Residential Permits Issuance and 311 Building Violations complaints

Dana Chermesh, federica B bianco

https://github.com/danachermesh/PUI2017_dcr346
Instructor: **Dr. Federica Bianco**

Abstract:

This study sought to analyze the possible correlation between the number of residential permits issuance and buildings violations, represented by 311 building related complaints. Using several sources of data including Census Bureau, NYC open data, 311 and NYC spatial data, a descriptive analysis and regression models were conducted to better understand the two urban factors. The results were insignificant, which not necessarily mean the correlation between the two isn't exist, rather than different or further methods could have better explain it. A more meaningful negative correlation was detected between renter-occupied housing units and BV complaints.

Introduction:

New York City is a rapidly renewing urban area, with an escalating demand for housing and increasing housing costs. My motivation for this research was to analyze how does the number of residential permits issuance correlate with 311 complaints related to buildings violations, if at all, and by this, ideally, to identify areas that should get more attention regarding building codes and building use validation. I relate to permit issuance as an indicator of urban renewal, due to the fact that the majority of construction in New York City requires a Department of Buildings permit, this to make sure that the plans are in compliance with building code. I pre assumed that an area with a very low number of residential permits issued over a year (meaning, an area that is less developing / renewing) will also show a relatively large number of building violation. I also guessed there are highly renewing areas with large number of permits issued and a large number of building violation complaints. Additionally, I was interested in assessing the role of renter / owner occupancy ratio on building violations complaints.

Data:

This research rely on data from year 2016, and focus on residential information only. Any personal information was

Dana Chermesh is with NYU Center for Urban Science & Progress
federica B bianco is with NYU Center for Urban Science & Progress

excluded. The analysis was performed in the granularity level of Zip codes, which seemed a reasonable geographical unit to observe urban renewal trends.

The study required the use of several data sources. First, Permit Issuance data were obtained from **DOB permits issuance open data** and were cleaned to include Residential permits only, then were filtered again to include only **New Buildings** (NB) and **massive Alternation** (AL) permit types, ignoring plumbing, signs, equipment etc. permit types, that are insignificant to this research. The permits data were normalized by the overall number of occupied housing units, obtained from the US Census Bureau, American Fact Finder website, using the ACS 5 years estimate data. Data for 2016 do not exist in the zip code geographical level; for that reason I used from year 2013, assuming the change in the number of housing units is not meaningful. All data were grouped by zip code to count the number of permits issued in each zip code in 2016.

The data of the **Department of Buildings (DOB) violations** are divided to more than a hundred complaint categories, most of them are meaningless to urban renewal. It was hard to define the exact categories that will best contribute to this analysis. In order to avoid misinterpretations, **311 complaints** data were selected instead. The 311 data were filtered to include only **Building** related complaints in year 2016. The 311 complaints are also divided by complaint descriptor; the descriptors were included in this analysis are:

- Illegal Conversion Of Residential Building/Space
- Illegal Commercial Use In Resident Zone
- Zoning - Non-Conforming/Illegal Vehicle Storage
- No Certificate Of Occupancy/Illegal/Contrary To CO
- SRO - Illegal Work/No Permit/Change In Occupancy/Use
- ROOFING
- PORCH/BALCONY
- SKYLIGHT
- GUTTER/LEADER
- FENCING

311 is a relatively new citizens-city engagement system, of which not all citizens are taking advantage or aware. To overcome this bias the 311 data were normalized by dividing each zip code's number of building-related complaints by its overall number of 311 complaints. Due to the large size of the 311 dataset the overall number of complaints in 2016 was assessed by extracting two months only from that year, January and June, proxies for winter overall complaints and summer overall complaints respectively (seeIpython notebook). The data were grouped by zip code to count their number of

building violation complaints. The weakness of the 311 data is that even when neutralizing the bias in the citizens' use of the 311 system, it is harder to address the differences between citizens' involvement and engagement level in the city and their feeling about complaining.

Additionally, for the second part of the analysis, data of number of renter-occupied housing units and owner-occupied housing units were obtained, also from the American Fact finder website. Finally, New York City's zip codes shapefiles were included, also obtained from NYC open data website.

