Spatial Selection and the Statistics of Neighborhoods

Luís M. A. Bettencourt
Joe Hand
José Lobo

SFI WORKING PAPER: 2015-06-020

SFI Working Papers contain accounts of scientific work of the author(s) and do not necessarily represent the views of the Santa Fe Institute. We accept papers intended for publication in peer-reviewed journals or proceedings volumes, but not papers that have already appeared in print. Except for papers by our external faculty, papers must be based on work done at SFI, inspired by an invited visit to or collaboration at SFI, or funded by an SFI grant.

©NOTICE: This working paper is included by permission of the contributing author(s) as a means to ensure timely distribution of the scholarly and technical work on a non-commercial basis. Copyright and all rights therein are maintained by the author(s). It is understood that all persons copying this information will adhere to the terms and constraints invoked by each author's copyright. These works may be reposted only with the explicit permission of the copyright holder.

www.santafe.edu

SANTA FE INSTITUTE

Spatial Selection and the Statistics of Neighborhoods

Luís M. A. Bettencourt,[1*], Joe Hand[1] and José Lobo[2]

[1] Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe NM 87501, USA,
[2] School of Sustainability, Arizona State University, 800 Cady Mall, Tempe, AZ 85281, USA.

*To whom correspondence should be addressed; E-mail: bettencourt@santafe.edu

April 7, 2015

Biological and social systems are often characterized by the emergence of general macroscopic patterns within a structure of local variations. Such variations - whether in an ecosystem or a city - express not only statistical accidents but also a rich history of innovation, selection and resulting local adaptations. For these reasons, it has remained a challenge to analyze the structure of complex systems and characterize how much information they contain at different scales of organization. Here we develop a unifying framework for studying the local heterogeneity of complex systems across scales. We show how methods from evolutionary biology and statistical learning theory can be used to quantify how much information is encoded at local levels and how complexity builds up from coarse-grained simple patterns to rich local structures. To illustrate our approach, we apply these ideas to the neighborhood structure of US cities. We observe a strong pattern of local heterogeneity in household income across over 900 cities and 200,000 neighborhoods within a simple and general statistical pattern at the metropolitan level. In this way, we identify variable strengths of local selection by income and quantify the complexity of explanation needed to account for different neighborhood structures observed across US urban areas.

Complexity has been defined as the presence of *structure with variations* (*1*). Such a description is deceptively simple, though. It glosses over the fact that local structures in a patch of forest or on a street in a large city represent not mere statistical fluctuations but a rich history of serendipity and adaptation (*2–6*). This point is well understood by biologists, ecologists and social scientists. Such variations matter: they contain information (*4*) and should not, in general, be averaged over as is typically done in statistical physics (*7*) or in models that assume representative behavior. These variations are also the source of the diversity that drives evolutionary processes (*4, 5, 8*). Thus, to attain a more complete understanding of biological and social systems we must develop methods to analyze how information is encoded and evolves across different scales of organization (*4, 5*).

There are two related approaches to this problem. We can (i) start from variable local patterns and average them to obtain a more *coarse-grained* description of the system, usually at longer spatial or temporal scales. Or, (ii) we can proceed in the opposite direction and generate detailed local patterns from more aggregated statistics. Historically, these two approaches have been developed separately, in distinct disciplines and motivated by different questions.

The coarse-graining approach is the basis for the renormalization group (RG) in statistical physics (*1, 7*). RG methods are the essential tools to analyze phase-transitions and have resulted in powerful ideas of universality, often invoked to explain complex systems (*9*). Universality states that many microscopic descriptions result in the same macroscopic dynamics, thus identifying classes of equivalent models. Proceeding in this direction leads to *information loss* as local states are replaced by averages over larger scales (*9*). Because of this essential feature, RG methods are not invertible: we cannot recover more detailed systems from their coarse-grained versions. Thus, to proceed in the opposite direction (fine-graining) we must specify more information as new scales (degrees of freedom, in the language of physics) are considered (*4*). In different but parallel ways, such methods have been developed in statistical learning theory (*10*) and in evolutionary biology (*11*).

Here, we unify these approaches and show how they lead to the quantification of the patterning of complex systems across scales, measured in units of information. We provide an empirical illustration through the analysis of household income statistics in neighborhoods of US cities. This is motivated not only by the emerging availability of high-quality data for millions of neighborhoods across the world, but also because such analysis addresses important questions of socioeconomic and urban development (*2, 3, 12, 13*) throughout history (*14, 15*). The application of these ideas to other locally heterogeneous systems, such as ecosystems (*6*) or neural networks (*16*), is straightforward but requires datasets of comparable scope and quality.

Consider the pattern of household income in New York City neighborhoods (*9*), Figure 1A and Figs. S1-S19. We observe strong heterogeneity at different spatial scales, from adjacent neighborhoods with different average household incomes to larger recognizable patches of wealth and poverty, e.g. the Upper East Side (the richest part of Manhattan) or the Bronx, Fig. 1B. This spatial heterogeneity is long-lived, persisting for decades or longer through several economic cycles and substantial demographic turnover.

Such rich and detailed pattern contrasts with the simple normal distribution for (the logarithm of) household income across the entire city, Fig. 1C. This coarse-grained statistics is common to all US metropolitan areas: the distribution of income across all cities is well described by a lognormal, see Figs. S18-S19 (*9*). Moreover, the two parameters characterizing this distribution are themselves simple and general. The mean obeys a scaling relation (*17, 18*), well parameterized by a power-law of the form $\langle y\rangle(N) = y_0 N^{\delta}$, with δ a general system-wide parameter (*18*), while the variance of the distribution is also a simple general number, see Figs. S20-S21 (*9*).

Thus, a "universal" statistical regularity emerges at the city-wide scale as the result of averaging a rich pattern of local variations. While this coarse-grained statistic is interesting (*9, 19, 20*), we now focus on the structure of the variations using the regularity as a reference. Specifically, we quantify the complexity of the pattern of variations at the neighborhood level by comparing probability distributions at different levels of aggregation. We write

$$p(y_\ell|n_j) = w_{\ell,j}\, p(y_\ell), \tag{1}$$

where $p(y_\ell, n_j)$ is the distribution of income y (in discrete bins labeled by ℓ (*9*)) in neighborhood n_j (the colorful patches in Fig 1A), and $p(y_\ell)$ is the income distribution at a more aggregate level (Fig. 1C) that we take to be the city. Eq. 1 defines the *weights* $w_{\ell,j} \equiv p(y_\ell|n_j)/p(y_\ell) \geq 0$, which transform one distribution into the other, Fig. S22. With this definition, the average weights over income obey $\langle w_j\rangle = \sum_\ell w_{\ell,j} p(y_\ell) = 1, \forall_j$.

Eq.1 is recognizable from two different perspectives. First, it is the haploid model of population genetics (*11*), also known as the "replicator" equation in evolutionary game theory (*21*). In that context, the two distributions are related across time (not space) and the weights w_ℓ are the *fitness* for allele ℓ, expressing its differential propagation into the next generation. The stronger the deviation of w_ℓ away from unity the stronger the selection for allele ℓ, corresponding to high fitness if $w_\ell \gg 1$, and vice-versa, if $w_\ell \ll 1$. This interpretation gives a formal correspondence between evolutionary biology (in time) and neighborhood structure (in space).

Second, Eq. 1 is a form of Bayes' relation, which leads to the interpretation of $w_{\ell,j}$ as probability ratios

$$p(y_\ell|n_j) = \frac{p(n_j|y_\ell)p(y_\ell)}{p(n_j)} \to w_{\ell,j} = \frac{p(n_j|y_\ell)}{p(n_j)} = \frac{p(y_\ell|n_j)}{p(y_\ell)} = \frac{p(n_j, y_\ell)}{p(y_\ell)\, p(n_j)}. \tag{2}$$

Here, $p(n_j|y_\ell)$ is the probability for a household in the city to reside in neighborhood n_j given that they have income y_ℓ while $p(n_j)$ is the probability to be in neighborhood n_j (*9*). This second perspective leads to another powerful correspondence between probability theory and statistics and associated methods of inference, and neighborhood structure in complex systems. In this context, $\ln w_{\ell,j}$ in Eq. 2 is the (non-averaged) Shannon *information* (*22*) between neighborhood j and the distribution of income y across the city. To see this more explicitly consider

$$\langle \ln w_j\rangle = \sum_\ell p(y_\ell|n_j) \ln \frac{p(y_\ell|n_j)}{p(y_\ell)} = D_{\mathrm{KL}}\left[p(y|n_j)||p(y)\right]. \tag{3}$$

This is the Kullback-Leibler divergence, D_{KL}, between the local neighborhood j and the city-wide distributions of income. D_{KL} is a fundamental quantity in information theory from which many other quantities can be derived (*22, 23*). For each neighborhood, this is the amount of information needed to describe its statistical pattern of income, given that we start by knowing the aggregate income distribution across the city. Atypical neighborhoods, with income distributions very different from the city as a whole, will require a longer explanation (more information), whereas neighborhoods that reflect the city require no further considerations and $\langle \ln w_j \rangle \to 0$. Thus, this measure expresses the strength of neighborhood effects (*3, 24–27*) relative to city-wide dynamics measured in units of information (*28*).

Fig. 2A shows the strength of neighborhood selection across New York City, measured by $\langle \ln w_j \rangle$ (see Figs. S23-S39 for examples from other cities). We observe a very mixed pattern of local selection with many neighborhoods reflecting the distribution of income for the city as a whole, but with a significant fraction of others manifesting primarily a strong local flavor. Comparing Figs. 1A and 2A reveals that the most atypical neighborhoods tend to have both the highest and the lowest average household incomes. It turns out that this is a general pattern of selection across all US cities. This effect can be assessed more directly via

$$\langle \ln w_\ell \rangle = \sum_j p(n_j|y_\ell) \ln \frac{p(n_j|y_\ell)}{p(n_j)} = D_{\mathrm{KL}} \left[p(n|y_\ell) || p(n) \right]. \tag{4}$$

This is the average information necessary to explain the distribution of specific income ranges ℓ across the city, given that we know its neighborhood structure. In the absence of neighborhood effects (i.e. of spatial selection), this quantity is zero. Thus, its magnitude quantifies the differential strength of neighborhood effects for different income levels in each city. Fig. 2B shows that the strength of neighborhood effects is highest for the highest income group, followed by the lowest. Mid-incomes are spatially the most mixed.

These effects are summarized by a single measure that captures the overall strength of neighborhood effects for each city in units of information, Fig. 2C. This is the total (mutual) information between neighborhood structure and personal income, given as the average of the previous quantities over the remaining variable,

$$\begin{aligned} \langle \ln w \rangle \quad &= \sum_j p(n_j) D_{\mathrm{KL}} \left[p(y|n_j) || p(y) \right] = \sum_\ell p(y_\ell) D_{\mathrm{KL}} \left[p(n|y_\ell) || p(n) \right] \\ &= \sum_{j,\ell} p(n_j, y_\ell) \ln w_{j,\ell} = I(n; y). \end{aligned} \tag{5}$$

If every neighborhood is a microcosm of the city as a whole, there are no (statistical) differences between neighborhoods and $I(n; y) = 0$. Conversely, in cities where every neighborhood has its own unique flavor $I(n; y)$ will be larger. How large depends on the relative amount of information needed to describe the system at the local level, Fig 1A, vs. as a whole, Fig. 1C. Thus, the mutual information $I(n; y)$ gives a measure of how well a coarse-grained pattern describes a complex system, vs. how important it is to consider all of its local features. In other words, $I(n; y)$ quantifies the average complexity of any theory of local effects.

Finally, we would like to use the strength of local selection to predict patterns of income, such as those in Fig. 1A. To do this, we evaluate the deviations in each local patch versus the system as whole. Consider the deviation in the average of some function of local characteristics, for each neighborhood $\bar{z}_i = \sum_\ell z(y_\ell)p(y_\ell|n_i)$ from that for the entire system $\bar{z} = \sum_\ell z(y_\ell)p(y_\ell)$. Using Eq. 1, we write

$$\Delta \bar{z}_j = \sum_\ell z(y_\ell) \left(w_{\ell,j} - 1 \right) p(y_\ell) = \mathrm{covar}(w_j, z) = \beta_{w_j,z} \sigma_z^2, \tag{6}$$

where the regression coefficient, $\beta_{w_j,z} = \mathrm{covar}(w_j, z)/\sigma_z^2$. This is the famous Price's equation (*8, 29, 30*) now applied to spatial selection. The simplest instance of this relations is for $z = y$, $\Delta \bar{y}_j = \mathrm{covar}(w_j, y)$, which re-derives the neighborhood patterns of Fig 1A, see (*9*) and Figs. S40-S42. Any other function of neighborhood characteristics can be computed using the spatial Price equation.

Selection is a general process by which individuals learn and adapt to their environment by acquiring information (*4, 5, 11*), generating local heterogeneities within a context of broader statistical regularities (*1*). We have shown how these *deviations* occur within general coarse-grained *structures* and how to account for information embedded in the structure of complex systems at multiple scales. In this way, we can satisfy at once the contention of two frequently opposing approaches to the study of biological and social systems and express quantitatively how coarse-grained universality can co-exist with local exceptionality.

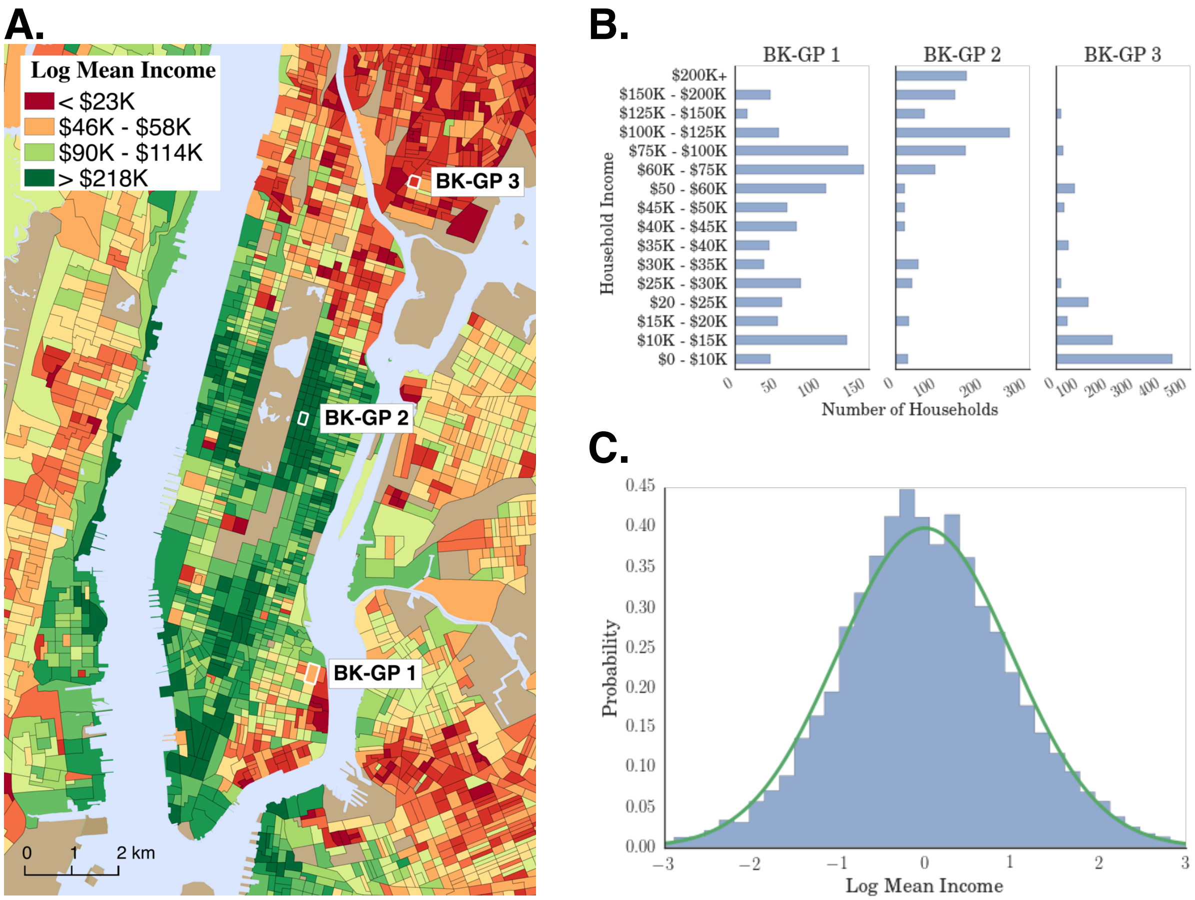

Figure 1: **The heterogeneity of neighborhoods in New York City**. See Figs. S1-S19 for other metropolitan areas. **A**: Average household income in New York City census block groups. **B**: Income distributions in selected neighborhood shown in Fig. 1A. **C**: The city-wide distribution of household income is well described by a lognormal distribution (green line), which we show in SOM is a good general model for the household income distribution in all U.S. MSAs. The mean and variance of this distribution obey simple scaling relations (*9*). Data was compiled by the US Census 5-year American Community Survey (2010 release) comprising of over 200,000 block groups nationwide and about 14,000 in the New York Metropolitan Statistical Area (MSA). See (*9*) for Materials and Methods and a discussion of neighborhood units.

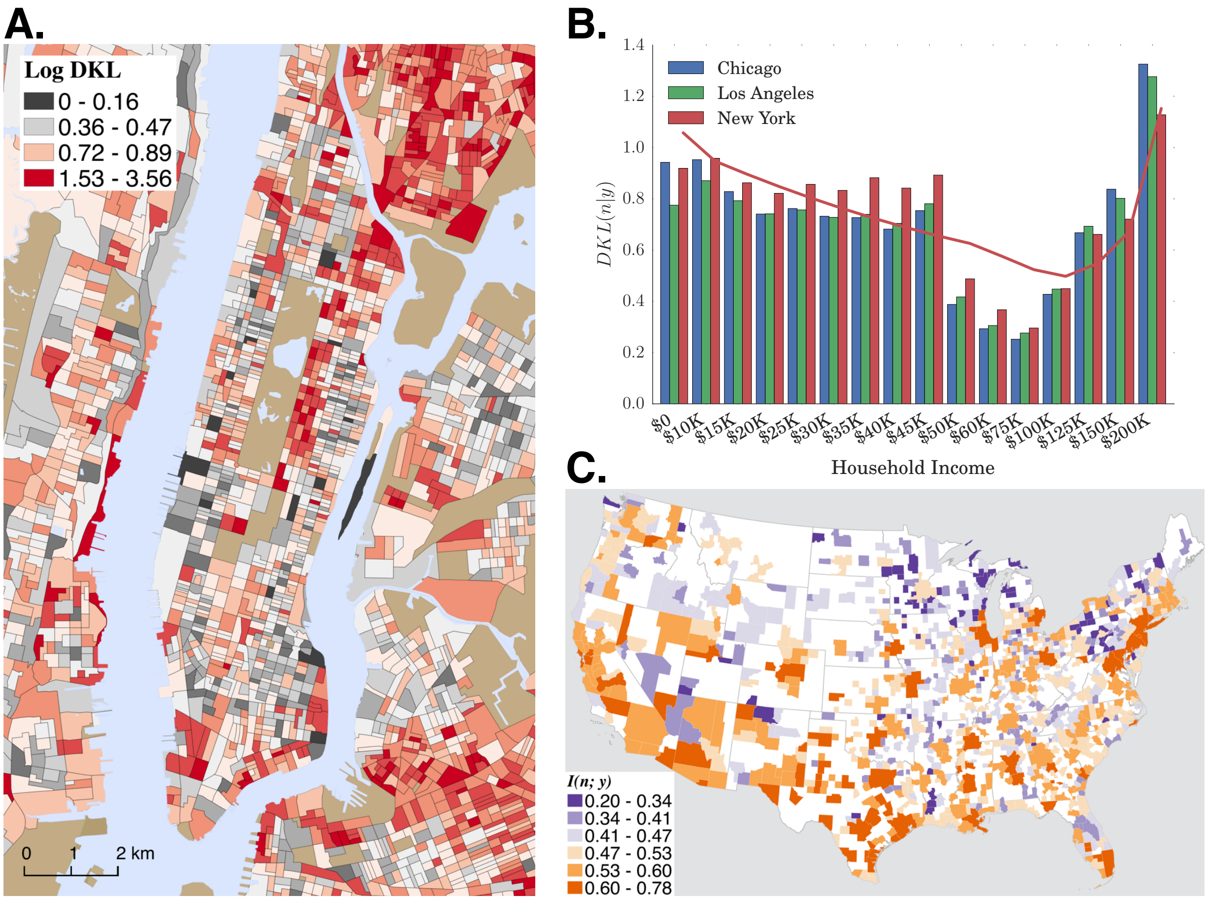

Figure 2: **Spatial Selection Measured in Units of Information**. **A**: The information, $\langle \ln_2 w_j \rangle$, necessary to explain the household income distribution of different neighborhood, given the city-wide income statistics, Fig. 1C. (See Figs. S23-S39 for other metropolitan areas). The entropy of the income distribution for the New York metropolitan area is $H(y) = 3.89$. Thus neighborhoods in darker grey require very little additional information, while those in red may demand as much explanation as the city-wide pattern of income itself. Comparing to Fig. 1A, note that it is both the poorest and richest neighborhoods that tend to require more information (red). **B**: The average intensity of local selection by income, measured by $\langle \ln_2 w_\ell \rangle$, for several large US MSAs. The richest income brackets experience the strongest spatial selection, followed by the lowest incomes. Incomes near the middle of the distribution are more spatially mixed and may provide some social connectivity between poorer and richer households. The variation of $\langle \ln_2 w_\ell \rangle$ with income can be reasonably fitted by a quadratic form (red line, for New York MSA) with $\langle \ln_2 w_\ell \rangle(y) = 6.074 \times 10^{-11} y^2 - 1.167 \times 10^{-5} y + 1.058$. **C**: The average strength of neighborhood effects across urban areas in the US measured by the mutual information $I(y; n)$ (942 units). Orange denotes stronger average neighborhood selection, where the distribution of income in each of the city's neighborhoods is less like that of the city as a whole, and vice versa (purple). Cities with low $I(y; n) \to 0$ have less distinguishable neighborhood structure by income.

References and Notes

1. N. Goldenfeld, L. P. Kadanoff, Simple Lessons from Complexity. *Science* **284**, 87-89 (1999).

2. W. J. Wilson, *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy*, (Univ. of Chicago Press, Chicago, 1987).

3. R. J. Sampson, *Great American City* (Chicago University Press, Chicago, 2012).

4. J. Maynard Smith, E. Szathmáry *The Major Transitions in Evolution*. (Oxford University Press, Oxford, England, 1995).

5. A. Wagner, W. Rosen, Spaces of the possible: universal Darwinism and the wall between technological and biological evolution. *J. Roy. Soc. Interface* **11**, 1-12 (2014).

6. S. T. A. Pickett, P. S. White, *The Ecology of Natural Disturbance and Patch Dynamics* (Academic Press, New York, 1985).

7. J. Zinn-Justin, *Quantum field theory and critical phenomena* (Clarendon Press, Oxford, 2002).

8. A. Gardner, The Price equation. *Curr. Biol.* **18**, R198-R202 (2008).

9. See Supplementary Online Materials for more information and discussion.

10. V. N. Vapnik, *Statistical Learning Theory* (Wiley-Interscience, Hoboken, NJ, 1998).

11. S. A. Frank, Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory. *J. Evol. Biol.* **25**, 2377-2396 (2012).

12. Y. M. Ioannides, L. D. Loury, Job information networks, neighborhood effects, and inequality. *J. Econ Lit.* **42**, 1056-1093 (2004).

13. J. T. Rothwell, D. S. Massey, Geographic effects on intergenerational income mobility. *Econ. Geo.* **91**, 83-106 (2015).

14. A. Kearns, M. Parkinson, The significance of neighborhood. *Urban Studies* **38**, 2103-2110 (2001).

15. M. E. Smith, The archaeological study of neighborhoods and districts in ancient cities. *J. Anthropol. Archaeol.* **29**, 137-154 (2010).

16. G. Buzsáki, K. Mizuseki, The log-dynamic brain: how skewed distributions affect network operations. *Nature Reviews Neuroscience* **15**, 264-278 (2014).

17. L. M. A. Bettencourt *et al.*, Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7301 (2007).

18. L. M. A. Bettencourt, The Origin of Scaling in Cities. *Science* **340**, 1438-1441 (2013).

19. A. J. Harrison, Earnings by size: a tale of two distributions. *Rev. Econ. Studies* **48,** 621-631 (1981).

20. F. A. Cowell, *Measuring Inequality*, (Prentice Hall, London, 1995).

21. K. M. Page, M. A. Nowak, Unifying Evolutionary Dynamics. *J. theor. Biol.* **219,** 93-98 (2002).

22. T. M. Cover, J. A. Thomas, *Elements of Information Theory* (John Wiley & Sons, New York, 1991).

23. S. Kullback, *Information Theory and Statistics* (John Wiley and Sons, NY, 1959).

24. R. J. Sampson, S. W. Raudenbush, F. Earls, Neighborhoods and violent crime: a multilevel study of collective efficacy. *Science* **277**, 918-924 (1997).

25. J. Ludwig *et al.*, Neighborhood effects on the long-term well-being of low-income adults. *Science* **337**, 1505 (2012).

26. P. Sharkey, J. W. Faber, Where, When, Why, and For Whom Do Residential Contexts Matter? Moving Away from the Dichotomous Understanding of Neighborhood Effects. Ann. Rev. Soc. **40**, 559 (2014).

27. Y. M. Ioannides, G. Topa, Neighborhood effects: accomplishments and looking beyond them. *J. Reg. Sci.* **50**, 343-362 (2010).

28. S. F. Reardon, K. Bischoff, Income Inequality and Income Segregation. *Am. J. Soc.* **116**, 1092-1153 (2011).

29. G. R. Price, Selection and Covariance. *Nature* **227**, 520-521 (1970).

30. S.A Frank, Natural Selection IV. The Price equation. *J. Evol. Biol.* **25**, 1002-1019 (2012).

31. We thank Clio Andris, Elisabeth Bruch, Rob Sampson and Mario Small for discussions. This research is partially supported by the John D. and Catherine T. MacArthur Foundation (grant 13-105749-000-USP) and by the Arizona State University/Santa Fe Institute Center for Biosocial Complex Systems.

