



Applied Data Science, NYU CUSP, Fall 2017  
Matt Dwyer, Emily Padvorac, Rebecca Scheidegger, Dana Chermesh

# School Performance and Crime Rates

# 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**.

## II | Context and Motivation

- Previous research in other cities has found **decreases in standardized test performance and IQ scores after violent incidents** in neighborhood, regardless if a child was present
- Previous research has not explored how **non-violent crimes** may impact school quality

### III | Data Inventory

Data	Source	Comments
* <b>Crime data</b>	Historical New York Crime Dataset	<b>Five crime sub-categories</b> Theft; Murder; Drugs; Assault; Crimes against children
* <b>School quality and Performance data</b>	NYC Department of Education	<b>Overall quality grade</b> - Student progress, performance, school environment <b>A, B, C, D, E &gt;&gt; 1-5</b>

# IV | Methodology

1. **Descriptive Statistics** // Mapping variables (by census tract)
2. **Linear Regression** // All crimes & Drugs only (by PUMA)
3. **Multivariate Regression** // Sub-categories (by PUMA)
4. **Buffer Zone Analysis**

# IV | Methodology // Geography

1. Descriptive Statistics

2. Linear Regression

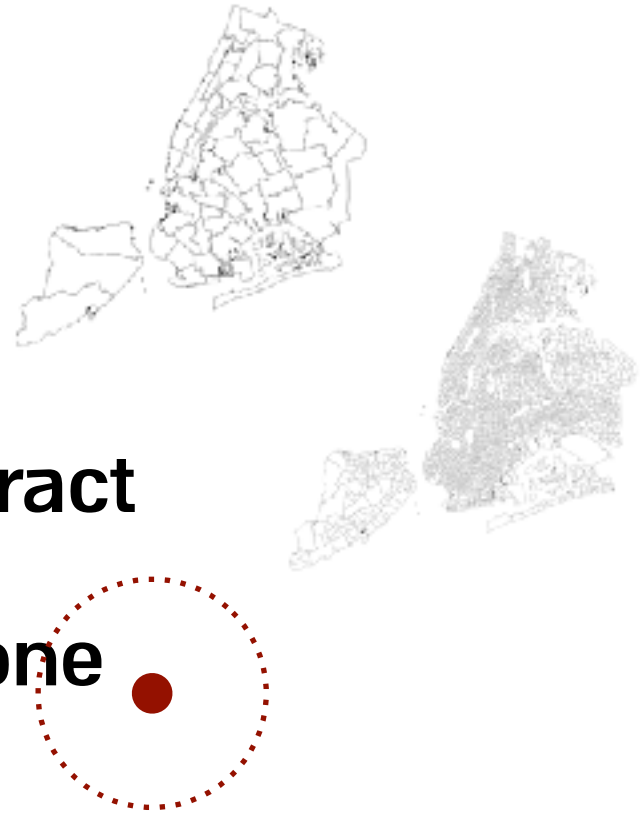
3. Multivariate Regression

4. Buffer Zone Analysis

- **PUMA**

- **Census tract**

- **Buffer Zone**



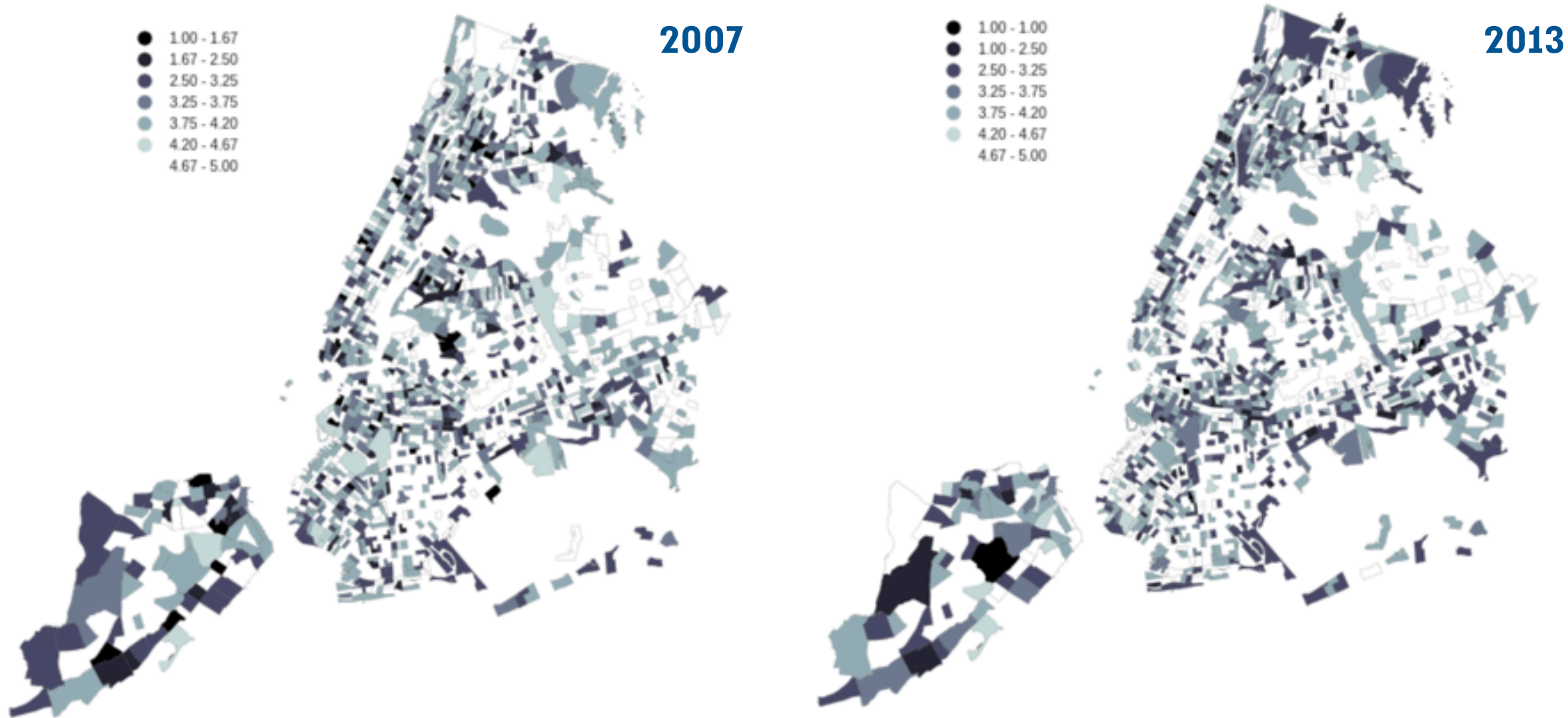
# 1.Descriptive Statistics

---

Mapping School Ratings and Crime



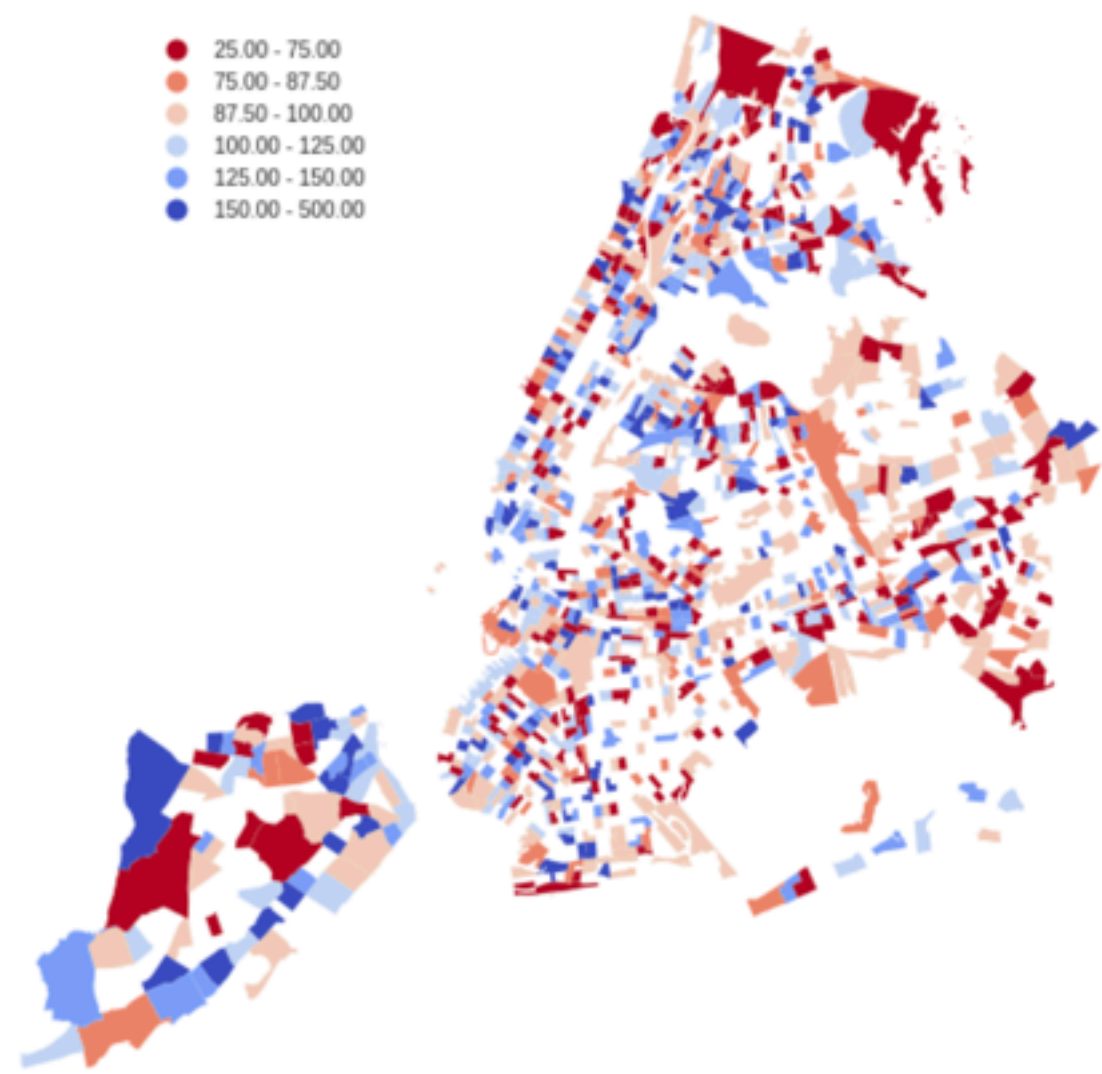
# **\*\* School Rating (1-5) by Census tract**



