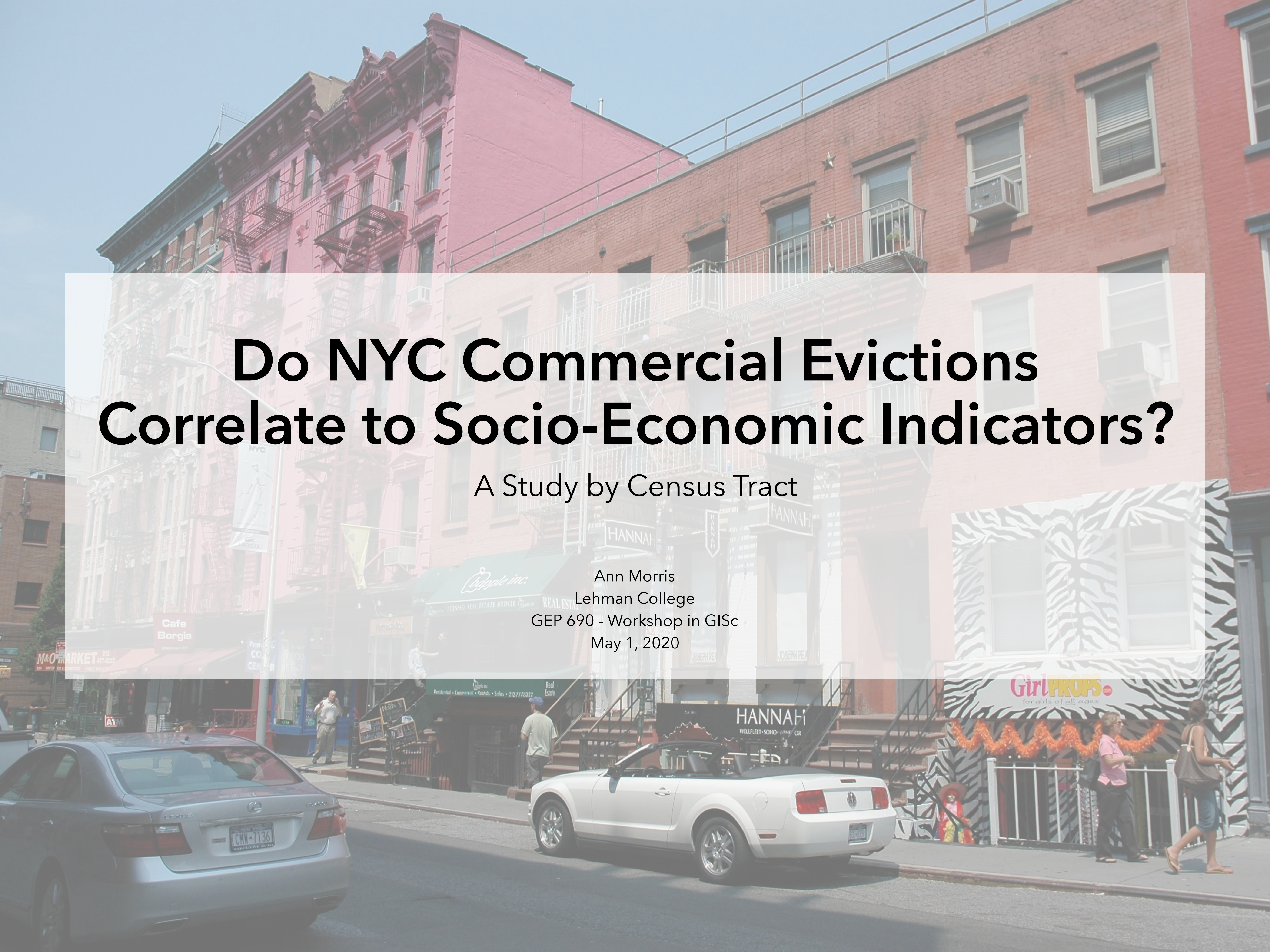


Do NYC Commercial Evictions Correlate to Socio-Economic Indicators?

A Study by Census Tract

Ann Morris
Lehman College
GEP 690 - Workshop in GISc
May 1, 2020

Background

It is frequently suggested that retail businesses in New York City have difficulty renewing their releases due to rising rents. Rising rents, along with tenant displacement and commercial turnover, are among what Samuel Stein, author of *Capital City: Gentrification and the Real Estate Stat*e, considers to be the core attributes of gentrification (2019, Chapter 2, 49-50, para. 4). In extreme cases, tenants are evicted, unexpectedly changing the fabric of a neighborhood in ways that its residents may resent. This debate has been in progress for some time, and has recently intensified with the passage of New York State's 2019 rent laws.

There are several reasons why a landlord might evict a tenant, non-payment of rent being the most obvious. But are evictions isolated incidents, wholly dependent upon the relationship between landlord and tenant and the tenant's finances, or they related to conditions in a neighborhood? And if there is a relationship between evictions and changing conditions, are the evictions symptomatic of improving or declining conditions?

There has been so much discussion of gentrification in connection with New York City that it seems unthinkable to consider that any neighborhood might be in decline. In his book *The Divided City: Poverty and Prosperity in Urban America*, Alan Mallach describes a study he did on gentrification in Indianapolis. He found while there was intense discussion of gentrification in that city, that its effects, namely increased housing prices, decreased poverty rates, and a larger college-educated population, were limited to just five of about 200 census tracts. By his chosen metrics, most of the rest of the city was actually in decline (Mallach, 2018, Chapter 6, para. 2). New York City is large and diverse, but parts of it are much less in the spotlight than others. Therefore I thought it might be instructive to approach the topic without prejudice as to what "everyone knows."

For the purpose of this study, I will define improving conditions as rising median household incomes, education levels, housing values, home ownership levels, as well as declining poverty and unemployment rates, comparing the Decennial Census figures from 2000 to those from the American Community Survey of 2018. I am choosing 2000 as my starting point because the timeframe will include the 2008 financial crisis, and because 18 years is long enough to contemplate the idea that technological and political forces may have led to structural economic change.

The questions addressed by this study are briefly summarized on the next slide. I will be analyzing them using spatial autocorrelation and regression testing techniques.

Research Questions and Methodology

1. Does the rate of commercial evictions in a census tract correlate to socio-economic indicators such as median household income, education levels, poverty rate, homeownership, unemployment, or housing value?

2. Are commercial evictions in each census tract correlated with the extent of change that has occurred in these indicators between the years of 2000 and 2018?

This study will explore these questions using spatial autocorrelation regression testing techniques.

Data Sources

Description	Source	Format	Coordinate System	Temporal Resolution
Census Tract Socio-Economic Variables 2000: Total population, occupied housing units, owner-occupied housing units, population in labor force, population in labor force and unemployed, median household income, median housing value, educational attainment, poverty level.	Neighborhood Change Database (2000 variables redrawn according to 2010 borders)	.CSV		2000
Census Tract Socio-Economic Variables 2018: Total population, occupied housing units, owner-occupied housing units, population in labor force, population in labor force and unemployed, median household income, median housing value, educational attainment, poverty level.	American Community Survey	.CSV		2018
Census Tract Boundaries	NYC Department of City Planning	.SHP	New York Long Island FIPS State Plane (2263)	2010
Primary Land Use Tax Lot Output (PLUTO): Commercial area, total units, residential units.	NYC Department of City Planning	.CSV		2020
NYC Commercial Evictions	NYC Open Data	.CSV		2017-2020

