Exploring the Relationship Between Household Income and High School Graduation Rates in Georgia Counties

Ashley Allen, Ida De Vierno, Emilie Pourchet
Georgia Institute of Technology

ECON 3161
Dr. Shatakshee Dhongde
Fall 2017

Abstract: To better understand the factors that contributed to the rise in high school graduation rates between 2008 and 2015, this study explored the relationship between average household income and average four-year graduation rates for public high schools in Georgia counties in 2015. Additional economic factors including teacher salary, spending per pupil, local education revenue, maintenance and operation (M&O) tax, poverty rate and Gini index as well as social factors, including unemployment, class size, educational attainment and teen pregnancy rate were studied. Four ordinary least squares (OLS) models were generated to understand the impact of change in household income on high school graduation rates at all income levels and at the income extremes. Analysis of the results shows that household income is not statistically significant throughout the models for our full data set. When looking solely, at the upper and lower income brackets, household income is significant at the 5% level.

1. Introduction

In 2015, president Obama announced that high school graduation rates had reached an all time high of 83% in the United States (Office of the Press Secretary, 2016). After winning the presidential election in 2008, he promised to work on improving the education system in the United States. As he worked to achieve this goal, graduation rate trends varied across the United States and some states saw better improvements than others. The state of Georgia saw its graduation rate rise by 10% to reach a high of 80% (Georgia Department of Education, 2017). This rise in Georgia graduation rates prompted us to look further into the elements that may have influenced that number. As outlined in the UN's sustainable development goal on education, the quality of education received can greatly alter a child's life down the line. The goal of this paper is to look at some of the characteristics that may influence the quality of education as determined by the graduation rate and explore their effect on that metric.

The United Nation's goal on education is to ensure inclusive and quality education for all. Even within the US, current high school graduation rates greatly vary between states, with Florida experiencing the lowest rate at 65.8% while Nebraska has the highest rate at 93.8% (U.S. News & World Report, 2017). By understanding the factors that have the strongest correlation with higher graduation rates, more policies can be set in place to ensure that no child has an unequal opportunity at quality education based on their location. This aligns with the United Nation's goals to achieve their target of quality education by ensuring that all children complete quality primary and secondary education.

We believe that better economic conditions in a given Georgia county will be indicative of higher quality education in that county. In counties where household income is low, the amount of income tax revenue available to fund public education will be reduced and educational quality will be lower. In counties where household incomes are high, funding of the amount of income tax revenue available to fund public education will be greater and educational quality will be higher. The economic conditions in Georgia counties will be measured by several variables with a focus on household income. Specifically, we hypothesize that household income will be positively correlated with four-year high school graduation rates for public high schools in Georgia counties. In addition to household income, other economic and social factors were examined. Simple regression and multiple regression analyses were performed to test the validity of this hypothesis.

2. Literature Review

In order to better understand the relationships between household income, spending per student and teacher salary on graduation rates, research was done to review some of the literature that already exists on the subject. By looking at previous research, we hoped to learn more about the possible variables that could influence graduation rates. Several studies investigating the relationship between income status of the students and their graduation rates were found. Another study examined the effect of higher salaries on scores in math and readings, and the last study looks at the effect that increases in spending on students has on students of high and low economic status.

Stark, Noel and McFarland (2015) examines various characteristics of students who never graduated high school including the income of their households in 2012. The Department of Education found that dropout rates of students living in low income households (5.9%) were much greater than those of students living in medium to high income households (1.3%). Low incomes were classified as families earning in the lowest 20% of all family incomes, medium incomes were classified as families earning between 20 and 80% of all family incomes, and high income families as earning in the top 20% of all family incomes. The compendium also states that all three income categories show a decline in high school dropout rates since the mid 1970s. The figure in **Appendix A**, generated by the authors of the paper, shows a general decline in dropout rates but highlights the remaining gap between dropout rates of students of high income families and those of low income families. This study allows us to conclude that the coefficient of the income variable in our regression model should be positive: as income increases, the dependent variable, graduation rates, should rise.

Kearney and Levine (2016) also look into income as having effects on high school graduation rates. This report looks at how greater income levels could lead to a perpetual economic disadvantage, and affect perceived returns on investment in education from the perspective of low income students. They look at areas of high income inequality and evaluate the effect of this inequality on rates of graduation among students from low income households.

This report finds that individuals coming from less fortunate backgrounds are less likely to graduate if they live in an area with a large income inequality gap. This research differs from the paper mentioned above in that it looks at income inequality rather than income itself. That is, the gap between low and high incomes of the families of the students, rather than the incomes of the families themselves. This adds yet another dimension to this problem.

Darling-Hammond (2000) looks at the ways in which teacher salaries as influenced by state investments improved student education. The level of education was analyzed through reading and mathematics scores. This report discusses substantial investments in teaching in the states of North Carolina and Connecticut, two states with relatively high impoverished student populations. These states' investments in teaching increased teacher salaries which resulted in higher reading and mathematics scores for both states. This creates a positive relationship between teacher salaries and reading and mathematics scores, which we hypothesize are also positively correlated with graduation rates. What is interesting about this study is that it looks at two states in greater detail as opposed to looking at income classes, something this paper intends to do with the state of Georgia.

