Computer vision uncovers predictors of physical urban change

Nikhil Naik[a,b,1], Scott Duke Kominers[c,d,e,f,g,h], Ramesh Raskar[b], Edward L. Glaeser[e,h,i], and César A. Hidalgo[j]

[a]Joint Center for History and Economics, Harvard University, Cambridge, MA 02138; [b]Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139; [c]Society of Fellows, Harvard University, Cambridge MA 02138; [d]Harvard Business School, Boston, MA 02163; [e]Department of Economics, Harvard University, Cambridge, MA 02138; [f]Center of Mathematical Sciences and Applications, Harvard University, Cambridge, MA 02138; [g]Center for Research on Computation and Society, Harvard University, Cambridge, MA 02138; [h]National Bureau of Economic Research, Cambridge, MA 02138; [i]John F. Kennedy School of Government, Harvard University, Cambridge, MA 02138; and [j]Collective Learning Group, Media Lab, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Jose A. Scheinkman, Columbia University, New York, NY, and approved May 25, 2017 (received for review November 17, 2016)

Which neighborhoods experience physical improvements? In this paper, we introduce a computer vision method to measure changes in the physical appearances of neighborhoods from time-series street-level imagery. We connect changes in the physical appearance of five US cities with economic and demographic data and find three factors that predict neighborhood improvement. First, neighborhoods that are densely populated by college-educated adults are more likely to experience physical improvements—an observation that is compatible with the economic literature linking human capital and local success. Second, neighborhoods with better initial appearances experience, on average, larger positive improvements—an observation that is consistent with "tipping" theories of urban change. Third, neighborhood improvement correlates positively with physical proximity to the central business district and to other physically attractive neighborhoods—an observation that is consistent with the "invasion" theories of urban sociology. Together, our results provide support for three classical theories of urban change and illustrate the value of using computer vision methods and street-level imagery to understand the physical dynamics of cities.

urban economics | gentrification | urban studies | computer vision | neighborhood effects

For more than a century, urban planners, economists, sociologists, and architects have advanced theories connecting the dynamics of a neighborhood's physical appearance to its location, demographics, and built infrastructure.

The tipping theory of Schelling (1) and Grodzins (2) suggests that neighborhoods in bad physical condition will get progressively worse, whereas nicer areas will get better. Economic theories of urban change at the city level often emphasize population density and education (3–6), and it is natural to hypothesize that agglomeration of human capital will predict neighborhood-level improvements as well. Theories from urban sociology, such as the invasion theory of Burgess (7), however, emphasize locations and social networks, predicting that improvements in a city's appearance should be spatially clustered, and that improvements should occur both near the central business districts (CBDs) and near other physically attractive neighborhoods.

To test theories of physical neighborhood change, we need to quantify neighborhood appearance at different points in time. Historically, however, methods to quantify neighborhood appearance have not been scalable. The empirical literature on urban appearance, which was pioneered by urban planners such as Lynch (8), Rapoport (9), and Nasar (10), as well as by psychologists such as Milgram (11), has relied on interviews, low-throughput visual perception surveys, and manual evaluation of images. Those methods, however, can only be used to collect data on a few neighborhoods and have limited spatial resolution. In the past decade, new data on urban appearance have emerged in the form of "street view" imagery (12). As of 2016, Google Street View has photographed more than 3,000 cities from 106 countries at the street level. Recent approaches to quantify urban appearance, such as those of Rundle et al. (13), Hwang and Sampson (14), and Salesses et al. (15), leverage this large online corpus of street-level imagery but still rely on manual data curation, limiting throughput.

The appearance of street-level imagery sources has been paralleled by significant advances in the field of computer vision. Tasks such as automatically classifying and labeling images are now much easier, thanks in part to the availability of more comprehensive training datasets and new machine learning algorithms (16). These advances have led to an emerging literature at the intersection between computer vision, urban planning, urban sociology, and urban economics.

In 2011, the Massachusetts Institute of Technology (MIT) Place Pulse project (15) began collecting a massive crowdsourced dataset on urban appearance by asking people to select images from pairs in response to evaluative questions (such as "Which place looks safer?"). Naik et al. (17) used the Place Pulse data to train a computer vision algorithm called Streetscore that accurately predicts human-derived ratings for the perception of a streetscape's safety (also see refs. 18 and 19). Using Streetscore, Naik et al. (17) scored more than 1 million images from 21 cities in the northeastern United States, creating the largest

Significance

We develop a computer vision method to measure changes in the physical appearances of neighborhoods from street-level imagery. We correlate the measured changes with neighborhood characteristics to determine which characteristics predict neighborhood improvement. We find that both education and population density predict improvements in neighborhood infrastructure, in support of theories of human capital agglomeration. Neighborhoods with better initial appearances experience more substantial upgrading, as predicted by the tipping theory of urban change. Finally, we observe more improvement in neighborhoods closer to both city centers and other physically attractive neighborhoods, in agreement with the invasion theory of urban sociology. Our results show how computer vision techniques, in combination with traditional methods, can be used to explore the dynamics of urban change.

Author contributions: N.N., S.D.K., E.L.G., and C.A.H. designed research and experiments; N.N., S.D.K., E.L.G., and C.A.H. performed research and experiments; R.R. and E.L.G. contributed new analytic tools; N.N., S.D.K., E.L.G., and C.A.H. analyzed data; and N.N., S.D.K., E.L.G., and C.A.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. Email: naik@mit.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1619003114/-/DCSupplemental.

high-resolution dataset of urban appearance to date. Been et al. (20) used the Streetscore dataset to show that streets with higher Streetscores in New York are more likely to have been designated as historical districts. Harvey and Aultman-Hall (21) examined the skeletal aspects of neighborhoods to show that narrow streets with high building densities are perceived as safer than wider streets with few buildings. Nadai et al. (22) used Streetscore and mobile phone data to investigate whether safer-looking neighborhoods are more lively.

Moreover, crowdsourcing and computer vision methods have been used along with street-level imagery to identify geographically distinctive architectural elements (23), develop unique city signatures (24), and predict socioeconomic indicators (25, 26). Taken together, the range of findings illustrates how computer vision methods can be used to improve the quantitative study of urban appearance and space.

In this paper, we create a high-resolution dataset of physical urban change for five major US cities and use it to study the

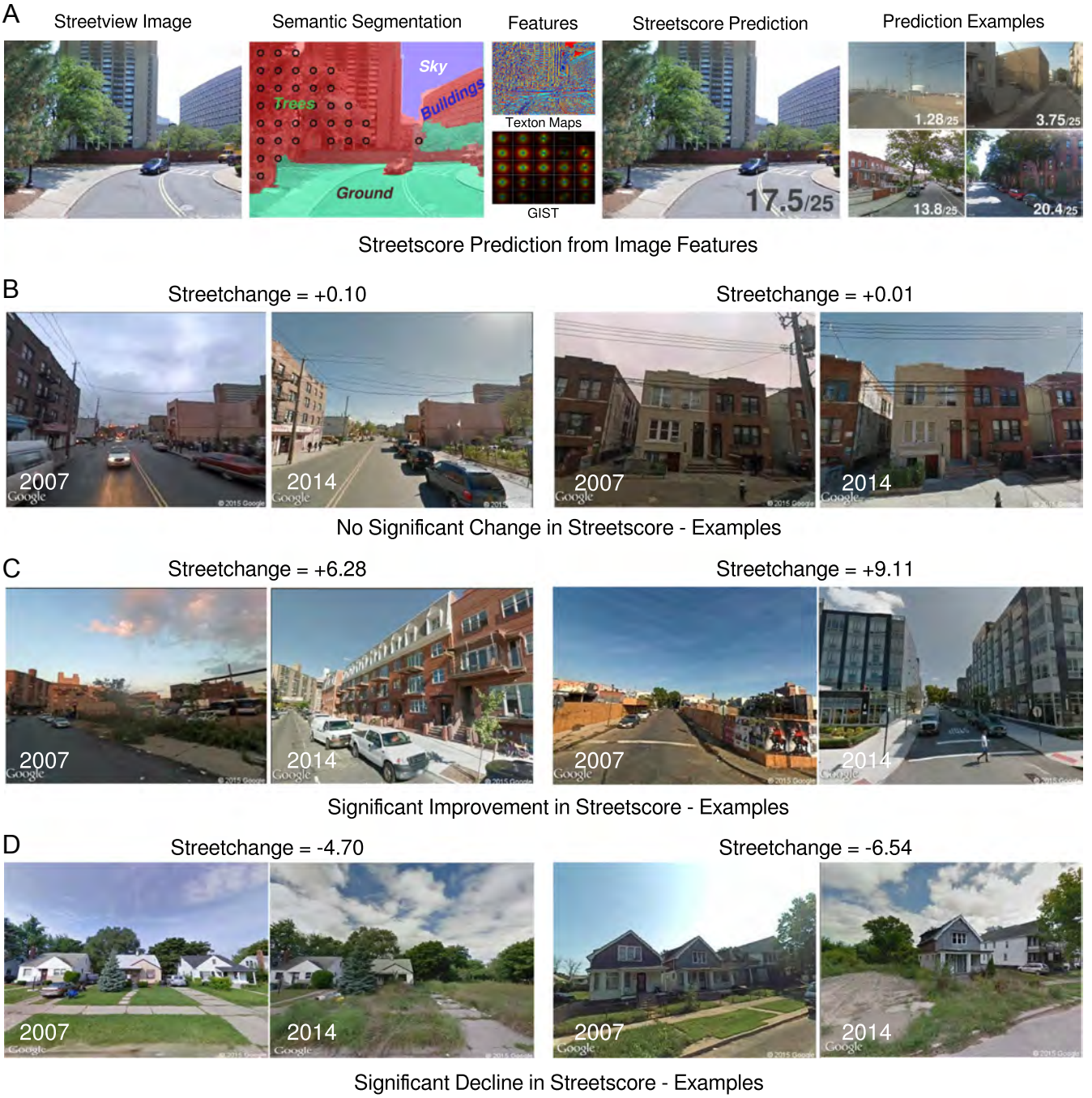

Fig. 1. Computing Streetchange: (*A*) We calculate Streetscore, a metric for perceived safety of a streetscape, using a regression model based on two image features: GIST and texton maps. We calculate those features from pixels of four object categories—ground, buildings, trees, and sky—which are inferred using semantic segmentation. (*B–D*) We calculate the Streetchange of a street block as the difference between the Streetscores of a pair of images captured in 2007 and 2014. (*B*) The Streetchange metric is not affected by seasonal and weather changes. (*C*) Large positive Streetchange is typically associated with major construction. (*D*) Large negative Streetchange is associated with urban decay. Insets courtesy of Google, Inc.

Table 1. Summary statistics ($N = 2{,}513$)

Variables	Description	Mean	SD	Minimum	Maximum
Streetscore 2007	Mean Streetscore 2007 of all sampled street blocks within a census tract	7.757	2.587	1.681	18.930
Streetchange 2007–2014	Mean Streetchange 2007–2014 of all sampled street blocks within a census tract	1.390	0.779	−4.076	6.121
Adjacent Streetscore 2007	Mean Streetscore 2007 of all boundary-adjacent census tracts	7.787	2.309	2.548	17.240
Log population density 2000	Log of population density within a census tract, as reported by the 2000 US Census	−4.655	1.220	−15.290	−2.480
Adjacent log population density 2000	Mean of log population density 2000 for all boundary-adjacent census tracts	−4.508	0.883	−11.090	−2.730
Share college education 2000	The share of adults within a census tract that have a four-year college degree, as reported by the 2000 US Census	0.254	0.216	0	1
Adjacent share college education 2000	Mean of share college education 2000 for all boundary-adjacent census tracts	0.251	0.191	0.115	1
Distance to CBD	The distance of the census tract from the central business district, in miles	5.123	2.685	0	9.997

determinants of physical improvements in neighborhoods. We use our data to test three theories of urban change. We find that, in agreement with economic theories of human capital agglomeration, neighborhoods that are densely populated by highly educated individuals are more likely to experience positive urban change. Also, in agreement with the invasion theory (7) of urban sociology, we find that neighborhoods are more likely to improve in physical appearance when they are proximate to a CBD and/or other neighborhoods perceived as safe. Finally, we find evidence for a weak version of the neighborhood tipping theory (1, 2), as the neighborhoods that had the best appearances at the beginning experienced the largest improvements (however, we do not find that neighborhoods with initially low scores deteriorated—they just improved less). Our findings illustrate how computer vision methods, together with demographic and economic data, can be used to study physical urban change.

Data and Methods

We obtained 360° panorama images of streetscapes from five US cities using the Google Street View application programming interface. Each panorama was associated with a unique identifier ("panoid"), latitude, longitude, and time stamp (which specified the month and year of image capture). We extracted an image cutout from each panorama by specifying the heading and pitch of the camera relative to the Street View vehicle. We obtained a total of 1,645,760 image cutouts for street blocks in Baltimore, Boston, Detroit, New York, and Washington, DC, captured in 2007 (the "2007 panel") and 2014 (the "2014 panel").* We matched image cutouts from the 2007 and 2014 panels by using their geographical locations (i.e., latitude and longitude) and by choosing the same heading and pitch. This process gave us images that show the same place, from the same point of view, but in different years (Fig. 1 *B–D*).†

We calculated the perception of safety—called "Streetscore"—for each image using a variant of the Naik et al. algorithm (17) trained on a crowdsourced study of people's perception of safety (15) based on 2,920 images from Boston and New York and 186,188 pairwise comparisons. The Streetscore computation process included three steps (Fig. 1*A*). First, we segmented images into four "geometric" classes: ground (which contains streets, sidewalks, and landscaping), buildings, trees, and sky (27). Next, we created feature vectors characterizing each geometric class using two image features: GIST (28) and texton maps (29). Roughly speaking, these features encode the shapes and textures present in an image. Finally, we used the features of streets and buildings to predict the Streetscore of an image using support vector regression (30). We ignored the features of trees and sky to minimize seasonal effects (weather, time of day, and time of year). The predicted Streetscore of a Street View image ranges from 0 to 25, with 0 being the most unsafe-looking street scene in the sample and 25 the most safe-looking scene. Next, we computed changes in Streetscores between images in the 2007 and 2014 panels, to obtain Streetchange (Fig. 1 *B–D*). A positive value of Streetchange is indicative of upgrading in physical appearance, whereas a negative value of Streetchange is indicative of decline. (For details on the methods, see *SI Appendix*.)

We validated Streetchange using three sources: a survey conducted on Amazon Mechanical Turk (AMT), a survey of graduate students in MIT's School of Architecture and Planning, and data from Boston's Planning and Development Authority (BPDA).

Participants gave informed consent for all human subject studies. Experiments were approved by the Massachusetts Institute of Technology's Committee on the Use of Humans as Experimental Subjects (MIT COUHES). The AMT study was conducted in accordance with the requirements of MIT COUHES.

We found strong agreement between Streetchange and both (*i*) human assessments and (*ii*) new urban development. In the AMT validation, workers were presented two image pairs, drawn from a pool of 1,565, and asked to select the one showing more physical change. The binned ranked scores provided by the AMT workers had a strong correlation with absolute Streetchange (Spearman correlation = 72%, P-value $< 1 \times 10^{-5}$). In the School of Architecture and Planning student validation we presented students with 150 image pairs and asked them to classify images into positive and negative physical change ($N = 3$). The students agreed with Streetchange in 74% of cases. Finally, we collected building project data from BPDA and correlated Streetchange with total new square footage built per square mile (at census-tract level) during the sample period (2012–2014). We found a significant and positive correlation between Streetchange and new square footage—one SD increase in log

*For the street blocks that lack images for either 2007 or 2014 we completed the 2007 and 2014 panels using images from the closest years for which data were available. As a result, 5% of the images in the 2007 panel are from either 2008 or 2009. Similarly, 12% of the images in the 2014 panel are from 2013.

†We reduced our data to eliminate pairs containing over-exposed, blurred, or occluded images (for details, see *SI Appendix*).

Table 2. Relationship between social characteristics and changes in Streetscore

	Coefficients for Streetscore 2007			Coefficients for Streetchange 2007–2014		
Independent variables	(1)	(2)	(3)	(4)	(5)	(6)
Share college education 2000	2.547***		2.657***	0.657***		0.703***
	(0.740)		(0.668)	(0.106)		(0.105)
Log population density 2000		0.740***	0.832***		0.056***	0.084***
		(0.095)	(0.107)		(0.020)	(0.024)
Streetscore 2007				0.027***	0.033**	0.013
				(0.010)	(0.014)	(0.012)

All models control for city fixed effects. $^{***}P < 0.01$, $^{**}P < 0.05$.

total square footage corresponds to roughly half an SD increase in Streetchange (see *SI Appendix* for details).

To relate the Streetscore indicators of neighborhood appearance to socioeconomic composition, we aggregated the Streetscore and Streetchange variables at the census-tract level and obtained tract characteristic data from the 2000 US Census, adjusted to the 2010 census-tract boundaries (31). For summary statistics, see Table 1.

Results

We begin by presenting the cross-sectional demographic and economic correlates of cities' physical appearances and changes in appearance, as estimated by 2007 Streetscore and Streetchange between 2007 and 2014 (Table 2). All regressions include city fixed effects and hold up in multivariate specifications. Additionally, in all regressions we have corrected for spatial correlation in standard errors following Conley (32) using STATA routines developed by Hsiang (33). For each census tract we consider population density, level of education (share of college educated adults), median income, housing price, rental costs, housing vacancy, race, and poverty. From all of these variables the two strongest correlates of perception of safety are population density and education, so we present a table (Table 2) summarizing the coefficients of these two variables. (For a table with all controls see *SI Appendix*, Table S4.)

Column 2 of Table 2 shows that Streetscores improve by 0.74 with the log of population density. This represents about one-quarter of an SD of Streetscore (2.6). Because Streetscores are roughly linear in log density, the overall relationship is concave, meaning that perceived safety rises with density but the effect levels off. This fact lends some support to the idea that perceived safety increases with "eyes on the street" (34). However, our finding does not imply that dense urban spaces are seen as safer than low-density suburban or rural areas, because we do not have such low-density spaces in our sample. Our results suggest only that in five (generally dense) eastern US cities, spaces with high population densities are perceived as being safer than urban spaces with low population densities.

