FRONT MATTER

Title

- Full title: **Information, Spatial Selection and the Statistics of Neighborhoods**
- Short title: **The Statistics of Neighborhoods**

Authors

L. M. A. Bettencourt,[1,2]* J. Hand,[1] J. Lobo[3]

Affiliations

[1]Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe NM 87501, USA.

[2]Department of Ecology and Evolution, and Mansueto Institute for Urban Innovation, University of Chicago, Chicago IL 60637, USA.

[3]School of Sustainability, Arizona State University, 800 Cady Mall, Tempe, AZ 85281, USA.

* Corresponding Author: bettencourt@santafe.edu

Abstract

Complex systems are recognizable by the coexistence of general macroscopic patterns emerging out of a rich structure of local variations. This poses a challenge for statistical analyses and theoretical descriptions of these systems, which tend to be either too aggregated or too local, depending on the disciplinary perspective. Using mathematical methods from evolutionary biology and statistical learning, we develop a framework for quantifying how much information is encoded in the structure of complex systems at different scales. We illustrate this approach through the analysis of the neighborhood structure of over 900 US metropolitan areas, which display strong spatial variation in household income within a simple city-wide income distribution. By treating the formation of neighborhood structure as a process of spatial selection, we quantify the complexity of explanation needed to account for the local economic heterogeneity observed across all US urban areas and each of its neighborhoods. We find that spatial selection is, in general, income group dependent with richer and poorer households appearing spatially more segregated than middle-income groups. These findings emphasize the versatility and power of information theoretic quantities to characterize problems of diversity, inequality and spatial sorting in a wide range of complex systems across scales.

One Sentence Summary: Complex patterns in any system can be read in terms of their relative information providing a bridge between sciences studying ecosystems and human societies at different scales.

MAIN TEXT

Introduction

Complex systems – such as cities, ecosystems or biological organisms - are often recognizable by the presence of *structure with variations* (*1*). Such a classical description is deceptively simple though, as it glosses over the fact that local structures in a patch of forest or on a street of a great city represent not just mere random fluctuations but a long cumulative history of serendipity and adaptation (*2–4*). This distinction is central to any theoretical approach to complex systems: How to account for mechanisms of change and adaptation not only on large scales, but also locally? What is chance and what is necessity (*5*)?

The point of the present paper is to develop a systematic approach to the analysis of how individual agent characteristics (traits) form patterns in complex systems across scales. We will use data from US cities to motivate and illustrate our results, but the approach is general and applicable to any heterogeneous population where individual traits are sorted differentially into groups.

Arguably, the main conceptual obstacle to building theories of complex systems at different scales is clearly patent in the distinct approaches to the same problem developed by different disciplines (*6*). For example, in order to model macroscopic regularities, statistical physics emphasizes the behavior of averaged quantities, over large populations and spatiotemporal scales (*7–9*). In other words, it is in the "thermodynamic limit" that the strongest general statements can be made about the macroscopic behavior of a system (*9*). Short of this situation the formalism still applies, but results become more contingent (*10*).

But this and other "large-system" limits are much less productive of insights in many other important situations. For example, they are anathema to natural historians, anthropologists or ethnographers, who study primarily the detailed behavior of organisms and people in small groups and specific natural and social contexts.

Can these macro and microscopic approaches be reconciled? What formal methodologies allow us to study the same system –say, a city or an ecosystem – and systematically bridge these two perspectives?

To answer this question – but also pose it in a more formal light – we propose an approach based on information (*11*).

In practice, there are two strategies commonly used to tackle the problem of studying complex systems across scales: We can (i) start from variable local patterns and average them to obtain more *coarse-grained* descriptions of the system, usually for larger populations and on longer spatial or temporal scales (*12*, *13*). Or (ii) we can proceed in the opposite direction and generate detailed local patterns from more aggregated statistical theories (*14*, *15*). Historically these two approaches have been developed separately, in distinct disciplines and motivated by different questions. Nevertheless, every discipline, from physics to economics, seeks generative connections between "micromotives" and "macrobehaviors" (*9*, *16–18*).

Each point of departure has its clear advantages. Approaching complex systems from the "bottom-up" emphasizes that local variations matter because they contain critical information about how agents operate within systems (*16*, *19*, *20*). This includes issues of choice and agency in human societies, as well as adaptation, accident and serendipity in all types of system. This information should not, in general, be averaged over as is typically done in statistical physics (*8*, *12*, *13*) or in economic models that assume representative agent behavior (*21*). This is not only a point of principle, it is a necessity for building general theories of change, for example reflecting patterns of interaction between agents (*22*, *23*) or energy and resource flows in networks (*24–26*), which in turn feed evolutionary and adaptive dynamics.

The counterpoint to this position, favoring a "top-down" view that starts with large average aggregates, is well supported by the common empirical observation of simpler statistical regularities and the associated resounding successes of creating predictive science to derive them, especially in statistical physics (*9*, *13*). This "coarse-graining" approach tells us specifically that, in many known systems, most local details do not contribute to macroscopic behavior (*12*, *13*). This is the basis for the renormalization group (RG) in statistical physics, which constitutes the essential tool to analyze phase-transitions in bulk materials and has resulted in powerful ideas of universality (*8*, *12*, *13*).

It follows that proceeding in the direction of coarse-graining leads to *information loss* as local states are replaced by averages over larger scales (*8*), see Supplementary Text 2.1 for formal discussion. Because of this essential feature, RG methods are not invertible. Conversely, proceeding in the opposite direction ("fine-graining") requires that we specify *more information* as new degrees of freedom on finer scales are considered (*15*, *27*). In different but parallel ways, such "fine-graining" methods have been developed in statistical learning theory and inference (*28*, *29*) and in evolutionary biology (*30*, *31*) and are also part of powerful generative statistical models applied to signal reconstruction (*27*, *32*).

These general considerations tell us that we can navigate theories of complex systems at different scales by keeping track of their information content. Below, we show how this insight can be used as a general unifying strategy for specifying different theories of complex systems (specifically, cities) at distinct scales, and how these must transmute into each other as we adjust the magnification of our lens.

In pursuing this approach, we were motivated by attempting to account for some of the complexity of human behavior in cities, and specifically the observed patterns of spatial sorting of households into neighborhoods by personal income (*23*, *33*–*37*). The differential sorting of households into urban neighborhoods, whether by income, race or any other characteristic, is a classic problem in sociology and economics (*16*, *23*), approached originally by showing that seemingly innocuous local decisions can lead to extreme macroscopic patterns of segregation (*16*, *19*). We will show, however, that observed patterns of economic sorting in US metropolitan areas are, to a large extent, place and income group specific and therefore require different decision models to explain them in practice (*19*, *38*).

Results

Information and Spatial Selection

As an illustration of the problem, consider the complex pattern of household income in New York City neighborhoods (Figure 1A); see Figs. S1-S17 for other cities. 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 (generally rather poor), Fig. 1B. As is well known, this spatial heterogeneity is long-lived, persisting for decades or longer, through many economic cycles and substantial demographic turnover: thus, these patterns are not the result of accidental fluctuations, but rather of decades of choice and adaptation by successive waves of individuals.

Such rich and detailed patterns contrasts with the simple normal distribution for (the logarithm of) household income across the entire metropolitan area, Fig. 1C. This coarse-grained statistic is common to all US metropolitan areas: the distribution of income across all cities is well described by a lognormal (except at the top tail, which is censored in this data), see Supplementary Text S2.2 and Figures S18-S19. Moreover, the two parameters characterizing this distribution are themselves simple and general, see Supplementary Text 2.3 for background. The mean obeys a scaling relation (*26*, *39*), well parameterized by a power-law of the form $\langle Y(N,t)|N\rangle = Y_0(t)N^{\beta}(t)$, see Fig. S20, with general system-wide parameters $Y_o(t), \beta$ (*18*); where the exponent $\beta > 1$, expresses urban agglomeration effects (*21*, *26*). Its value is predictable from urban scaling theory, which describes the city in terms of interdependent networks of people, organizations and infrastructure (*26*). The variance of the logarithmic income, see Fig. S21, is also a simple general number, associated with the cumulative volatility of incomes over time and setting the level of inequality within the urban area, as measured, for example, by the Gini coefficient or Theil's *T* index.

Thus, a seemingly "universal" statistical regularity, Fig. 1C, emerges at the city-wide scale as the result of averaging over a rich pattern of local neighborhood variations, Fig. 1A-B. While this coarse-grained statistic is interesting and reveals important aspects of urban socioeconomic dynamics (*40–42*) REF, we now focus on the structure of the neighborhood variations using this macroscopic regularity as a reference. Specifically, we quantify the complexity of the pattern of variations at the neighborhood level by comparing income probability distributions at different levels of spatial aggregation.

To do this explicitly, we write

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

where $p(y_\ell | n_j)$ is the distribution (normalized share) of income *y*, in discrete bins labeled by ℓ, *see* Supplementary Materials S1.4, in neighborhood n_j (the different colorful patches in Fig 1A), and $p(y_\ell)$ is the income distribution at a more aggregate level (Fig. 1C), which we take to be metropolitan area throughout this paper. Eq. 1 *defines* the weights, $w_{\ell j} \equiv p(y_\ell | n_j)/p(y_\ell)$, which transform one distribution into the other, see Fig. S22. With this definition, the average weights over income obey $\langle w_j \rangle = \sum_\ell w_{\ell j} p(y_\ell)$=1, for all neighborhoods *j*.

Eq. 1 is well known and can be readily recognized from two different perspectives. First, it is the haploid model of population genetics (*43*), also known as the "replicator" equation in evolutionary game theory (*44*). In that context, the two distributions are related across time (not space) and the weights $w_{\ell j}$ are the *fitness* of a trait (allele) ℓ, expressing its differential propagation into the next generation over the period *j*. The stronger the deviation of $w_{\ell j}$ away from the average (unity), the stronger the selection for allele ℓ. This corresponds to high fitness if $w_{\ell j} > 1$, and vice-versa if $w_{\ell j} < 1$. When $w_{\ell j} = 1$, the dynamics is *neutral*, and there is no selection over the specific time period (generation) indexed by j. This interpretation gives a mathematical correspondence between genetic evolutionary dynamics (in time) and neighborhood sorting (in space).

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

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

Here, $p(n_j|y_\ell)$ is the probability for a person in the city to reside in neighborhood n_j, given that they have income y_ℓ, while $p(n_j)$ is the (income independent) probability to live in neighborhood n_j, see Supplementary Material S1.4), estimated as the ratio of its population to that of the city, see Supplementary Text. This second perspective leads to another powerful correspondence between probability theory, inference, and neighborhood structure in complex systems. In this context, $\log w_{\ell j}$, in Eq. 2 is the (non-averaged) Shannon *mutual information* (*11*, *45*) between neighborhood j and the distribution of income y. To see this more explicitly consider the average of $\log w_{\ell j}$ over income groups

$$\langle \log w_j \rangle = \sum_\ell p(y_\ell|n_j) \log \frac{p(y_\ell|n_j)}{p(y_\ell)} = D_{KL}[p(y|n_j)||p(y)]. \quad (3)$$

This is the Kullback-Leibler divergence, D_{KL}, between the distributions of income city-wide and in neighborhood j. The Kullback-Leibler divergence is a fundamental quantity in information theory from which many other important quantities can be derived (*11*, *46*, *47*). For each neighborhood j, 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, from this perspective, with income distributions very different from the city as a whole, will require a longer explanation (more information), whereas neighborhoods that already reflect the city-wide pattern require no further description. In other words, atypical neighborhoods require local theories of neighborhood effects, in addition to an explanation of the city wide distribution of traits, such as Fig. 1C. Thus, $\langle \log w_j \rangle$ expresses the strength of neighborhood (income) effects (*23*, *34*, *48*, *49*) in each neighborhood j relative to city-wide dynamics, measured in units of information (*11*, *45*, *47*).

Fig. 2A shows the strength of selection by income for each neighborhood across New York City, measured by $\langle \log w_j \rangle$; see Figs. S23-S39 for 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 (dark grey), but also with a significant fraction of others manifesting primarily a strong local flavor (red). We verified that the magnitude of the observed differences could not be the result of a purely random process of sorting, by building maps analogous to Fig. 2A using the same corresponding population sizes in the city and neighborhoods, but drawing at random individuals from the metropolitan income distribution into each place, see Fig S40.

Comparing Figs. 1A and 2A suggests 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 metropolitan areas that we can quantify systematically via the average of $\log w_{\ell j}$ over neighborhoods j,

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

This quantity is the average information necessary to explain the distribution of specific income ranges y_ℓ across the city, given that we know its neighborhood structure. In the absence of neighborhood effects (i.e. of spatial sorting), this quantity is zero. Thus, its magnitude quantifies the differential average strength of *neighborhood effects for different income levels* in each city. Fig. 2B shows that the neighborhood effects are strongest for the highest income group, followed by the lowest; Middle-incomes are observed to be spatially the most mixed and thus less determined by specific neighborhoods. This is an interesting finding because it shows that different income groups exercise different kinds of choices – by preference or necessity - in terms of residential location. Thus, any realistic model of residential choice in US cities needs to be an explicit function of income levels.

These two effects are summarized in turn by a single fundamental quantity that captures the overall strength of neighborhood effects for each city in units of information, Fig. 2C. This is the total (mutual) information, $I(y; n) = \langle \log w \rangle$, between neighborhood structure and y, given as the average of the previous quantities over the remaining variable,

$$\langle \log w \rangle = \sum_j p(n_j) D_{KL}[p(y|n_j)||p(y)] = \sum_\ell p(y_\ell) D_{KL}[p(n|y_\ell)||p(n)]$$

$$= \sum_{\ell,j} p\left(y_\ell, n_j\right) \log w_{\ell j} = I(y;n). \quad (5)$$

If every neighborhood were a microcosm of the city as a whole, then all income groups would be spatially well-mixed and there would be no (income) neighborhood effects, leading to $I(y;n) = 0$. Conversely, in cities where every neighborhood has its own unique flavor, not at all like the distribution of traits across the city, there is strong sorting of incomes by neighborhood and $I(y;n)$ will be large. How large depends on the relative amount of information needed to describe the system at the local level, Fig 1A, versus as a whole, Fig. 1C. Thus, the mutual information $I(y;n)$ gives a measure of how well a coarse-grained pattern describes a complex system observed at a more disaggregated level. In other words, $I(y,n)$ quantifies the average complexity of any theory of *local* neighborhood effects versus a theory of the same quantity at the metropolitan level, in this case connecting urban neighborhoods with income distribution.

The top and bottom ranked metropolitan areas in the US by the magnitude of $I(y,n)$ are shown in Tables S1-S4. We see that Dallas TX, followed by New York City and New Orleans, LA show the highest $I(y,n)$ and that many cities of Texas show in general strong income segregation by neighborhood. This is particularly interesting because these cities are currently among the fastest growing in the nation so that at least some of the observed income segregation is the result of recent residential choices. Smaller cities, especially in parts of the Midwest (e.g. Wisconsin) but also in other states, show the lowest neighborhood segregation by income.

Below we will show how informational quantities are also well suited to quantify issues of inequality and how these quantities have the special property of being recursive across scales, generalizing the two-level decomposition illustrated above to a larger number of sorting steps.

Connections to Models of Income Inequality and the Aggregation Problem

We have demonstrated so far how information theoretic quantities arise naturally to describe the strength of local sorting effects, through the comparison between (income) distributions at different levels of spatial and population aggregation.

A similar approach aimed at characterizing economic *inequality* is commonly used in econometrics (*45*, *50*). Seeking to adopt the desirable properties of information theoretic quantities in the characterization of economic inequality, Henri Theil (*45*) proposed an index, *T*, which is defined analogously to Eqs (3-4), see Supplementary Text S2.4,

$$T\,(y) = \sum_{i=1}^{N} \frac{y_i}{N\,\bar{y}} \log \frac{y_i}{\bar{y}} = \sum_{\ell}\; q(y_\ell) \log \frac{q(y_\ell)}{p(y_\ell)} = \; D_{KL}[q(y)||p(y)]. \tag{6}$$

The Theil *T* index compares the income share, $q(y_\ell) = \frac{y_\ell}{\bar{y}} p(y_\ell)$ to the corresponding population share, $p(y_\ell)$, of a population structured into income bins labeled by ℓ and where $\bar{y}$ is the average income in the total population. The first expression in Eq. (6), which is the most common definition of *T*, is obtained when we take groups, ℓ, to become single individuals, i, in a population of size *N*. The second equality refers to arbitrary finite groups and makes *T*'s informational interpretation explicit, see also Supplementary Text Section 2.1 for a more detailed discussion.

Both *T* and the spatial information theoretic quantities in Eq. (3-5) have interesting and important properties under (dis)aggregation, which, in fact, motivated their original use for the study of inequality (*45*, *50*). If we disaggregate the population into a number of groups, n_j, the Theil index can be decomposed into two terms as

$$T(y) = T(n) + \sum_j p(n_j)\, T(y|n_j), \tag{7}$$

see Supplementary Text S2.4 for detailed derivation. This expression shows how the total inequality in a population can be expressed as the *inequality across groups*, n_j, plus the average *inequality within each group,* weighted by that group's relative size.

The information quantities defined above have a similar property that extents their definitions to an arbitrary number of (dis)aggregation levels. For example, if we take the groups n_j to be further decomposed into groups m_i, we can define divergences analogous to the ones above and a generalized mutual information $I(y; m, n)$, which obeys the multi-level relation, see Supplementary Text S2.4 for derivation,

$$I(y; m, n) = \; I(y; n) + \; \sum_j \; p(n_j) I(y; m|n_j), \tag{8}$$

which shows that a total information in a pattern is that contained in the first level of disaggregation plus the average information contained in each of its subgroups.

We see therefore that information quantities provide a systematic and recursive way to characterize both inequality and spatial selection in structured heterogeneous populations with given traits (*50*).

Neighborhood Choice and the Spatial Price Equation

The problem of neighborhood choice and the associated observation of strong spatial segregation as a(n unintended) consequence of individual preferences is a classic problem in sociology and spatial economics, especially in the context of American cities. Thomas Shelling famously proposed a simple stylized model of individual choice to illustrate how personal "micromotives" can create "macrodynamics" at the neighborhood and city levels (*16*). His original model was very simple and far from realistic (*19*, *23*, *33*), but it had the virtue of showing that a population divided into non-overlapping racial groups would necessarily become strongly racially segregated as a result of individuals moving between neighborhoods so as not to become part of the local racial minority.

Since then, many authors have analyzed the assumptions of this sort of model and worked towards adding greater realism to neighborhood choices, as observed for example in population surveys. Bruch and Mare (*19*), in particular, conducted a systematic analysis of the consequences of different choice models and clearly concluded that more heterogeneous and gradual residential preferences lead to very different spatial outcomes from Shelling's original findings, and that much more mixed neighborhoods can result from even a minimal relaxing of the original model specifications.

Though these results refer to discrete choice patterns associated with traits such as race and ethnicity, if is interesting to compare them to the perspective following from our results above. First, patterns of spatial sorting by income are in general neighborhood dependent (Fig. 2A), so that there is no model of choice that fits all local cases. Second, we see in addition, when we average over neighborhoods, that choices (or at least outcomes) are income group dependent, Fig. 2B. This means that richest and poorest households are more spatially segregated than middle-income groups (*49*, *51–53*). This points to different motives for the spatial concentration of different groups at the neighborhood level, with high-income groups presumably exerting their economic prerogative to concentrate into the same neighborhoods, while poorest households possibly have no choice but to live where rents are low (*38*, *49*, *54*). It is poignant, as a result, that most of the space of US cities (red in Figs. 1A, S1-17) is occupied by poor and typically strongly economically segregated neighborhoods.

It is also in this light that the behavior of middle-income groups is most interesting, showing the least amount of spatial sorting with corresponding individuals more likely to be present in both rich and poor neighborhoods. By the same token, it is interesting to stress that neighborhoods that concentrate only middle-income groups are rare. Thus, we believe that these and related results (*49*, *51–53*, *55*), should motivate more realistic models of residential choice where both contingencies and preferences become dependent on place, population traits and their levels of expression, as emphasized by Bruch and Mare (*19*).

We end by showing how, given models of residential preference that do depend on each neighborhood and each income group specified by the weights $w_{\ell i}$, we can predict complex patterns of income. This then generates the mosaic structures in Fig. 1A, where only the average neighborhood income is given.

To do this in practice, we evaluate the deviations in each local patch j versus the system as whole. Consider the deviation in the average of some function of local characteristics, $z(y)$, for each neighborhood $\bar{z}_j = \sum_j z(y_\ell)\, p(y_\ell|n_j)$ from that for the entire system $\bar{z} = \sum_\ell z(y_\ell)\, p(y_\ell)$. Using Eq. 1, we write $\Delta\bar{z}_j = \bar{z}_j - \bar{z}$, as

$$\Delta\bar{z}_j = \sum_\ell z(y_\ell)\left(w_{\ell j} - 1\right)p(y_\ell) = \mathrm{covar}(w_j, z). \quad (7)$$

This is the famous Price equation (*47*, *56*), here expressing spatial selection, see Supplementary Text 2.5. The simplest instance of this relations is for $z(y_\ell) = y_\ell$, which derives the neighborhood patterns of Fig. 1A; see Supplementary Text 2.5 and Figs. S41-S43. Any other function of neighborhood characteristics, $z = f(y)$ besides the mean, can also be computed using the spatial Price equation (*47*).

We conclude then that the specification of place and trait-dependent models of neighborhood choice is equivalent to the knowledge of the information weights, $\log w_{\ell j}$.

Discussion and Outlook

In summary, we showed how complex patterns of population traits can arise from more aggregate general distributions in meta-populations through processes of spatial selection. We also demonstrated how the type and strength of these patterns are most naturally measured in terms of information, a fundamental quantity that provides a unifying common language across all complex systems.

Selection is a general process by which individuals learn and adapt to their environment by acquiring information (*44*, *47*, *57*, *58*). As Price emphasized when proposing his eponymous equation (*56*), processes of selection apply generally in many different complex systems and can be described by the same mathematical formalism. When applied to spatial sorting, this approach provides a means to study how local heterogeneities can arise within a context of broader statistical regularities (*1*).

The quantification of processes of selection in terms of information is still relatively new (*30*, *31*, *47*, *59*). We hope that the results described here provide new additional perspectives into this fundamental relationship. To this end, we have shown how to account for information embedded in the spatial structure of complex systems at different scales, thus accounting for the complexity of an overall pattern in terms of both local and global mechanisms. In this way, we hope to bridge two frequently opposing perspectives, by showing how coarse-grained "universality" can co-exist with local exceptionality. This is particularly poignant for cities, where mesmerizing diversity at the neighborhood levels coexists with relatively simple regularities at the level of functional cities and urban systems (*26*, *39*, *60*). The application of these methods to other locally heterogeneous systems, such as ecosystems or neural networks, is straightforward but requires datasets of comparable scope and quality.

In the context of cities, we emphasized the generality of the approach developed here by connecting it to well-known precedents, such as the Theil Index of economic inequality and mathematical models of residential choice. Theil's T, like the indices of spatial selection introduced here, has the special property of self-similarity (recursion) under successive population partitioning, which allows the systematic study of patterns of inequality and spatial selection across a hierarchy of scales. This important property is not shared by other measures of inequality, such as the Gini index. In principle, the property of self-similarity under agglomeration can be used to actively discover specific scales, places and special characteristics (e.g. barriers or connections) at which systems become more strongly sorted (*48*, *61*, *55*). Consequently, this approach applied across scales has the potential to help identify the mechanisms by which inequality and exclusion take place in practice and become entrenched, or indeed how they may be reversed and mitigated.

Related to these issues, models of residential choice implicit in these empirical patterns have recently evolved to represent more natural and empirically supported household decisions, characterized by more realistic local contexts, personal traits, and continuous levels of preference (*19*, *62*). Estimating these personal and contextual characteristics from empirical data requires a systematic methodology such as the one introduced here, which at once measures the complexity (i.e. the amount of information) necessary at different scales. The recursive property of informational quantities tells us how much of the explanation for an overall pattern may be macroscopic or microscopic by helping us keep track of the information content of models at different scales. The results of Figs. 1 and 2, show specifically that the aggregate distribution of income for metropolitan New York City is a poor model for most of the city's neighborhoods, especially its richest and poorest places.

An important dimension to be explored in future studies is how sorting patterns form and change explicitly over time. For example, data are becoming more available that may enable comparisons of income patterns in the same neighborhoods over time (*63–66*), through the direct and comprehensive access to the choices and movements of households. Several recent studies in this direction in US and Canadian cities – using different methods -- suggest that neighborhoods are becoming more segregated by income, a phenomenon known as *neighborhood polarization* (*49*, *67*), see Supplementary Text 2.6. This corresponds to stronger neighborhood effects and suggests that the mutual information between cities' neighborhood structure and income distributions is increasing over time.

A more systematic understanding of spatial population sorting by personal income and other household characteristics remains at the root of some of the most challenging problems for urban science and policy, including the causes and consequences of economic inequality (*51*, *67*, *68*), ethnic and racial segregation, disparate access to opportunity (*65*, *69*) and spatially concentrated (dis)advantage (*33*, *51*, *70*), including issues of crime and violence (*23*, *33*, *34*, *62*, *71*). Extensions of present models and analytical approaches to more urban systems (nations) and several other demographic dimensions (income, race, education, gender,…) remain necessary to make urban policy and practice more effective in the face of radically different challenges faced by specific individuals and places.

Acknowledgements:

We thank Clio Andris, Christa Brelsford, 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.

References

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

2. C. Darwin, E. Mayr, *On the origin of species* (Harvard Univ. Press, Cambridge, Mass., 18. print., a facs. of the first ed., 2003).

3. L. Mumford, *The city in history: its origins, its transformations, and its prospects* (Harcourt, San Diego ;New York ;London, 1989), *A Harvest book*.

4. J. Jacobs, *The death and life of great American cities* (Vintage Books, New York, Vintage Books ed., 1992).

5. J. Monod, A. Wainhouse, *Chance and necessity: an essay on the natural philosophy of modern biology* (Penguin, London, 1997).

6. P. W. Anderson, More Is Different. *Science*. **177**, 393–396 (1972).

7. L. D. Landau, E. M. Lifšic, L. P. Pitaevskij, L. D. Landau, *Statistical physics, Part 1* (Elsevier [u.a.], Amsterdam [u.a], 3. ed., 2008), *Course of theoretical physics*.

8. L. P. Kadanoff, *Statistical physics: statics, dynamics and renormalization* (World Scientific, Singapore ; River Edge, N.J, 2000).

9. S.-K. Ma, *Statistical Mechanics* (Wspc, Philadelphia, 1985).

10. R. Balescu, *Statistical dynamics: matter out of equilibrium* (Imperial College Press, London, 1997).

11. T. M. Cover, *Elements of information theory* (Wiley, New York, 1991), *Wiley series in telecommunications*.

12. N. Goldenfeld, *Lectures on phase transitions and the renormalization group* (Addison-Wesley, Advanced Book Program, Reading, Mass, 1992), *Frontiers in physics*.

13. J. Zinn-Justin, *Phase transitions and renormalization group* (Oxford University Press, Oxford ; New York, 2007), *Oxford graduate texts*.

14. J. B. Tenenbaum, T. L. Griffiths, C. Kemp, Theory-based Bayesian models of inductive learning and reasoning. *Trends Cogn. Sci.* **10**, 309–318 (2006).

15. M. Elad, A. Feuer, Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Process.* **6**, 1646–1658 (1997).

16. T. C. Schelling, *Micromotives and macrobehavior* (Norton, New York, [New ed.] with a new preface and the Nobel Lecture., 2006), *Fels lectures on public policy analysis*.

17. N. Chater, J. B. Tenenbaum, A. Yuille, Probabilistic models of cognition: Conceptual foundations. *Trends Cogn. Sci.* **10**, 287–291 (2006).

18. G. Duranton, D. Puga, in *Handbook of Regional and Urban Economics* (Elsevier, 2004; http://linkinghub.elsevier.com/retrieve/pii/S1574008004800051), vol. 4, pp. 2063–2117.

19. E. E. Bruch, R. D. Mare, Neighborhood Choice and Neighborhood Change. *Am. J. Sociol.* **112**, 667–709 (2006).

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

21. E. L. Glaeser, *Cities, agglomeration, and spatial equilibrium* (Oxford University Press, Oxford, 2008), *Lindahl lectures*.

22. M. O. Jackson, *Social and economic networks* (Princeton Univ. Press, Princeton, NJ, 2008).

23. R. J. Sampson, *Great American city: Chicago and the enduring neighborhood effect* (Univ. of Chicago Press, Chicago, Ill., 2012).