Finally, Jackson, Johnson and Persico (2014), a paper from the National Bureau of Economic Research analyzes the effect of school finance reforms on spending distribution in schools. This report analyzes the structural changes that occurred in education spending in the U.S. as a result of large school finance reforms from the 1970s and 1980s. They specifically look at children born between 1955 and 1985 and analyze long term outcomes. They found that a 20% increase in spending per pupil every year for all years of K-12 education led to 0.9 extra completed years of education. These effects were exclusive to lower income families, as more well-off students did not see an increase in graduation rates. This implies that spending per student has a positive correlation with graduation rates for lower income families, but relatively no correlation for graduation rates of higher income families.

The goal of this paper is to confirm the finding that higher household incomes lead to higher four-year graduation rates by looking exclusively at the state of Georgia. By exclusively looking at Georgia, this report hopes to obtain results that are specific to Georgia since states have wide discrepancies between spending per student and teacher salary. This allows the state to tailor its public policies toward variables relevant to improved graduation rates specifically for the state. In addition, this paper aims to look at graduation rates using data from various counties, something that was not done in the above mentioned research. This is an attempt at linking several economic and social variables as having a combined significant impact on graduation rates, as opposed to looking at variables separately.

3. Data

The data used in this analysis was obtained from a variety of government sources including the U.S. Bureau of Labor Statistics, the U.S. Census Bureau, Open Georgia, the Georgia Department of Education, the Georgia Department of Revenue and the Georgia Department of Health. Data was collected for public high schools in all 159 Georgia counties. Our analysis is impacted by missing data for nine counties resulting in a reduction of the sample size for the simple and multiple regression models. This may impact the quality of our models.

Table 1. Variables and Data Sources Used in Regression Model

Variable	*Classification*	*Abbreviation*	*Year(s)*	*Source*
High School Graduation Rate	Dependent	*gradrate*	2015	Georgia Department of Education
Avg. Household Income	Independent	*houseinc*	2015	U.S. Census Bureau
Avg. Teacher Salary	Independent	*teachsal*	2010-2015*	Open Georgia
Avg. Spending Per Pupil	Independent	*spp*	2015	Georgia Department of Education
Local Education Revenue	Independent	*localrev*	2015	Georgia Department of Education
M&O School District Tax Rate	Independent	*motax*	2015	Georgia Department of Revenue
Poverty Rate	Independent	*povrate*	2009-2013*	U.S. Census Bureau
Gini Index	Independent	*gini*	2009-2013*	U.S. Census Bureau
Unemployment Rate	Independent	*unemployment*	2015	U.S. Bureau of Labor Statistics
Avg. Class Size	Independent	*classsize*	2016	Georgia Department of Education
Educational Attainment (High School)	Independent	*eahs*	2015	U.S. Census Bureau
Teen Pregnancy Rate	Independent	*teenpreg*	1997	Georgia Department of Public Health

* = averaged across the years indicated

To control for other factors that may be affecting high school graduation rates, additional economic and social variables were considered . The first eight variables in **Table 1** were added as economic variables. The last four variables in the table were added as social variables. By adding the additional independent variables shown into the overall analysis, more of the variation in high school graduation rates can be explained. Average teacher salary is a measure of the average salaries of grade 9 - 12 public school teachers. This variable was chosen because higher teacher pay will attract higher quality graduates into the teaching profession and improve student outcomes (Dolton & Marcenaro-Gutierrez, 2011). Average spending per pupil is a measure of the average amount of money spent on each student. It was included in the regression because a significant relationship has been found between increased per-pupil expenditure and improved student achievement (Hedges & Greenwald, 1996). Local education revenue is the percentage of education revenue that is funded by the local government. This variable was chosen based on the belief that the extent to which a school district relies on local funding, as opposed to state or federal funding, may impact its ability to provide quality education and thus its graduation rate. Gini index is a measure of income inequality based on an index from 0 to 100, with 0 denoting perfect income equality and 100 denoting perfect income inequality. This variable was chosen because greater income inequality is associated with lower high school graduation rates (Kearny & Levine, 2016). Unemployment rate is a measure of the percentage of the labor force which is unemployed. This variable was included because it’s believed that areas of high unemployment likely have a large number of high school dropouts. The M&O tax is the maintenance and operations tax demanded of each household by the county. **Appendix B** gives more detailed description and justification for all variables included in the regression. Holding all other factors constant, it is expected that poverty rate, Gini index, unemployment rate, average class size and teen pregnancy rate will be inversely related to high school graduation rates, while median household income, average teacher salary, average spending per pupil, M&O tax rate, local revenue and educational attainment are expected to be directly related to high school graduation rates.

3.1 Descriptive Statistics

Table 2. Descriptive Statistics Table Excluding Dummy Variables

Variable	***Obs***	***Mean***	***Std. Dev.***	***Min***	***Max***
gradrate	158	83.60	7.26	34.90	97.30
houseinc	159	42,510.31	11,747.57	25,941.00	97,866.00
avgsal	157	46,854.68	5,017.73	28,649.83	56,237.76
spp	158	336.25	123.09	114.27	1027.89
localrev	159	34.96	11.21	15.72	73.81
motax	159	16.49	2.46	7.96	25.32
povrate	159	22.06	6.76	7.60	44.50
gini	159	0.45	0.04	0.36	0.55
unemployment	159	6.40	1.29	4.30	10.80
classsize	159	12.43	2.61	4.30	20.10
eahs	159	80.28	5.98	61.00	93.90
teenpreg	153	62.86	22.32	18.20	152.50

Descriptive statistics for each of the variables included in the regression are included in **Table 2** above. Nearly all of Georgia's 159 counties have data for each variable. Only nine counties - about 0.56% of our total data - will have to be excluded for incomplete information. The regressand, *gradrate*, shows wide variation between its minimum value of 34.90% and its maximum value of 97.30%. This variation further justifies that public high school graduation rates in Georgia are worth studying because there is substantial variation in graduation rates across the state. *houseinc* as well as the other regressors being study also show notable range. This is important because it indicates that the independent variables being studied do vary from county to county and thus may be useful in explaining variation in the regressand.