The second robust correlate of perceived safety is education (Table 2, column 1). As the share of the population with a college degree increases by 20% (one SD), perceived safety rises by 0.51, or one-sixth of an SD in Streetscore. We suspect that the relationship reflects the tendency of educated people to be willing to pay for neighborhoods that appear safer, rather than the ability of educated residents to make a neighborhood feel safe.

We now move to changes in physical appearance—the primary contribution of this paper. Columns 4, 5, and 6 of Table 2 show the correlations between initial social characteristics, as measured in the 2000 US Census, and neighborhood Streetchange as measured between 2007 and 2014.

Column 4 of Table 2 examines education, again controlling for initial Streetscore. The observed impact of education on Streetchange seems to be large. A one-SD increase in share with college degree in 2000 (20%) is associated with an increase in Streetchange of 0.13, or about one-sixth of an SD. Just as skilled cities have done particularly well over the last 50 y, skilled neighborhoods seem to have experienced more physical improvement.

Column 5 of Table 2 shows that—controlling for initial Streetscore—as the log of density increases by 1 the growth in Streetscore increases by 0.06 points. The estimated impact of log

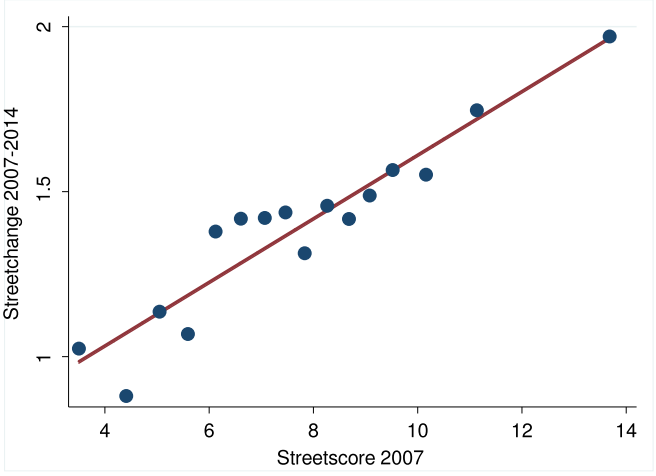

Fig. 2. Evidence of neighborhood tipping: We test the tipping model of neighborhood change. We group the data into 16 bins based on the initial value of Streetscore and plot the average Streetchange in each bin against the average initial Streetscore.

Table 3. Evidence of invasion

	Coefficients for Streetchange 2007–2014			
Independent variables	(1)	(2)	(3)	(4)
Distance to CBD	−0.042***	−0.050***	−0.051***	−0.036***
	(0.011)	(0.011)	(0.011)	(0.011)
Adjacent Streetscore 2007	0.063***			0.049**
	(0.019)			(0.019)
Adjacent log population density 2000		0.115**		0.093**
		(0.046)		(0.046)
Adjacent share college education 2000			0.620***	0.626***
			(0.167)	(0.172)

All models control for Streetscore 2007, Share college education 2000, and Log population density 2000 of a given census tract—along with city fixed effects. $^{***}P < 0.01$, $^{**}P < 0.05$.

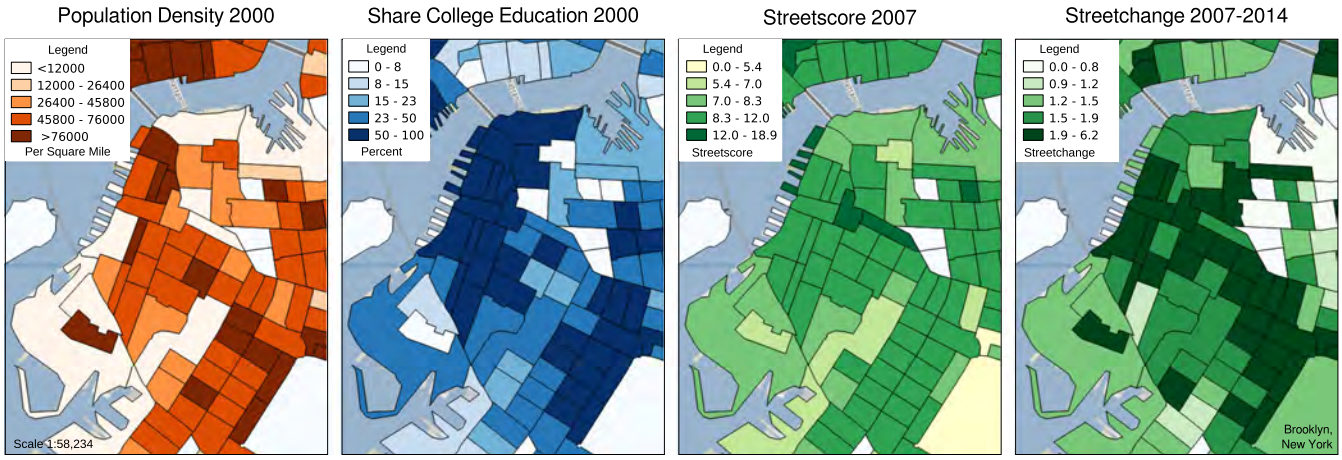

Fig. 3. The correlates of physical upgrading in neighborhoods: Positive urban change occurs in geographically and physically attractive areas with dense, highly educated populations, as illustrated here for Brooklyn, New York.

density on the Streetchange over a 7-y period is about 1/12 of the impact of log density on the level of Streetscore in 2007. Density does seem to predict growth in Streetscore over the sample period, but the relationship is far weaker than the connection between density and the level of Streetscore in 2007. We did not find any robust relationships between Streetchange and median income, housing price, or rental costs; this suggests that the education effect is more likely to reflect skills than income (*SI Appendix*, Table S4).

The finding that variables that predict the level of Streetscore in 2007 also predict the change in Streetscore between 2007 and 2014 seems to support a positive feedback loop—the essence of tipping models (1, 2). Tipping is also suggested by the positive correlation between 2007 Streetscore and Streetchange.[‡] However, we find a linear relationship, rather than the nonlinear relationship suggested by the original tipping theory (2). Moreover, tipping models suggest that initially unattractive neighborhoods get worse over time—and that is not found in our data. The mean Streetchange by decile of 2007 Streetscore is positive even for the areas with the lowest scores (Fig. 2). It is not clear whether this represents tipping or a pattern in which visually safer areas are being upgraded first and faster. We suspect that the lack of downward movement may be particular to the time period under consideration. Despite the Great Recession, 2007–2014 was a relatively good time period for many of America's eastern cities, and this may explain why we do not see declines even for less-attractive neighborhoods. Still, the data do show the overall pattern predicted by tipping models, in which upward growth is faster in initially better areas.

Next, we test for invasion (7) by regressing changes in Streetscore on characteristics of bordering neighborhoods and proximity to the CBD, after controlling for the predictors identified in Table 2 (2007 Streetscore, log of density, and education).[§] The invasion hypothesis is just one of the reasons why areas may improve more when they have attractive neighbors—perhaps the most natural explanation is just that areas that are worse or better than their neighbors tend to mean-revert to the norm for their sections of the city. We test for the importance of location within the city by looking at the impact of proximity to the CBD. Column 1 of Table 3 shows that as the distance to the CBD increases by 1 mile expected Streetscore growth falls by 0.04 points. Appearance upgrading is strongest closer to the city center, paralleling Kolko's (36) finding that economic gentrification is more pronounced closer to the city center.

The original invasion hypothesis postulated a process under which low-income areas would gradually make their ways out from the center to nearby suburbs. The current pattern is instead one in which the central city sees particularly large upgrades in perceived street safety. One interpretation is that we are currently witnessing the reversal of the process described by Burgess (7): City centers, which have always had a strong fundamental asset—proximity to jobs—are experiencing physical change that expresses a reversion to that fundamental.

Although the data do not suggest decay emanating out from a center, the core idea of the invasion hypothesis—that neighborhoods spill over into each other—is readily confirmed in the data. Column 1 of Table 3 also shows the effect of average Streetscore in surrounding areas on Streetchange. Notably, the coefficient on neighboring scores is more than double the impact of the neighborhood's own score, implying that almost 1/10 of the Streetscore difference between a neighborhood and its neighbors is eliminated over a 7-y period. Because most of the movement over the sample period is positive, the regression should be interpreted as meaning that growth is faster in areas with more attractive neighbors. This strong convergence is exactly the prediction of the invasion theory.

Column 2 of Table 3 examines the effect of adjacent density. Because adjacent density is highly correlated with adjacent Streetscore, it is not surprising to see that there is also a robust correlation here, although the connection is not as strong as with adjacent Streetscores. Column 3 of Table 3 looks at average share of the population with college degrees in adjacent areas. The relationship is positive and robust. As the share increases by 20%, Streetscore increases by 0.12 points. This again corroborates the results of Kolko (36), who found that gentrification is faster in areas with more educated neighbors. These findings point to a process of neighborhood spillovers and convergence, which are, in a sense, at the heart of the invasion hypothesis.[¶]

[‡] Without the other controls, for each extra point of Streetscore in 2007, Streetscore growth is 0.04 points higher over the next 7 y.

[§] CBD locations were based on the coding of Cortright and Mahmoudi (35).

[¶] In our working paper (37) we also looked at the "filtering" hypothesis, which suggests the importance of the age of the building stock: Areas should gradually decline until they are upgraded. To test the hypothesis that building age shapes streetscape change, we regressed Streetchange on the shares of the building stock (as of the year 2000) built during different decades, controlling for 2007 Streetscore, log of density, and education. We found at best limited support for the filtering hypothesis (*SI Appendix*, Table S5).

Fig. 3 illustrates the relationship between location, education, population density, and physical improvement in neighborhoods from Brooklyn, New York. In *SI Appendix*, Figs. S9–S28 we provide similar map visualizations for all cities in our dataset.

Conclusion

For decades, scholars from the social sciences and the humanities have discussed the importance of urban appearance and the factors that may contribute to physical urban change. Here, we test theories of urban change using Streetchange, a metric for change in urban appearance obtained from street-level imagery with a computer vision algorithm.

The data show that population density and education in both neighborhoods and their surrounding areas robustly predict improvements in neighborhoods' physical environments; other variables show less correlation. The results also show strong support for the invasion hypothesis of neighborhood change (7), which emphasizes spillovers across neighborhoods.

Our work suggests several open questions for future work. Is the correlation between density and perceived safety true more generally, or does it mean-revert after a certain point? Does tipping appear when we examine cities with static or declining levels of Streetscore? We hope that future research, enabled in our part by our dataset and methods, can help address these questions and continue exploring the links between the physical city and the humans that reside there.

ACKNOWLEDGMENTS. We thank Gary Becker, Jörn Boehnke, Graeme Campbell, Steven Durlauf, Ingrid Gould Ellen, James Evans, Jay Garlapati, Lars Hansen, John William Hatfield, James Heckman, John Eric Humphries, Jackelyn Hwang, Paul Kominers, Michael Luca, Mia Petkova, Priya Ramaswamy, Matthew Resseger, Robert Sampson, Scott Stern, Zak Stone, Erik Strand, and Nina Tobio for helpful comments. This work was supported by the International Growth Center, the Alfred P. Sloan Foundation, and a Star Family Challenge grant (to N.N., S.D.K., and E.L.G.); National Science Foundation Grants CCF-1216095 and SES-1459912 (to S.D.K.), the Harvard Milton Fund, the Ng Fund of the Harvard Center of Mathematical Sciences and Applications, and the Human Capital and Economic Opportunity Working Group sponsored by the Institute for New Economic Thinking (S.D.K.); the Taubman Center for State and Local Government (E.L.G.); as well as the Google Living Labs Award and a gift from Facebook (to C.A.H.).

1. Schelling TC (1969) Models of segregation. *Am Econ Rev* 59:488–493.
2. Grodzins M (1957) Metropolitan segregation. *Sci Am* 197:33–41.
3. Glaeser EL, Scheinkman JA, Shleifer A (1995) Economic growth in a cross-section of cities. *J Monet Econ* 36:117–143.
4. Bettencourt LM (2013) The origins of scaling in cities. *Science* 340:1438–1441.
5. Ciccone A, Hall RE (1996) Productivity and the density of economic activity. *Am Econ Rev* 86:54–70.
6. Glaeser EL, Gottlieb JD (2009) The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *J Econ Lit* 47:983–1028.
7. Burgess EW (1925) *The Growth of the City* (Univ of Chicago Press, Chicago).
8. Lynch K (1960) *The Image of the City* (MIT Press, Cambridge, MA).
9. Rapoport A (1969) *House Form and Culture* (Prentice Hall, Englewood Cliffs, NJ).
10. Nasar JL (1998) *The Evaluative Image of the City* (Sage Publications, Thousand Oaks, CA).
11. Milgram S (1976) Psychological maps of Paris. *Environmental Psychology: People and Their Physical Settings*, eds Proshansky HM, Ittelson WH, Rivlin LG (Holt, Rinehart and Winston, New York), pp 104–124.
12. Anguelov D, et al. (2010) Google street view: Capturing the world at street level. *IEEE Comput* 43:32–38.
13. Rundle AG, Bader MD, Richards CA, Neckerman KM, Teitler JO (2011) Using Google street view to audit neighborhood environments. *Am J Prev Med* 40: 94–100.
14. Hwang J, Sampson RJ (2014) Divergent pathways of gentrification: Racial inequality and the social order of renewal in Chicago neighborhoods. *Am Sociol Rev* 79:726–751.
15. Salesses P, Schechtner K, Hidalgo CA (2013) The collaborative image of the city: Mapping the inequality of urban perception. *PLoS One* 8:e68400.
16. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521:436–444.
17. Naik N, Philipoom J, Raskar R, Hidalgo CA (2014) Streetscore – Predicting the perceived safety of one million streetscapes. *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (IEEE, Washington, DC), pp 793–799.
18. Ordonez V, Berg TL (2014) Learning high-level judgments of urban perception. *Computer Vision – ECCV 2014*. Lecture Notes in Computer Science, eds Fleet D, Pajdla T, Schiele B, Tuytelaars T (Springer, Cham, Switzerland), Vol 8694, pp 494–510.
19. Porzi L, Rota Bulò S, Lepri B, Ricci E (2015) Predicting and understanding urban perception with convolutional neural networks. *Proceedings of the 23rd ACM International Conference on Multimedia* (ACM, New York), pp 139–148.
20. Been V, Ellen IG, Gedal M, Glaeser EL, McCabe BJ (2016) Preserving history or restricting development? The heterogeneous effects of historic districts on local housing markets in New York City. *J Urban Econ* 92:16–30.
21. Harvey C, Aultman-Hall L (2016) Measuring urban streetscapes for livability: A review of approaches. *Prof Geogr* 68:149–158.
22. De Nadai M, et al. (2016) Are safer looking neighborhoods more lively? A multimodal investigation into urban life. *Proceedings of the 2016 ACM on Multimedia Conference* (ACM, New York), pp 1127–1135.
23. Doersch C, Singh S, Gupta A, Sivic J, Efros AA (2015) What makes Paris look like Paris? *Commun ACM* 58:103–110.
24. Zhou B, Liu L, Oliva A, Torralba A (2014) Recognizing city identity via attribute analysis of geo-tagged images. *Computer Vision – ECCV 2014*. Lecture Notes in Computer Science, eds Fleet D, Pajdla T, Schiele B, Tuytelaars T (Springer, Cham, Switzerland), Vol 8691, pp 519–534.
25. Arietta SM, Efros AA, Ramamoorthi R, Agrawala M (2014) City forensics: Using visual elements to predict non-visual city attributes. *IEEE Trans Vis Comput Graph* 20:2624–2633.
26. Glaeser EL, Kominers SD, Luca M, Naik N (2016) Big data and big cities: The promises and limitations of improved measures of urban life. *Econ Inq*, 10.1111/ecin.12364.
27. Hoiem D, Efros AA, Hebert M (2008) Putting objects in perspective. *Int J Comput Vis* 80:3–15.
28. Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. *Int J Comput Vis* 42:145–175.
29. Malik J, Belongie S, Leung T, Shi J (2001) Contour and texture analysis for image segmentation. *Int J Comput Vis* 43:7–27.
30. Schölkopf B, Smola AJ, Williamson RC, Bartlett PL (2000) New support vector algorithms. *Neural Comput* 12:1207–1245.
31. Logan JR, Xu Z, Stults BJ (2014) Interpolating US decennial census tract data from as early as 1970 to 2010: A longitudinal tract database. *Prof Geogr* 66:412–420.
32. Conley TG (2008) Spatial econometrics. *The New Palgrave Dictionary of Economics*, eds Durlauf SN, Blume LE (Palgrave Macmillan, New York).
33. Hsiang SM (2010) Temperatures and cyclones strongly associated with economic production in the Caribbean and Central America. *Proc Natl Acad Sci USA* 107:15367–15372.
34. Jacobs J (1961) *The Death and Life of Great American Cities* (Vintage, New York).
35. Cortright J, Mahmoudi D (2014) *Neighborhood change, 1970 to 2010: Transition and growth in urban high poverty neighborhoods* (Impresa Consulting, Portland, OR).
36. Kolko J (2010) The determinants of gentrification. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=985714.
37. Naik N, Kominers SD, Raskar R, Glaeser EL, Hidalgo CA (2015) *Do people shape cities, or do cities shape people? The co-evolution of physical, social, and economic change in five major US cities. NBER Working Paper 21620* (National Bureau of Economic Research, Cambridge, MA).

Supporting Information: Computer Vision Uncovers Predictors of Physical Urban Change

Nikhil Naik*, Scott Duke Kominers, Ramesh Raskar, Edward L. Glaeser, and César A. Hidalgo

Contents

1. Data and Methods

Using a dataset consisting of human-coded image comparisons, we train a computer vision algorithm to predict perceived safety of individual street scenes ("Streetscore"); we then obtain "Streetchange" by comparing that measurement across Street View images of the same locations from 2007 (the "2007 image") and 2014 (the "2014 image").