Methodology:

The first step of the analysis was to observe and describe the data of both primarily variables, Permit Issuance and 311 Building Violation complaints. The distribution of the normalized variables were viewed to assess the possible similarity in their statistical behaviour:
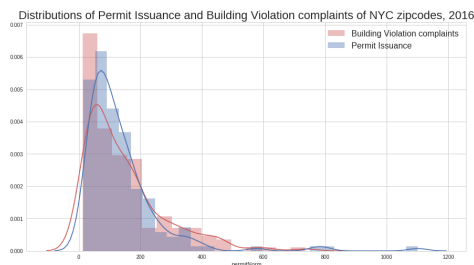

Figure 1. **Distribution of Permit Issuance and Building Violation complaints of NYC zip codes, 2016** The distributions of the two variables seem to be not very different. Performing AD test inferred that the distributions do not follow a Gaussian distribution

Then, a bar plot of the two overlayered variables by all zip codes was observed. The graph didn't give a further understanding of the data nor the possible relationship between the variables and its magnitude or direction.
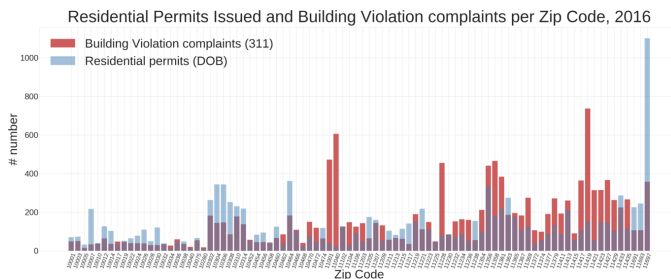

Figure 2. **Residential Permits Issued and Building Violation complaints per zip code, 2016** The bar plot of the two variables does not give much information about any possible relationship between them.

A linear model was then built to assess the power of permit issuance with predicting building violation complaints. As shown in the model summary below, the R-squared is extremely low, although the p-value of the regressor is below the significance threshold which imply for its existing impact on Building Violations complaints. Figure 4 is a scatter of the data including the fitted line.

OLS Regression Results

Dep. Variable:	BVnorm	R-squared:	0.059
Model:	OLS	Adj. R-squared:	0.054
Method:	Least Squares	F-statistic:	10.94
Date:	Thu, 14 Dec 2017	Prob (F-statistic):	0.00114
Time:	17:21:16	Log-Likelihood:	-1092.8
No. Observations:	175	AIC:	2190.
Df Residuals:	173	BIC:	2196.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]
Intercept	114.0167	13.572	8.401	0.000	87.229 140.804
permitNorm	0.2336	0.071	3.308	0.001	0.094 0.373

Omnibus:	74.918	Durbin-Watson:	0.927
Prob(Omnibus):	0.000	Jarque-Bera (JB):	207.932
Skew:	1.835	Prob(JB):	7.05e-46
Kurtosis:	6.879	Cond. No.	275.

Figure 3. **A linear regression model of Building violation complaints over Residential permits issuance**
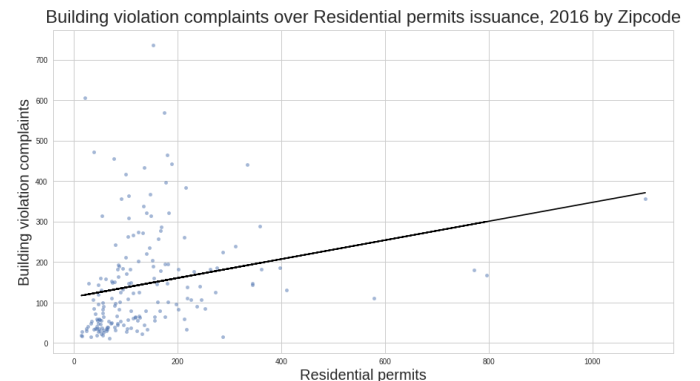

Figure 4. **Building Violation complaints over Residential permit issuance + fitted line** Permit Issuance linear regression model has very low predictive power on number of Building Violation. The permits' apparent outliers seem to have a high impact on the fitted line.