Literature Review

Author	Year	Study Area	Data	Unit	Purpose	Method	Result
Al-Yami	2017	Meatpacking District, NYC (104-building Gansevoort Market Historic District)	(1) Land use and ownership change history from Greenwich Village Society for Historic Preservation website, (2) Geographical data from NYC Dept. of City Planning (tax lots), (3) building footprints from DOITT, (4) orthoimagery from USGS website	Census tracts, Tax lots	Gentrification study. Gentrification defined as a change in land use for more upscale businesses.	Mapping and color-coding historical land use and ownership changes over time, by tax lot. Renovation and conversion of building usage also studied. A geocorrected orthoimage was used to establish boundaries of study area.	The author identified waves of land use change in the distrit: between 1900 and 1938, between 1939 and 1969, and 1970 to 1993. Through color coding the changes in land use, the author was able to see a transition to a more service-oriented economy. The Meatpacking District is certainly gentrified, bu the changes in land usage are less recent than might be supposed: gentrification was underway from the 1970s. Since the 1990s, the focus has been on more and more fashionable businesses.
Benediktsson	2015	Brooklyn	Census data (ethnicity, per-capita income)	Census tracts	Gentrification; impact on local business and ethnicity	Mapping, data analysis, and interviews with local business owners Authors looked at correlation between rising incomes and number of Hispanic-owned businesses and determined that the first prefigured the second.	Decline in Hispanic-owned businesses occurred in places where the income rose and the Hispanic population declined. This displacement is not attributable to ethnic succession. Business Improvement Districts can help to stem this loss.
Freeman	2009	US	Decennial Census, data taken from Neighborhood Change Database, 1970-2000	Census tracts	Does a neighborhood's racial or income composition change as a result of gentrification? What is the relationship between different kinds of segregation and gentrification?	Gentrification considered at both the neighborhood and metropolitan levels. Trends in diversity studied over decades. Indexes calculated to measure balances of race, income, and education for gentrifying and non-gentrifying neighborhoods. Regression testing conducted with graphs, not maps produced.	At the neighborhood level, gentrification does not decrease diversity. It may or may not increase diversity. At the metropolitan level, gentrification reduces income segregation, but may increase racial segregation.
Meltzer	2016	NYC. Three neighborhoods studied at finer-grained level: East Harlem, Astoria, and Sunset Park.	(1) National Establishment Time Series Database, (2) PLUTO, (3) Neighborhood Change Database, (4) NAICS	Census tracts	What kinds of businesses survive or fail with gentrification?	Regression tests measuring business retention/ displacement and neighborhood gentrification levels citywide. Author calculated rates of retention and displacement for each property in sample, one for each five year period between 1990 and 2011. Analysis repeated for East Harlem, Astoria and Sunset Park to determine if nuances would be different at more local levels.	Citywide, gentrifying neighborhoods do not necessarily experience higher rates of business displacement. Retail spaces do remain vacant longer in gentrifying areas, and the mix of business types does change as a neighborhood gentrifies. Results for the neighborhood drill downs varied.

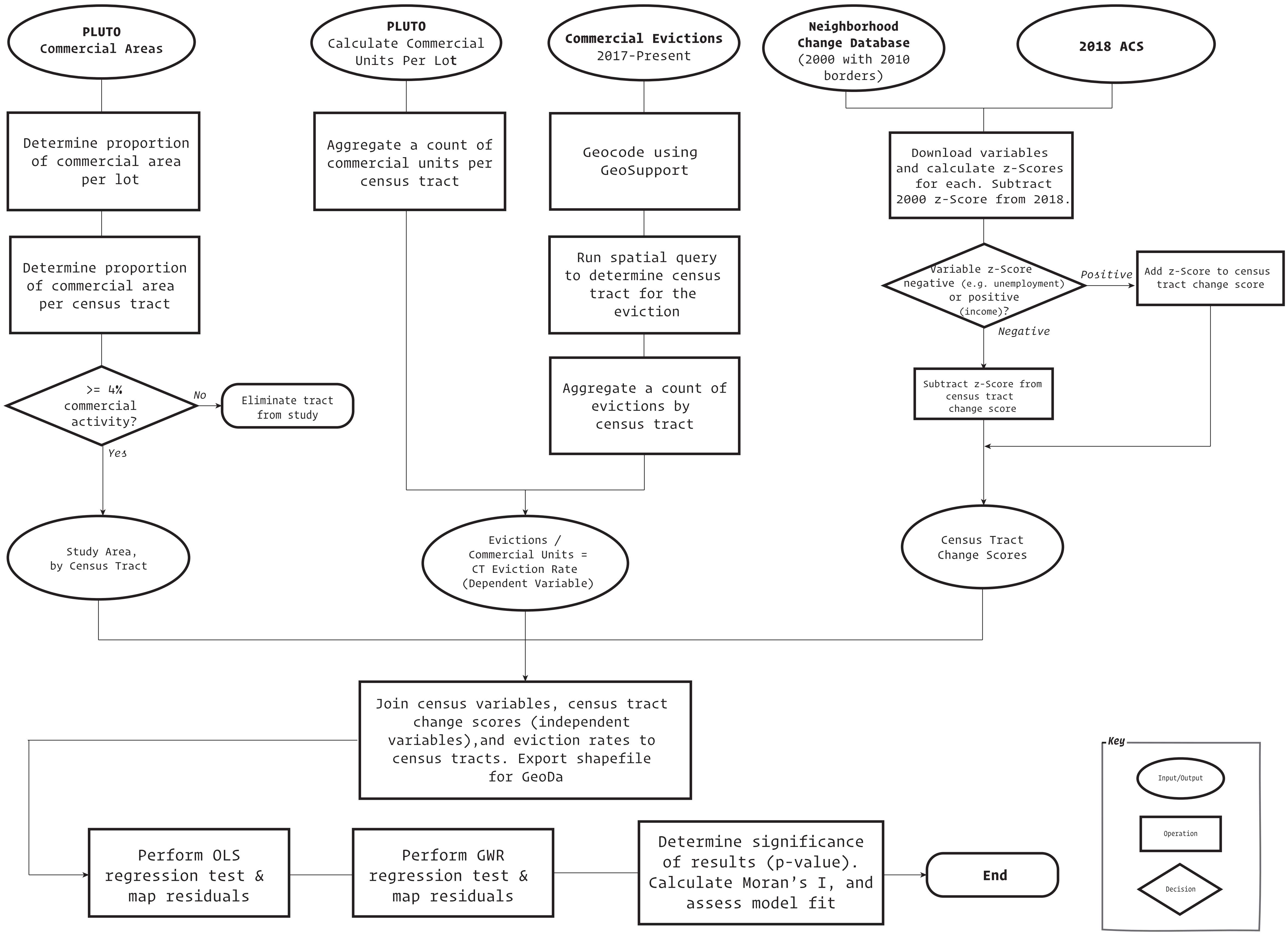
The Process

藥業(集團)有限公司
WING FAT COMPANY INC.
祭祀供品
美倫公司
212-964-6888
MEI LUN CO.
61B Bayard St.
老正川
SICHUAN CUISINE INC.
Mango Mango
美麗屋
HARMONY GIFT CENTER
OLD SICHUAN CUISINE INC.
THE ORIGINAL CHINATOWN ICE CREAM FACTORY
美倫

High Level Process Flow

1. Download and geocode the evictions.
2. Define the study area: Select census tracts based on their level of commercial activity.
3. Exploratory data analysis: Map the Census variables.
4. Calculate the census tract change score.
5. Conduct ordinary least squares and geographically weighted regression tests.
6. Assess model fit, and map the residuals.

