Relationship between School Quality & Crime Rates in New York

Emily Padvorac, Dana Chermesh, Matt Dwyer, Rebecca Scheidegger

I. Introduction

This research seeks to explore the relationship between school quality and crime in New York City. The goal is to better understand if crimes near schools negatively impact school performance or if low quality schools result in areas with higher levels of crime. To do this, data will be used from 2007 and 2013, as well as the changes between those years, to additionally assess how areas may have changed over time. Our hypothesis is that crime near schools negatively impacts school performance, particularly when looking at certain categories of crimes, such as drug-related offenses.

II. Context & Motivation

Both academics and the media have sought to explore how student performance may be impacted by local environments (Badger, 2014). In particular, many have considered the effects of violent crime on student achievement. Research conducted in Chicago found that test scores declined following violent incidents in the immediate area, even if a child did not personally witness the event (Sharkey, 2010; Burdick-Will, 2013). Similar research found that violent incidents in the community results in lower IQ test results for adolescents (Delaney-Black et al., 2002). Less explored is how other categories of crime, such as non-violent crimes like drug-related offenses, may relate to educational achievement. This research proposes to explore how violent crimes affect school quality in New York, as well as exploring how other non-violent categories of crime may relate to local school quality.

III. Data

School Data

The NYC Department of Education (DOE) annually evaluates K-12 schools on a variety of indicators, resulting in a singular performance grade for each school. From 2007 to 2013, the metric used was by the DOE was called Progress Reports. The reports considered three weighted categories: student progress (60%), student performance (25%), and school environment (15%). Based on a school's performance relative to its peers, it was assigned a letter grade (A, B, C, D, F). For this analysis, the letter grade was converted to a numerical value (A=5, B=4, C=3, D=2, F=1). The Progress Report datasets are publically available through the NYC DOE website, as well as a detailed methodology of how scores are calculated (New York City Department of Education, 2009). After 2013, the Progress Reports were replaced with a new methodology and currently only three years of data are available. Due to this, in order to maximize the range of the years compared, we focused on data from 2007 and 2013. Lastly, NYC Public Schools shapefiles were obtained for spatial descriptive analysis and were merged with the school quality data.

Crime Data

Crime data was accessed through the dataset 'NYPD Complaint Data Historic', available through NYC Open Data. The dataset lists felony, misdemeanor, and violation crimes from 2006 to 2016 that were reported to New York Police Department (NYPD). Different categories of crimes were developed by offense descriptions for the analysis. The five categories constructed are: Drug Crimes, Murder, Theft, Crimes Against Children, and Assault. A full list of specific offenses considered for each category is available in Appendix A. To match the school data, only crimes from 2007 and 2013 were extracted.

IV. Methodology & Results

Merging Data

Data was first merged on the geographic level of Public Use Microdata Areas (PUMA). For school quality data, the average school quality in 2007 and 2013 was calculated per PUMA. For crimes, both total crime and crime per the categories outlined above were summed for each PUMA. Further, the difference in crime counts and school rating was calculated by subtracting the 2007 value from the 2013 value per PUMA, ensuring that a decrease would result in a negative number.

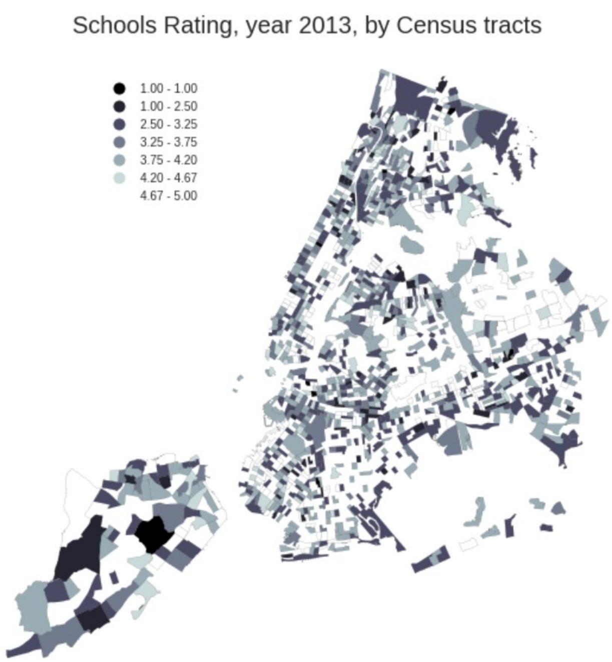
The data was not normalized by population for two reasons. First, PUMAs are designed to contain approximately 100,000 people. Second, crimes do not necessarily occur where people live. For example, though the residential population of Midtown Manhattan is low, the number of crimes that occurs is high, because of a significant day-time population due to work and tourism. It therefore seemed presumptive to conclude that a PUMA with a higher population would be likely to have crime rates proportionally higher than a PUMA with a lower population, so such a scaling was not applied.