Table 3. Descriptive Statistics Table Including Dummy Variables

Variable	*Obs*	*Mean*	*Std. Dev.*	*Min*	*Max*
gradrate	61	84.04	5.71	70.90	97.20
houseinc	61	47,372.20	17,163.88	28,328.00	97,866.00
avgsal	61	47,120.63	4,798.70	29,941.75	56,237.76
spp	61	314.90	100.64	140.48	737.12
localrev	61	36.23	10.11	15.72	57.63
motax	61	17.08	2.38	12.50	25.32
povrate	61	21.24	9.06	7.60	38.70
gini	61	0.45	0.04	0.38	0.55
unemployment	61	6.28	1.50	4.30	10.60
classsize	61	12.98	2.84	6.10	20.10
eahs	61	81.31	7.53	61.00	93.90
teenpreg	61	64.51	26.24	18.90	152.50
lowinc	61	0.49	0.50	0	1

To better explain variation in public high school graduation rates, the data was split into two income brackets based on average household income. The division of data was carried out in the same fashion as Stark et al., 2015. Using the standards set forth is that paper, counties with average household incomes among the top 20% - $51,081.00 per year and above - were considered to be high-income and counties with average household incomes among the bottom 20% - $34,139 per year and below - were considered to be low-income. A sub-dataset containing data from the high and low-income counties only was generated and the descriptive statistics for this dataset are shown in **Table 3** above. There are 31 high-income counties and 30 low-income counties in the data set (See **Appendix C**).The dummy variable *lowinc* is a binary variable which is given a value of 1 to denote a low-income county or a value of 0 to denote a high-income county. As shown in **Table 3**, when considering only the income extremes of the state, substantial variation in the regressand and all regressors being studied is maintained.

3.2 Gauss-Markov Assumptions

Linear in Parameters

The model is linear in parameters such that the following is true:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Random Sampling

Data for all variables was collected from nearly all of Georgia's 159 counties from state and federal government data sources. All counties that had data available were considered for the analyses conducted, satisfying the condition of random sampling.

No Perfect Collinearity

To assess collinearity between regressors, correlations between all explanatory variables were computed using STATA. The correlation coefficient computations, shown in **Appendix D** reveal that some correlation does exist between regressors in the model. The correlations exist at levels less than 1.0, thus the no perfect collinearity assumption is satisfied.

Zero Conditional Mean

In the simple linear regression model, the expected value of the error term, *u*, is nonzero thus the zero conditional mean assumption is not satisfied. To address this issue, several additional explanatory variables were added to generate multiple linear regression models. In the multiple linear regression models, the expected value of the error term is zero, thus for those models, the zero conditional mean assumption is satisfied.

Homoscedasticity

In the simple linear regression model, the variance in the error term is nonzero, violating the homoscedasticity assumption. Additional explanatory variables were added to generate multiple linear regression models and in these multiple linear regression models, the variance in the error term is zero. The multiple linear regression models satisfy the homoscedasticity assumption.

3.3 Scatter Plot: Graduation Rate vs. ln(*houseinc*)

A scatter plot of *gradrate* vs. *lhouseinc* is shown in **Appendix E**. The graph shows a slightly linear correlation between the two variables with outliers above and below the regression line at all values of *lhouseinc*. The distribution of data points is very loose at lower values of *lhouseinc* and becomes a bit tighter around the regression line as *lhouseinc* values rise. This indicates that *lhousinc* may better explain variation in high school graduation rates for high-income counties than for low-income counties.

3.4 Functional Form

In addition to the above scatter plot, a scatter plot of the simple linear regression *gradrate* vs. $lhouseinc^2$ was also generated in hopes of capturing the data points that could not be explained linearly as with the original model (**Appendix F**). Using the quadratic functional form the main variable, *lhousinc* and $lhouseinc^2$ are no longer significant. For this reason, the original linear model discussed above was used throughout the paper.

4. Results

A regression was run on all counties in Georgia. The first model is a simple linear regression of high school graduation rates versus ln (average household income).

Model 1: *gradrate = 16.80 + 6.29 (lhouseinc)* + u

In **Table 4** below, Model 1 shows a positive relationship between household income and high school graduation rates as a 1% increase in household income results in roughly 6.29 unit increase in graduation rate which is consistent with our hypothesis. This relationship is shown to be significant at the 5% level and explains merely 4% of the variation in high school graduation rates amongst all counties which implies that there are other variables that may have a more significant impact on graduation rates in Georgia.