The next phase was to detect the outliers of the data, setting the threshold at two standard deviations from the mean. The results were interesting and surprising, having the five outliers with the highest number of permits issued

being zip codes in **Staten Island** and **Rockaway, Queens**. Similar results were obtained in the building violations complaints sense. Understanding that in each of NYC borough the analyzed phenomena might perform differently, I re-conducted the analysis on a shortened dataset including Manhattan and Brooklyn only. The results of the new linear models were even less significant than the whole city's, with **R-squared** of **0.047** in **Brooklyn** and an insignificant p-value and **R-squared** of **0.022** for **Manhattan** and even a less significant p-value.

Building violation complaints over Renter-occupied Housing units, 2016 by Zipcode

Figure 6. **Building violation complaints over renter-occupied housing units % + fitted line** The fitted line seems to explain the data in some level with R-square of 0.379. The correlation is negative which means the higher the percentage of renter-occupied housing units the lower the number of 311 complaints regarding building violations.
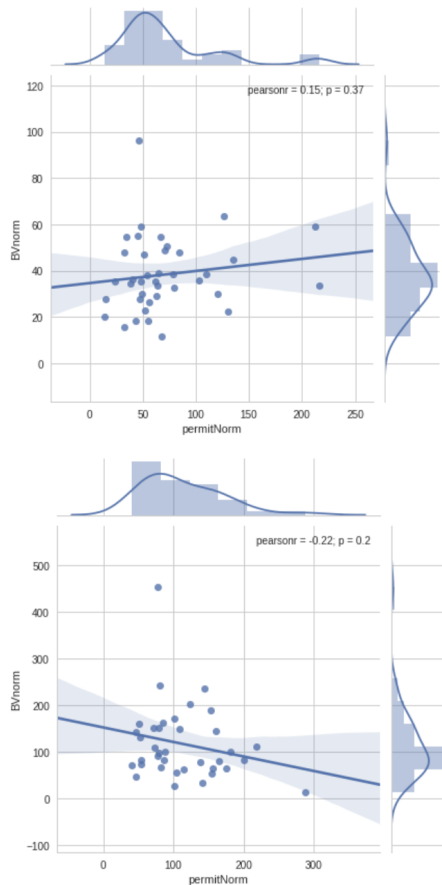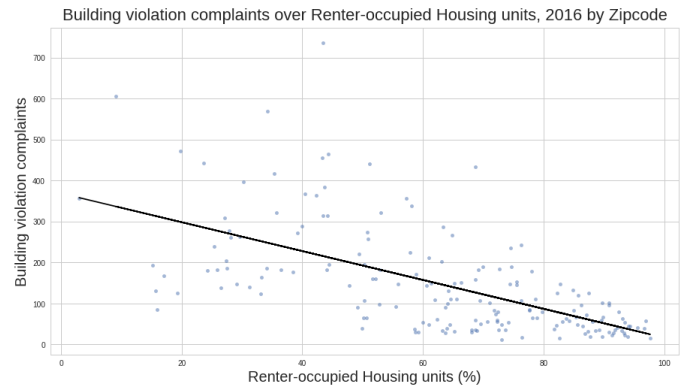

Figure 5. **Building Violation complaints over Residential permit issuance, Manhattan only (upper) and Brooklyn only (lower)** The fitted lines nor the confidence areas do not seem to explain the data sufficiently. The direction of the correlation between the two boroughs selected is opposite, with Manhattan revealed as possitive (extremely low) correlation and Brooklyn having negative (extremely low) correlation.
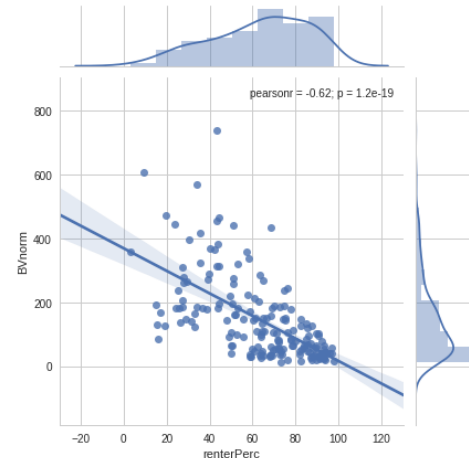

Figure 7. **Building violation complaints over renter-occupied housing units % + Fitted line + confidence level** The fitted line seems to explain the data in some level with R-square of 0.379. The correlation is negative which means the higher the percentage of renter-occupied housing units the lower the number of 311 complaints regarding building violations.

