Decoding the city: multiscale spatial information in urban income distributions

Supplementary Online Material

Luís M. A. Bettencourt,[1,2*], Ivanna Rodriguez[1], Jordan T. Kemp[3] and José Lobo[4]

[1] University of Chicago, Chicago IL 60637, USA,

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

[3] Vienna Complexity Science Hub, Vienna, 1030, Austria

[4] Institute for New Economic Thinking at the Martin School, University of Oxford, UK

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

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

April 15, 2026

Contents

List of Tables

List of Figures

1 Supplementary Text

1.1 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

$$w_{\ell,j} = \frac{p(y_\ell, n_j)}{p(y_\ell)\, p(n_j)} = \frac{p(y_\ell|n_j)}{p(y_\ell)}. \tag{S1}$$

Taking the logarithm, we obtain

$$\log p(y_\ell|n_j) = \log w_{\ell j} + \log p(y_\ell) = \log \frac{p(y_\ell, n_j)}{p(y_\ell)\, p(n_j)} + \log p(y_\ell), \tag{S2}$$

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

$$h(y_\ell) = h(y_\ell|n_j) + i(y_\ell|n_j), \tag{S3}$$

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,

$$H(y) = H(y|n) + I(y; n). \tag{S4}$$

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

$$\langle \log w \rangle = I(y; n) = \sum_{\ell, j} p(y_\ell, n_j) \log \frac{p(y_\ell, n_j)}{p(y_\ell)\, p(n_j)}. \tag{S5}$$

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 (*37, 38*). 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” (*39*) 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 (*39*). 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) (*39*). 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

$$p(y_\ell) = \bar{w}_{\ell j} p(y_\ell | n_j), \tag{S6}$$

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

$$\bar{w}_{\ell j} = -\log \frac{p(y_\ell, n_j)}{p(y_\ell)\, p(n_j)} = \log \frac{p(y_\ell)\, p(n_j)}{p(y_\ell, n_j)} = -i(y_\ell, n_j). \tag{S7}$$

Thus, we write

$$h(y_\ell | n_j) = h(y_\ell) - i(y_\ell | n_j) \rightarrow H(y|n) = H(y) - I(y; n), \tag{S8}$$

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 (*5*).

1.2 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 - S5 and Figures S11 - S15 show similar maps (average household income by block group and the $\langle \log w_j \rangle$ for each neighborhood) for other large US metropolitan areas, including a larger map of New York City.

1.3 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 S6 and S7 present the results of comparing the goodness of fit of the lognormal distribution to that of other alternative distributions using the Bayesian Information Criterion (*2*) 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 (*3*)). The lognormal, on the other hand, is well known to fit well the body of distributions of income (*1, 4*) 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 S8 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 S9 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. (*5*).

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 (*5*), the degree of cell phone urban social networks (*6*) and mobility (*7*) 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 (*8, 9*).

1.4 Information, Theil indices, and the Aggregation Problem

The Theil index of inequality

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, to issues of economic inequality in a human population,

$$H(p) = \sum_\ell p(y_\ell) \log 1/p(y_\ell), \tag{S9}$$

where $p(y_\ell)$ is a (discrete) population density such that $\sum_\ell p(y_\ell) = 1$. The Shannon entropy is minimized when all incomes are the same and maximal when they are most dispersed, so that it is a measure of spread or disorder.

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

$$T(y) = \frac{1}{N} \sum_{k=1}^{N} \frac{y_k}{\bar{y}} \log \frac{y_k}{\bar{y}}, \tag{S10}$$

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

$$T(y) = \sum_\ell q(y_\ell) \log \frac{q(y_\ell)}{p(y_\ell)} = D\left[q(y)||p(y)\right], \tag{S11}$$

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.

Another way to write the Theil index is in terms of a sum over persons, not income shares. Consider the faction of total income for person k, $s_k = \frac{y_k}{N\bar{y}}$. Then the Theil index is

$$T(y) = \sum_k s_k \log(N s_k) = H_u - H(s), \tag{S12}$$

where $H_u = \log N$ is the entropy of the uniform (equal income) distribution, which is maximal, and $H(s) = -\sum_k s_k \log s_k$.