Table 4: Model for All Counties

Independent Variables	*Model 1*	*Model 2*	*Model 3*	*Model 4*
lhouseinc	6.29*** (2.37)	2.79 (3.01)		-0.85 (5.18)
lavgsal		22.42*** (4.93)		21.92*** (5.13)
lspp		3.65** (1.62)		2.85 (1.79)
localrev		0.17*** (0.05)		0.18*** (0.05)
motax		-0.25 (0.23)		-0.33 (0.24)
povrate		0.20 (0.17)		0.12 (0.18)
gini		-41.83** (19.33)		-47.58** (20.46)
unemployment			-1.44*** (0.57)	-0.57 (0.57)
classsize			0.08 (0.24)	0.12 (0.23)
eahs			0.09 (0.10)	0.05 (0.09)
teenpreg			-0.01 (0.03)	0.02 (0.03)
Intercept	16.80 (25.18)	-194.92 (69.01)	84.89 (8.75)	-144.25 (78.13)
No. of Observations	158	156	152	150
R^2	0.04	0.25	0.07	0.27

Significance: 10%*, 5%**, 1%*** ( ) = standard error

The second model generated is a multiple linear regression model using economic regressors. In this model, high school graduation rate is regressed against ln(*houseinc*), ln(*avgsal*), ln(*spp*), *localrev*, *motax*, *povrate* and *gini*.

Model 2: *gradrate* = -194.92 +2.79 *(lhouseinc)* + 22.42 *(lavgsal)* + 3.65 *(lspp)* + 0.17 *(localrev)* -0.25 *(motax)* + *0.20 (povrate)* - 41.83 *(gini)* + u

Moving from Model 1 to our economic model, Model 2, the positive relationship between *lhouseinc* and *gradrate* is maintained, but the influence of *lhouseinc* in the model is diminished as a 1% increase in *lhouseinc* only results in roughly 2.79 unit increase in *gradrate.*

gini is shown to be the strongest predictor of *gradrate* among the economic variables as a one unit decrease in *gini* results in a 41.83 unit increase in *gradrate*. This is inconsistent with our hypothesis that household income would be the strongest predictor of high school graduation rates. Also surprising is the negative correlation shown between M&O tax and high school graduation rates, and the positive correlation between poverty rates and graduation rates. However both M&O tax and poverty rates are not significant and the coefficient can be considered zero. The remaining economic variables are all significant at 5%. A 1% increase in *lavgsal* increases the *gradrate* by 22.42 units , a 1% increase in *lspp* increases *gradrate* by 3.65 units and a 1 unit increase in *localrev* increases *gradrate* by 0.17 units. Thus the three variables only have a slight impact on graduation rates. This model explains 25% of the variation in graduation rates.

The third model is another multiple regression model, but this time using social regressors. In this model, *gradrate* is regressed against *unemployment*, *classsize*, *eahs*, and *teenpreg*.

Model 3: *gradrate* = 84.89 - 1.44 *(unemployment)* + 0.08 *(classsize)* + 0.09 *(eahs)* - 0.01 *(teenpreg)* + u

Model 3 shows *unemployment* to be the strongest social variable predictor of high school graduation rates as 1 unit decrease in *unemployment* increases *gradrate* by 1.44 units. The remaining variables are all insignificant even at the 10% level thus even though it may seem surprising that *classsize* has a positive correlation with *gradrate*, the coefficient can be considered 0. This also applies to *teenpreg* and *eahs*. This model only explains 7% of the graduation rates in Georgia, which implies that the chosen social variables have a smaller impact on graduation rates than economic variables.

The final model is a multiple regression model combining the economic and social regressors used in the previous two models. In this model, *gradrate* is regressed against ln (*houseinc*), ln(*avgsal*), ln(*spp*), *localrev*, *motax*, *povrate*, *gini* , *unemployment*, *classsize*, *eahs*, and *teenpreg*.

Model 4: *gradrate* = -144.25 - 0.85 *(lhouseinc)* + 21.92 *(lavgsal)* + 2.85 *(lspp)* + 0.18 *(localrev)* - 0.33 *(motax)* - 47.58 *(gini)* - 0.57 *(unemployment)* + 0.12 *(classsize)* + 0.05 *(eahs)* + 0.02 *(teenpreg)* + 0.12 *(povrate)* + u

In the combined model, Model 4, the positive relationship between *lhouseinc* and *gradrate* is no longer retained. However *lhouseinc* is now insignificant which can explain change in sign. Similarly to Model 2, *gini*, rather than *lhouseinc*, is shown to be the strongest predictor of *gradrates* as a unit decrease in *gini* increases *gradrate* by 47.58 units. Similarly to model 3, the social variables chosen and even *unemployment,* which was originally significant, are now insignificant. Thus the social variables do not explain the impact on graduation rates. The regression shows that only three variables – *lavgsal*, *localrev* and *gini* – are significant. A 1% increase in *lavgsal* will result in a 22.92 unit increase in *gradrate* while a 1 unit increase in *localrev* will result in a 0.18 unit increase in *gradrate*. This shows that *localrev* has a very slight impact on *gradrate*. This model explains 27% of the graduation rates in Georgia which is only slightly higher that Model 2.

5. Extensions

5.1 F-tests

From the table, we note that adding the social variables to Model 4 only has a 0.02 increase on R^2 as compared to Model 2 which only has economic variables. Thus conducting an F-test between these two models will help verify whether the social variables have a joint significance on Model 4. To result in a more accurate test, six counties were excluded from Model 4, so that both models would have the same observation numbers. The R^2 value for the adjusted Model 4 was still 0.27 despite 6 counties being excluded. The calculated F-value is

$$\frac{0.27 - 0.25}{1 - 0.27} * \frac{138}{4} = 2.3$$

The value for significance at the 5% level of $F_{4,138}$ is 2.37. Since 2.37 > 2.3, the social variables are not jointly significant at the 5% level. The value for significance at the 10% level of $F_{4,138}$ is 1.94. Since 2.3 > 1.94, the social variables are significant at the 10% level.