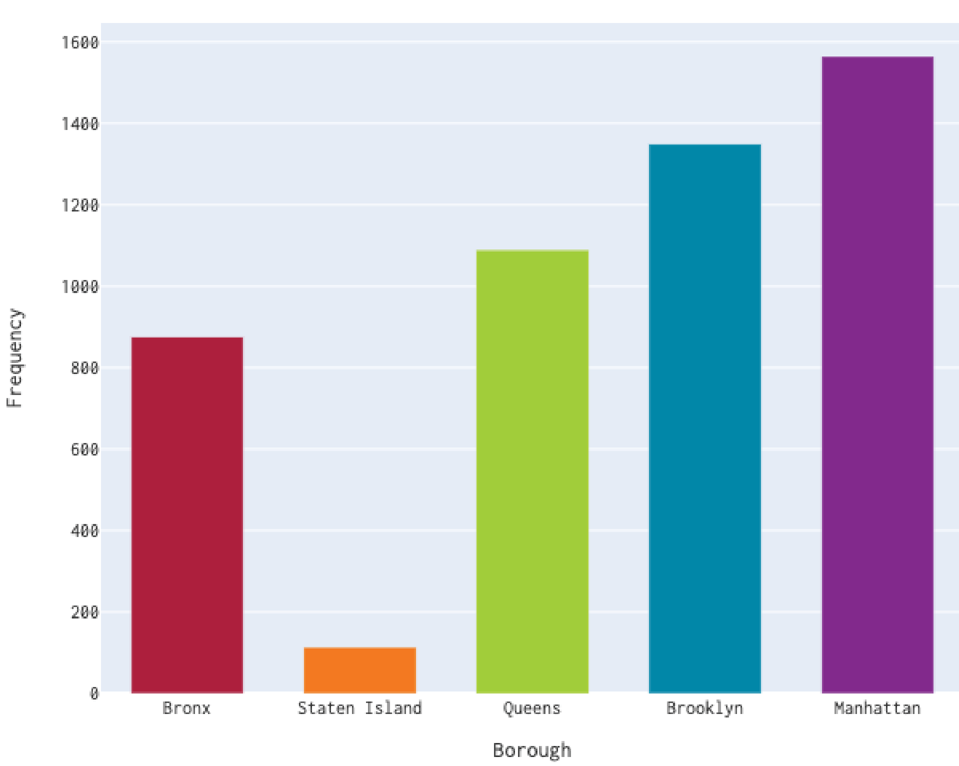
Detailed Process Flow

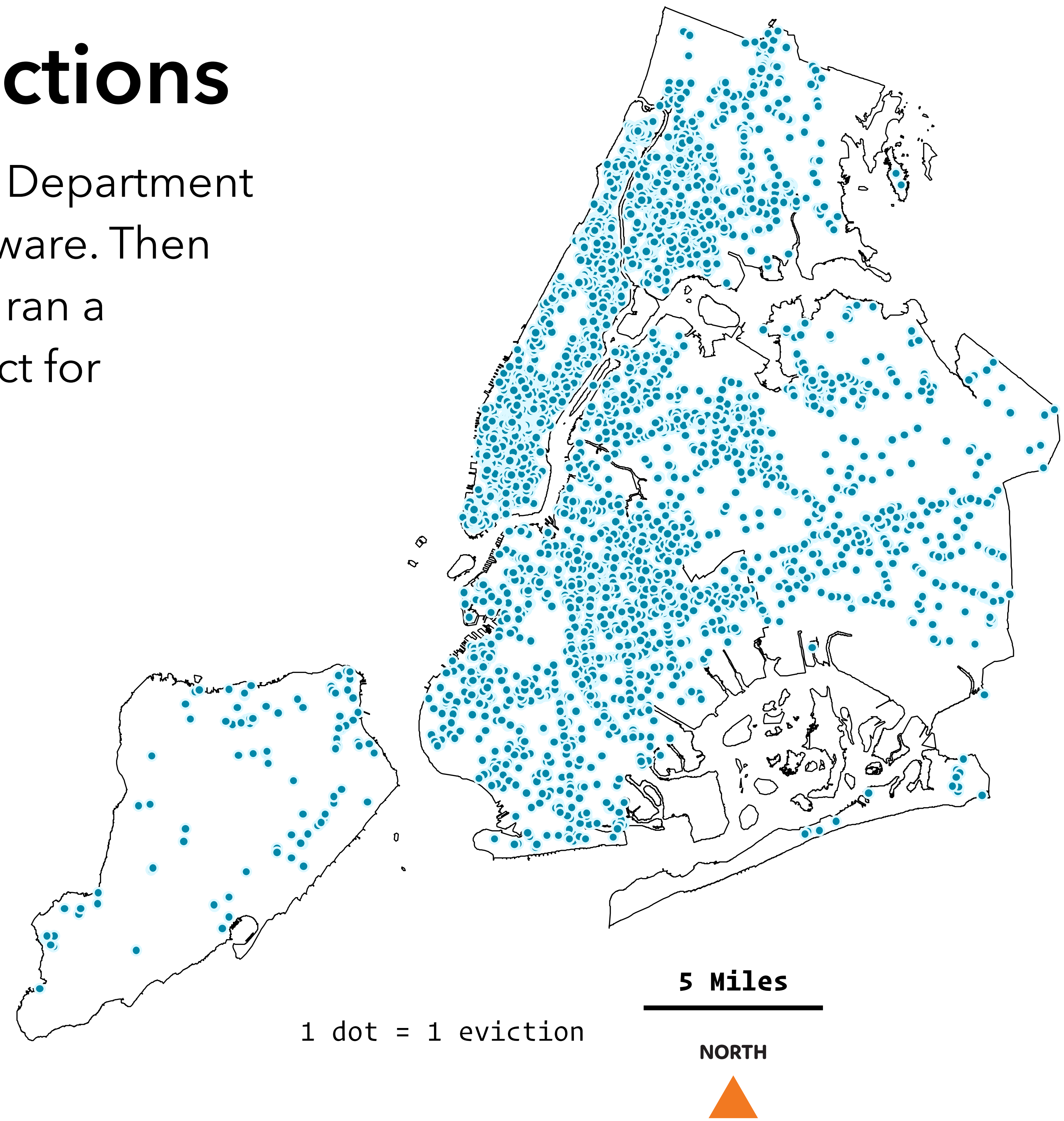


Geocoding the Evictions

I geocoded the evictions using the Department of City Planning's GeoSupport software. Then I situated the points on a map, and ran a spatial query to assign a census tract for each eviction.

NYC Evictions by Borough: 2017-Present





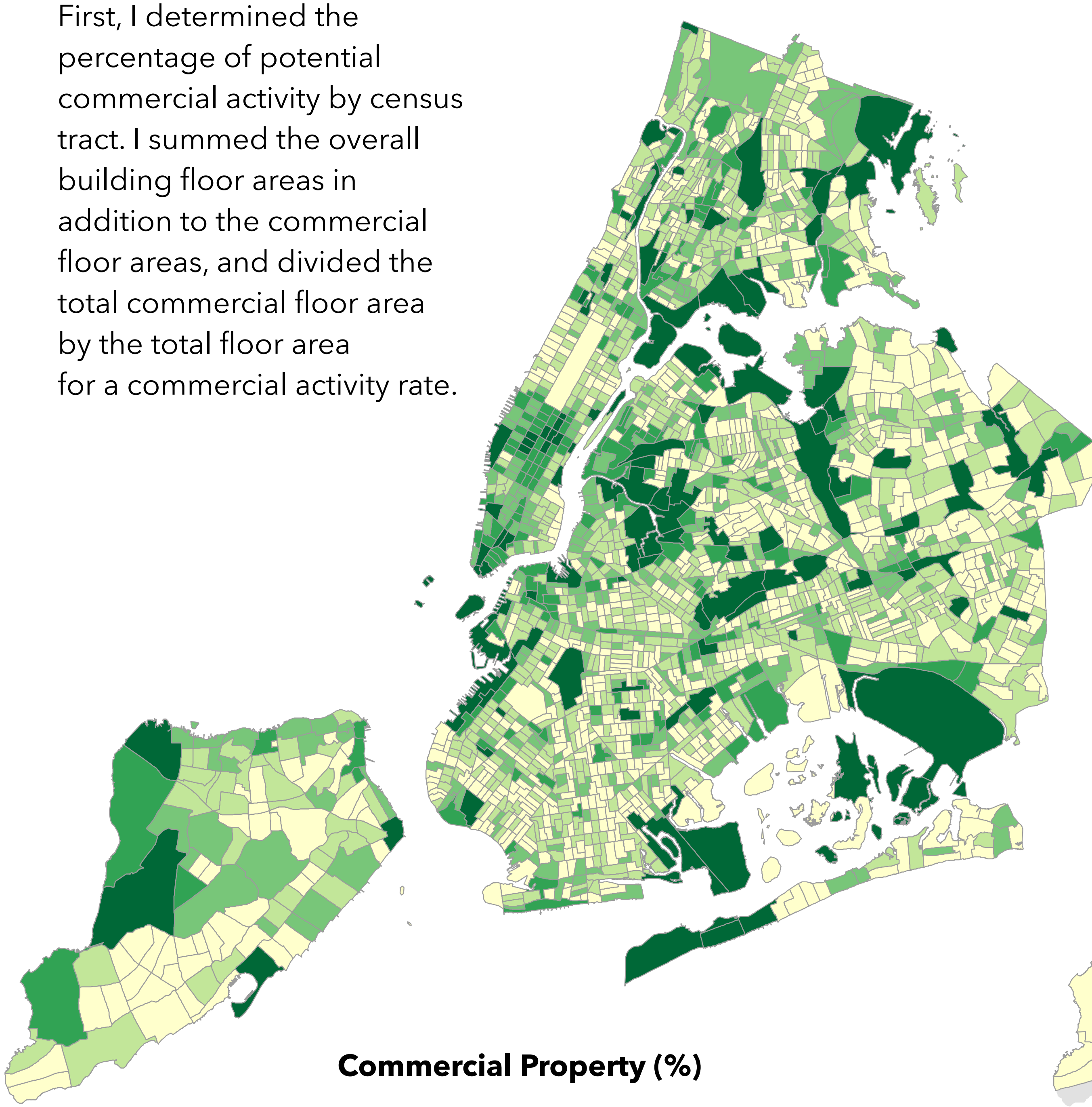
Source: Pending, Scheduled, and Executed Evictions, NYC Open Data, 2017-Present