Theil's H_T as an index of segregation

A related, but different, quantity to Theil's T index of inequality, is Theil's H_T (or simply H, but do not confuse to the Shannon entropy), a measure of segregation (*19*), also known as the spatial information theory index. This is usually defined in the context of a partition of a population into groups, g, corresponding to different races/ethnicities. We can define a probability of belonging to each group as $p(g)$, and compare it among (spatial) sub-populations, such as all neighborhoods i, $p(g|n_i)$. With these definitions, we can compute the Shannon entropy of the aggregate population, $H(g)$ and of each of its neighborhoods, $H(g|n_i)$, and also note that each neighborhood has a population share $p(n_i)$, as defined in the main text. Then the Theil index of segregation, H_T, is defined as

$$H_T = \frac{H(g) - \sum_i p_i H(g|n_i)}{H(g)} = \frac{I(g,n)}{H(g)}. \quad \text{(S13)}$$

We see that, given the same quantities and spatial definitions, H_T coincides with our mutual information, and normalizes it by its maximum possible value $H(g)$. This index is well known in sociology and geography, where it has been used to study segregation across scales (*19*).

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,

$$p(y_\ell) = \sum_i p(y_\ell|n_i) p(n_i), \tag{S14}$$

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

$$q(y_\ell) = \sum_i \frac{y_\ell}{\bar{y}} p(y_\ell|n_i) p(n_i) = \sum_{i=1} \frac{y_\ell}{\bar{y}_i} p(y_\ell|n_i)\ \frac{\bar{y}_i}{\bar{y}} p(n_i) = \sum_i q(y_\ell|n_i) q(n_i). \tag{S15}$$

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

$$\log \frac{q(y_\ell)}{p(y_\ell)} = \log \frac{y_\ell}{\bar{y}} = \log \frac{y_\ell}{\bar{y}_i} + \log \frac{\bar{y}_i}{\bar{y}} = \log \frac{q(y_\ell|n_i)}{p(y_\ell|n_i)} + \log \frac{q(n_i)}{p(n_i)}, \tag{S16}$$

we obtain the hierarchical decomposition,

$$\begin{aligned} T(y) \quad &= \sum_i q(n_i) \log \frac{q(n_i)}{p(n_i)} + \sum_i q(n_i) \sum_\ell q(y_\ell|n_i) \log \frac{q(y_\ell|n_i)}{p(y_\ell|n_i)} \\ &\equiv T(n) + \sum_i q(n_i) T(y|n_i), \end{aligned} \tag{S17}$$

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

$$p(y_\ell|n_i, m_j) = w'_{\ell ij} p(y_\ell|n_i) = w'_{\ell ij}\ w_{\ell i} p(y_\ell) \equiv w''_{\ell ij} p(y_\ell) \tag{S18}$$

with

$$w_{\ell i} = \frac{p(y_\ell|n_i)}{p(y_\ell)} = \frac{p(y_\ell, n_i)}{p(n_i)p(y_\ell)}, \tag{S19}$$

$$w'_{\ell ij} = \frac{p(y_\ell|n_i, m_j)}{p(y_\ell|n_i)} = \frac{p(y_\ell, m_j|n_i)}{p(m_j|n_i)p(y_\ell|n_i)} = \frac{p(y_\ell, m_j, n_i)}{p(m_j, n_i)p(y_\ell|n_i)}. \tag{S20}$$

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

$$\langle \log w''_{ij} \rangle \equiv D(p(y|n_i, m_j)||p(y)) = \sum_\ell p(y_\ell|n_i, m_j) \log \frac{p(y_\ell|n_i, m_j)}{p(y)} \tag{S21}$$

$$= \sum_\ell p(y_\ell|n_i, m_j) \left[\log \frac{p(y_\ell|n_i, m_j)}{p(y|n_i)} + \log \frac{p(y_\ell|n_i)}{p(y)} \right], \tag{S22}$$

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

$$\begin{aligned} I(y; n, m) &= \sum_{i,j} p(n_i, m_j) \langle \log w''_{ij} \rangle = \sum_{i,j} p(n_i) p(m_j|n_i) \sum_\ell D(p(y|n_i, m_j)||p(y)) \\ &= \sum_i p(n_i) D(p(y|n_i)||p(y)) \\ &\qquad + \sum_{i,j} p(n_i, m_j) D(p(y|n_i, m_j)||p(y|n_i)) \\ &= \sum_i p(n_i) \langle \log w_i \rangle + \sum_{ij} p(n_i, m_j) \langle \log w'_{ij} \rangle \qquad \text{(S23)} \\ &= I(y; n) + \sum_i p(n_i) I(y; m|n_i), \end{aligned}$$

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.