24. S. P. Borgatti, Centrality and network flow. *Soc. Netw.* **27**, 55–71 (2005).

25. W. Warntz, The topology of a socio-economic terrain and spatial flows. *Pap. Reg. Sci. Assoc.* **17**, 47–61 (1966).

26. L. M. A. Bettencourt, The Origins of Scaling in Cities. *Science*. **340**, 1438–1441 (2013).

27. Jianchao Yang, J. Wright, T. S. Huang, Yi Ma, Image Super-Resolution Via Sparse Representation. *IEEE Trans. Image Process.* **19**, 2861–2873 (2010).

28. D. J. C. MacKay, *Information theory, inference, and learning algorithms* (Cambridge University Press, Cambridge, UK ; New York, 2003).

29. V. N. Vapnik, *Statistical learning theory*, vol. 1.

30. M. C. Donaldson-Matasci, C. T. Bergstrom, M. Lachmann, The fitness value of information. *Oikos*. **119**, 219–230 (2010).

31. O. Rivoire, S. Leibler, The Value of Information for Populations in Varying Environments. *J. Stat. Phys.* **142**, 1124–1166 (2011).

32. R. Chartrand, Exact Reconstruction of Sparse Signals via Nonconvex Minimization. *IEEE Signal Process. Lett.* **14**, 707–710 (2007).

33. W. J. Wilson, Studying Inner-City Social Dislocations: The Challenge of Public Agenda Research: 1990 Presidential Address. *Am. Sociol. Rev.* **56**, 1–14 (1991).

34. J. Intrator, J. Tannen, D. S. Massey, Segregation by race and income in the United States 1970–2010. *Soc. Sci. Res.* **60**, 45–60 (2016).

35. E. E. Bruch, How Population Structure Shapes Neighborhood Segregation. *Am. J. Sociol.* **119**, 1221–1278 (2014).

36. G. Firebaugh, F. Acciai, For blacks in America, the gap in neighborhood poverty has declined faster than segregation. *Proc. Natl. Acad. Sci.* **113**, 13372–13377 (2016).

37. B. A. Lee *et al.*, Neighborhood Income Composition by Household Race and Income, 1990–2009. *Ann. Am. Acad. Pol. Soc. Sci.* **660**, 78–97 (2015).

38. M. J. Rich, *Federal Policymaking and the Poor: National Goals, Local Choices, and Distributional Outcomes* (Princeton University Press, 2014).

39. L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G. B. West, Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci.* **104**, 7301–7306 (2007).

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

41. L. M. A. Bettencourt, J. Lobo, D. Strumsky, G. B. West, Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities. *PLoS ONE*. **5**, e13541 (2010).

42. A. Gomez-Lievano, H. Youn, L. M. A. Bettencourt, The Statistics of Urban Scaling and Their Connection to Zipf's Law. *PLoS ONE*. **7**, e40393 (2012).

43. S. A. Frank, Natural selection. III. Selection versus transmission and the levels of selection*: Selection versus Transmission. *J. Evol. Biol.* **25**, 227–243 (2012).

44. K. M. Page, M. A. Nowak, Unifying Evolutionary Dynamics. *J. Theor. Biol.* **219**, 93–98 (2002).

45. H. Theil, *Economics and information theory* (North-Holland, Amsterdam, 1967).

46. S. Kullback, *Information theory and statistics* (Dover Publications, Mineola, N.Y, 1997), *Dover books on mathematics*.

47. 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).

48. J. Ludwig *et al.*, Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults. *Science*. **337**, 1505–1510 (2012).

49. S. F. Reardon, K. Bischoff, Income Inequality and Income Segregation. *Am. J. Sociol.* **116**, 1092–1153 (2011).

50. E. Roberto, The Divergence Index: A Decomposable Measure of Segregation and Inequality. *ArXiv150801167 Phys. Stat* (2015) (available at http://arxiv.org/abs/1508.01167).

51. L. J. Krivo *et al.*, Social Isolation of Disadvantage and Advantage: The Reproduction of Inequality in Urban Space. *Soc. Forces*. **92**, 141–164 (2013).

52. S. Musterd, W. P. van Gent, M. Das, J. Latten, Adaptive behaviour in urban space: Residential mobility in response to social distance. *Urban Stud.* **53**, 227–246 (2016).

53. C. D. Solari, Affluent Neighborhood Persistence and Change in U.S. Cities. *City Community*. **11**, 370–388 (2012).

54. B. A. Lee, G. Firebaugh, J. Iceland, S. A. Matthews, A. Owens, Assisted Housing and Income Segregation among Neighborhoods in U.S. Metropolitan Areas. *Ann. Am. Acad. Pol. Soc. Sci.* **660**, 98–116 (2015).

55. E. Roberto, J. Hwang, Barriers to Integration: Institutionalized Boundaries and the Spatial Structure of Residential Segregation. *ArXiv150902574 Phys. Stat* (2015) (available at http://arxiv.org/abs/1509.02574).

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

57. R. R. Nelson, S. G. Winter, *An Evolutionary Theory of Economic Change* (Harvard University Press, 2009).

58. A. Smith, A. B. Krueger, *The Wealth of Nations* (Bantam Classics, New York, N.Y, Reprint edition., 2003).

59. M. Hilbert, The More You Know, the More You Can Grow: An Information Theoretic Approach to Growth in the Information Age. *Entropy*. **19**, 82 (2017).

60. M. Batty, The Size, Scale, and Shape of Cities. *Science*. **319**, 769–771 (2008).

61. A. Owens, Neighborhoods on the Rise: A Typology of Neighborhoods Experiencing Socioeconomic Ascent. *City Community*. **11**, 345–369 (2012).

62. M. Besbris, J. W. Faber, P. Rich, P. Sharkey, Effect of neighborhood stigma on economic transactions. *Proc. Natl. Acad. Sci.* **112**, 4994–4998 (2015).

63. C. R. Browning, K. A. Cagney, B. Boettner, in *Handbook of the Life Course*, M. J. Shanahan, J. T. Mortimer, M. K. Johnson, Eds. (Springer International Publishing, 2016; http://link.springer.com/chapter/10.1007/978-3-319-20880-0_26), *Handbooks of Sociology and Social Research*, pp. 597–620.

64. S. J. South, Y. Huang, A. Spring, K. Crowder, Neighborhood Attainment over the Adult Life Course. *Am. Sociol. Rev.* **81**, 1276–1304 (2016).

65. R. Chetty *et al.*, The fading American dream: Trends in absolute income mobility since 1940. *Science*. **356**, 398–406 (2017).

66. J. R. Hipp, A Dynamic View of Neighborhoods: The Reciprocal Relationship between Crime and Neighborhood Structural Characteristics. *Soc. Probl.* **57**, 205–230 (2010).

67. W.-H. Chen, J. Myles, G. Picot, Why Have Poorer Neighbourhoods Stagnated Economically while the Richer Have Flourished? Neighbourhood Income Inequality in Canadian Cities. *Urban Stud.* **49**, 877–896 (2012).

68. G. Firebaugh, C. R. Farrell, Still Large, but Narrowing: The Sizable Decline in Racial Neighborhood Inequality in Metropolitan America, 1980–2010. *Demography*. **53**, 139–164 (2016).

69. M. C. Lens, Measuring the geography of opportunity. *Prog. Hum. Geogr.* **41**, 3–25 (2017).

70. D. S. Elliott *et al.*, The Effects of Neighborhood Disadvantage on Adolescent Development. *J. Res. Crime Delinquency*. **33**, 389–426 (1996).

71. W. J. Wilson, *When Work Disappears: The World of the New Urban Poor* (Knopf Doubleday Publishing Group, 2011).

72. J. R. Hipp, Block, Tract, and Levels of Aggregation: Neighborhood Structure and Crime and Disorder as a Case in Point. *Am. Sociol. Rev.* **72**, 659–680 (2007).

Materials and Methods

Experimental Design

In this study, we measured and analyzed spatial patterns of household income distribution in US cities (metropolitan statistical areas). We used geo-referenced data at the household level (such as income and population) for the United States, reported at the *Census Block Group* level by the 2006-2010 5-year American Community Survey, see Supplementary Text S1.1. Block groups are statistical subdivisions of US 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 nation. 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. Section S1.1 in Supplementary Materials provides more detail about data sources.

In the main text we used the term *neighborhood* to refer to Census Block Groups, as shown in Figure 1A and S1-17. 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 S1-17. 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 (*23*, *72*) and Supplementary Text S1.2 for discussions. Our aim here is to demonstrate effects of spatial selection at any given set of scales. A systematic study of the strength of spatial selection at different local scales (neighborhood definitions) is beyond the scope of the present manuscript and will be presented elsewhere. Sections S1.2 and S1.3 in Supplementary Materials Text discuss the issues of units of analysis, neighborhoods and data limitations further.

Statistical Analysis

Data are provided in terms of household counts within fixed income bins, as shown in Figure 1B. We used these counts to define frequencies in each income bin and normalized these by total population counts in each neighborhood and metropolitan level for every city in the US. This produces standard frequency estimators for the various distributions necessary to compute Eqs. 1-5. We verified that the patterns of income in each neighborhood, and the values of the KL-Divergences and Mutual Information could not have been created by a random (uniform) draw from the statistics of income at the metropolitan level, as shown in the Supplementary Material, Figs S40. Section S1.3 in Supplementary Materials discusses some of the limitations of the American Community Survey data, especially for high-income households. Section S1.4 provides a detailed procedure and set of mathematical expressions for estimating the various probability distributions used in the main text.

References

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

2. C. Darwin, E. Mayr, *On the origin of species* (Harvard Univ. Press, Cambridge, Mass., 18. print., a facs. of the first ed., 2003).

3. L. Mumford, *The city in history: its origins, its transformations, and its prospects* (Harcourt, San Diego ;New York ;London, 1989), *A Harvest book*.

4. J. Jacobs, *The death and life of great American cities* (Vintage Books, New York, Vintage Books ed., 1992).

5. J. Monod, A. Wainhouse, *Chance and necessity: an essay on the natural philosophy of modern biology* (Penguin, London, 1997).

6. P. W. Anderson, More Is Different. *Science*. **177**, 393–396 (1972).

7. L. D. Landau, E. M. Lifšic, L. P. Pitaevskij, L. D. Landau, *Statistical physics, Part 1* (Elsevier [u.a.], Amsterdam [u.a], 3. ed., 2008), *Course of theoretical physics*.

8. L. P. Kadanoff, *Statistical physics: statics, dynamics and renormalization* (World Scientific, Singapore ; River Edge, N.J, 2000).

9. S.-K. Ma, *Statistical Mechanics* (Wspc, Philadelphia, 1985).

10. R. Balescu, *Statistical dynamics: matter out of equilibrium* (Imperial College Press, London, 1997).

11. T. M. Cover, *Elements of information theory* (Wiley, New York, 1991), *Wiley series in telecommunications*.

12. N. Goldenfeld, *Lectures on phase transitions and the renormalization group* (Addison-Wesley, Advanced Book Program, Reading, Mass, 1992), *Frontiers in physics*.

13. J. Zinn-Justin, *Phase transitions and renormalization group* (Oxford University Press, Oxford ; New York, 2007), *Oxford graduate texts*.

14. J. B. Tenenbaum, T. L. Griffiths, C. Kemp, Theory-based Bayesian models of inductive learning and reasoning. *Trends Cogn. Sci.* **10**, 309–318 (2006).

15. M. Elad, A. Feuer, Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Process.* **6**, 1646–1658 (1997).

16. T. C. Schelling, *Micromotives and macrobehavior* (Norton, New York, [New ed.] with a new preface and the Nobel Lecture., 2006), *Fels lectures on public policy analysis*.

17. N. Chater, J. B. Tenenbaum, A. Yuille, Probabilistic models of cognition: Conceptual foundations. *Trends Cogn. Sci.* **10**, 287–291 (2006).

18. G. Duranton, D. Puga, in *Handbook of Regional and Urban Economics* (Elsevier, 2004; http://linkinghub.elsevier.com/retrieve/pii/S1574008004800051), vol. 4, pp. 2063–2117.

19. E. E. Bruch, R. D. Mare, Neighborhood Choice and Neighborhood Change. *Am. J. Sociol.* **112**, 667–709 (2006).

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

21. E. L. Glaeser, *Cities, agglomeration, and spatial equilibrium* (Oxford University Press, Oxford, 2008), *Lindahl lectures*.

22. M. O. Jackson, *Social and economic networks* (Princeton Univ. Press, Princeton, NJ, 2008).

23. R. J. Sampson, *Great American city: Chicago and the enduring neighborhood effect* (Univ. of Chicago Press, Chicago, Ill., 2012).

24. S. P. Borgatti, Centrality and network flow. *Soc. Netw.* **27**, 55–71 (2005).

25. W. Warntz, The topology of a socio-economic terrain and spatial flows. *Pap. Reg. Sci. Assoc.* **17**, 47–61 (1966).

26. L. M. A. Bettencourt, The Origins of Scaling in Cities. *Science*. **340**, 1438–1441 (2013).

27. Jianchao Yang, J. Wright, T. S. Huang, Yi Ma, Image Super-Resolution Via Sparse Representation. *IEEE Trans. Image Process.* **19**, 2861–2873 (2010).

28. D. J. C. MacKay, *Information theory, inference, and learning algorithms* (Cambridge University Press, Cambridge, UK ; New York, 2003).

29. V. N. Vapnik, *Statistical learning theory*, vol. 1.

30. M. C. Donaldson-Matasci, C. T. Bergstrom, M. Lachmann, The fitness value of information. *Oikos*. **119**, 219–230 (2010).

31. O. Rivoire, S. Leibler, The Value of Information for Populations in Varying Environments. *J. Stat. Phys.* **142**, 1124–1166 (2011).

32. R. Chartrand, Exact Reconstruction of Sparse Signals via Nonconvex Minimization. *IEEE Signal Process. Lett.* **14**, 707–710 (2007).

33. W. J. Wilson, Studying Inner-City Social Dislocations: The Challenge of Public Agenda Research: 1990 Presidential Address. *Am. Sociol. Rev.* **56**, 1–14 (1991).

34. J. Intrator, J. Tannen, D. S. Massey, Segregation by race and income in the United States 1970–2010. *Soc. Sci. Res.* **60**, 45–60 (2016).

35. E. E. Bruch, How Population Structure Shapes Neighborhood Segregation. *Am. J. Sociol.* **119**, 1221–1278 (2014).

36. G. Firebaugh, F. Acciai, For blacks in America, the gap in neighborhood poverty has declined faster than segregation. *Proc. Natl. Acad. Sci.* **113**, 13372–13377 (2016).

37. B. A. Lee *et al.*, Neighborhood Income Composition by Household Race and Income, 1990–2009. *Ann. Am. Acad. Pol. Soc. Sci.* **660**, 78–97 (2015).

38. M. J. Rich, *Federal Policymaking and the Poor: National Goals, Local Choices, and Distributional Outcomes* (Princeton University Press, 2014).

39. L. M. A. Bettencourt, J. Lobo, D. Helbing, C. Kühnert, G. B. West, Growth, innovation, scaling, and the pace of life in cities. *Proc. Natl. Acad. Sci.* **104**, 7301–7306 (2007).

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

41. L. M. A. Bettencourt, J. Lobo, D. Strumsky, G. B. West, Urban Scaling and Its Deviations: Revealing the Structure of Wealth, Innovation and Crime across Cities. *PLoS ONE*. **5**, e13541 (2010).

42. A. Gomez-Lievano, H. Youn, L. M. A. Bettencourt, The Statistics of Urban Scaling and Their Connection to Zipf's Law. *PLoS ONE*. **7**, e40393 (2012).

43. S. A. Frank, Natural selection. III. Selection versus transmission and the levels of selection*: Selection versus Transmission. *J. Evol. Biol.* **25**, 227–243 (2012).

44. K. M. Page, M. A. Nowak, Unifying Evolutionary Dynamics. *J. Theor. Biol.* **219**, 93–98 (2002).

45. H. Theil, *Economics and information theory* (North-Holland, Amsterdam, 1967).

46. S. Kullback, *Information theory and statistics* (Dover Publications, Mineola, N.Y, 1997), *Dover books on mathematics*.

47. 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).

48. J. Ludwig *et al.*, Neighborhood Effects on the Long-Term Well-Being of Low-Income Adults. *Science*. **337**, 1505–1510 (2012).

49. S. F. Reardon, K. Bischoff, Income Inequality and Income Segregation. *Am. J. Sociol.* **116**, 1092–1153 (2011).

50. E. Roberto, The Divergence Index: A Decomposable Measure of Segregation and Inequality. *ArXiv150801167 Phys. Stat* (2015) (available at http://arxiv.org/abs/1508.01167).

51. L. J. Krivo *et al.*, Social Isolation of Disadvantage and Advantage: The Reproduction of Inequality in Urban Space. *Soc. Forces*. **92**, 141–164 (2013).

52. S. Musterd, W. P. van Gent, M. Das, J. Latten, Adaptive behaviour in urban space: Residential mobility in response to social distance. *Urban Stud.* **53**, 227–246 (2016).

53. C. D. Solari, Affluent Neighborhood Persistence and Change in U.S. Cities. *City Community*. **11**, 370–388 (2012).

54. B. A. Lee, G. Firebaugh, J. Iceland, S. A. Matthews, A. Owens, Assisted Housing and Income Segregation among Neighborhoods in U.S. Metropolitan Areas. *Ann. Am. Acad. Pol. Soc. Sci.* **660**, 98–116 (2015).

55. E. Roberto, J. Hwang, Barriers to Integration: Institutionalized Boundaries and the Spatial Structure of Residential Segregation. *ArXiv150902574 Phys. Stat* (2015) (available at http://arxiv.org/abs/1509.02574).

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

57. R. R. Nelson, S. G. Winter, *An Evolutionary Theory of Economic Change* (Harvard University Press, 2009).

58. A. Smith, A. B. Krueger, *The Wealth of Nations* (Bantam Classics, New York, N.Y, Reprint edition., 2003).

59. M. Hilbert, The More You Know, the More You Can Grow: An Information Theoretic Approach to Growth in the Information Age. *Entropy*. **19**, 82 (2017).

60. M. Batty, The Size, Scale, and Shape of Cities. *Science*. **319**, 769–771 (2008).

61. A. Owens, Neighborhoods on the Rise: A Typology of Neighborhoods Experiencing Socioeconomic Ascent. *City Community*. **11**, 345–369 (2012).

62. M. Besbris, J. W. Faber, P. Rich, P. Sharkey, Effect of neighborhood stigma on economic transactions. *Proc. Natl. Acad. Sci.* **112**, 4994–4998 (2015).

63. C. R. Browning, K. A. Cagney, B. Boettner, in *Handbook of the Life Course*, M. J. Shanahan, J. T. Mortimer, M. K. Johnson, Eds. (Springer International Publishing, 2016; http://link.springer.com/chapter/10.1007/978-3-319-20880-0_26), *Handbooks of Sociology and Social Research*, pp. 597–620.

64. S. J. South, Y. Huang, A. Spring, K. Crowder, Neighborhood Attainment over the Adult Life Course. *Am. Sociol. Rev.* **81**, 1276–1304 (2016).

65. R. Chetty *et al.*, The fading American dream: Trends in absolute income mobility since 1940. *Science*. **356**, 398–406 (2017).

66. J. R. Hipp, A Dynamic View of Neighborhoods: The Reciprocal Relationship between Crime and Neighborhood Structural Characteristics. *Soc. Probl.* **57**, 205–230 (2010).

67. W.-H. Chen, J. Myles, G. Picot, Why Have Poorer Neighbourhoods Stagnated Economically while the Richer Have Flourished? Neighbourhood Income Inequality in Canadian Cities. *Urban Stud.* **49**, 877–896 (2012).

68. G. Firebaugh, C. R. Farrell, Still Large, but Narrowing: The Sizable Decline in Racial Neighborhood Inequality in Metropolitan America, 1980–2010. *Demography*. **53**, 139–164 (2016).

69. M. C. Lens, Measuring the geography of opportunity. *Prog. Hum. Geogr.* **41**, 3–25 (2017).

70. D. S. Elliott *et al.*, The Effects of Neighborhood Disadvantage on Adolescent Development. *J. Res. Crime Delinquency*. **33**, 389–426 (1996).

71. W. J. Wilson, *When Work Disappears: The World of the New Urban Poor* (Knopf Doubleday Publishing Group, 2011).

72. J. R. Hipp, Block, Tract, and Levels of Aggregation: Neighborhood Structure and Crime and Disorder as a Case in Point. *Am. Sociol. Rev.* **72**, 659–680 (2007).

Figure Captions

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 very good general model for the household income distribution in all U.S. MSAs. The mean and variance of this distribution obey simple scaling relations, see Supplementary Materials S2.2. 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 Materials and Methods and Supplementary Sections S1.1-1.3 for more details and a discussion of neighborhood units.

Figure 2: **Spatial Selection Measured in Units of Information**. **A.** The information, $< \ln_2 w_j >$, 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 Shannon 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 local explanations that are as comprehensive 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 $< \ln_2 w_\ell >$, 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 $< \ln_2 w_\ell >$ with income can be reasonably fitted by a quadratic form (red line, for New York MSA) with $< \ln_2 w_\ell >=) = 1.058\ - 1.167 \text{ x } 10^{-5}\text{y} +\ 6.074 \text{ x } 10^{-11}\ \text{y}^2$. **C**. The average strength of neighborhood effects across urban areas in the US measured by the mutual information $I(y;n)$ (942 metropolitan and micropolitan areas). 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 $I(y;n) \rightarrow 0$ have less distinguishable neighborhood structure by income.

Figure 1

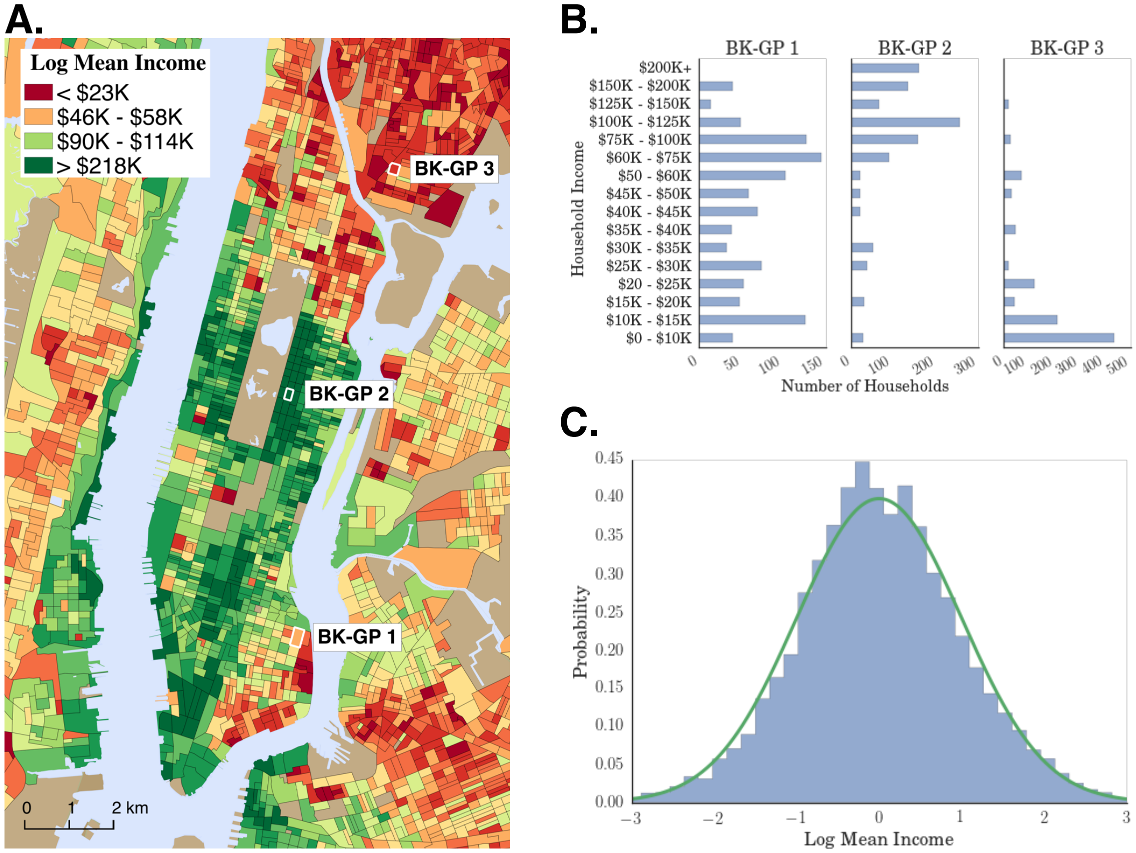
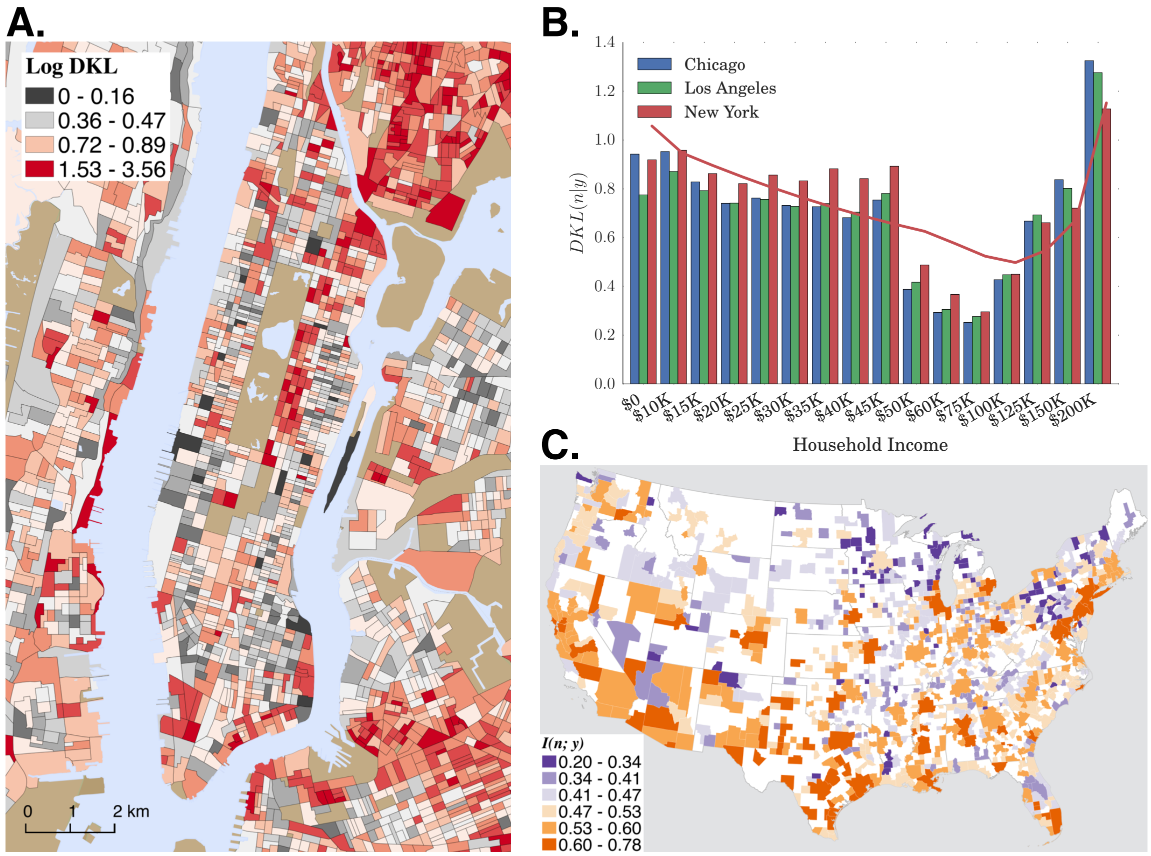

A.
Log Mean Income
< $23K
$46K - $58K
$90K - $114K
> $218K
BK-GP 3
BK-GP 2
BK-GP 1
0
1
2 km
B.
BK-GP 1
BK-GP 2
BK-GP 3
$200K+
$150K - $200K
$125K - $150K
$100K - $125K
$75K - $100K
$60K - $75K
$50 - $60K
$45K - $50K
$40K - $45K
$35K - $40K
$30K - $35K
$25K - $30K
$20 - $25K
$15K - $20K
$10K - $15K
$0 - $10K
Household Income
0
50
100
150
0
100
200
300
0
100
200
300
400
500
Number of Households
C.
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Probability
−3
−2
−1
0
1
2
3
Log Mean Income

Figure 2


A.
Log DKL
0 - 0.16
0.36 - 0.47
0.72 - 0.89
1.53 - 3.56
0
1
2 km
B.
Chicago
Los Angeles
New York
DKL(n|y)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
$0
$10K
$15K
$20K
$25K
$30K
$35K
$40K
$45K
$50K
$60K
$75K
$100K
$125K
$150K
$200K
Household Income
C.
I(n; y)
0.20 - 0.34
0.34 - 0.41
0.41 - 0.47
0.47 - 0.53
0.53 - 0.60
0.60 - 0.78

Information, 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

June 19, 2017

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 (*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 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

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

where we identify the $\log w_{\ell,j}$ term as the specific mutual information (before averaging) between the states y_ℓ and n_j. Moreover, note that the specific Shannon entropies are $h(y_\ell|n_j) = -\log p(y_\ell|n_j)$ and $h(y_\ell) = -\log p(y_\ell)$ (*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, $\langle \log w \rangle = I(y;n)$ and is given by

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

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 levels 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

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

Thus, we write

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

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*).