Establishing the Study Area

5 Miles

NORTH

First, I determined the percentage of potential commercial activity by census tract. I summed the overall building floor areas in addition to the commercial floor areas, and divided the total commercial floor area by the total floor area for a commercial activity rate.



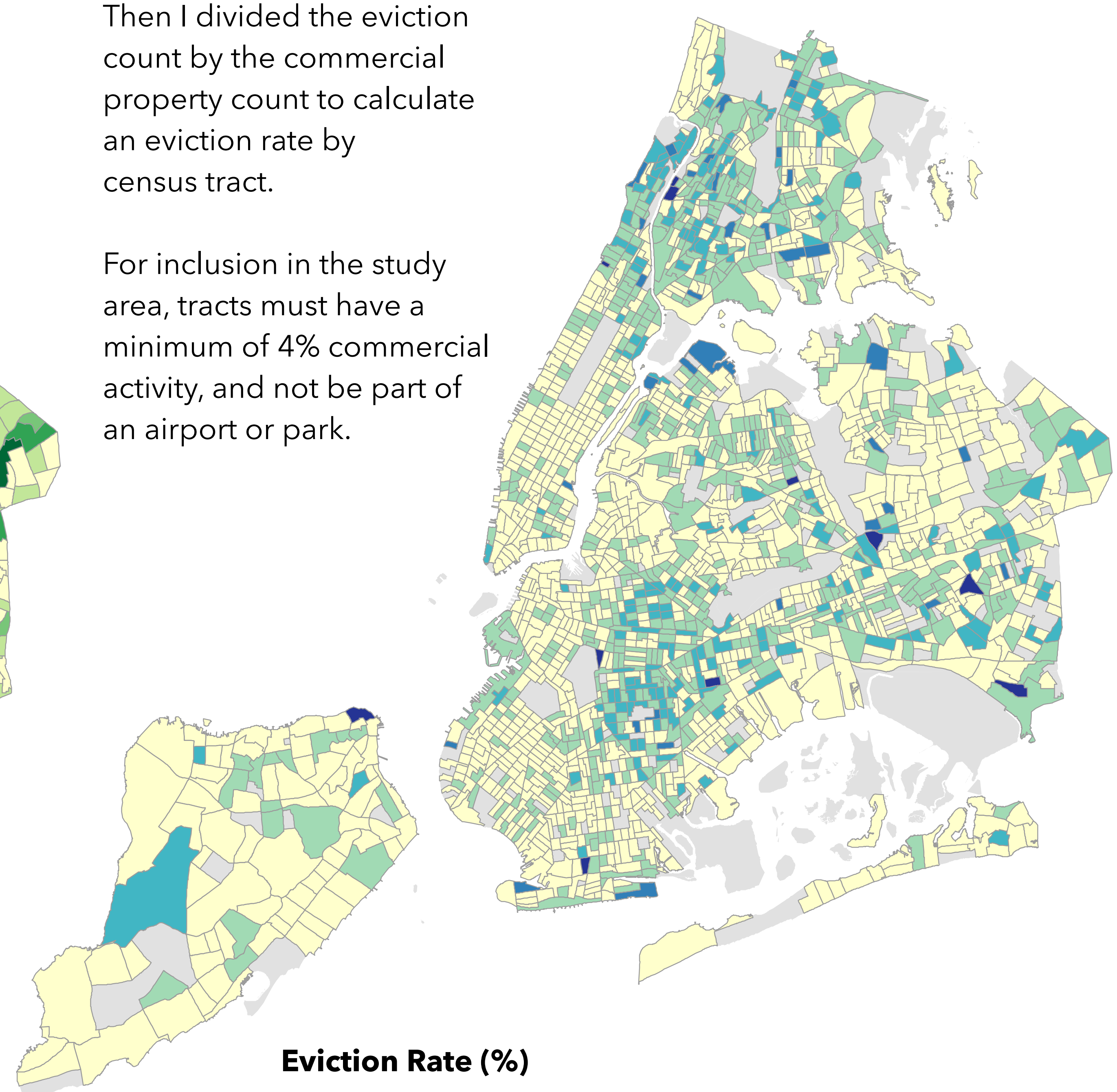
Commercial Property (%)

- 0 - 13.02
- 13.03 - 26.77
- 26.78 - 46.11
- 46.12 - 73.78
- 73.79 - 100

Source: PLUTO 19v2

Then I divided the eviction count by the commercial property count to calculate an eviction rate by census tract.

For inclusion in the study area, tracts must have a minimum of 4% commercial activity, and not be part of an airport or park.



Eviction Rate (%)

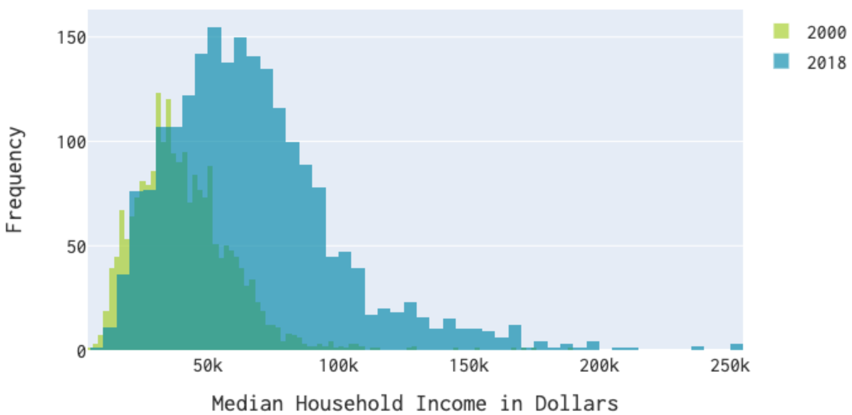
- 0 - 1.77
- 1.78 - 5.19
- 5.2 - 11.86
- 11.87 - 25.8
- 25.81 - 53.85
- = Excluded Areas

Sources: Pending, Scheduled, and Executed Evictions, NYC Open Data, 2017-Present

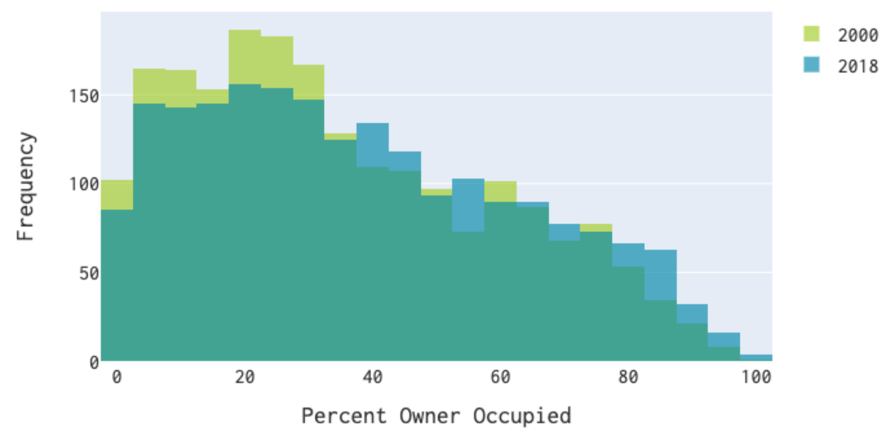
Selected Census Figures:
A 2000 to 2018 Comparison