Spatial Selection and the Statistics of Neighborhoods

Supplementary Online Material

Luís M. A. Bettencourt,[1*], Joe Hand[1] and José Lobo[2]

[1] Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe NM 87501, USA,
[2] School of Sustainability, Arizona State University, 800 Cady Mall, Tempe, AZ 85281, USA.
*To whom correspondence should be addressed; E-mail: bettencourt@santafe.edu

April 7, 2015

Contents

List of Figures

List of Tables

1 Materials and Methods

1.1 Data Sources

Geo-referenced data at the household level (such as income and population) for the United States is reported at the *Census Block Group* level through the 2006-2010 5-year American Community Survey (*31*). Block groups are statistical subdivisions of Census Tracts, which in turn are the basic data collection units for the population census. The boundaries of block groups are generally set so that they contain between 600 and 3,000 people, with a typical size of about 1,500 (or 500 households). Block Groups are spatially contiguous and tile the entire country. Data are aggregated into urban areas defined as Core-Based Statistical Areas (CBSA), which include Micropolitan Statistical Areas and Metropolitan Statistical Areas. Micropolitan and metropolitan areas consist of a core county, or set of counties, with an urban area having a population of at least 10,000 people plus adjacent counties having a high degree of social and economic integration with the core counties as measured through commuting ties. Counties are the primary legal divisions of States in the U.S., many of which are functioning governmental units whose powers and functions vary from state to state. Counties differ greatly in their areal expansion and populations size. For simplicity we refer to micropolitan and metropolitan areas together as *Metropolitan Areas*: there are 942 such areas currently in the USA.

1.2 Units of Analysis: Neighborhood Definitions

In the main text we used the colloquial term *neighborhood* to refer to Census Block Groups, as shown in Figure 1A and S1 - S17. Block groups provide an exhaustive tiling of the entire national territory of the United States and its population. In denser areas Block Groups correspond to smaller land areas, as can be clearly seen in the maps of Fig. 1A and S1 - S17. Adopting block groups as proxies for neighborhoods is convenient because they are consistently defined by the US Census Bureau (and similarly by other national census around the world) and provide a universal standard for the study of small area statistics across an entire nation. For these reasons, they are the most common proxies for social units at this scale (neighborhoods). However, sociologists with a rich knowledge of social dynamics at the local level have debated the advantages and disadvantages of several neighborhood definitions and have in different detailed studies adopted different units of analysis, see for example (*3*) and Hipp (*32*) for discussions.

Our aim here is to demonstrate effects of spatial selection at any given scale. A systematic study of the strength of spatial selection at different scales (neighborhood definitions) is beyond the scope of the present manuscript and will be presented elsewhere.

1.3 Data limitations

The American Community Survey (ACS) and the Decennial Census collect household data in small spatial units that allow us to characterize patterns of spatial selection in neighborhoods. The ACS is a statistical survey conducted by the US Census Bureau, sent to approximately 250,000 addresses monthly (or about 3 million per year). Unlike the population census (which is strictly a population count), the ACS collects socioeconomic information (for example, on household income). The data are collected primarily by mail, with follow-ups by telephone and personal visits. ACS data are used to make yearly estimates for counties which are then aggregated to provide estimates for States and metropolitan areas.[1] ACS data has an important reporting limitation when it comes to the upper tail of the income distribution: the number of households is listed only for a data bin set by a minimum value ($>$ \$200k per household in 2010). We estimate the average income in these upper bins using the constraint provided by the Price Equation, see below.

It has been often shown empirically (*33*) that, at higher levels of spatial aggregation, the upper tail income distribution deviates from the lognormal pattern reported in Fig. 1C and

[1]For detailed information on the American Community Survey go to `www.census.gov/acs/`.

detailed below. Such statistics do, in fact, often follow a Pareto (power-law) distribution for the top richest fraction of 1% (*33*). Consideration of a finer distribution in this regime is likely to produce even higher atypical values of information for neighborhoods that concentrate such high incomes. In this sense, even though many richer neighborhoods appear the most atypical from the point of view of their income distribution relative to the city at large, Fig. 2B, it is likely that this effect is underestimated as a result of the way data for these incomes are reported.

1.4 Practical Estimation of Probabilities

Here, we provide an explicit version of the probability distributions introduced in the main text and the procedure by which they are estimated from discretely binned data.

Let N be the the total number of households in a given city, or the size of that city, for short. Let N_j be the number of households in neighborhood j, across all income levels. Then $n_{j,\ell}$ is the number of households in neighborhood j, with income (in the interval denoted by) ℓ. N_ℓ is, correspondingly, the total number of households in the city with income in the interval indexed by ℓ. These quantities obey several simple sum rules:

$$\sum_j N_j = N, \qquad \sum_\ell N_\ell = N, \tag{S1}$$

$$\sum_j n_{j,\ell} = N_\ell, \qquad \sum_\ell n_{j,\ell} = N_j. \tag{S2}$$

Having defined these quantities, which are the ones typically reported by the U.S. Census Bureau, we can provide simple frequency estimators for the several probability densities introduced in the main manuscript. The simplest is $p(n_j)$, the probability of living in a specific neighborhood, which is

$$p(n_j) = \frac{N_j}{N}. \tag{S3}$$

Analogously the probability of belonging to a given income level, ℓ, is

$$p(y_\ell) = \frac{N_\ell}{N}. \tag{S4}$$

The conditional distribution for being in a given neighborhood j given income ℓ is

$$p(n_j|y_\ell) = \frac{n_{j,\ell}}{N_j}. \tag{S5}$$

From this and Bayes' relation it follows that

$$p(y_\ell|n_j) = \frac{p(n_j|y_\ell)}{p(n_j)} p(y_\ell) = \frac{n_{j,\ell}}{N_j}. \tag{S6}$$

The weights $w_{j,\ell}$ are given by

$$w_{j,\ell} = N \frac{n_{j,\ell}}{N_\ell N_j}. \tag{S7}$$

Finally, we can check that the properties of the conditional probabilities hold, under these definitions,

$$\sum_j p(n_j|y_\ell) = \sum_j \frac{n_{j,\ell}}{N_\ell} = \frac{1}{N_\ell} N_\ell = 1. \tag{S8}$$

$$\sum_\ell p(n_j|y_\ell) N_\ell = \sum_\ell \frac{n_{j,\ell}}{N_\ell} N_\ell = \sum_\ell n_{j,\ell} = N_j. \tag{S9}$$

$$\sum_\ell p(y_\ell|n_j) = \sum_\ell \frac{n_{j,\ell}}{N_j} = \frac{1}{N_j} N_j = 1. \tag{S10}$$

$$\sum_j p(y_\ell|n_j) N_j = \sum_j \frac{n_{j,\ell}}{N_j} N_j = \sum_j n_{j,\ell} = N_\ell. \tag{S11}$$

2 Supplementary Text

2.1 The Statistics of Urban Income and Urban Scaling Relations

In Figure 1C, we showed that the frequency distribution of average household income in New York City (MSA) is visually well described by a lognormal distribution (green line). Here we demonstrate that this is a general property of all US metropolitan areas and show how the two parameters of the distribution (the mean logarithmic income and its logarithmic variance) express scaling relations with city size.

Figures S18 and S19 present the results of comparing the goodness of fit of the lognormal distribution to that of other alternative distributions using the Bayesian Information Criterion (*34*) for each city. In the vast majority of cities (83%) the lognormal is the best distribution. Many other plausible distributions manifestly fail to even occasionally fit the data. In a small number of cases, we find reasonable fits to the data using an exponential Weibull distribution, but there has not been much work providing a theoretical justification for such a distribution in other studies of income distributions (for a notable exception see reference (*35*)). The lognormal, on the other hand, is well known to fit well the body of distributions of income (*33, 36*) and is generally explained in terms of models of multiplicative random growth. The extreme 1% wealthiest part of the frequency distribution has been known to deviate from the lognormal pattern at the national level but, as discussed above, this regime of urban wealth is not well represented in the ACS survey data.

The lognormal is characterized by two parameters, the mean of the log-household income for each city and its variance. Figure S20 shows the correlation between the log-mean income vs household size for each city (MSA). This relationship is a well known urban scaling relation (*17,18*), $y(N) = y_0 N^{\delta}$, characterizing many urban systems around the world, which share the same scaling approximate exponent $\delta > 0$. Figure S21 shows the scaling plot for the log-variance. We see that the existence of such a scaling relation is less clear, in the sense that the relationship is noisier, and may be consistent with no variation of this parameter with city size, as has been e.g. observed for violent crime in Ref. (*37*).

A fuller exposition and analysis of these results will be presented elsewhere. Nevertheless, we would like to emphasize that the present results, in conjunction to other recent research involving crime (*37*), the degree of cell phone urban social networks (*38*) and mobility (*39*) point to a general form of the statistics of urban indicators (a sort of statistical universality), that may also hold not only for contemporary cities, but throughout history (*40, 41*).

2.2 Spatial Selection, Neighborhood Effects and Income Polarization

In this section, we briefly discuss how our approach and results relate to relevant work in sociology and economics on neighborhood effects and the spatial characteristics of household income distributions. A fuller treatment of these issues and of the spatial selection approach developed here to analyzing the empirics of neighborhood income inequality will be pursued elsewhere.

Differences between neighborhoods are perhaps the clearest manifestation of the spatial heterogeneity of urban areas, that is, the uneven and complex distribution of individuals and households within cities (*42*). The question of how the composition of a population affects the sorting of individuals by place of residence, what sociologist term ”residential selection”, has been a long-standing question for sociology. In its earliest terms, somewhat simplistic by today’s standards, Park and Burgess (*43*) proposed a explanation for spatial urban patterning in direct analogy to darwinian selection, an approach known as urban ecology. Thinking in sociology has come a long way since then, but echoes of these first attempts to conceptualize the issue remain even as a new literature on neighborhood effects has emerged with a strong empirical base, especially in Chicago (*3*), (*44–46*).

The importance of neighborhood selection has been emphasized in this literature because of its consequences or ”contextual effects”. This refers to the way in which individuals’ social, economic and health outcomes are affected by the physical and socioeconomic characteristics of their residential communities (*47–50*). Income differentials are a major determinant of spatial residential selection and the associated ”neighborhood effects” as higher-income individuals tend to want to live next to other higher-income individuals while low-income individuals may have fewer choices, increasingly tied to residing next to other poor households (*51*), see also Fig. 2B. This residential selection is often associated with changes in real estate market valuations and tax revenue bases which together have been proposed as a means to sustain cycles of increasing neighborhood polarization (*49*). While many of these patterns and their temporal change is being reveled by new data and detailed studies at the neighborhood level, much remains to be done toward a general understanding of the social and economic causes and consequences of spatial selection in cities.

As the evidence and concern mounts for growing income inequality at the national level (*53–55*), so it has for the growing income segregation in US urban areas (*52,56–60*). Residential selection on the basis of income is related to income inequality but also to the ability and willingness of individuals to act on preferences regarding who they reside next to. However, measuring income segregation in urban areas is not a straightforward matter. The workhorse metric for income inequality, the Gini Index, suffers from several deficiencies when measured at a spatially disaggregated level, such as neighborhoods. For one thing, the Gini is sensitive to the number of income categories used when constructing the measure. The typical manner in which the index is constructed assumes that the spatial units of observation are similar in population size (but U.S. census tracts or block groups differ in their population size). But most importantly, the Gini Index cannot distinguish between the effects of an overall increase in income inequality and increasing income differentiation inside neighborhoods (*49, 61*). As an alternative approach, a variety of studies have turned to entropy-based measures as these are able to capture how individuals or households are distributed across various income groups within neighborhoods (*49, 62–65*). But while purely justified on statistical grounds, the use of entropy measures to capture income inequality across and within neighborhoods is not typically grounded on a firm theoretical framework.

In this light, we emphasize that the measures introduced here are not new ad-hoc socio-economic indices but follow inevitably from treating neighborhood heterogeneity as an instance of spatial selection defined as the relationship between income distributions at two different spatial levels of analysis. Nevertheless, we note that our informational measures of spatial selection are close relatives of the Rank-Order Information Theory Index (*49*), which compares the variation in family incomes within neighborhoods (census tracts) to the variation in family incomes in the metropolitan area in which the tracts are embedded. Although formally and quantitatively different, our results agree qualitatively with those of (*49*), in that we also find increasing income segregation between neighborhoods in US metropolitan areas over the last twenty years. This phenomenon is often referred to as *neighborhood polarization*, and is very visible e.g. in Detroit, Figs. S6 - S28, St. Louis, S17 - S39 or even Austin, S2 - S24, where poor and rich section if the city are clearly physically separated almost as a dipole. In other cities the overall spatial pattern of rich and poor neighborhoods is often more mixed spatially.

The consequences of any selection process on the distribution of a characteristic of interest in a population can be separated, using Price's equation (*66*), into two components: the direct effects of selection and those of transmission over time. Here, we focused on the purely spatial aspect of the problem and so have not formally analyzed temporal transmission, which can clearly be included in the context of the Price equation (see Section 2.4). The derivation of the Price equation, Eq. 6 in the main text, not only provides a statistical description of the process leading to spatial heterogeneity within cities, it can also be used to predict actual neighborhood income patterns. This mathematical account of selection allows us to express neighborhood heterogeneity in terms of the mathematics of evolution and information, thereby connecting the diversity of patterns in urban neighborhoods to the study of of how structure, complexity and diversity arise in other complex systems (*67, 68*).

2.3 Fine-graining, Information and Learning

In the main text we introduced, Eq. 1, the relation

$$p(y_\ell|n_j) = w_{\ell,j}p(y_\ell),$$

for the distribution of some individual trait y (such as income) at two levels of (spatial) aggregation. Here, we show more explicitly why the weights, $w_{\ell,j}$, should be interpreted in terms of information and how their specification is a process of information gain, i.e., of learning.

In the main text, we used the interpretation of Eq. 1 as Bayes' relation to write the weights as

$$w_{\ell,j} = \frac{p(y_\ell, n_j)}{p(y_\ell)\, p(n_j)} = \frac{p(y_\ell|n_j)}{p(y_\ell)}. \tag{S12}$$

Taking the logarithm, we obtain

$$\ln p(y_\ell|n_j) = \ln w_{\ell,j} + \ln p(y_\ell) = \ln \frac{p(y_\ell, n_j)}{p(y_\ell)\, p(n_j)} + \ln p(y_\ell), \tag{S13}$$

where we identify the $\ln w_{\ell,j}$ term as the specific mutual information (before averaging) between the states y_ℓ and n_j. Moreover, note that the specific Shannon entropies are $h(y_\ell|n_j) = -\ln p(y_\ell|n_j)$ and $h(y_\ell) = -\ln p(y_\ell)$ (*22*). We can then write

$$h(y_\ell) = h(y_\ell|n_j) + i(y_\ell|n_j), \tag{S14}$$

which states that the (higher) entropy of the city wide income distribution is equal to the lower entropy of the same distribution in each neighborhood plus the mutual information that such neighborhood has on the city wide distribution. This statement is usually presented in averaged form (where all three quantities are provably positive (*22*)), by tracing under the joint $p(y_\ell, n_j)$ as,

$$H(y) = H(y|n) + I(y;n). \tag{S15}$$

Here, $\langle \ln w \rangle = I(y;n)$ and is given by

$$\langle \ln w \rangle = I(y;n) = \sum_{\ell,j} p(y_\ell, n_j) \ln \frac{p(y_\ell, n_j)}{p(y_\ell)\, p(n_j)}. \tag{S16}$$

Thus, the operation of disaggregating the structure of the system as a whole to smaller spatial units requires in general the *addition of information* (or "structure") to that present in the averaged distribution across the city. What this means is that there is, in general, higher complexity of system spatial configurations at the more disaggregated level. In turn, the advent of this local complexity is associated with the breaking of spatial symmetries of the system (*69, 70*). As a consequence, we conclude that the process by which (spatial) complexity arises is driven by selection associated with successive level of symmetry breaking. Intuitively, this is why local models of neighborhood structure, typical of social scientific approaches, must contain more information than coarse-grained models, based on statistical physics approaches.

This raises an interesting question of how to do the opposite, namely how to obtain the aggregated distribution from that of the smaller spatial pieces. This operation is known in statistical physics as ”coarse-graining” (*71*) and is at the basis of some of the most important results for the behavior of systems undergoing critical phenomena, via the application of renormalization group techniques (*71*). These methods perform successive levels of spatial (and sometimes temporal) averaging to obtain the large-scale (averaged) behavior of a physical system. For most systems, this procedure either leads to the uninteresting outcomes of an increasingly uniform or an increasingly noisy system (it is said that the system flows towards zero or infinite temperature, respectively, under coarse-graining). But at phase transitions - critical phenomena when the global properties of the system change coherently, such as a liquid-vapor transition - the operations of coarsening lead to systems that are spatially self-similar, regardless of a number of details of the microscopic physics (irrelevant operators) (*71*). In our case, cities obtained as averages over neighborhoods, emerge as a kind of self-similar structure out of this kind of procedure, Fig. 1C, as they are characterized by the same simple statistics although with parameters that themselves depend on city size (scaling) (*17,18*).

To see what is entailed by coarse-graining in terms of the framework developed in this section we simply write the inverse of Eq. 1 as

$$p(y_\ell) = \bar{w}_{\ell,j} p(y_\ell | n_j), \tag{S17}$$

and by taking logarithms and comparing to Eq. 1 we readily identify

$$\bar{w}_{\ell,j} = -\ln \frac{p(y_\ell, n_j)}{p(y_\ell)\, p(n_j)} = \ln \frac{p(y_\ell)\, p(n_j)}{p(y_\ell, n_j)} = -i(y_\ell, n_j). \tag{S18}$$

Thus, we write

$$h(y_\ell|n_j) = h(y_\ell) - i(y_\ell|n_j) \to H(y|n) = H(y) - I(y;n), \tag{S19}$$

where the last relation is obtained under averaging, as above, under the joint distribution. As might have been expected, we see that the operation of coarse-graining entails the *removal of information* present at the neighborhood level to obtain a spatially averaged distribution. This corresponds to the common intuition that averaging can mask important or revealing detail. How much information is ”thrown away” in this process is quantified on average by the mutual information between units of analysis at different levels of aggregation and the variable(s) of interest. Thus, the mutual information $I(y;n)$ is a city-wide average measure of the strength of neighborhood effects. It should be clear that such transformation maps potentially very complex patterns, such as those of Fig. 1A, to relatively simple ones, such as those of Fig. 1C. The formal treatment of this operation and its more common uses in statistical physics will be presented elsewhere. It should nevertheless be clear that such coarse-graining operations typically lead to simpler aggregate statistics and can, under certain specific conditions, result in Zipfian scale-free phenomena in ways that generalize approaches to criticality in physical systems (*37*).

2.4 Spatial Selection and the Price Equation

In this section, we clarify the use of the Price equation in spatial selection and contrast it with its most common use in evolutionary dynamics. As George Price himself emphasized referring to his eponymous equation: ”The mathematics given here applies not only to genetical selection but to selection in general” (*72*): spatial selection, therefore, is no exception.

In the canonical formulation of the Price equation, the evolution of the average value of any population characteristic z (household income in our case) is the result of the variation introduced by two distinct terms. Following Frank (*73*) we write

$$\Delta \bar{z} = \Delta_{\rm s} \bar{z} + \Delta_{\rm t} \bar{z} = \sum_i (w_i - 1) z_i p(z_i) + \sum_i \Delta z_i p'(z_i), \tag{S20}$$

respectively. The first term, $\Delta_{\rm s}\bar{z}$, is associated with selection whereas the second, $\Delta_{\rm t}\bar{z}$, encodes all other processes that may change the value of z during the process. In practice, in processes of evolutionary biology (*73*) this second term is often invoked to account for errors in the *transmission* of genetic information between generations (e.g. copying errors due to mutations). In our process of spatial selection we have assumed that such a term is zero, $\Delta_{\rm t}\bar{z} = 0$, as there is no change in income during the process of neighborhood choice.

The selection term is often written in a number of equivalent ways:

$$\Delta_s \bar{z} = \sum_i (w_i - 1) z_i p(z_i) = \text{covar}(w, z) = \beta_{wz} \sigma_w^2 = \beta_{zw} \sigma_z^2. \tag{S21}$$

The third term results from the first via the standard definition of covariance (*74*). The last two terms express the covariance between w and z in terms of the product of the variance of each variable (σ_w^2 and σ_z^2) and that of the *regression* (*74*) of the variable w on z, β_{wz} and vice-versa. It is worth noting that, for a lognormal distribution, the Gini-coefficient provides a measure of income inequality, which is only a function of the $\sigma^2_{\ln y}$. Thus, neighborhood polarization (deviations of average neighborhood household income from the city-wide mean) will be correlated to larger city-wide inequality, through the correlation of $\ln y$ on the selection strengths w, Fig. 2B.

We have also used the Price equation to predict the pattern of neighborhood wealth shown in Figure 1A. This allows us to derive an estimate of the mean wealth in the richest bin (a quantity censured in the data) and the distribution of incomes in each of the coarse income bins provided by the ACS, see Figures S40 - S42.

The Census reports household income figures at the block group level in two different ways: i) the aggregate household income (and total households to calculate the mean income) and ii) the number of households in sixteen discrete income bins, Fig. 1B. All households with an income greater than \$200,000 are grouped into the last bin. To estimate the mean income value of the richest bin in each neighborhood, we determine the amount of income missing for each neighborhood, ϵ_j, by comparing the reported mean income to the value calculated using the minimum value of each income bin:

$$\bar{y}_j = \frac{\sum_\ell \bar{y}_{\ell,j} n_{\ell,j}}{N_j} = \frac{\sum_\ell y_{\min,\ell} n_{\ell,j}}{N_j} + \epsilon_j \quad \text{(S22)}$$

where $y_{\min,\ell}$ is the minimum value of income in the bin ℓ (e.g. \$0 for the bin \$0 - \$10K). In this way we can estimate an upper bound on the mean income value of the richest bin in each neighborhood to make up for the missing income, as seen in Figures S40 and S41. We observe that there were some neighborhoods where any value above the lower bound resulted in too much total income.

For neighborhoods without households reported in the richest bin, we adjust the mean income calculation using the other bins. We explored the difference between the detailed sum of every actual income and the calculation using the given bins, where we sum the number of households in each bin times the corresponding average income. Because the latter is not given we are free to estimated it to derive the given total. In practice, we have parameterized the total income in each bin in terms of a parameter a defined as

$$a_j = \frac{\bar{y}_j - \bar{y}_{\min,j}}{\bar{y}_{\mathrm{m},j} - \bar{y}_{\min,j}} \quad \text{(S23)}$$

where $\bar{y}_j$ is the reported mean average income for each neighborhood, $\bar{y}_{\min,j}$ is its value computed using the lower bound in each income bin and $\bar{y}_{\mathrm{m},j}$ is the same value computed using the midpoint of income in each bin. Because the distribution is strongly skewed no specific point within the income bin (smallest or mid-point), is likely to give the actual correct result, which then can be estimated through a. Figures S40 - S42, show this process for New York City.

2.5 Average Household Income & Information Maps for US Urban Areas

In the main text we illustrated the diversity of income across American urban areas using a map of New York City, because we thought that this would be the best known case to most readers. Figures S1 - S17 and Figures S23 - S39 show similar maps (average household income by block group and the D_{KL} for each neighborhood) for other large US metropolitan areas, including a larger map of New York City.

3 Supplementary Tables

Table S1: Top 10 US Metropolitan Areas by $I(y;n)$

City	Mutual Information	Total Population
Dallas, TX	0.697	6,154,265
New York City, NY	0.689	18,700,715
New Orleans, LA	0.685	1,105,020
Reno, NV	0.681	416,860
College Station, TX	0.680	219,058
Morgantown, WV	0.677	125,691
Memphis, TN	0.671	1,301,248
Midland, TX	0.667	132,103
Fresno, CA	0.666	908,830
San Antonio, TX	0.665	2,057,782

Table S2: Lowest 10 US Metropolitan Areas by $I(y;n)$

City	Mutual Information	Total Population
Mount Vernon, WA	0.378	115,231
Hinesville, GA	0.372	76,996
Palm Coast, FL	0.362	91,806
Wausau, WI	0.357	132,644
Glens Falls, NY	0.328	128,795
Dover, DE	0.327	156,918
Coeur d'Alene, ID	0.323	134,851
Mankato, MN	0.319	94,990
Sheboygan, WI	0.315	115,328
St. George, UT	0.310	134,033

Table S3: Top 10 US Micropolitan Areas by $I(y;n)$

City	Mutual Information	Total Population
Lamesa, TX	0.774	13,853
Beeville, TX	0.763	31,896
Bay City, TX	0.723	36,647
Hobbs, NM	0.710	62,503
Edwards, CO	0.690	57,832
Wauchula, FL	0.680	27,521
Greenville, MS	0.651	52,455
Arcadia, FL	0.649	34,557
Clewiston, FL	0.648	39,030
Clovis, NM	0.645	46,924

Table S4: Lowest 10 US Micropolitan Areas by $I(y;n)$

City	Mutual Information	Total Population
Sayre, PA	0.260	62,415
Huntingdon, PA	0.252	45,830
Cadillac, MI	0.250	47,615
Bradford, PA	0.245	43,853
DeRidder, LA	0.241	35,000
Platteville, WI	0.235	50,716
Menomonie, WI	0.230	43,365
Miami, OK	0.229	32,193
Natchitoches, LA	0.222	39,274
Baraboo, WI	0.206	60,957

4 Supplementary Figures

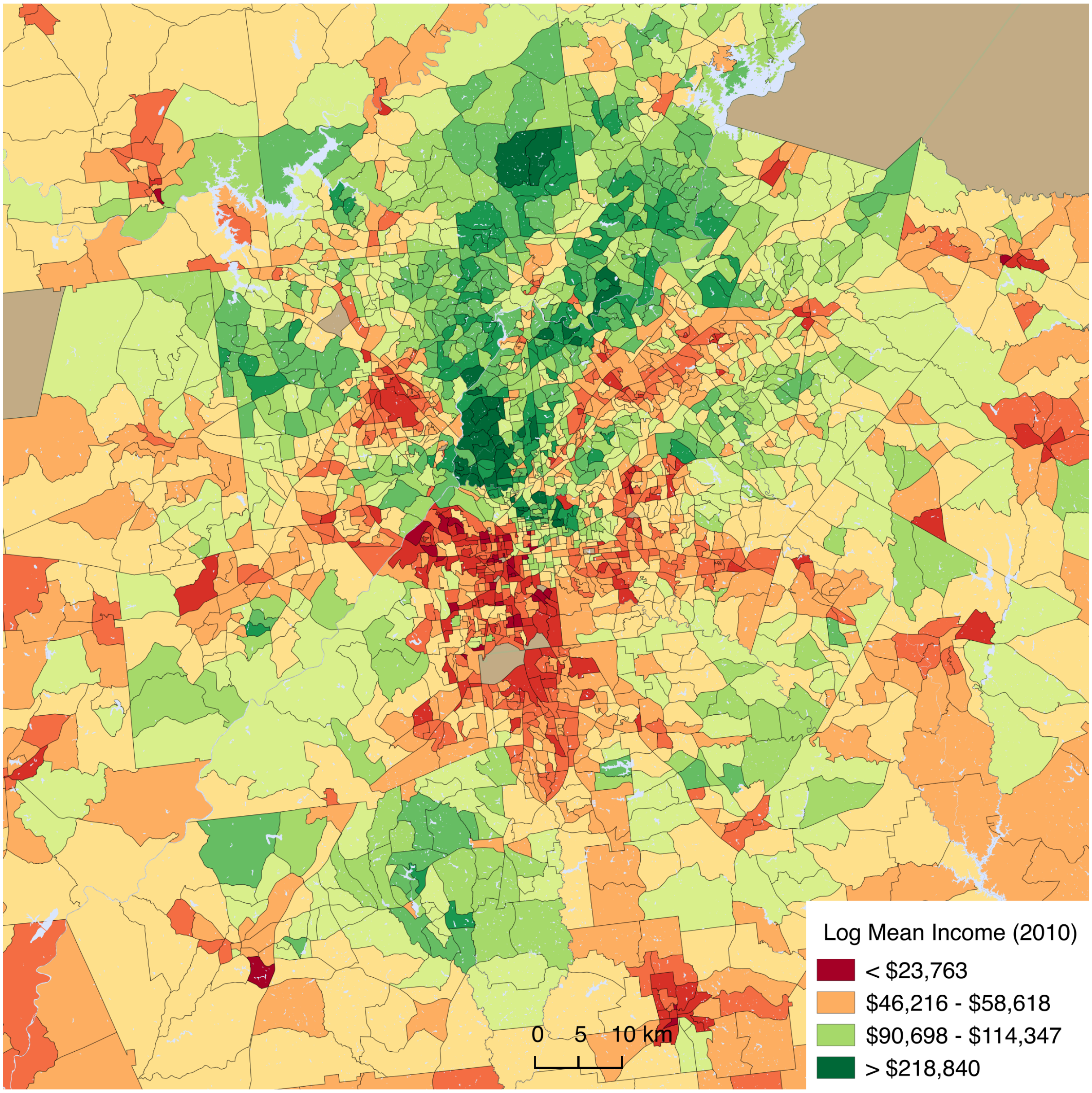

Figure S1: Atlanta, GA Mean Household Income (2010)