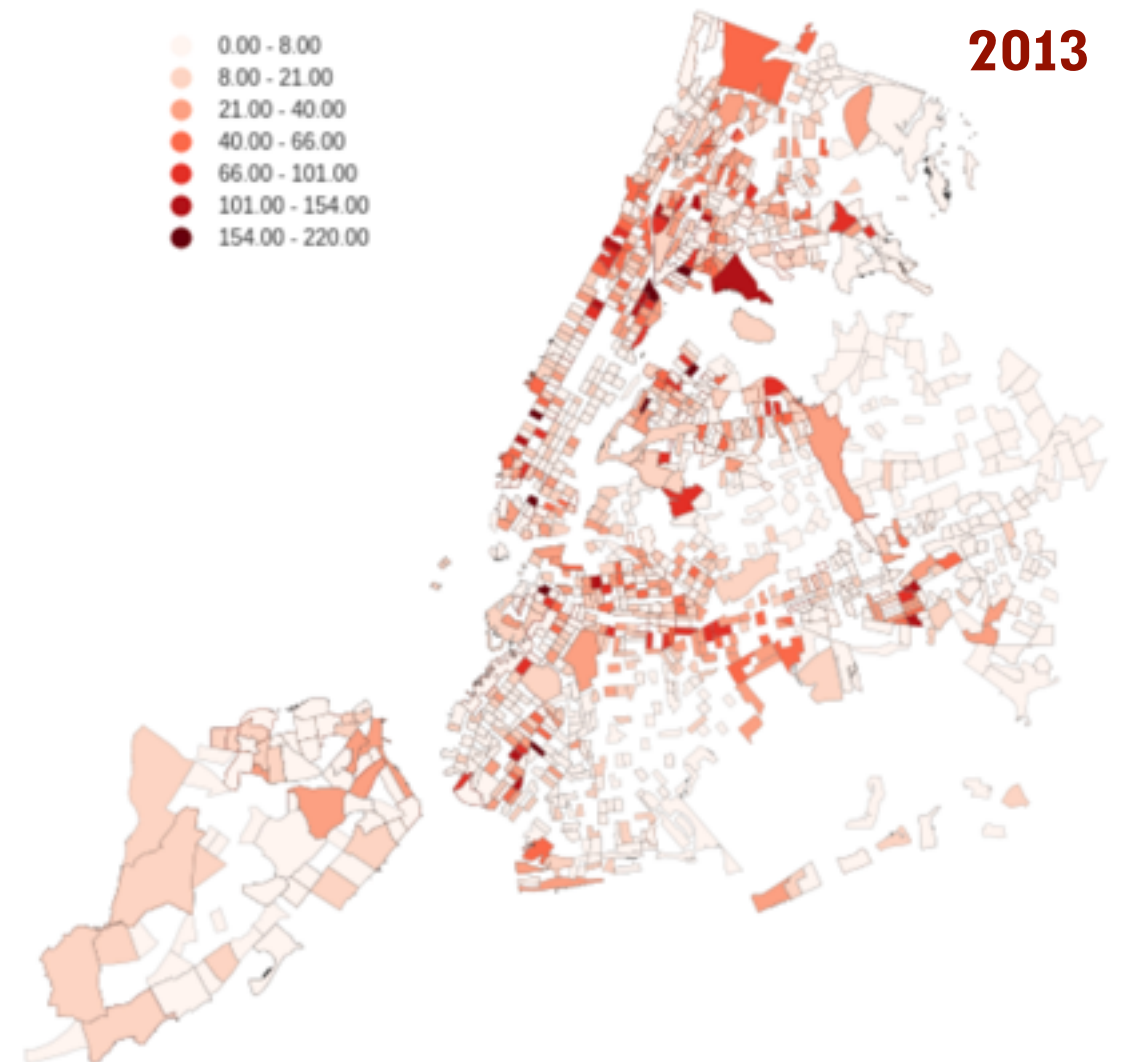