2.2 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 - S40 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.

2.3 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*).

2.4 Information, the Theil Inequality Index and the Aggregation Problem

The Theil Index as a Kulback-Leibler divergence

This section expands on the more succinct arguments given in the main text about Theil's T index of inequality and its connection to information theoretic quantities. Theil explicitly seeked to connect Shannon's measure of entropy, H,

$$H(p) = \sum_{\ell} p(y_\ell) \log 1/p(y_\ell), \tag{S20}$$

where $p(y_\ell)$ is a (discrete) population density such that $\sum_\ell p(y_\ell) = 1$, to issues of economic inequality in a human population.

To do this, he defined an income share density, $q(y_\ell)$. This follows from the standard density, $p(y_\ell)$, which is the share of the population in some group indexed by ℓ, as $q(y_\ell) = \frac{y_\ell}{\bar{y}} p(y_\ell)$, where $\bar{y} = \sum_\ell y_\ell p(y_\ell)$ is the average income per capita in the population.

The Theil index T then compares the income shares of individuals to the case where income is distributed uniformly, that is $y_i = y_j$ for all i, j, in terms of a difference of entropies as

$$T(y) = \frac{1}{N} \sum_{k=1}^{N} \frac{y_k}{\bar{y}} \log \frac{y_k}{\bar{y}}, \tag{S21}$$

where the sum is over individuals in a population of size N. Taking the same population to be aggregated into groups indexed by ℓ, we can write the same expression more elegantly as

$$T(y) = \sum_{\ell} q(y_\ell) \log \frac{q(y_\ell)}{p(y_\ell)} = D_{KL}\left[q(y) || p(y)\right], \tag{S22}$$

which is the Kulback-Leibler divergence between the density of income and population shares in the groups. This expression reveals the informational character of the Theil index and leads to the interpretation of T as the information error (measured in bits) between describing the distribution of income via assuming it is proportional to the population in each group, that is to the case of absolute equality.

We can also now compare and contrast the approach to patterning of distributions across scales in the main paper, and Theil's measure of inequality. While the former explores how a distribution (of income or of any other quantity) changes across (spatial) scales of aggregation, revealing a pattern of sorting or selection, the latter compares two different types of distribution at the same scale, manifesting how two different quantities, say income and population, are distributed into population groups in the same way or otherwise.

Because both approaches rely on the comparison of distributions via information quantities, they share some common properties, to which we now turn.

The Aggregation Problem, the Theil Index and Information across Scales

The definition of the Theil index, discussed above, was originally motivated by information theory and its properties under multilevel set (dis)aggregation. This means more specifically that if we successively decompose a population into sets, and sets of sets etc, in a non-overlapping hierarchical way, we can compute inequality sequentially.

To see this explicitly, consider the fact that we can write a total distribution of income $p(y_\ell)$, using the same distribution within groups n_i, $p(y_\ell|n_i)$ as,

$$p(y_\ell) = \sum_i p(y_\ell|n_i) p(n_i), \tag{S23}$$

where $p(n_i)$ is the probability of an individual in the population belonging to set i. These sets can be spatial, such as neighborhoods, but they need not be; in his book Theil uses the example of racial groups. If we define N_i as the population of set n_i then $p(n_i) = N_i/N$ and $\bar{y}_i = \sum_\ell y_\ell p(y_\ell|n_i)$ is the average income in set n_i. It follows immediately that we can decompose the income share distribution $q(y_\ell)$ is the same way, so that

$$q(y_\ell) = \sum_i \frac{y_\ell}{\bar{y}} p(y_\ell|n_i)p(n_i) = \sum_{i=1} \frac{y_\ell}{\bar{y}_i} p(y_\ell|n_i)\ \frac{\bar{y}_i}{\bar{y}} p(n_i) = \sum_i q(y_\ell|n_i)q(n_i). \tag{S24}$$

Introducing this expression into the informational definition for T, and noting that

$$\log \frac{q(y_\ell)}{p(y_\ell)} = \log \frac{y_\ell}{\bar{y}} = \log \frac{y_\ell}{\bar{y}_i} + \log \frac{\bar{y}_i}{\bar{y}} = \log \frac{q(y_\ell|n_i)}{p(y_\ell|n_i)} + \log \frac{q(n_i)}{p(n_i)}, \tag{S25}$$

we obtain the hierarchical decomposition,

$$\begin{aligned} T(y) \quad &= \sum_i q(n_i) \log \frac{q(n_i)}{p(n_i)} + \sum_i q(n_i) \sum_\ell q(y_\ell|n_i) \log \frac{q(y_\ell|n_i)}{p(y_\ell|n_i)} \\ &\equiv T(n) + \sum_i q(n_i) T(y|n_i), \end{aligned} \tag{S26}$$

where the first term is the inequality across sets and the second term is the set average of the inequality index within each subgroup.

Naturally, the patterning functions developed in the main text have an analogous aggregation property. To see this consider two levels of (dis)aggregation, from a general (metropolitan) population into sets n_i, which are then further decomposed into sets m_j. This implies two successive levels of selection, which we can write as

$$p(y_\ell|n_i, m_j) = w'_{\ell ij} p(y_\ell|n_i) = w'_{\ell ij}\ w_{\ell i} p(y_\ell) \equiv w''_{\ell ij} p(y_\ell) \tag{S27}$$

with

$$w_{\ell i} = \frac{p(y_\ell|n_i)}{p(y_\ell)} = \frac{p(y_\ell, n_i)}{p(n_i)p(y_\ell)}, \tag{S28}$$

$$w'_{\ell ij} = \frac{p(y_\ell|n_i, m_j)}{p(y_\ell|n_i)} = \frac{p(y_\ell, m_j|n_i)}{p(m_j|n_i)p(y_\ell|n_i)} = \frac{p(y_\ell, m_j, n_i)}{p(m_j, n_i)p(y_\ell|n_i)}. \tag{S29}$$

This also implies the identity $\log w''_{\ell ij} = \log w'_{\ell ij} + \log w_{\ell i}$, which will reappear below under averaging. Now, using the definition of the conditional densities, we can then write the following quantities

$$\begin{aligned}
\langle \log w''_{ij} \rangle \quad & \equiv D_{KL}(p(y|n_i, m_j)||p(y)) = \sum_{\ell} p(y_\ell|n_i, m_j) \log \frac{p(y_\ell|n_i, m_j)}{p(y)} && \text{(S30)} \\
& = \sum_{\ell} p(y_\ell|n_i, m_j) \left[\log \frac{p(y_\ell|n_i, m_j)}{p(y|n_i)} + \log \frac{p(y_\ell|n_i)}{p(y)} \right], && \text{(S31)}
\end{aligned}$$

which can then be averaged over the two-set level decomposition, to give the multi-information

$$\begin{aligned}
I(y; n, m) \quad & = \sum_{i,j} p(n_i, m_j) \langle \log w''_{ij} \rangle = \sum_{i,j} p(n_i) p(m_j|n_i) \sum_{\ell} D_{KL}(p(y|n_i, m_j)||p(y)) \\
& = \sum_{i} p(n_i) D_{KL}(p(y|n_i)||p(y)) \\
& \qquad\qquad + \sum_{i,j} p(n_i, m_j) D_{KL}(p(y|n_i, m_j)||p(y|n_i)) \\
& = \sum_{i} p(n_i) \langle \log w_i \rangle + \sum_{ij} p(n_i, m_j) \langle \log w'_{ij} \rangle && \text{(S32)} \\
& = I(y; n) + \sum_{i} p(n_i) I(y; m|n_i),
\end{aligned}$$

which shows how the information in the pattern is contained in the two levels of selection and how each level contributes according to the respective set probabilities.

Thus, in our view, information theoretic quantities are the most natural way to express differences in distribution of either income in a population (revealing issues of inequality) or its patterning across scales (sorting, or selection). A family of quantities, sharing analogous properties under (dis)aggregation into sets can in this way be created that reveals how populations are structures across quantities and scales in a systematic quantitative manner, measured in units of information.

2.5 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

$$\Delta \bar{z} = \Delta_{\mathrm{s}} \bar{z} + \Delta_{\mathrm{t}} \bar{z} = \sum_i (w_i - 1) z_i p(z_i) + \sum_i \Delta z_i p'(z_i), \tag{S33}$$

respectively. The first term, $\Delta_{\mathrm{s}}\bar{z}$, is associated with selection whereas the second, $\Delta_{\mathrm{t}}\bar{z}$, 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, $\Delta_{\mathrm{t}}\bar{z} = 0$, 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:

$$\Delta_s \bar{z} = \sum_i (w_i - 1) z_i p(z_i) = \mathrm{covar}(w, z) = \beta_{wz} \sigma_w^2 = \beta_{zw} \sigma_z^2. \tag{S34}$$

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 $\sigma^2_{\log y}$. 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 $\log y$ 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 S41 - S43.

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:

$$\bar{y}_j = \frac{\sum_\ell \bar{y}_{\ell,j} n_{\ell,j}}{N_j} = \frac{\sum_\ell y_{\min,\ell} n_{\ell,j}}{N_j} + \epsilon_j \tag{S35}$$

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 S41 and S42. We observe that there are 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

$$a_j = \frac{\bar{y}_j - \bar{y}_{\mathrm{min},j}}{\bar{y}_{\mathrm{m},j} - \bar{y}_{\mathrm{min},j}} \tag{S36}$$

where $\bar{y}_j$ is the reported mean average income for each neighborhood, $\bar{y}_{\mathrm{min},j}$ 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 S41 - S43, show this process for New York City.

2.6 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 are 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. This requires a development of socioeconomic theories of spatial selection that, neighborhood by neighborhood, can account for the differential amounts of spatial sorting quantified here.

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 household incomes within neighborhoods (census tracts) to the variation in household 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 - S40 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 easily be accounted for 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*).

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

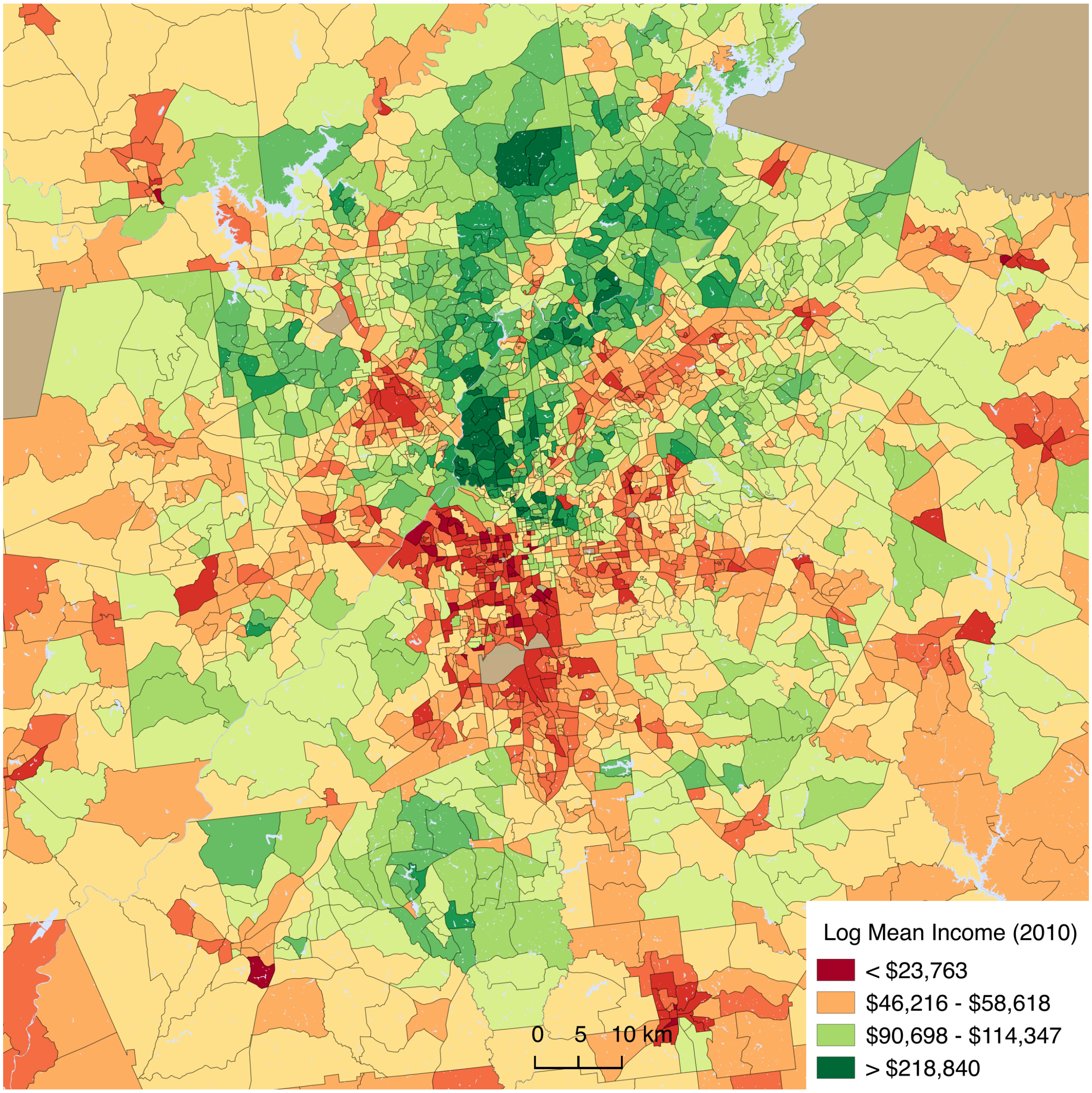


Figure S1: Atlanta, GA Mean Household Income (2010)

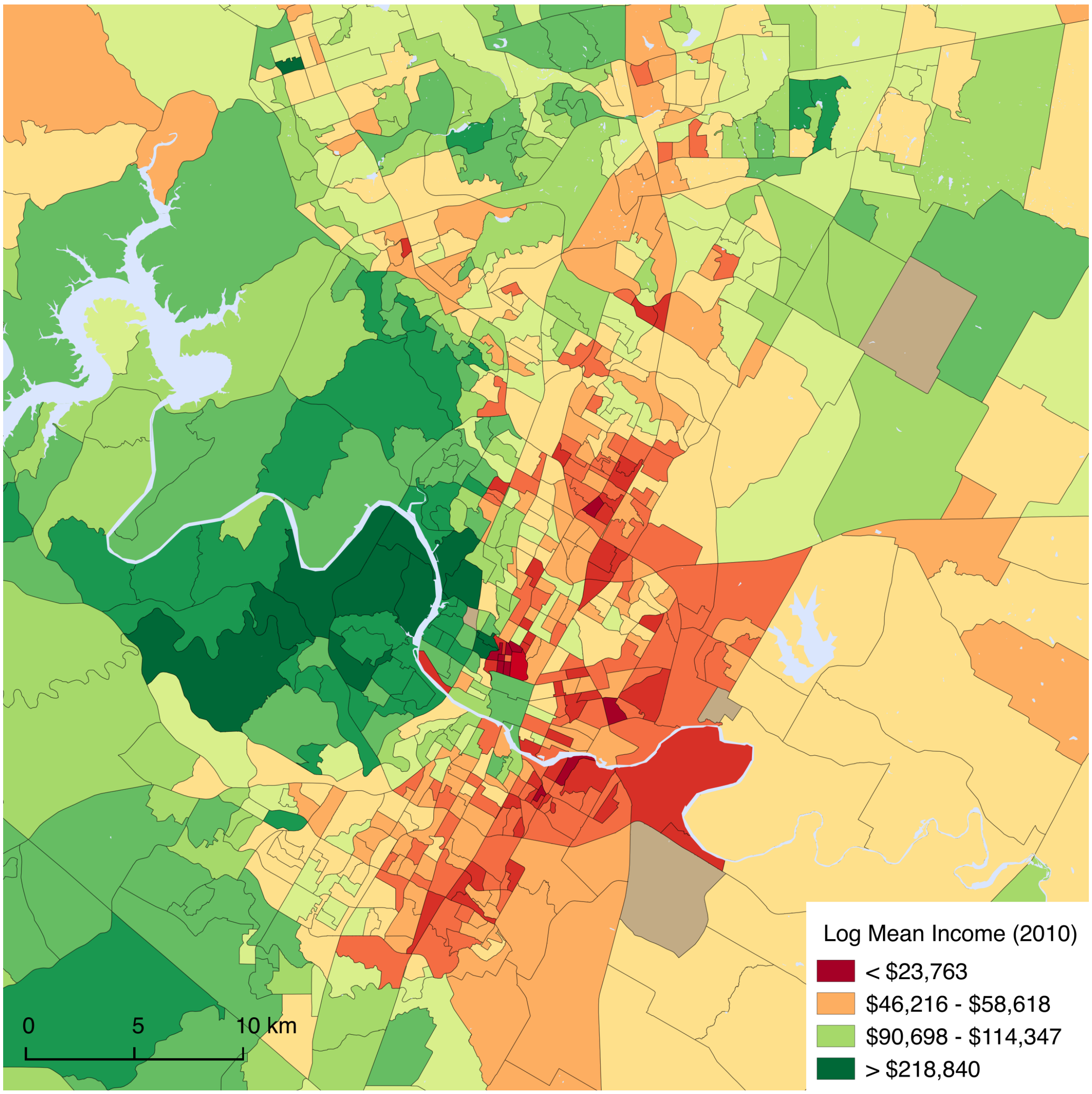


Figure S2: Austin, TX Mean Household Income (2010)

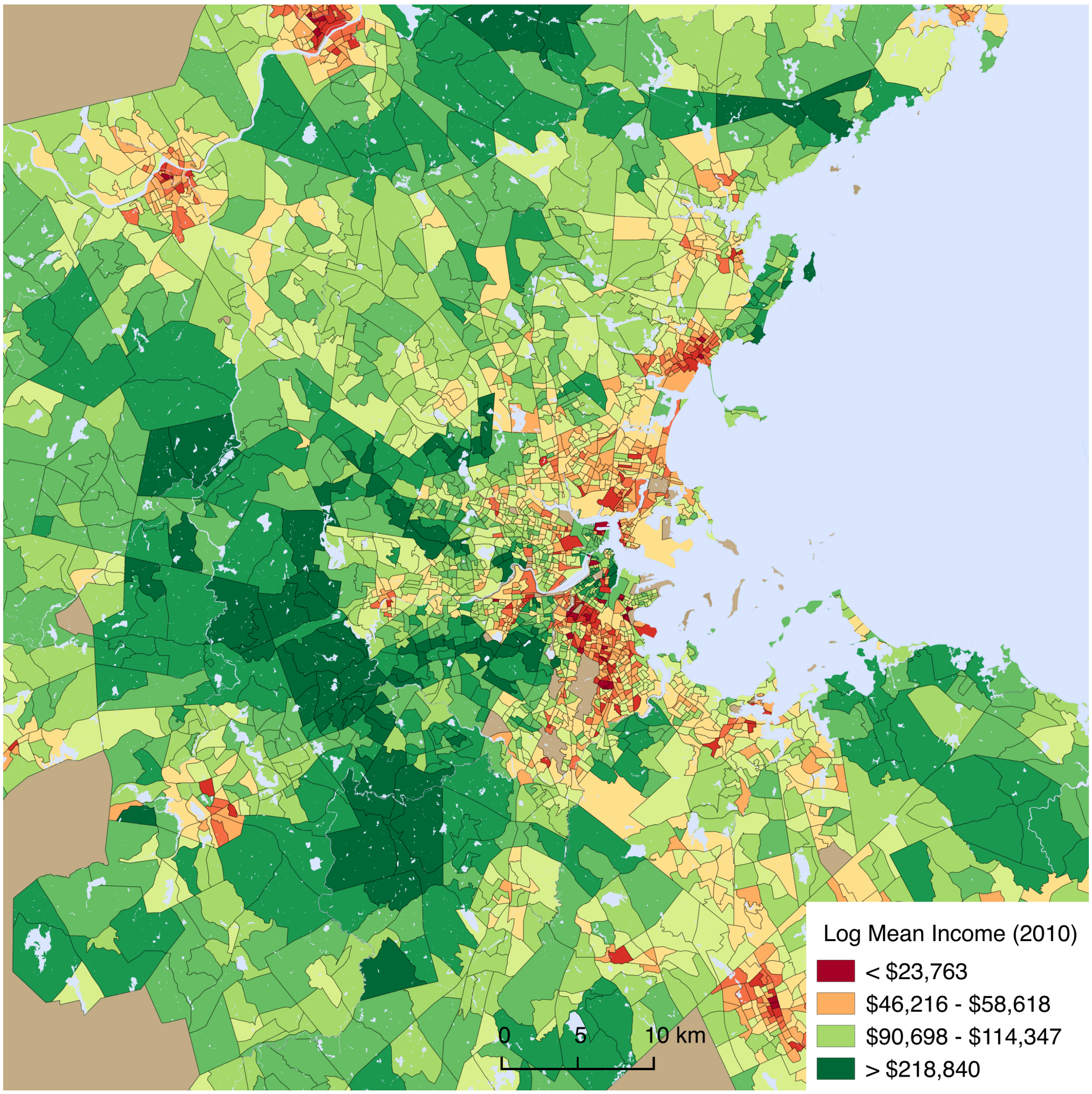


Figure S3: Boston, MA Mean Household Income (2010)

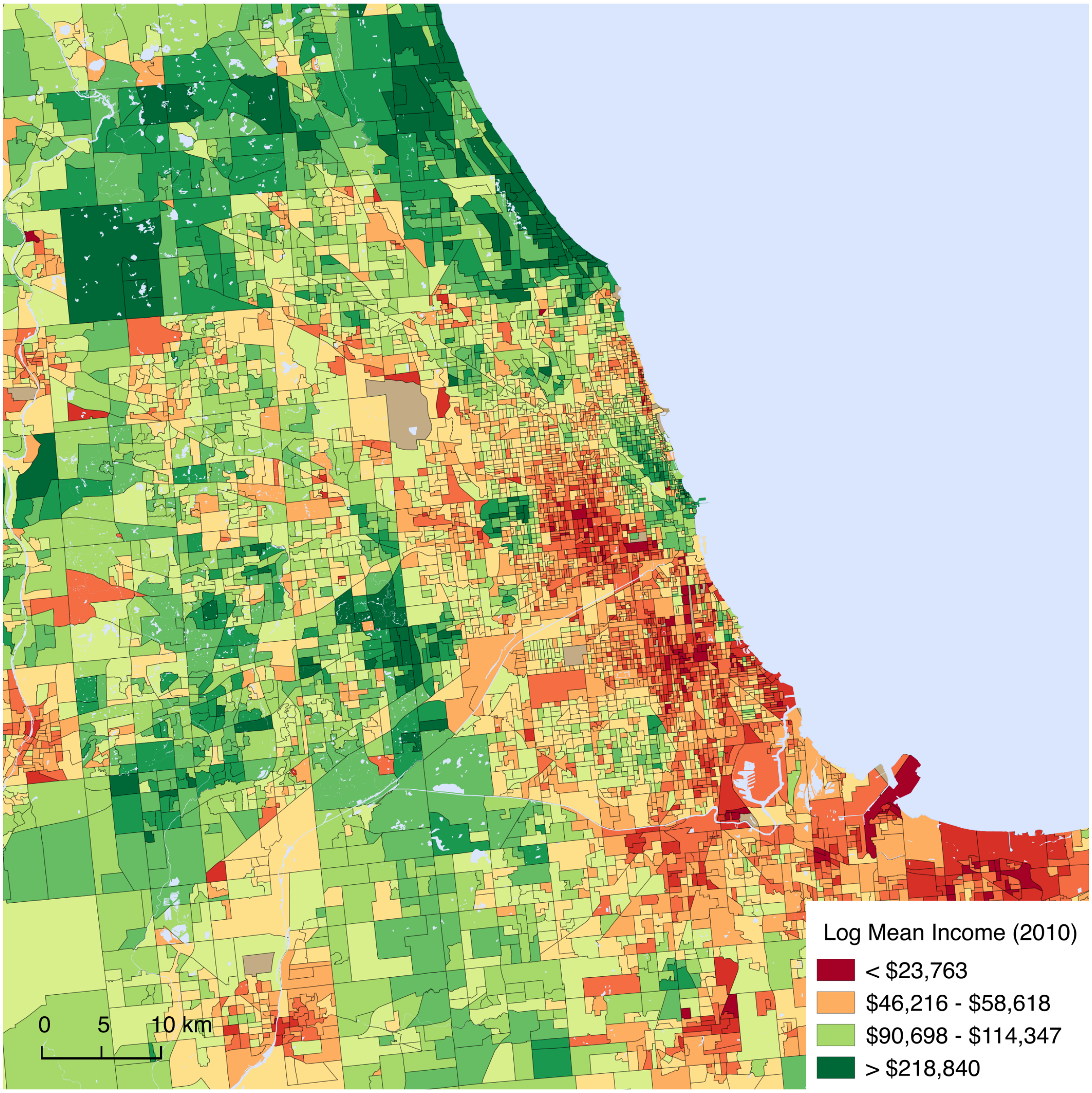


Figure S4: Chicago, IL Mean Household Income (2010)

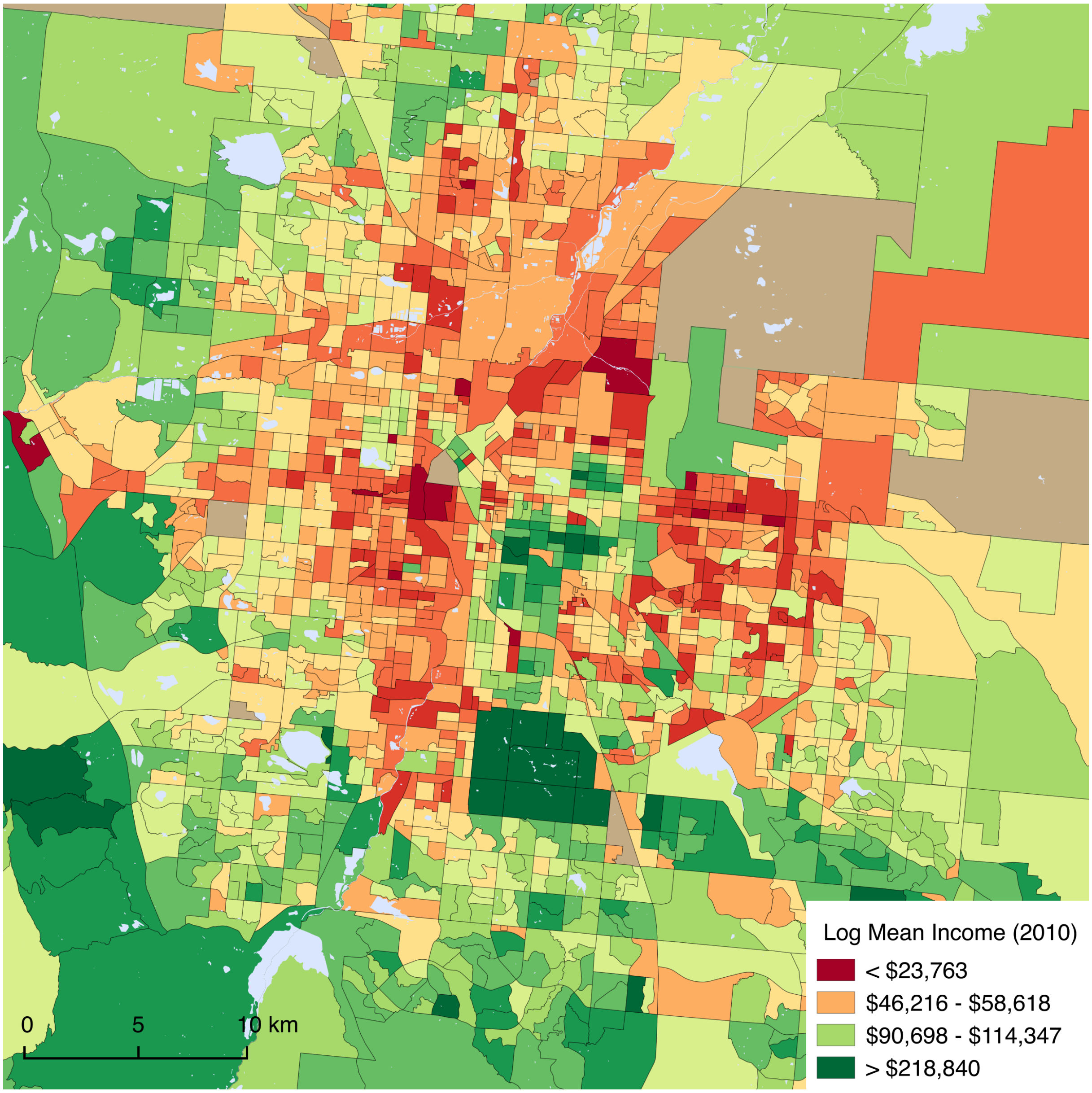


Figure S5: Denver, CO Mean Household Income (2010)