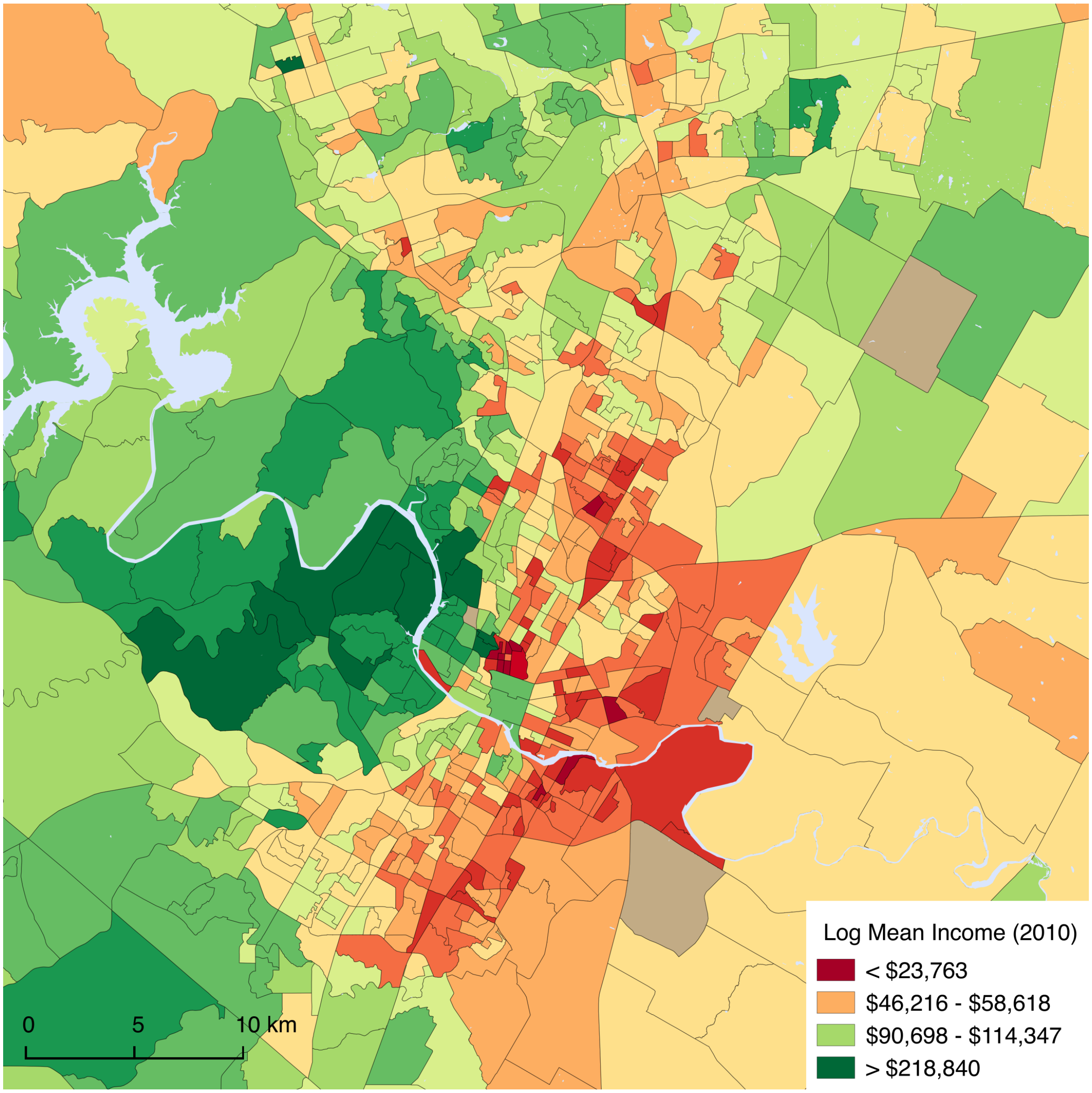

Figure S2: Austin, TX Mean Household Income (2010)

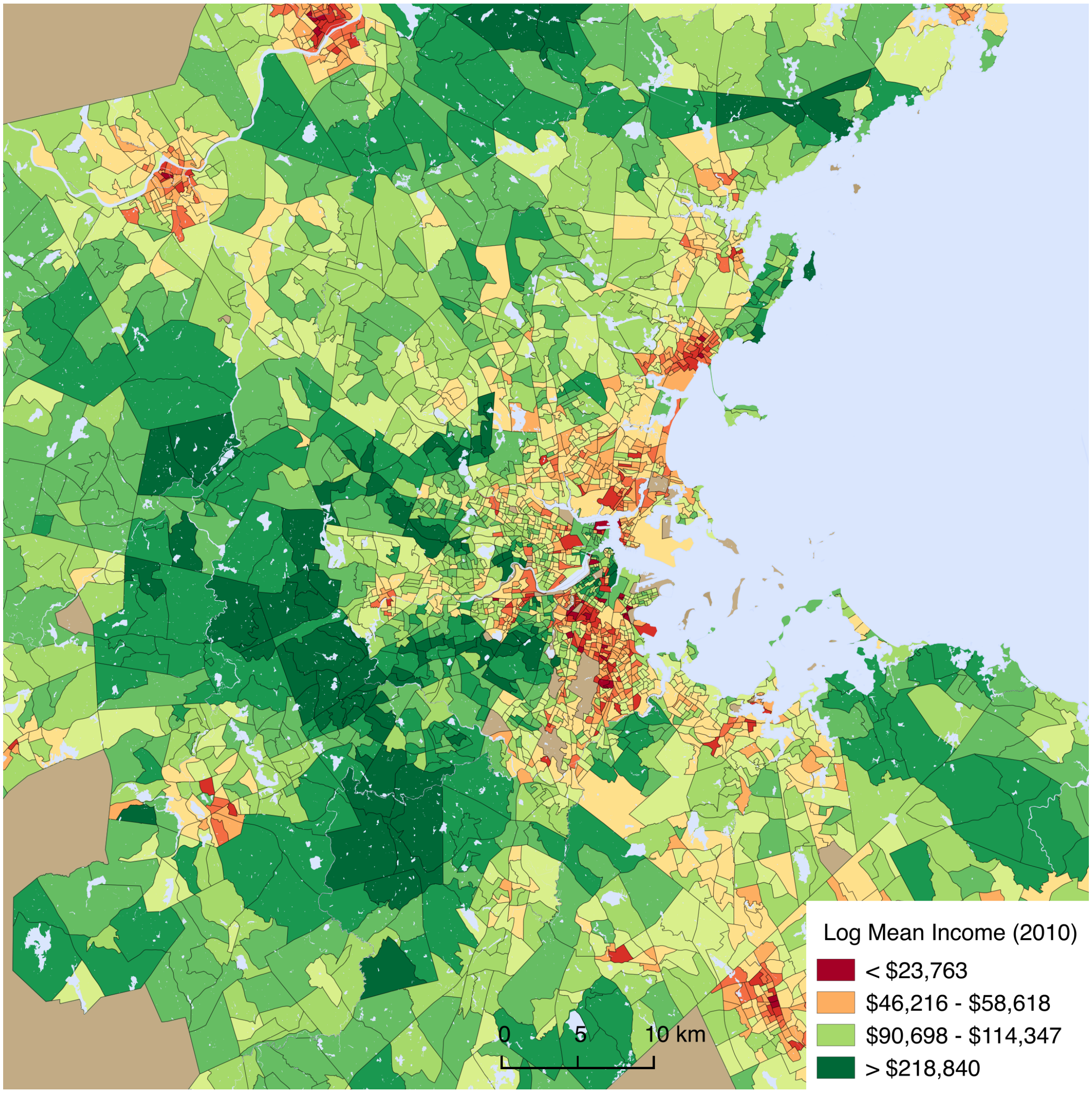

Figure S3: Boston, MA Mean Household Income (2010)

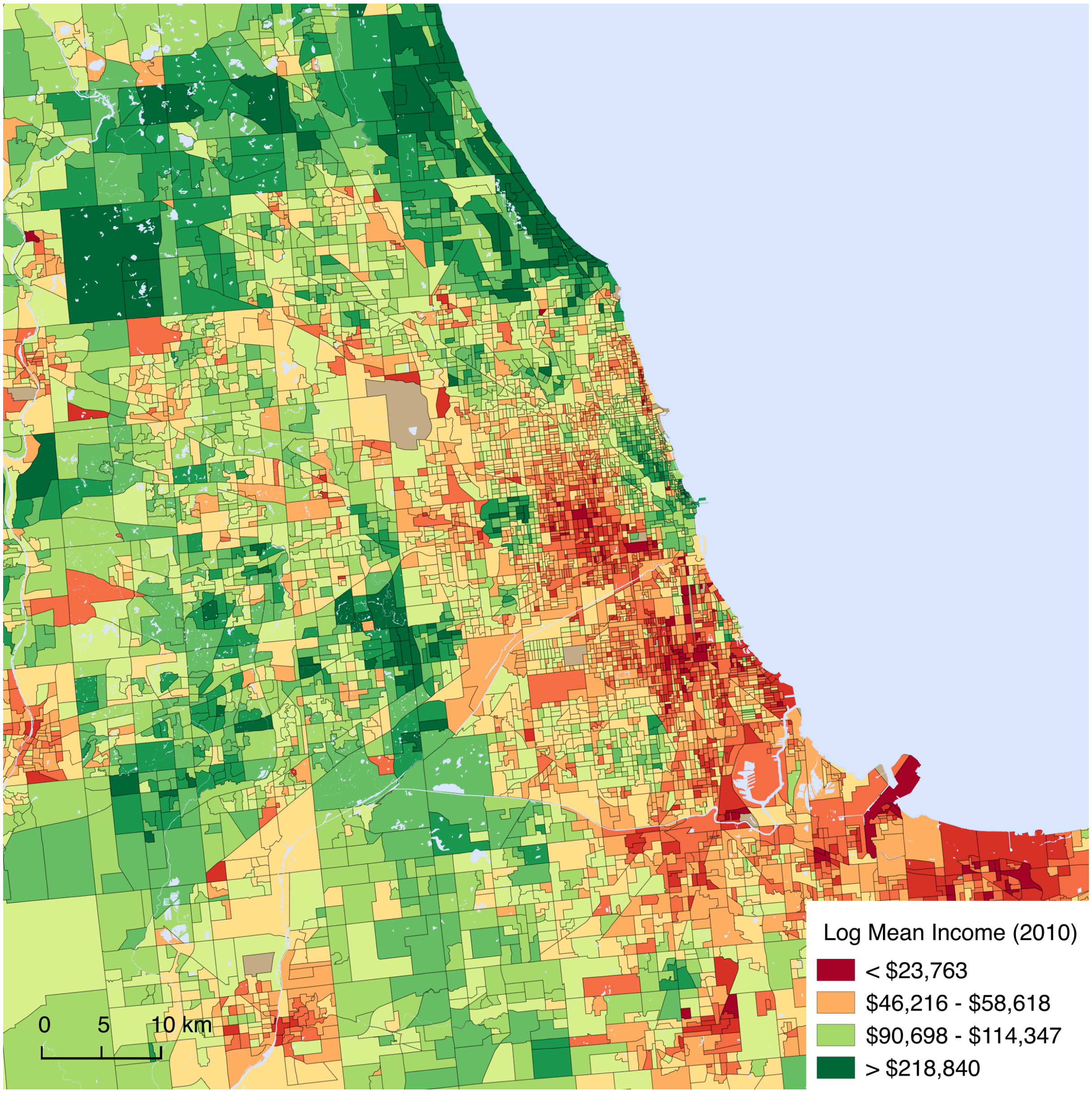

Figure S4: Chicago, IL Mean Household Income (2010)

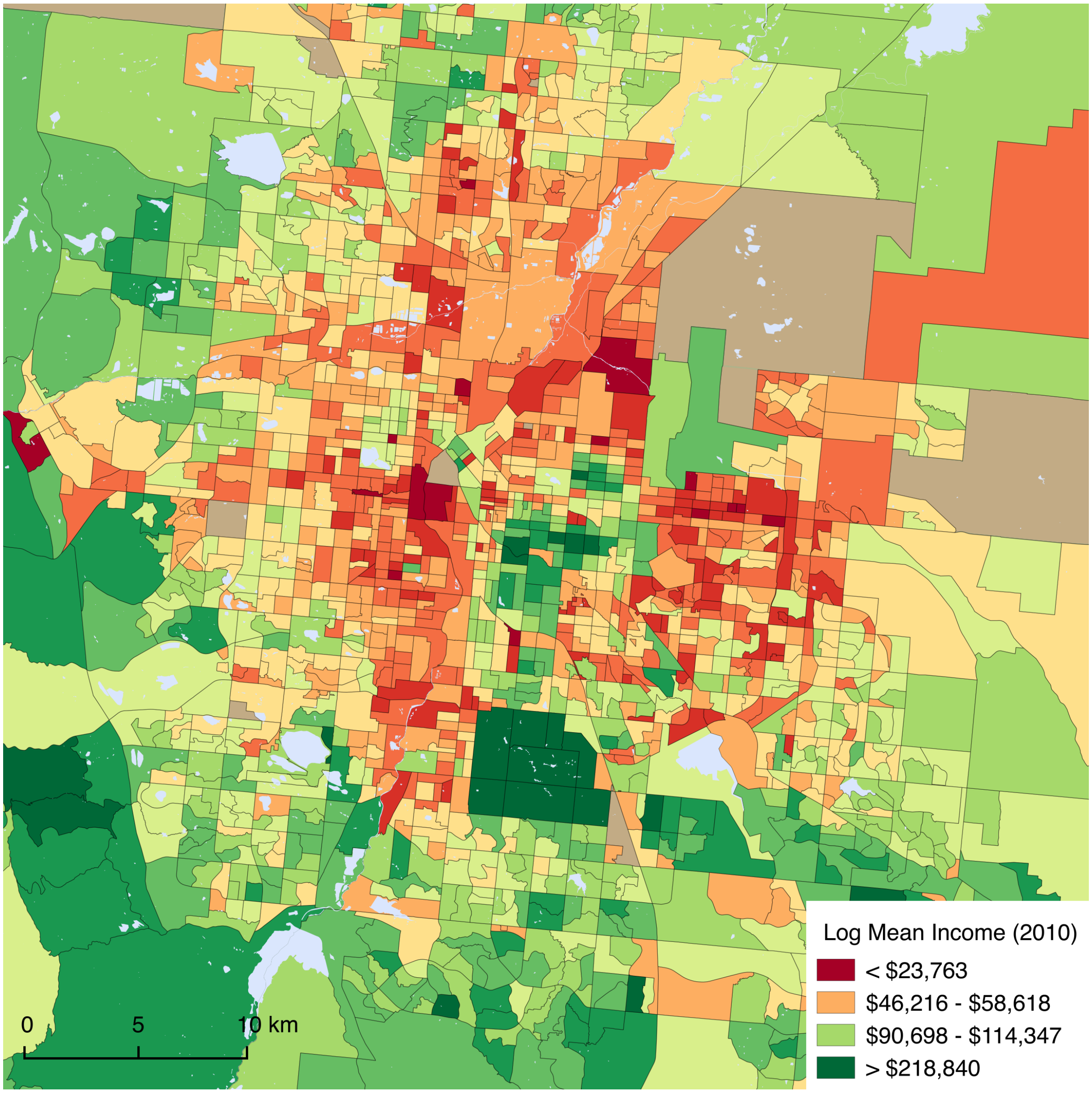

Figure S5: Denver, CO Mean Household Income (2010)

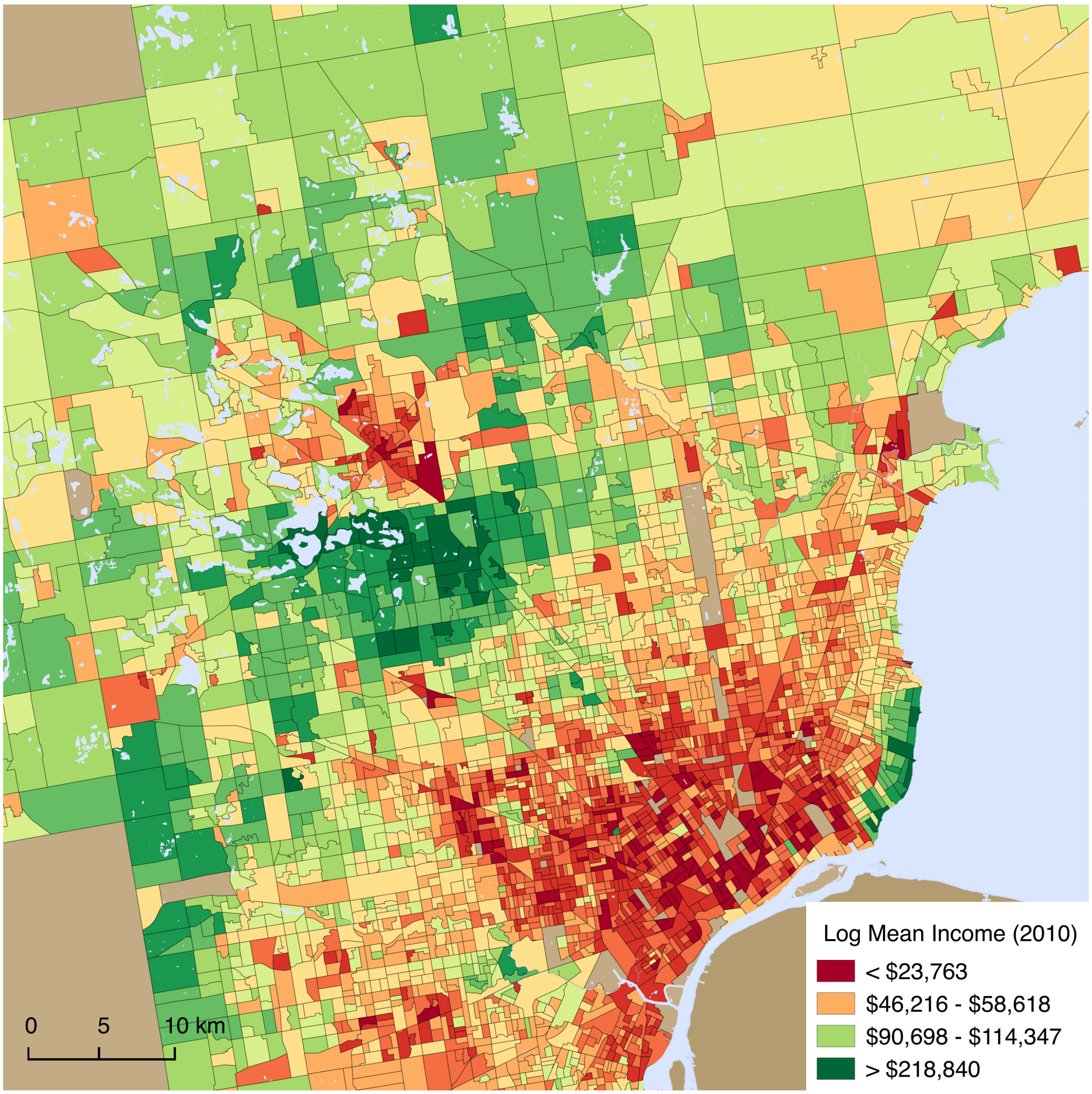

Figure S6: Detroit, MI Mean Household Income (2010)

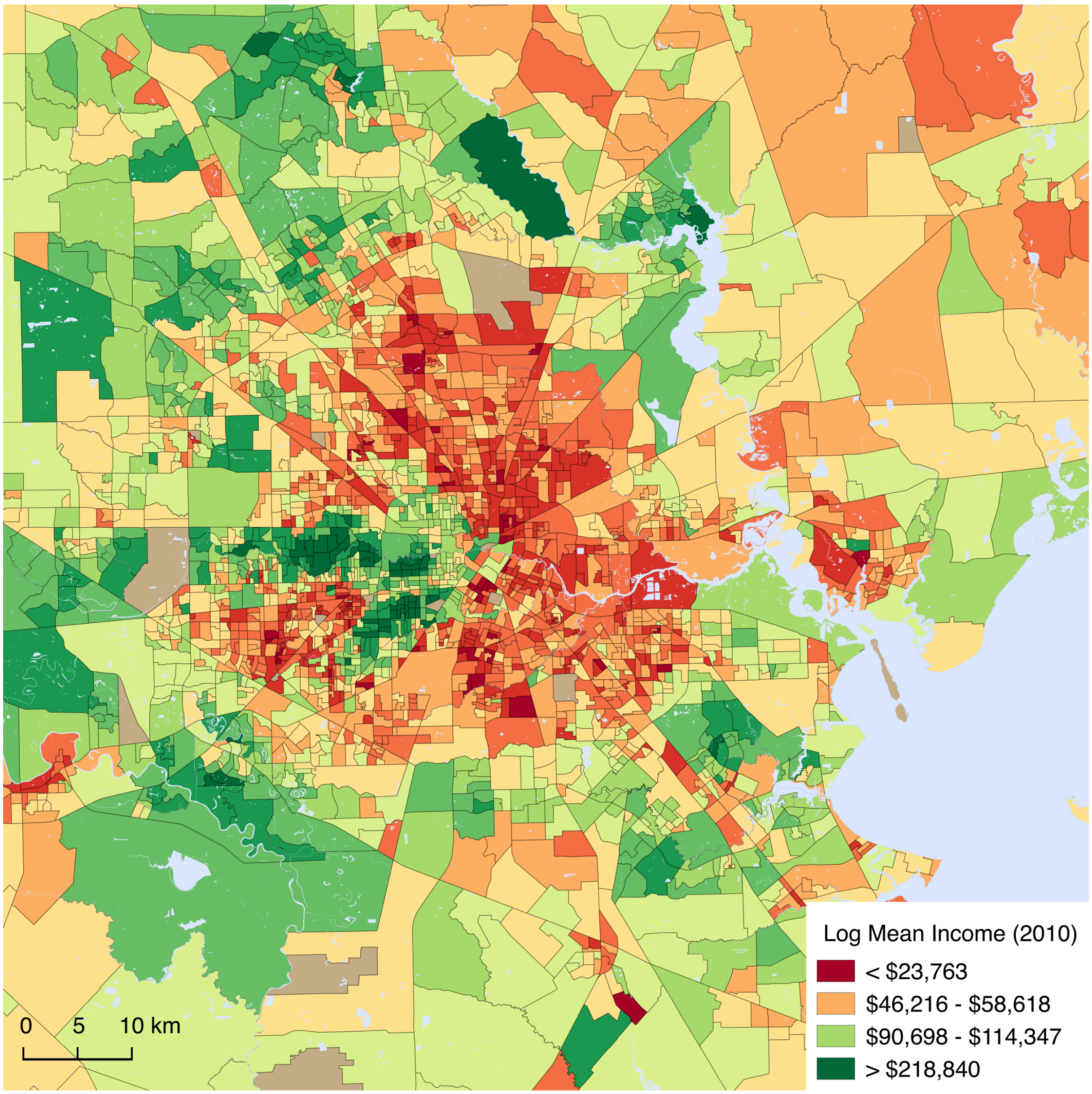

Figure S7: Houston, TX Mean Household Income (2010)

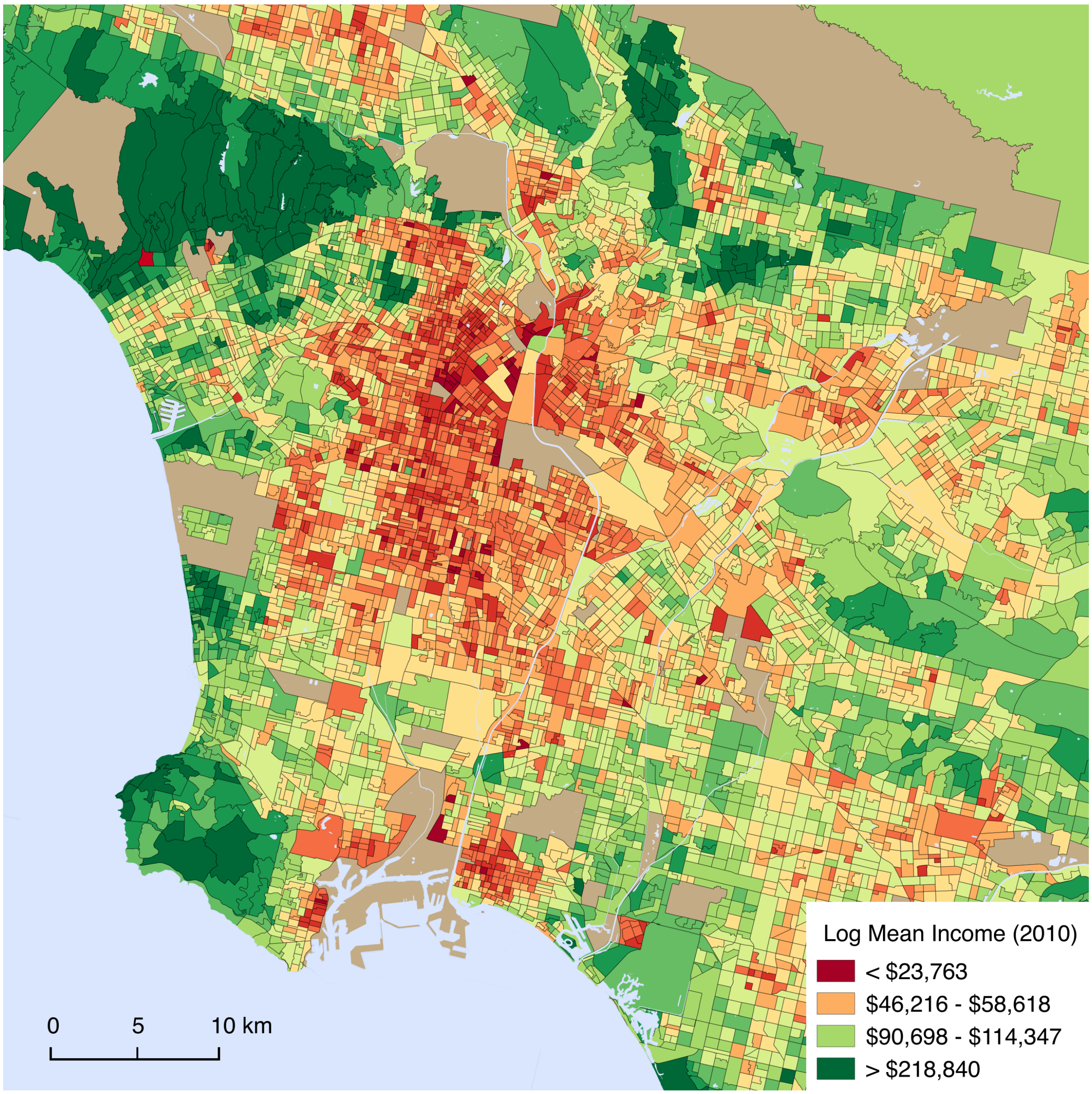

Figure S8: Los Angeles, CA Mean Household Income (2010)