Sources: Neighborhood Change Database,
2000 Census Reinterpolated to 2010 Boundaries;
American Community Survey, 2018

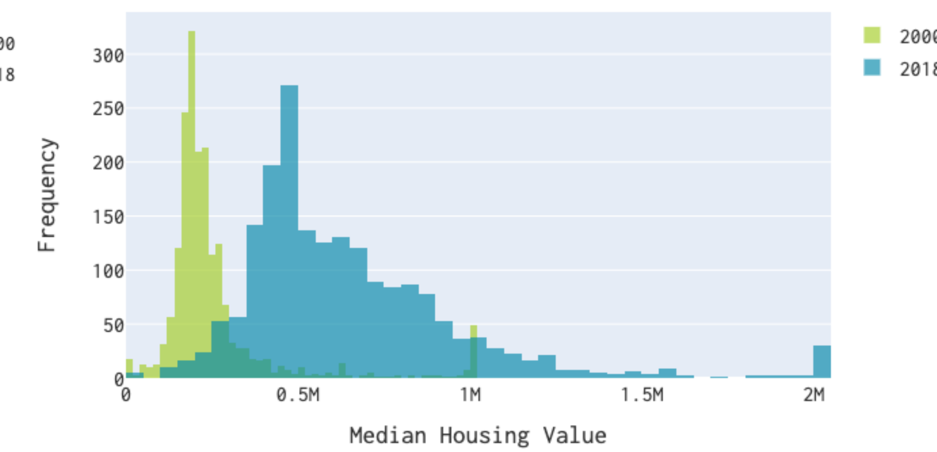
Median Household Income



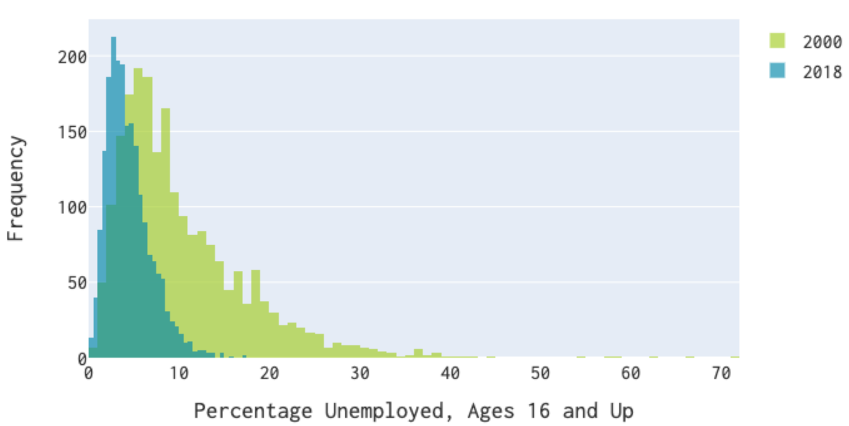
Percentage Owner Occupied



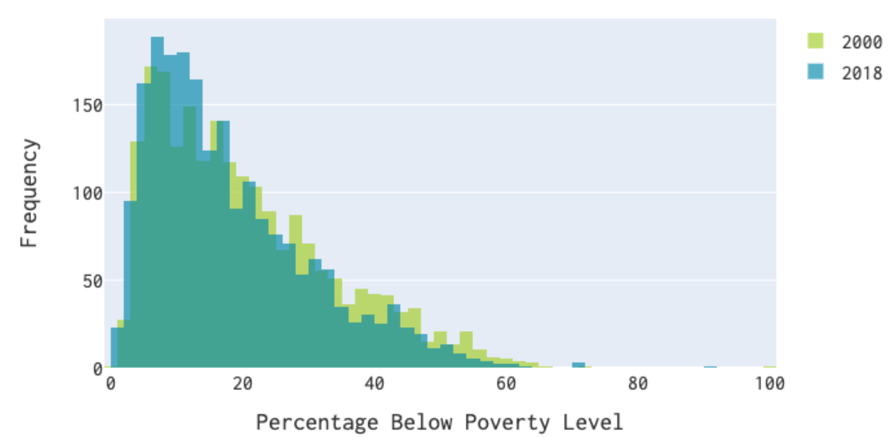
Median Housing Value



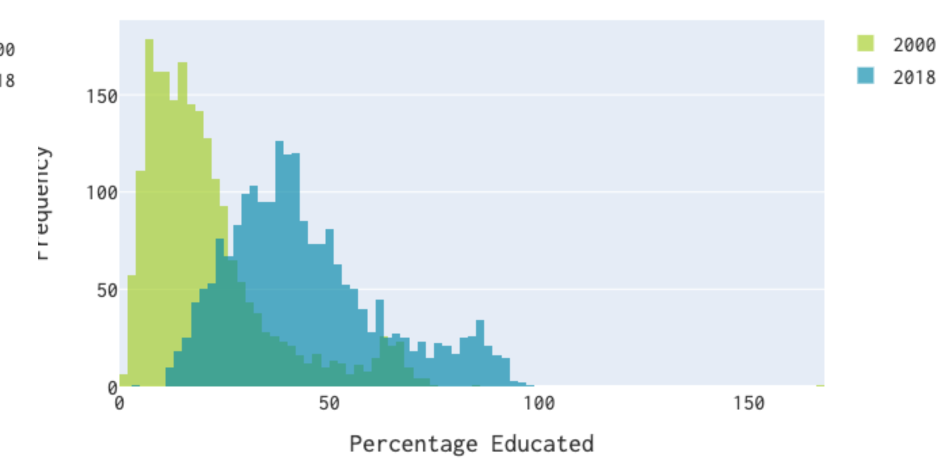
Percentage Unemployed, 16+



Percentage Below Poverty Level



Percentage Educated *



Each of these dual histograms shows the distribution of a census variable for 2000 (in lime green) and 2018 (in blue). The blue green areas show overlap.

Broadly, all of these indicators show "improvement." This is not true of all census tracts, and improvement is in the eye of the beholder. Nevertheless, the overall picture is not one of decline. An attempt to explain inequality would need to examine additional variables.

* Defined here as completion of an Associates Degree

Calculating the Census Tract Change Score

1. I calculated the z-Score for each variable, in both of the years 2000 and 2018. (A z-Score is the number of standard deviations a value is from the mean.)