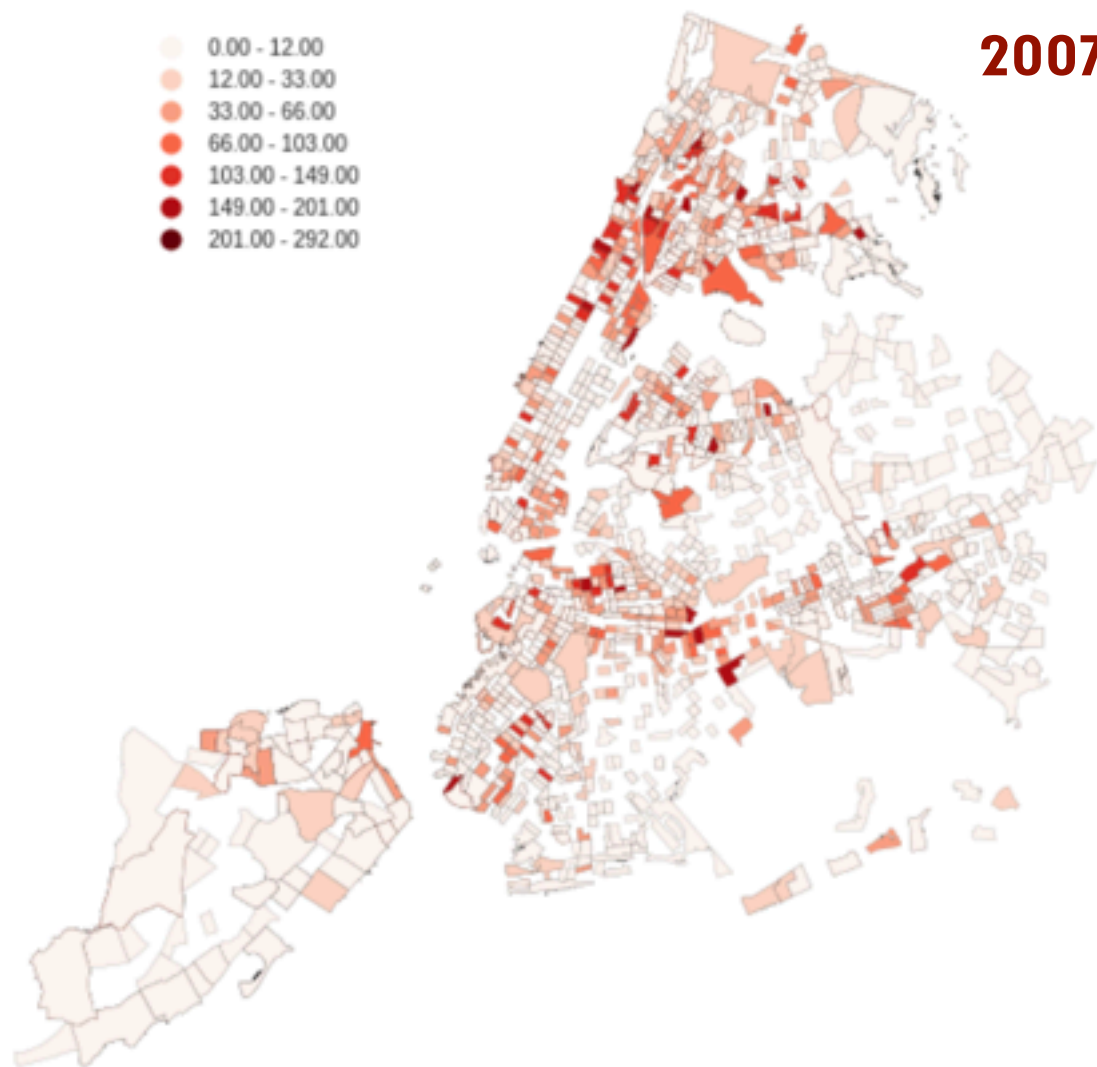