1.1. The Street View Image Dataset

We first describe the Street View image dataset used in this paper. We obtained 360° panorama images of streetscapes from five US cities using the Google Street View Application Programming Interface (API). Each panorama was associated with a unique identifier ("panoid"), latitude, longitude, and time stamp (which specified

*E-mail: naik@fas.harvard.edu

the month and year of image capture). We extracted an image cutout from each panorama by specifying the heading and pitch of the camera relative to the Street View vehicle. We obtained a total of 1,645,760 image cutouts for street blocks in Baltimore, Boston, Detroit, New York, and Washington DC, captured in years 2007 (the "2007 panel") and 2014 (the "2014 panel"). For the street blocks that lack images for either 2007 or 2014, we completed the "2007" and "2014" panels using images from the closest years for which data was available. As a result, 5% of the images in the "2007" panel are from either 2008 or 2009. Similarly, 12% of the images in the "2014" panel are from 2013. We matched image cutouts from the 2007 and 2014 panels by using their geographical locations (i.e. latitude and longitude) and by choosing the same heading and pitch. This process gave us images that show the same place, from the same point of view, but in different years. A large majority of images in our dataset were captured between the months of April and August, to avoid a change of season between the two images of the same location.

Next, we describe the computer vision algorithm used for obtaining Streetchange—a measure for physical urban change—from the 2007 and 2014 image panels.

1.2. Image Feature Extraction

Our computer vision algorithms work with a number of structured and unstructured features of the image data: First, we used the Geometric Layout algorithm [1] to assign pixel-wise semantic labels in four geometric classes: "ground," "buildings," "trees," and "sky." Next, we extracted two different image features separately for the pixels of the four geometric classes:

- We generated a texton dictionary [2] by convolving the images with a Gaussian filter bank and clustering their responses together; every pixel was then assigned to the nearest cluster center, creating a texton map. We computed 512-dimensional histograms with texton maps of the four geometric classes. We call this feature Geometric Texton Histograms (GTH).
- We calculated the GIST feature descriptor [3]—a global image feature that provides a low-dimensional representation of the spatial layout properties of a scene—for each of the geometric classes.

1.3. Removal of Unsuitable Images

A small number of Street View image pairs in the sample were unsuitable for comparison. In particular: some images were over-exposed, out of focus, or blurred; others had significant changes in greenery coverage likely driven by seasonal changes rather than urban foliage improvements. To eliminate these unsuitable pairs, we used a series of automated data cleaning methods:

- First, we removed over-exposed images, which typically result from the sun shining directly into the camera. To identify over-exposed pixels, we converted each image to the CIELAB color space, in which the L channel represents lightness and a, b channels represent the color. The color channels were combined as $C = (a, b)^T$. We computed an over-exposure value $\mathcal{M}$ (between 0 and 1) at each pixel following methods introduced by Guo et al. [4]: At pixel i,

$$\mathcal{M}_i = \frac{1}{2} \cdot \tanh\left(\delta \cdot \Big((L_i - L_T) + (C_T - ||C_i||_2)\Big)\right),$$

  with the constants set to $\delta = 1/60$, $L_T = 80$, and $C_T = 40$. We obtained the mean $\mathcal{M}_i^{\text{sky}}$ of $\mathcal{M}_i$ over all the pixels that belonged to the "sky" geometric class, as predicted by the Geometric Layout algorithm [1]. We discarded the image pairs in which at least one of the images i had $\mathcal{M}_i^{\text{sky}} > 0.85$, indicating over-exposure. We discarded all image pairs containing at least one over-exposed image.

Figure S1. The Streetchange algorithm is robust to large weather and seasonal changes. In this example, our algorithm assigns a small Streetchange value to the image pairs, even though there is a drastic change in weather between the two images. Images courtesy of Google, Inc.

- Second, we removed images that were out-of-focus or contained motion blur. To detect such images, we computed the Absolute Central Moment (ACMO) of each image, a statistical measure that allows a simultaneous optimization of both focus and exposure [5]. If the normalized value of ACMO was less than 0.2, we labeled the image as blurred. We discarded all image pairs containing at least one blurred image.
- Finally, we discarded all image pairs in which the number of pixels in the image occupied by the "tree" object class (again, as predicted by the Geometric Layout algorithm [1]) changed by more than 10% between the 2007 image and the 2014 image. This process eliminated image pairs in which only one of the images had significant occlusion of buildings by trees.

1.4. Streetchange Calculation

Having removed the images unsuitable for urban change detection, we predicted the "Streetscores" of the remaining images using a support vector regression model trained with computer vision features and aggregate scores obtained from the crowdsourced study by Salesses et al. [6], as described next.

Salesses et al. [6] created an online crowdsourced game in which participants were shown images of streetscapes randomly chosen from New York, Boston, Linz and Salzburg. Participants were asked to choose one of the two images in response to three questions: "Which place looks safer?", "Which place looks more upper class?", and "Which place looks unique?". In the Salesses et al. [6] study, 7,872 unique participants from 91 countries provided 186,188 comparisons ("clicks") of image pairs drawn from a pool of 4,109 images for the question "Which place looks safer?"

Following Naik et al. [7], we converted the 186,188 pairwise comparisons for the question "Which place looks safer?" to ranked scores using a Bayesian ranking algorithm called Trueskill [8]. We call the Trueskill perceived safety score for each image that image's *Streetscore*; these images' Streetscores are "true scores" derived from aggregations of human assessments. We used the Streetscores obtained from human assessments to train a machine learning model that uses the GTH and GIST features of the corresponding images (described in Section 1.2) to predict how humans would score the perceived safety of Street View images. As we are only seeking to predict

(a) Image Pairs with Location Errors - Examples

(b) Image Pairs with Significant Occlusion Errors - Examples

Figure S2. A human operator eliminated the small fraction of invalid image pairs containing location-coding errors or significant occlusion of buildings by large vehicles. Images courtesy of Google, Inc.

the human perception of American cities, we restricted the training sample to the 2,920 human-coded images from New York and Boston.

We used ν-Support Vector Regression (ν-SVR) [9] to predict image Streetscores. Given a set of training images with feature vectors $\mathbf{x}$ and Streetscores $q \in \mathbb{R}$, ν-SVR with a linear kernel generates a weight vector $\mathbf{w}$ and a bias term b under a set of constraints. The two variables $(\mathbf{w}, b)$ are used to predict the Streetscore for a new image with feature vector $\mathbf{x}'$ by evaluating $q' = \mathbf{w} \cdot \mathbf{x}' + b$. We measured the accuracy of our predictor using the Coefficient of Determination (R^2). We obtained $R^2 = 0.57$ over fivefold cross-validation on the training set.

In this paper we dropped the Geometric Color Histogram (GCH) features used by Naik et al. [7], since GCH features were more sensitive to weather changes than GIST and texton histograms. Dropping the GCH features, however, did not substantially reduce the predictor accuracy—the R^2 dropped from 0.5884 to 0.5709.

Next, we used the Streetscore predictor to calculate urban change from image pairs. As our predictor is a weight vector trained using image features on top of the four geometric classes (ground, building, trees and sky), we were able to compute the contribution of each geometric class to the Streetscore of each image. We chose to discard the contribution of the "trees" and "sky" classes since their scores depend on the season and weather at the time of image capture. Note that the "trees" class contains only large trees (and not landscaping), allowing us to account for changes in the built environment due to landscaping as part of the "ground" class. Figure S1 shows examples of image pairs with large weather and seasonal changes which have been accurately scored by our algorithm.

After computing the Streetscore for each image in a 2007–2014 image pair, we calculated "Streetchange" as the difference between the 2014 image's Streetscore and the 2007 image's Streetscore.

1.5. Removing Erroneous Pairs

While we were able to computationally eliminate pairs containing over-exposed, blurred, or occluded images, we discovered two additional sources of error that made a small number of image pairs invalid for Streetchange calculation. The first source of error was incorrect location information for one or both images in an image pair. For these images, the geographic coordinates (latitude and longitude) obtained from Google Street View did not

Figure S3. The Streetchange algorithm is robust to the change in Street View image quality between years 2007 and 2014. Images courtesy of Google, Inc.

match with the actual geographic coordinates of the locations at which the images were captured (Figure S2-(a)). The second source of error was partial or complete occlusion of buildings by large vehicles (Figure S2-(b)), which were not removed by the procedure described in Section 1.3.

Our algorithm calculated a large positive or negative Streetchange value for image pairs containing location-coding or vehicle-occlusion errors, since the two images in such pairs look very different (Figure S2). Due to large variation in visual appearance within and across image pairs that contained these errors, we were unable to automatically eliminate such image pairs with a computer vision algorithm. Therefore, a human operator observed image pairs whose Streetchange value was larger than four standard deviations of change in Streetscore (which amounted to only 1,849 image pairs out of a total of 822,880—less than 0.23% of the sample). The operator manually eliminated image pairs which contained location-coding errors or vehicle-occlusion errors.

1.6. Image Quality and Streetchange

We also note that there is a difference in image quality between Street View images captured in year 2007 and Street View images from year 2008 or after (due to improvements in Google's imaging hardware). But the Streetchange algorithm is robust to change in image quality from 2007 and 2014 (Figure S3) for a few reasons. First, a significant fraction of images in the training set are also from 2007, which helps to mitigate the effect of change in image quality. Second, we process the images at a resolution of 400×300 pixels; the difference in quality is not significant at this resolution. Finally, a higher fraction of 2007 Street View images (as compared to year 2014) tend to be over-exposed; but we discard such images from our calculations.

1.7. Validating Streetchange

We validated our final Streetchange measures using three sources: a survey on Amazon Mechanical Turk (AMT), a survey of graduate students in MIT's School of Architecture and Planning, and data from Boston's Planning and Development Authority.

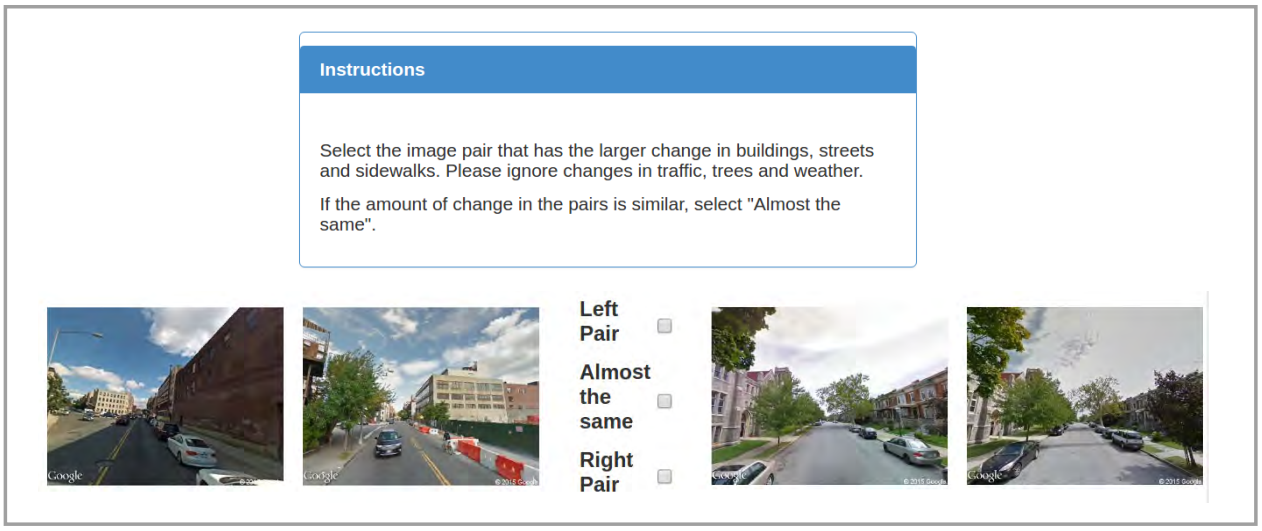

Figure S4. Screenshot of the Amazon Mechanical Turk experimental interface used for validating Streetchange. Insets courtesy of Google, Inc.

Validation with human observers: For the validation experiment on AMT, we selected 1,565 image pairs (roughly 1% of the final sample) using inverse transform sampling on Streetchanges. We presented AMT workers with two image pairs side-by-side (Figure S4) drawn randomly from the 1,565 image pairs and asked the following question:

> *Select the image pair that has the larger change in buildings, streets and sidewalks. Please ignore changes in traffic, trees and weather. If the amount of change in the pairs is similar, select "Almost the same".*

We obtained 28,170 pairwise comparisons for the 1,565 image pairs from 116 users—36 pairwise comparisons on average. We converted these pairwise comparisons to ranked scores using the Microsoft Trueskill algorithm [8]. Trueskill converges to a stable estimate of ranked scores after 12–36 comparisons, so we had enough comparisons to obtain accurate scores. We call this score *AMT–Streetchange*. A higher value of *AMT–Streetchange* is indicative of a larger absolute physical change in the image pair, as observed by the AMT users.

Next, we obtained binned ranks between 1 and 30 for both *AMT–Streetchange* and the absolute value of *Streetchange* output of our algorithm. Comparing the two, we found a Spearman's rank correlation of 0.72 ($p < 1 \times 10^{-5}$) between them. These results indicate that the algorithm outputs on Streetchanges are consistent with human judgments on changes in urban environment.

For the validation experiment with graduate students in MIT's School of Architecture and Planning, we presented the students (number of participants = 3) with 100 image pairs, where 50 image pairs contained large negative Streetchange and 50 image pairs contained large positive Streetchange. We asked them to choose if the image pair shows signs of positive change, negative change, or no change, and aggregated their responses with a majority rule. The students agreed with positive Streetchange (as scored by our algorithm) for 88% of the image pairs and they agreed with negative Streetchange for 59% of the image pairs. The lower agreement in the negative change was a result of students classifying demolition of blighted properties as a positive change (while our algorithm tends to classify them as negative).

Table S1. Summary Statistics for the Boston Validation Experiment (Section 1.7) (N = 222)

Variables	Description	Mean	SD	Min	Max
Streetchange 2007–2014	Mean Streetchange 2007–2014 of all sampled street blocks within a census tract	1.290	0.509	−1.599	2.613
Log Total Square Footage 2012–2014	Total square footage built per square mile within a census tract	13.527	1.548	10.109	17.202

Table S2. Streetscore: Generalization Performance (R^2)

Train \ Test	New York	Boston
New York	**0.5399**	0.5033
Boston	0.5028	**0.5384**

Validation with infrastructure development data: In the third validation experiment, we tested the relationship between improvements in Streetscore and improvements in infrastructure, using data on new developments in Boston (for summary statistics, see Table S1). We collected data on all public and private building projects from the Boston Planning and Development Agency (BPDA). We computed the total new square footage built per square mile for each census tract during the period 2012–2014 and tested its relationship with Streetchange 2007–2014. We expect census tracts where more square footage was built during 2012–2014 to be associated with a higher Streetchange, due to the physical improvements in these neighborhoods between the 2007 and 2014 image panels. And indeed, we find that infrastructure improvements are positively and significantly associated to Streetchange—one standard deviation increase in log total square footage corresponds to roughly half a standard deviation increase in Streetchange. We estimate:

$$\textit{Streetchange 2007 - 2014} = \underset{(0.069)}{1.620} + \underset{(0.035)}{0.159^{***}} \cdot \textit{Log total square footage} \tag{1}$$

These results provide empirical evidence for the connection between improvements in Streetscore and improvements in infrastructure.

1.8. Streetscore: Generalization Performance

Computer vision algorithms might have difficulty generalizing to out-of-sample data. Since we compute Streetscores for images from Baltimore, Detroit, and Washington DC using an algorithm trained with images from Boston and New York, we would like to estimate whether the Streetscore predictor can generalize without a significant drop in accuracy. So we performed an experiment where we trained a Streetscore predictor using images just from New York and measured the accuracy of its predictions on Boston images in our dataset, and vice versa. We found that the R^2 drops by only 0.036 on average during cross-city prediction (Table S2).

2. Regressions

We calculated Streetchange for 2007–2014 image pairs sampled uniformly from Baltimore, Boston, Detroit, New York, and Washington DC. Using tract boundaries from the 2010 US Census, we aggregated 2007 Streetscore and 2007–2014 Streetchange across each census tract. We obtained census tract characteristic data from the 2000 US Census, adjusted to the 2010 census tract boundaries [10]. For summary statistics, see Table S3.

2.1. Do Social Characteristics Predict Changes in Streetscore?

We now present the cross-sectional demographic and economic correlates of the 2007 Streetscore and 2007–2014 Streetchange (Table S4). For each census tract, we considered the following socioeconomic indicators from the 2000 US Census: population density, level of education, median income, housing price and rental costs, housing vacancy, race, and poverty. We find that the socioeconomic characteristics that best predict higher Streetscore in 2007—density and education—are also the best predictors of increases in Streetscore between 2007 and 2014. These relationships hold regardless of whether we control for the 2007 Streetscore or other variables. We find that other variables, such as income, housing prices, rent, race, and poverty have little or no predictive power in our context.

2.2. The Filtering Hypothesis of Urban Change

In addition to the invasion and tipping hypotheses of urban change discussed in main text, we evaluated the filtering hypothesis [11] of urban change. The filtering hypothesis suggests cycles in which neighborhoods gradually decay until they get upgraded. To test the hypothesis that building age shapes streetscape change, we regressed the 2007 Streetscore and 2007–2014 Streetchange on the shares of the building stock (as of the year 2000) built during different decades. The data grouped together all buildings erected before 1939. There is weak support for the filtering hypothesis in our dataset (Table S5): we found that neighborhoods with newer housing stocks score higher than neighborhoods built in the 1950s. However, we cannot rule out the possibility that our finding is also reflective of differences in the perception of various architectural styles, as neighborhoods built before 1939 (prior to the widespread adoption of modernist architecture in the US) also score highly.

3. Additional Examples and Map Visualizations

Figures S5–S7 present additional examples of positive Streetchange from the five cities in our dataset. The examples show that Streetchange is able to detect both upgrading and new construction. Figure S8 shows additional examples of negative Streetchange, which are associated with urban blight and decline in upkeep.

Figures S9–S28 present map visualizations for Log Population Density 2000, Share College Education 2000, Streetscore 2007, and Streetchange 2007–2014 for the five cities in our dataset.

References

[1] Hoiem D, Efros AA, Hebert M (2008) Putting objects in perspective. *International Journal of Computer Vision* 80(1):3–15.