Descriptive Statistics

To understand school ratings in a higher spatial resolution, each school was spatially joined to census tracts, with counts ranging from zero to nine schools in each census tract. The ratings for both years were analyzed, as well as the ratings change between the years, then averaged by census tract and colored to display the spatial distribution of school ratings, where a higher rate is represented by a lighter color (Figure 1).

Across the city, the school quality scores per census tract ranged from 1 to 5 in both 2007 and 2013, as shown in the maps of Figure 1. Also, the map in Figure 2, which looks at the percentage change, shows different trends in different census tracts, with no apparent spatial correlation. Red colors indicate negative change, while the blue colors reveal progress in school quality.

Fig 1: Maps of School Rating by Census tracts, 2007, 2013



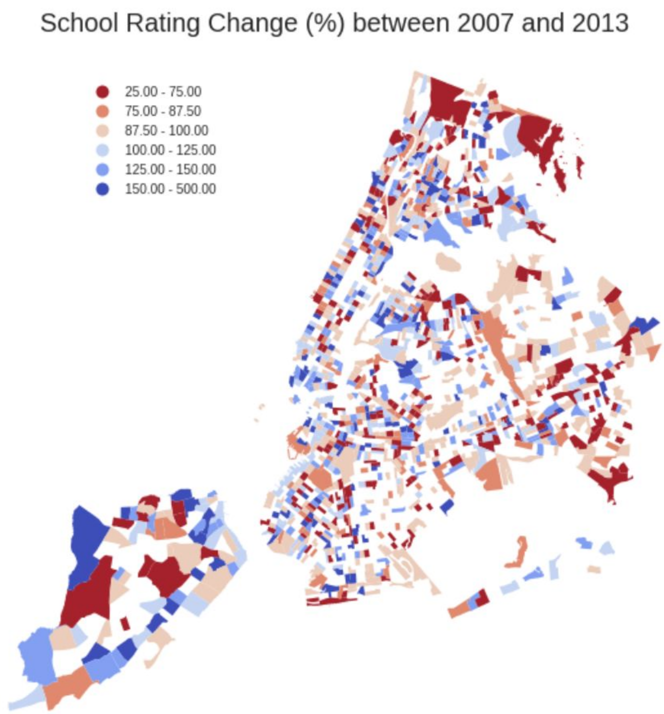
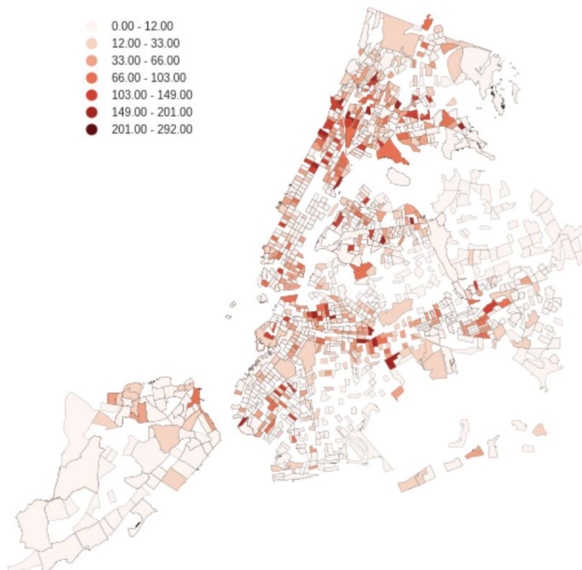


Fig 2: Map of School Rating Change (%) between 2007 and 2013, by Census tracts

Crime data were also spatially observed in the Census tract level, in order to get a better sense of its distribution across the city and of the possible relationship of it to the spatial behaviour of schools rating. For this instance we choose to use only *drug crimes* data, under the assumption that they might have a significant impact on youth who live among them. Each map in figure 3 shows the number of drug crimes per Census tract, with higher number represented by darker red color.



