

Introduction and Hypothesis

When searching up news articles related to “schools in Philadelphia” one will find articles related to increased crimes, un-fair employee compensation, or teachers leaving their profession.

As I was reading these articles, I began to wonder if any of these factors have a negative (or positive) impact on a child’s education. Would student attendance rate decrease if there was an increase in certain school crimes? Are some races more likely to have a decrease in attendance rate as compared to other races? In this report I am going to investigate the relationship between a student’s attendance rate and a group of independent variables and see if there is some type of correlation between any of the variables.

I hypothesize that students from minority groups (such as African Americans, Latinos, and Asians) are more likely to have a decrease in attendance rate as compared to other races (such as White individuals). I also expect to see a decrease in school attendance rate as the increase in low-income family attendance increases.

NOTE: When I mention that an increase in a certain race decreases/increases attendance rates I am referring to the racism or negative experience that the particular group might face rather than saying a person’s skin color affects attendance.

Mapping and Introduction Page

I have also produced maps in the “**Introduction – Maps**” tab which visualizes the Philadelphia school data by location. No direct conclusions should be made by just looking at the map since some variables were categorized when displayed.

Description of Main Variables

For my analysis I am using the data collected from various schools in Philadelphia, with a given zip code. The data includes details such as the percentage of each race enrolled in the school, number of crimes, etc. The output/dependent variable that I will be measuring is the “Attendance Rates” variable. This variable measures the average percentage of student attendance with-in the school.

Table 1: Summary Statistic of Variables					
Statistic	N	Mean	St. Dev.	Min	Max
Attendance Rates	209	92.710	3.770	77.940	97.480
African American Students	209	59.635	32.198	1.140	98.720
White Students	209	13.994	19.906	0.080	89.700
Asian Students	209	6.191	9.887	0.000	57.960
Latino Students	209	16.865	21.909	0.160	92.040
Students of other races	209	3.283	2.528	0.100	15.120
Number of Total Suspensions	209	75.266	101.201	0.000	667.000
Average Teacher Salary	209	68,949.580	4,419.240	46,694.000	78,322.400
Drug-related crimes	209	1.108	2.504	0.000	15.750
Moral misconduct-related crimes	209	0.537	0.648	0.000	4.000
Assault-related crimes	209	9.543	11.262	0.250	87.250
Weapon-related crimes	209	1.982	2.275	0.000	14.000
Theft-related crimes	209	0.783	1.290	0.000	9.250
Percentage of Low-income families attending school	209	80.489	13.104	45.440	94.960

Table 1: A summary statistic of the variables used in this analysis.
The variables highlighted in yellow are the main independent variables that I will be measuring.

My website allows users to see a summary statistic for each variable on the “**Description of Main Variables Tab**”. The user gets to choose from the drop-down menu which variable they would like to see a summary statistic of. I created this variable so a user can get the opportunity to get a better understanding about the variables that they would like to look at.

Regression Analysis

In my website I have also included a tab called “**Regression**” which allows users to select their chosen variables and see how they all affect one another.

One interesting thing I noticed in my regression output is that white students had a greater fall in Attendance rates as compared to students from other races. While the coefficient drop for all students was huge (-9). Students from white ethnic backgrounds had a negative coefficient of -**0.954**, Asian students had a coefficient of -**0.941** and African American students had a coefficient of -**0.926**

		Dependent variable:
		Attendance
1 unit change in African American students increases/decreases Attendance by:	-0.926	
	(4.018)	
1 unit change in White students increases/decreases Attendance by:	-0.954	
	(4.018)	
1 unit change in Asian students increases/decreases Attendance by:	-0.923	
	(4.020)	
1 unit change in Latino students increases/decreases Attendance by:	-0.941	
	(4.018)	
1 unit change in students of other races increases/decreases Attendance by:	-0.342	
	(4.030)	
1 unit change in low-income family percentage increases/decreases Attendance by:	-0.083	
	(0.025)	
Constant	190.661	
	(401.574)	
Observations	209	
R ²	0.222	
Note:	*p<0.1; **p<0.05; ***p<0.01	

Table 2

I was surprised to see that a one unit change in White students would cause the highest drop in attendance rate. I believe that the “low-income families” variable has something to do with this because when I remove that variable from the regression then African Americans have the highest negative coefficient.

Correlation Plots

I also created a “**Correlation Plots**” tab in my website and this allows the user to see a visual on how the dependent (Attendance Rate) and independent variables (Race and low-income families) correlate visually. It is important to note that the visualizations from these graphs are comparisons of two variables only, so the analysis and conclusions taken from here should not directly be compared to the regression since the regression takes into consideration multiple independent variables in its output.

To view the different correlation plots the user has to select the independent variable from the drop-down menu. This would then show the correlation plot of the independent and dependent variable.

The correlation between Attendance rates and African American students/Latino students was negative which made sense and supported my hypothesis partially.