1.5 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.

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 (*10*). 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 (*11*) 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 (*12–14*).

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 (*15–18*). 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 (*19*), 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 (*17*). 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 (*21–23*), so it has for the growing income segregation in US urban areas (*20,24–28*). 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 (*17, 29*). 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 (*17, 30–33*). 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 socioeconomic 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 (*17*), 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 (*17*), 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. S4, S14, St. Louis, S5, S15, where poor and rich section of 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 mathematical account of selection developed here thus 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 (*35, 36*).

2 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

Supplementary Figures

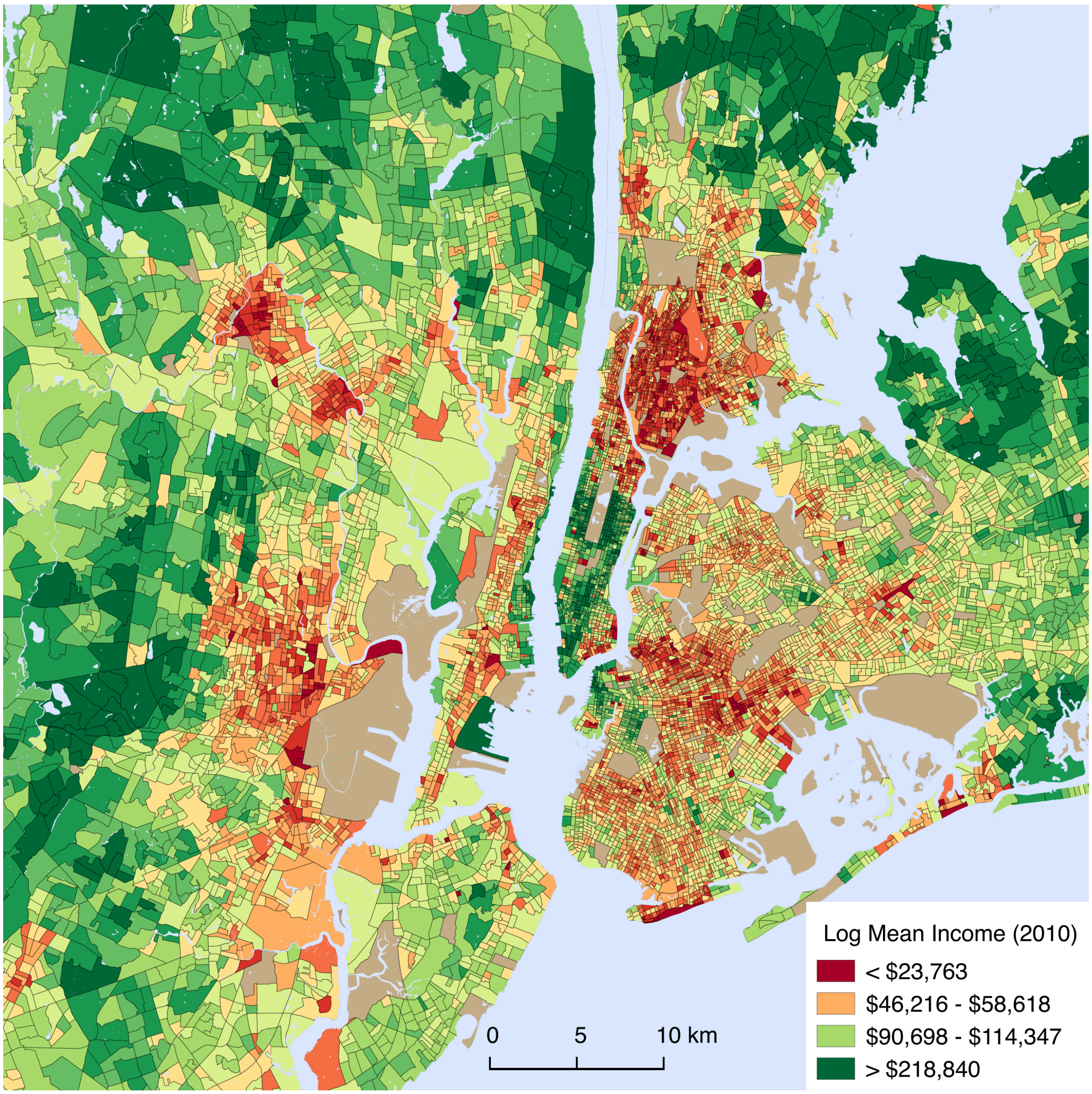

Figure S1: New York, NY Mean Household Income (2010)

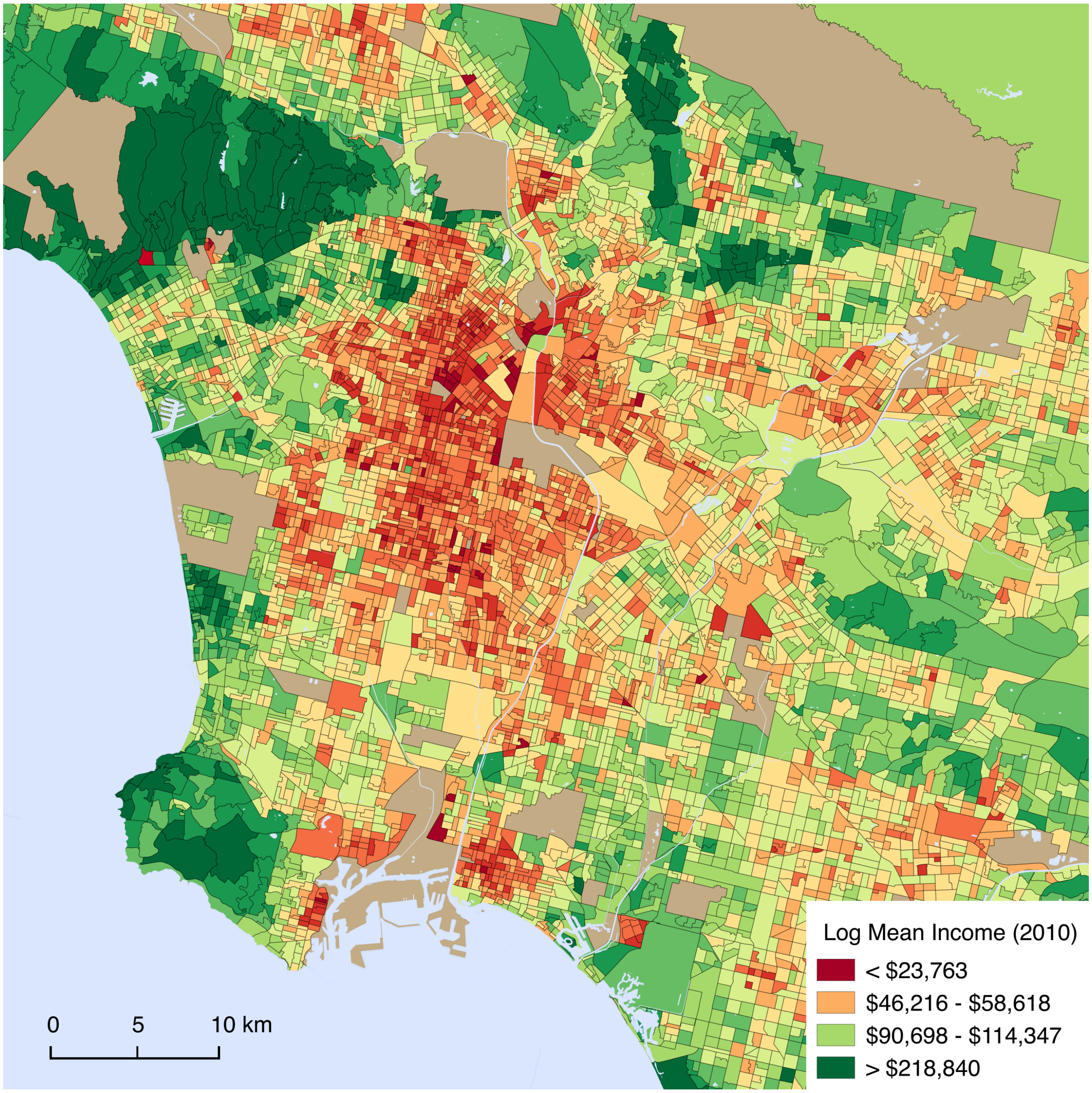

Figure S2: Los Angeles, CA Mean Household Income (2010)