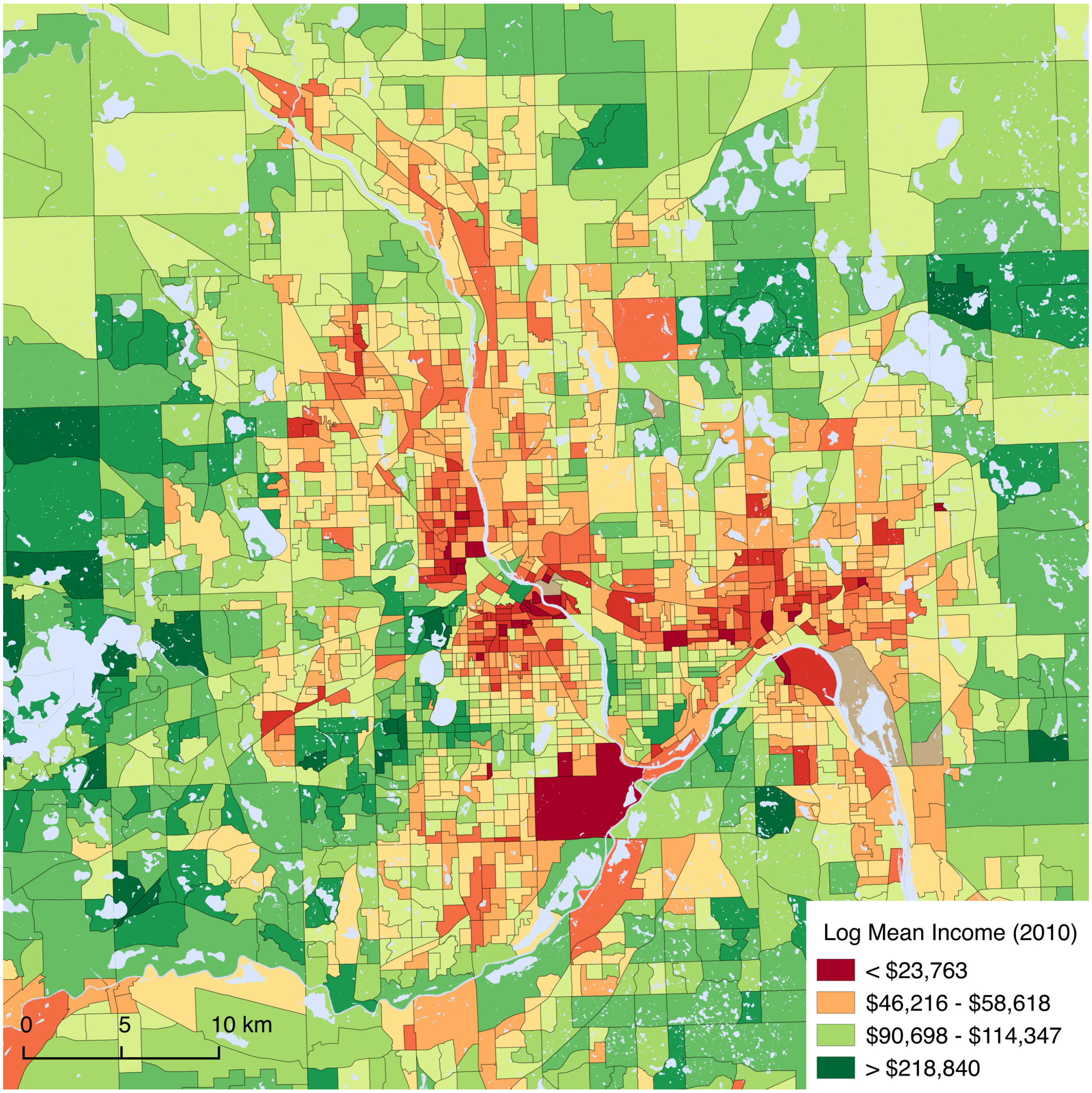

Figure S9: Minneapolis, MN Mean Household Income (2010)

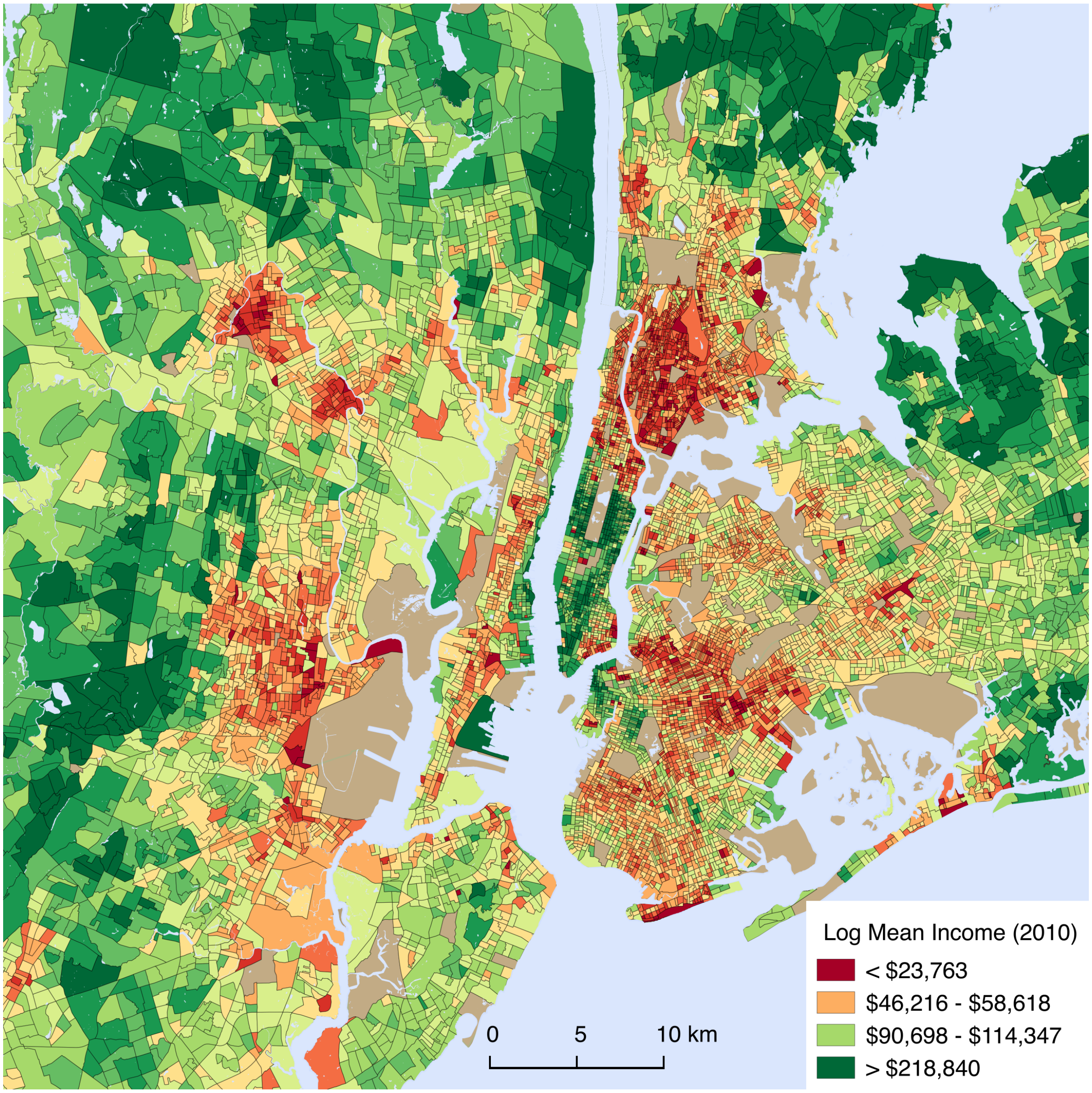

Figure S10: New York City, NY Mean Household Income (2010)

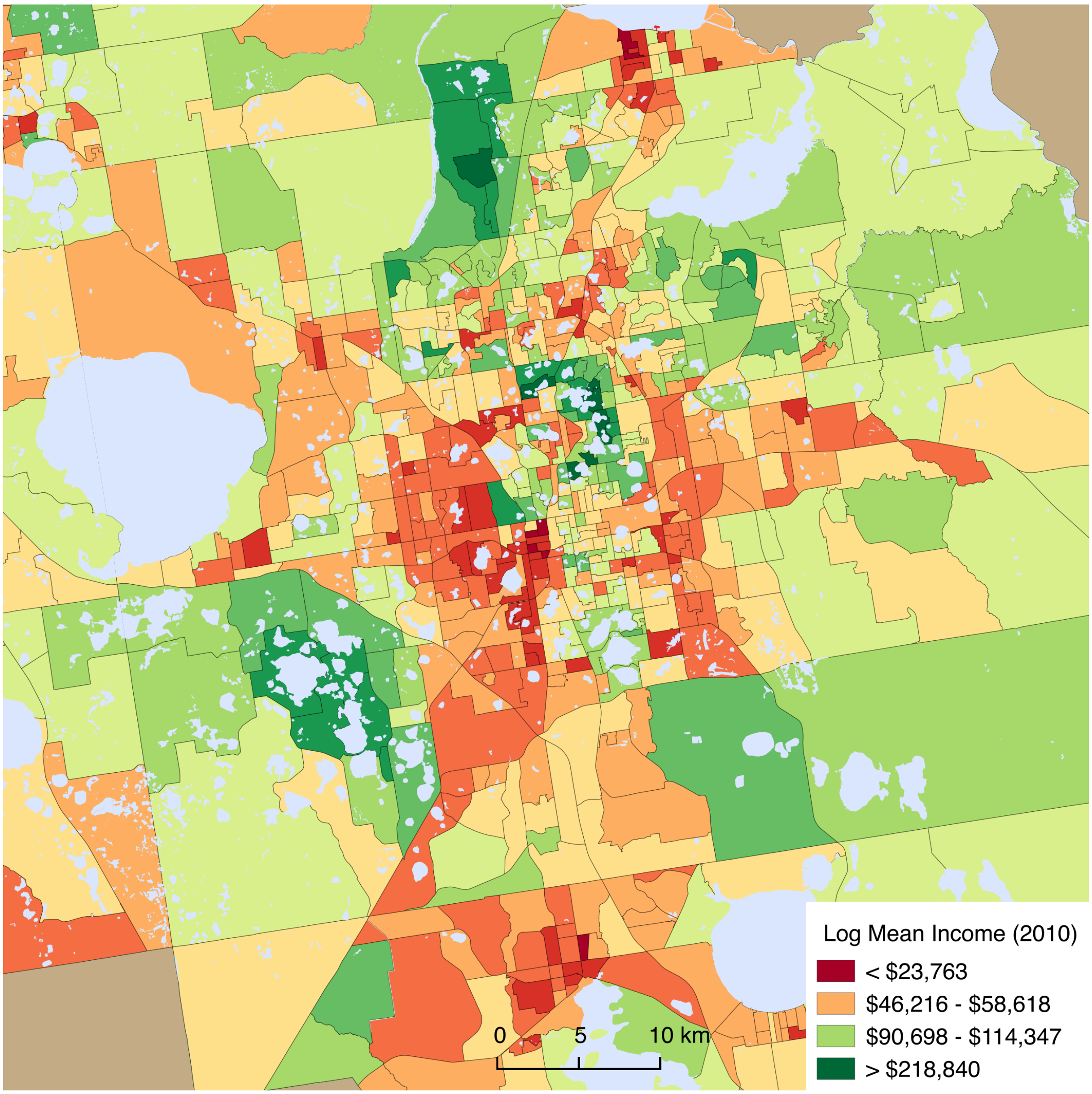

Figure S11: Orlando, FL Mean Household Income (2010)

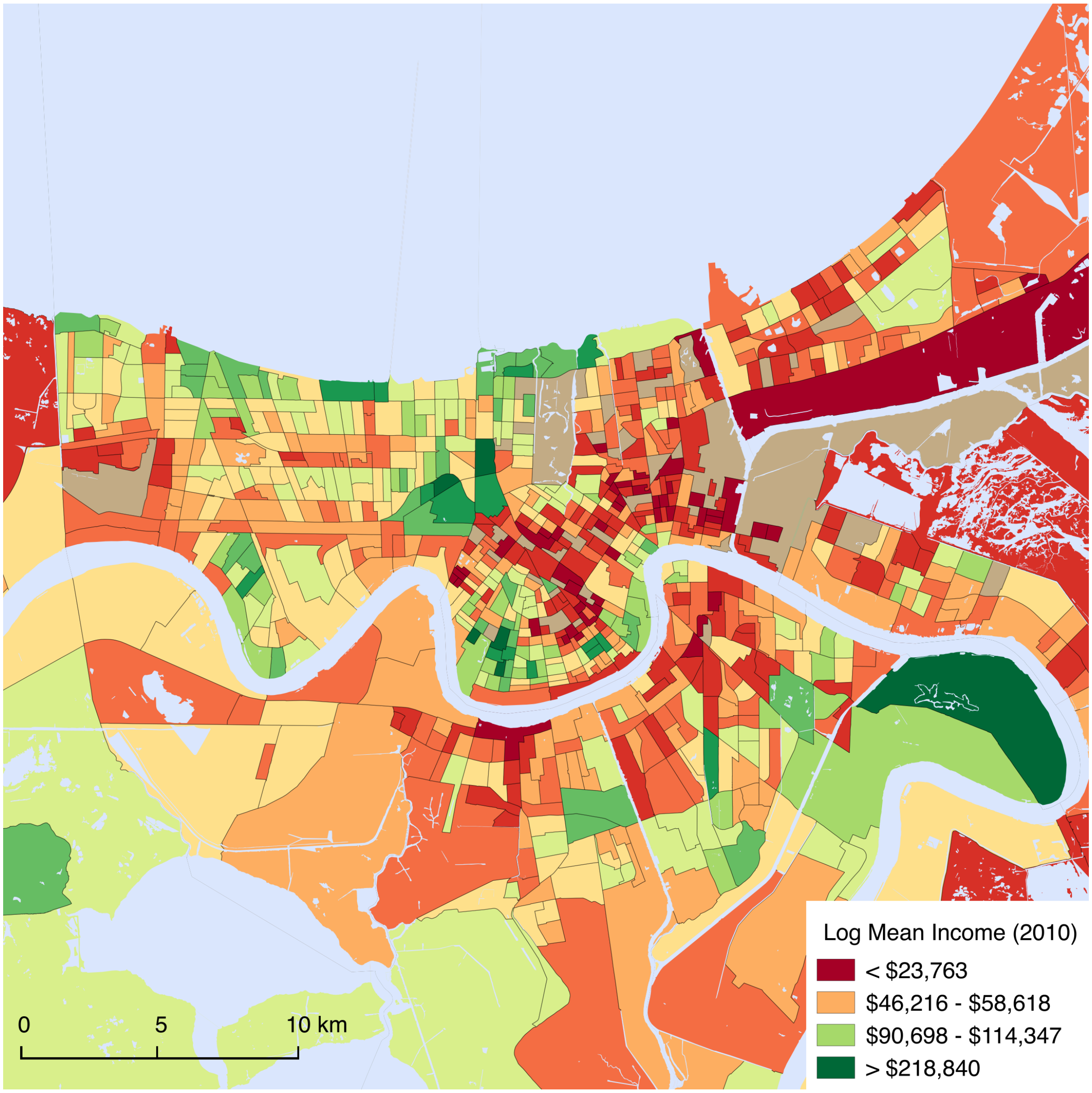

Figure S12: New Orleans, LA Mean Household Income (2010)

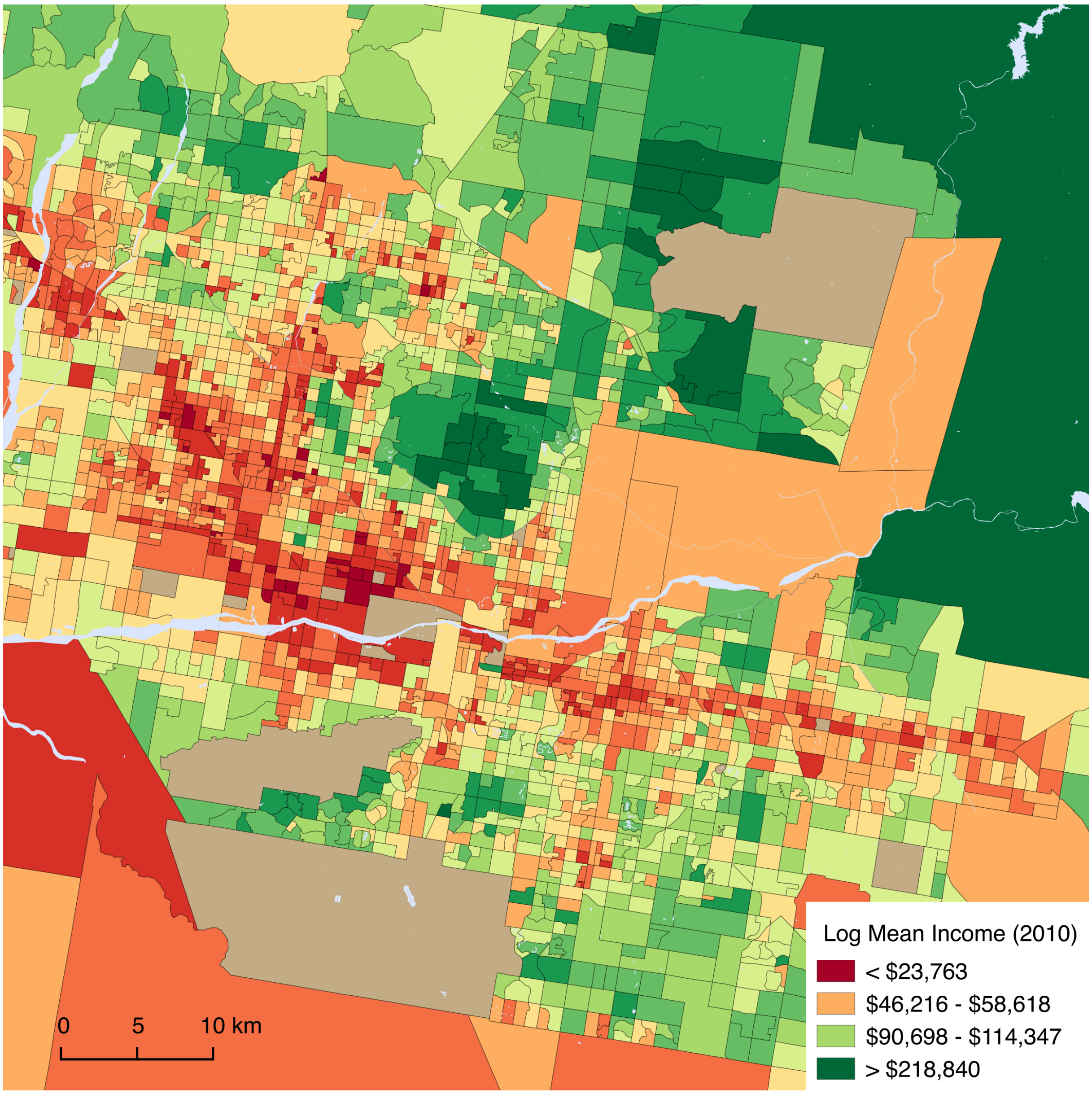

Figure S13: Phoenix, AZ Mean Household Income (2010)

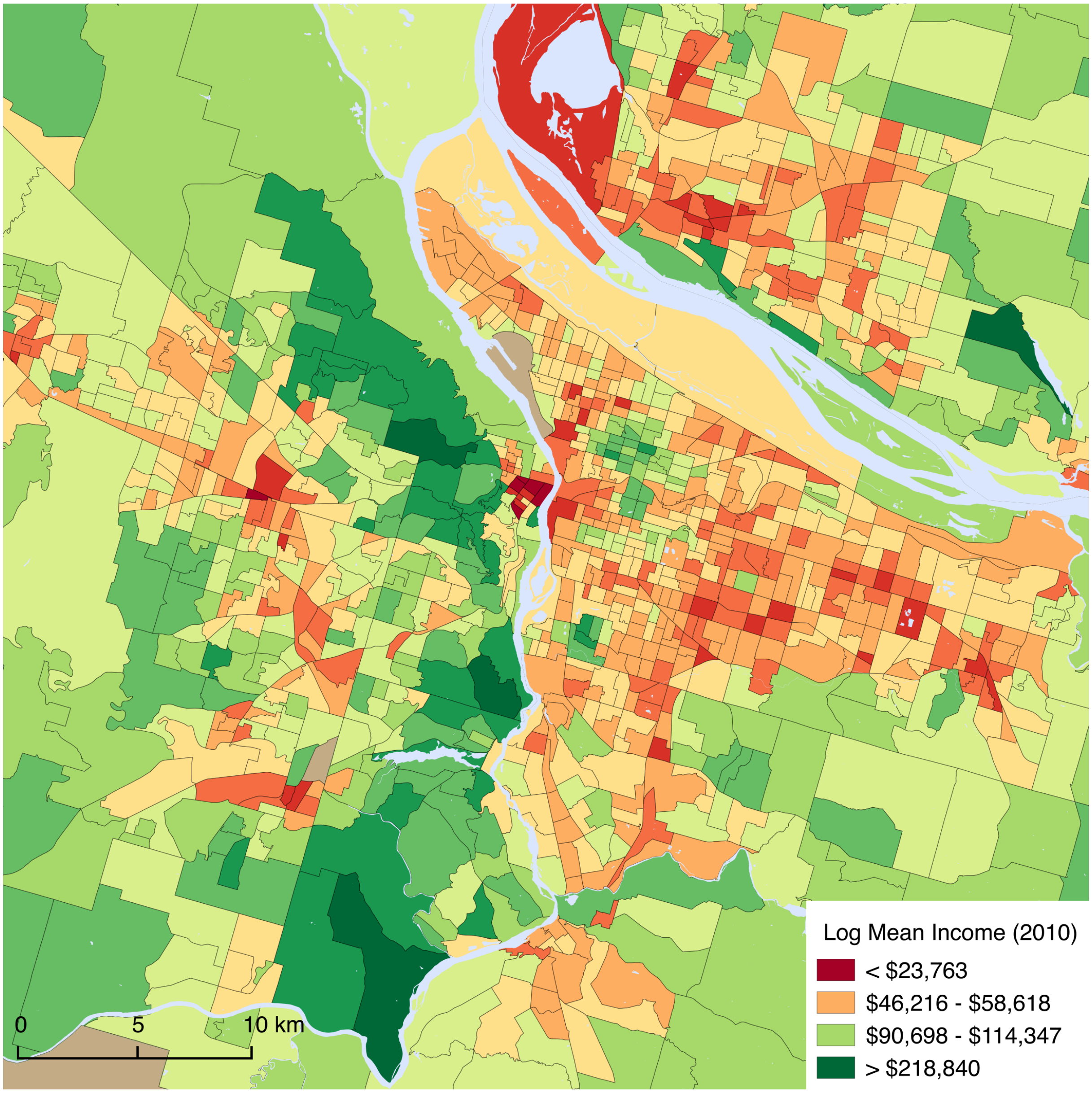

Figure S14: Portland, OR Mean Household Income (2010)

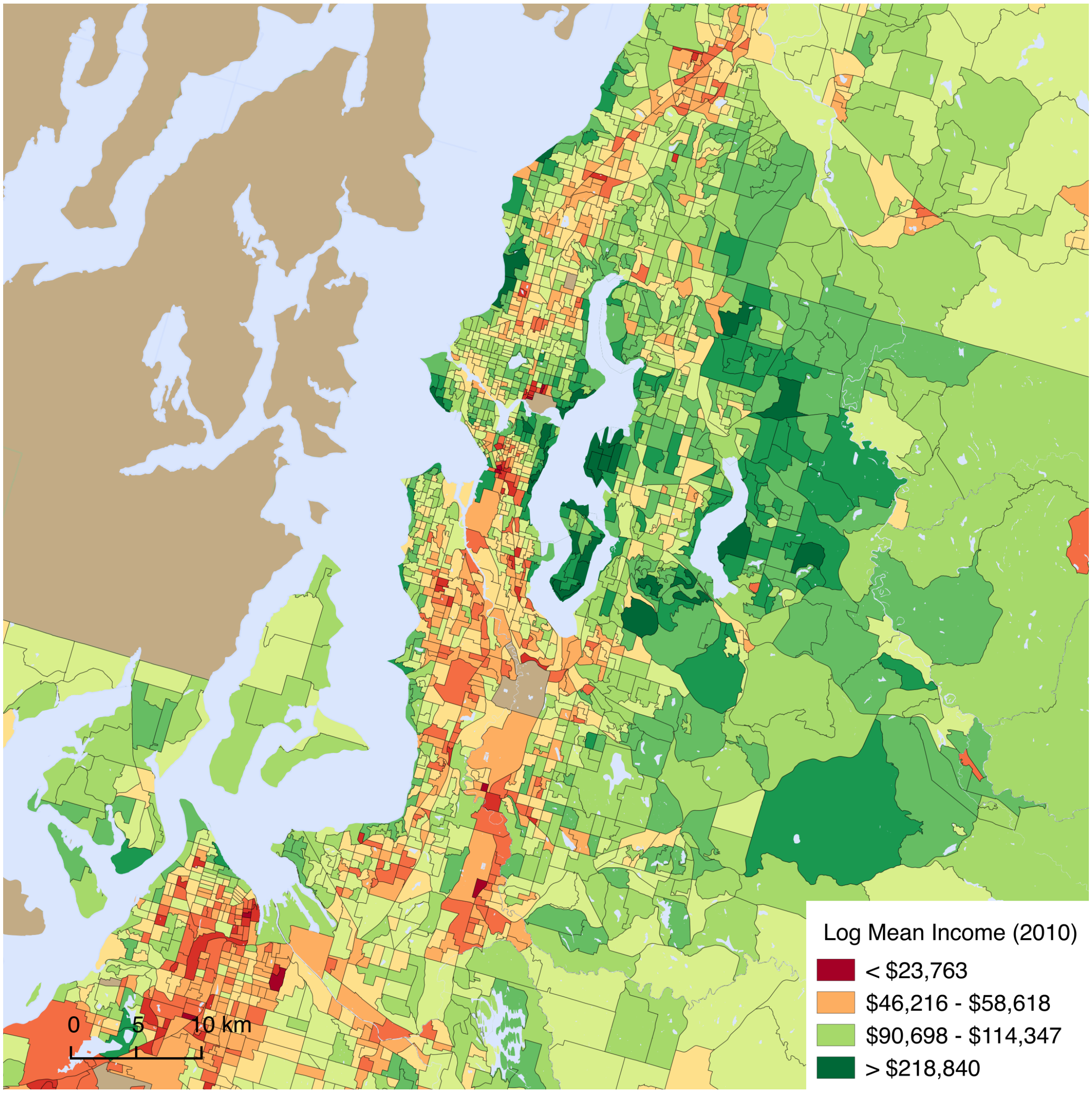

Figure S15: Seattle, WA Mean Household Income (2010)