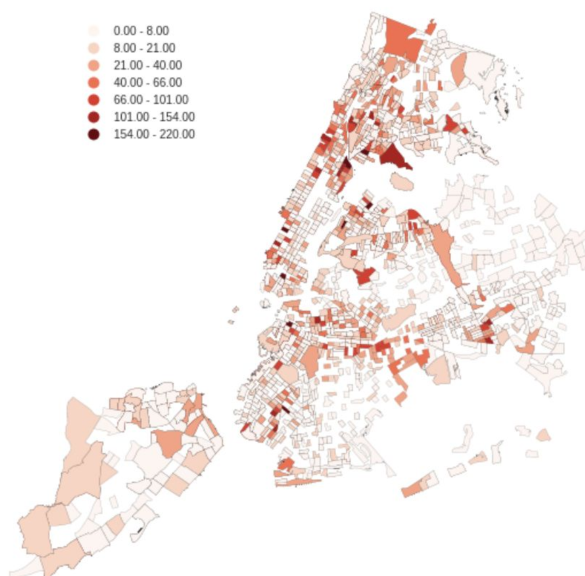


Fig 3: Maps of Drug Crimes by Census tracts, 2007 and 2013

Linear Regressions

In this analysis, linear regression was performed on both crime and school datasets for the two considered years, the crime being the regressors and the school quality being the dependent variable. Linear regression between school quality scores and number of crimes for

2007 produced an R-squared value of 0.029 (Figure 4). Linear regression between school quality scores and number of crimes for 2013 produced an R-Square value of 0.026 (Fig 5). Our R-squared value for for 2007 was slightly higher than our value for 2013, but not to a significant degree.

OLS Regression Results

Dep. Variable:	Rating07	**R-squared:**	0.029
Model:	OLS	**Adj. R-squared:**	0.011
Method:	Least Squares	**F-statistic:**	1.587
Date:	Sat, 09 Dec 2017	**Prob (F-statistic):**	0.213
Time:	22:06:28	**Log-Likelihood:**	-21.005
No. Observations:	55	**AIC:**	46.01
Df Residuals:	53	**BIC:**	50.03
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	3.8773	0.118	32.744	0.000	3.640	4.115
all_crimes07	-1.452e-05	1.15e-05	-1.260	0.213	-3.76e-05	8.6e-06

Omnibus:	0.515	**Durbin-Watson:**	1.163
Prob(Omnibus):	0.773	**Jarque-Bera (JB):**	0.459
Skew:	0.212	**Prob(JB):**	0.795
Kurtosis:	2.855	**Cond. No.**	2.50e+04

Fig 4: Regression results for 2007

OLS Regression Results

Dep. Variable:	Rating13	**R-squared:**	0.026
Model:	OLS	**Adj. R-squared:**	0.007
Method:	Least Squares	**F-statistic:**	1.389
Date:	Sat, 09 Dec 2017	**Prob (F-statistic):**	0.244
Time:	20:34:37	**Log-Likelihood:**	-28.002
No. Observations:	55	**AIC:**	60.00
Df Residuals:	53	**BIC:**	64.02
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	3.9839	0.129	30.879	0.000	3.725	4.243
all_crimes13	-1.585e-05	1.34e-05	-1.179	0.244	-4.28e-05	1.11e-05

Omnibus:	2.546	**Durbin-Watson:**	1.115
Prob(Omnibus):	0.280	**Jarque-Bera (JB):**	2.306
Skew:	0.409	**Prob(JB):**	0.316
Kurtosis:	2.420	**Cond. No.**	2.24e+04

Fig 5: Regression results for 2013

Multivariate Regression

A multivariate regression was performed on the crime and school data by using the five previously described crime categories as regressors and the school rating as the dependent variable. The resulting R-squared for 2007 crime and school ratings was low, but considerably higher for 2013 with a value of 0.41. The change in ratings and crime however had the lowest R-squared of 0.106, indicating that only for the year 2013 do the selected crime categories explain some amount of school performance. The coefficients associated with each category of crime demonstrate a slight negative trend, consistent with the notion of higher crime rates negatively impacting a school's performance, but the low value of each indicate that such a relationship is quite weak.

