12/9/2016

Housing Code Violations in the City of Providence

David Reich
Sociology 1340, Brown University

An exploration of spatial patterns in the quality of the residential built environment

Introduction

'Blight' is an important term in the fields of urban studies and city planning. Accusations of blight in the 1950s and 60s facilitated the destruction and disintegration of city neighborhoods across the United States during the period of "urban renewal". Even today, the term is powerful. In Boston, for example, the term 'blight' carries a heavy legal weight: Brian Golden, director of the Boston Planning and Development Agency (formerly the Boston Redevelopment Agency, or BRA), has said that "blight is the key which unlocks the tools,"[1] meaning that it is the legal designation that, when applied to various neighborhoods, empowers city governments to take more extreme actions. However, 'blight' is a prescriptive, rather than descriptive, designation. It is often applied to neighborhoods on the basis of political expediency rather than empirical fact. The term is, perhaps, only a political tool, and as such is susceptible to abuse. A rigorous definition of blight would enforce a more rigorous approach to city planning and would provide a grounding for many important conversations.

Blight, however, is difficult to define. In his foundational paper of 1967, G. E. Breger describes urban blight in terms of three unifying concepts: depreciation, real property, and nonacceptance[2]. Blight, for Breger, is the depreciation of the conditions of real property from a state of acceptance to one of nonacceptance. As such, it is a process rather than a condition, and not necessarily a linear one. Breger indicates that the physical condition of the built environment, "the capacity of realty to render service," is the critical element. He does not, however, propose a quantitative technique for measuring or estimating blight.

Weaver and Bagchi-Sen, 46 years later, proposed a quantitative implementation of Breger's theory. They selected housing code violations as an approach to the problem. A

---

[1] https://www.bostonglobe.com/business/2015/09/16/blight/UhOLZdHdL9A9ms1qPiKXAK/story.html

[2] Breger, G. E., (1967), The Concept and Causes of Urban Blight, Land Economics, 43, issue 4, p. 369-376, http://EconPapers.repec.org/RePEc:uwp:landec:v:43:y:1967:i:4:p:369-376.

complex, interdependent process as blight is impossible to reduce fully to a single variable without loss of information. Nonetheless Weaver and Bagchi-Sen hold that certain measurements, such as housing code violations, are more useful than others, are theoretically motivated, and constitute a metric that allows research to move forward, if imperfectly, in the face of a difficult problem. Housing code violations are “physical manifestations of urban decline” that, unlike the subjective appearance of a neighborhood, can be measured quantitatively. Weaver and Bagchi-Sen posited that the ratio of housing code violations to total units of housing, at a given geographic area and level, can be used to describe the blight of that area[3]. In fact, they named this measure “blight per unit” or BPU..

Conducting research based on housing code violations is useful in two ways. First, it can reveal spatial information about blight in the city, allowing the visualization of clustering. High BPU measurements will generally, though not always, correspond with the areas considered ‘blighted’ by city officials. Furthermore, housing code enforcement is known to be an effective strategy for preventing abandonment[4], and is among the most significant pieces of information for predicting it. Second, and more subtly, it can reveal information about the city government’s creation and enforcement of the housing code laws themselves - and can provide feedback to those practices. It is unclear, from easily-accessible public resources, how the City of Providence enforces housing code violations. It seems that inspections may be only on the basis of violation reports[5]. If this is the case, and even if not, and if research based on housing code violations returns an unexpected result, it may indicate an issue in the procedures for investigating housing code violations. It may, therefore, indicate unaddressed conditions in some neighborhoods or over-enforcement in others.

---

[3] R.C. Weaver, Sharmistha Bagchi-Sen, Spatial analysis of urban decline: The geography of blight, Applied Geography, Volume 40, June 2013, Pages 61-70, ISSN 0143-6228, http://dx.doi.org/10.1016/j.apgeog.2013.01.011.
[4] Yin, Li; Silverman, Robert M. 2015. "Housing Abandonment and Demolition: Exploring the Use of Micro-Level and Multi-Year Models." ISPRS Int. J. Geo-Inf. 4, no. 3: 1184-1200. http://www.mdpi.com/2220-9964/4/3/1184
[5] https://www.providenceri.com/DIS/code-enforcement

From this theoretical background, it is reasonable to continue under the principle that housing code violations provide a useful, though incomplete, proxy for blight, and that they will provide spatial data with various clear and subtle interpretations.