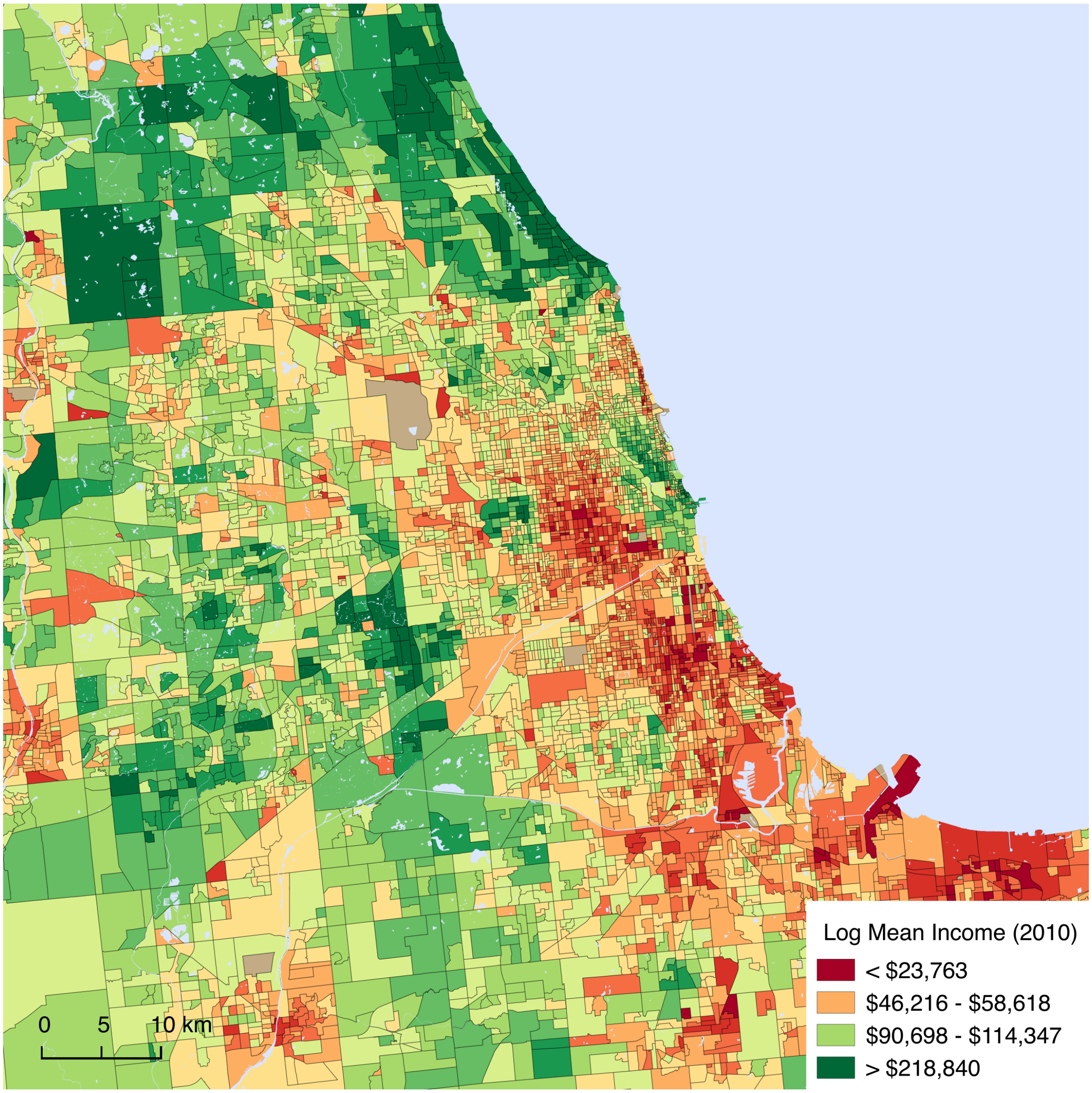

Figure S3: Chicago, IL Mean Household Income (2010)