OLS Regression Results

Dep. Variable:	df['Rating07']	**R-squared:**	0.205
Model:	OLS	**Adj. R-squared:**	0.124
Method:	Least Squares	**F-statistic:**	2.524
Date:	Sat, 09 Dec 2017	**Prob (F-statistic):**	0.0413
Time:	22:36:12	**Log-Likelihood:**	-15.514
No. Observations:	55	**AIC:**	43.03
Df Residuals:	49	**BIC:**	55.07
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	3.8485	0.117	32.763	0.000	3.612	4.085
df['theft_crimes07']	2.603e-05	2.58e-05	1.009	0.318	-2.58e-05	7.79e-05
df['assault_crimes07']	2.672e-05	0.000	0.119	0.906	-0.000	0.000
df['child_crimes07']	-0.0043	0.005	-0.793	0.432	-0.015	0.007
df['drug_crimes07']	-6.354e-05	0.000	-0.621	0.537	-0.000	0.000
df['murder_crimes07']	-0.0164	0.009	-1.804	0.077	-0.035	0.002

Omnibus:	1.385	**Durbin-Watson:**	1.032
Prob(Omnibus):	0.500	**Jarque-Bera (JB):**	0.920
Skew:	-0.313	**Prob(JB):**	0.631
Kurtosis:	3.096	**Cond. No.**	9.77e+03

Fig 6: Multivariate regression, 2007

OLS Regression Results

Dep. Variable:	df['Rating13']	**R-squared:**	0.410
Model:	OLS	**Adj. R-squared:**	0.350
Method:	Least Squares	**F-statistic:**	6.815
Date:	Sat, 09 Dec 2017	**Prob (F-statistic):**	6.77e-05
Time:	22:38:02	**Log-Likelihood:**	-14.195
No. Observations:	55	**AIC:**	40.39
Df Residuals:	49	**BIC:**	52.43
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4.0973	0.106	38.722	0.000	3.885	4.310
df['thieving13']	6.569e-05	2.61e-05	2.521	0.015	1.33e-05	0.000
df['assault13']	-0.0004	0.000	-1.753	0.086	-0.001	5.42e-05
df['child13']	-0.0094	0.007	-1.316	0.194	-0.024	0.005
df['drug13']	0.0002	0.000	1.107	0.274	-0.000	0.001
df['murder13']	-0.0259	0.017	-1.548	0.128	-0.059	0.008

Omnibus:	1.107	**Durbin-Watson:**	1.786
Prob(Omnibus):	0.575	**Jarque-Bera (JB):**	0.975
Skew:	-0.097	**Prob(JB):**	0.614
Kurtosis:	2.377	**Cond. No.**	8.82e+03

Fig 7: Multivariate regression, 2013

OLS Regression Results

Dep. Variable:	df['RatingChange']	**R-squared:**	0.106
Model:	OLS	**Adj. R-squared:**	0.015
Method:	Least Squares	**F-statistic:**	1.159
Date:	Sat, 09 Dec 2017	**Prob (F-statistic):**	0.343
Time:	22:44:48	**Log-Likelihood:**	-27.794
No. Observations:	55	**AIC:**	67.59
Df Residuals:	49	**BIC:**	79.63
Df Model:	5		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.1130	0.071	1.595	0.117	-0.029	0.255
df['theftChange']	-7.65e-05	0.000	-0.381	0.705	-0.000	0.000
df['assaultChange']	-0.0008	0.001	-1.524	0.134	-0.002	0.000
df['childChange']	-0.0105	0.007	-1.592	0.118	-0.024	0.003
df['drugChange']	-8.169e-05	0.000	-0.518	0.607	-0.000	0.000
df['murderChange']	-0.0019	0.014	-0.137	0.892	-0.031	0.027

Omnibus:	14.734	**Durbin-Watson:**	1.534
Prob(Omnibus):	0.001	**Jarque-Bera (JB):**	18.870
Skew:	0.990	**Prob(JB):**	7.99e-05
Kurtosis:	5.078	**Cond. No.**	613.

Fig 8: Multivariate regression, change 2007 to 2013

Buffer Zone Analysis

After running the various regressions at a PUMA level, we hypothesized that perhaps they are too large a geography to produce meaningful results. As an alternative, we created 0.5 mile buffer zones around each school and summed up the number of all types of crime in the zone and only drug-related offenses in the buffer zone. The results for both produced low R-squared values. For the analysis considering all types of crime within a 0.5 mile buffer and school quality, R-squared was 0.002 (Figure 9). When just looking at drug-related offenses within 0.5 miles, the R-squared was 0.005 (Figure 10).