This paper will describe the data and procedures used for a spatial analysis of housing code violations in Providence, and will discuss the possible interpretations and future directions of such analysis.

Data

The data for this project was acquired by Professor Sungu-Eryilmaz from the City of Providence through the OpenPVD initiative for transparency and the sharing of data by the municipality. However, although the data were available, they were presented as-is, without interpretation or explanation. The data seemed to represent all recorded housing code violations for the period of 2006-2016, although it is possible that the set was not complete. For each violation, a wide range of data was included, though not all fields were filled entirely. There were 60,555 violations included in the dataset.

Each violation was given a complete address, for the most part in a useable format, including the neighborhood. Identification numbers were provided for the violation, the inspection, the notice of violation, the lien, and the building. A date of ownership was provided. The violation was listed as “code”, “abated”, ”pending”, or “reissued”, and was assigned a type. However, they type categories seemed to reference different Rhode Island and Providence legal documents, and were entirely unclear; they did not seem to align with categorizations in the published laws easily found online. Analysis of this categorization field will probably be necessary for a truly rigorous analysis of this data. A brief English description of each violation was also provided. This field seems to have been hand-entered, and is neither complete nor systematized. Additionally, data were provided about the date of inspection, the notice of violation, and the lien, if any, placed on the property.

Although the dataset was remarkably large and very useful for studying blight through housing code violations, there were nonetheless several possible issues. First, it was unclear how exactly these data were acquired by the City of Providence. That is, their inspection and reporting procedures were not clarified. If, as is perhaps the case, the inspections were conducted on the basis of violations reported to the City by tenants or neighbors, then there is certainly a sampling bias that hinders the interpretation of data. In fact, there are a number of possible sampling biases. For example, certain populations may be less likely to report violations. These populations include very low-income renters who are concerned with alienating their landlords; temporary residents, like college students, with little interest in the property's future after their departure; homeowners, who are uninterested in calling the Department of Inspection and Standards on themselves; and groups of people, perhaps recent immigrants, who lack the knowledge that violations can be reported. Therefore, an analysis of blight using housing code violations as a proxy will underestimate levels of blight in neighborhoods where such groups are concentrated. This could conceivably be corrected for statistically, but it would be difficult to do so in a robust and objective way.

The dataset was relatively amenable to geocoding. The procedure used the Brown University geocoding resource, and all violations with a "match" score of less than 87.5 were replaced[6], as were all violations that had been located outside of Providence city limits. 57,728 of the original 60,555 violations remained, a loss of less than 5% of the data.

Other data for the project included U.S. census data, at the block group level, for Providence. It was collected through the SimplyMap interface[7]. Data was chosen for 2010,

---

[6] As in Weaver and Bagchi-Sen

[7] http://geographicresearch.com/simplymap/

both to provide access to the decennial census and to fall close to the middle of the range of code violation data. It should be indicated, however, that this data is six years out-of-date and therefore may not be entirely reliable for the project's purposes. Providence has changed significantly in that time, and although the overall trends are likely still to be relevant, details of the results may not be entirely accurate.

Analysis and Results

The spatial analysis of the data was composed of four parts. The first was the investigation of spatial autocorrelation of housing code violations within the city of Providence. The second was the investigation of local autocorrelation and clustering behavior of these data. The third was a spatial-statistical regression to identify and quantify the most important values for predicting a block group's BPU from other variables. The fourth was a spatial, though more qualitative, analysis of sub-categories of housing code violation, seeking to investigate whether certain neighborhoods were disproportionately prone to certain types of violation.

1 - Global autocorrelation

To begin the project, BPU was calculated for each of Providence's 154 block groups (see appendix, fig. 1, for a map). For each block group, BPU was set as the ratio of housing code violations falling within its borders to its total number of housing units. A global Moran's *i* was run on those data, using an inverse distance conceptualization of spatial relationships and a Euclidean distance method. The global Moran's index was 0.52365, with a p-value less than 0.000001 and a variance of 0.001821. Clearly the data are spatially autocorrelated.