2. Then I subtracted the z-Scores for each 2000 variable from the z-Scores for each 2018 variable.

3. Finally, I totaled the changes to the z-Scores as follows:

2000 to 2018 Neighborhood Change Score =

Change_in_zScore_Median_Household_Income +

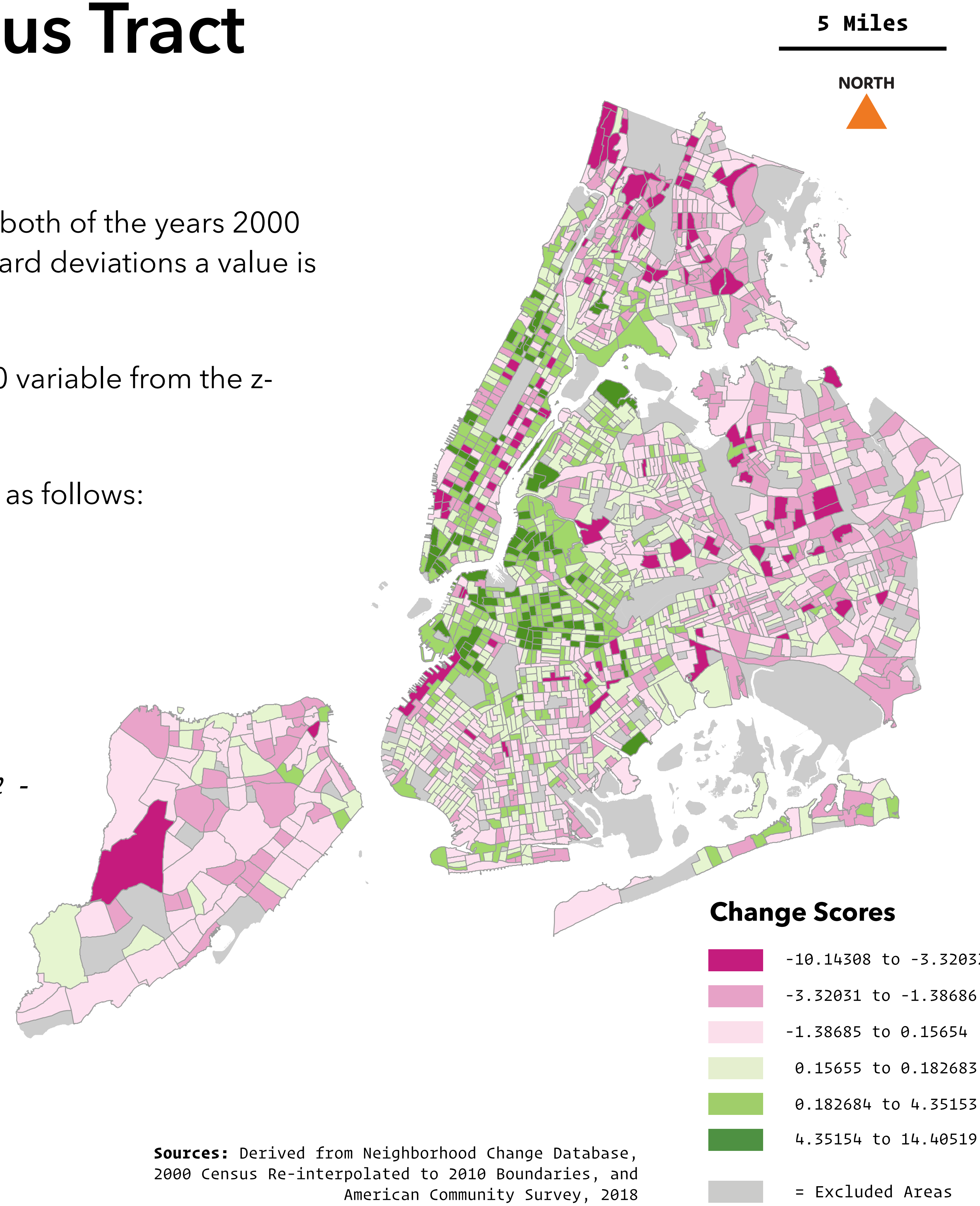
Change_in_zScore_Percent_Owner_Occupied +

Change_in_zScore_Median_Housing_Value +

Change_in_zScore_Percent_Educated -

Change_in_zScore_Percent_Below_Poverty_Level -

Change_in_zScore_Percent_Unemployed



Sources: Derived from Neighborhood Change Database, 2000 Census Re-interpolated to 2010 Boundaries, and American Community Survey, 2018

The results are mapped at right. It is important to understand that this is not a measure of overall neighborhood health, but a relative measure of change. Some of the areas in pink were and remain quite affluent by any reasonable standard.

The Results

Bivariate Regression Tests Predicting Eviction Rates

	Median Household Income	Percent Owner Occupied	Median Housing Value	Percentage Unemployed	Percentage Below Poverty Line	Percentage Educated	Change Score
Moran's I	-0.149	-0.101	-0.148	0.144	0.125	-0.137	-0.021
R-Squared	0.028944	0.023857	0.038335	0.035782	0.026000	0.020720	0.006390
Adj. R-Squared	0.028448	0.023358	0.037844	0.035290	0.025502	0.020220	0.005883
F-Statistic	58.3322	47.8282	78.0121	72.625	52.2403	41.4069	12.5866
Prob (F-Statistic)	3.44036E-14	6.2771E-12	2.22696E-18	3.08134E-17	7.00795E-13	1.55022E-10	0.000397683
AIC	10771.8	10782.1	10752.8	10758	10777.8	10788.4	10816.8
Coefficient	-1.92555E-05	-0.0241366	-2.05236E-06	0.294544	0.0503829	-0.0289136	-0.127834
P-Value	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00040

Above are the results of bivariate regression tests. The variables above are the independent variables, with the eviction rate being the dependent variable.

Moran's I is a measure of spatial autocorrelation that ranges between -1 and +1. Values nearer to -1 or +1 suggest spatial autocorrelation—perfect randomness or perfect clustering—while values nearer to zero suggest no clustering.

The Moran's I for the Change Score is very near zero, suggesting there is no clustering.

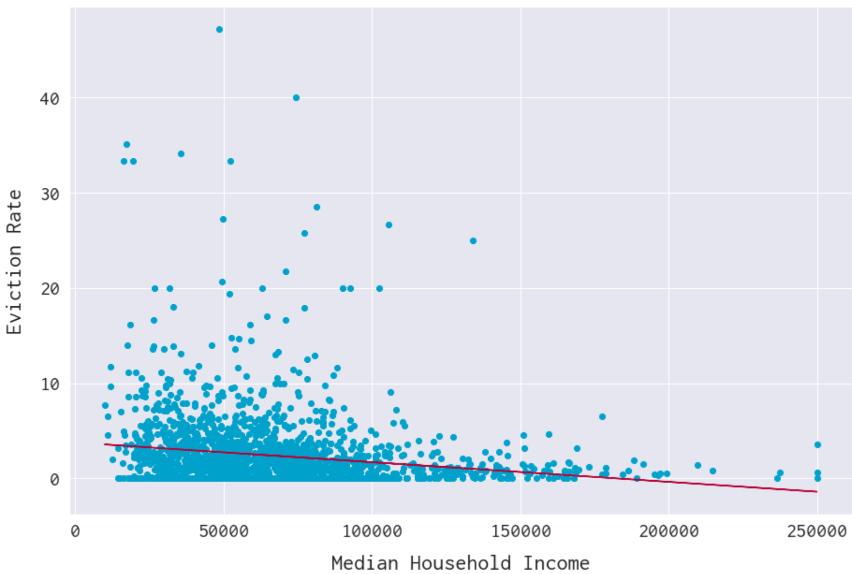
Other variables show moderately positive or negative Moran's I values, and so are perhaps somewhat spatially autocorrelated. All of them have p-values of 0, suggesting that the results are significant.

Median Household Income, Percentage Owner Occupied, Median Housing Value, and Percentage Educated have negative coefficients, and are therefore negatively correlated with eviction rates. Percentage Unemployed and Percentage Below Poverty Line have positive coefficients, and so are at least somewhat positively correlated with eviction rates.