[2] Malik J, Belongie S, Leung T, Shi J (2001) Contour and texture analysis for image segmentation. *International Journal of Computer Vision* 43(1):7–27.

[3] Oliva A, Torralba A (2001) Modeling the shape of the scene: A holistic representation of the spatial envelope. *International Journal of Computer Vision* 42(3):145–175.

[4] Guo D, Cheng Y, Zhuo S, Sim T (2010) Correcting over-exposure in photographs. *IEEE Conference on Computer Vision and Pattern Recognition* pp. 515–521.

[5] Shirvaikar MV (2004) An optimal measure for camera focus and exposure. *Southeastern Symposium on System Theory* pp. 472–475.

[6] Salesses P, Schechtner K, Hidalgo CA (2013) The collaborative image of the city: Mapping the inequality of urban perception. *PLoS One* 8(7):e68400.

[7] Naik N, Philipoom J, Raskar R, Hidalgo CA (2014) Streetscore – Predicting the perceived safety of one million streetscapes. *IEEE Conference on Computer Vision and Pattern Recognition Workshops* pp. 793–799.

[8] Herbrich R, Minka T, Graepel T (2006) Trueskill: A Bayesian skill rating system. *Advances in Neural Information Processing Systems* pp. 569–576.

[9] Schölkopf B, Smola AJ, Williamson RC, Bartlett PL (2000) New support vector algorithms. *Neural Computation* 12(5):1207–1245.

[10] Logan JR, Xu Z, Stults BJ (2014) Interpolating US decennial census tract data from as early as 1970 to 2010: A longitudinal tract database. *The Professional Geographer* 66(3):412–420.

[11] Margolis SE (1982) Depreciation of housing: An empirical consideration of the filtering hypothesis. *Review of Economics and Statistics* 64(1):90–96.

Table S3. Summary Statistics ($N = 2513$)

Variables	Description	Mean	SD	Min	Max
Streetscore 2007	Mean Streetscore 2007 of all sampled street blocks within a census tract	7.757	2.587	1.681	18.93
Streetchange 2007–2014	Mean Streetchange 2007–2014 of all sampled street blocks within a census tract	1.39	0.779	−4.076	6.121
Log Population Density 2000	Log of population density within a census tract	−4.655	1.22	−15.29	−2.48
Share College Education 2000	Share of adults with a four-year college degrees within a census tract	0.254	0.216	0	1
Log Median Income 2000	Log of median income of adults within a census tract	4.54	0.206	3.884	5.276
Log Monthly Rent 2000	Log of median monthly rent within a census tract	6.40	0.390	4.595	7.601
Log Housing Price 2000	Log of median housing price within a census tract	5.223	0.311	3.938	6
Poverty Rate 2000	Share of households under poverty line within a census tract	0.218	0.137	0	1
Share African-American 2000	Share of African-Americans within a census tract	0.366	0.371	0	1
Share Hispanic 2000	Share of Hispanics within a census tract	0.192	0.221	0	0.927
Share Vacant Units 2000	Share of vacant units within a census tract	0.036	0.055	0	0.348
Share Built 1990-2000	Share of housing stock built during 1990-2000 within a census tract	0.587	0.874	0	1
Share Built 1980-1989	Share of housing stock built during 1980-1989 within a census tract	0.404	0.573	0	0.659
Share Built 1970-1979	Share of housing stock built during 1970-1979 within a census tract	0.694	0.744	0	0.719
Share Built 1960-1969	Share of housing stock built during 1960-1969 within a census tract	0.123	0.983	0	1
Share Built 1950-1959	Share of housing stock built during 1950-1959 within a census tract	0.164	0.105	0	0.748
Share Built 1940-1949	Share of housing stock built during 1940-1949 within a census tract	0.171	0.956	0	0.815
Share Built before 1940	Share of housing stock built before 1940 within a census tract	0.394	0.201	0	1

All socioeconomic variables are from the 2000 US Census. Streetscore 2007 and Streetchange 2007–2014 are computed using the method described in Section 1 of this document. All data are aggregated at the census tract level.

Table S4. Relationship Between Social Characteristics and Changes in Streetscore

Independent Variables	**Coefficient**	
	Streetscore 2007	Streetchange 2007–2014
Share College Education 2000	2.024***	1.099***
	(0.483)	(0.162)
Log Population Density 2000	0.765***	0.080***
	(0.097)	(0.026)
Log Median Income 2000	0.186	−0.118
	(0.920)	(0.242)
Log Monthly Rent 2000	1.191***	−0.149
	(0.260)	(0.096)
Log Housing Price 2000	0.986***	0.175*
	(0.274)	(0.090)
Poverty Rate 2000	3.815***	0.300
	(1.114)	(0.312)
Share African-American 2000	0.151	−0.002**
	(0.225)	(0.001)
Share Hispanic 2000	0.878**	0.312***
	(0.365)	(0.118)
Share Vacant Units 2000	0.917***	0.090***
	(0.158)	(0.033)
Streetscore 2007		−0.012
		(0.013)

All models control for city fixed effects. $***p < 0.01, **p < 0.05, *p < 0.1$
Standard errors corrected for spatial correlation are in parentheses. Regressions are estimated with a constant that is not reported.

Table S5. Evidence of Filtering

Independent Variables	**Coefficient** Streetscore 2007	Streetchange 2007–2014
Share Built 1990-2000	4.390***	0.410
	(1.422)	(0.341)
Share Built 1980-1989	4.638***	0.677*
	(1.406)	(0.346)
Share Built 1970-1979	2.077**	0.079
	(0.948)	(0.259)
Share Built 1960-1969	2.367**	0.343
	(1.119)	(0.288)
Share Built 1940-1949	−0.195	-0.720**
	(1.227)	(0.324)
Share Built Before 1940	4.614***	0.323*
	(0.618)	(0.194)
Streetscore 2007		0.000
		(0.014)
Log Population Density 2000		0.106***
		(0.033)
Share College Education 2000		0.648***
		(0.106)

All models control for city fixed effects. $***p < 0.01, **p < 0.05, *p < 0.1$
Standard errors corrected for spatial correlation are in parentheses. Regressions are estimated with a constant that is not reported.

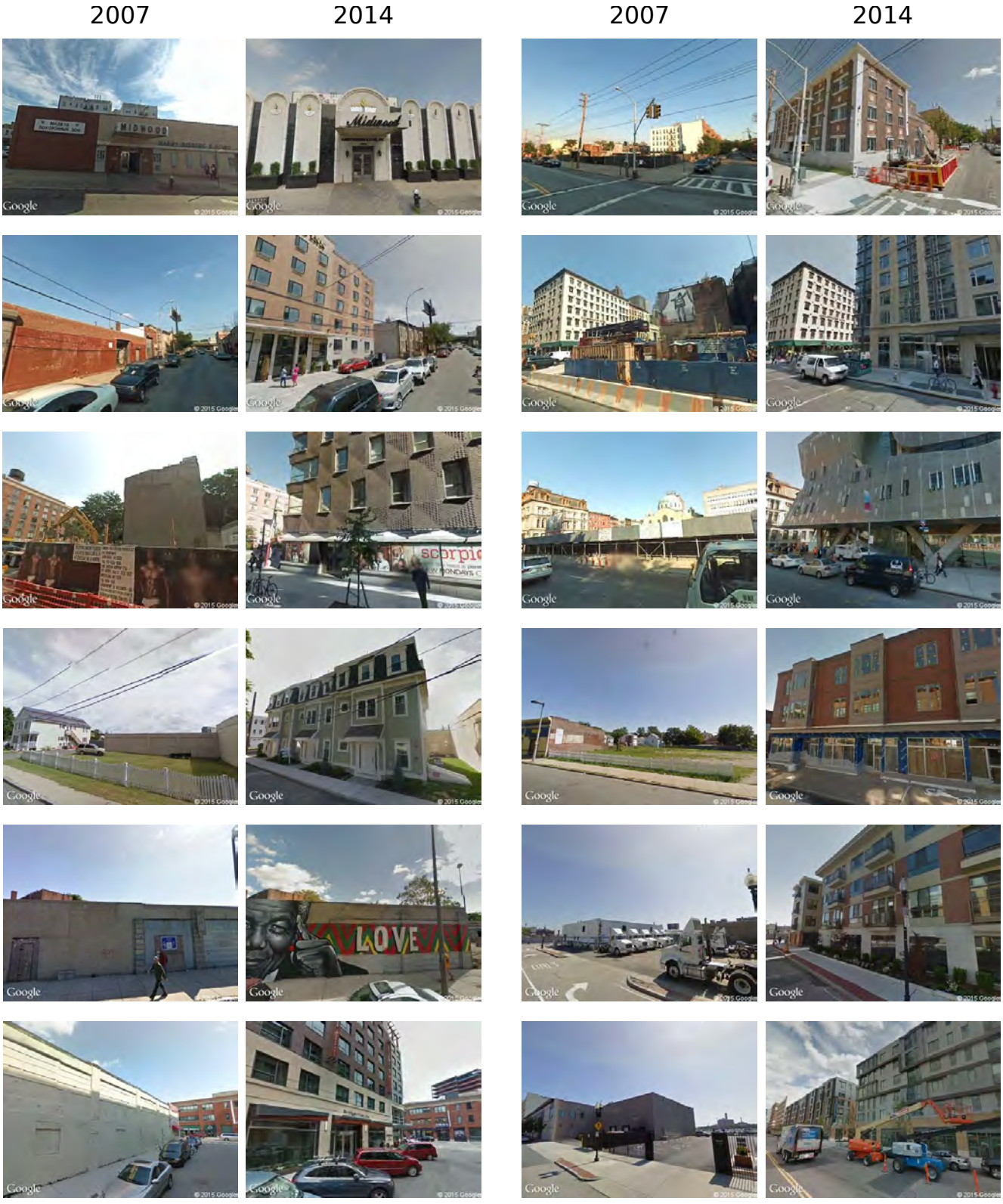

Figure S5. Additional examples of positive Streetchange. The first three rows show examples from New York City. The next three rows show examples from Boston. Images courtesy of Google, Inc.

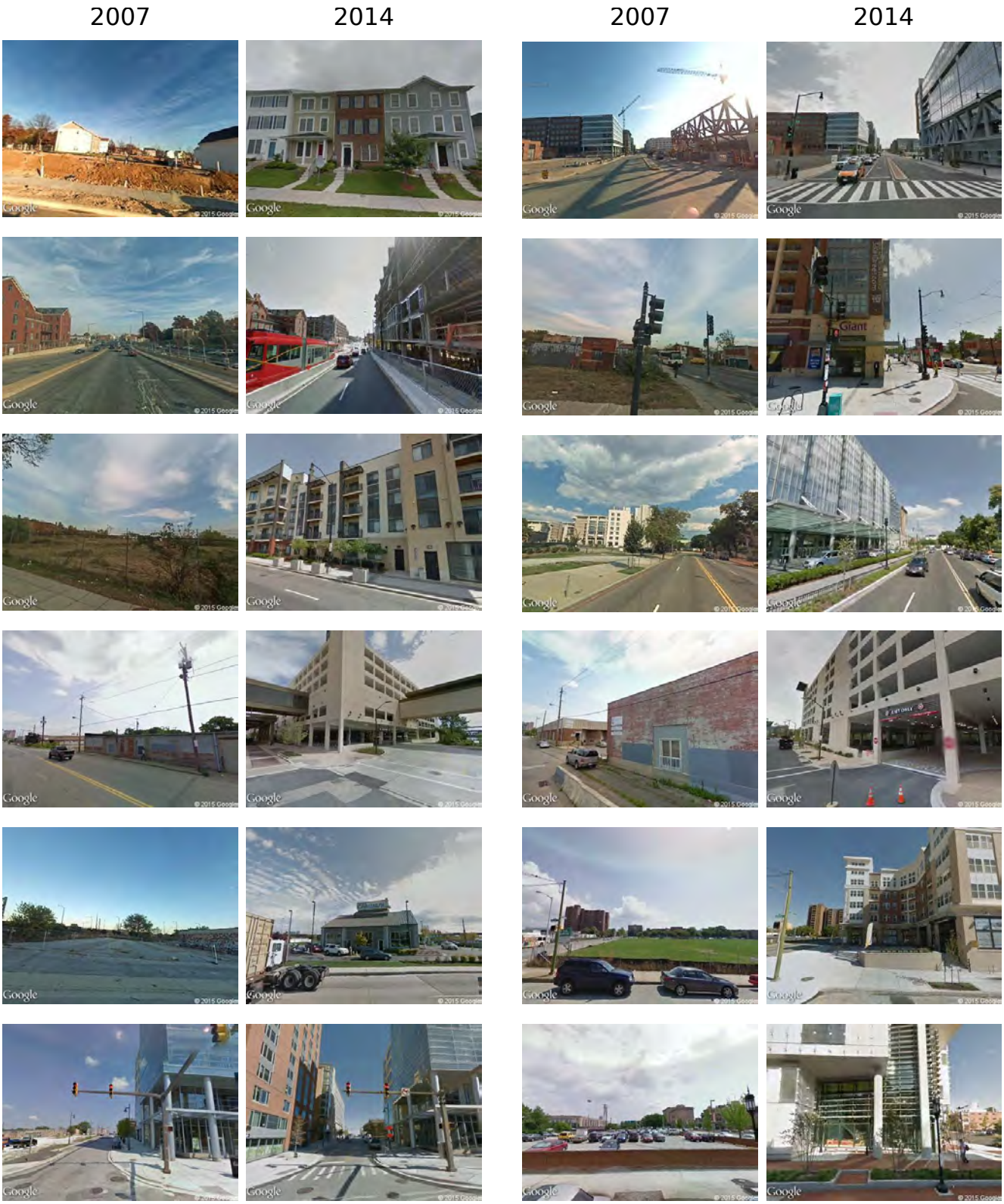

Figure S6. Additional examples of positive Streetchange. The first three rows show examples from Washington DC. The next three rows show examples from Baltimore. Images courtesy of Google, Inc.

Figure S7. Additional examples of positive Streetchange. All examples from Detroit. Images courtesy of Google, Inc.

Figure S8. Additional examples of negative Streetchange from the five cities in our dataset. Images courtesy of Google, Inc.

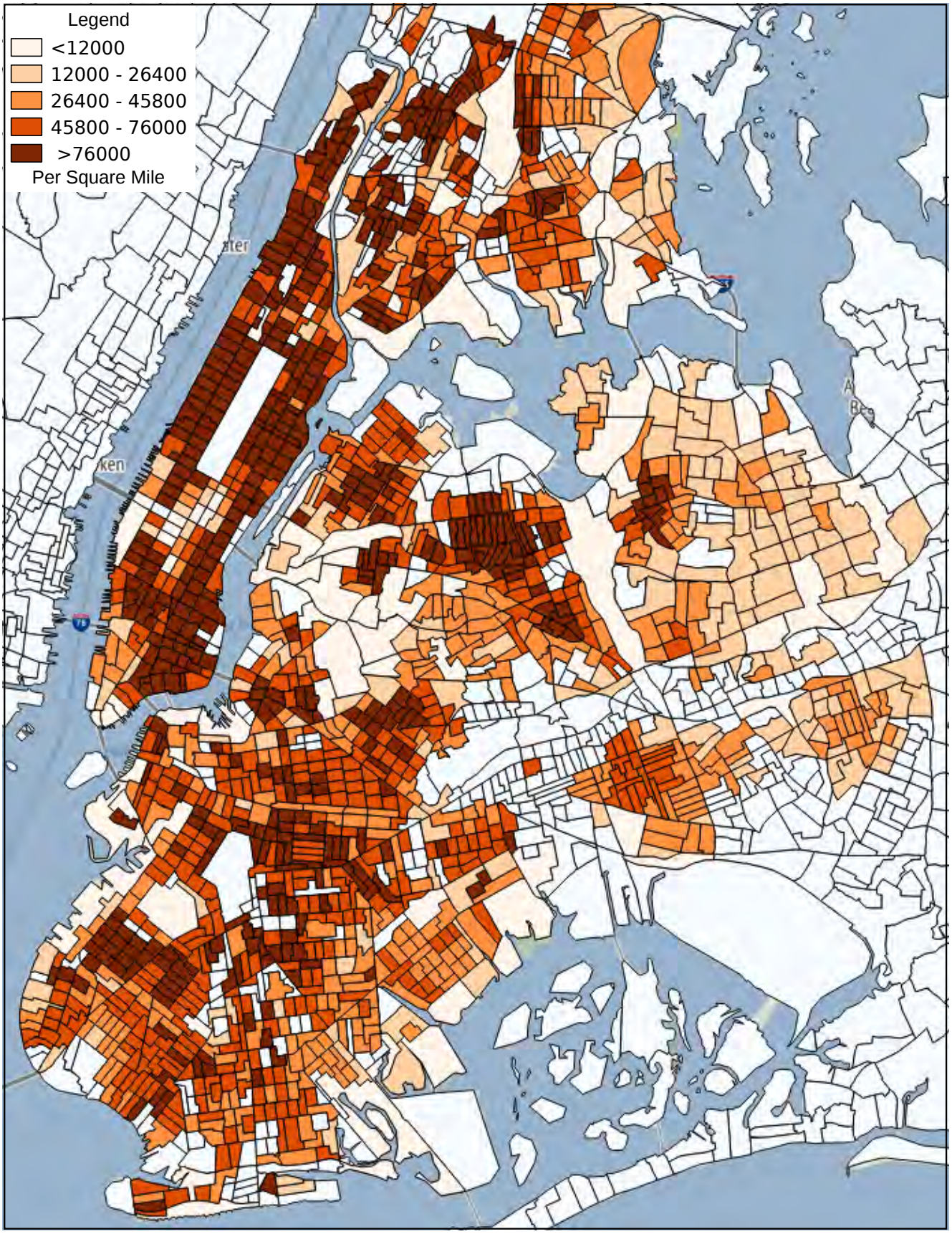

Figure S9. New York City: Log Population Density 2000.