2 - Clustering

After establishing the spatial autocorrelation of the BPU data, the next step was to identify clustering behavior - to note which regions exhibit significantly higher- or lower-than-average BPU. This part of the project identified the neighborhoods that are under strongest need of additional support by the City of Providence, not only in enforcement of the housing code, but in structural changes to better support economic development and prosperity. A local Moran's *i* was selected on the basis of its use in the literature[8,9].

The assessment of spatial clustering was done by running an Anselin's local Moran's *i* on the same block group data, this time using a first-order Queen's contiguity conceptualization of spatial relationships. It resulted in several clear clusters of BPU behavior. Most of the East Side, as well as Downtown Providence except the Jewelry District, formed a large low-low cluster - a region of block groups with low BPU measurements surrounded by like block groups.. A high-high cluster took up most of South Providence, with another smaller such cluster south of Atwells Avenue in Olneyville. There were two outliers. A low-high outlier (that is, a block group of unusually low BPU, surrounded by block groups of unusually high BPU) was found in the middle of the South Providence cluster in what seems, from online maps, to be an unremarkable section of the neighborhood. A high-low outlier was found next to the Pawtucket border in the north, in the block group including the North Burial Ground cemetery, the Peter Pan bus terminal, and the Miriam Hospital.

3 - Spatial-Statistical Regression

To identify the most important block group-level variables in terms of their co-association with BPU, several methods were tried. A non-spatial ordinary least squares model was

---

[8] Weaver and Bagchi-Sen

[9] Victoria C. Morckel, Spatial characteristics of housing abandonment, Applied Geography, Volume 48, March 2014, Pages 8-16, ISSN 0143-6228, http://dx.doi.org/10.1016/j.apgeog.2014.01.001

attempted first, but indicated a high degree of failure due to spatial autocorrelation. Then, both spatial lag and spatial error models were attempted. In both models, a wide range of possible variables were added and then slowly removed from the regression to identify the optimal trade-off between number of variables and predictive power. These variables represented factors such as race and ethnicity, educational attainment, household income, housing density, residential type, and median age. Various combinations of variables were tried and compared to isolate the few with the highest R^2. The process was thorough, though not exhaustive. After this process, the two models were compared, and the spatial lag model was seen to offer a higher predictive power, with an R^2 of 0.60. It became clear that the three most important variables were the percentage of housing units vacant, the percentage of non-vacant units occupied by renters rather than owners, and the percentage of white non-Hispanic residents. These three factors had coefficients of 4.84, -0.496, and -0.785, respectively. Each had a p-value of less than 0.05. The spatially lagged variable for BPU had a coefficient of 0.357, and an equally low p-value. The y-intercept was 0.510.

4 - Sub-Categorizations of Violations

Another useful, and more specific, application of the housing code violation data is its sub-categorization based on the type of violation and the analysis of the various sub-categories. Taking this approach provides a more fine-grained level of results. Although the City of Providence may learn from a general analysis which areas are disproportionately prone to violations of the housing code, in order to address these violations it may be more useful to know exactly what types of violations are taking place in which areas. For example, the City might choose to send a pest exterminator to an area based on a general concentration of violations there, when in fact rodent-related violations are clustered in a different neighborhood entirely.

However, the procedure of sub-categorization posed a set of problems when applied to this data. Because the data had not been categorized in any computer-legible way, but rather in hand-typed English, the procedure was difficult to automate - and because there were 60,555 recorded violations, impossible to conduct by hand. To solve this problem, an Excel script was used that categorized violations based on the presence of certain keywords in their descriptions. There were a total of 12 categories. For example, all violations whose descriptions included "mice", "mouse", "rat", "rodent", etc., were assigned to the "small mammal" category. Descriptions including "fire", "smoke", "carbon monoxide", etc., were assigned to the "fire hazard" category.