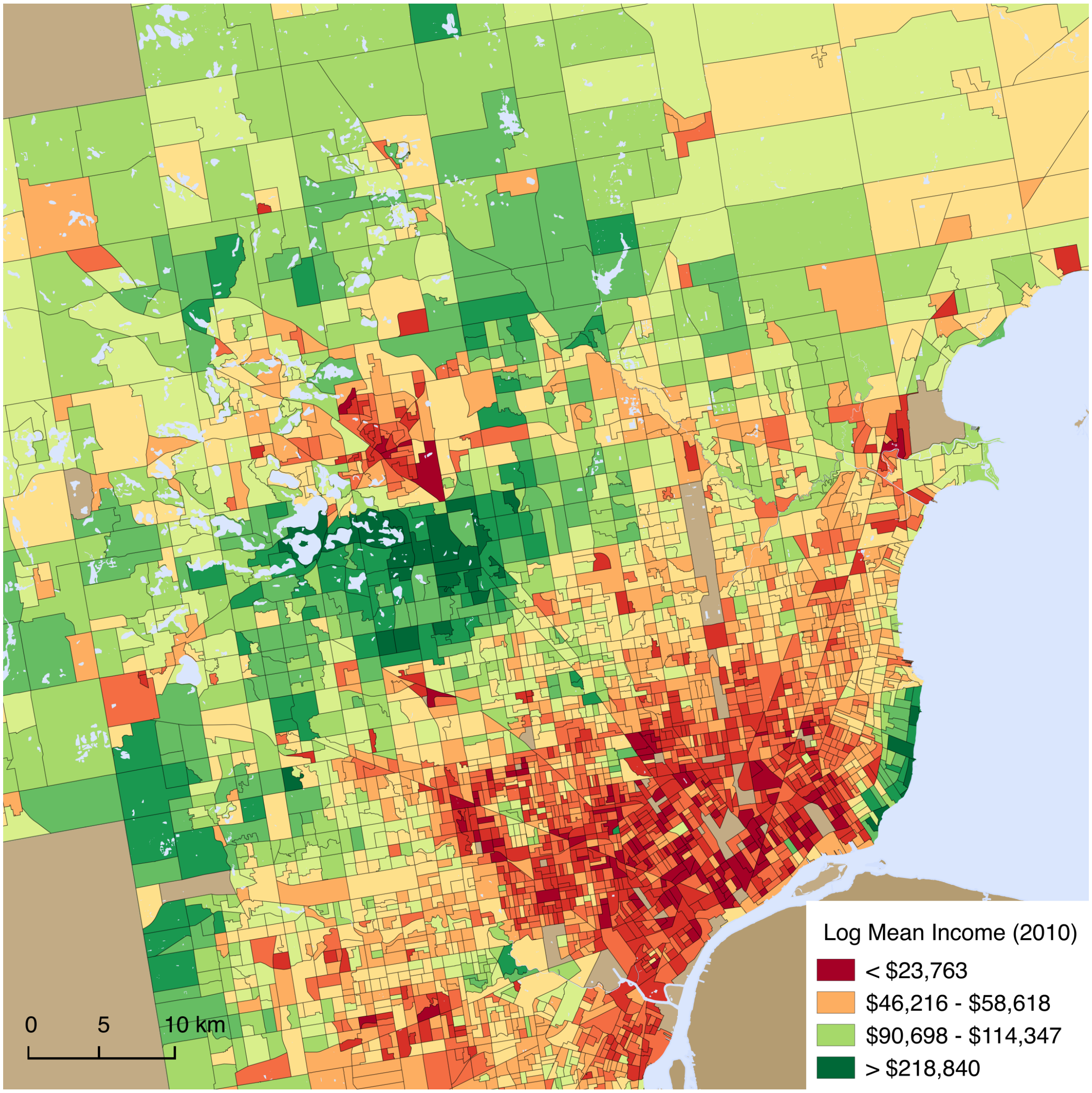


Figure S6: Detroit, MI Mean Household Income (2010)

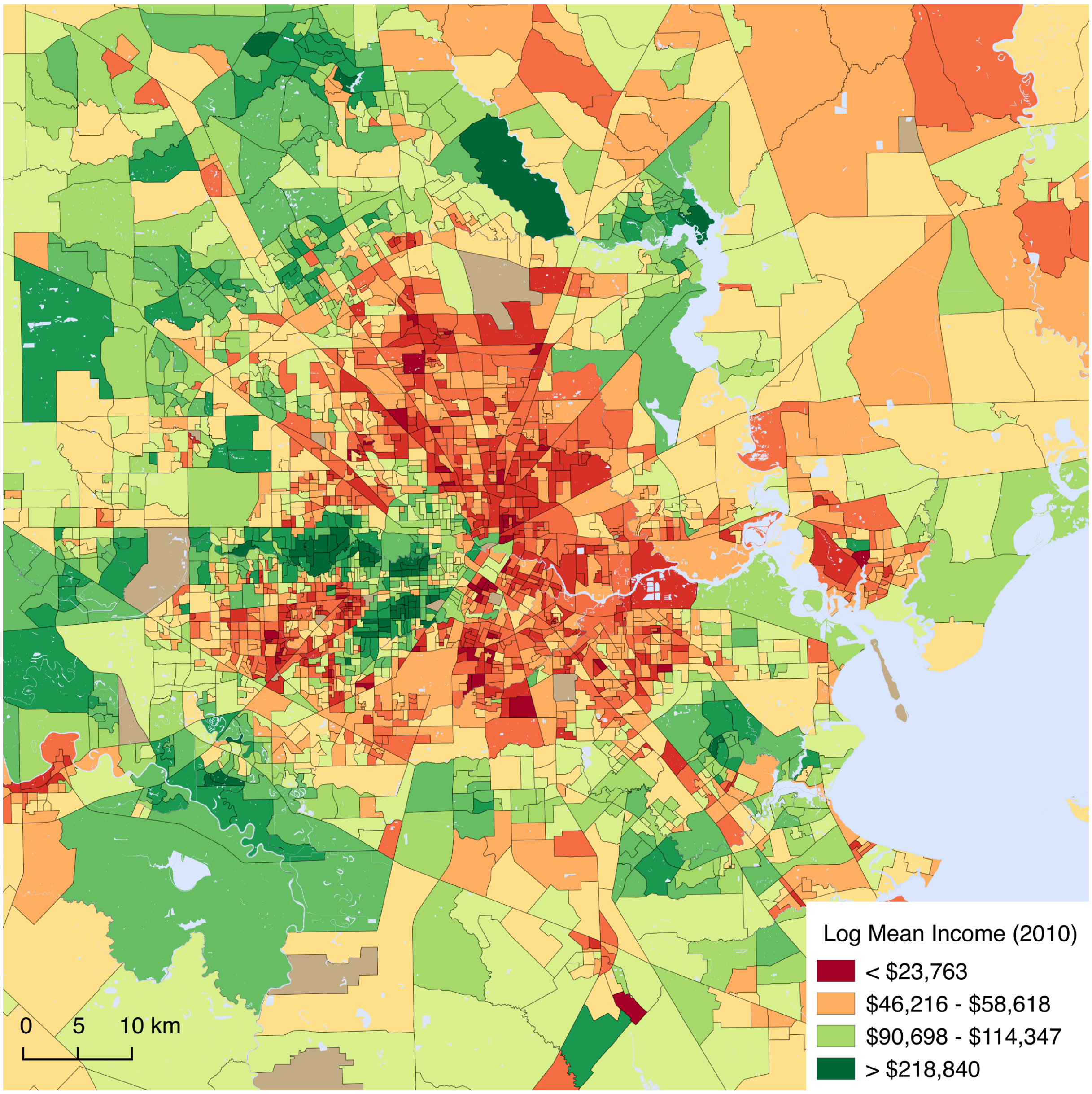


Figure S7: Houston, TX Mean Household Income (2010)

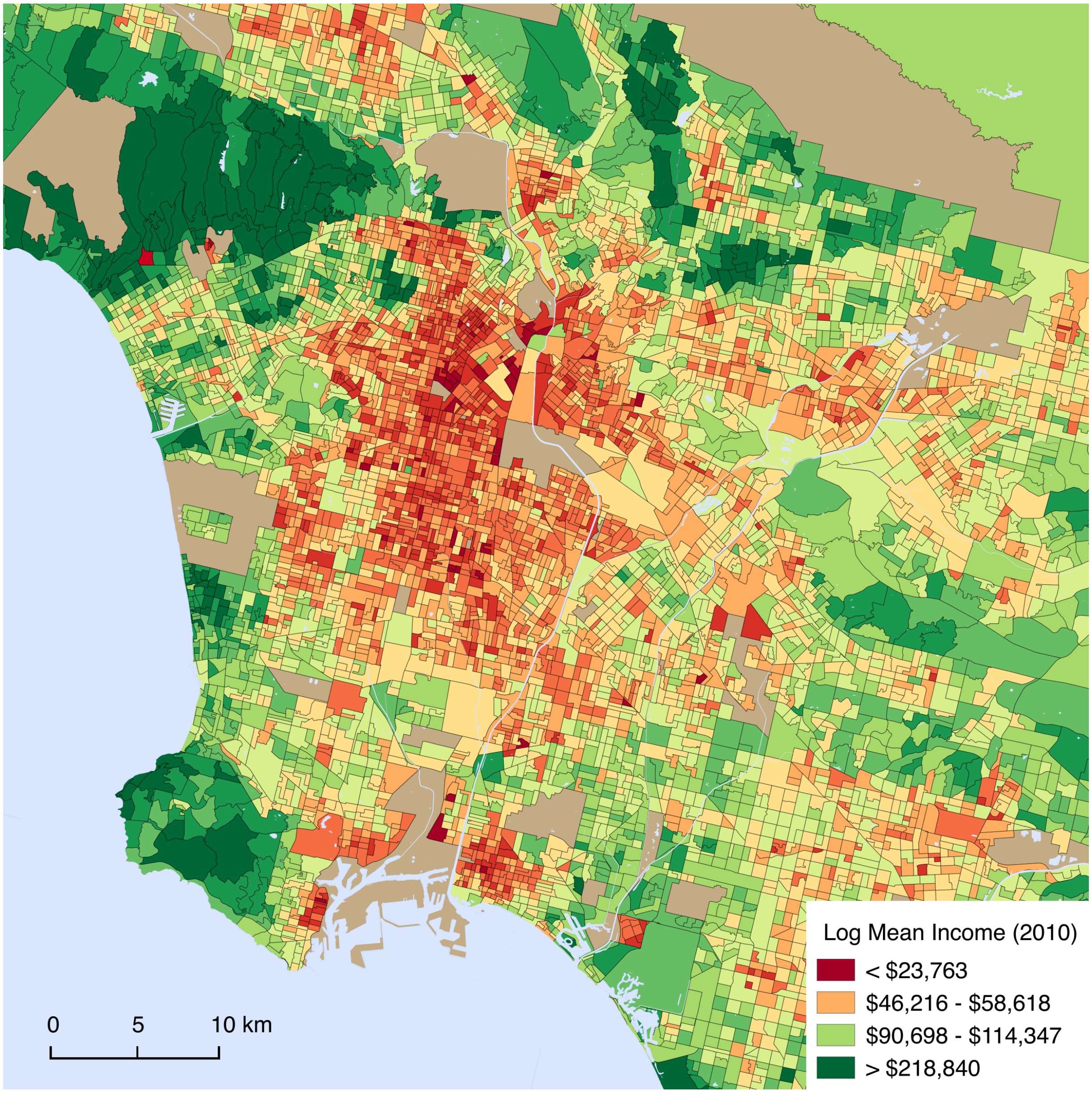


Figure S8: Los Angeles, CA Mean Household Income (2010)

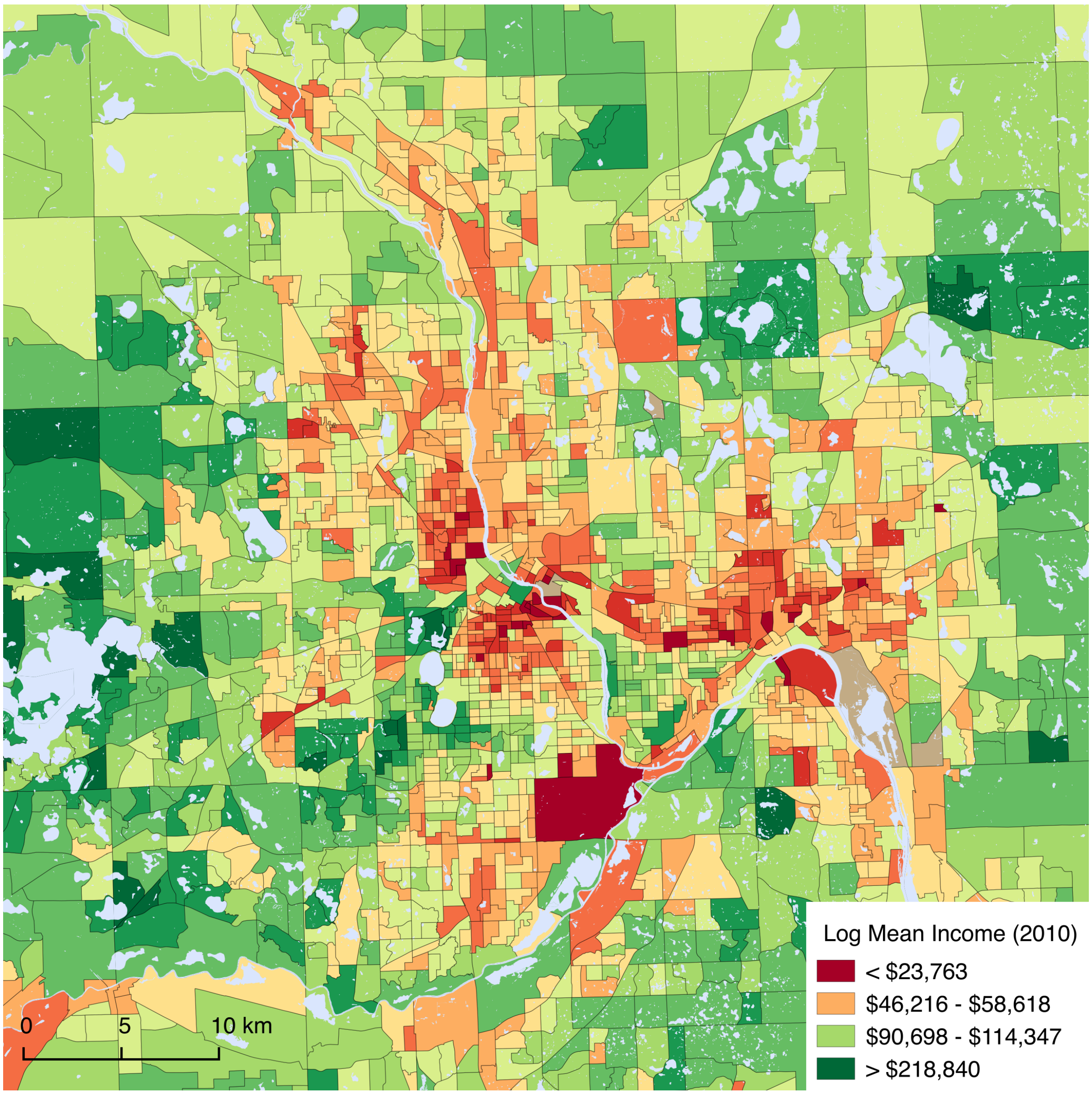


Figure S9: Minneapolis, MN Mean Household Income (2010)

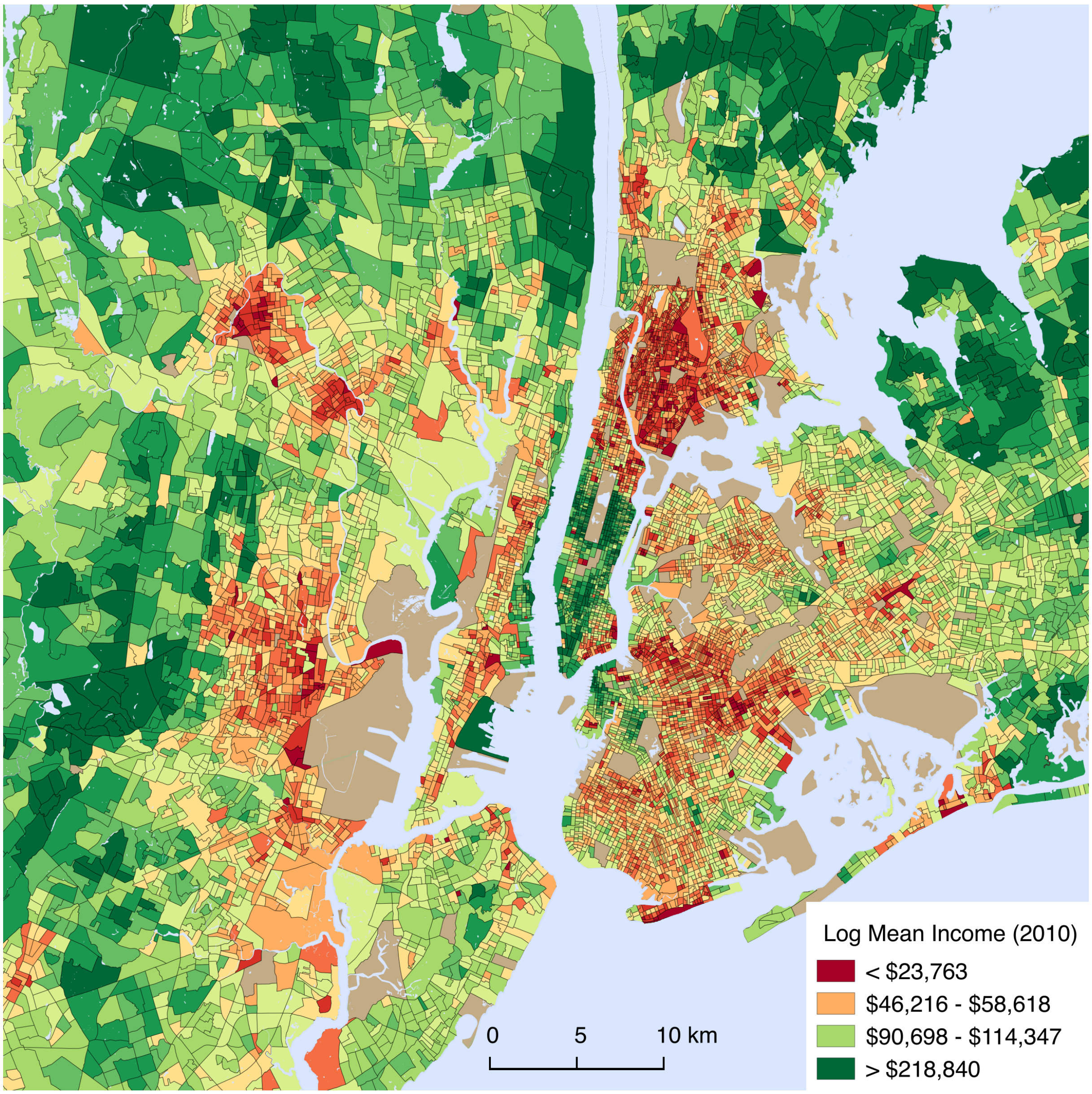


Figure S10: New York City, NY Mean Household Income (2010)

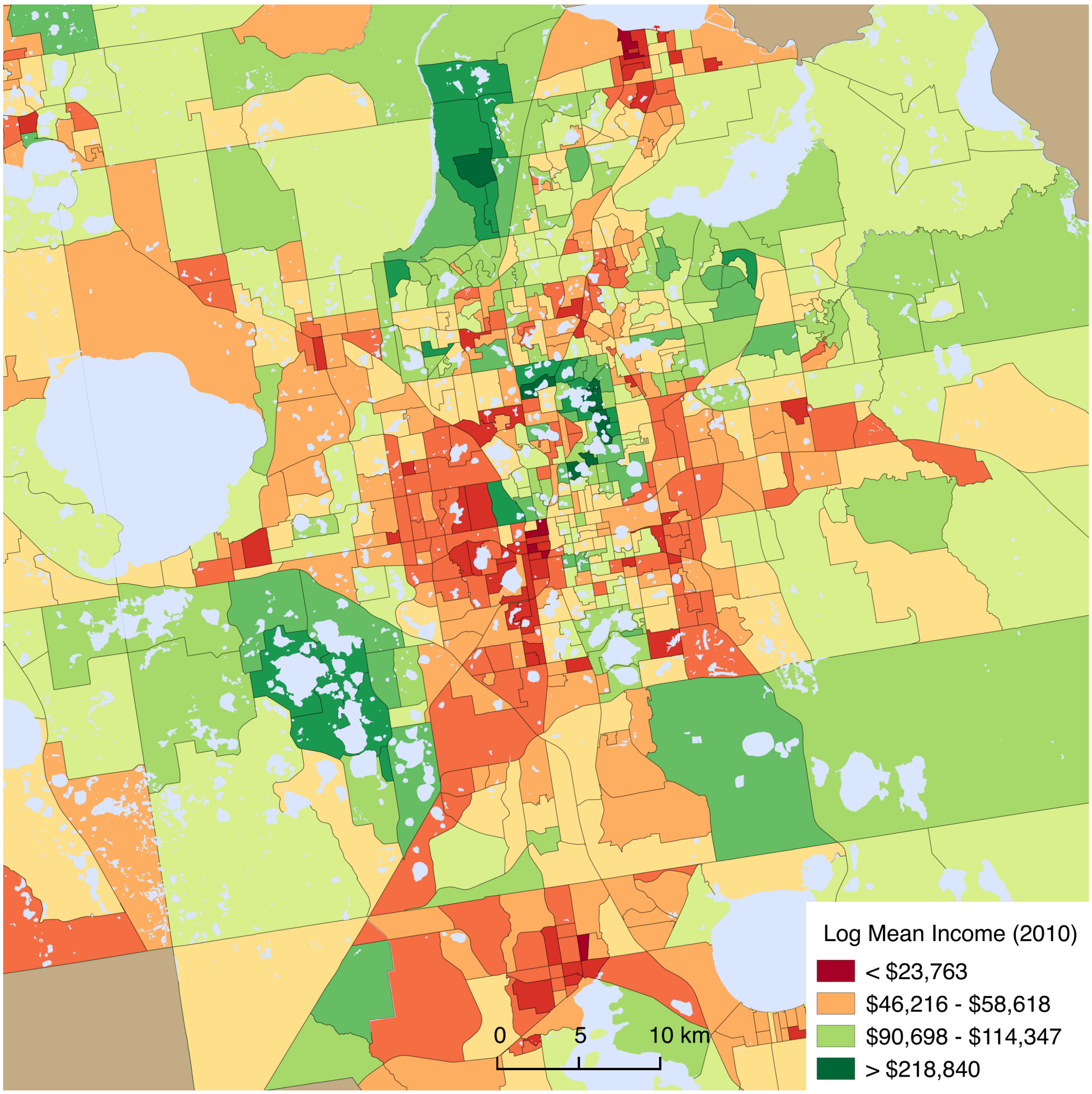


Figure S11: Orlando, FL Mean Household Income (2010)

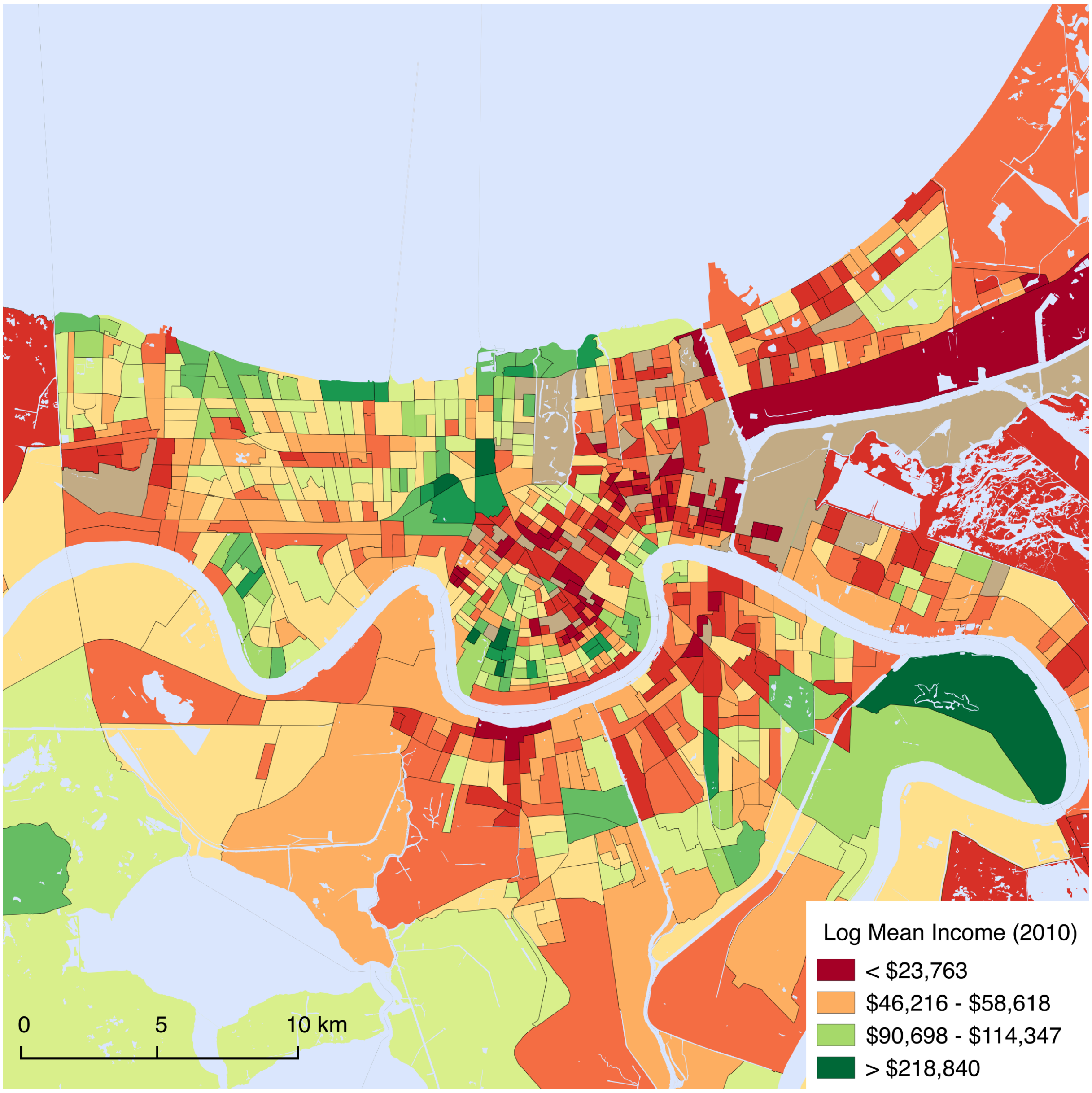


Figure S12: New Orleans, LA Mean Household Income (2010)