The social variables are only jointly significant at the 10% level. Thus, we can determine that Model 2 is our best model and that chosen social variables have a more limited impact on graduation rates as compared to chosen economic variables

5.4 Dummy Variables

As explained in the data section, a dummy variable (*lowinc*) was created for household incomes (**Appendix G**). This variable was given the value 1 if *houseinc* is in the bottom 20% of household incomes, and 0 if it is in the top 20%.

$$gradrate = 85.5 - 3.2lowinc$$

The average graduation rate for high income students is 85.5. The coefficient on *lowinc* states that the difference in graduation rate between low income and high income students is 3.2. The average graduation rate for low income student is 85.5 - 3.2 = 82.3. In this model, *lowinc* is significant at the 5% level. Thus, this confirms our claim that low income students do have, on average, a lower graduation rate than high income students and that reducing this income gap should increase graduation rates.

Adding the rest of our variables from the chosen Model 2, we obtain the following regression:

$$gradrate = \beta_0 + \gamma lowinc + \beta_1 lavgsal + \beta_2 lspp + \beta_3 localrev + \beta_4 motax + \beta_5 povrate + \beta_6 gini$$

In this model we see that *lowinc* is no longer statistically significant (**Appendix H**). This implies that there may be evidence of multicollinearity between *lowinc* and the other variables. *lowinc* and *povrate* have a very high correlation of 0.91, which causes *lowinc* to lose significance once povrate is added to the regression. Removing *lavgsal*, *localrev*, *povrate* from the model (**Appendix I**) results in *lowinc* remaining statistically significant at 5%. However the coefficient has decreased to -5.19 implying that the added economic factors have in fact worsened the gap in graduation rates. However we note that none of the variables other than *lowinc* are significant.

6. Conclusion

This paper sought to understand the factors that influence high school graduation rates with the hypothesis that household income would be positively correlated with high school graduation rates. Analysis of the full data set showed that household income is significant in simple linear regression with high school graduation rate, but loses significance when additional economic and social explanatory variables are added to the regression. Other variables such as gini index and average teacher salary had a much stronger correlation and significance with high school graduation rates, but their impact in explaining variation in high school graduation rates is limited by extremely small regression coefficients. Supported by our analysis of graduation rates among counties at all income levels, we can report that our hypothesis - household income is positively correlated with public high school graduation rates among Georgia counties- cannot be not be proven true for the full data set.

Analysis of the sub-dataset including only high and low-income counties, shows that students with low income, on average, have lower graduation rates than high-income students, but this gap is relatively small. Moreover, as explanatory variables included in our chosen regression model, Model 2 (**Table 4)**, were added to this regression, the gap between graduation rates of high and low-income counties is reduced and *lowinc* becomes insignificant. Though upon removal of variables contributing to high multicollinearity, *lowinc* becomes significant again. Supported by our analysis of graduation rates at the income extremes, we can report that our hypothesis is supported.

It is also worth noting that our models explain only a small variation in graduation rates, implying that there are multiple other factors outside of those discussed that could have a significant impact on graduation rates. This suggests more research must be done to understand what those other factors are. Economic policy targeted at reducing gini index or increasing teacher salaries may have small impacts on graduation rates, but to see significant improvement will require public policies which target factors outside of those studied in this paper that may be contributing substantially to variation in high school graduation rates.

References

Office of Press Secretary. (2016) President Obama Announces High School Graduation Rates Has Reached New High. Retrieved November 29, 2017, from https://obamawhitehouse.archives.gov/the-press-office/2016/10/17/fact-sheet-president-obama-announces-high-school-graduation-rate-has

Georgia Department of Education. (2017) Georgia's graduation rate increases for sixth straight year, tops 80 percent. Retrieved November 29, 2017, from http://www.gadoe.org/External-Affairs-and-Policy/communications/Pages/PressReleaseDetails.aspx?PressView=default&pid=567

Boyington, B (2017) See High School Graduation by States. Retrieved November 29 2017, from https://www.usnews.com/high-schools/best-high-schools/articles/2017-05-17/map-see-high-school-graduation-rates-by-state

Stark, P., Noel, A. M., & McFarland, J. (2015). Trends in High School Dropout and Completion Rates in the United States: 1972–2012 - Compendium Report. *National Center for Education Statistics*. Retrieved October 17, 2017, from https://nces.ed.gov/pubs2015/2015015.pdf.

Kearney, M. S., & Levine, P. B. (Spring 2016). Income Inequality, Social Mobility, and the Decision to Drop Out of High School. *Brookings Papers on Economic Activity*. Retrieved October 17, 017, from https://www.brookings.edu/wp-content/uploads/2016/03/kearneytextspring16bpea.pdf.