There were both advantages and drawbacks to this method of categorization. It enabled an automated analysis of a relatively uncooperative dataset, allowing an in-depth analysis perhaps more useful than the results of an overall analysis. However, it required a number of assumptions and contained a number of biases, which limit its authoritativeness regarding results. For example, there may be violations of a certain type whose descriptions happened not to include any of the assigned keywords, or in which the keywords were misspelled. Very common misspellings were included in keyword lists ("protective caotings", for example), but the lists could not be exhaustive. False positives also present a problem for this method. Additionally, the process could not categorize by severity: "House totally overgrown by vines" and "slight issue with weeds in the front yard" would both be categorized under "issues with vegetation", for example.

Despite the possible issues, the analysis proceeded. These categories were re-counted at the block group level, and each block group was given 24 new variables: 12 representing

BPU for each of the particular types of violation, and 12 representing the ratio of type-categorized BPU to overall BPU - that is, representing the ratio of a specific kind of violation to violations in general. The latter variables will be referred to as ‘type ratios’. For each type of violation, the average and standard deviation of the distribution of type ratios across block groups were calculated. The type ratios were then mapped: 12 maps were created, one for each type, highlighting all block groups with a type ratio more than 1.5 standard deviations above the mean for that type. The resulting maps are highly illustrative (see Appendix 1, fig. 3).

There are many interesting results to be gleaned from these maps, although each of them should be understood with some skepticism. In all cases, there are multiple possible interpretations, and in all cases, sampling or procedural biases may be responsible for the result. However, a few of the most interesting are the following:

The only type of violation that is disproportionately common on the East Side are those dealing with protective coatings (in most cases, paint). East Side residents may be unusually concerned with appearances, or else simply have a disproportionately low BPU of all other types.

Insect- and small mammal- related violations both have a relatively wide spread but both also exhibit some clustering, particularly in a handful of block groups in the western part of the city. Perhaps those neighborhoods have some set of characteristics that make them more vulnerable to infestation, like a poor garbage disposal system.

Vegetation-related violations were unusually common only in a certain area of northwest Providence, perhaps because that area is both more suburban (and therefore has more plant life in general) and lower-income than other suburban areas of the city.

Mold was disproportionately reported in only one block group and flooding in only two, but the two categorizations overlap. It is easy to imagine a causal link connecting the two factors, as mold will easily grow in flooded buildings, but that is not necessarily the explanation for this behavior.

Conclusions

As important as the results of this project are the avenues for future research that it opens. Overall, there are a number of other analyses or improvements that could be made: The data, for example, could be analyzed temporally, exploring the change in violation patterns over time or more heavily weighting recent violations. It could also be categorized by severity. The data might also be used to investigate the behavior of individual inspectors, checking for bias in enforcement patterns, and, perhaps, correcting for these biases in further projects.

The analysis of citywide clustering points clearly to areas of greater and lesser housing code violations per unit of housing, and therefore of greater and lesser urban blight. These areas colocate, at least broadly, with neighborhoods of greater and lesser economic privilege, and these findings could be used by the City of Providence to improve the allocation of its resources in fighting violations of the housing code and thereby fighting the decline of neighborhoods.

The statistical regression accomplishes a similar goal: it enables prediction of the need to improve housing code compliance based on related variables. The indication that vacancy, aside from adjacent violations, was the most variable most heavily-correlated with BPU indicates that Providence may do well to fight property deterioration by fighting vacancy, or the inverse.

Sub-categorized analysis indicated that certain types of housing code violation could be most effectively combated in certain neighborhoods, and visualizes which neighborhoods those are. The co-localization of certain types of violation indicates that they may be causally linked, but does not prove such a linkage or demonstrate how it might occur. This section of the project, though, holds the most room for improvement and for future extensions of the research - the violations might be more rigorously categorized, or more robust statistical evaluations and comparisons applied. If the research is extended in such a way, it offers the City of Providence data that may be useful for the intelligent enforcement of the housing code throughout our city.

Appendix 1 - Figures and Tables

Fig. 1 - Quantile Map of Housing Code Violations by Block Group.