OLS Regression Results

Dep. Variable:	Rating	**R-squared:**	0.005
Model:	OLS	**Adj. R-squared:**	0.004
Method:	Least Squares	**F-statistic:**	5.632
Date:	Sun, 10 Dec 2017	**Prob (F-statistic):**	0.0178
Time:	10:42:29	**Log-Likelihood:**	-1687.7
No. Observations:	1165	**AIC:**	3379.
Df Residuals:	1163	**BIC:**	3390.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	3.7664	0.040	94.538	0.000	3.688	3.845
Count_	-0.0002	9.5e-05	-2.373	0.018	-0.000	-3.9e-05

Omnibus:	67.631	**Durbin-Watson:**	1.853
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	78.610
Skew:	-0.635	**Prob(JB):**	8.51e-18
Kurtosis:	3.072	**Cond. No.**	553.

Fig 9: Buffer zone regression, 2007, Drug Crimes

OLS Regression Results

Dep. Variable:	Rating	**R-squared:**	0.002
Model:	OLS	**Adj. R-squared:**	0.001
Method:	Least Squares	**F-statistic:**	1.878
Date:	Sun, 10 Dec 2017	**Prob (F-statistic):**	0.171
Time:	10:43:46	**Log-Likelihood:**	-1689.6
No. Observations:	1165	**AIC:**	3383.
Df Residuals:	1163	**BIC:**	3393.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	3.7647	0.053	70.766	0.000	3.660	3.869
Count_	-2.122e-05	1.55e-05	-1.370	0.171	-5.16e-05	9.16e-06

Omnibus:	67.090	**Durbin-Watson:**	1.851
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	77.974
Skew:	-0.633	**Prob(JB):**	1.17e-17
Kurtosis:	3.043	**Cond. No.**	6.04e+03

Fig 10: Buffer zone regression, 2007, All Crimes

V. Conclusions

Limitations & Future Research

The linear and multivariate regression models at either a PUMA level or 0.5 mile buffer zone, did not give us a clear sense of the correlation between crime rates and school performance. Also, the map observations of the two phenomena do not reveal obvious spatial clustering and/or trends. Considering these results, there is a need to consider the limitations this research faced and opportunities for future research.

As with any crime related analysis, there is always the possibility of underreporting of different types of crime and under or over-reporting in different areas of the city. If crimes are going unreported, it may make the crime rate look lower than the reality and thus not provide an accurate gauge of the local environment around a school.

While averaging school ratings within a geographic area turned categorical values to continuous, looking at schools on an individual basis by utilizing small buffers around schools to count nearby crimes directly compared school performance and crime. The analysis above utilized an ordinary least squares regression, while a further analysis could instead use an ordinal regression to be able to predict the categorical school rating variables. Further, crimes could be weighted in severity to approximate the impact of different events on performance.

Another opportunity would be to look at more specific student academic performance metrics. As the school quality rating includes multiple factors in addition to academic achievement, it is possible that isolating one academic performance measure, such as just standardized test scores, may change the results of the analysis.

The change in school rating from 2007 to 2013 could also be tested for spatial autocorrelation to measure how the phenomenon of school improvement is clustered. The results could reveal patterns of change within neighborhoods, or may instead reveal more individualistic school by school improvement.

Ultimately, the results of this research did not find a conclusive relationship between school quality and crime rates using linear and multivariate regression models. As discussed, there exists opportunities to expand upon this work by considering additional methodological approaches.

References

Badger, Emily. 2014. Neighborhood violence is hurting black kids' test scores. Washington Post. https://www.washingtonpost.com/news/wonk/wp/2014/05/29/just-how-much-is-neighborhood-violence-hurting-black-kids-test-scores/?utm_term=.7ea5361dc5de

Burdick-Will, J. 2013. School violent crime and academic achievement in Chicago. Sociology of Education 86: 343-361. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3831577/

Delaney-Black, V., Covington, C., Ondersma, S.J., Nordstrom-Klee, B., Templin, T., Ager, J., Janisse, J., Sokol. R.J. 2002. Violence Exposure, Trauma, and IQ and/or Reading Deficits Among Urban Children. Arch Pediatr Adolesc Med. 2002;156(3):280–285. https://jamanetwork.com/journals/jamapediatrics/fullarticle/191640

Grogger, J. (1997). Local violence and educational attainment. Journal of Human Resources, 32(4), 659-682. https://link.springer.com/article/10.1023/B:CCFP.0000006292.61072.d2

New York City Department of Education. 2009. Educator Guide: New York City Progress Report. http://schools.nyc.gov/NR/rdonlyres/B027DF51-87CB-453C-9A80-5C37A7B12B86/0/EMS200708.pdf

Sharkey, P. 2010. The Acute Effect of Local Homicides on Children's Cognitive Performance. Proceedings of the National Academy of Sciences of the United States of America 107 (26):11733–11738. http://www.pnas.org/content/107/26/11733.full