# \*\* School Rating (1-5) by Census tract

Change between  
2007 and 2013



## \*\* Number of Drug crimes



## 2. Linear Regression

---

- Independent: **All Crimes**
- Independent: **Drug Crimes** Only

# OLS Regression Results

2007

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

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

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

# OLS Regression Results

2013

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

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

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

# 3. Multivariate Regression

---

Crimes Sub-Categories

2007

OLS Regression Results

Dep. Variable:	df['Rating07']	R-squared:	0.205
Model:	CLS	Adj. R-squared:	0.124
Method:	Least Squares	F-statistic:	2.524
Date:	Sat, 09 Dec 2017	Prob (F-statistic):	0.0413
Time:	22:06:12	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']	2.803e-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:	1.335	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

2013

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	-1.313	0.194	-0.024	0.005
df['drug13']	0.0002	0.000	1.107	0.274	-0.000	0.001
df['murder13']	-0.0250	0.017	-1.543	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

### 3. Multivariate Regression



# OLS Regression Results

<b>Dep. Variable:</b>	df['FatingChange']	<b>R-squared:</b>	0.106
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.015
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	1.159
<b>Date:</b>	Sat, 09 Dec 2017	<b>Prob (F-statistic):</b>	0.343
<b>Time:</b>	22:44:48	<b>Log-Likelihood:</b>	-27.794
<b>No. Observations:</b>	55	<b>AIC:</b>	67.69
<b>Df Residuals:</b>	49	<b>BIC:</b>	79.63
<b>Df Model:</b>	5		
<b>Covariance Type:</b>	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	0.113	-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

<b>Omnibus:</b>	14.734	<b>Durbin-Watson:</b>	1.534
<b>Prob(Omnibus):</b>	0.001	<b>Jarque-Bera (JB):</b>	18.870
<b>Skew:</b>	0.990	<b>Prob(JB):</b>	7.99e-06
<b>Kurtosis:</b>	5.078	<b>Cond. No.</b>	613.

Change between  
2007 and 2013

## 3. Multivariate Regression

## 4. Buffer Zone Analysis

---

- Independent: **All Crimes**
- Independent: **Drug Crimes** Only

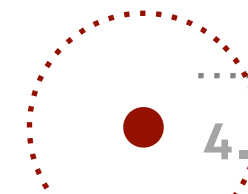
# OLS Regression Results

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

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

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

All Crimes; 2007



# OLS Regression Results

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

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

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

Drug Crimes; 2007

# V | Conclusion

1. **Linear and Multivariate Regression:** 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.
2. **Map observations:** The map observations of the two phenomena do not reveal obvious spatial clustering and/or trends

## VI | Further Work

1. Perform **Ordinal Logistic regression** on buffer zones
2. Consider **individual student performance** data



---

**\*\* Thank you \*\***

---