Darling-Hammond, L. (2000). Teacher Quality and Student Achievement: A Review of State Policy Evidence. *Education Policy Analysis Archives*. Retrieved October 17, 2017, from http://epaa.asu.edu/ojs/article/view/392

Jackson, K. C., Johnson, R., & Persico, C. (2014). THE EFFECT OF SCHOOL FINANCE REFORMS ON THE DISTRIBUTION OF SPENDING, ACADEMIC ACHIEVEMENT, AND ADULT OUTCOMES. *National Bureau of Economic Research*. Retrieved October 17, 2017, from http://www.nber.org/papers/w20118.pdf

Dolton, P., & Marcenaro-Gutierrez, O. (2011). Teachers' pay and pupil performance. *CentrePiece*. Retrieved November 27, 2017, from http://cep.lse.ac.uk/pubs/download/cp352.pdf

Hedges, L., Greenwald, R. (1996). "Have Times Changed? The Relation Between School Resources and Student Performance," in Gary Burless, ed., Does Money Matter? The Effect of School Resources on Student Achievement and Adult Success.(*Washington , D.C.: Brookings Institution Press)*, pp.74-92. Retrieved November 27, 2017.

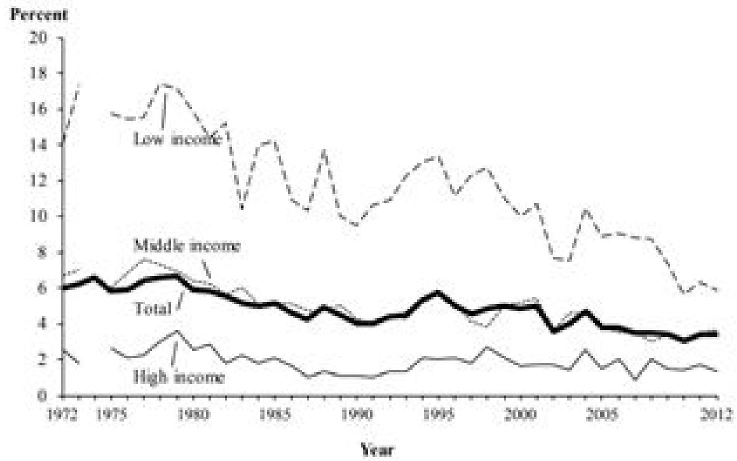
Davis, E., Ruthotto, I. (2015) Financing Georgia's Schools: A 2015 Briefing. In (pp. 17-19). Atlanta, GA: Georgia State University

Rumberger, R. W. (2013). Poverty and High School Dropouts: The Impact of Family and Community Poverty on High School Dropouts. *The SES Indicator,6*(2). Retrieved November 28, 2017, from http://www.apa.org/pi/ses/resources/indicator/2013/05/poverty-dropouts.aspx

Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small Classes in the Early Grades, Academic Achievement, and Graduating From High School. *Journal of Educational Psychology,97*(2), 214-223. doi:10.1037/0022-0663.97.2.214

Appendices

Appendix A. Percentage of high school dropouts among low, middle and high-income households between 1972 and 2012



Appendix B. Description and justification for explanatory variables included in regression

Variable	*Description*	*Justification for Inclusion*
Avg. Teacher Salary	Average salaries of grade 9 through 12 teachers	High teacher pay will attract higher quality graduates into the profession and improve student outcomes (Dolton & Marcenaro-Gutierrez, 2011)
Avg. Spending Per Pupil	Average amount of money the county spends on each student	Direct relationship found between per-pupil expenditure and student achievement (Hedges & Greenwald, 1996)
Local Education Revenue	Percentage of education revenue funded by local government	The extent to which a school district relies on local funding as opposed to state or federal funding may relate to four-year graduation rates
M&O School District Tax	Maintenance and operations tax demanded of each household by the county; These are an element of property tax which is the primary sources of funding for Georgia school systems. (Davis & Ruthotto, 2015)	Higher tax rates should generate greater education revenue and lead to more positive student outcomes.
Poverty Rate	Percentage of people living below the poverty line	Instability caused by poverty negatively impacts high school graduation rates (Rumberger, 2013)
Gini Index	Measure of income inequality based on	Greater income inequality is associated

	an index from 0 to 100.	with lower high school graduation rates (Kearney & Levine, 2016)
Unemployment Rate	Percentage of labor force which is not employed	Areas with higher unemployment rates may have more high school dropouts
Avg. Class Size	Average number of students per class in grade 9 through 12 classes	Smaller class sizes have direct correlation with higher graduation rates (Finn et al., 2005)
Educational Attainment	Percentage of adults aged 25 years and older who have completed at least a high school education	Having more high school graduates in the general population should make high school graduation more of an expectation for students in the area
Teen Pregnancy Rate	Pregnancy incidence among women aged 15 to 19 per 1,000 people in 1997. This the approximate time frame mothers having children graduating in 2015 would have been pregnant.	Children born to younger mothers may be less likely to graduate from high school

Appendix C. High-income and low-income counties included in sub-dataset

High-income		***Low-income***	
Barrow County	Harris County	Atkinson County	Randolph County
Bryan County	Henry County	Ben Hill County	Seminole County
Camden County	Houston County	Calhoun County	Stewart County
Cherokee County	Jackson County	Clarke County	Sumter County
Cobb County	Jones County	Clinch County	Talbot County
Columbia County	Lee County	Crisp County	Taylor Count
Coweta County	Monroe County	Dooly County	Telfair County
Dawson County	Morgan County	Early County	Terrell County
DeKalb County	Newton County	Emanuel County	Treutlen County
Douglas County	Oconee County	Evans county	Warren County
Effingham County	Paulding County	Hancock County	Wilcox County
Fayette County	Pickens County	Jefferson County	Wilkes County
Forsyth County	Pike County	Jenkins County	