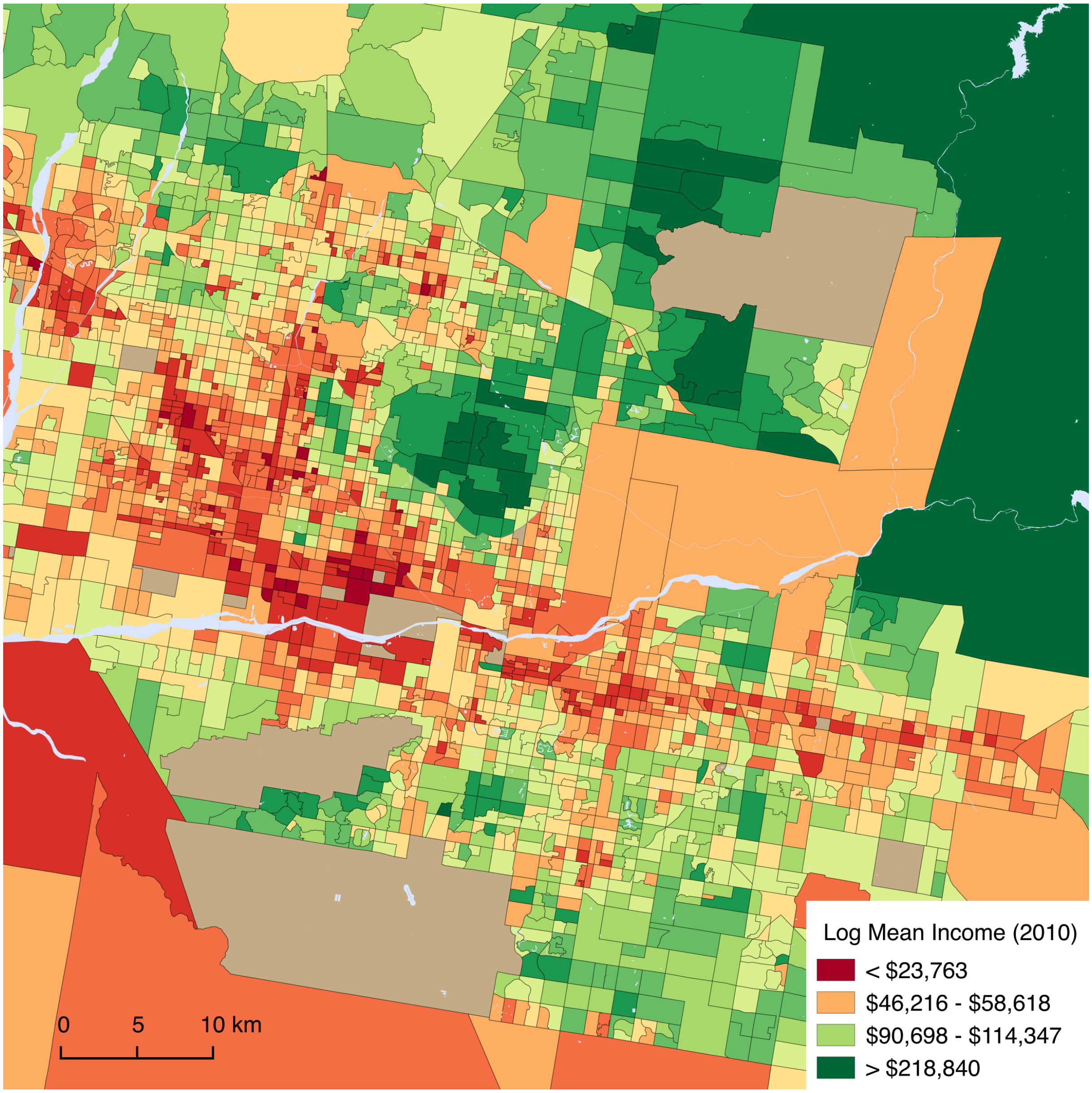


Figure S13: Phoenix, AZ Mean Household Income (2010)

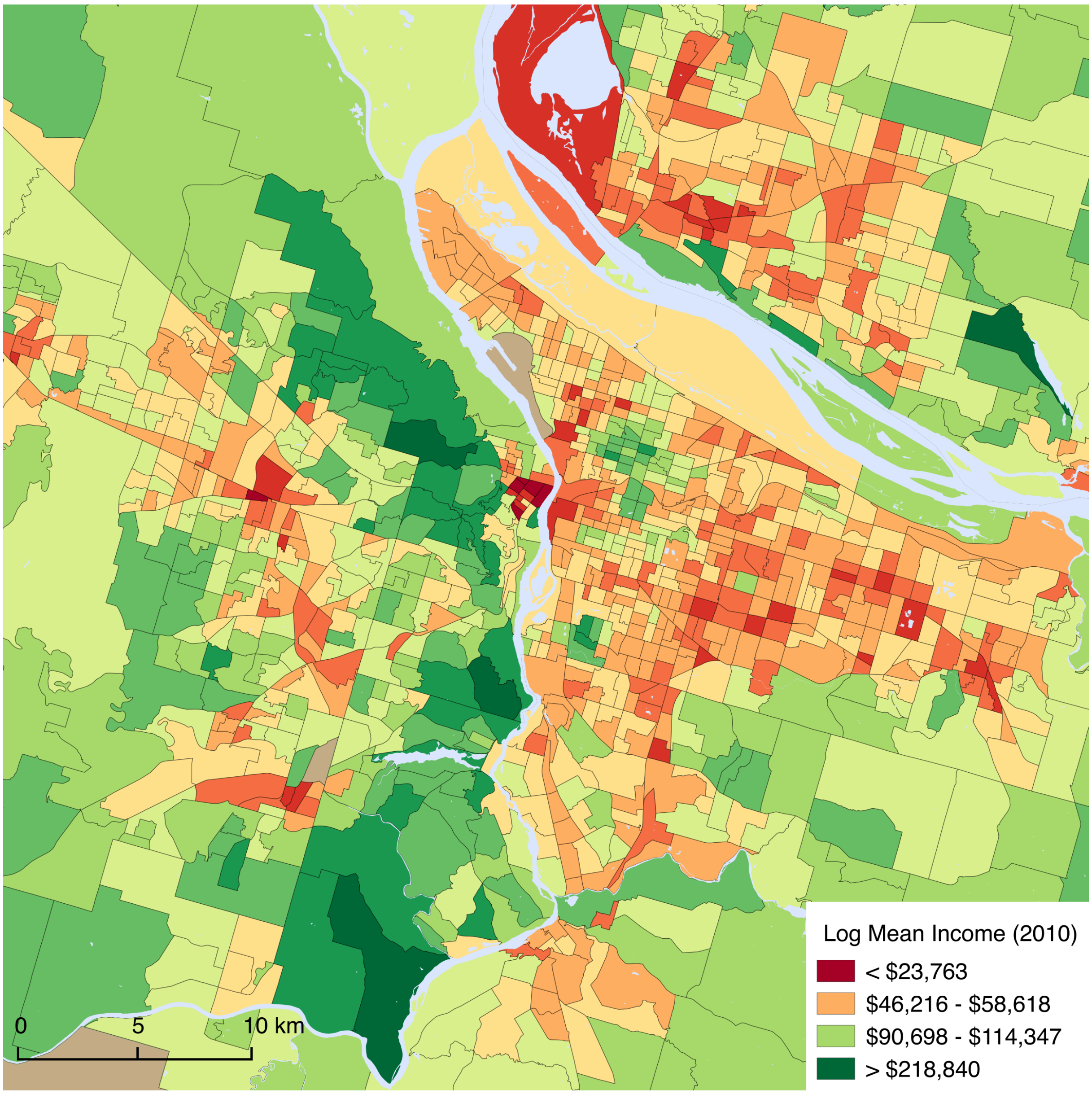


Figure S14: Portland, OR Mean Household Income (2010)

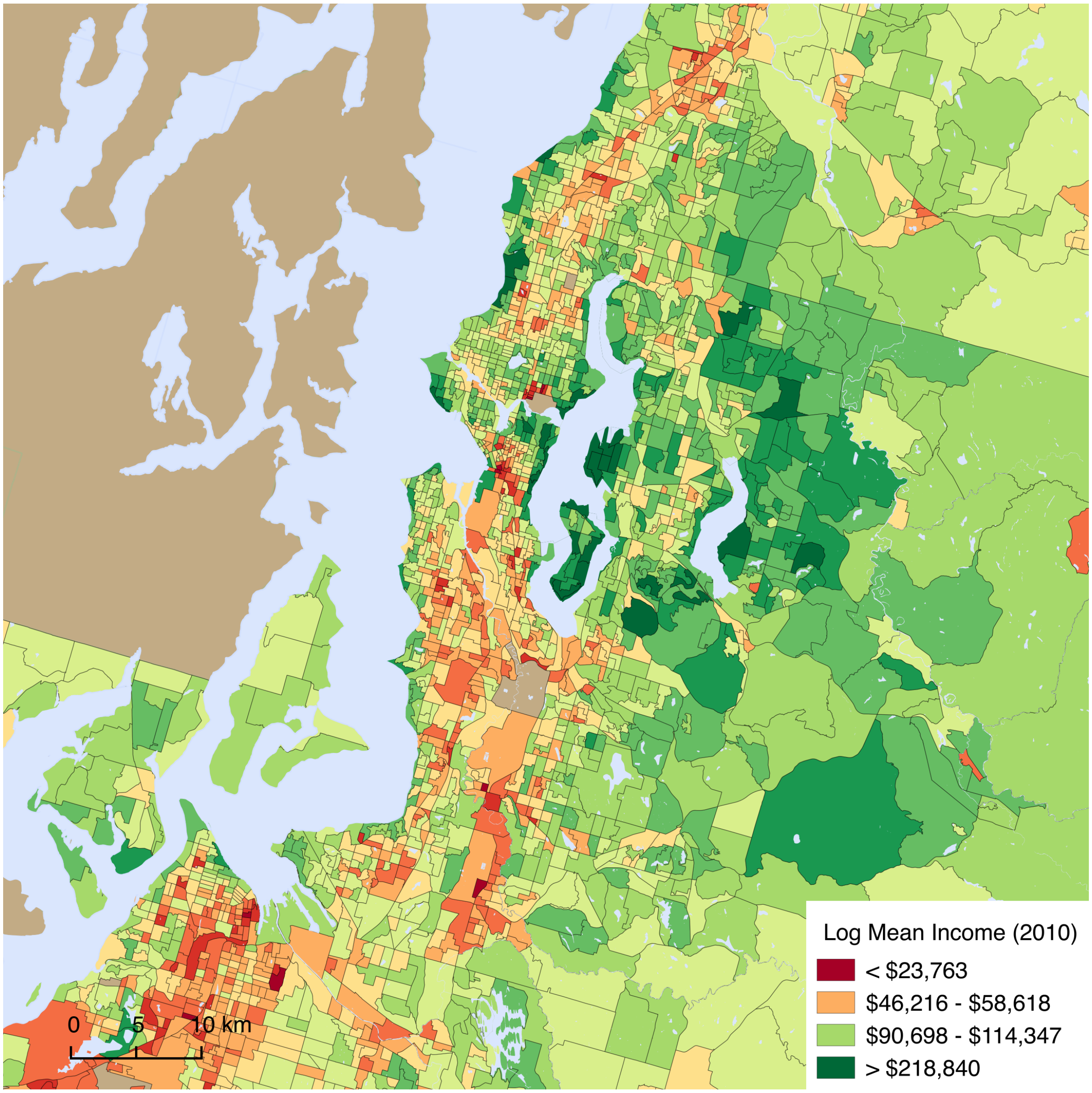


Figure S15: Seattle, WA Mean Household Income (2010)

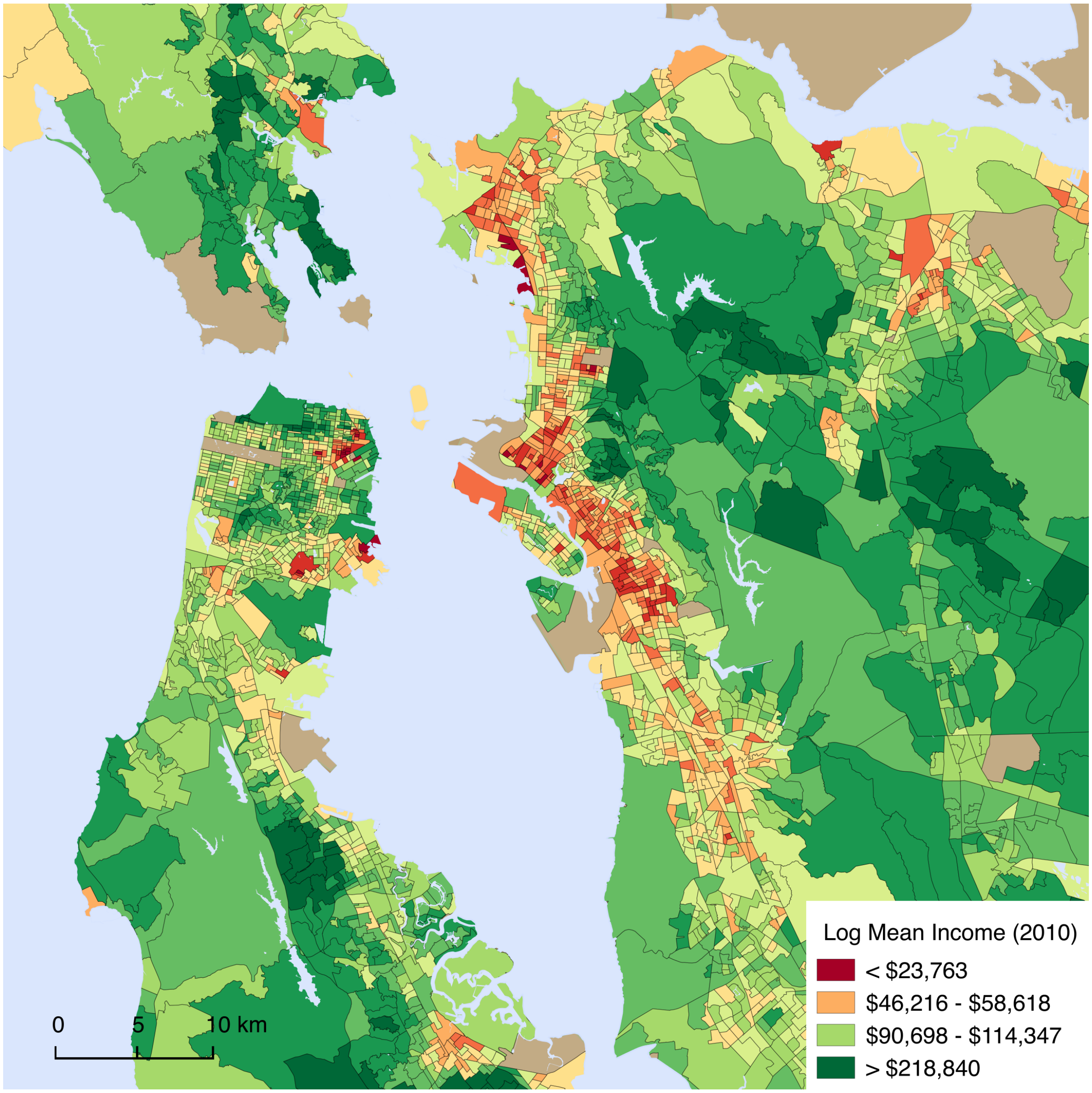


Figure S16: San Francisco, CA Mean Household Income (2010)

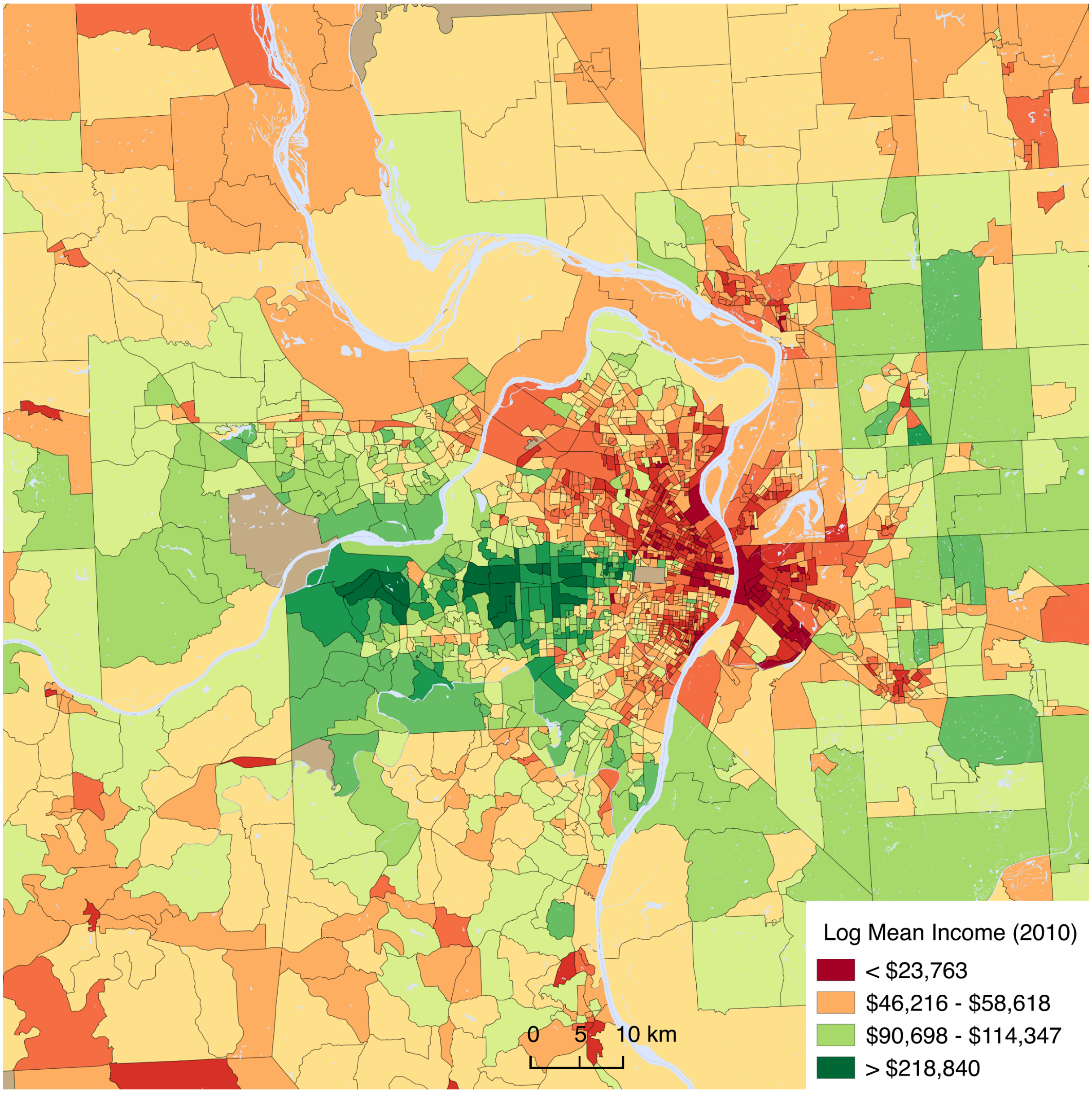


Figure S17: St. Louis, MO Mean Household Income (2010)

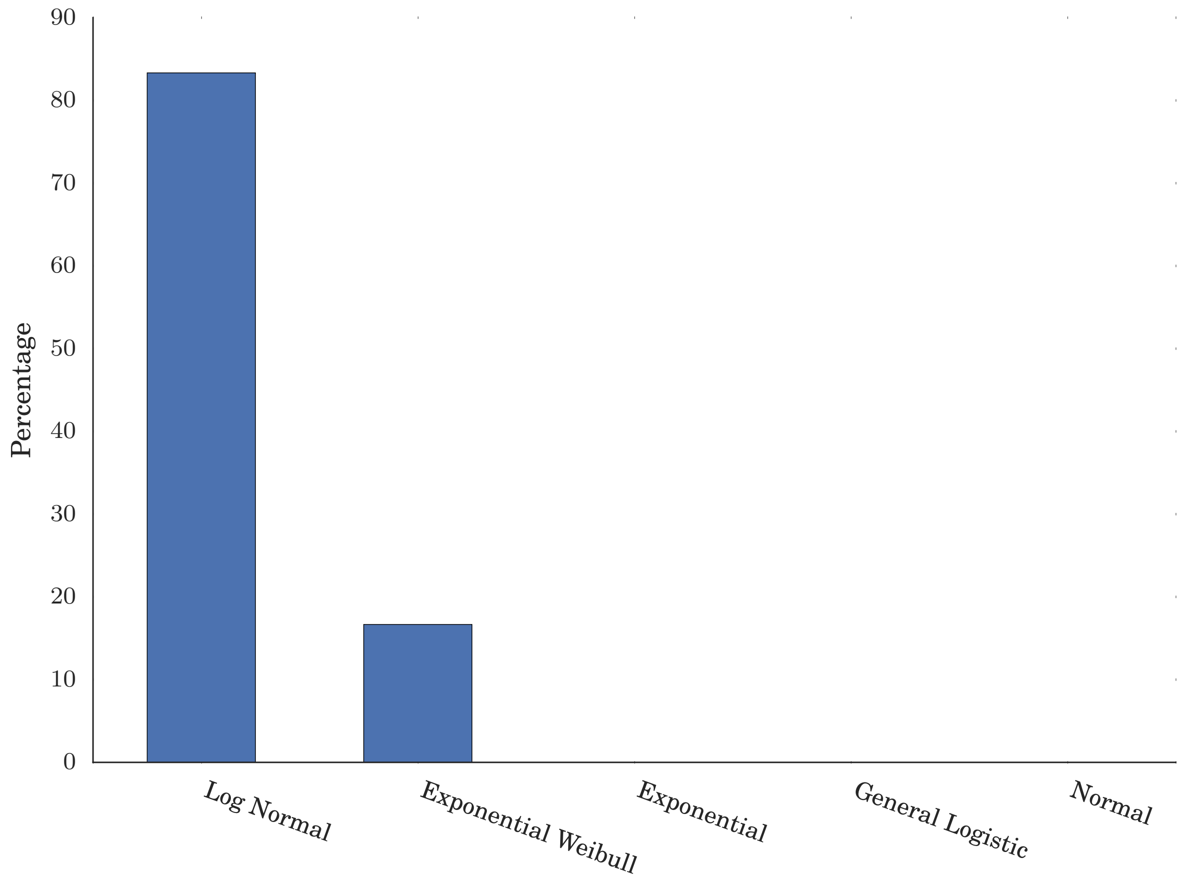


Figure S18: BIC fit test for the distribution of mean income in all CBSA's in the US, first place results.

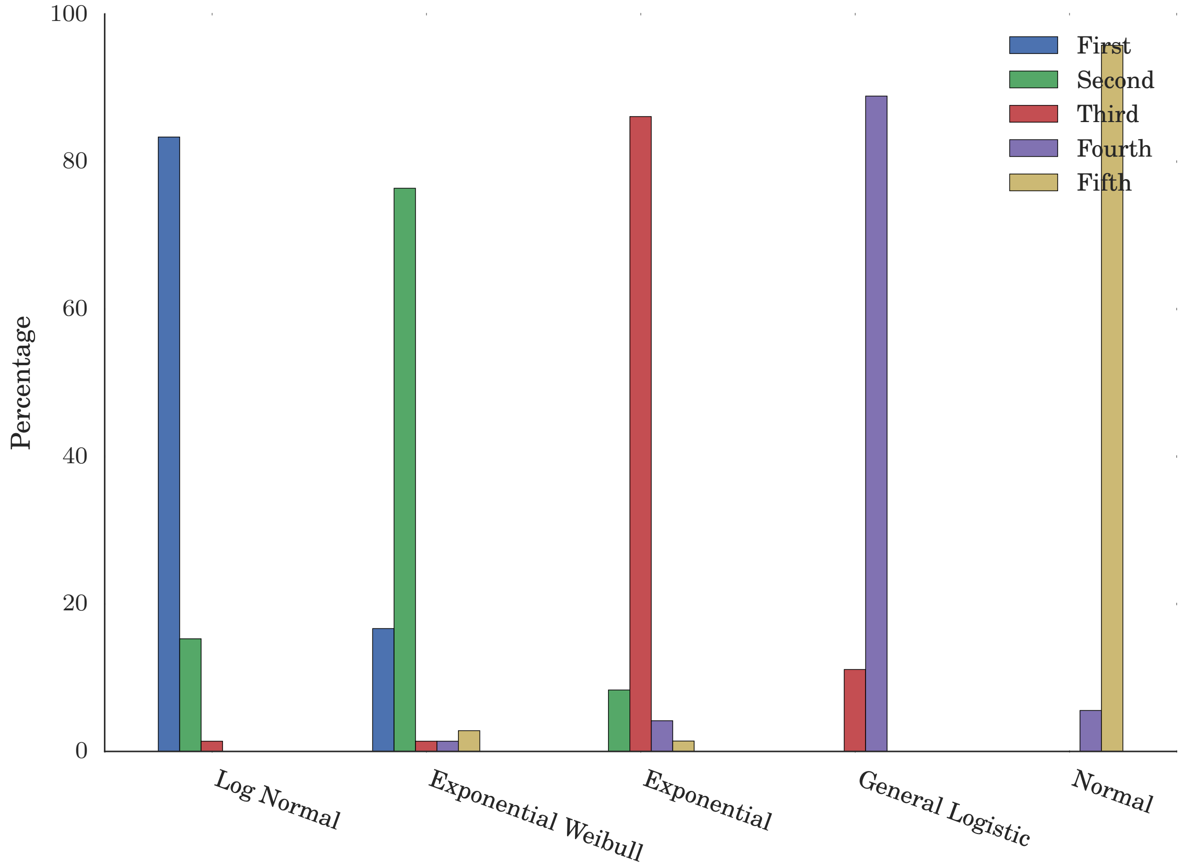


Figure S19: BIC fit test for the distribution of mean income in all CBSA's in the US

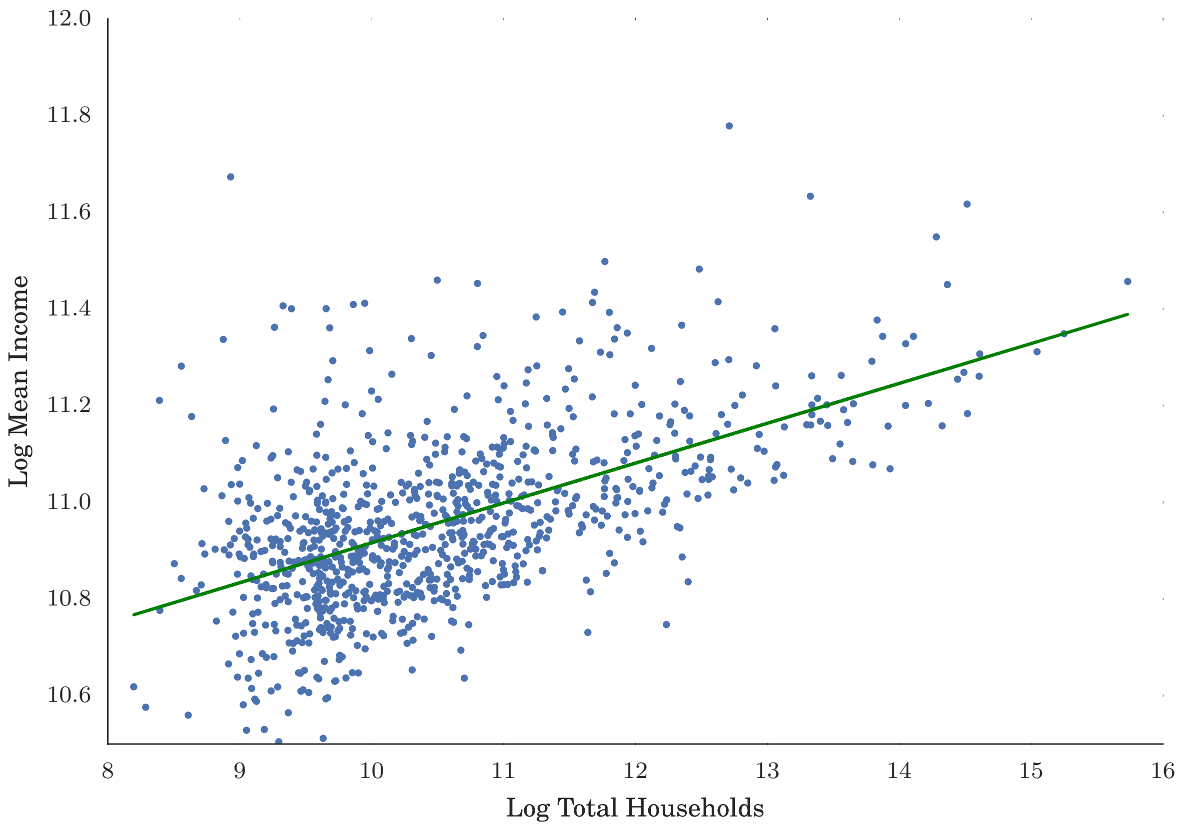


Figure S20: Scaling Mean. The mean income of a city, in relation to number of households, is characterized by an exponent of 0.0825 ($95\% CI = [0.075, 0.090]$, $R^2 = 0.317$).