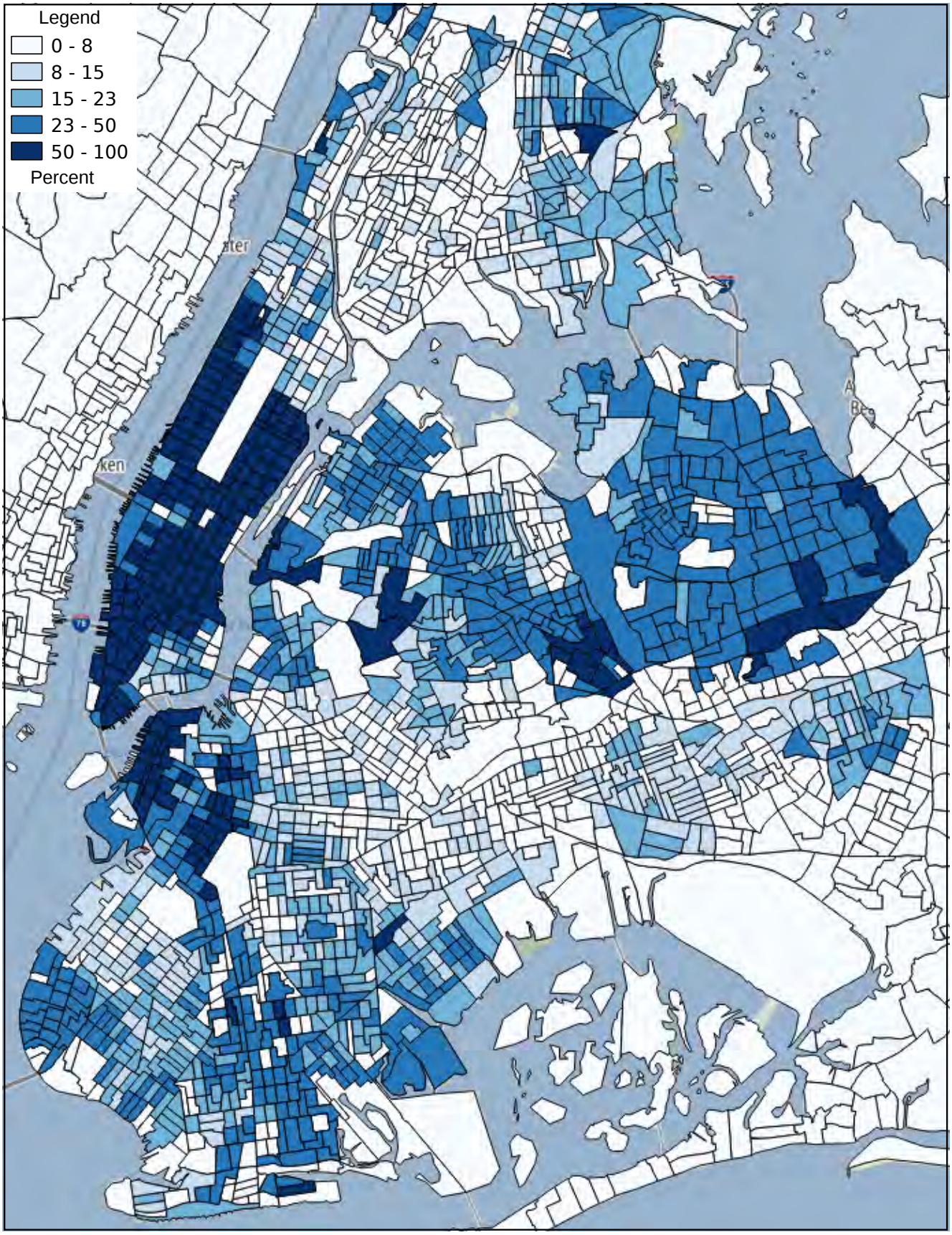

Figure S10. New York City: Share College Education 2000.

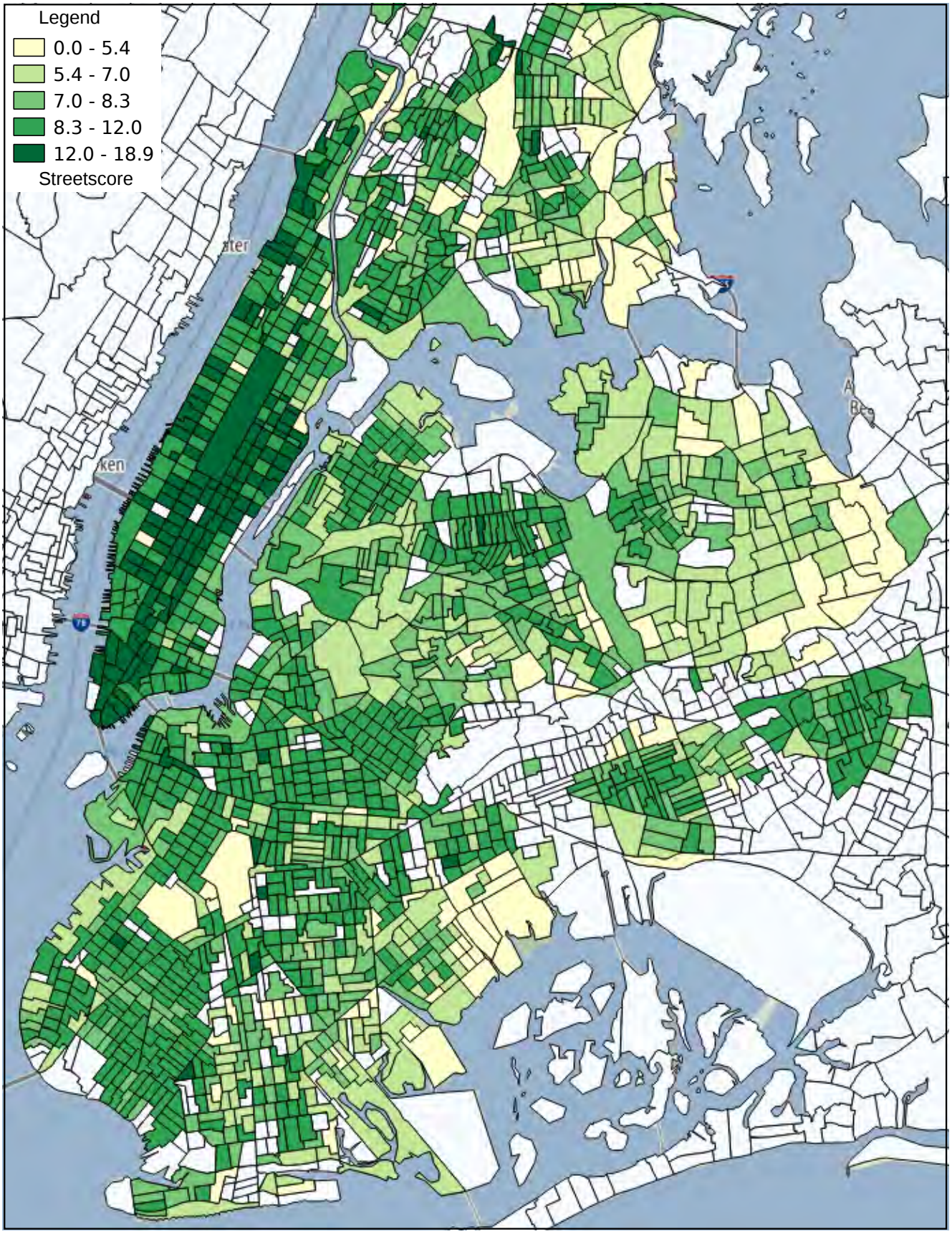

Figure S11. New York City: Streetscore 2007.

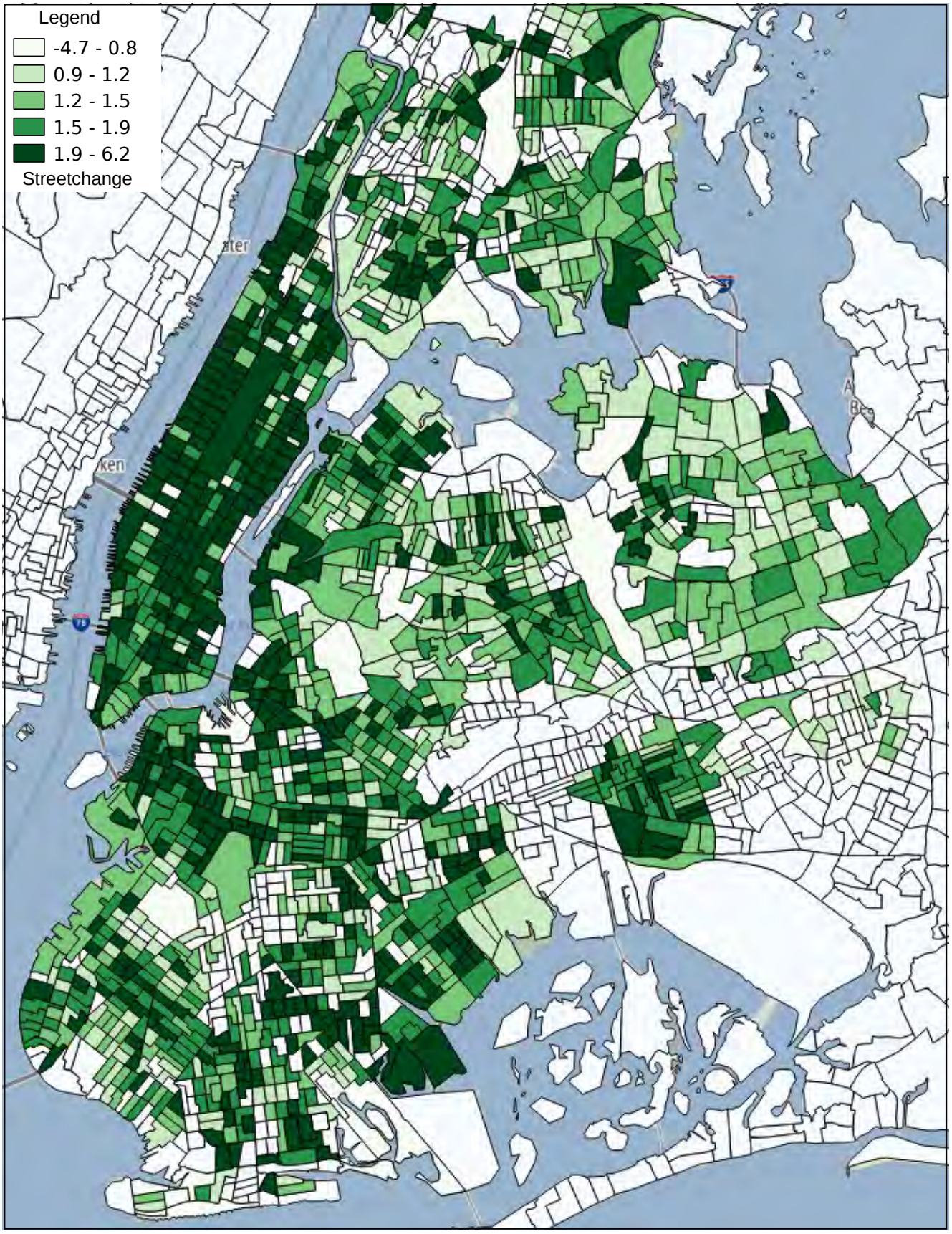

Figure S12. New York City: Streetchange 2007–2014.

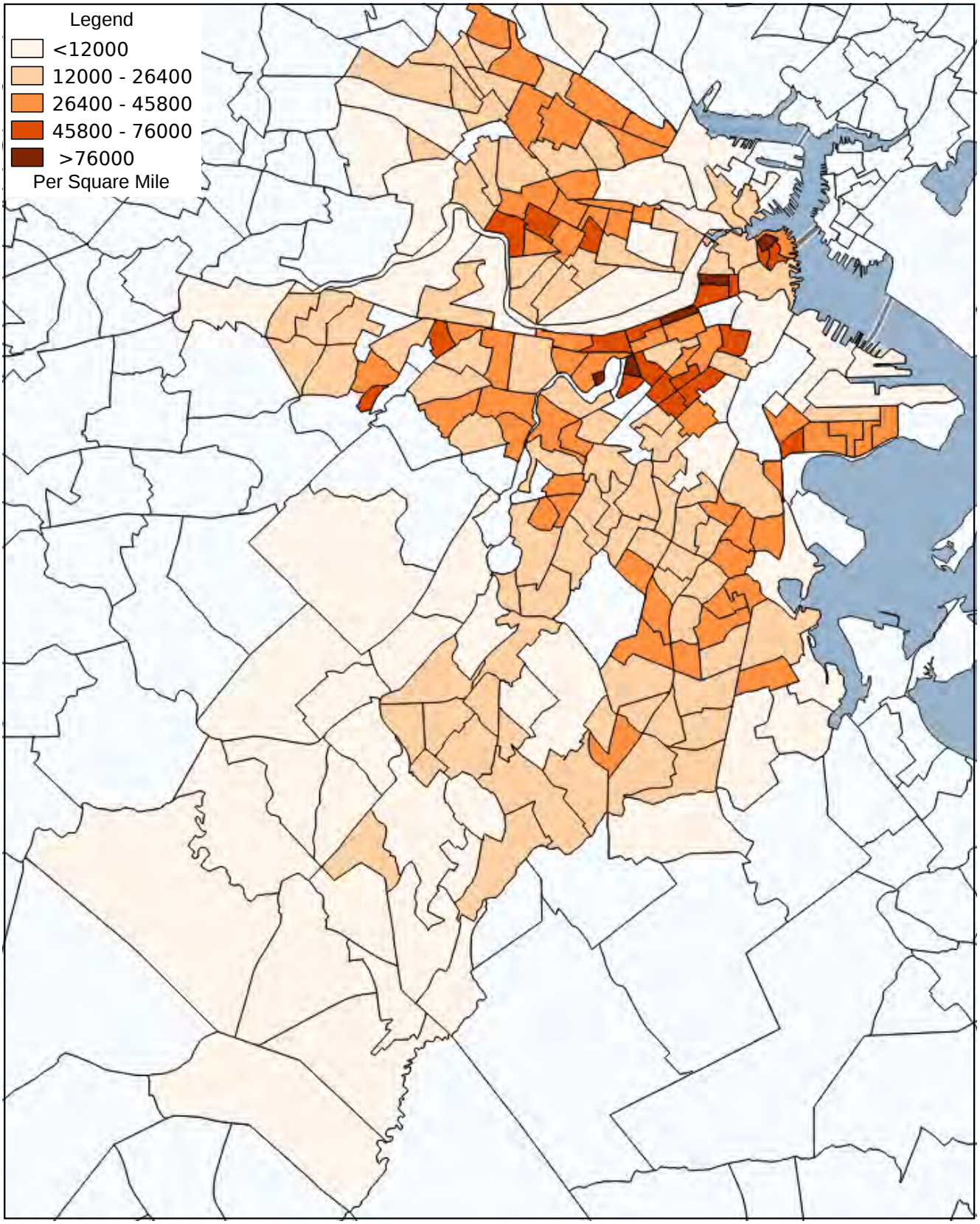

Figure S13. Boston: Log Population Density 2000.

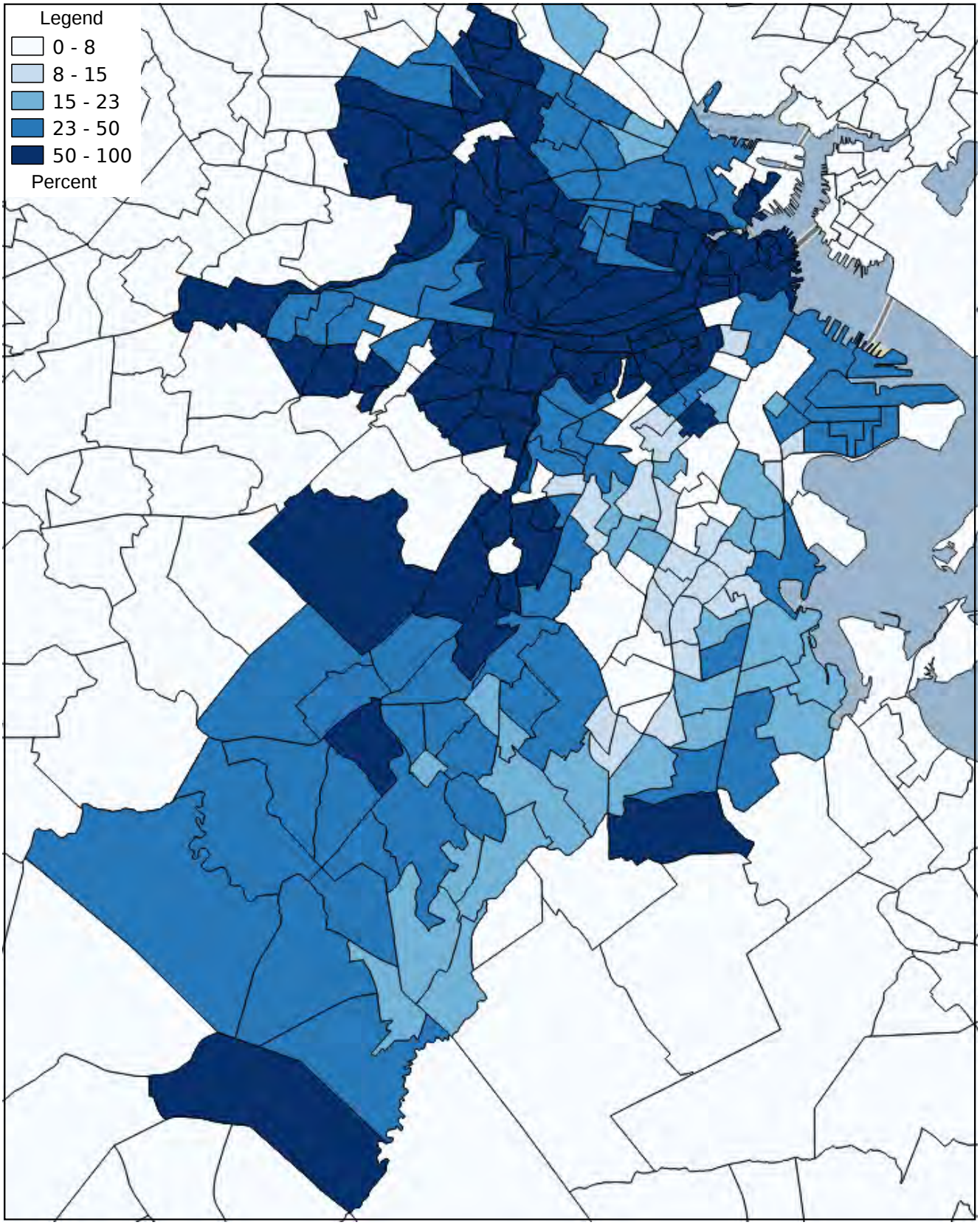

Figure S14. Boston: Share College Education 2000.