Appendix A: Crime Subcategories

Drug Crimes: 'Controlled substance, possessi', 'Controlled substance,possess.', 'Marijuana, possession 4 & 5', 'Controlled substance, intent to', 'Controlled substance, sale 3', 'Controlled substance, sale 1', 'Controlled substance, sale 5', 'Marijuana, sale 4 & 5', 'Possession hypodermic instrume', 'Drug paraphernalia, possesse', 'Sale school grounds 4', 'Controlled substance, intent t', 'Marijuana, sale 1, 2 & 3', 'Marijuana, possession 1, 2 & 3', 'Controlled substance, sale 2', 'Controlled substance, sale 4', 'Controlled substance, possess.-', 'Drug, injection of', 'Sales of prescription'

Crimes Against Children: 'Child, offenses against, unclass', 'Child, alcohol sale to', 'Dhild abandonment', 'Labor law,employing children', 'Education law'

Theft crimes: 'Larceny, grand by bank acct compromise-unclassified', 'Larceny, grand by open/compromise cell phone acct', 'Larceny, grand by open credit card (new acct)', 'Larceny, grand by open bank acct', 'Larceny, grand from person, uncl', 'Larceny, grand from building (non-residence) unattended', 'Larceny, grand by dishonest emp', 'Larceny, grand by theft of credit card', 'Larceny, grand by extortion', 'Larceny, grand by bank acct compromise-reproduced check', 'Larceny, petit from building,un', 'Larceny, petit by dishonest emp', 'Larceny, petit of license plate', 'Larceny, petit from open areas,', 'Larceny, petit from auto', 'Larceny, grand from open areas, unattended', 'Larceny, petit by credit card u', 'Larceny, petit from store-shopl', 'Larceny, petit by check use', 'Larceny, grand by acquiring los', 'Larceny, grand from vehicle/motorcycle', 'larceny, grand of vehicular/motorcycle accessories', 'robbery,residential common area', 'larceny, petit by false promise', 'larceny, petit of vehicle acces', 'burglary, residence,day', 'burglary, residence,night', 'robbery, open area unclassified', 'burglary, residence,unknown tim', 'larceny,grand from person,pick', 'burglary,commercial,day', 'robbery,unlicensed for hire vehicle', 'burglary,commercial,night', 'burglary,unclassified,day', 'larceny,grand person,neck chai', 'robbery,bar/restaurant', 'robbery,pocketbook/carried bag', 'robbery,commercial unclassified', 'robbery, chain store', 'larceny, grand from person, purs', 'robbery,bicycle', 'robbery, neckchain/jewelry', 'robbery,gas station', 'larceny,grand from store-shopl', 'robbery, car jacking', 'robbery,public place inside', 'robbery,bank', 'robbery,dwelling', 'larceny,grand by false promise', 'robbery,pharmacy', 'larceny,petit by acquiring los', 'burglary,unknown time', 'burglary, truck night', 'larceny, petit from truck', 'burglary,unclassified,night', 'robbery,check cashing business', 'burglary,truck day', 'larceny,petit of bicycle', 'robbery,bodega/convenience store', 'burglary, commercial,unknown ti', 'petit larceny-check from mailb', 'burglary,unclassified,unknown', 'larceny,grand of bicycle', 'robbery,licensed medallion cab', 'robbery,atm location', 'larceny,petit from coin machine', 'robbery,on bus/ or bus driver', 'grand larceny-check from mailb', 'robbery, doctor/dentist office', 'robbery, payroll', 'robbery,liquor store', 'larceny,grand from pier, unattended', 'robbery,hijacking', 'larceny,petit from pier', 'larceny,grand from boat, unattended', 'larceny,grand from truck, unattended', 'larceny, grand from coin machine', 'larceny,petit from parking met', 'larceny,petit of boat', 'robbery,of truck driver', 'larceny,petit from boat', 'Larceny, grand by identity theft-unclassified', 'Larceny, grand of boat', 'Larceny, grand from retail store, unattended', 'larceny, grand from residence, unattended'

Assault Crimes: 'Assault 3', 'Menacing, unclassified', 'Menacing, peace officer'

Murder Crimes: ‘Murder & Non-negl. Manslaughter’

Appendix B: Team Work Acknowledgement Statement

All team members contributed equally to this project. We jointly discussed our data and approach on how to apply analytical methods for crime and school quality data. We collaboratively obtained, cleaned, and analyzed the data as a team. This report and the presentation of this research were written by all team members.