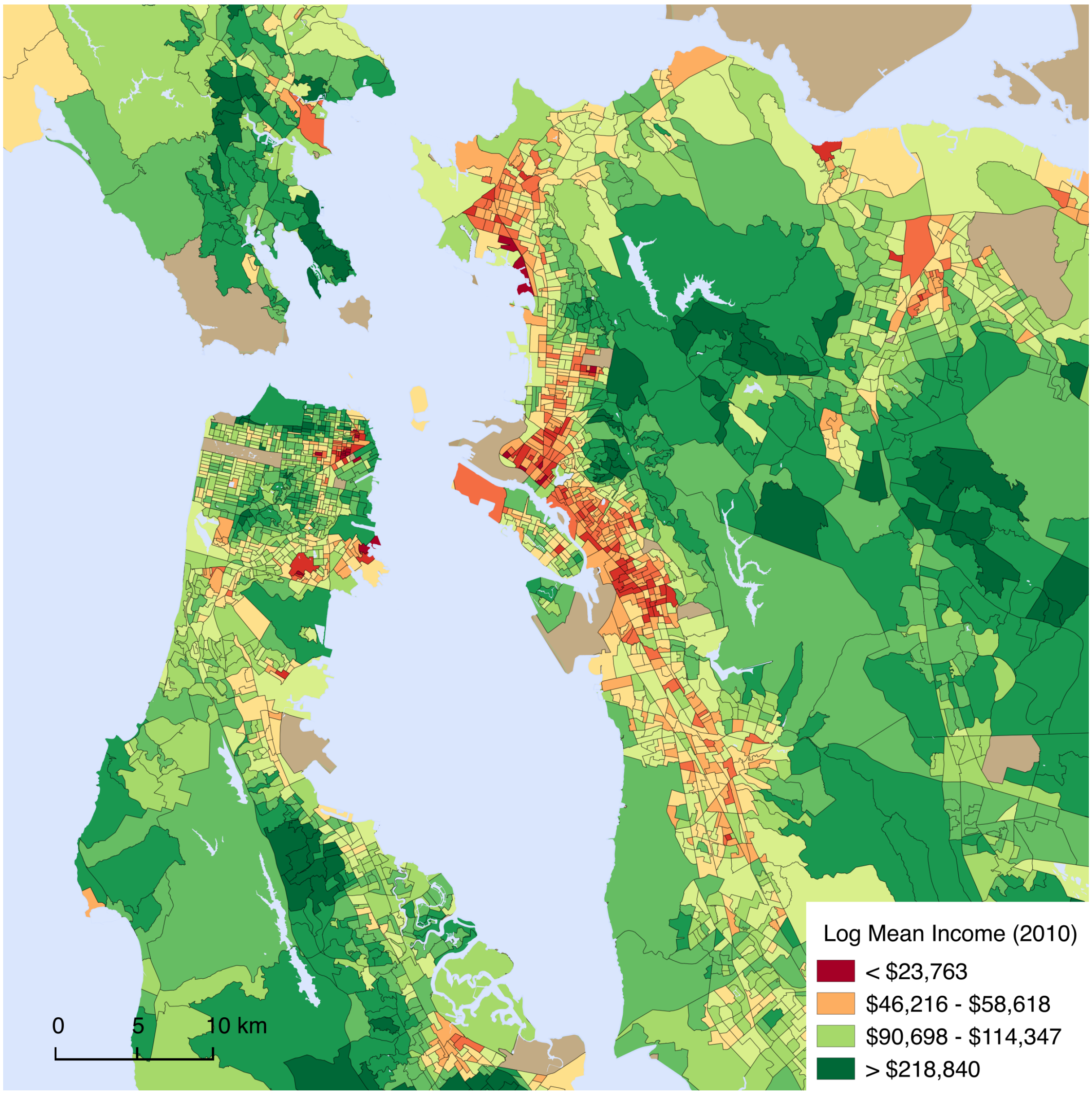

Figure S16: San Francisco, CA Mean Household Income (2010)

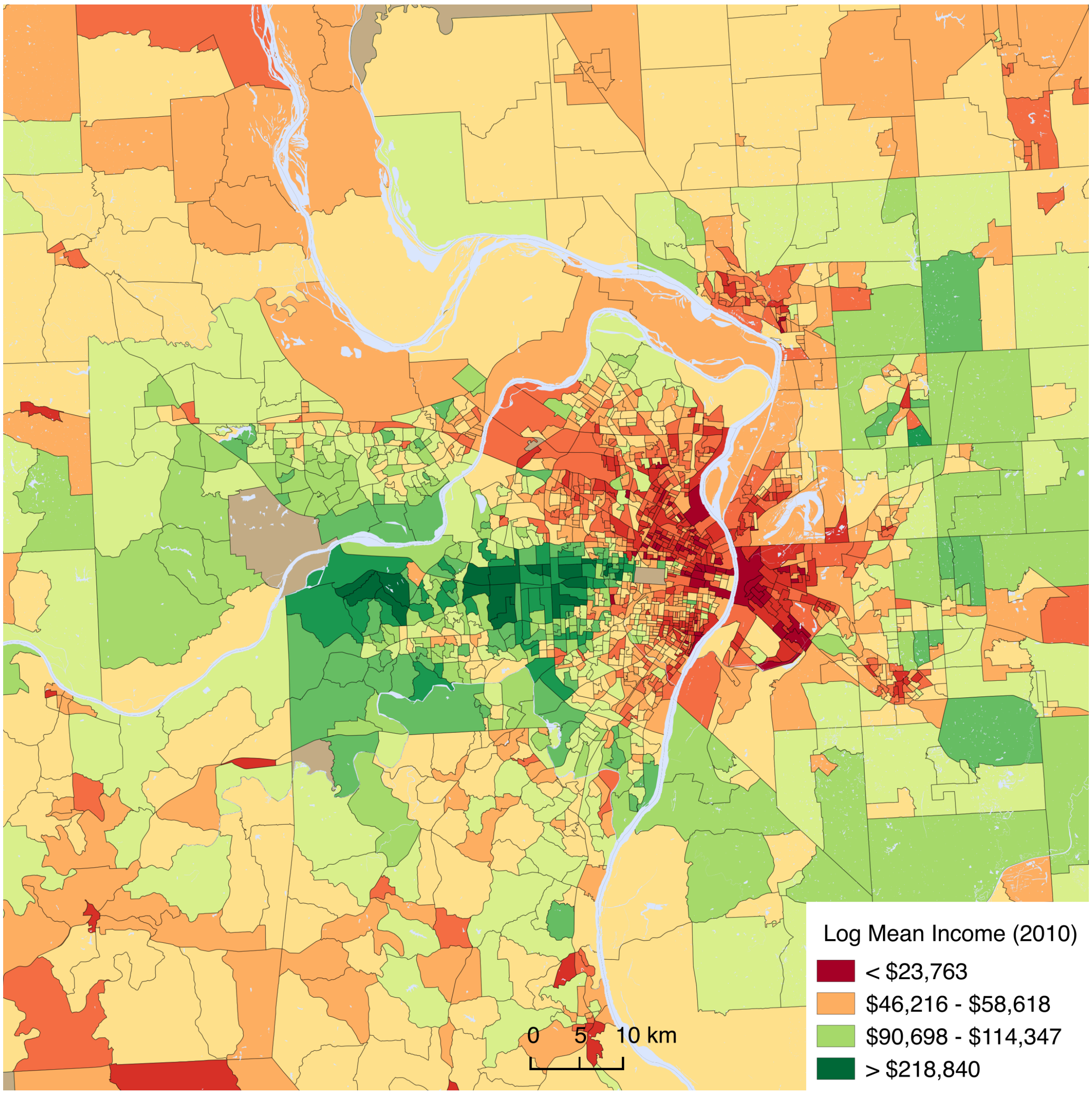

Figure S17: St. Louis, MO Mean Household Income (2010)

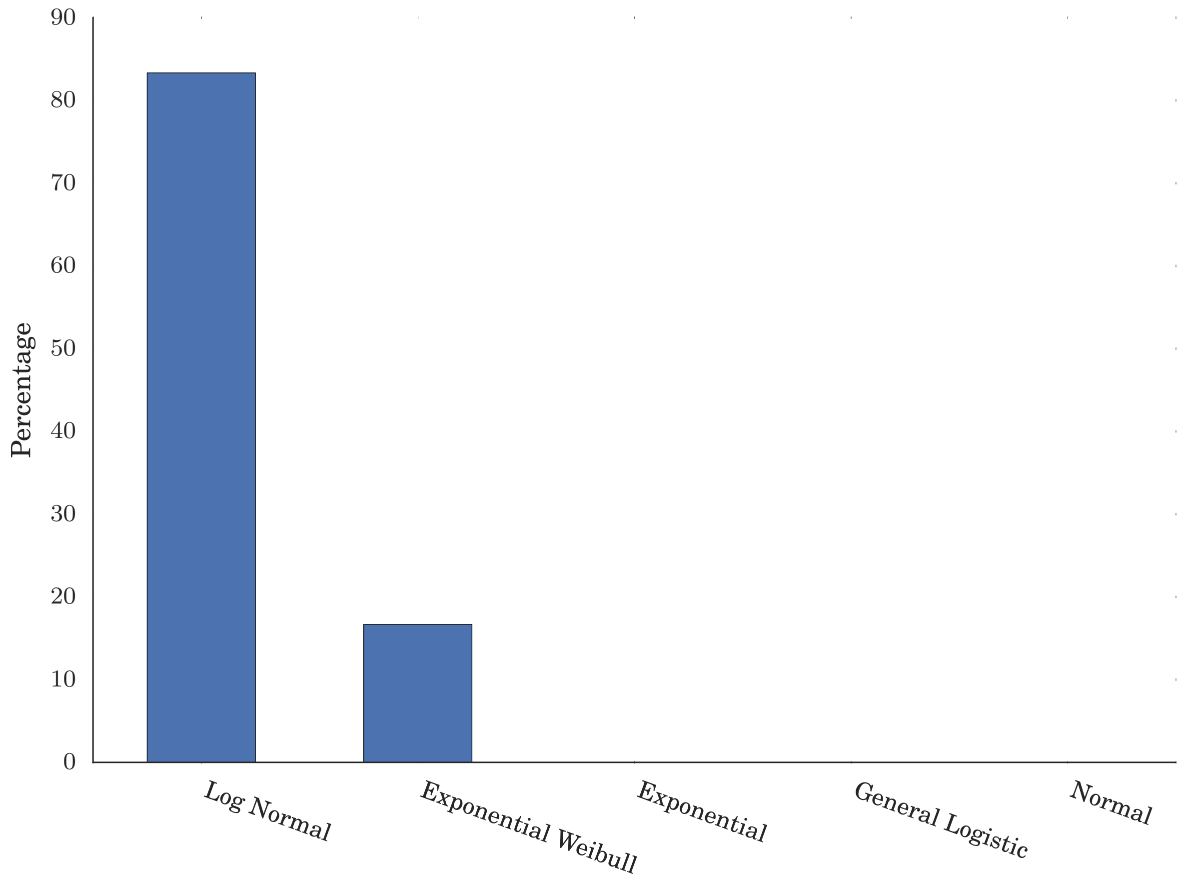

Figure S18: BIC fit test for the distribution of mean income in all CBSA's in the US, first place results.

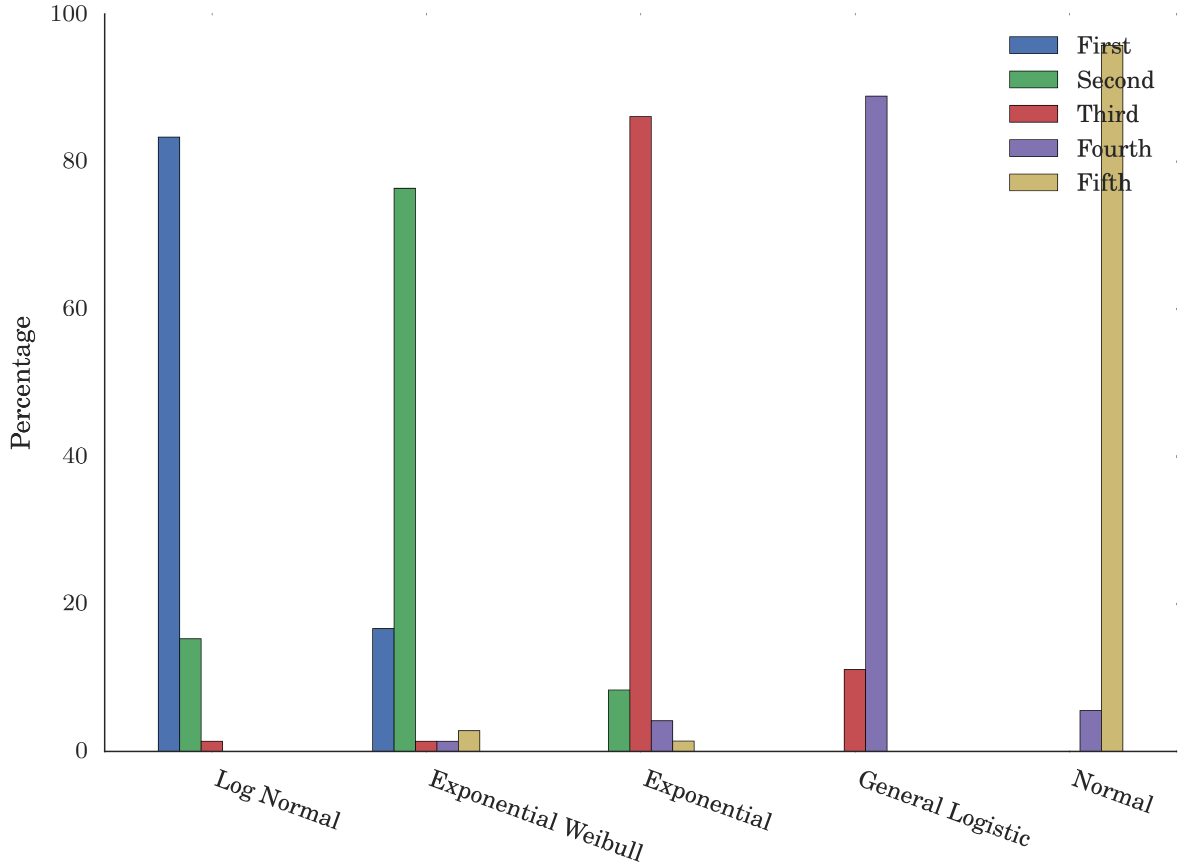

Figure S19: BIC fit test for the distribution of mean income in all CBSA's in the US

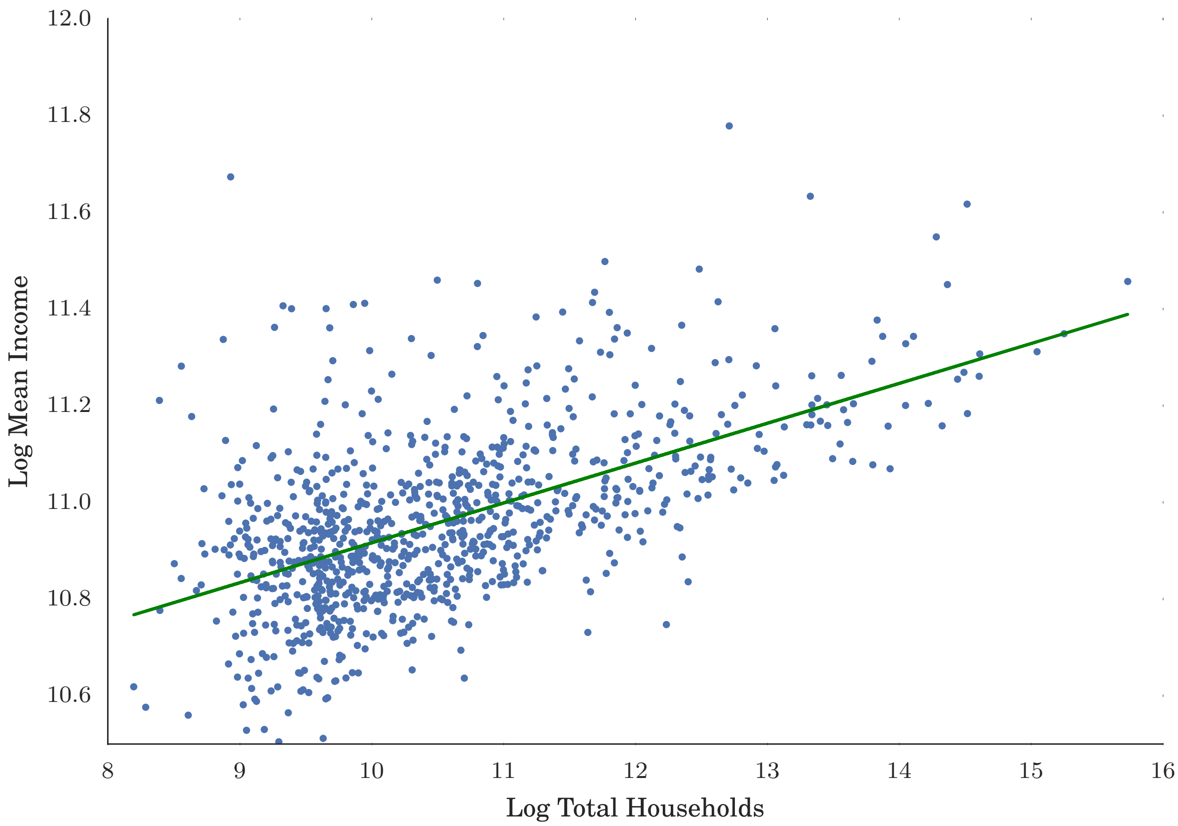

Figure S20: Scaling Mean. The mean income of a city, in relation to number of households, is characterized by an exponent of 0.0825 ($95\% CI = [0.075, 0.090]$, $R^2 = 0.317$).

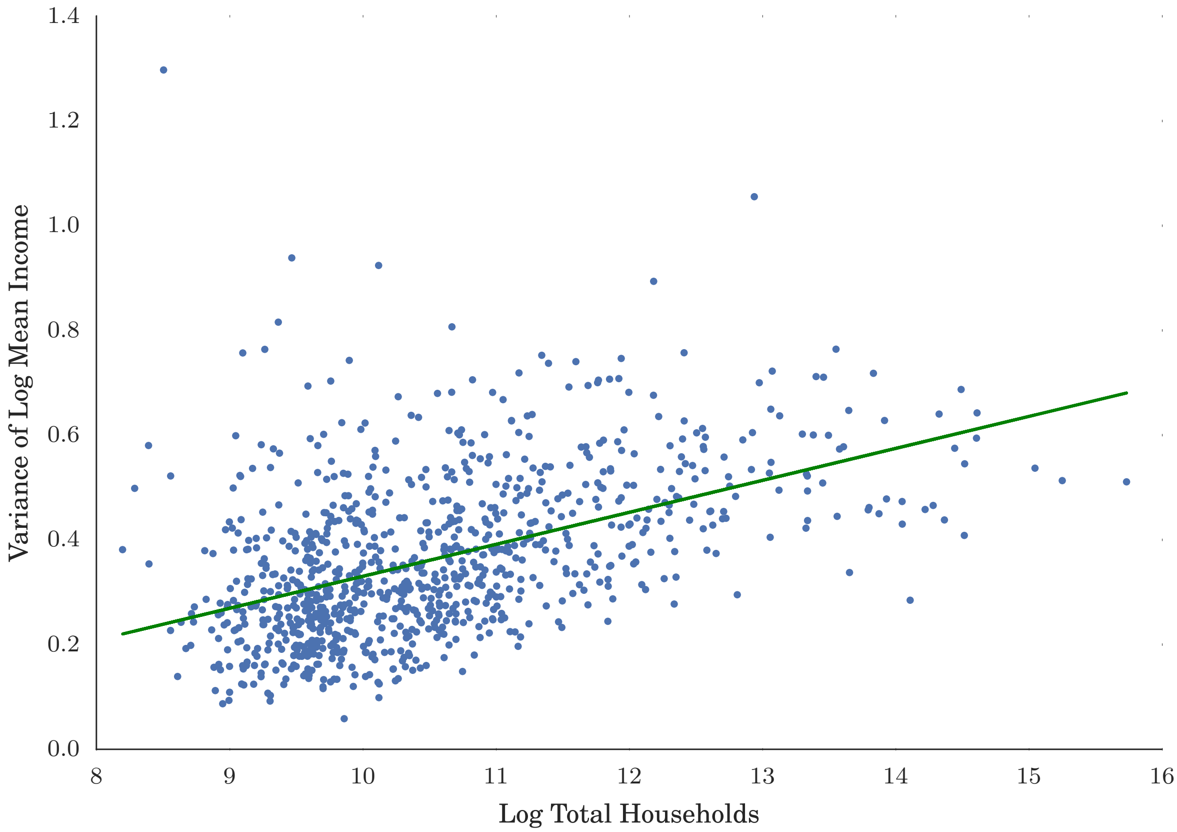

Figure S21: Scaling Variance. The variance of block group income for a city, in relation to number of households, is characterized by an exponent of 0.0611 ($95\% CI = [0.054, 0.068], R^2 = 0.235$).

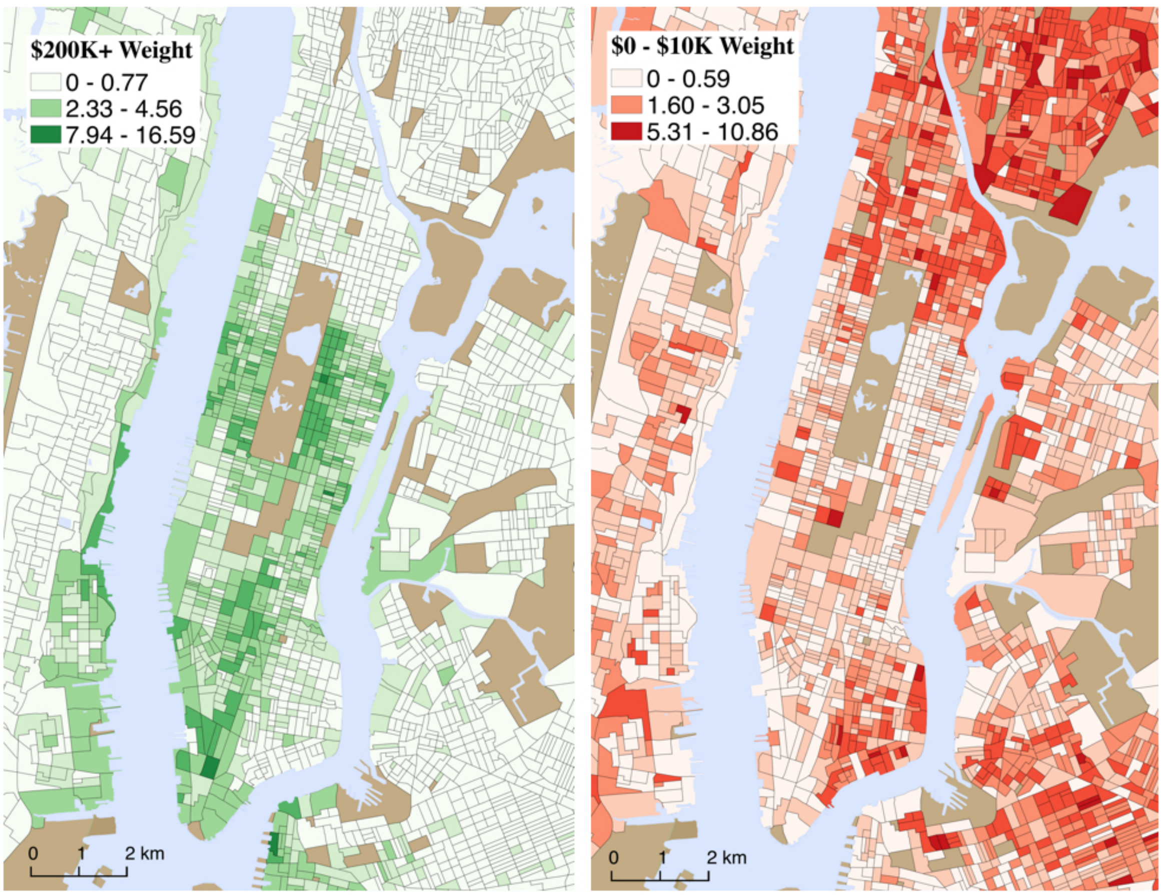

Figure S22: The weights $w_{\ell,j}$ for the richest (left, in green) and poorest (right, in red) income bins for New York City.