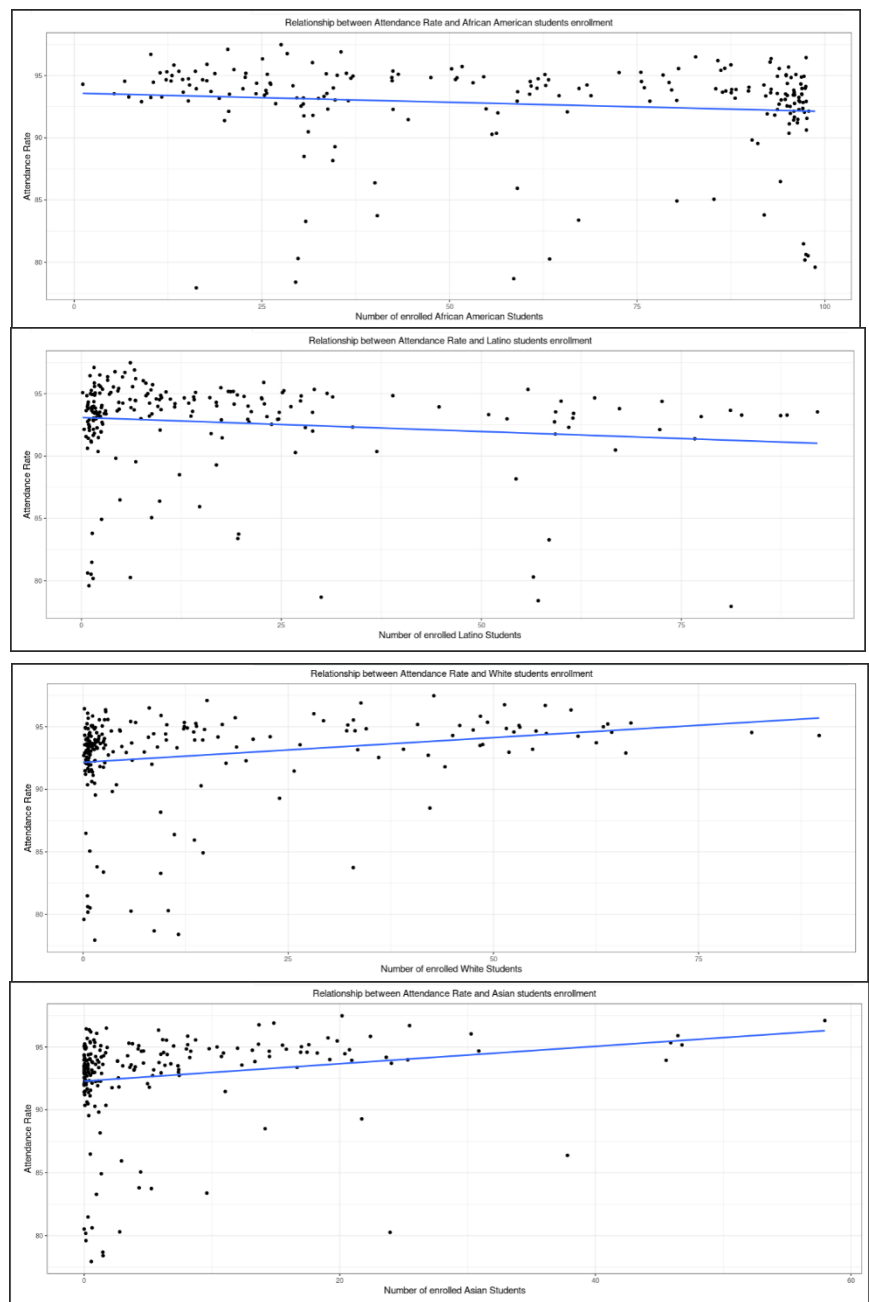
I did notice that the line of best fit for both graphs decreased gradually which made me realize that there might be another variable that could play a role in the decrease in student attendance rate.

Also, seeing the regression output confirmed these doubts.

I was surprised to see that the correlation between Attendance and Asian students was positive. I hypothesized that the correlation would be negative.

What I find interesting is that in my regression I got a negative coefficient for all races which meant that an increase in students for minority or white races would cause a decrease in attendance rates.

I believe that the correlation plots give a general understand about the relationships between the variables but to investigate how various variables affect each other simultaneously I would need to conduct a difference of means test or regression.



Conclusion

The relationship between student attendance rate and race is a complicated one. There are many other variables (such as income) that need to be taken into consideration when looking at what causes a decrease in attendance.

My hypothesis was proven to be wrong since my regression output showed that all races had a negative coefficient (a negative decrease in attendance rate) when family income was taken into account.

But when looking at attendance rate and race alone I could see that white and Asian student enrollment had a positive correlation with attendance rate, but Latino and African American students had negative correlations.