Fulton County	Rockdale County	Johnson County	
Gwinnett County	Walton County	Macon County	
Hall County		Mitchell County	

Appendix D. Test for no perfect collinearity among variables, STATA outputs

Full Dataset

	houseinc	avgsal	spp	localrev	motax	povrate
houseinc	1.0000					
avgsal	0.3705	1.0000				
spp	-0.1101	-0.1288	1.0000			
localrev	0.2910	0.1215	0.0693	1.0000		
motax	0.3488	0.0024	-0.1074	0.0769	1.0000	
povrate	-0.8168	-0.3492	-0.0229	-0.1727	-0.2747	1.0000
unemployment	-0.6186	-0.3352	0.0289	-0.2300	-0.1791	0.5694
eahs	0.7128	0.3703	-0.0649	0.2786	0.2441	-0.6438
classsize	0.3223	0.0678	0.0201	0.0008	0.2358	-0.2391
teenpreg	-0.4349	-0.1487	-0.0607	-0.0734	-0.0451	0.4978
gini	-0.4828	-0.2094	0.0392	0.1823	-0.2092	0.6623

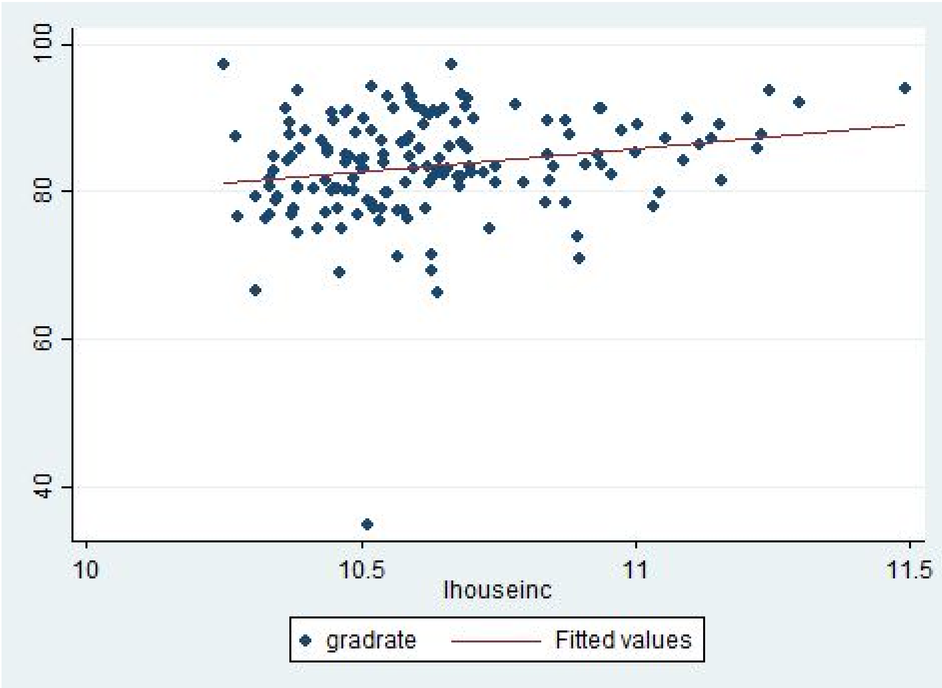
	unempl~t	eahs	classs~e	teenpreg	gini
unemployment	1.0000				
eahs	-0.4138	1.0000			
classsize	-0.1775	0.1949	1.0000		
teenpreg	0.2958	-0.4665	0.0175	1.0000	
gini	0.3110	-0.3089	-0.1865	0.3769	1.0000