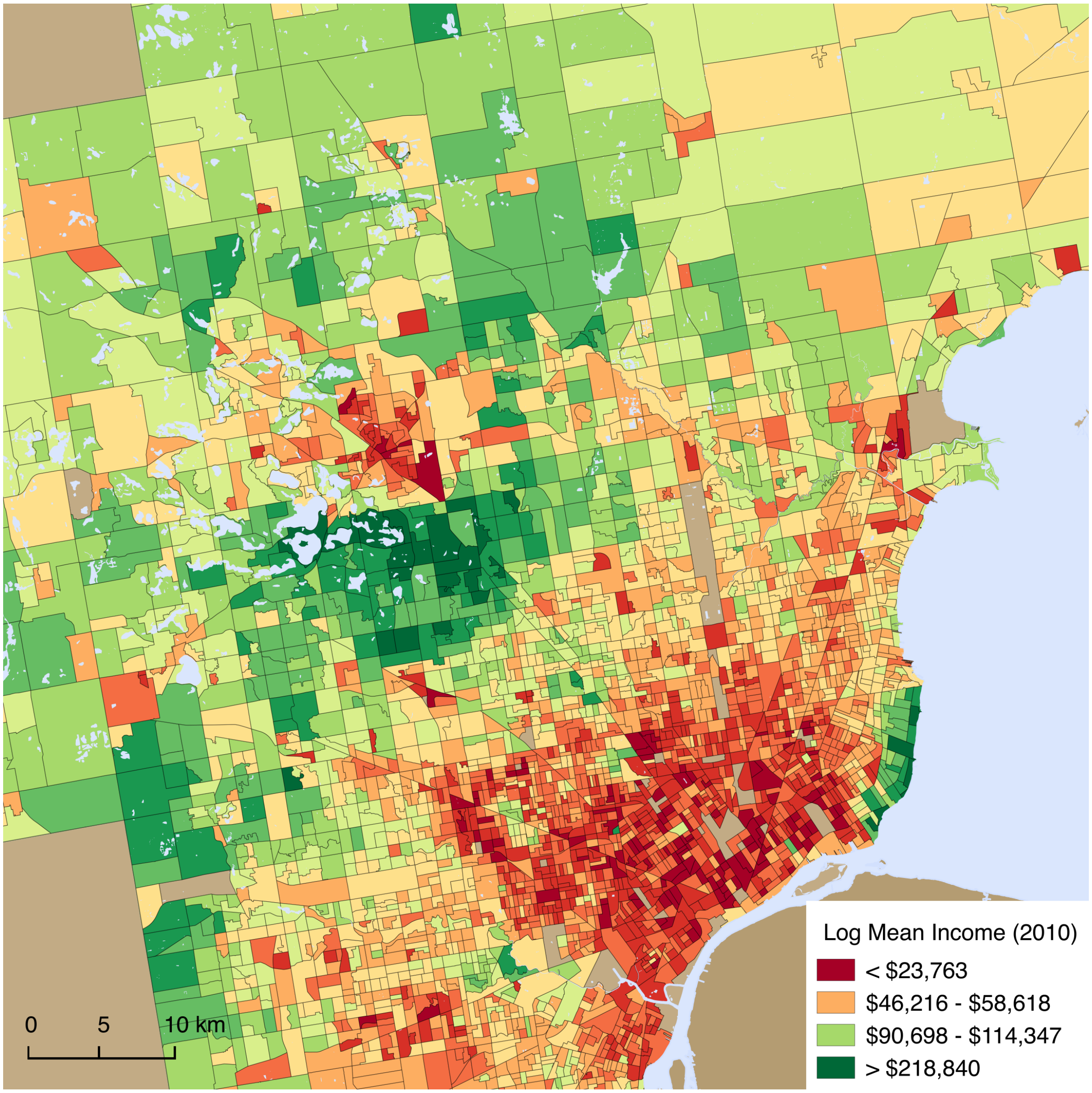

Figure S4: Detroit, MI Mean Household Income (2010)