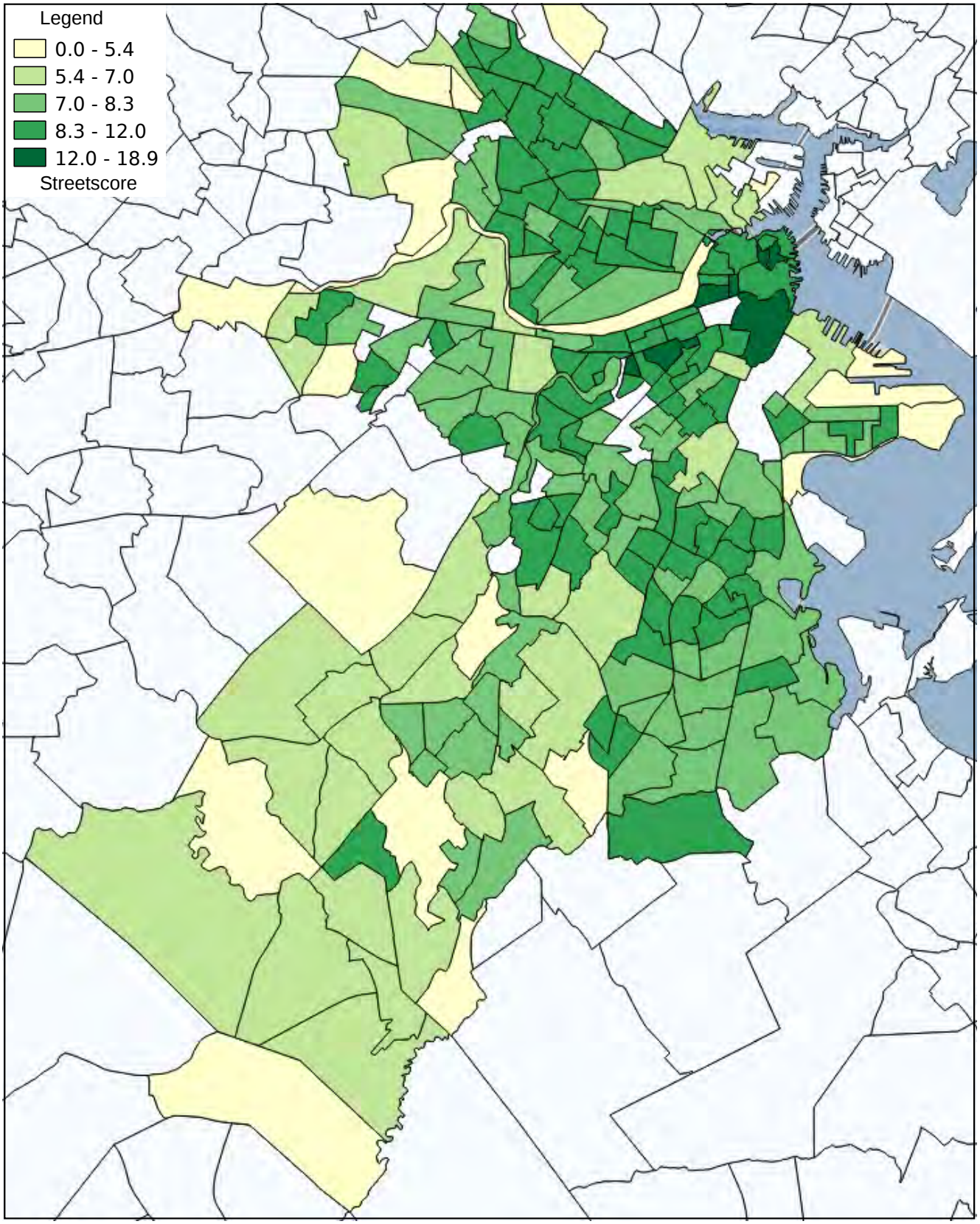

Figure S15. Boston: Streetscore 2007.

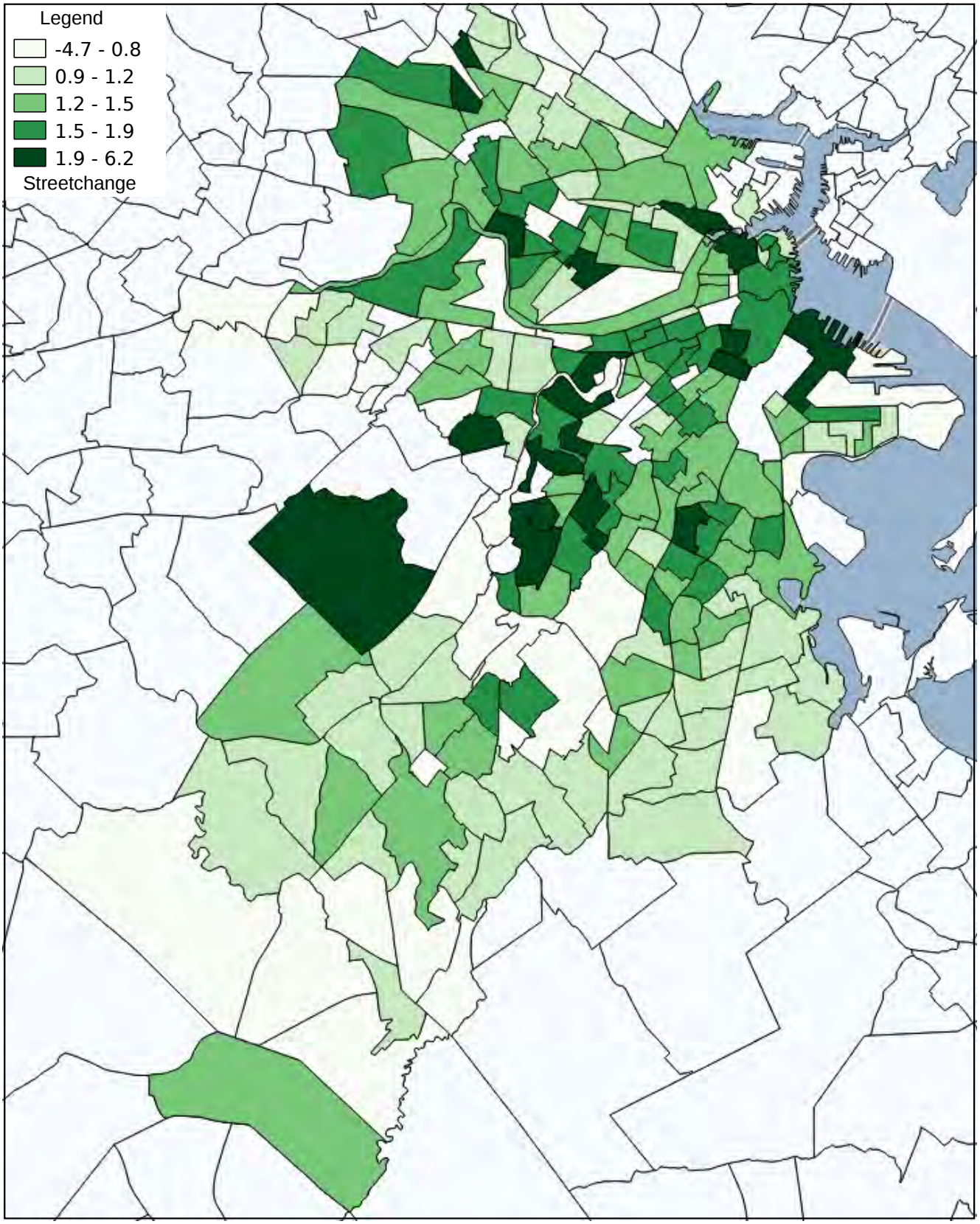

Figure S16. Boston: Streetchange 2007–2014.

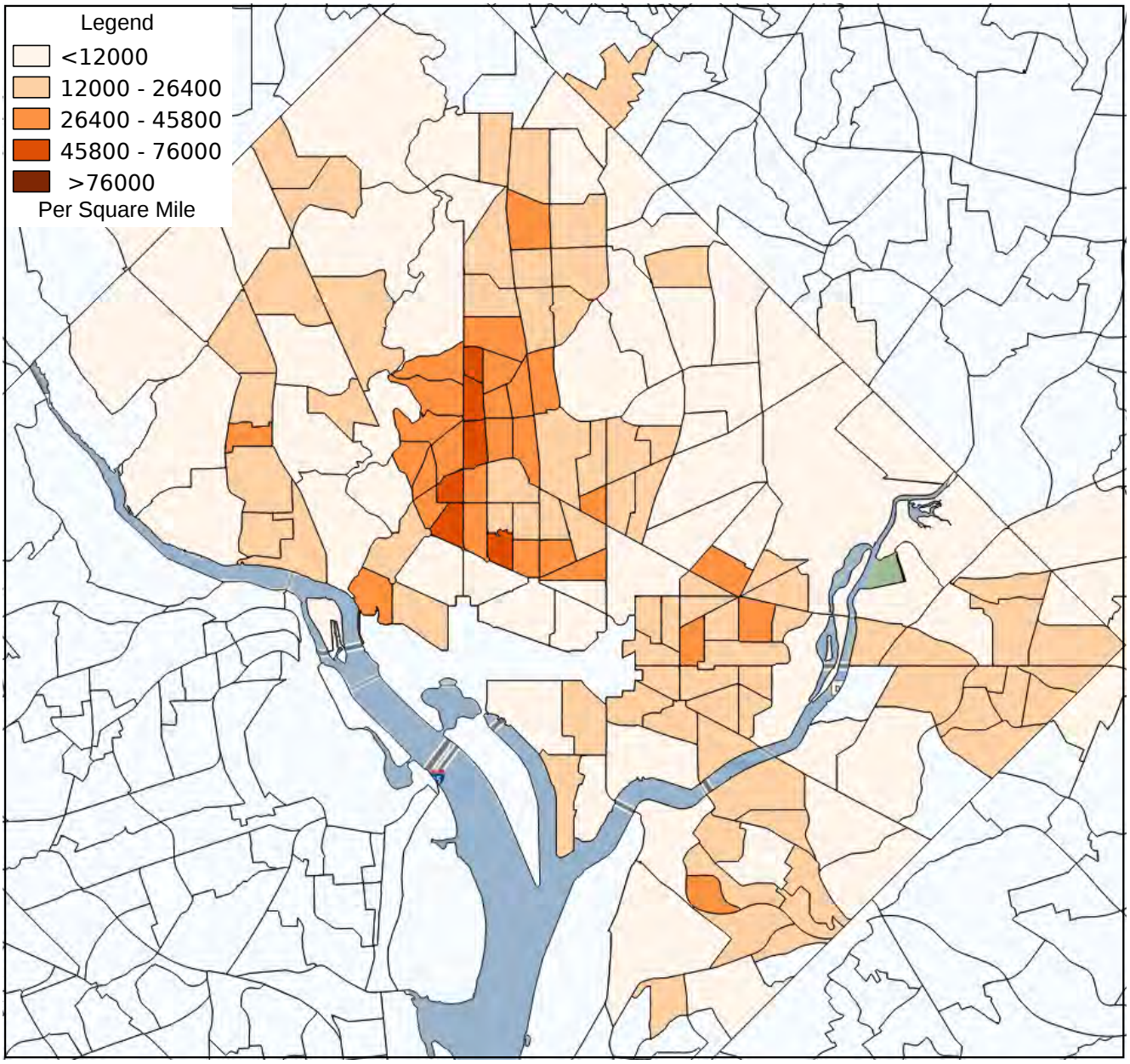

Figure S17. Washington DC: Log Population Density 2000.

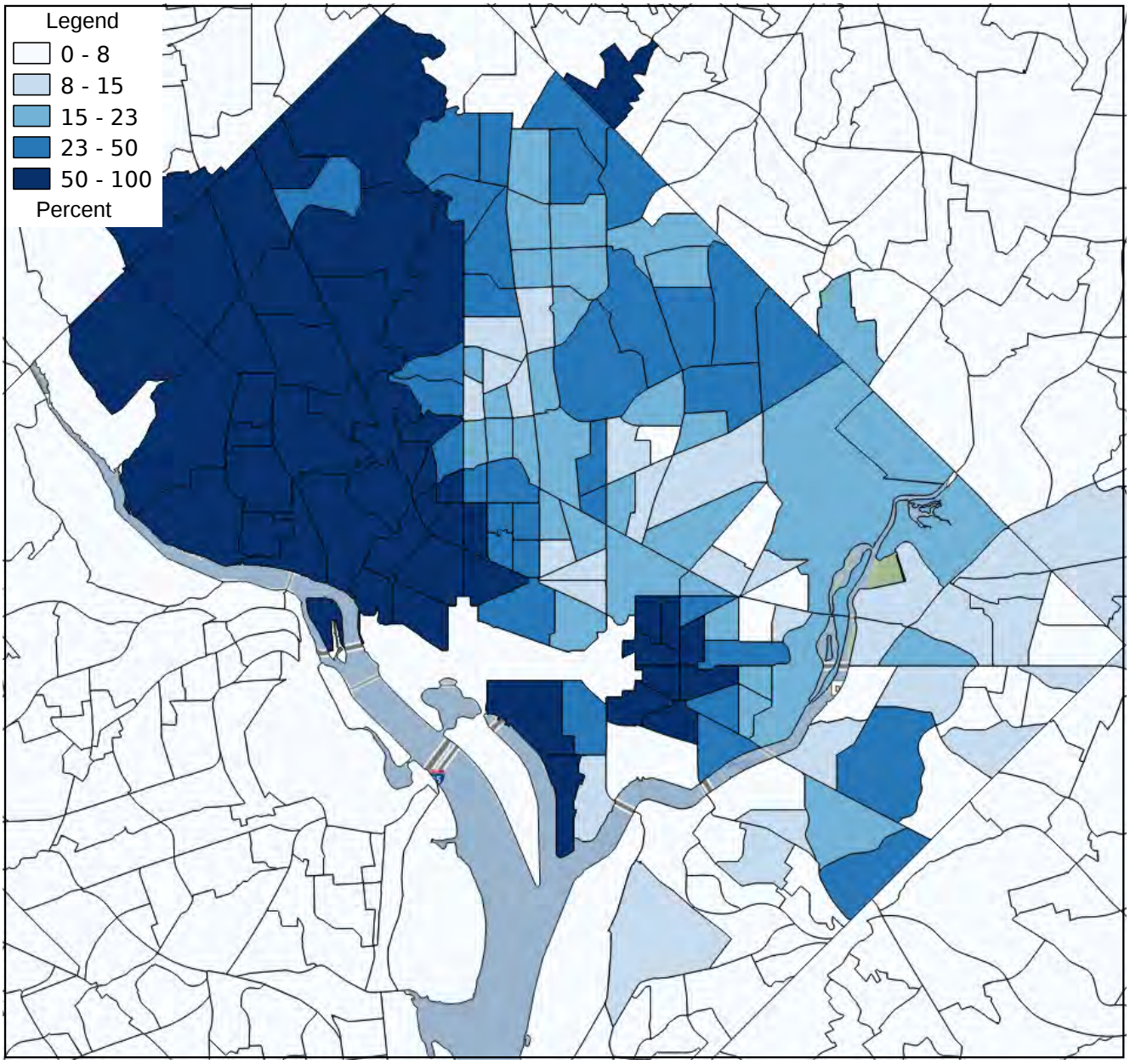

Figure S18. Washington DC: Share College Education 2000.

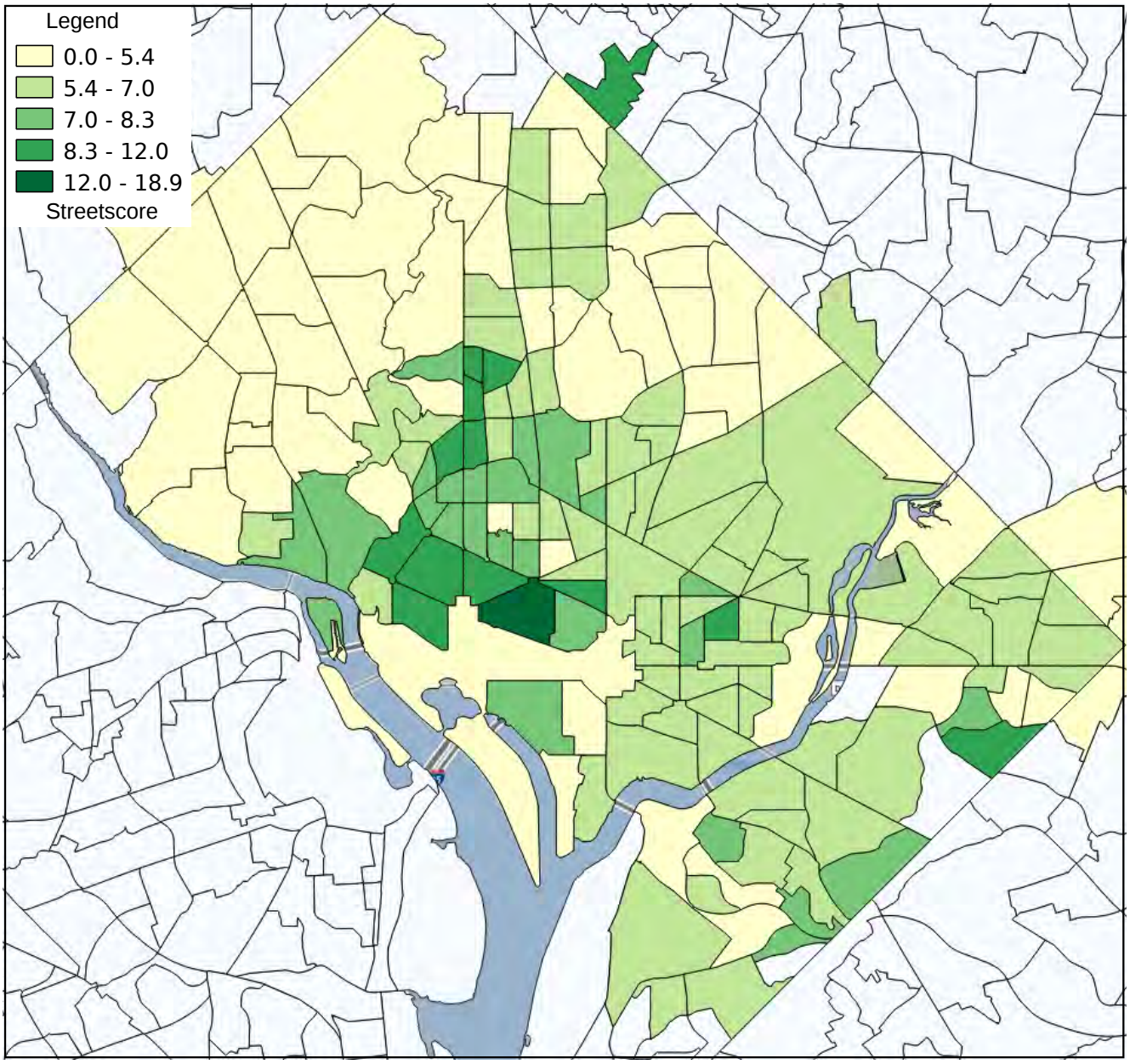

Figure S19. Washington DC: Streetscore 2007.