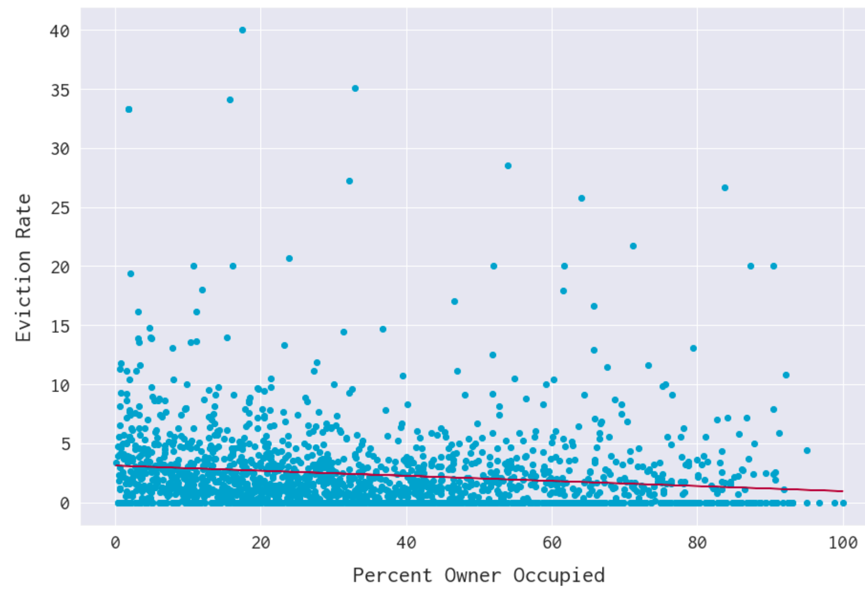
Bivariate Scatter Plots

Census variables independent; eviction rate dependent

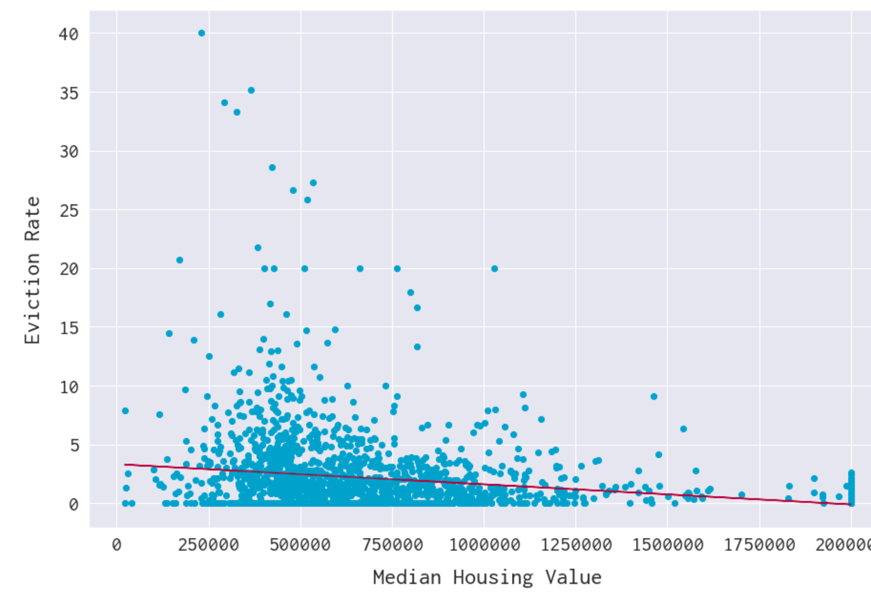
Median Household Income



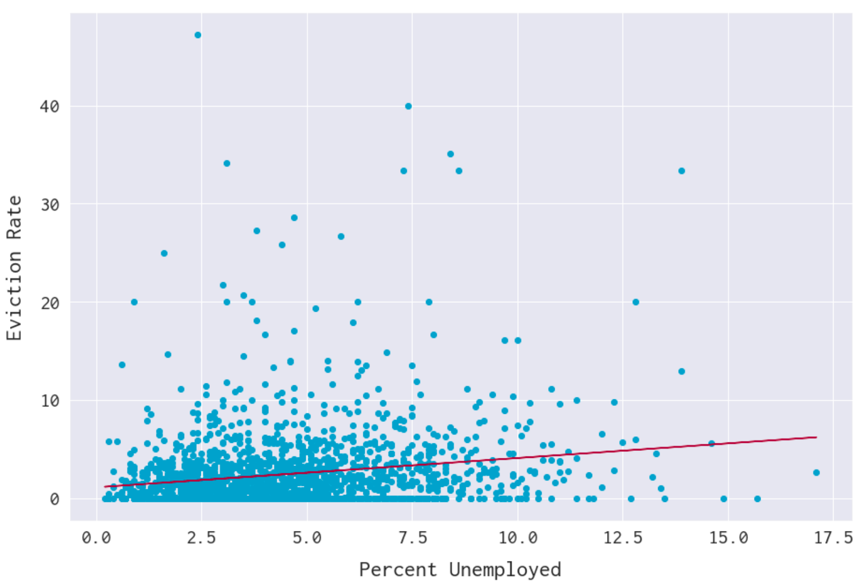
Percentage Owner Occupied



Median Housing Value

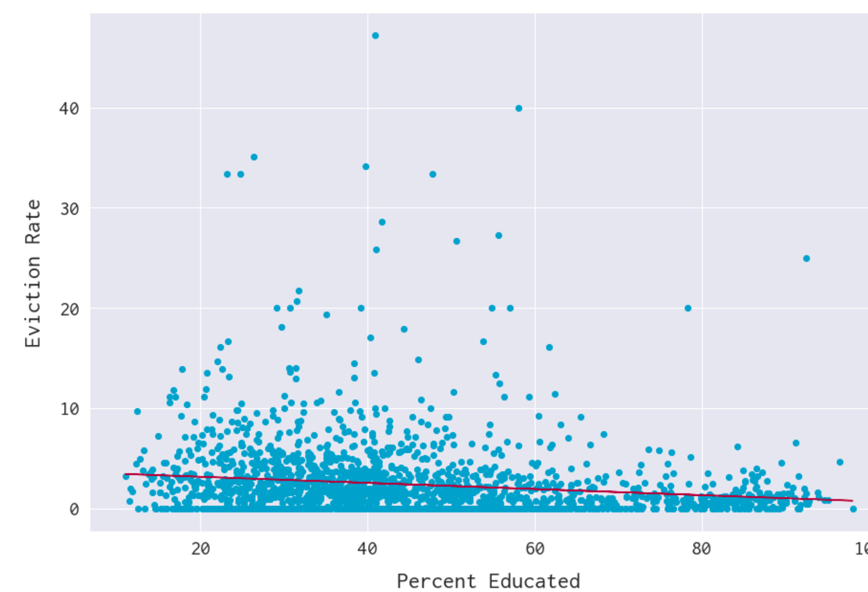


Percentage Unemployed, 16+



Percentage Below Poverty Level

Percentage Educated



A visual restatement of the previous slide. Unemployment percentage and poverty level seem most correlated.

Spatial Regression Model Results Predicting Eviction Rates (Queen Contiguity)

	OLS R2=0.068738 AIC=10699.9		Spatial Lag R2=0.083695 AIC=10679.9		Spatial Error R2=0.080643 AIC=10682.9	
	β	ρ	β	ρ	β	ρ
CONSTANT	2.84091	0.00000	2.36516	0.00000	2.89843	0.00000
Median HH Income	3.97928E-06	0.43860	4.22404E-06	0.40622	4.72851E-06	0.37505
Pct Owner Occupied	-0.0154723	0.00072	-0.0143003	0.00162	-0.0163042	0.00079
Median Housing Value	-1.63193E-06	0.00000	-1.47924E-06	0.00000	-1.56526E-06	0.00000
Pct Unemployed	0.192537	0.00000	0.168928	0.00001	0.172419	0.00001
Pct Below Poverty Line	0.00607723	0.56183	0.00574752	0.57950	0.00681107	0.52984
Pct Educated	-0.00337489	0.65958	-0.00219374	0.77253	-0.00420988	0.60512
Lag Coefficient (Rho)	-	-	0.163755	0.00000	-	-
Lag Coefficient (Lambda)	-	-	-	-	0.150332	0.00002
Moran's I	0.0569	0.00002	-	-	-	-

Since the Moran's I value for the change score did not suggest spatial autocorrelation, I eliminated it from the spatial autocorrelation tests. Above are the results for the remaining variables.

The Rho and Lambda tests for the spatial lag and spatial error tests are significant, suggesting there is spatial autocorrelation for the census variables. However, only Percent Owner Occupied, Median Housing Value, and Percentage Unemployed have p-values that indicate significance. Percent Owner Occupied and Median Housing Value are negatively correlated with eviction rates; Percentage Unemployed is positively correlated with eviction rates.

Spatial Lag has the lowest Akaike Info Criterion value, which suggests it is the model with the best fit. Likewise, it has the highest R-Squared value, also an indicator of fit.