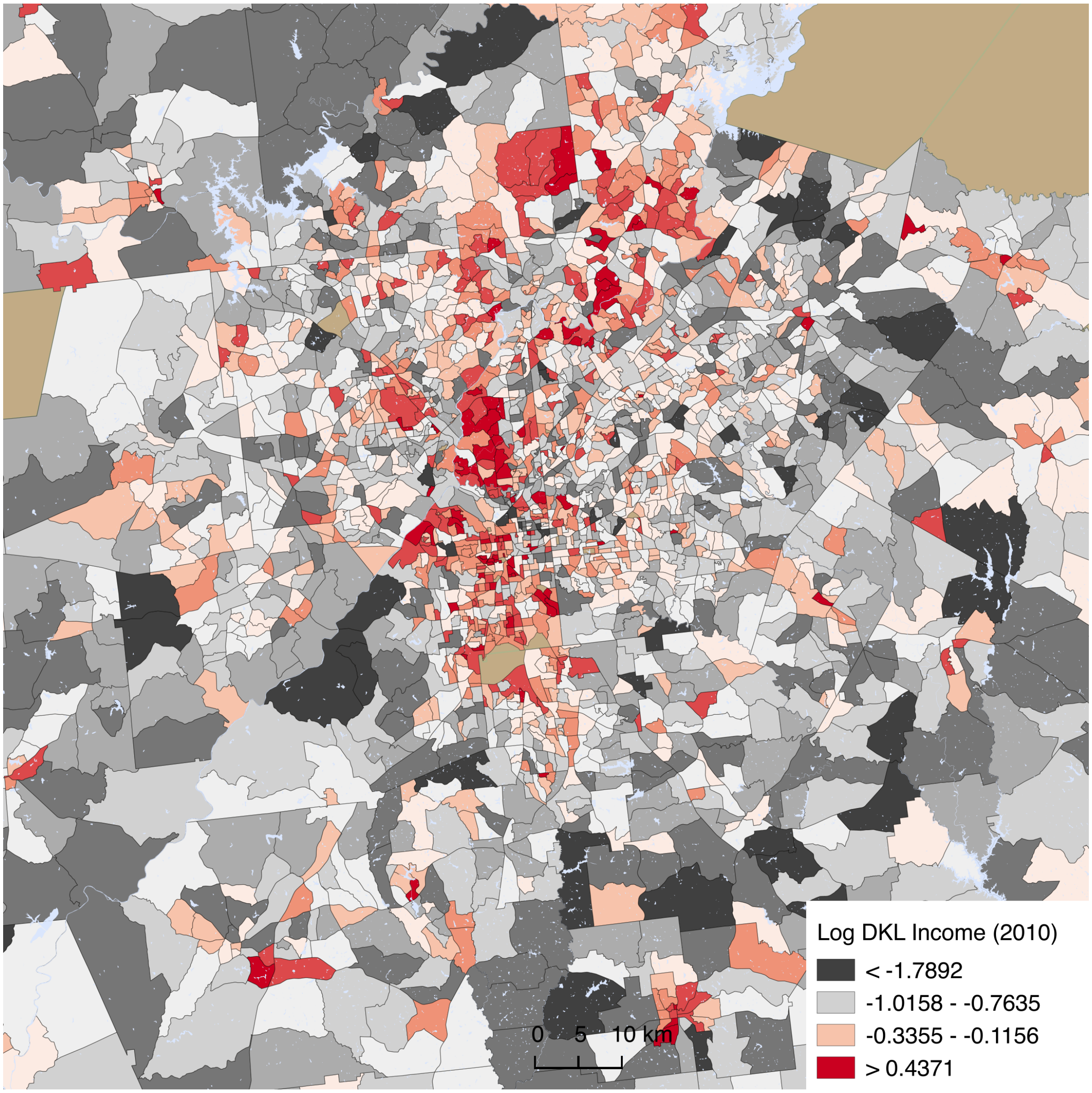

Figure S23: Atlanta, GA D_{KL} (2010)

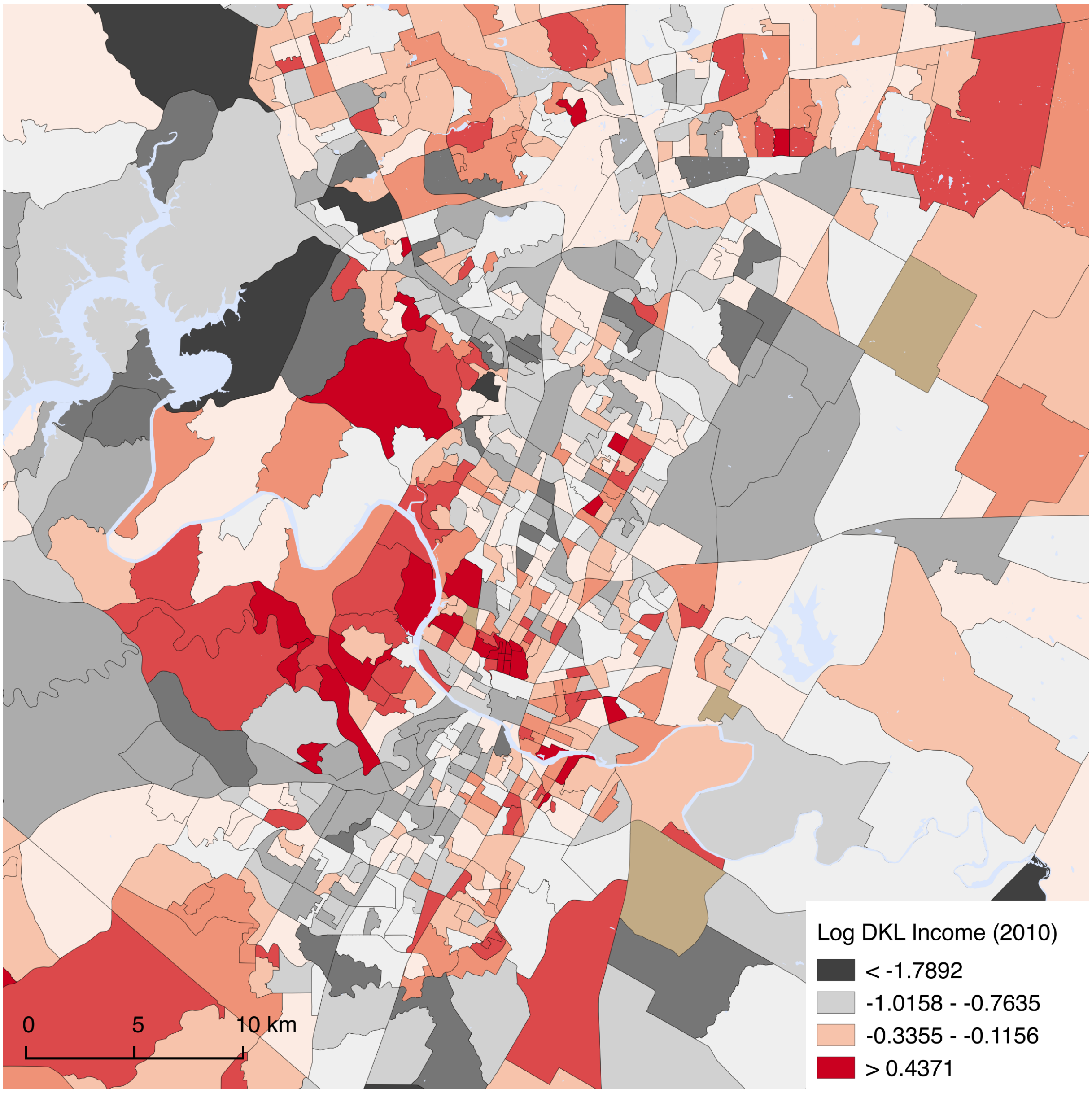

Figure S24: Austin, TX D_{KL} (2010)

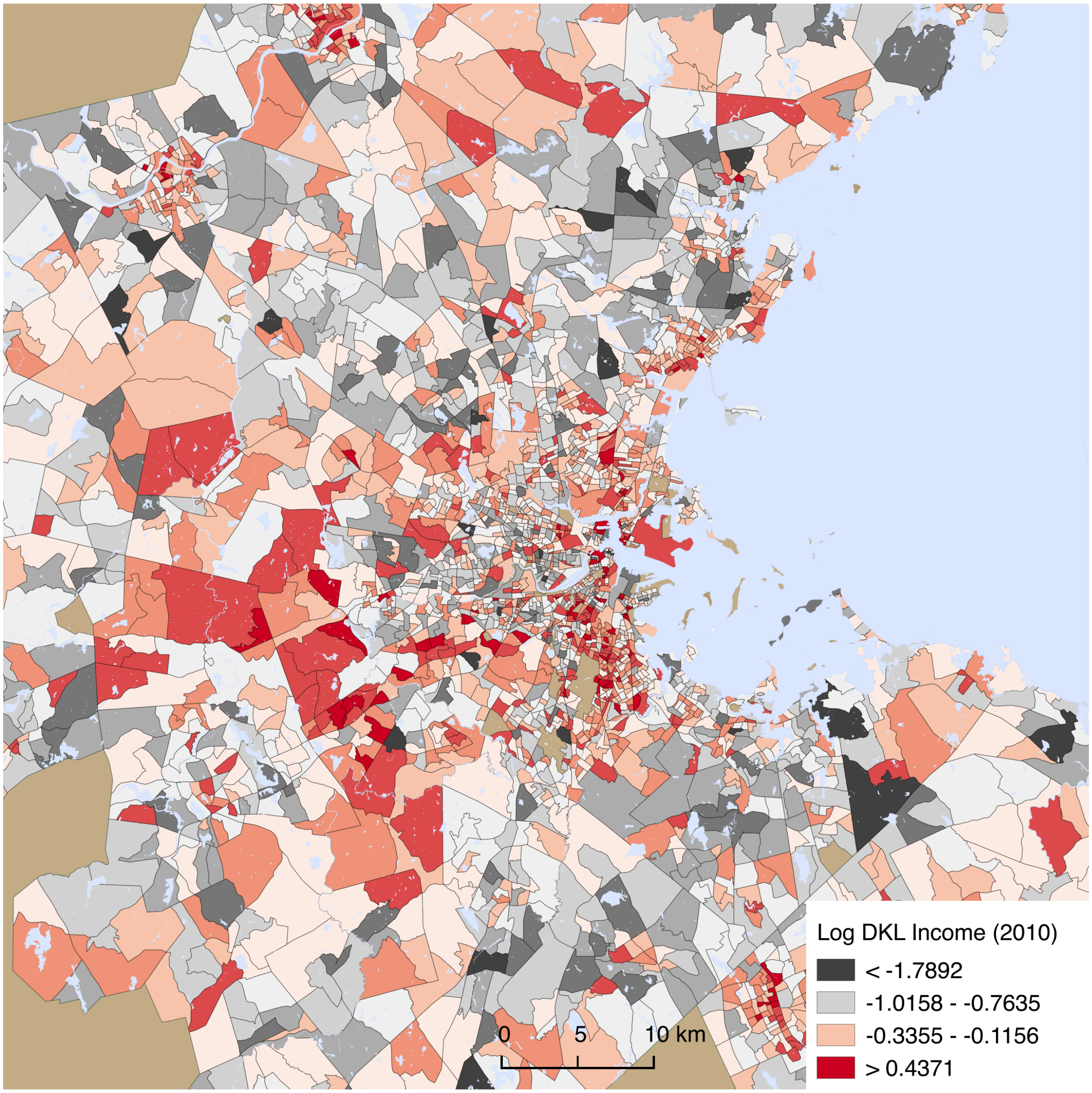

Figure S25: Boston, MA D_{KL} (2010)

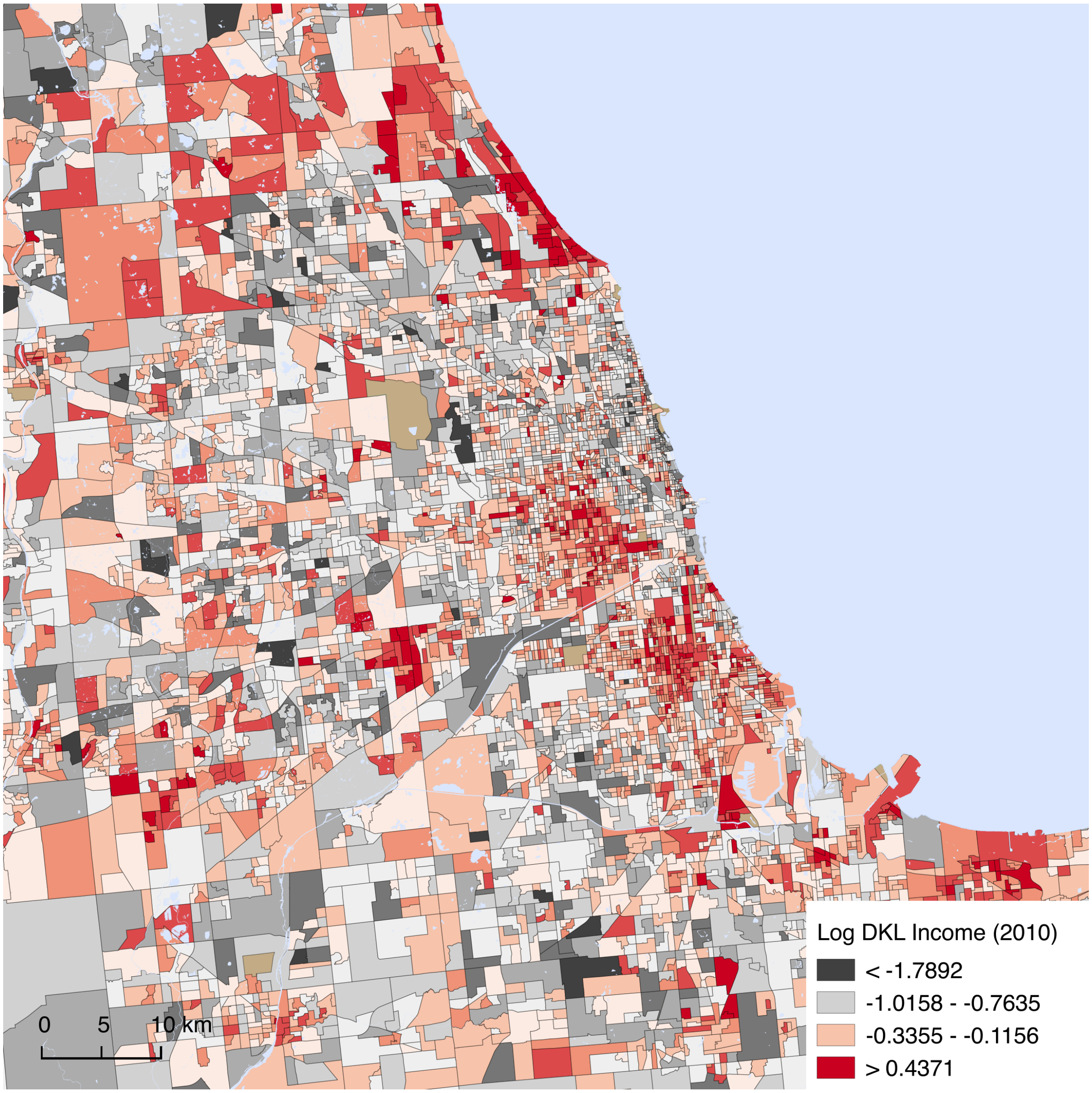

Figure S26: Chicago, IL D_{KL} (2010)

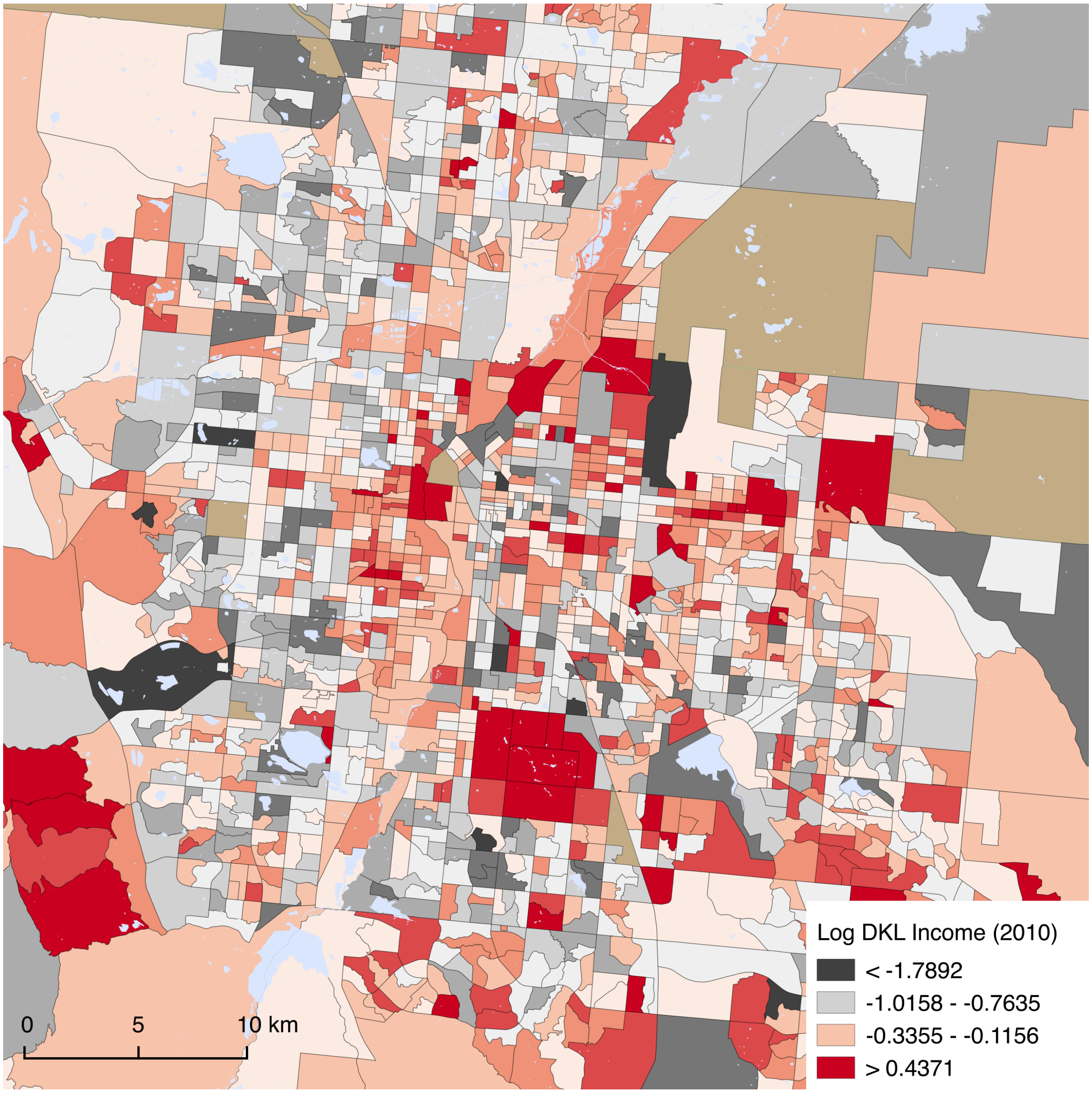

Figure S27: Denver, CO D_{KL} (2010)

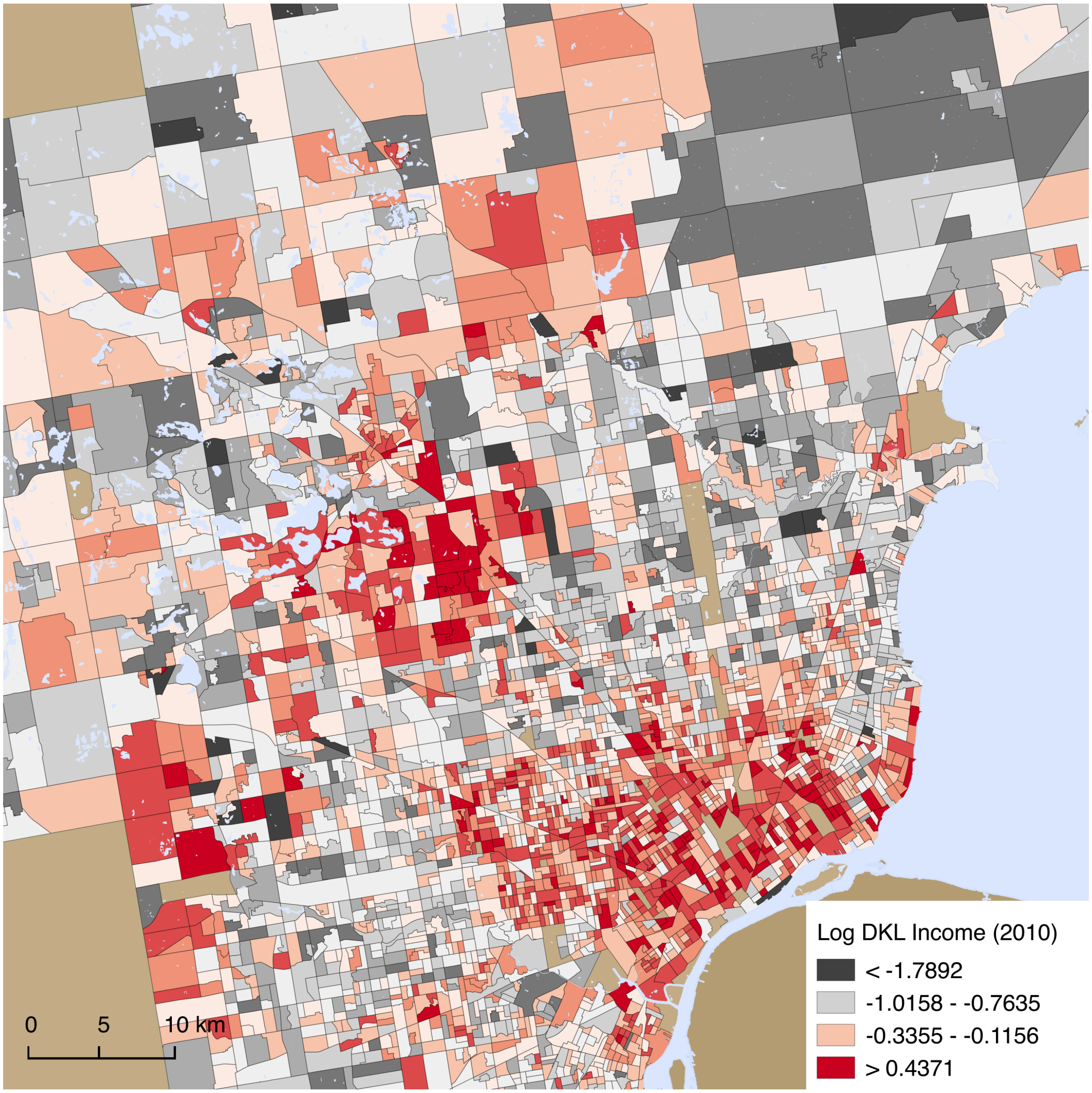

Figure S28: Detroit, MI D_{KL} (2010)

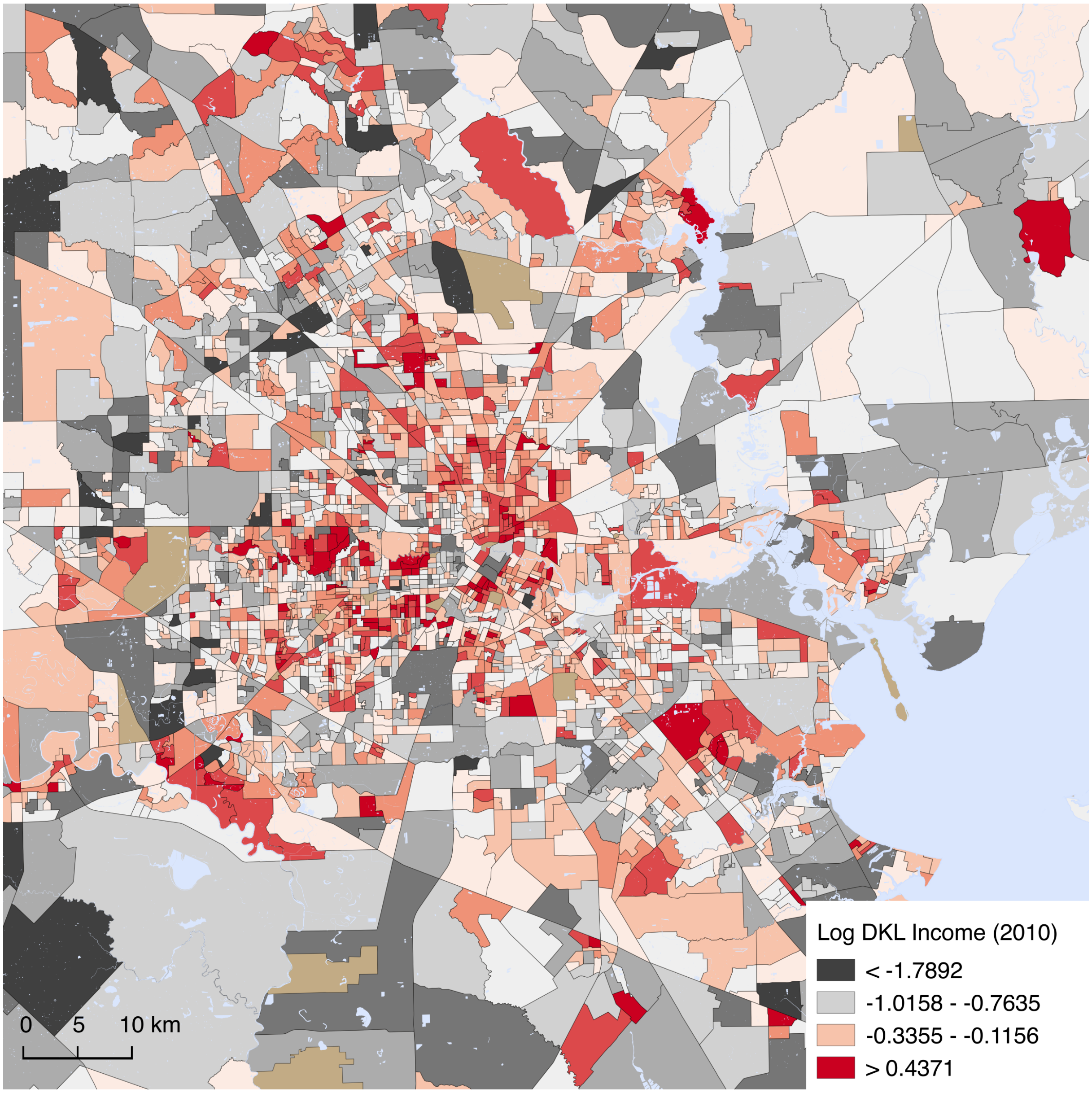

Figure S29: Houston, TX D_{KL} (2010)