Sub-Dataset with Dummy Variable

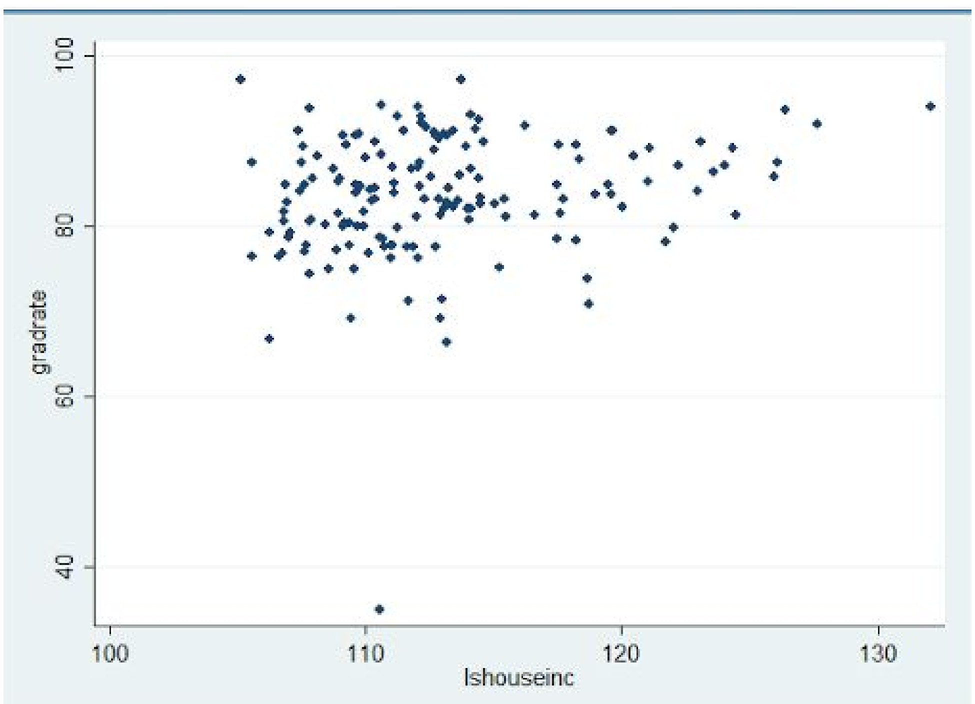
	houseinc	avgsal	spp	localrev	motax	gini	unempl~t	classs~e
houseinc	1.0000							
avgsal	0.5787	1.0000						
spp	-0.0518	0.0073	1.0000					
localrev	0.3819	0.2865	0.2592	1.0000				
motax	0.4097	0.2381	-0.0216	0.3247	1.0000			
gini	-0.6373	-0.4043	0.0281	0.0537	-0.2360	1.0000		
unemployment	-0.7460	-0.3520	0.0185	-0.3168	-0.2926	0.4091	1.0000	
classsize	0.4367	0.1633	-0.1032	0.2794	0.3782	-0.2629	-0.4133	1.0000
eahs	0.8624	0.6219	0.0478	0.3219	0.4506	-0.5358	-0.6353	0.4119
teenpreg	-0.6411	-0.4059	-0.0059	-0.1584	-0.2862	0.4977	0.3733	-0.0552
lowinc	-0.8979	-0.5593	0.0405	-0.3494	-0.4629	0.6774	0.7505	-0.4482

	eahs	teenpreg	lowinc
eahs	1.0000		
teenpreg	-0.6524	1.0000	
lowinc	-0.8356	0.6278	1.0000

Appendix E. Scatter plot, linear model: *gradrate* vs. *ln(houseinc)*



Appendix F. Scatter plot, quadratic model *gradrate* vs. *ln(houseinc)*^{2}



Appendix G: Stata output with dummy variable inclusion

```
. generate lowinc = .
(999 missing values generated)

. replace lowinc = 0 if lhouseinc >=10.84
(871 real changes made)

. replace lowinc = 1 if lhouseinc <= 10.44
(34 real changes made)

. regress gradrate lowinc

      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(1, 62)        =      4.84
       Model |  164.611321         1  164.611321   Prob > F        =    0.0316
    Residual |  2109.00727        62  34.0162463   R-squared       =    0.0724
-------------+----------------------------------   Adj R-squared   =    0.0574
       Total |  2273.61859        63   36.089184   Root MSE        =    5.8323

------------------------------------------------------------------------------
    gradrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lowinc |  -3.209091   1.458799    -2.20   0.032    -6.125188   -.2929939
       _cons |       85.5    1.04752    81.62   0.000     83.40604    87.59396
------------------------------------------------------------------------------
```

Appendix H: Multiple regression model with dummy variable

```
. regress gradrate lowinc lavgsal lspp localrev motax gini povrate

      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(7, 56)        =      3.13
       Model |  640.045836         7  91.4351194   Prob > F        =    0.0073
    Residual |  1633.57276        56  29.1709421   R-squared       =    0.2815
-------------+----------------------------------   Adj R-squared   =    0.1917
       Total |  2273.61859        63   36.089184   Root MSE        =     5.401

------------------------------------------------------------------------------
    gradrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lowinc |   -2.22953   4.061608    -0.55   0.585    -10.36591    5.906848
     lavgsal |   20.21691   7.061703     2.86   0.006     6.070622     34.3632
        lspp |   .6503652   2.335914     0.28   0.782    -4.029034    5.329764
    localrev |   .1693533   .0854183     1.98   0.052    -.0017601    .3404668
       motax |  -.6931718   .3456031    -2.01   0.050    -1.385498   -.0008456
        gini |  -14.17794   25.28555    -0.56   0.577    -64.83099     36.4751
     povrate |   .1168665   .2346881     0.50   0.620    -.3532703    .5870034
       _cons |   -126.462   76.92785    -1.64   0.106     -280.567    27.64298
```

Appendix I Restricted multiple regression model with dummy variable

```
. regress gradrate lowinc lspp motax gini

      Source |       SS           df       MS      Number of obs   =        64
-------------+----------------------------------   F(4, 59)        =      2.21
       Model |  295.759499         4  73.9398749   Prob > F        =    0.0793
    Residual |  1977.85909        59  33.5230355   R-squared       =    0.1301
-------------+----------------------------------   Adj R-squared   =    0.0711
       Total |  2273.61859        63   36.089184   Root MSE        =    5.7899

------------------------------------------------------------------------------
    gradrate |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lowinc |  -5.190302   2.074821    -2.50   0.015    -9.342009   -1.038594
        lspp |   2.389925   2.382623     1.00   0.320    -2.377692    7.157542
       motax |  -.5527364   .3341591    -1.65   0.103    -1.221387    .1159145
        gini |   12.03637   22.66201     0.53   0.597    -33.31021    57.38295
       _cons |   76.79289   17.72729     4.33   0.000     41.32067    112.2651
------------------------------------------------------------------------------
```