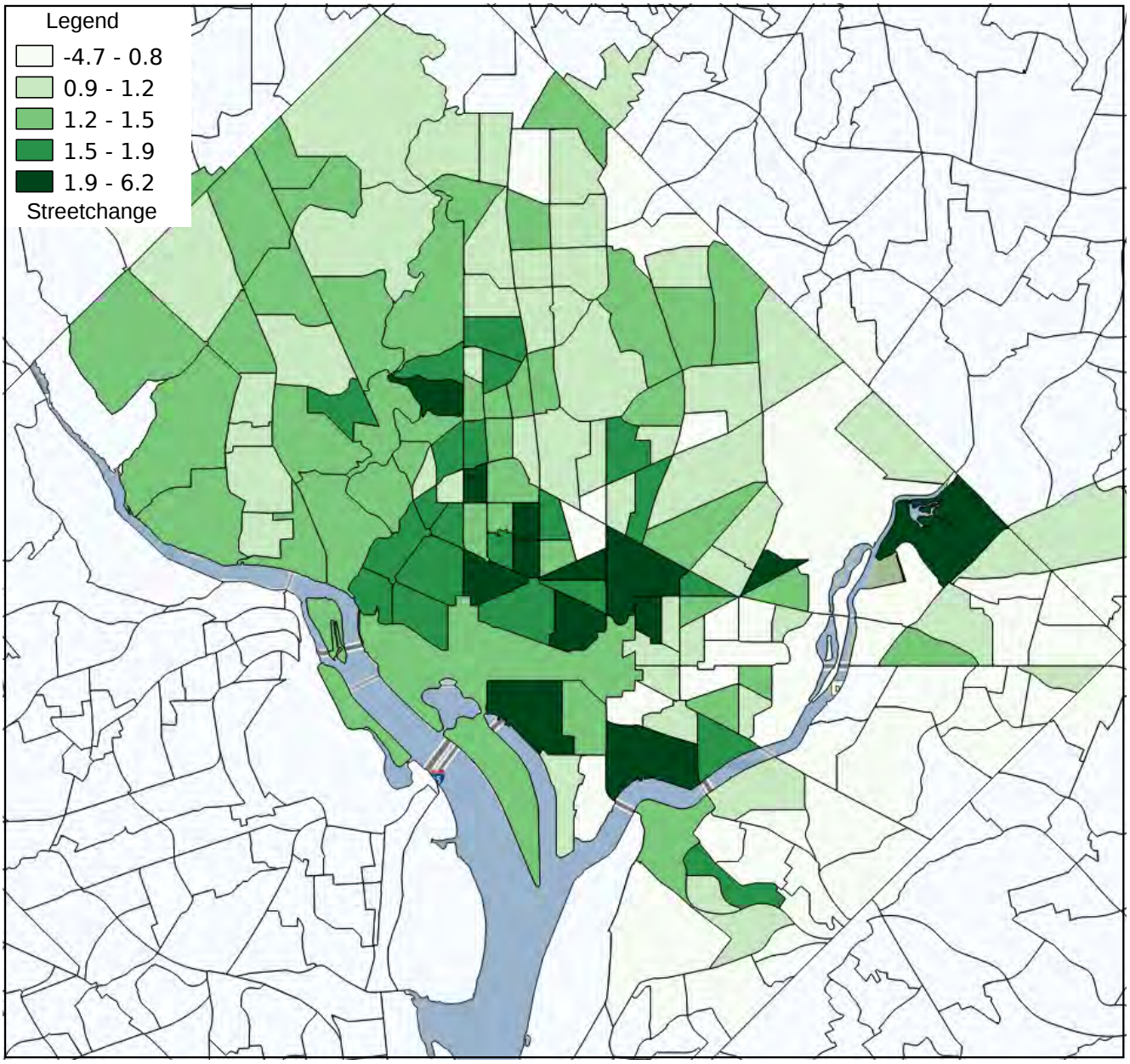

Figure S20. Washington DC: Streetchange 2007–2014.

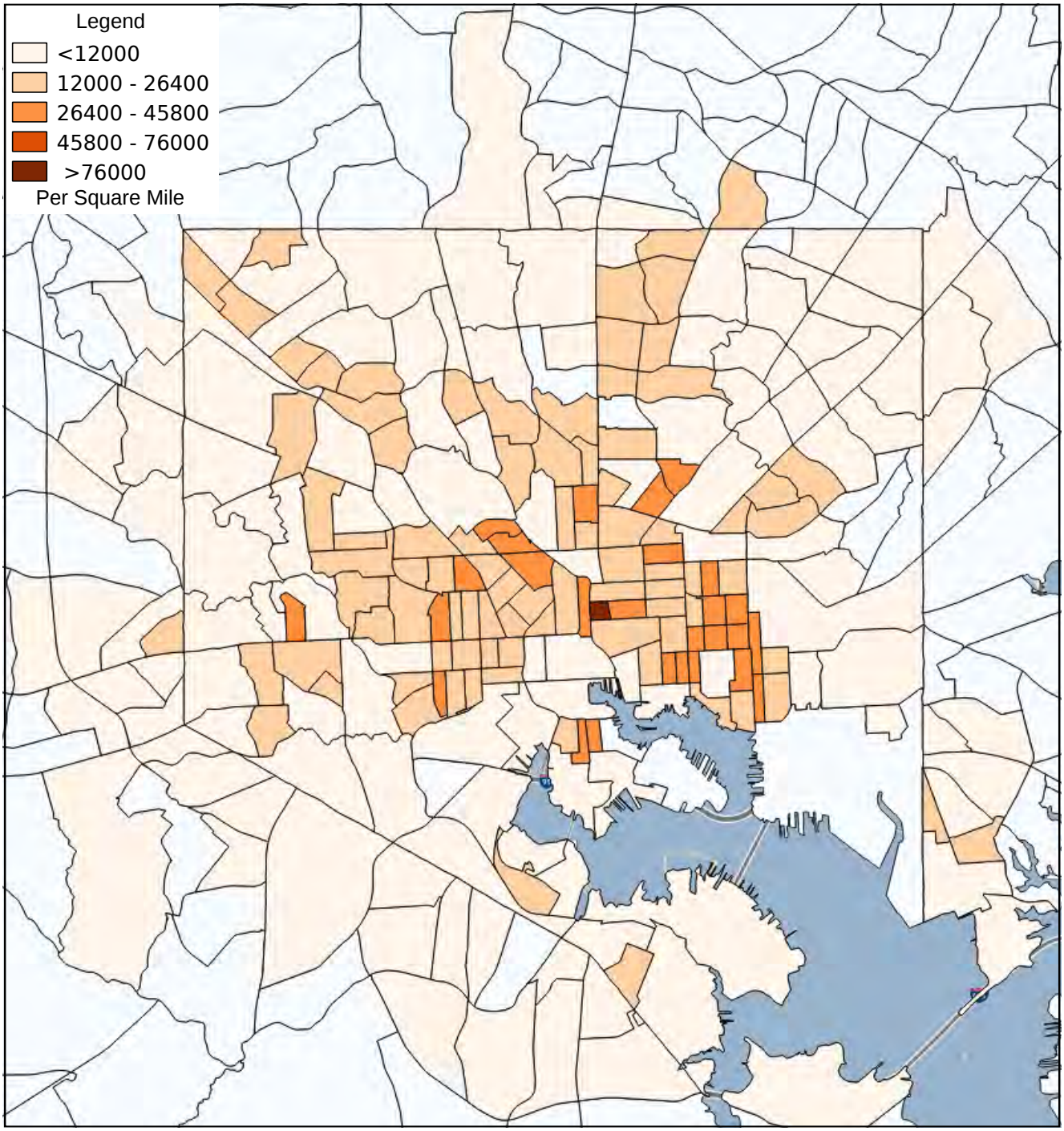

Figure S21. Baltimore: Log Population Density 2000.

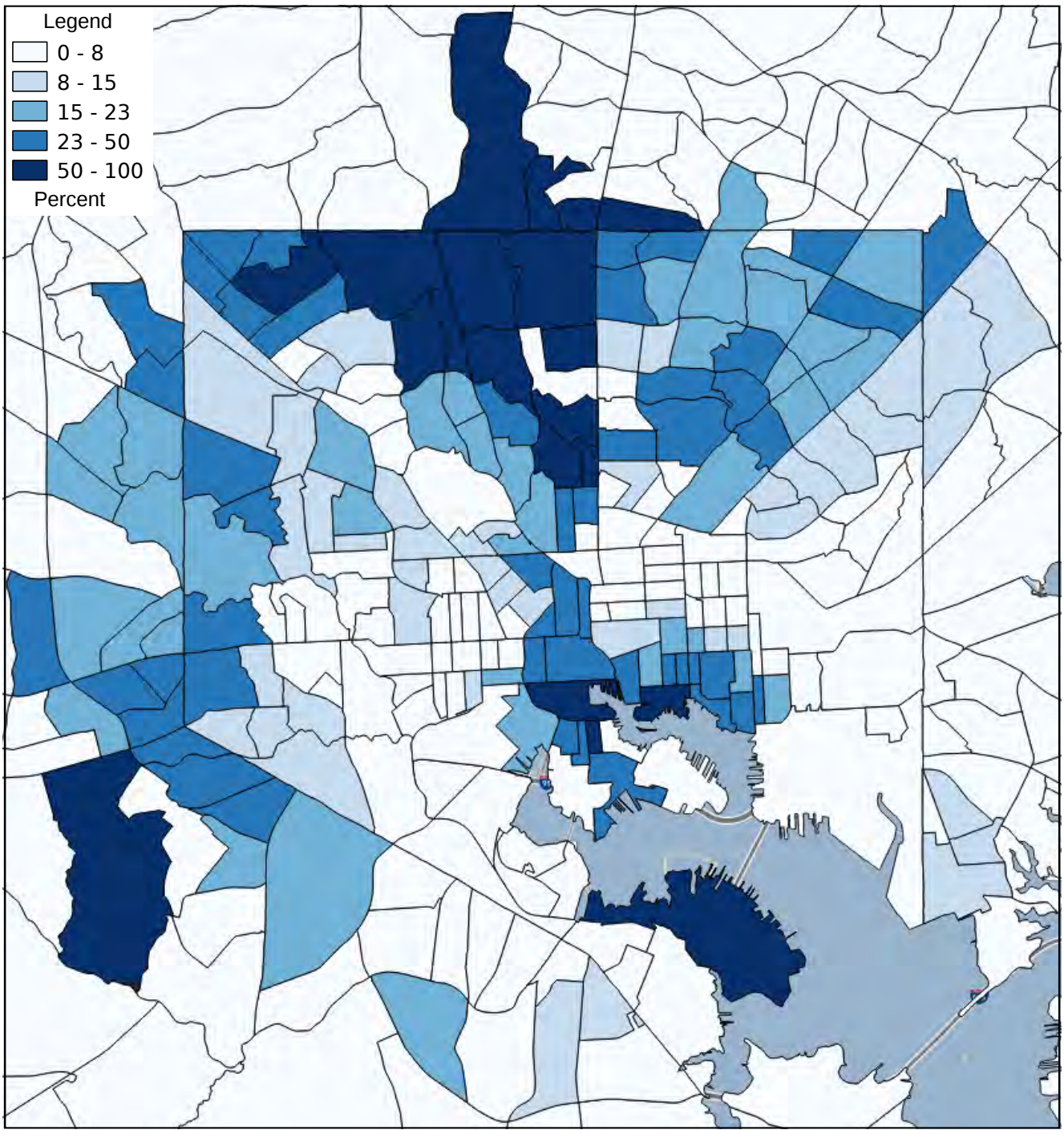

Figure S22. Baltimore: Share College Education 2000.

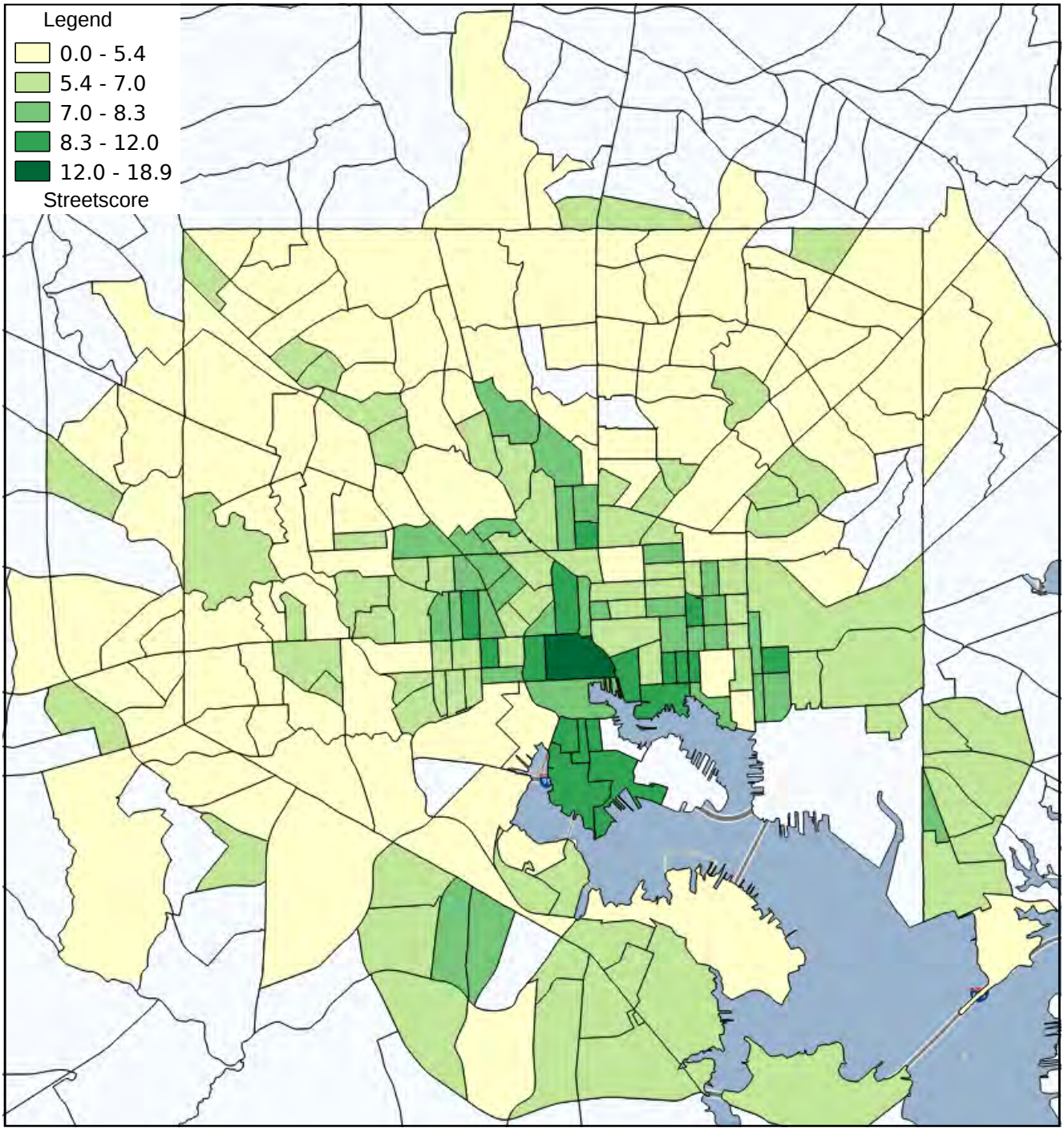

Figure S23. Baltimore: Streetscore 2007.