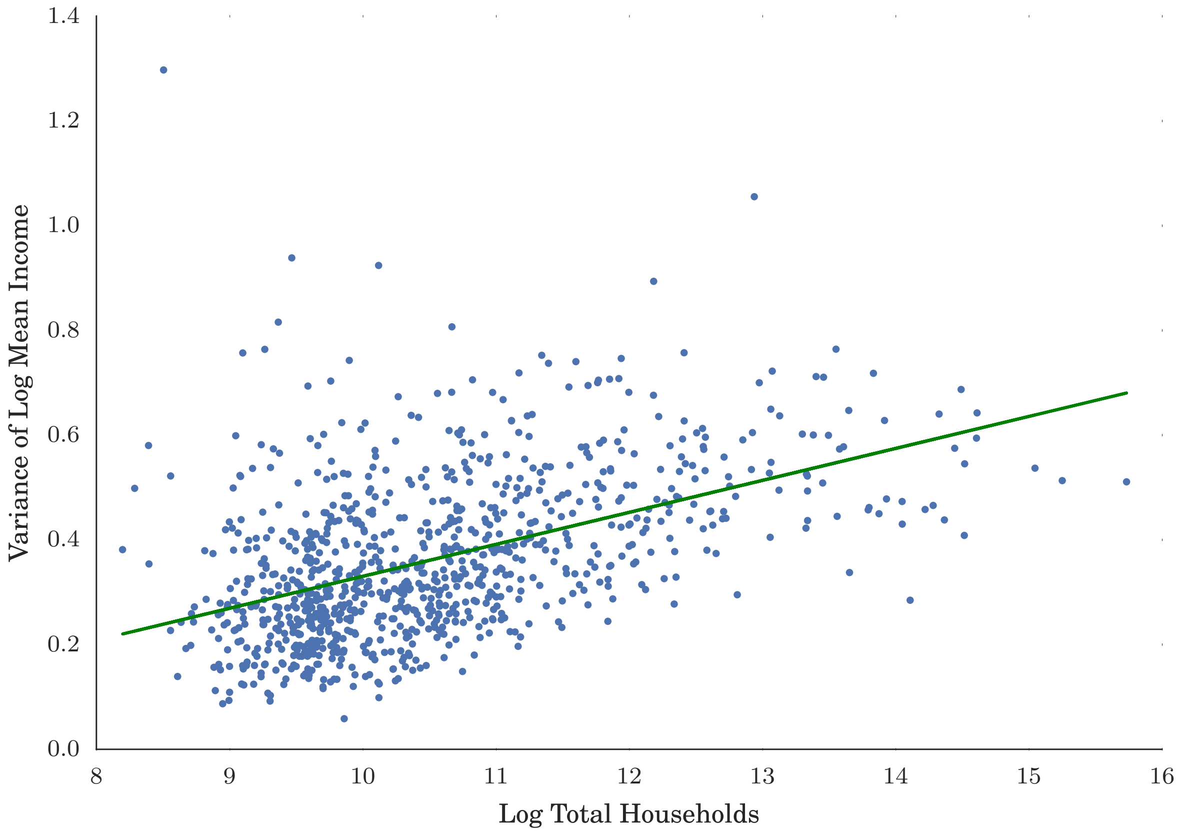


Figure S21: Scaling Variance. The variance of block group income for a city, in relation to number of households, is characterized by an exponent of 0.0611 ($95\% CI = [0.054, 0.068], R^2 = 0.235$).

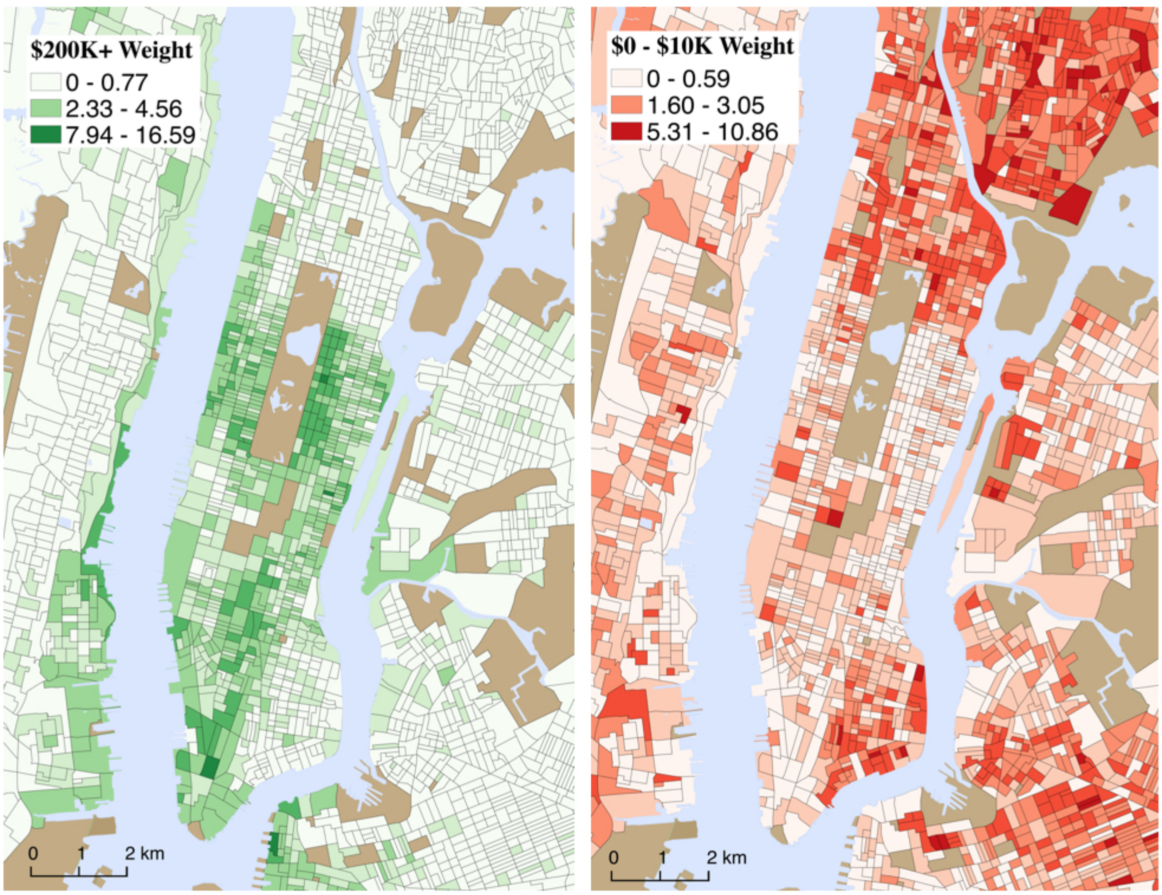


Figure S22: The weights $w_{\ell,j}$ for the richest (left, in green) and poorest (right, in red) income bins for New York City.

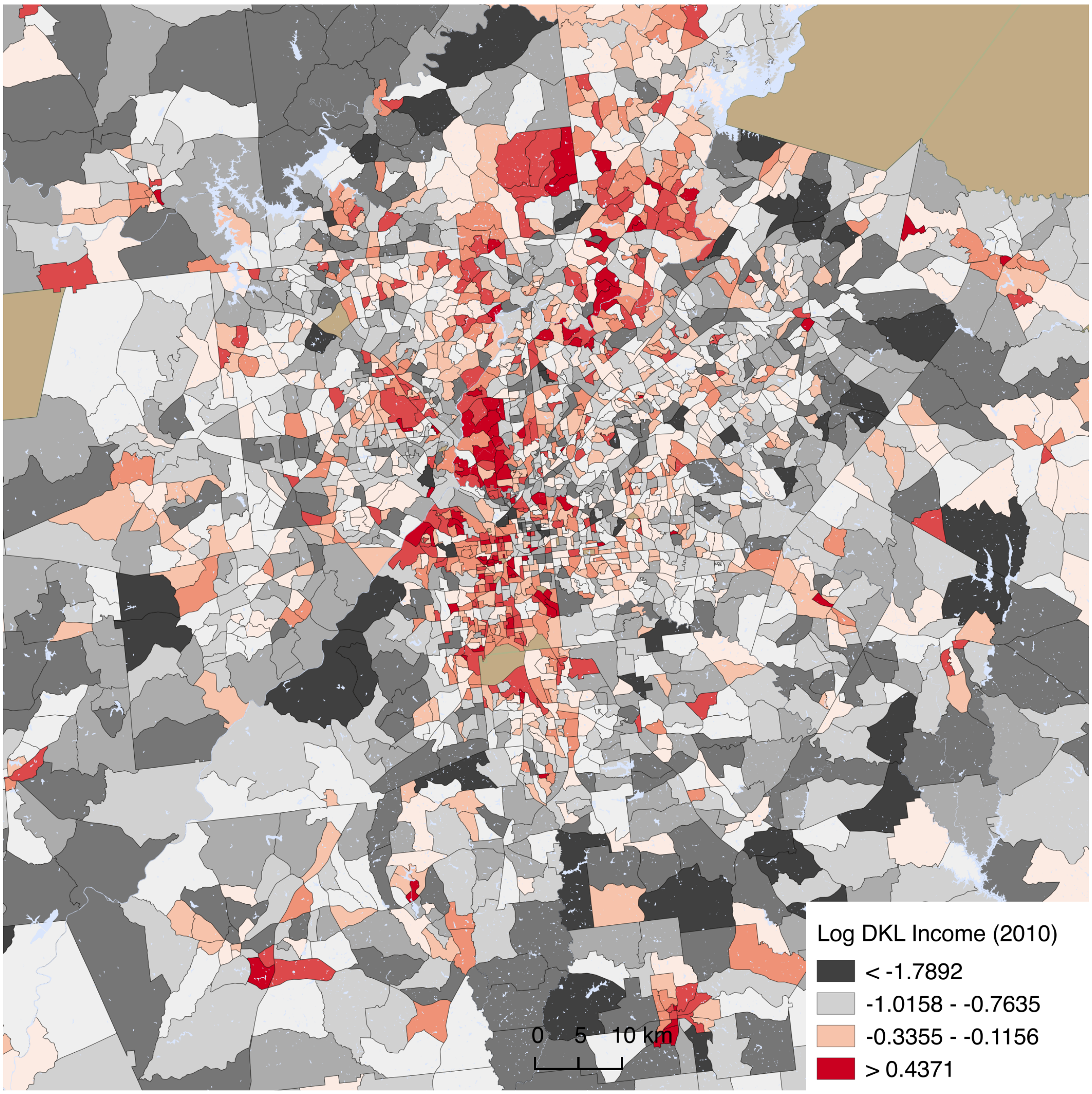


Figure S23: Atlanta, GA D_{KL} (2010)

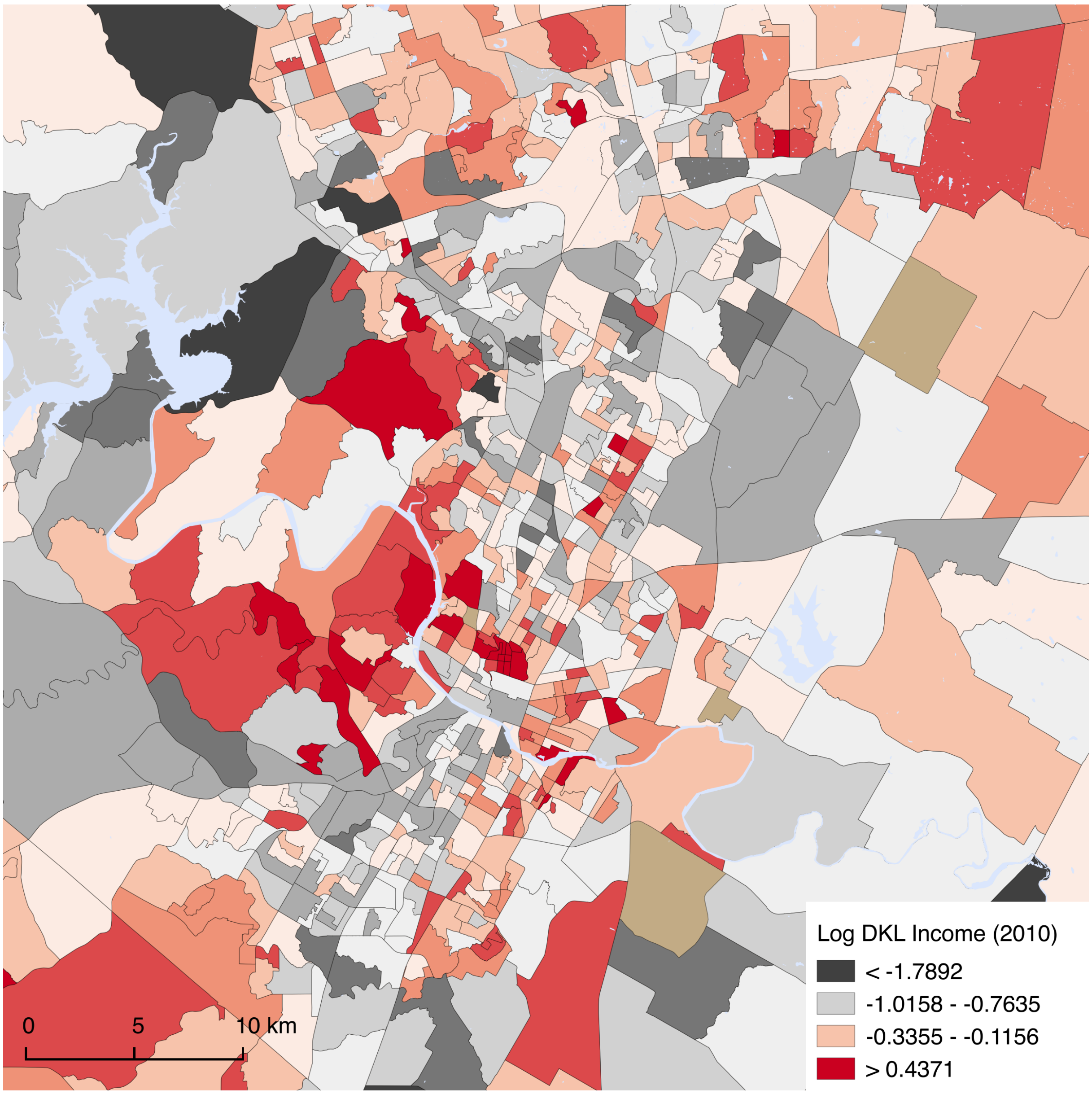


Figure S24: Austin, TX D_{KL} (2010)

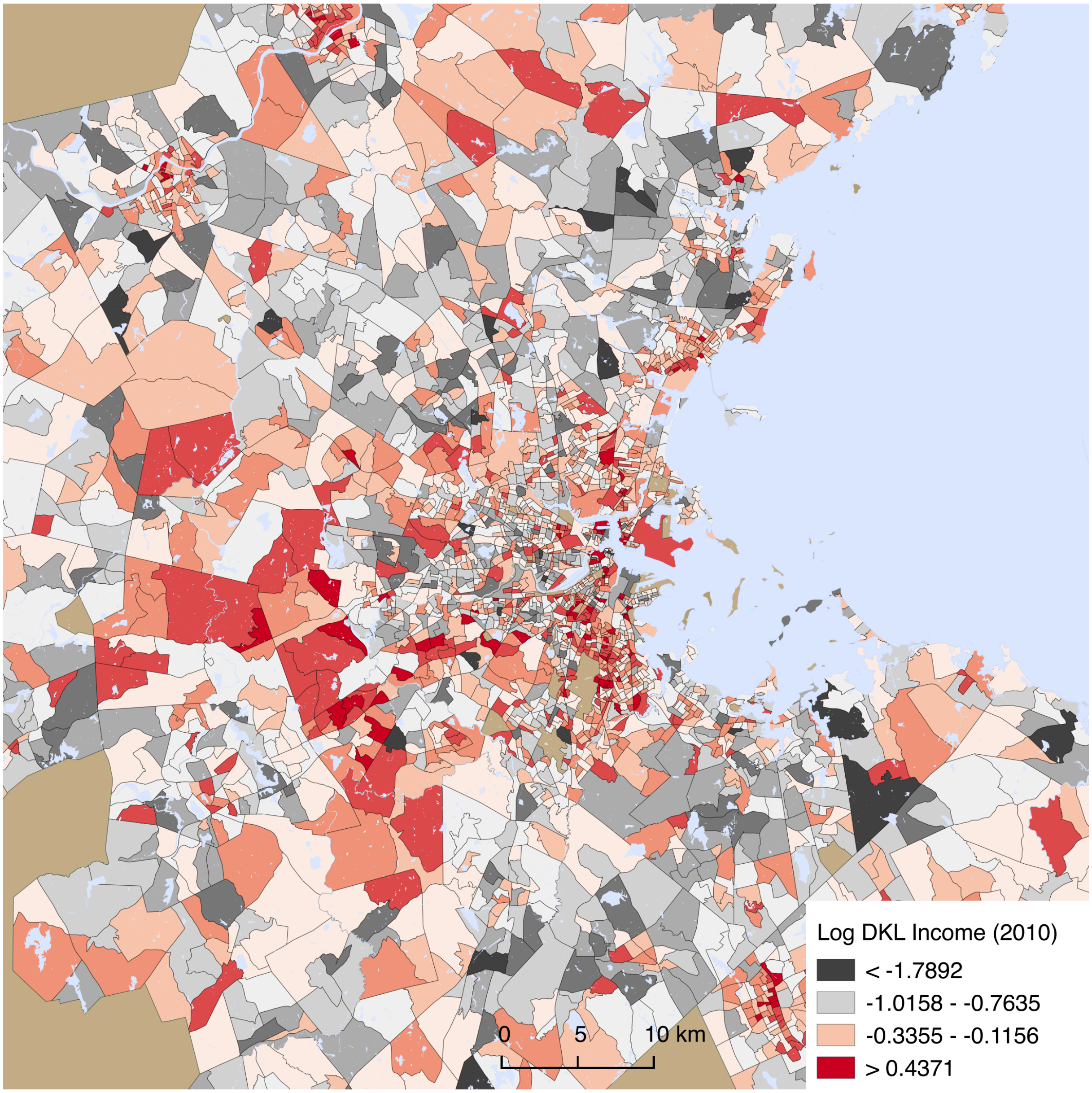


Figure S25: Boston, MA D_{KL} (2010)

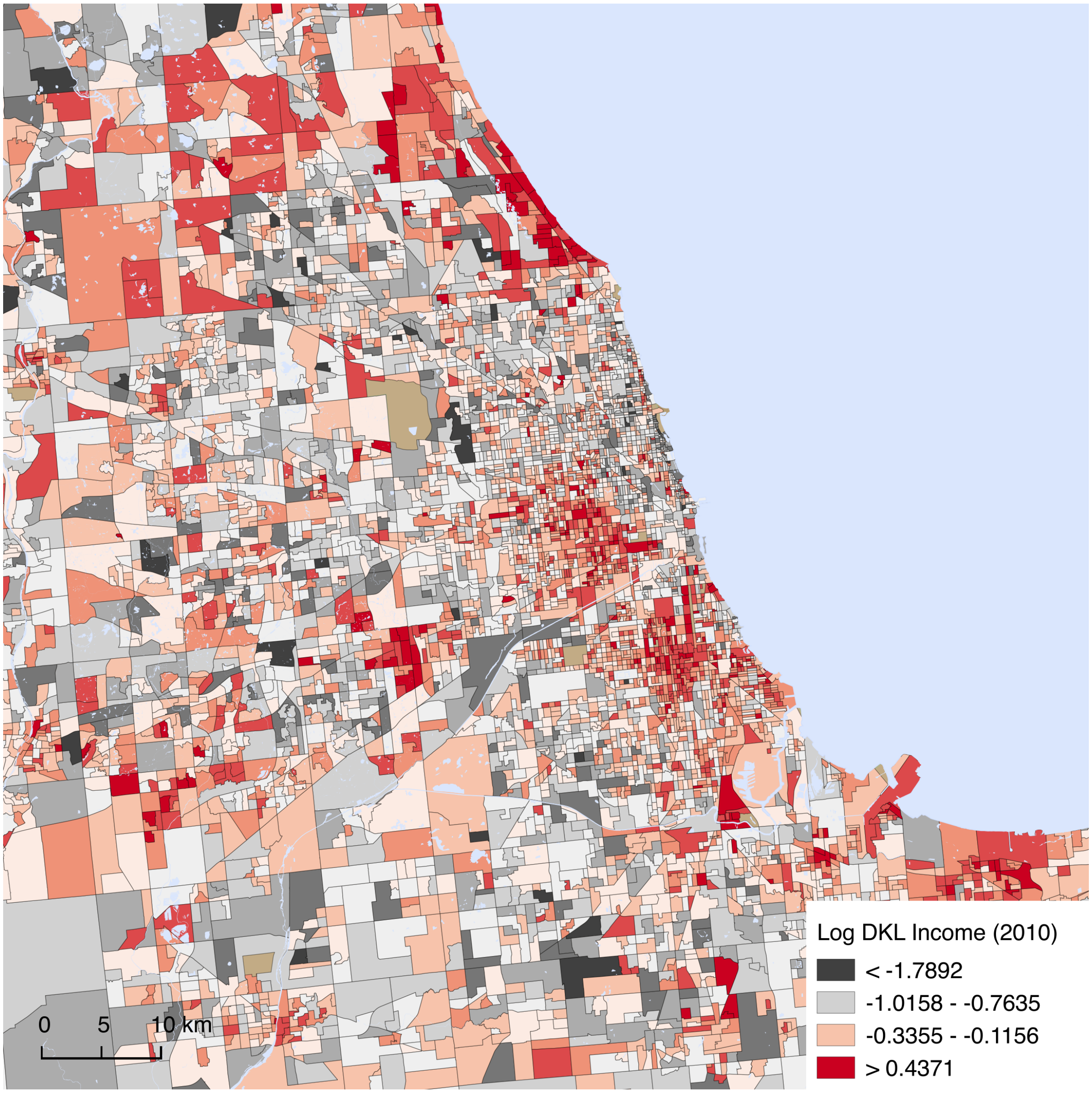


Figure S26: Chicago, IL D_{KL} (2010)

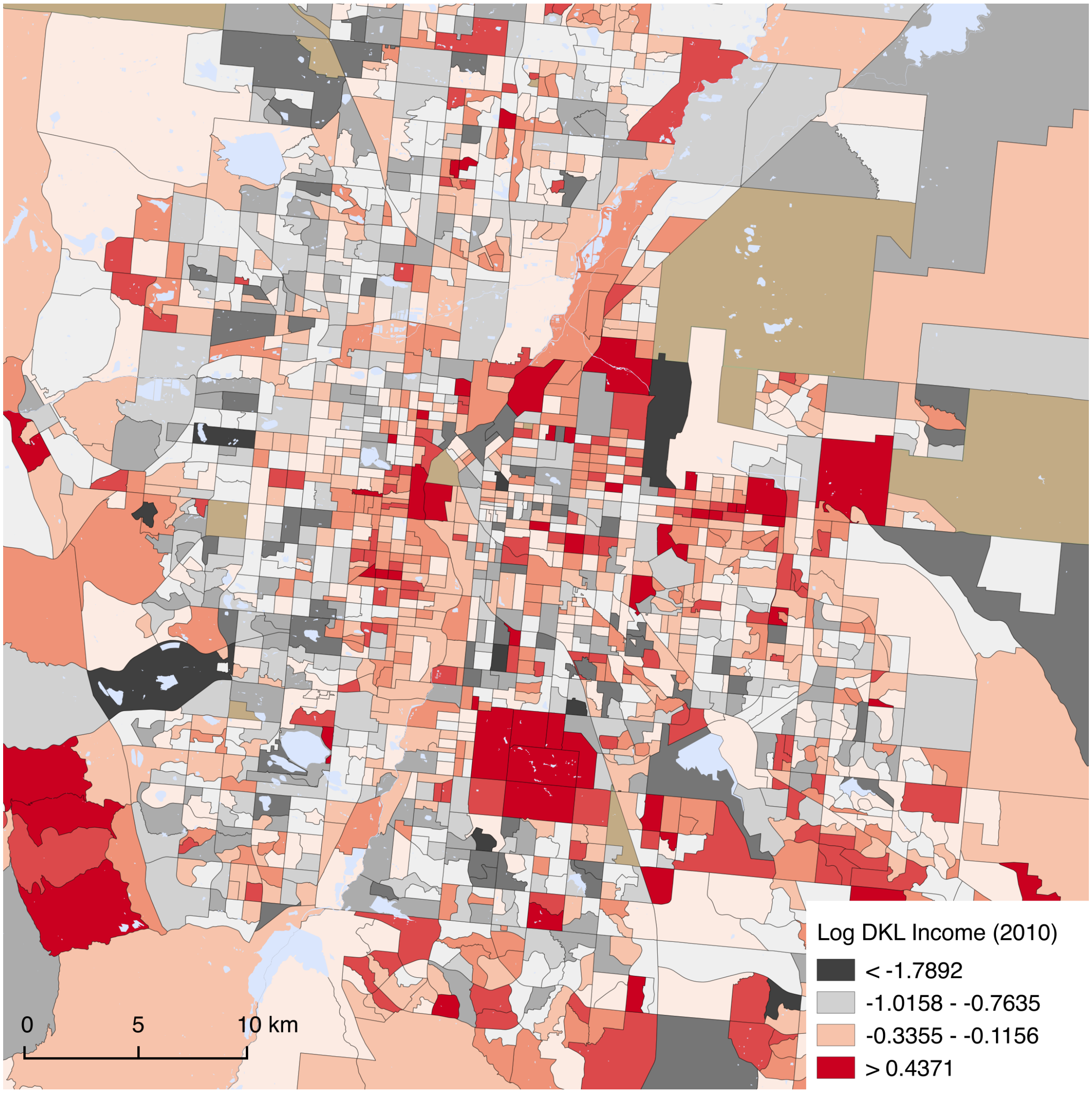


Figure S27: Denver, CO D_{KL} (2010)

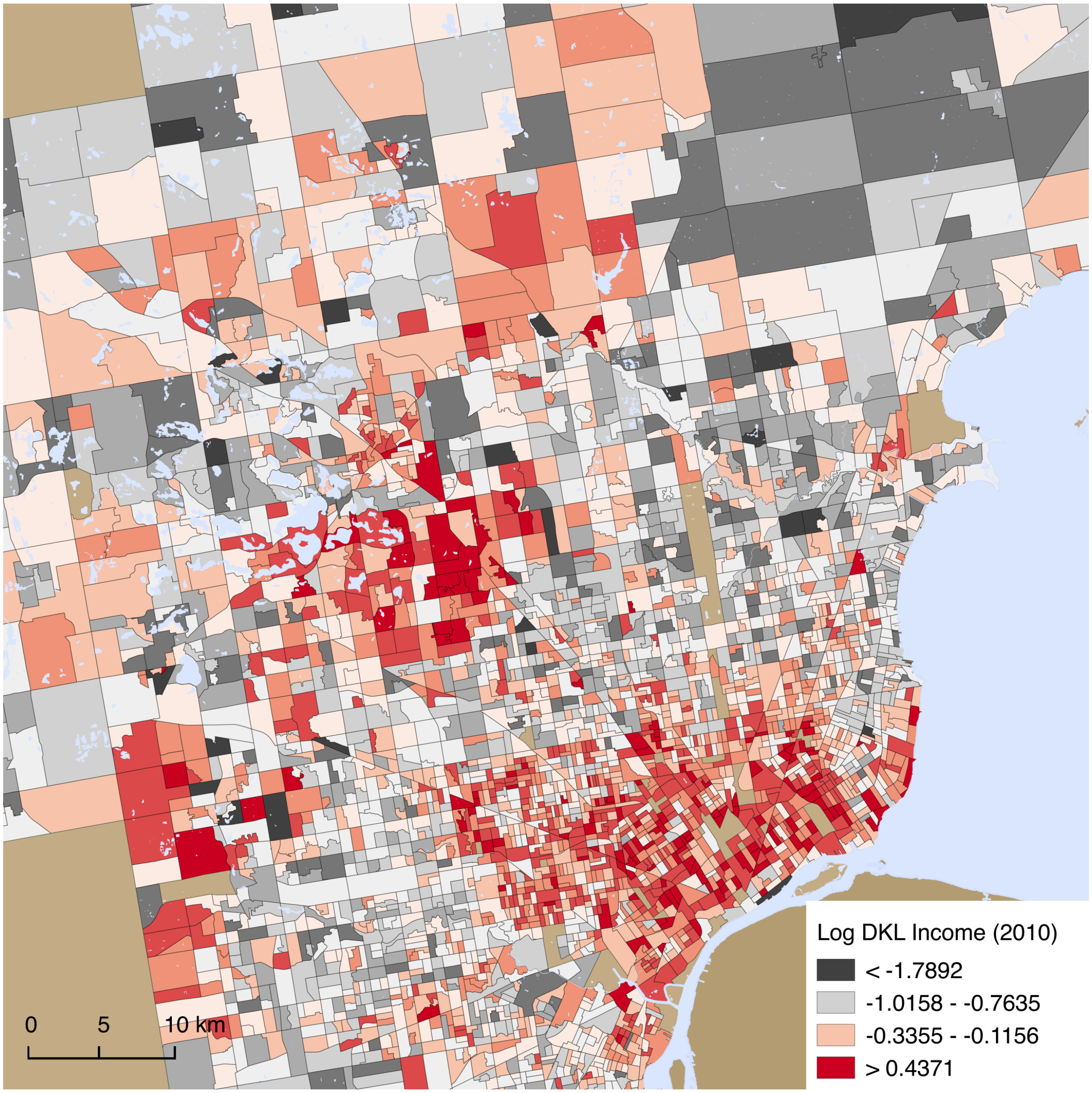


Figure S28: Detroit, MI D_{KL} (2010)