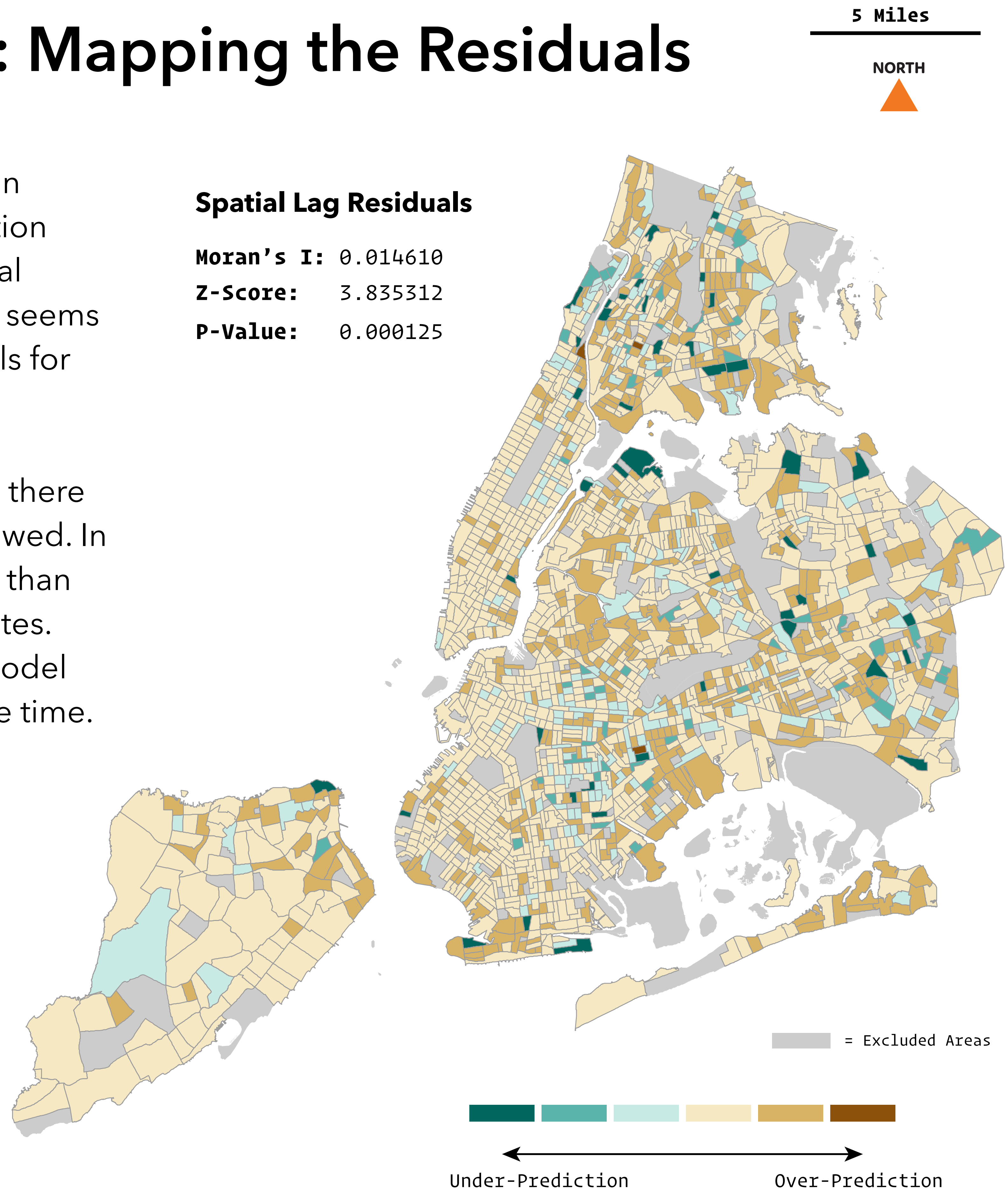
Assessing the Model: Mapping the Residuals

A residual is the difference between the observed value and the predicted value, in this case the difference between the eviction rate predicted by the model and the actual eviction rate. Since the Spatial Lag model seems like the best fit, I am mapping the residuals for that model here.

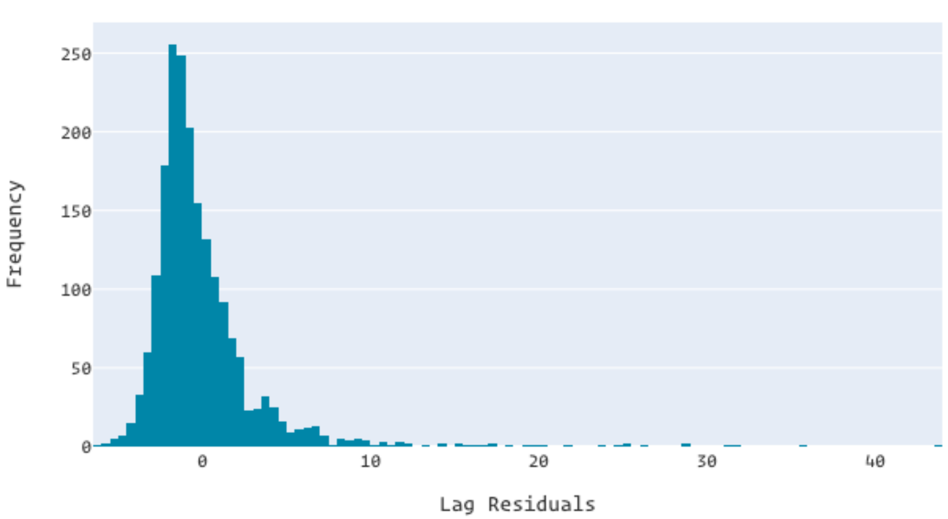
The Moran's I value of 0.014610 suggests there is some clustering, or that the model is flawed. In other words, there is some element other than those studied that predicts the eviction rates. The histogram below suggests that the model slightly over-predicts evictions most of the time.

Spatial Lag Residuals

Moran's I: 0.014610
Z-Score: 3.835312
P-Value: 0.000125



Lag Residuals Distribution



Conclusions

- There is no spatial autocorrelation between the neighborhood change index and the eviction rate.

- Other variables show some spatial autocorrelation with the eviction rate. Percentage owner occupied and median housing value are negatively correlated with the eviction rate. Percentage unemployed is positively correlated with the eviction rate. Results for the other variables were not significant.

- Judging from the R-squared and Akaike Info Criterion figures, the spatial lag model was the best fit. But the Moran's I for the residuals indicated clustering, suggesting that even this model was flawed. The model performs reasonably well in Manhattan below 96th Street, as well as much of Staten Island, but less well in the Bronx, Northern Brooklyn, Eastern Queens, Upper Manhattan, and Astoria.

- Explaining commercial eviction rates would require the examination of additional variables. A future project could attempt to identify and explore these variables using similar techniques.

Study Limitations

- Included in the commercial eviction records are storage unit and parking garage evictions. While these are indicators of economic stress, they weren't really what I was looking to study. I compromised by counting all of the evictions at a single parking garage or storage space as one eviction.

- This study uses freely available sources, such as PLUTO, for its supporting data. Due to competing priorities at different agencies, PLUTO is known to have some issues with accuracy (I interned at DCP), although it's unclear how much those affect this study. This is how I used PLUTO:

    - In order to do a study focusing on commercial evictions, it is necessary to ascertain the level of commercial activity in the unit of analysis. A single eviction in a tract with a small number of commercial properties will have a high eviction rate that is essentially meaningless, and these tracts needed to be excluded from the study. But the amount of commercial activity per census tract is hard to ascertain, at least with freely available data sources. PLUTO contains a commercial floor area field in addition an overall floor area field which I used as proxies for commercial activity. The derived commercial rate is therefore only as valid as the supporting data. To assess the level of commercial rent pressures, the NYC City Council has mandated that a "storefront tracker" database be created that would require the registration of all commercial properties (Surico, 2019). Needless to say, the existence of such a database really has the potential to improve the accuracy of studies like this one.

    - Likewise, the eviction rate was determined by dividing the eviction count per census tract by the number of commercial properties for that census tract, also obtained from PLUTO. Again, this result can only be as good as its inputs.

Sources

Mallach, Alan. *The divided city: Poverty and prosperity in urban America* [Kindle version]. Retrieved from Amazon.com.

Stein, Samuel. *Capital city: Gentrification and the real estate state [Kindle* version]. Retrieved from Amazon.com.

Surico, John. (2019, August 1). *Are small businesses really fleeing New York? This tool can tell.* CityLab. Retrieved from https://www.citylab.com/life/2019/08/vacant-storefront-tracker-law-nyc-retail-rent-control/595294/.