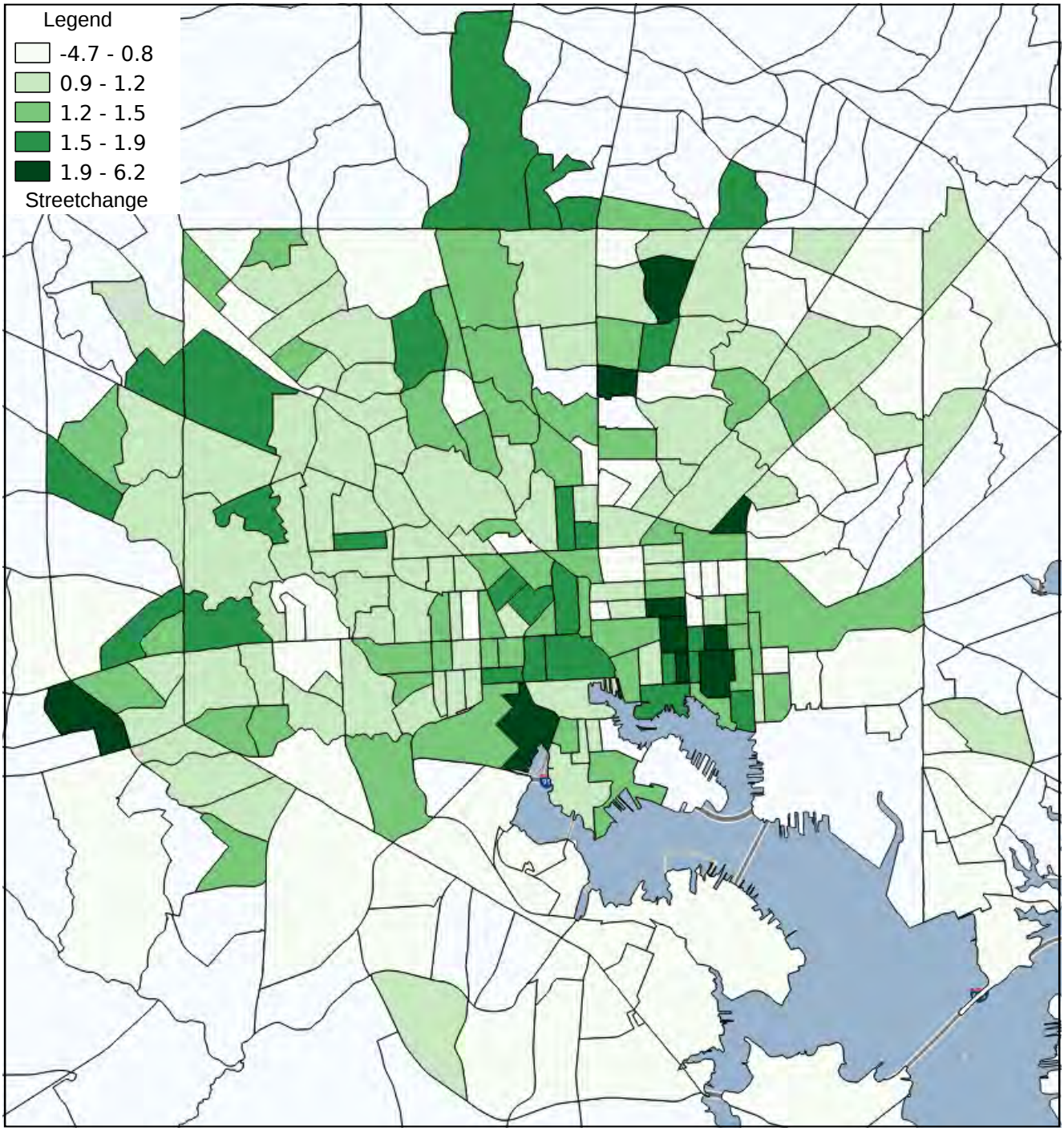

Figure S24. Baltimore: Streetchange 2007–2014.

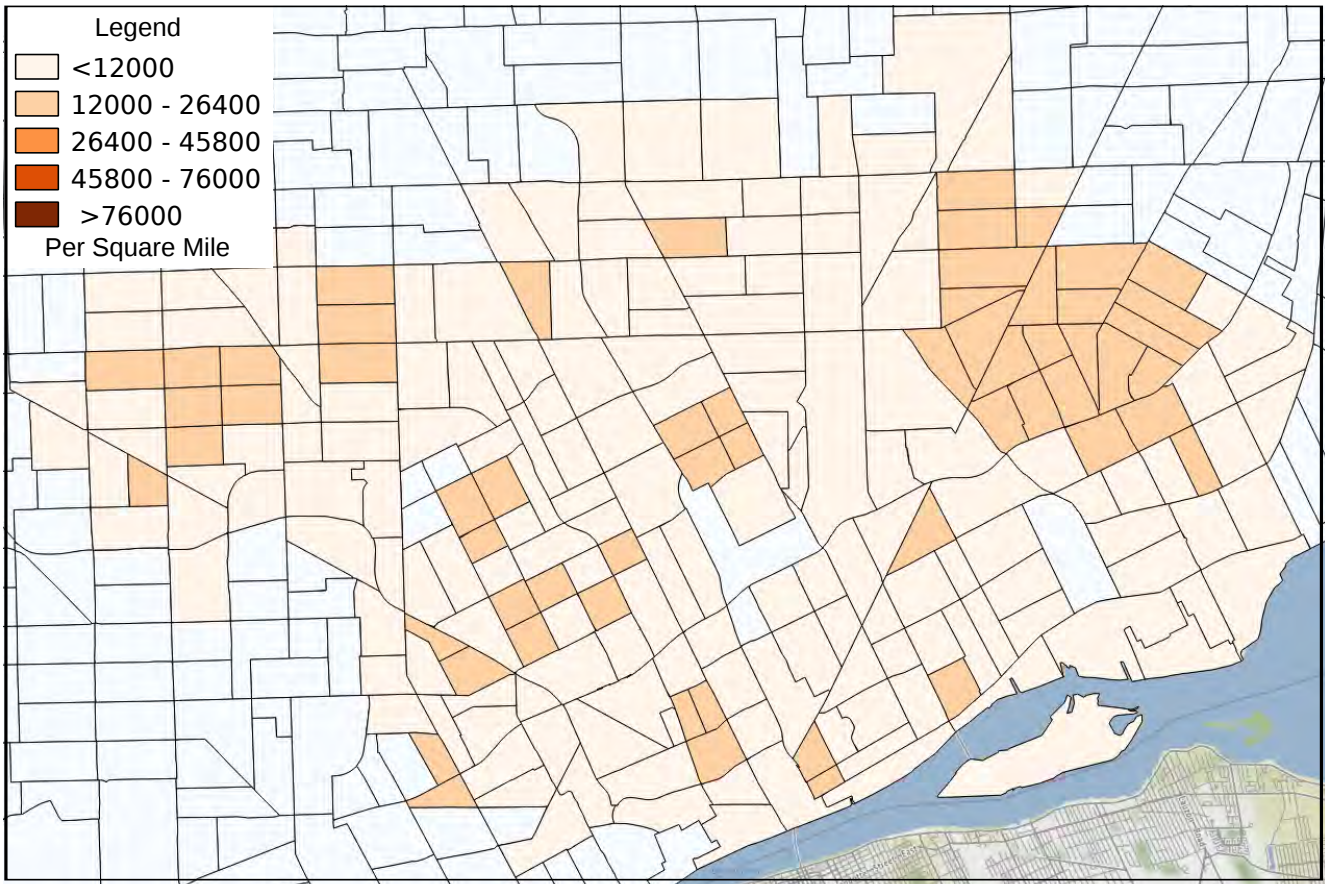

Figure S25. Detroit: Log Population Density 2000.

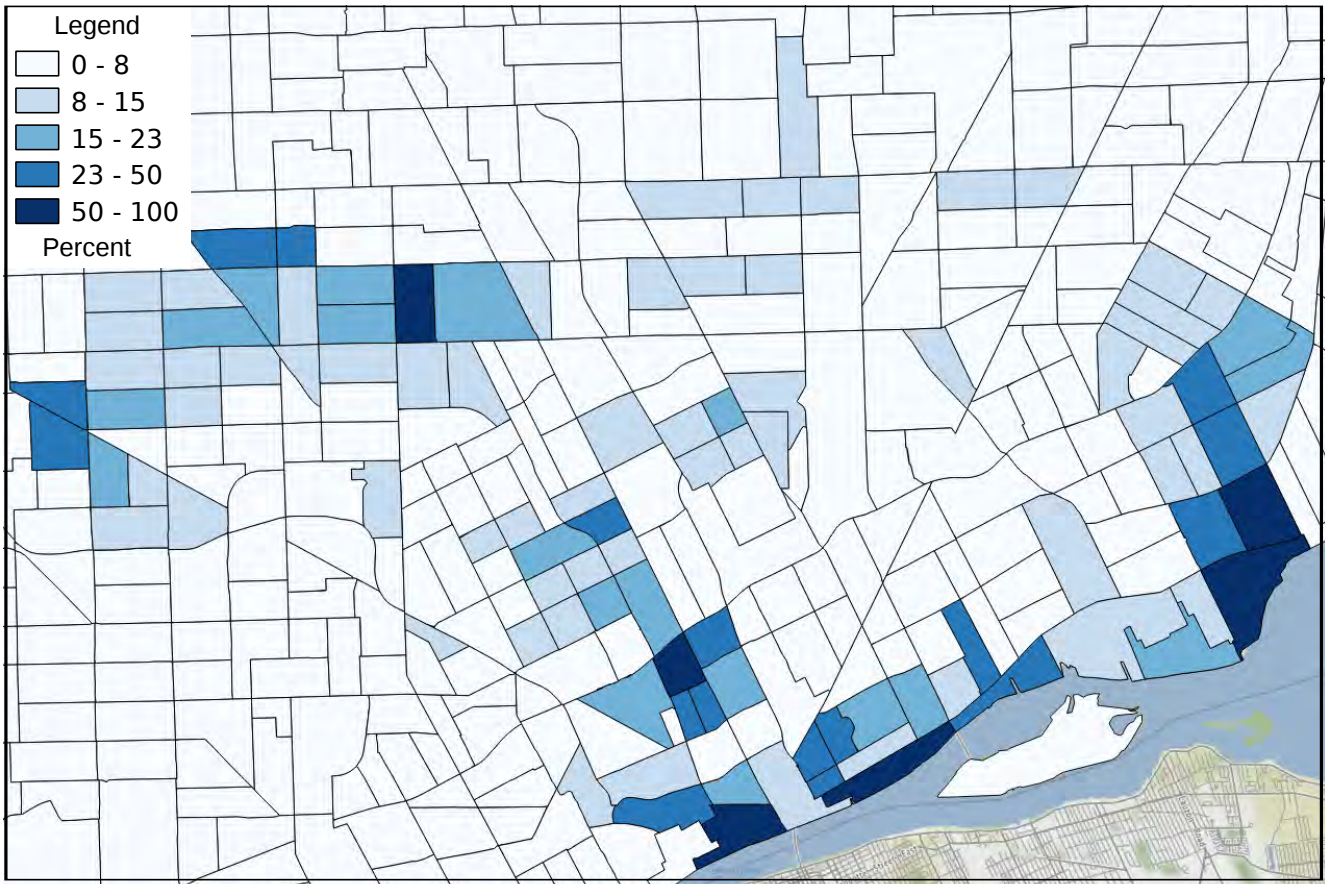

Figure S26. Detroit: Share College Education 2000.

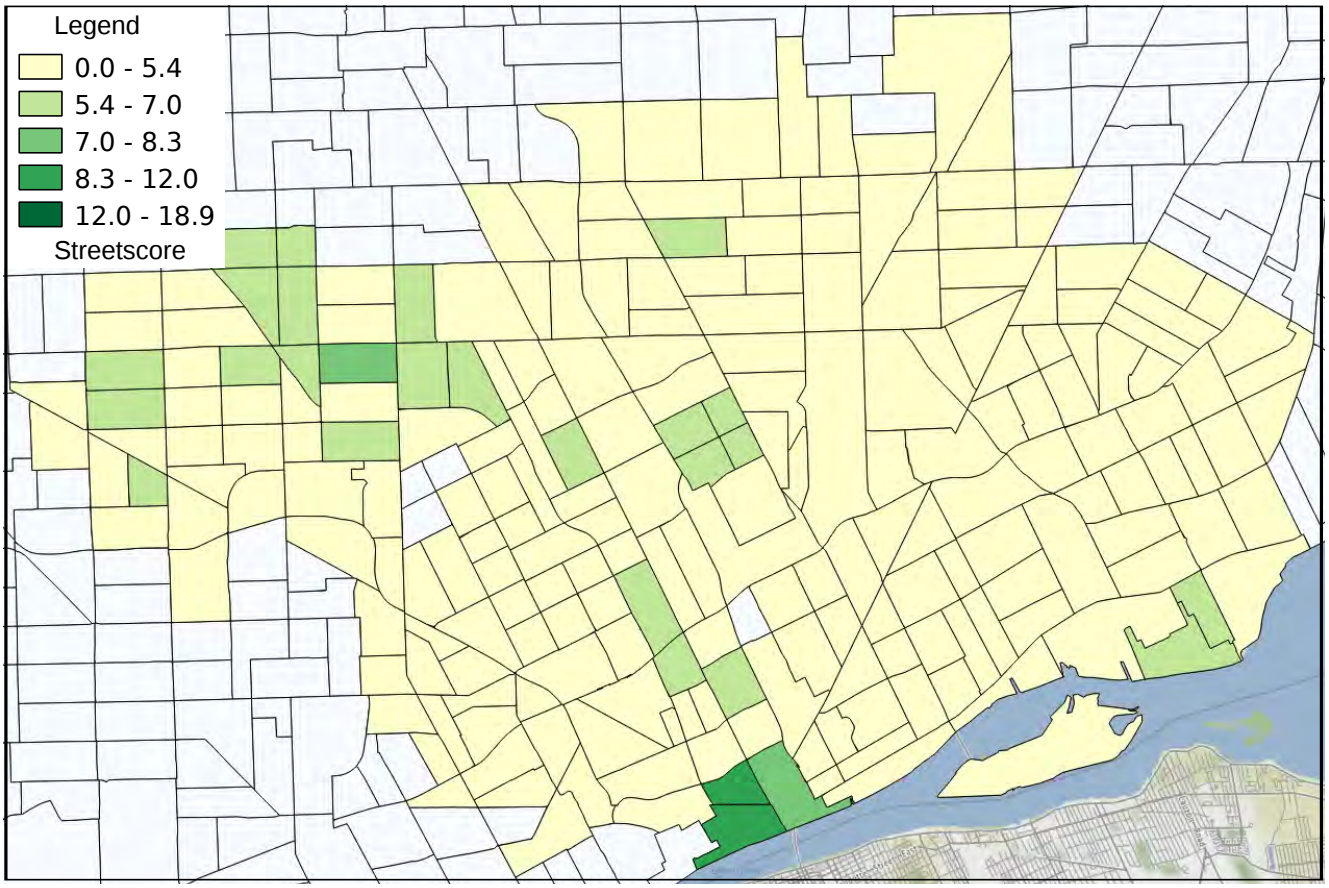

Figure S27. Detroit: Streetscore 2007.

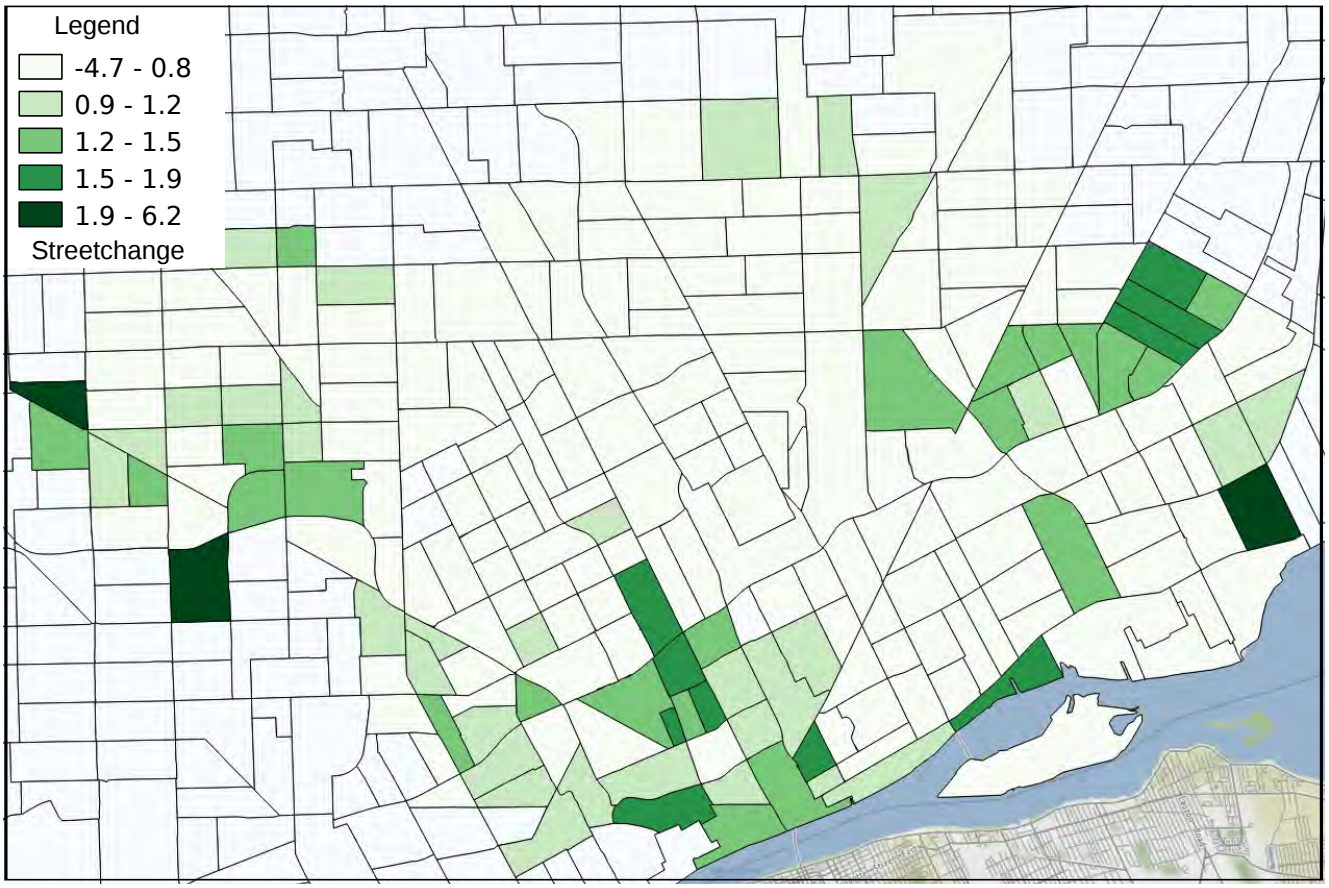

Figure S28. Detroit: Streetchange 2007–2014.