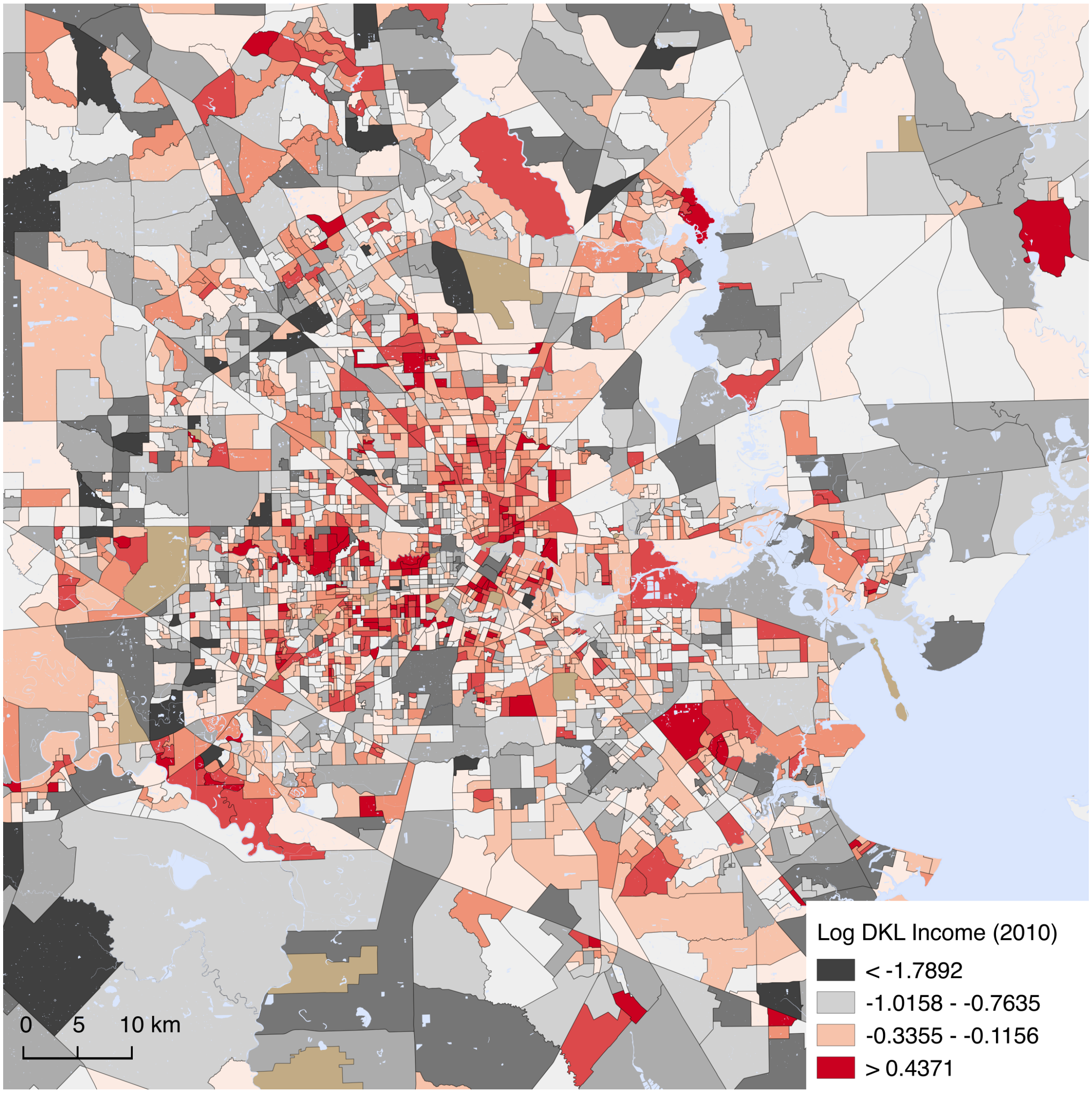


Figure S29: Houston, TX D_{KL} (2010)

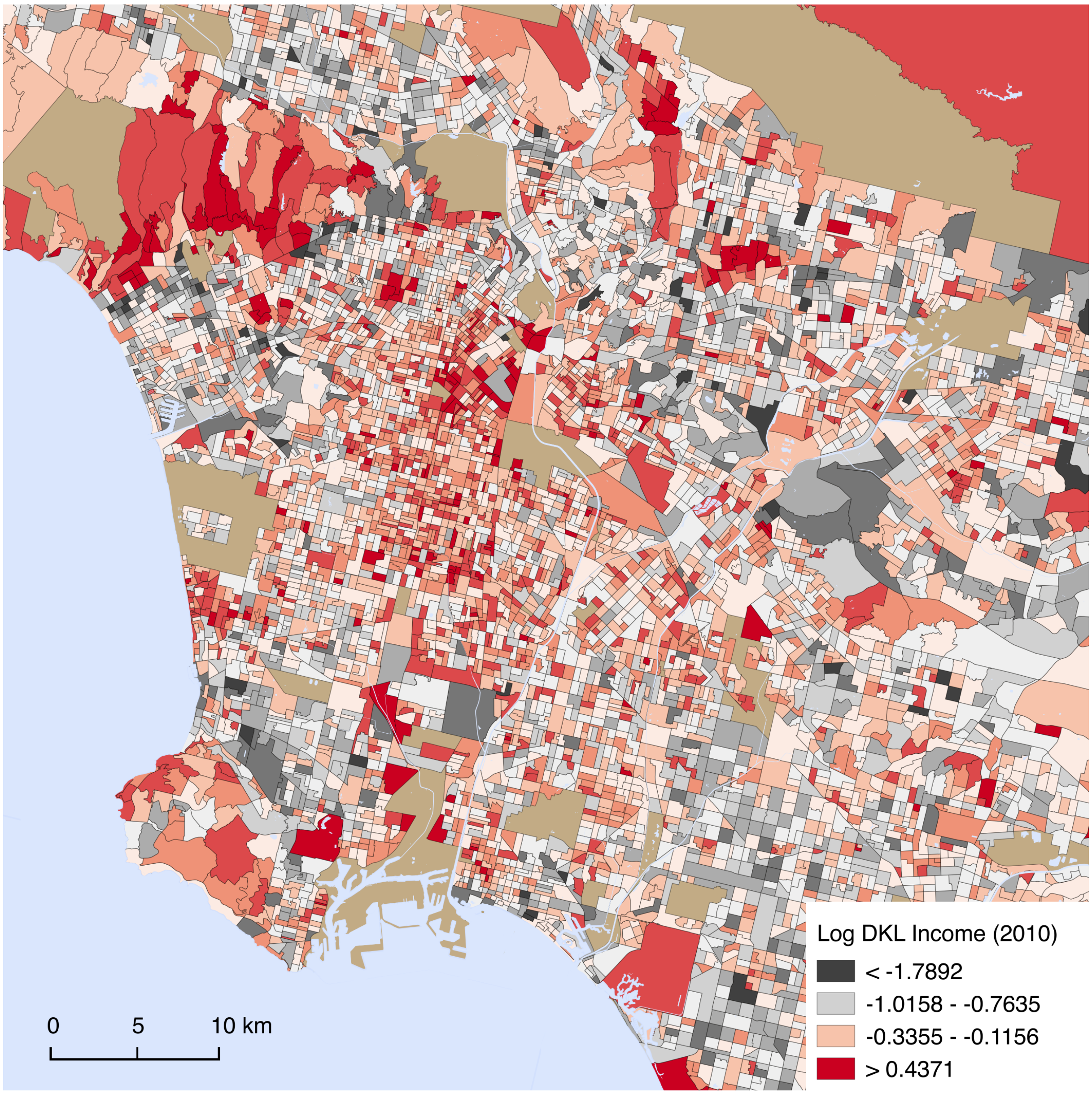


Figure S30: Los Angeles, CA D_{KL} (2010)

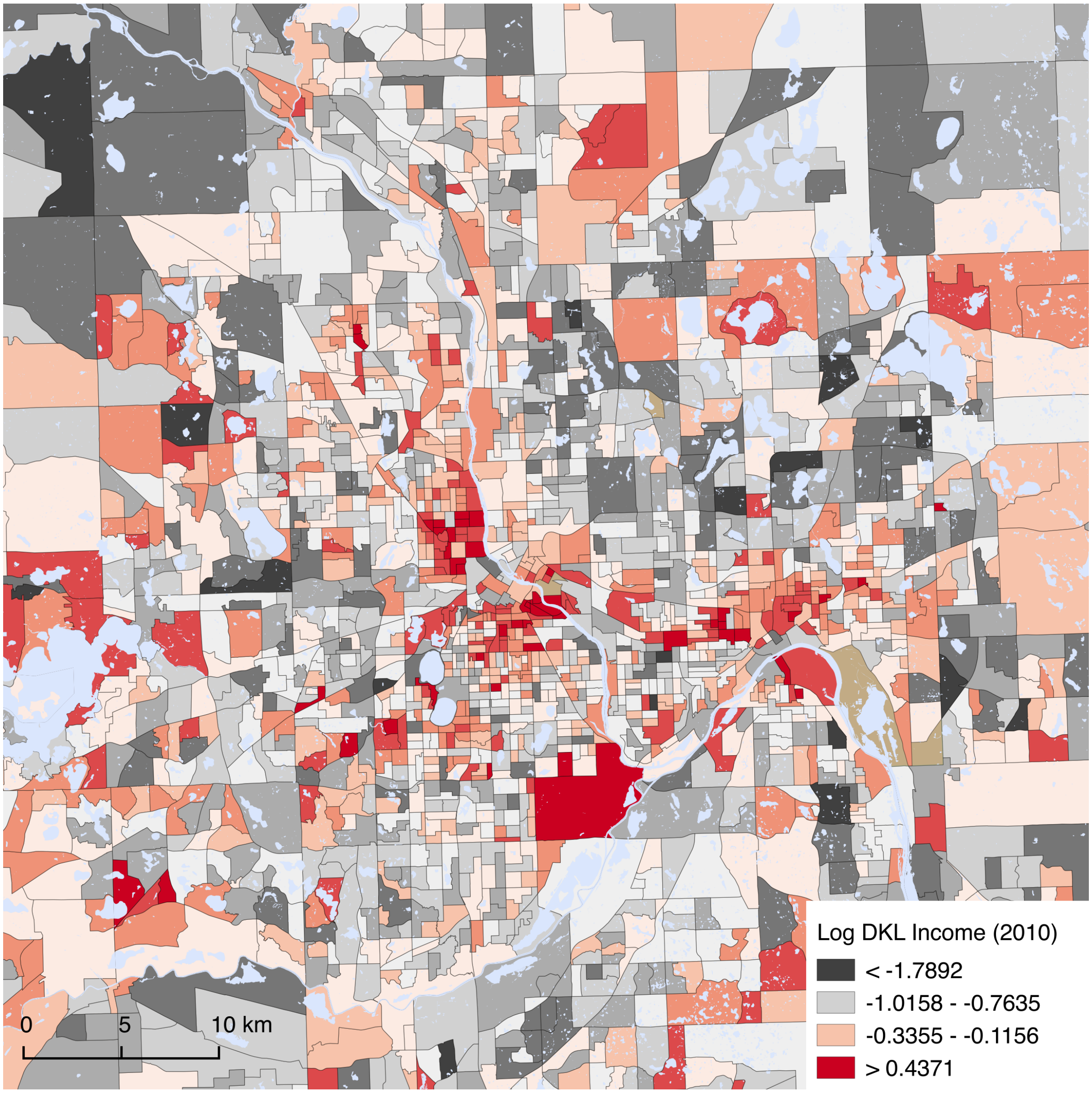


Figure S31: Minneapolis, MN D_{KL} (2010)

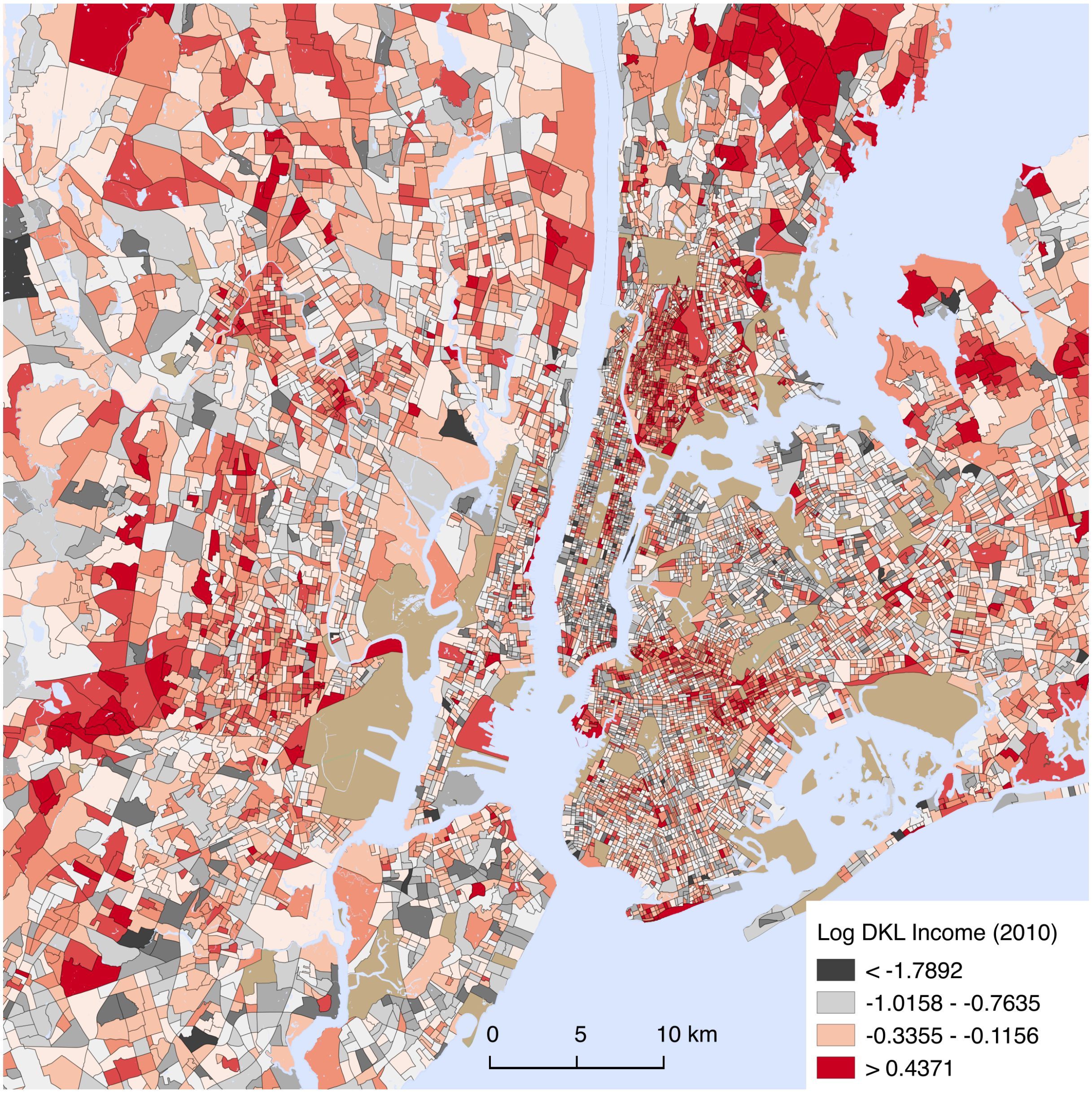


Figure S32: New York City, NY D_{KL} (2010)

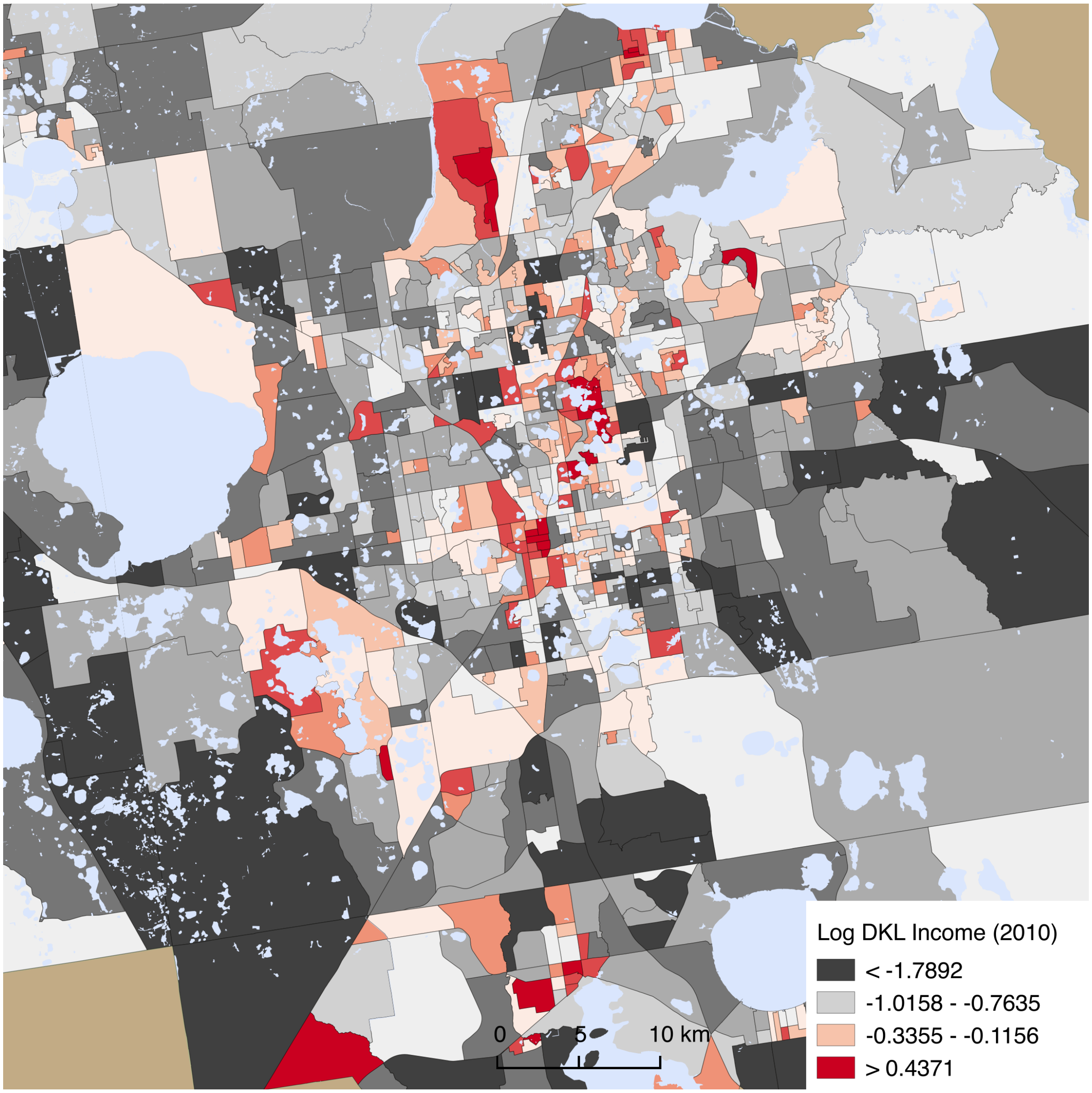


Figure S33: Orlando, FL D_{KL} (2010)

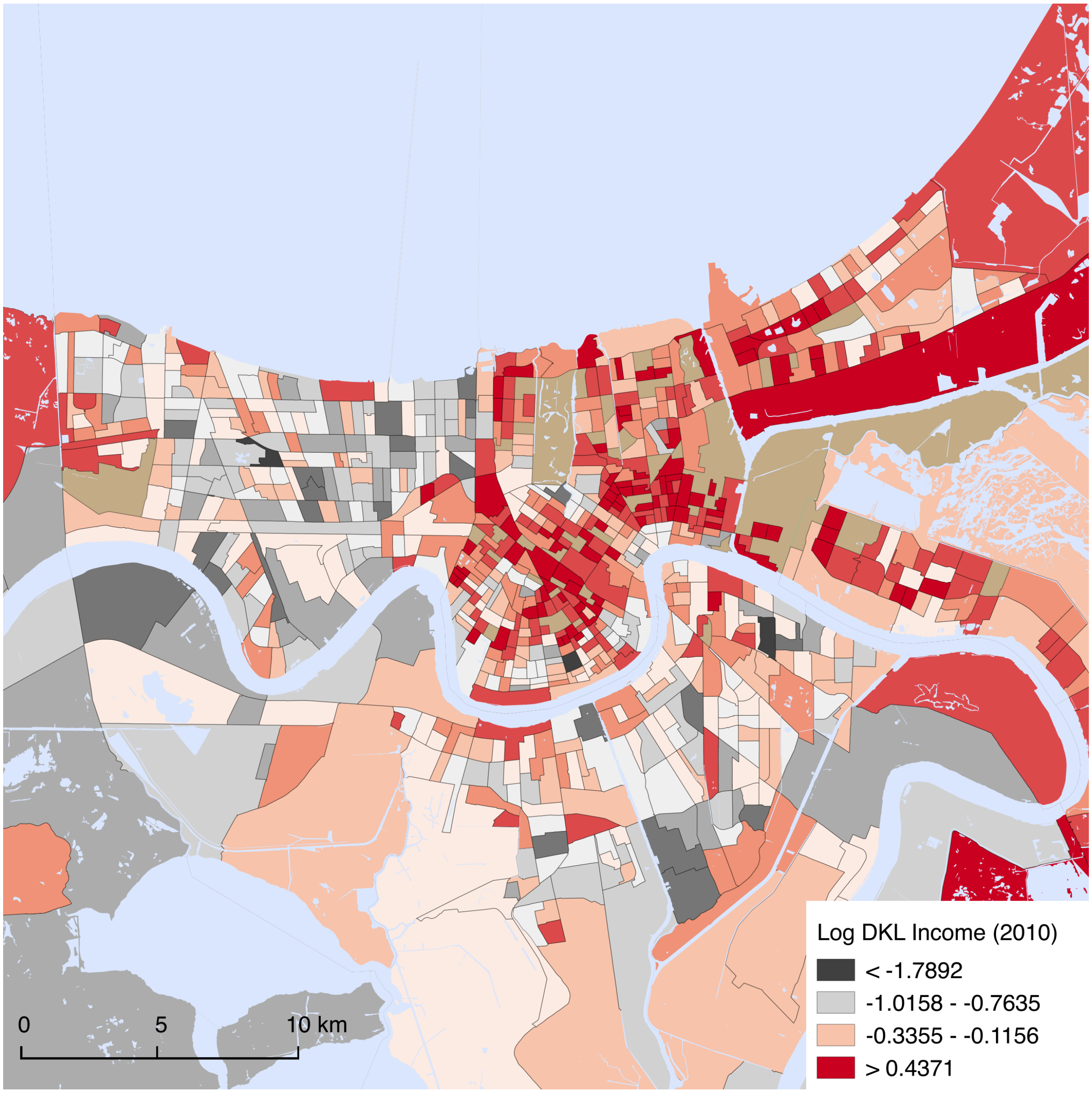


Figure S34: New Orleans, LA D_{KL} (2010)

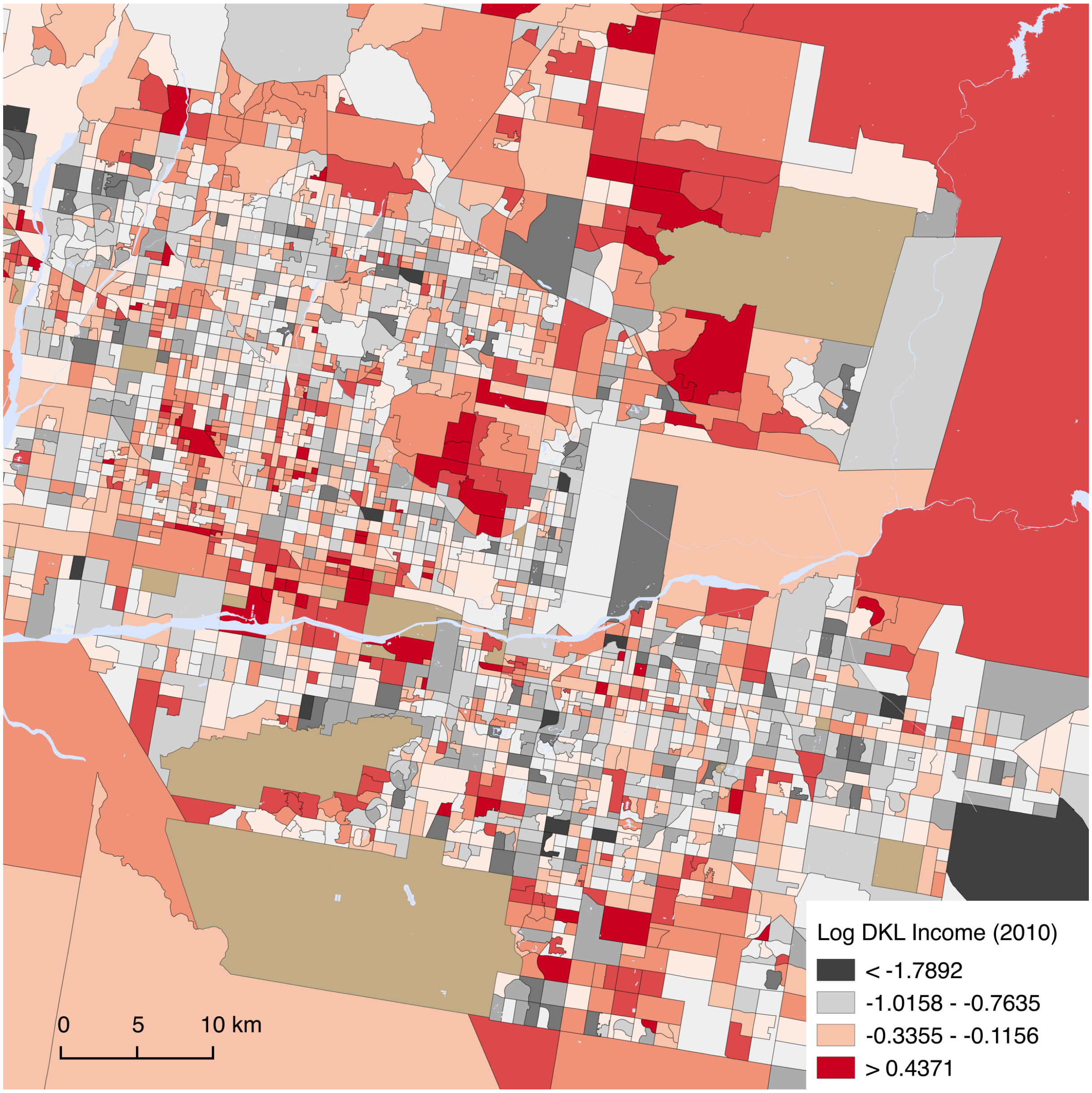


Figure S35: Phoenix, AZ D_{KL} (2010)