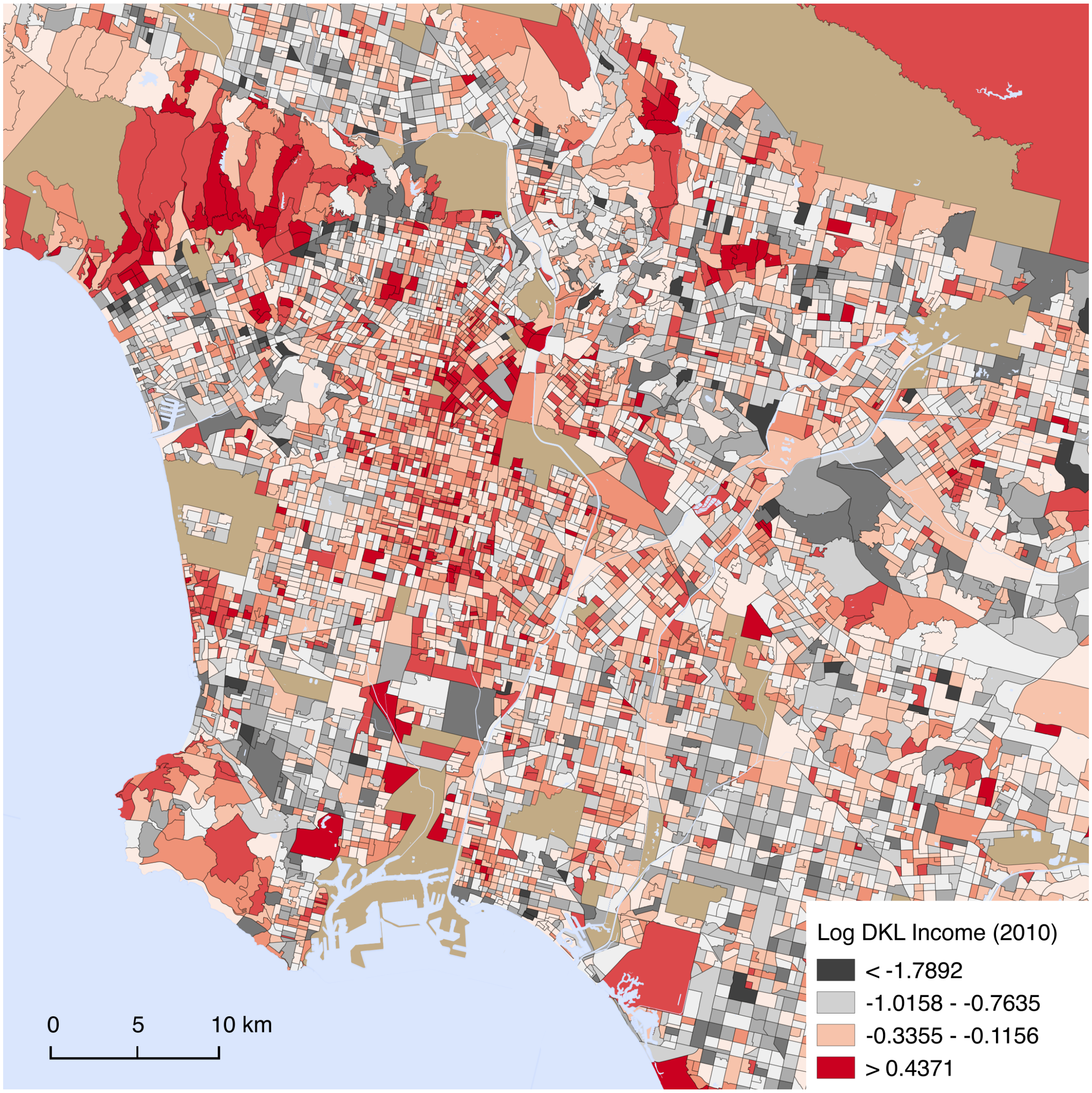

Figure S30: Los Angeles, CA D_{KL} (2010)

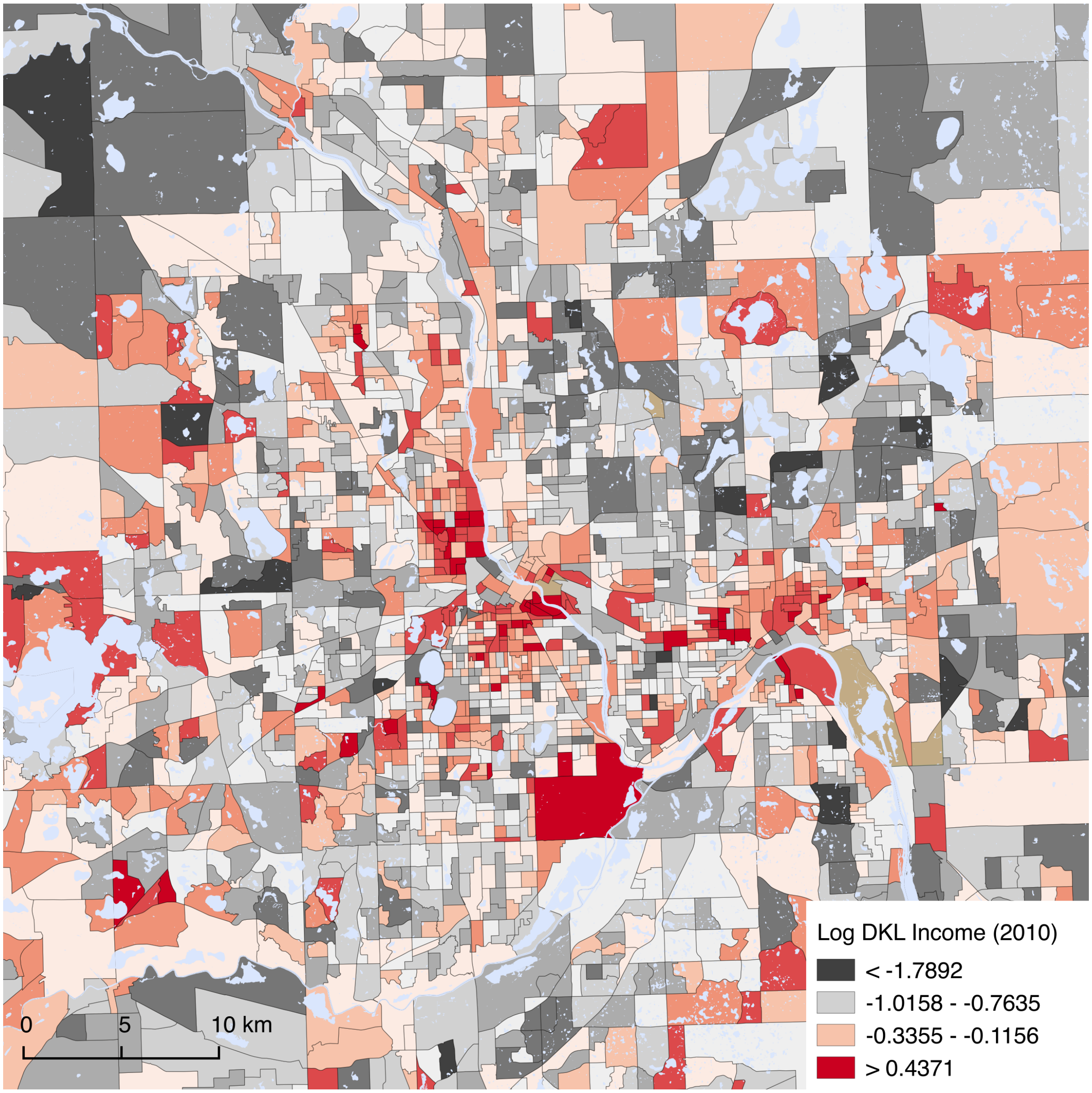

Figure S31: Minneapolis, MN D_{KL} (2010)

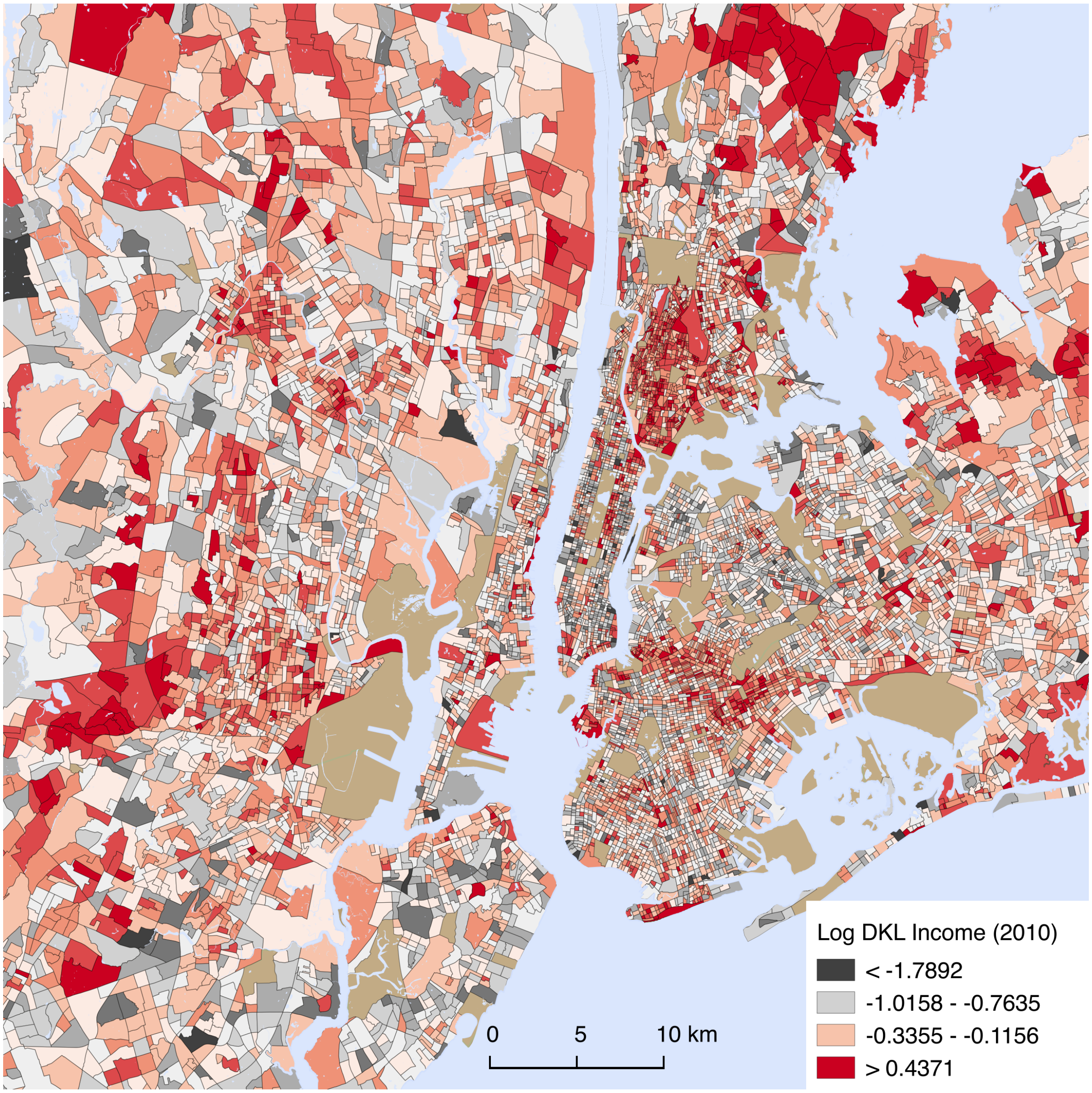

Figure S32: New York City, NY D_{KL} (2010)

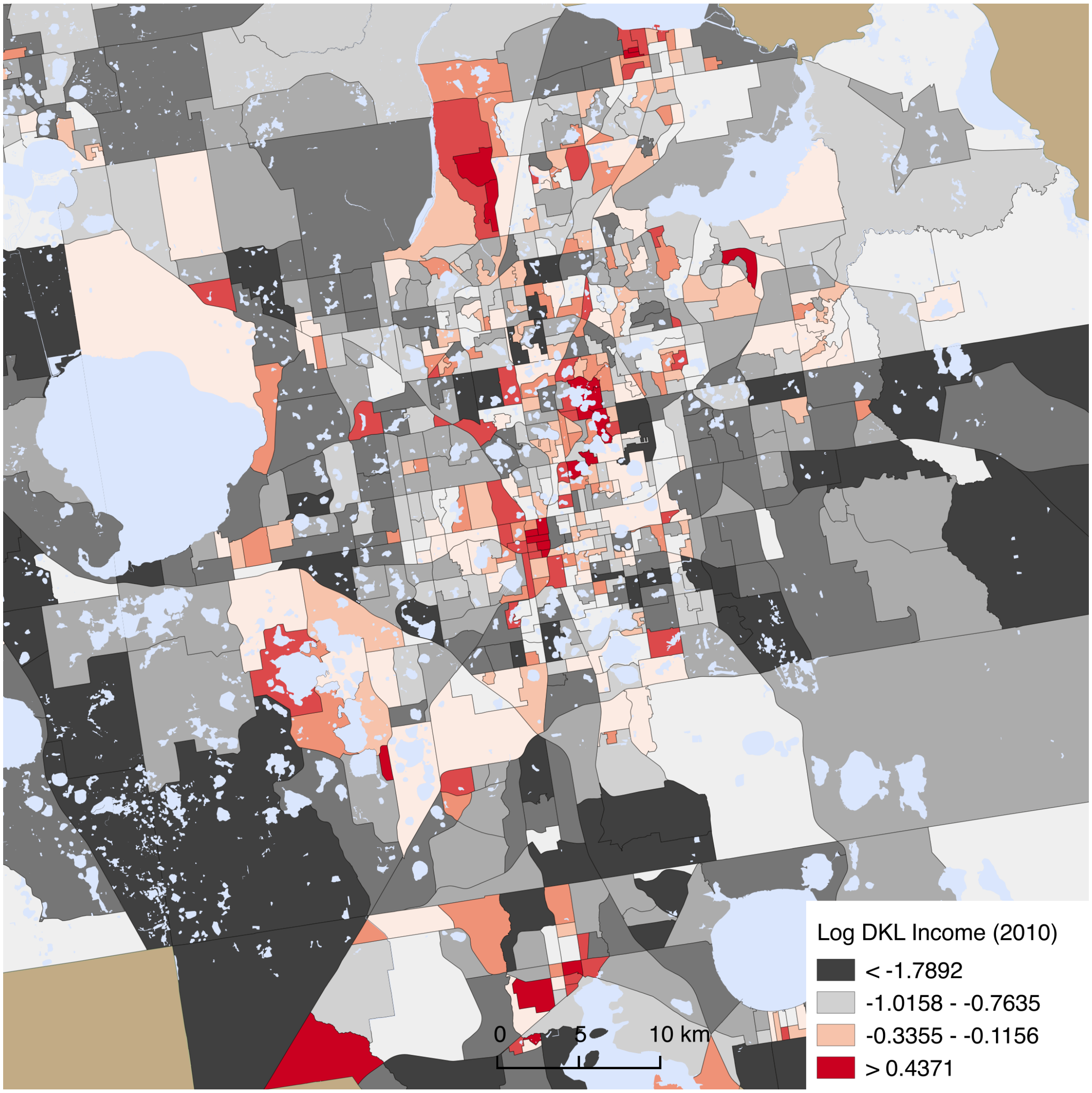

Figure S33: Orlando, FL D_{KL} (2010)

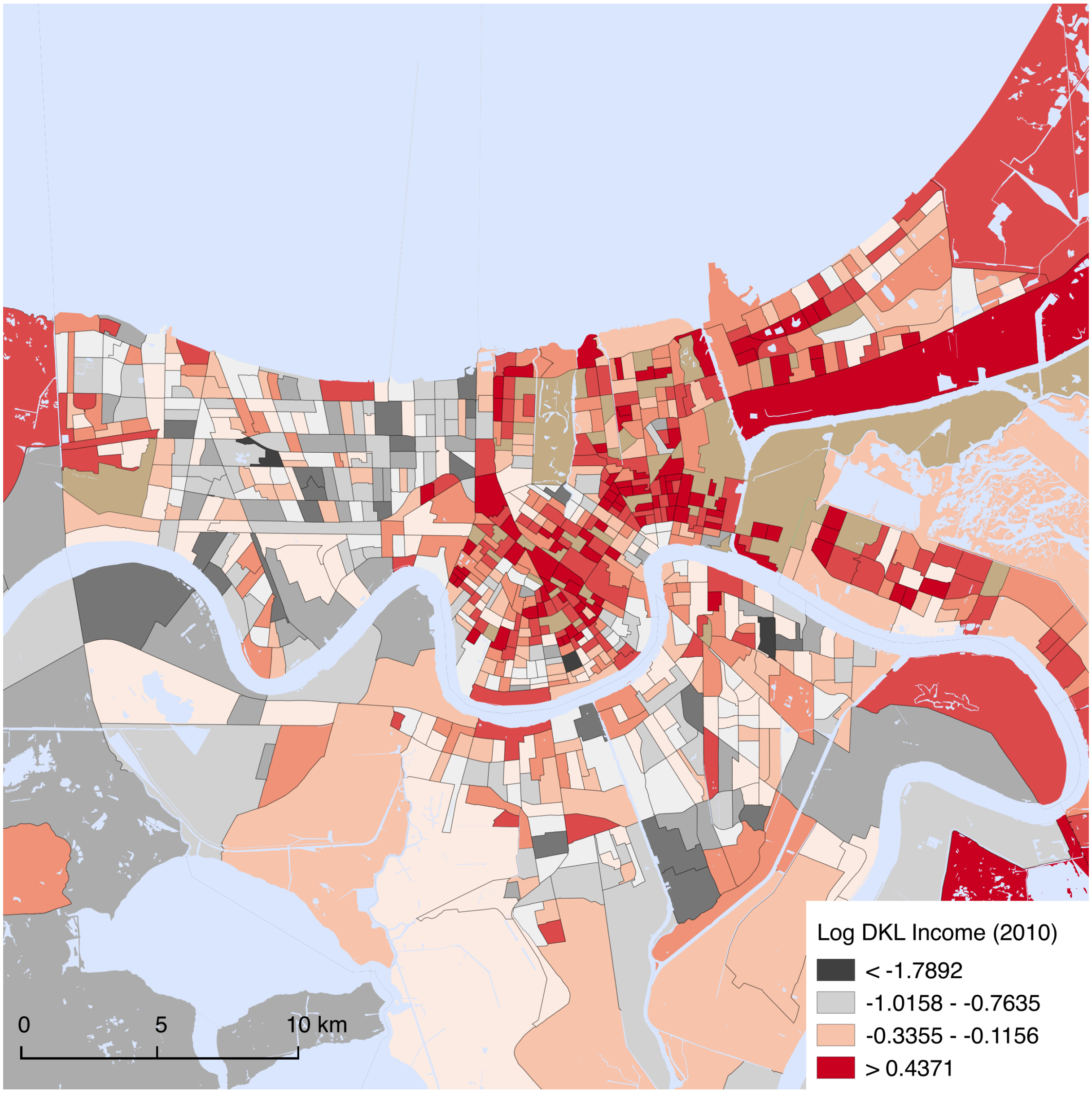

Figure S34: New Orleans, LA D_{KL} (2010)

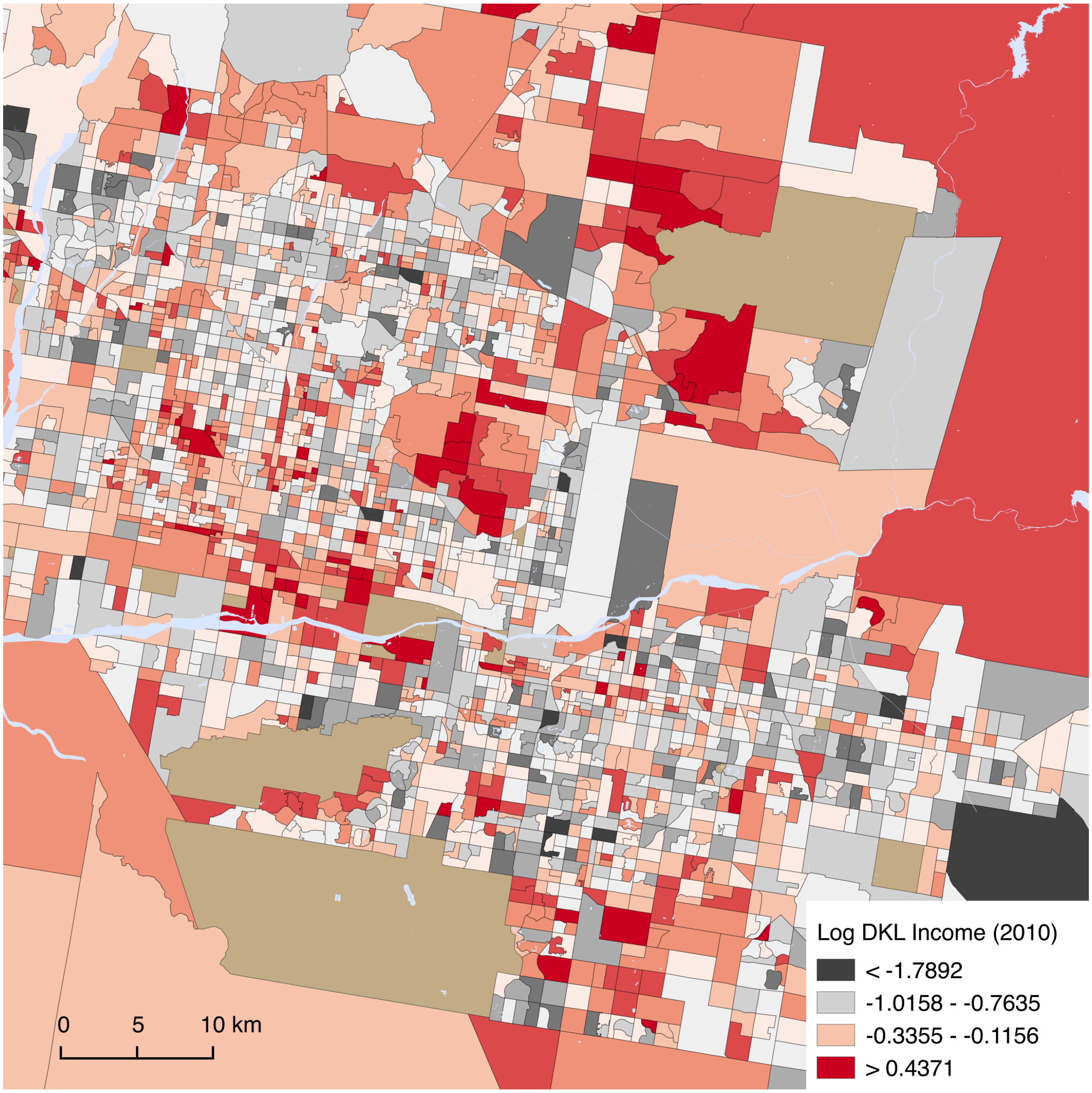

Figure S35: Phoenix, AZ D_{KL} (2010)

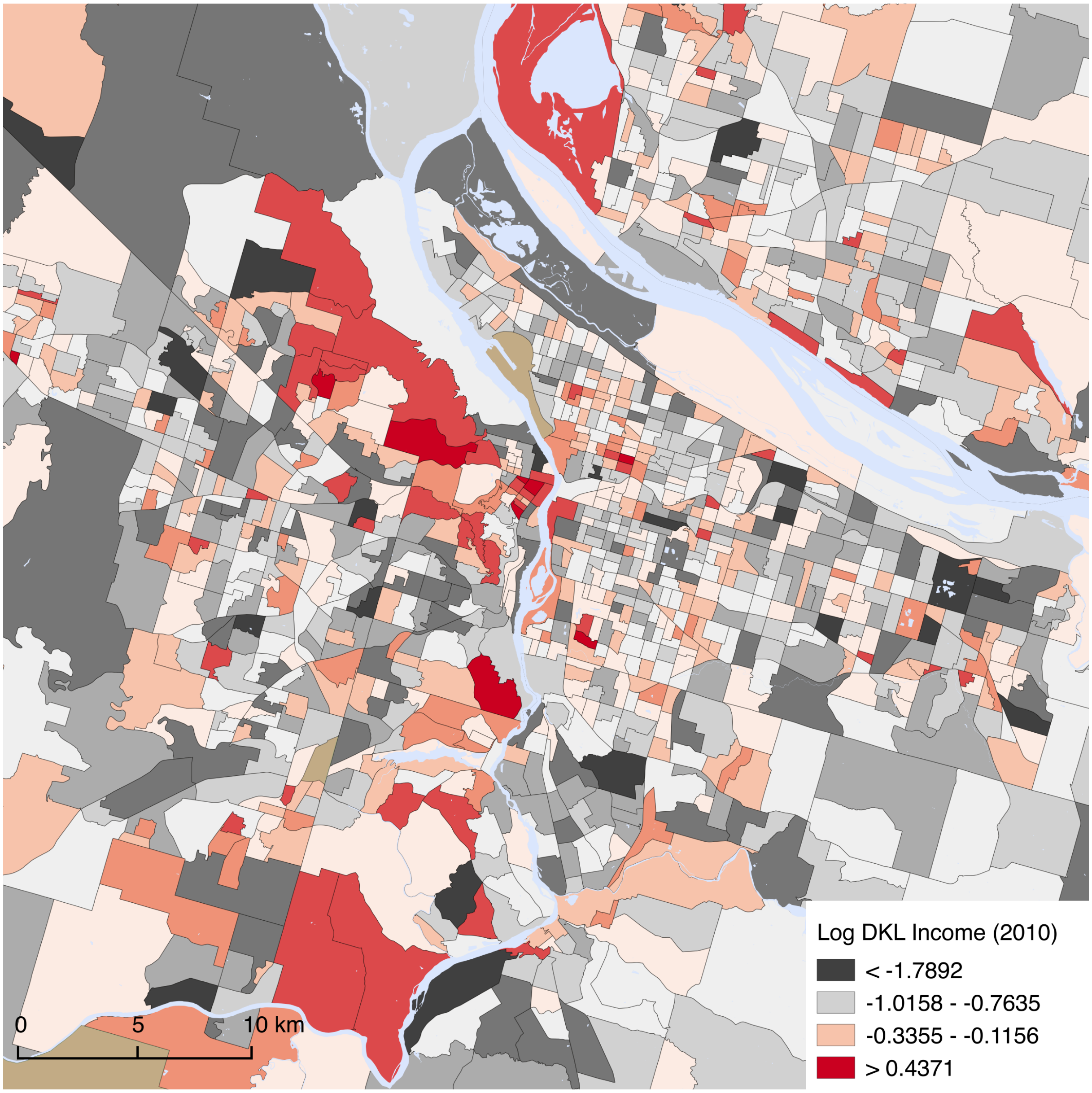

Figure S36: Portland, OR D_{KL} (2010)