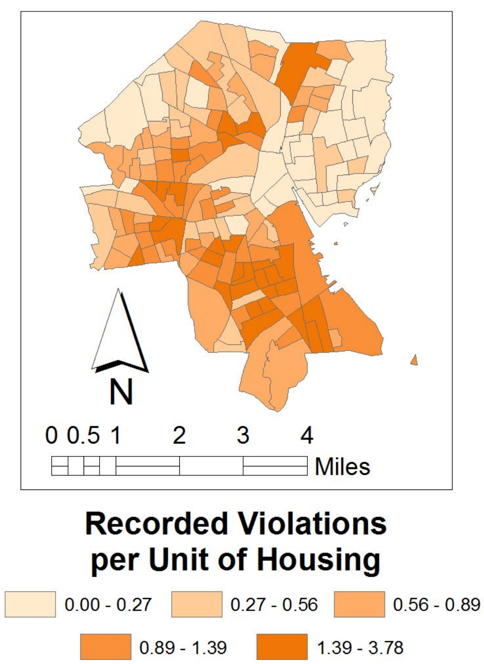


Fig. 2 - Clustering Map of Housing Code Violations by Block Group.

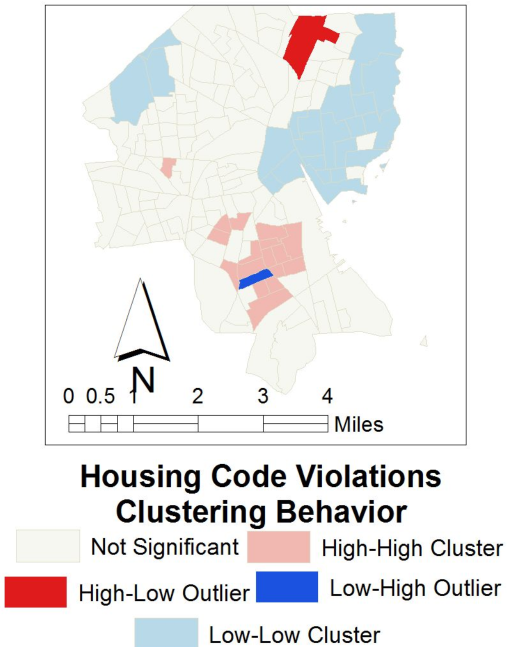
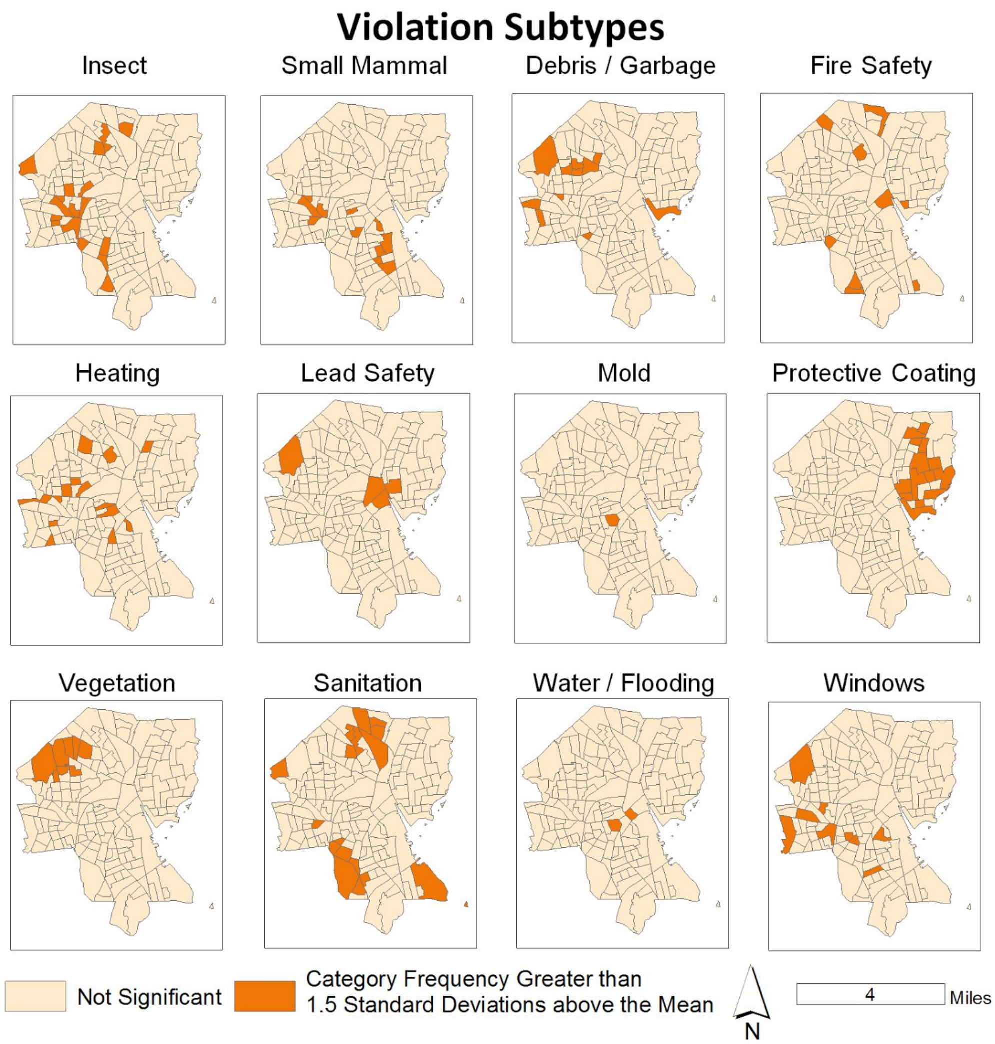