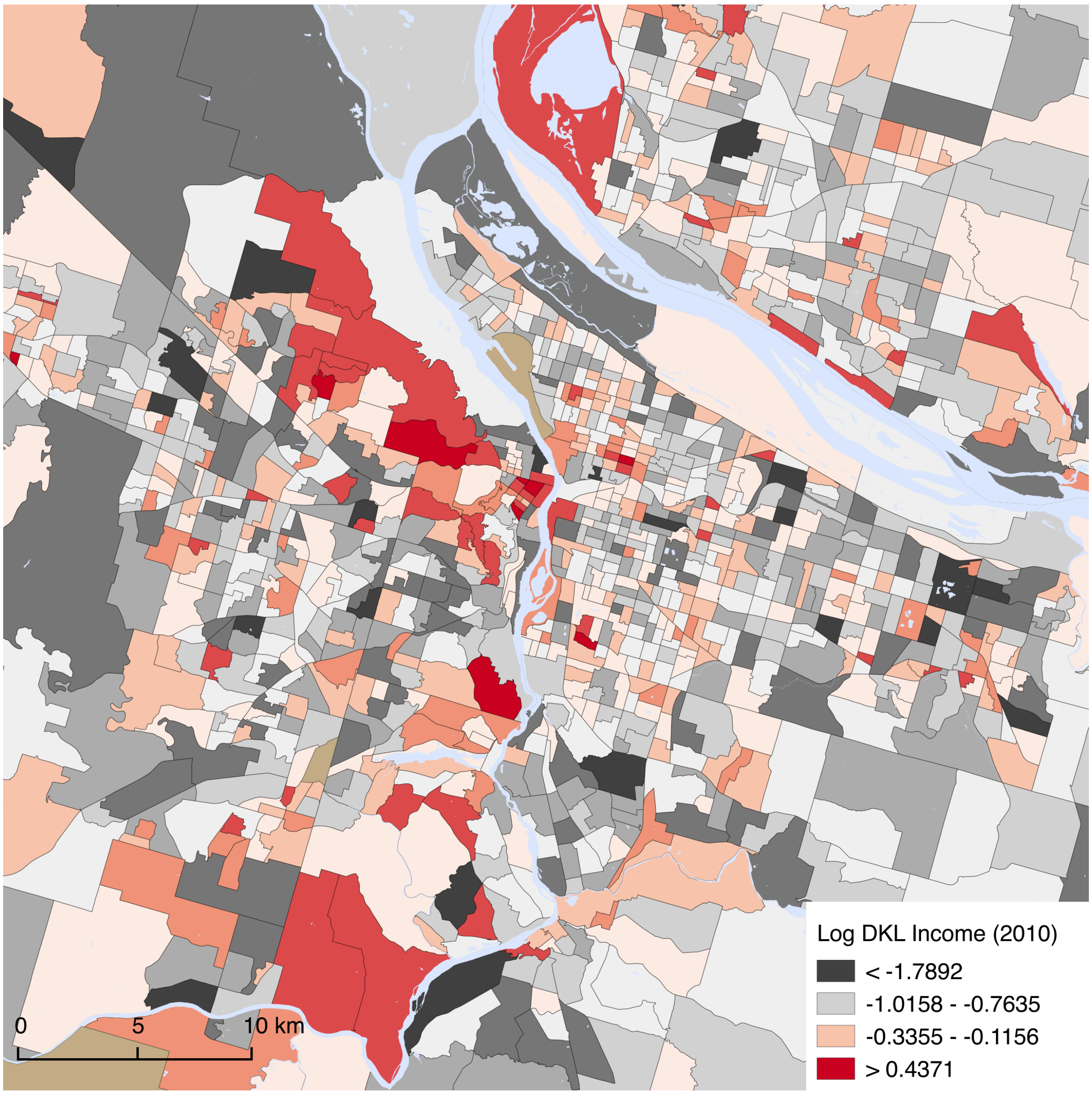


Figure S36: Portland, OR D_{KL} (2010)

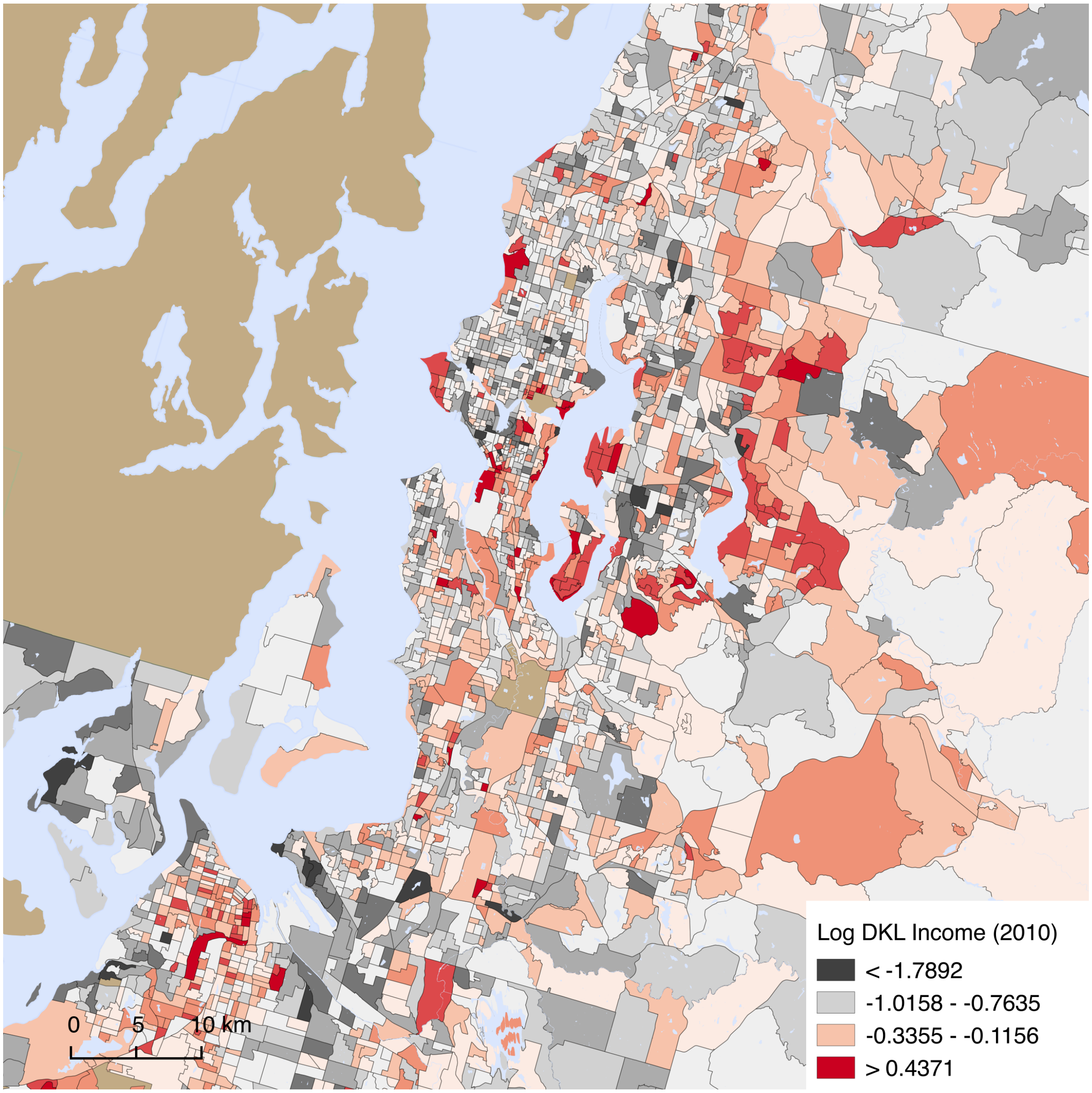


Figure S37: Seattle, WA D_{KL} (2010)

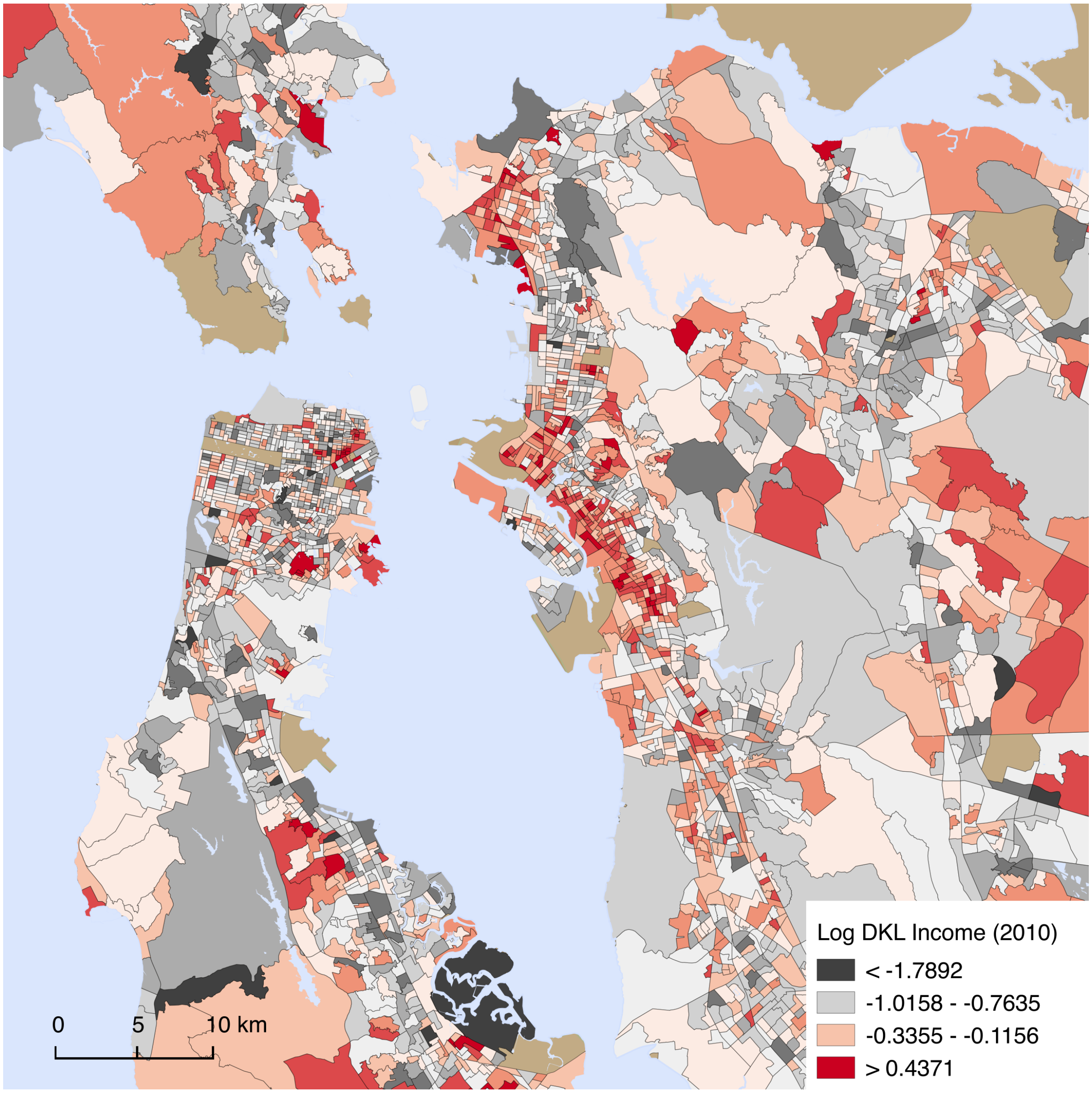


Figure S38: San Francisco, CA D_{KL} (2010)

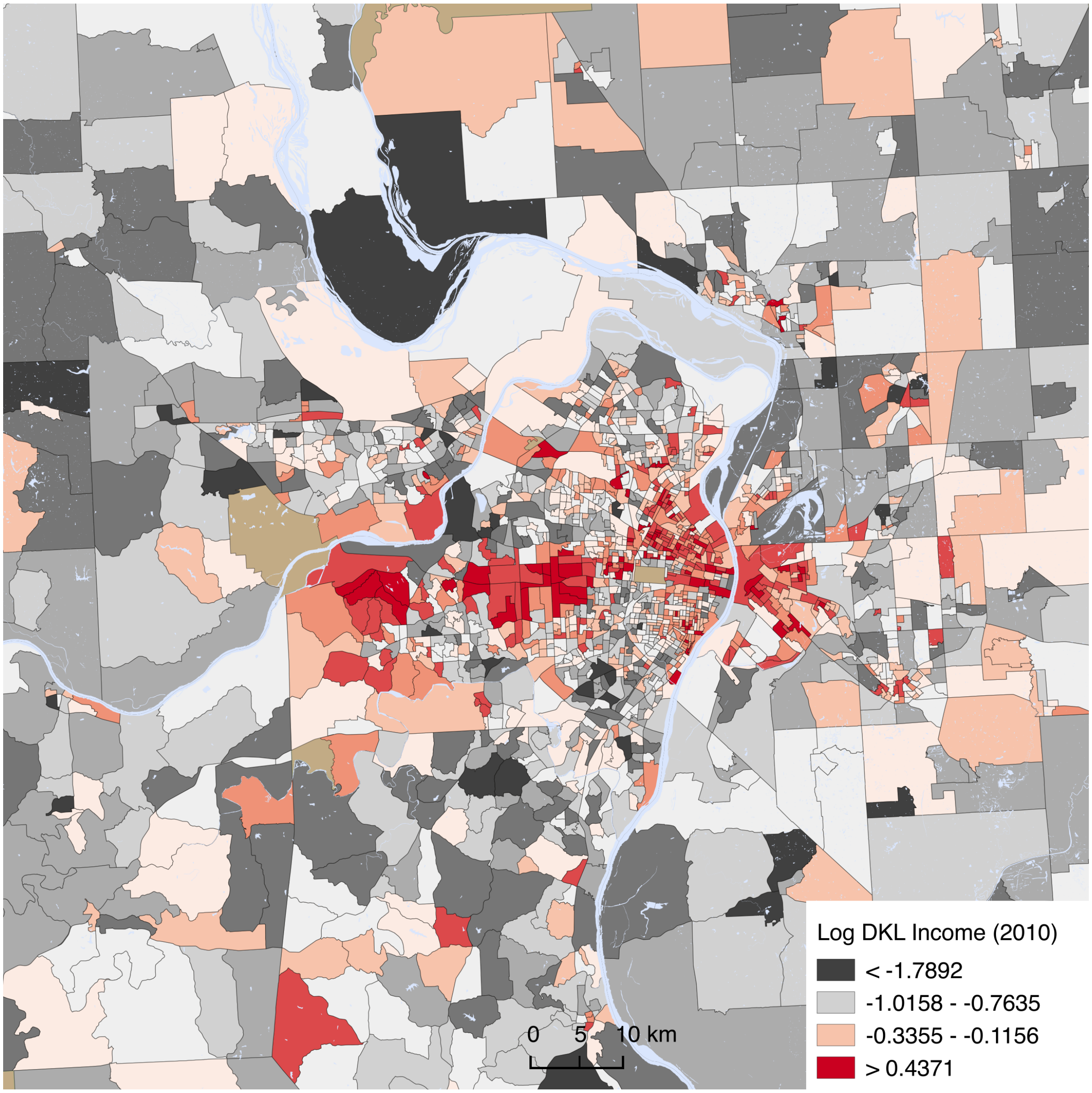


Figure S39: St. Louis, MO D_{KL} (2010)

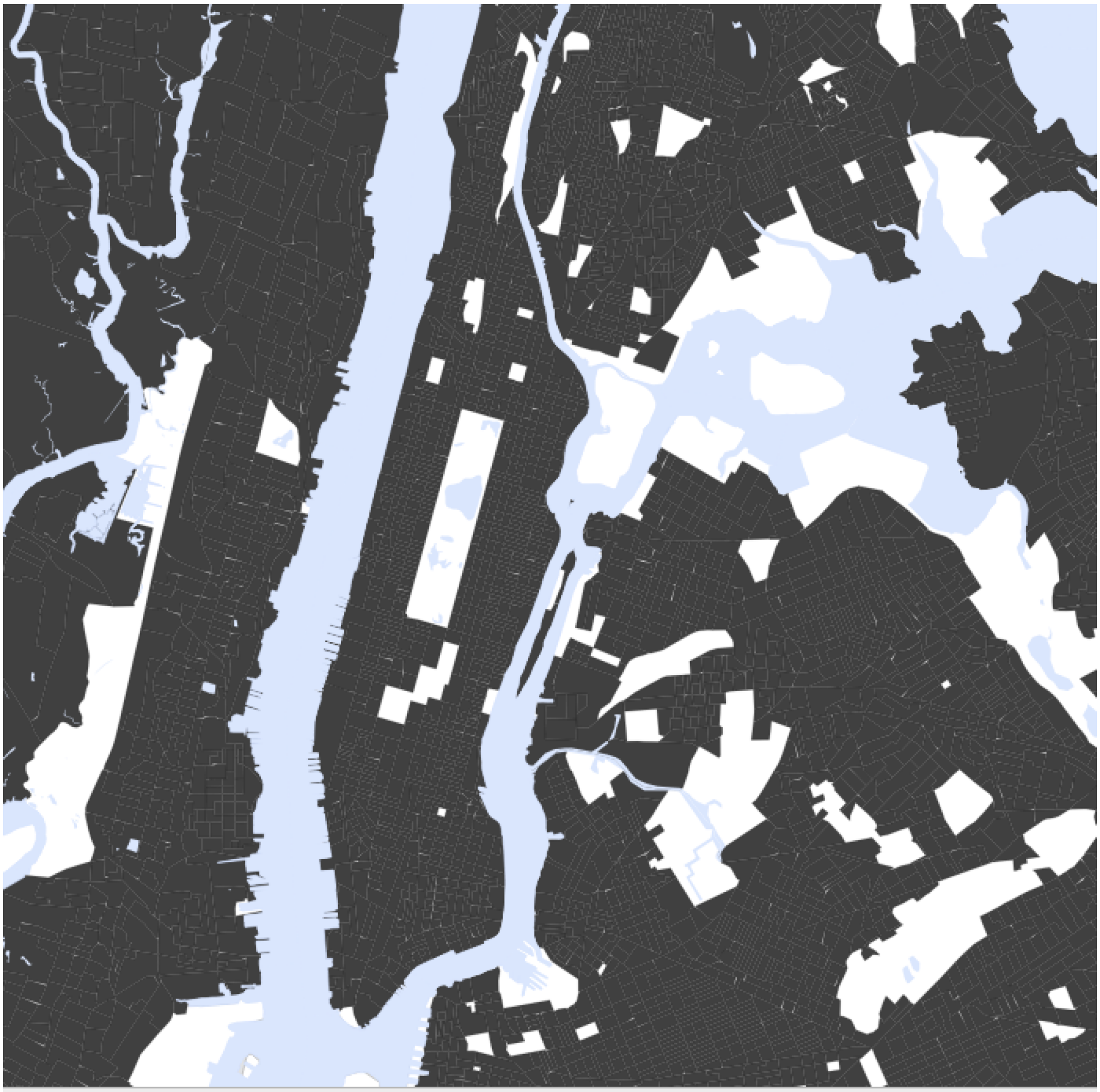


Figure S40: The D_{KL} for different neighborhoods obtained from random sampling of the metropolitan income distribution, c. f. Fig. 2A. We clearly observe that most observations of the strength of local selection that are not very small (> 0.03) cannot have arisen by chance.

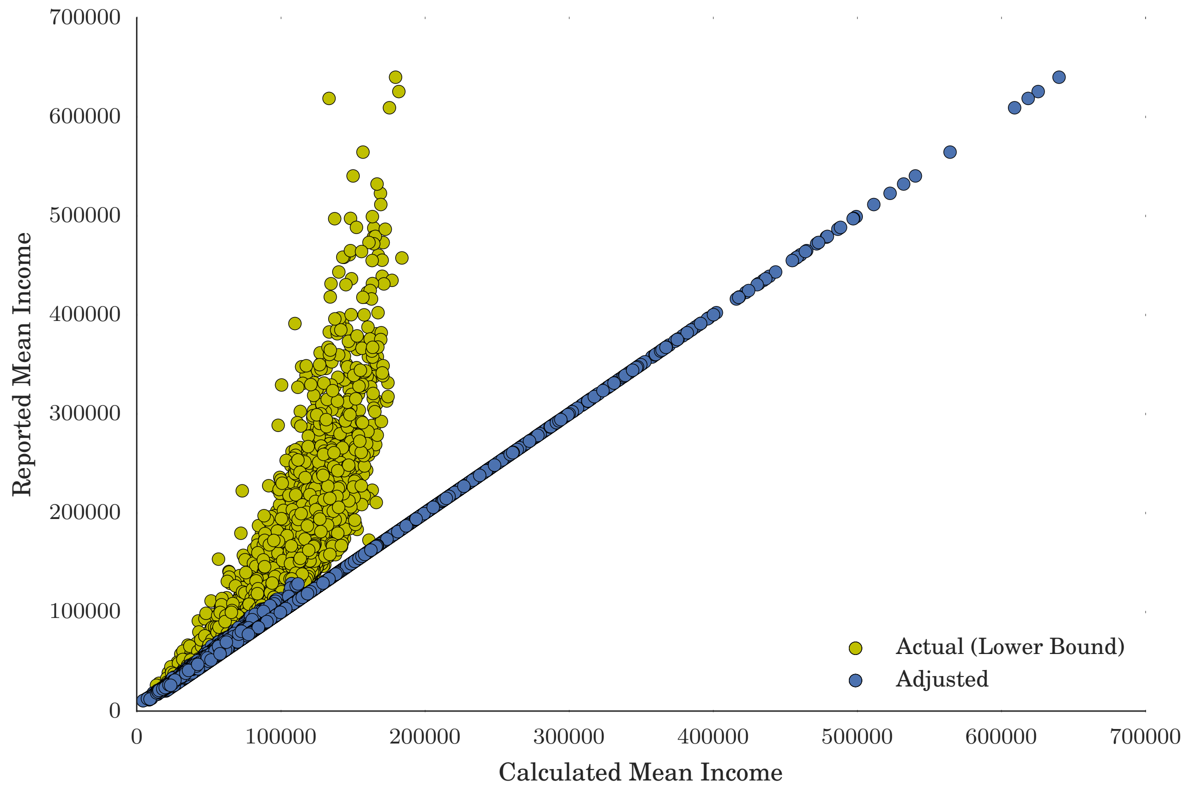


Figure S41: Household income for NYC block groups as reported by the Census and as calculated using the sixteen income bins. Using the lower bound of each income bin, we show the minimum mean income vs the reported incomes, in yellow. Then we increase the value of the top bin (previously $200,000) bin to account for the missing income, in blue.

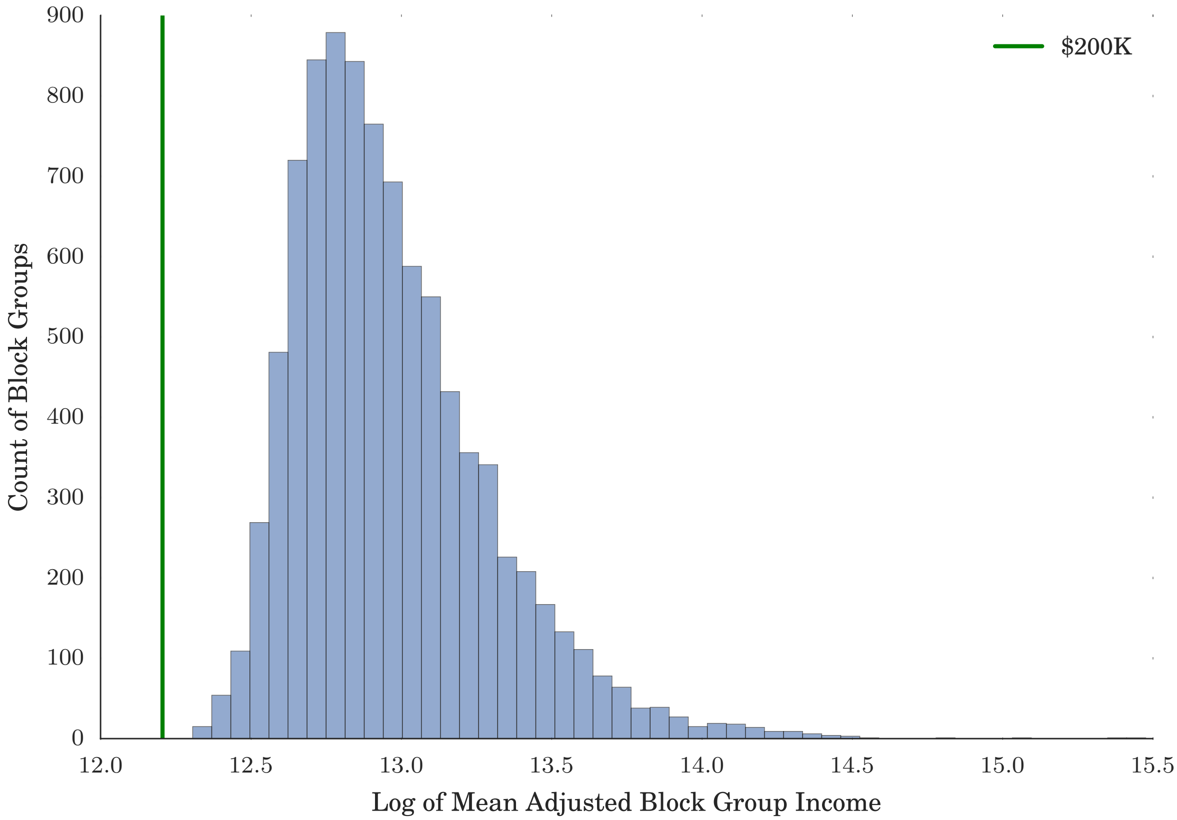


Figure S42: New values for the $200K+ bin used for adjusting the mean value.

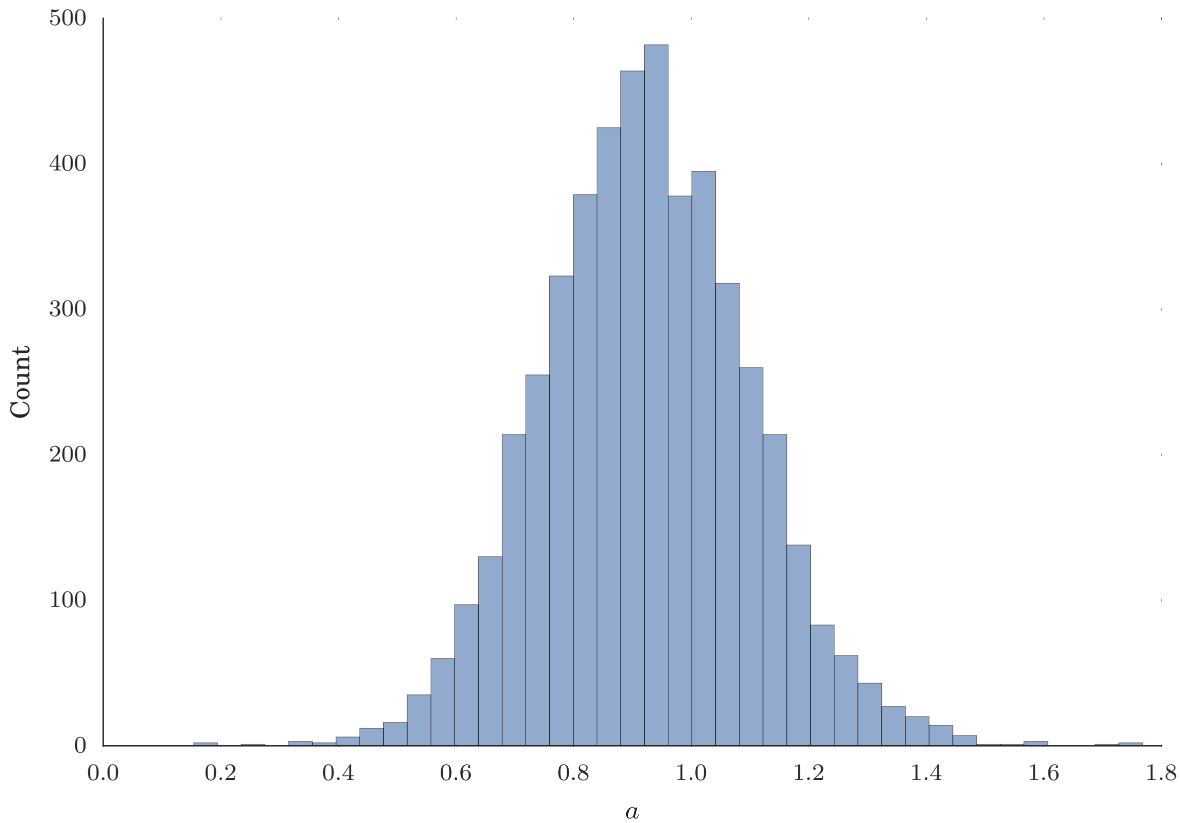


Figure S43: Price a value

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).