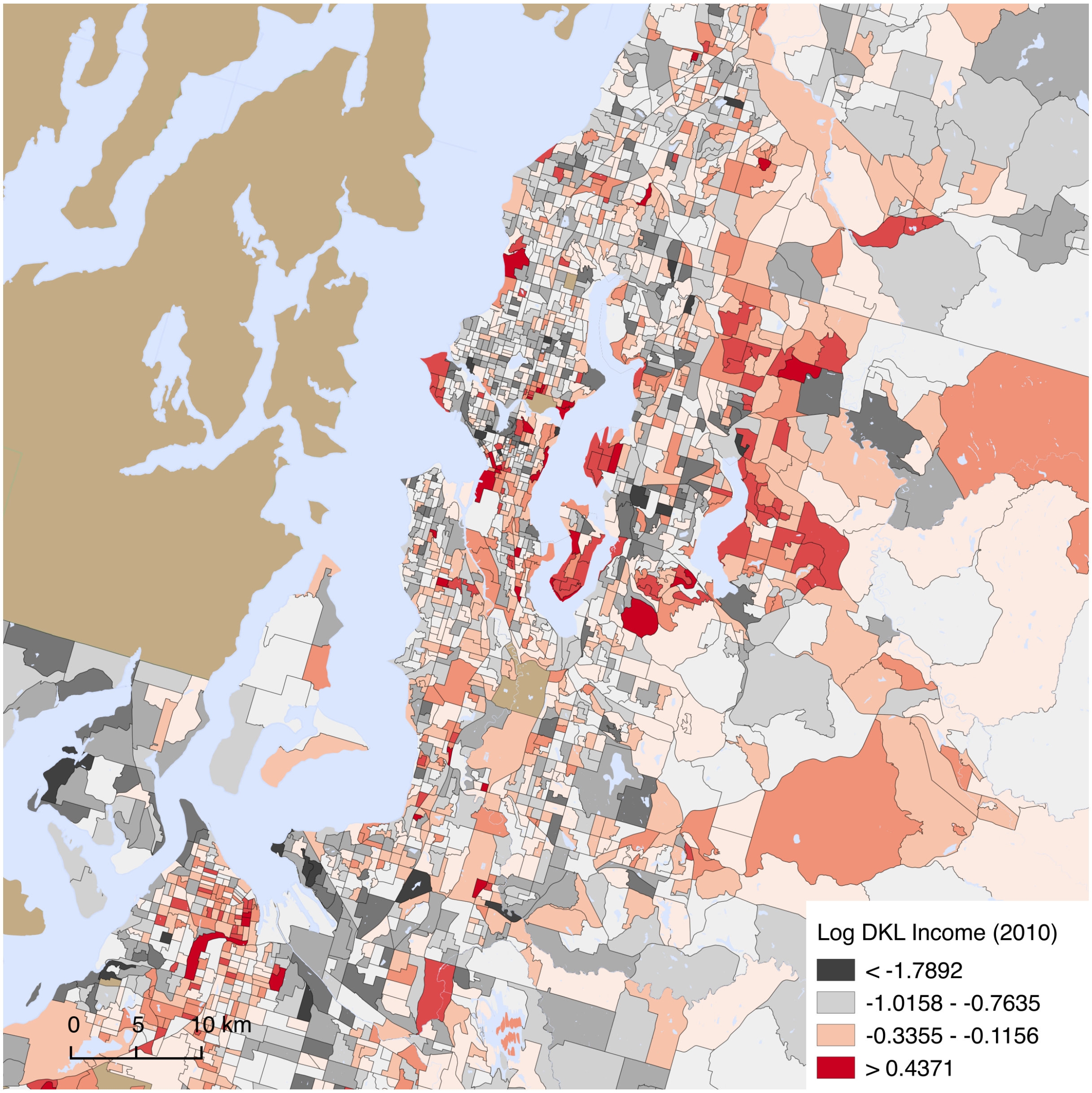

Figure S37: Seattle, WA D_{KL} (2010)

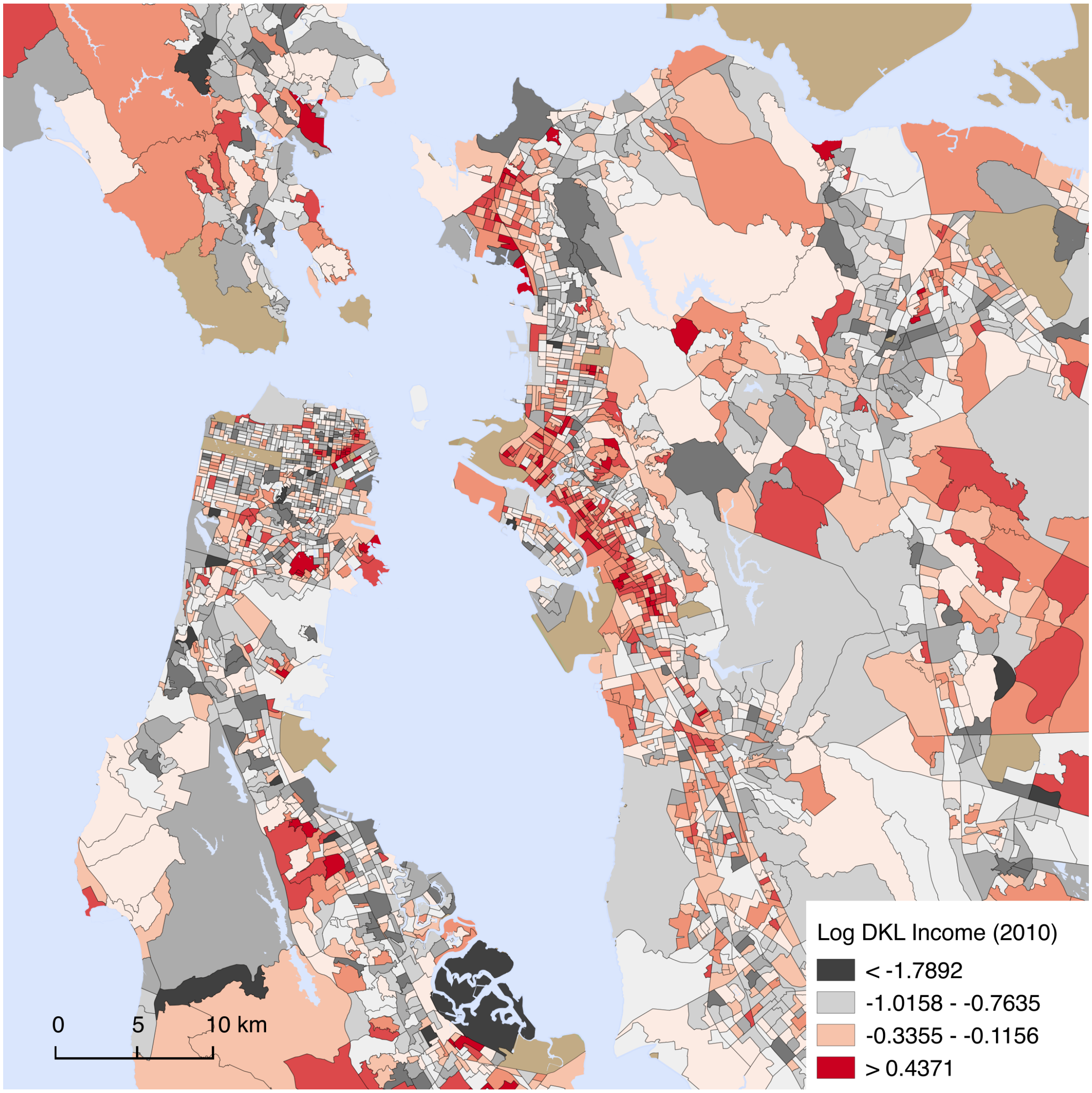

Figure S38: San Francisco, CA D_{KL} (2010)

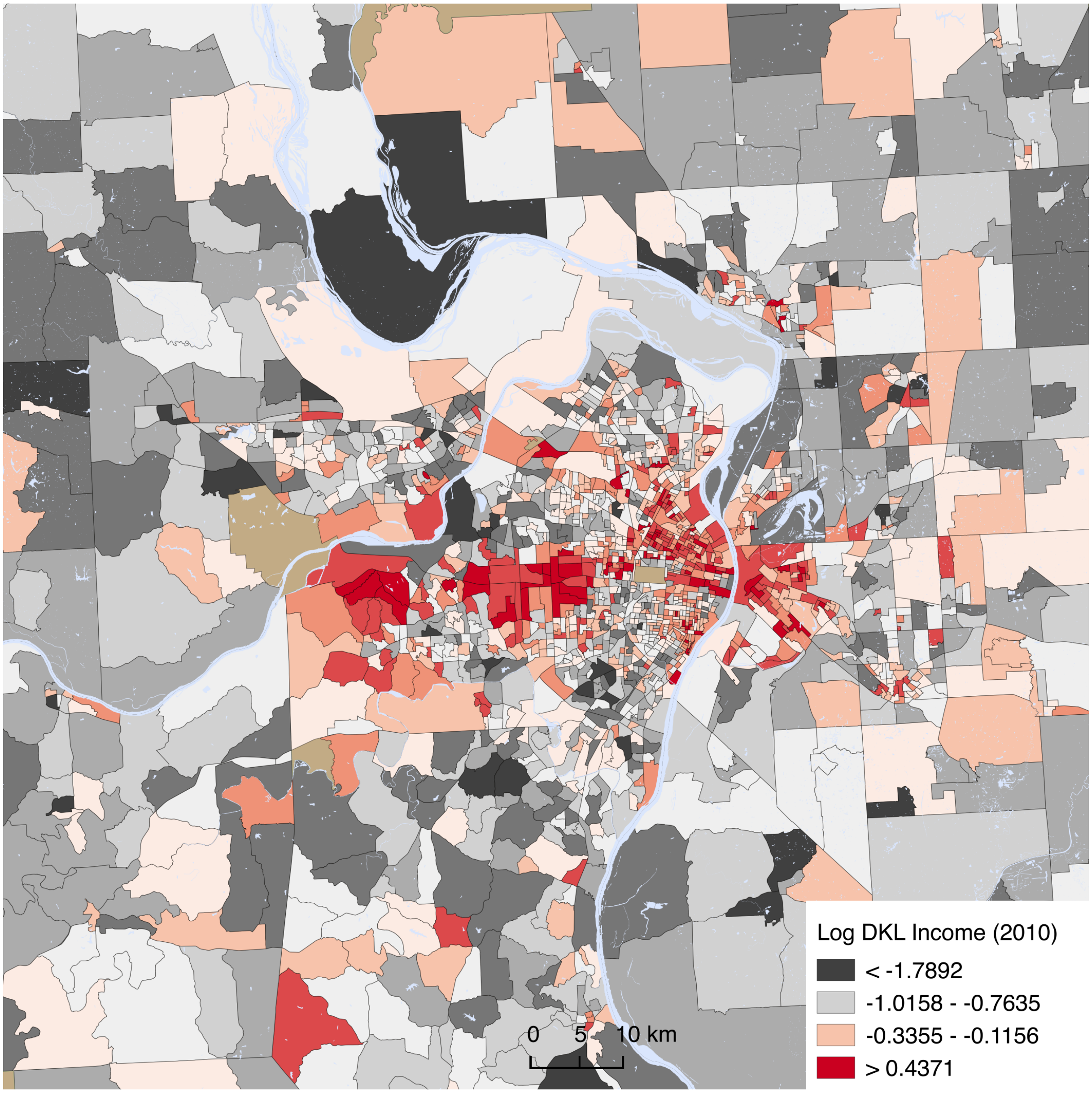

Figure S39: St. Louis, MO D_{KL} (2010)

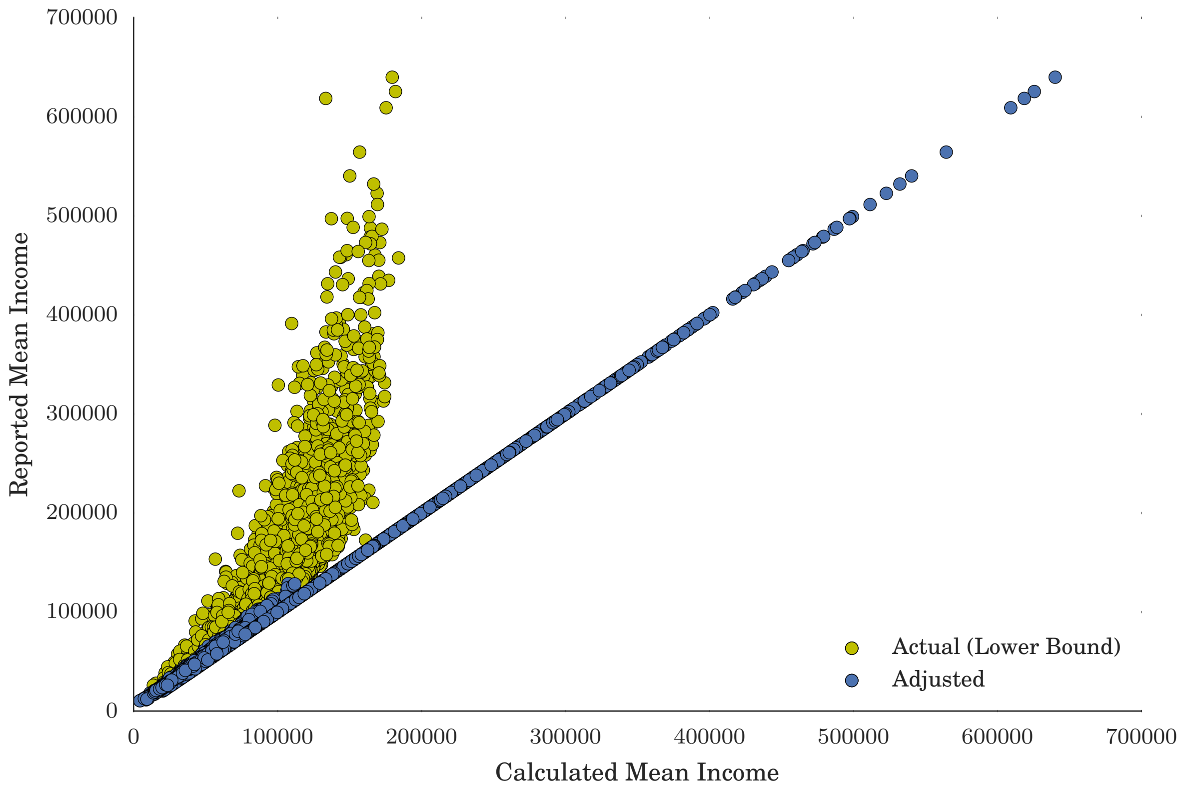

Figure S40: Household income for NYC block groups as reported by the Census and as calculated using the sixteen income bins. Using the lower bound of each income bin, we show the minimum mean income vs the reported incomes, in yellow. Then we increase the value of the top bin (previously $200,000) bin to account for the missing income, in blue.

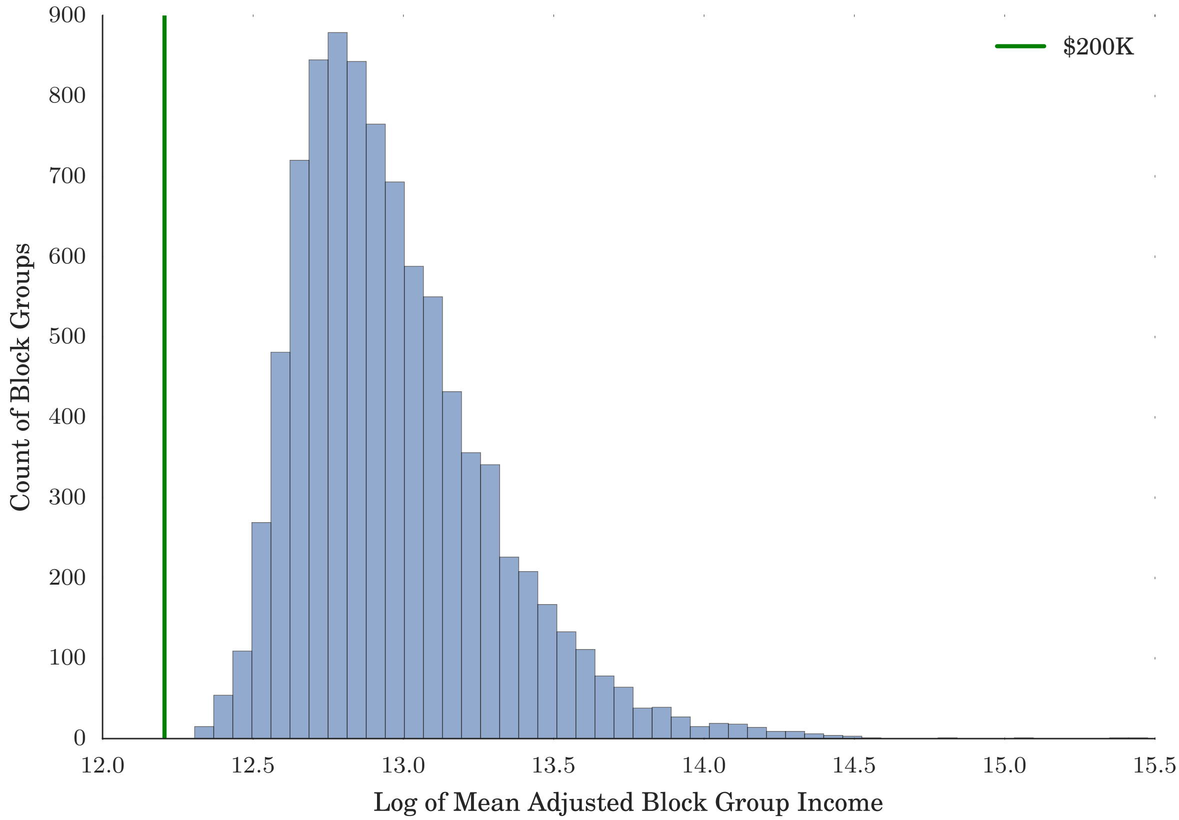

Figure S41: New values for the $200K+ bin used for adjusting the mean value.

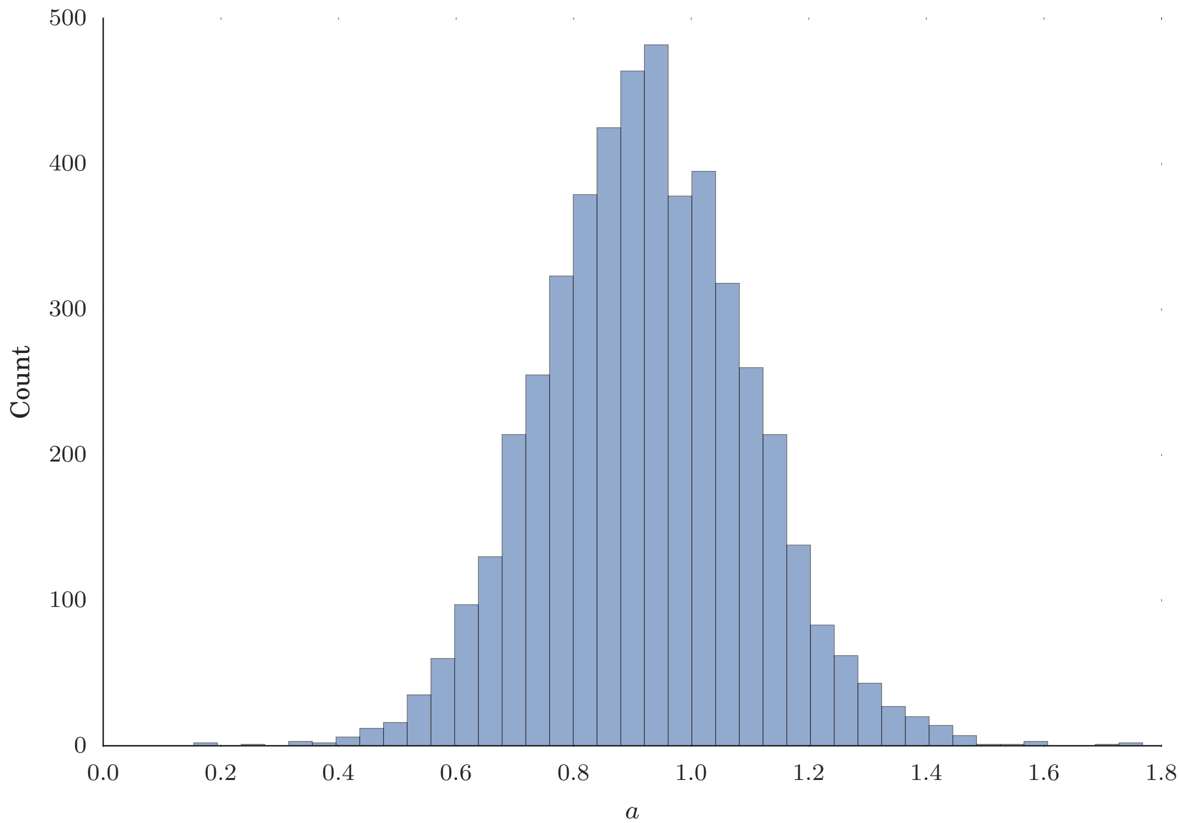

Figure S42: Price a value

References and Notes

31. Information about the American Community Survey 2010 data release (for the 5 years between 2006-2010) are available at `http://www.census.gov/acs/www/data_documentation/2010_release/`.

32. J. R. Hipp, Block, Tract, and Levels of Aggregation: Neighborhood Structure and Crime and Disorder as a Case in Point. *Ame. Soc. Rev.* **72**, 659-680 (1997).

33. E. W. Montroll, M. F. Shlesinger, On 1/f noise and other distributions with long tails. *Proc. Natl. Acad. Sci. U.S.A.* **79**, 3380-3383 (1982).

34. E. Wit, E. van den Heuvel, J. W. Romeijn, All models are wrong...: an introduction to model uncertainty. *Statistica Neerlandica* **66**, 217-236 (2012).

35. S. K. Singh, G. S. Maddala, A function for size distribution of incomes. *Econometrica* **44**, 963-970 (1976).

36. J. Aitchinson, J. A. Brown, *The Lognormal Distribution* (Cambridge Univ. Press, Cambridge, 1957).

37. A. Gomez-Lievano, H. Youn, L. M. A. Bettencourt, The Statistics of Urban Scaling and Their Connection to Zip's Law. *PLoS ONE* **7(7)**, e40393 (2012) `doi:10.1371/journal.pone.0040393`.

38. M. Schläpfer *et al.*, The scaling of human interactions with city size. *J. R. Soc. Interface* **11**, 20130789 (2014) `doi:10.1098/rsif.2013.0789`.

39. P. Wang *et al.*, Understanding Road Usage Patterns in Urban Areas. *Scientific Reports* **2**, Article number: 1001 (2012) `doi:10.1038/srep01001`.

40. S. G. Ortman, A. H. F. Cabaniss, J. O. Sturm, L. M. A. Bettencourt, The Pre-History of Urban Scaling. *PLoS ONE* **9(2)**, e87902 (2014) `doi:10.1371/journal.pone.0087902`.

41. S. G. Ortman *et al.*, Settlement scaling and increasing returns in an ancient society. *Sci. Adv.* **1**, e1400066 (2015) `doi:10.1126/sciadv.1400066`.

42. D. S. Massey, N. A. Denton, The dimensions of residential segregation. *Social Forces* **67**, 281-315 (1988).

43. R. Park, E. W. Burgess, R. D. McKenzie, *The City* (Univ. of Chicago Press, Chicago, 1925).

44. W. J. Wilson, *The Truly Disadvantaged: The Inner City, the Underclass, and Public Policy* (Univ. of Chicago Press, Chicago, 1987).

45. R. J. Sampson, P. Sharkey, Neighborhood Selection and the Social Reproduction of Concentrated Racial Inequality. *Demography* **45**, 1-29 (2008).

46. E. E. Bruch, How population structure shapes neighborhood segregation. *Amer. J. Soc.* **119**, 1221-1278 (2014).

47. C. E. Ross, J. Mirowksy, S. Pribesh, Powerlessness and the amplification of threat: neighborhood disadvantage, disorder, and mistrust, *Ame. Soc. Rev.* **66**, 443-478 (2001).

48. R. J. Sampson, J. D. Morenoff, T. Gannon-Rowley, Assessing neighborhood effects: social processes and new directions in research, *Ann. Rev. Soc.* **28**, 443-478 (2002).

49. S. F. Reardon, K. Bischoff, Income inequality and income segregation, *Amer. J. Soc.* **116**, 1092-1153 (2011).

50. J. Ludwig *et al.*, Neighborhood effects on the long-term well-being of low-income adults, *Science* **337**, 1505-1510 (2012).

51. S. F. Reardon, D. OSullivan, measures of spatial segregation, *Soc. Metho.* **34**, 121-162 (2004).

52. S. F. Reardon, B. Bischoff, ”Growth in the Residential Segregation of Families by Income, 1970-2009” (Russell Sage Foundation, Providence, RI, 2011).

53. T. Piketty, E. Saez, Income inequality in the United States, 1913-1998, *Qua. J. Econ.* **118**, 1-39 (2003).

54. T. Piketty, *Capital in the 21st Century* (Belknap Press, Cambridge, MA, 2014).

55. T. Piketty, E. Saez, Inequality in the long run, *Science* **344**, 838-843 (2014).

56. D. S. Massey, M. J. Fischer, The geography of inequality in the United States, 1950-2000. *Brookings-Wharton Papers on Urban Affairs*, 1-40 (2003).

57. T. Watson, Inequality and the measurement of residential segregation by income in American neighborhoods. *Rev. Inc. Wealth* **55**, 820-844 (2009).

58. D. H. Weinberg, ”U.S. neighborhood income inequality in the 2005-2009 period” (American Community Survey Reports, U.S. Census Bureau, Washington, D.C. 2011).

59. E. Kneebone, C. Nadeau, A. Berube, ”The re-emergence of concentrated poverty: metropolitan trends in the 2000s” (Metropolitan Policy Program, Brookings Institution, Washington, D.C., 2011).

60. R. Fry, P. Taylor, ”The Rise of Residential Segregation by Income” (Pew Research Center, Washington, D.C., 2012).

61. A. Walks, ”Income Inequality and Polarization in Canadas Cities: An Examination and New Form of Measurement” (Research Paper 227, Cities Centre, University of Toronto, 2013).

62. B. Harsman, J. M. Quigley, The spatial segregation of ethnic and demographic groups: comparative evidence from Stockholm and San Francisco. *J. Urb. Econ.* **37**, 1-16 (1995).

63. M. J. Fischer, The relative importance of income and race in determining residential outcomes in U.S. urban areas 1970-2000. *Urban Affairs Review* **38**, 669-696 (2003).

64. E. Talen, Neighborhood-level social diversity: insights from Chicago. *J. Ame. Plan. Asso.* **72**, 431-446 (2006).

65. G. C. Galster, J. C. Booza, J. M. Cutsinger, Income diversity within neighborhoods and very low-income families. *Cityscape* **10**, 257-300 (2008).

66. S. A. Frank, *Foundations of Social Evolution* (Princeton Univ. Press, Princeton, 1998).

67. J. T. Bonner, *The Evolution of Complexity by Means of Natural Selection* (Princeton Univ. Press, Princeton, 1988).

68. J. K. Parrish, L. Edelstein-Keshet, Complexity, pattern, and evolutionary trade-offs in animal aggregation. *Science* **284**, 99-101 (1999).

69. A. M. Turing, The Chemical Basis of Morphogenesis. *Philosophical Transactions of the Royal Society of London B* **237**, 37-72 (1952).

70. P. W. Anderson, More is Different. *Science* **177**, 393-396 (1972).

71. N. Goldenfeld, ”Lectures On Phase Transitions And The Renormalization Group” (Frontiers in Physics, Addison-Wesley, Boston, MA, 1992).

72. G. R. Price, Extension of covariance selection mathematics. *Ann. Hum. Genet.* **35**, 485490 (1972).

73. S. A. Frank, Natural selection. IV. The Price equation, *Journal of Evolutionary Biology* **25**, 10021019 (2012).

74. M. Lynch, B. Walsh, *Genetics and Analysis of Quantitative Characters* (Sinauer Associates, Sunderland, MA, 1998).