To consider a different possible affective variable on building violations complaints a new linear model was built using renter-occupied housing units percentage as the regressor. The results were much more compelling, with **R-squared** of **0.379** and a **negative correlation** with **-3.52 slope**, as shown in Figures 6,7:

Lastly, a spatial view of the analyzed variables was observed, revealing some possible spatial clustering of 311 building violations complaints, a less clear spatial distribution of permit issuance and an obvious spatial clustering of renter-occupied housing unit percentage. The outliers detected in the first part of the analysis are clearly observed, too.

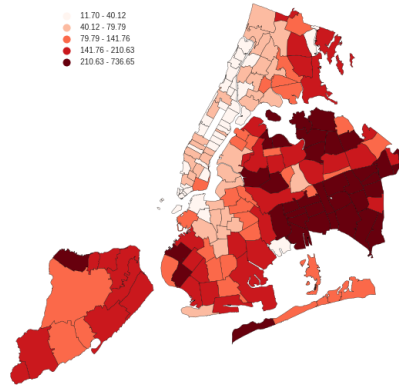NYC Building Violation complaints by zip code, 2016

NYC Renter-occupied Housing units % by zip code, 2016

Figure 8.
Building Violation complaints 2016
of 311 Building Violation complaints of NYC zip codes, year 2016. The darker the color the more complaints received. The data seem to be clustered at some level, with South and North Queens with the highest number of BV complaints.

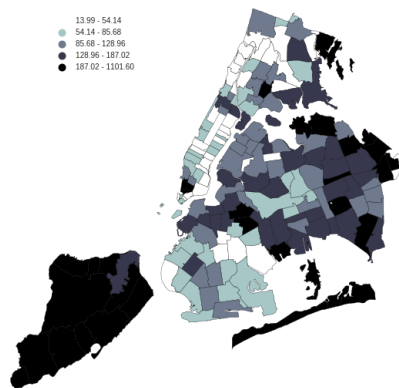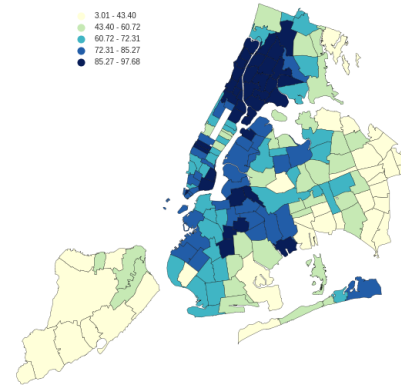NYC Residential Permit Issuance by zip code, 2016

Figure 10. **NYC**
occupied Housing units %, 2016   Map
of Renter-occupied Housing units percentage, by zip codes, year 2016. The blue colors stand for high percentage of renter-occupied housing units, while light green-yellow colors stand for the opposite. The map infer the data is spatially clustered, with high rate of renter-occupied HU in Harlem and the Bronx and most of Brooklyn. Medium rate of renter-occupied HU is less clustered.

One limitation of this research is the use of 311 complaints regarding building violations as representative for actual building / use violations, which are not compliant. Furthermore, as discussed in the data section, it is hard to assess the willingness to complaint about citizens issues across the city. Further work could analyze **DOB approved violations** rather that 311 data.

Another possibly powerful approach that could have made this study more accurate is to assess the housing units-per-permit, and by this to weight the affect of each permit on the city as a whole. This could prevent areas with single family houses such as Staten Island to be mistakenly revealed as significantly renewing areas.

Also, spatial autocorrelation tests such as Global and local Moran's I should be included in further research for all considered variables.   Map

To conclude, renter/owner occupancy percentage has some predictive power of building violations 311 complaints, as opposed to permit issuance as was analyzed in this study.

Figure 9.
Residential Permit Issuance 2016
of Residential permits issued by DOB, by zip codes. The darker the color the more permits issued. The data don't reveal an obvious clustering, although Staten Island and Rockaway zip codes have significantly higher number of permits than the rest of the city.

Conclusions and Limitations*:*

The initial analysis' results were not significant enough to consider permit issuance as a reliable predictor for building violations complaints. The assumption according to which higher number of permits issued will result in less building violation complaints did not approved in this research. Even when dividing the data into two sub samples of Manhattan and Brooklyn, there was no consistency detected in the behaviour of the variables nor their relationship.