N
0 0.5 1 2 3 4
Miles
Housing Code Violations
Clustering Behavior
Not Significant
High-High Cluster
High-Low Outlier
Low-High Outlier
Low-Low Cluster

Fig. 3 - Housing Code Violations by Type



Appendix 2 - Academic Sources:

1: R.C. Weaver, Sharmistha Bagchi-Sen, Spatial analysis of urban decline: The geography of blight, Applied Geography, Volume 40, June 2013, Pages 61-70, ISSN 0143-6228, http://dx.doi.org/10.1016/j.apgeog.2013.01.011.

2: Breger, G. E., (1967), The Concept and Causes of Urban Blight, Land Economics, 43, issue 4, p. 369-376, http://EconPapers.repec.org/RePEc:uwp:landec:v:43:y:1967:i:4:p:369-376.

3: Sheila U. Appel, Derek Botti, James Jamison, Leslie Plant, Jing Y. Shyr, Lav R. Varshney, (November 2014), Predictive analytics can facilitate proactive property vacancy policies for cities, Technological Forecasting and Social Change, Volume 89, Pages 161-173, ISSN 0040-1625, http://dx.doi.org/10.1016/j.techfore.2013.08.028.

4: Amy Hiller, Dennis Culhane, Tony Smith, C. Dana Tomlin (2003), Predicting Housing Abandonment with the Philadelphia Neighborhood Information System, U. Penn. ScholarlyCommons, http://repository.upenn.edu/cgi/viewcontent.cgi?article=1006&context=cplan_papers

5: Yin, Li; Silverman, Robert M. 2015. "Housing Abandonment and Demolition: Exploring the Use of Micro-Level and Multi-Year Models." ISPRS Int. J. Geo-Inf. 4, no. 3: 1184-1200. http://www.mdpi.com/2220-9964/4/3/1184

6: Victoria C. Morckel, Spatial characteristics of housing abandonment, Applied Geography, Volume 48, March 2014, Pages 8-16, ISSN 0143-6228, http://dx.doi.org/10.1016/j.apgeog.2014.01.001

7: Accordino, J. and Johnson, G. T. (2000), Addressing the Vacant and Abandoned Property Problem. Journal of Urban Affairs, 22: 301–315. doi:10.1111/0735-2166.00058

8: Anselin, L. (1995), Local Indicators of Spatial Association—LISA. Geographical Analysis, 27: 93–115. doi:10.1111/j.1538-4632.1995.tb00338.x

9: Haifeng Zhang, Eric S. McCord, A spatial analysis of the impact of housing foreclosures on residential burglary, Applied Geography, Volume 54, October 2014, Pages 27-34, ISSN 0143-6228, http://dx.doi.org/10.1016/j.apgeog.2014.07.007.