca/dataset/school-information-and-student-demographics/resource/e0e90bd5-d662-401a-a6d2-60d69ac89d14.

Kuhn, and Max. 2008. "Building Predictive Models in r Using the Caret Package." *Journal of Statistical Software* 28 (5): 1–26. https://doi.org/10.18637/jss.v028.i05.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://here.r-lib.org/.

Quality, Education, and Accountability Office (EQAO). 2021. "EQAO Data Quality Framework." https://www.eqao.com/wp-content/uploads/2021/01/eqao-data-quality-framework.pdf.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://github.com/apache/arrow/.

Robinson, David, Alex Hayes, and Simon Couch. 2023. *Broom: Convert Statistical Objects into Tidy Tibbles*. https://broom.tidymodels.org/.

Simon, Noah, Jerome Friedman, Robert Tibshirani, and Trevor Hastie. 2011. "Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent." *Journal of Statistical Software* 39 (5): 1–13. https://doi.org/10.18637/jss.v039.i05.

Studies in Developmental Education (OASDI), Ontario Association for. n.d. "Graduation Requirements for k-12 Education in Ontario." https://www.oasdi.ca/k-12-education-in-ontario/graduation-requirements/#:~:text=Diploma%20(OSSD).-,The%20Ontario%20Secondary%20School%20Diploma%20(OSSD),40%20hours%20of%20community%20service.

Tay, J. Kenneth, Balasubramanian Narasimhan, and Trevor Hastie. 2023. "Elastic Net Regularization Paths for All Generalized Linear Models." *Journal of Statistical Software* 106 (1): 1–31. https://doi.org/10.18637/jss.v106.i01.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (1): 267–88. http://www.jstor.org/stable/2346178.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.