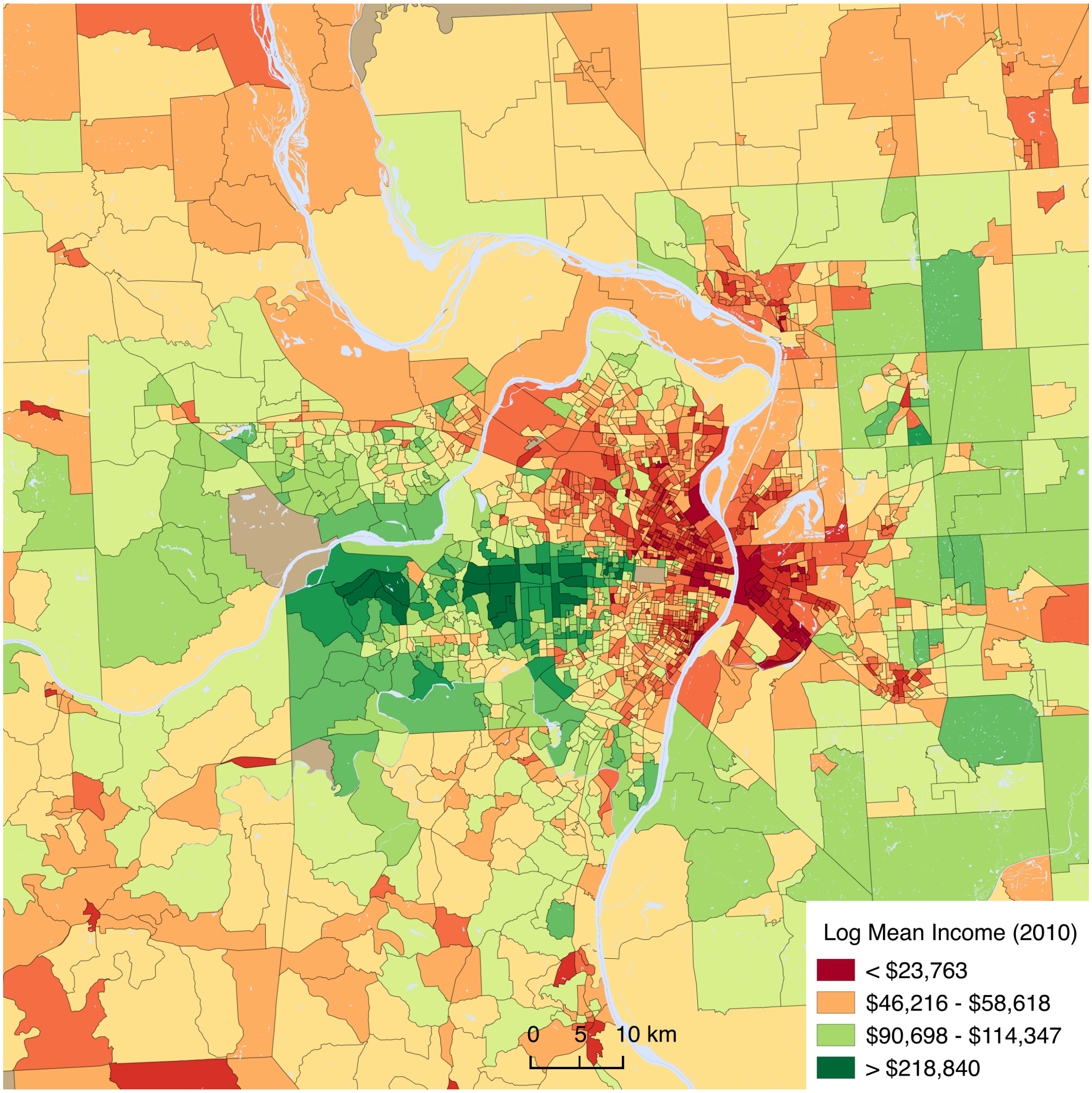

Figure S5: St Louis, MO Mean Household Income (2010)

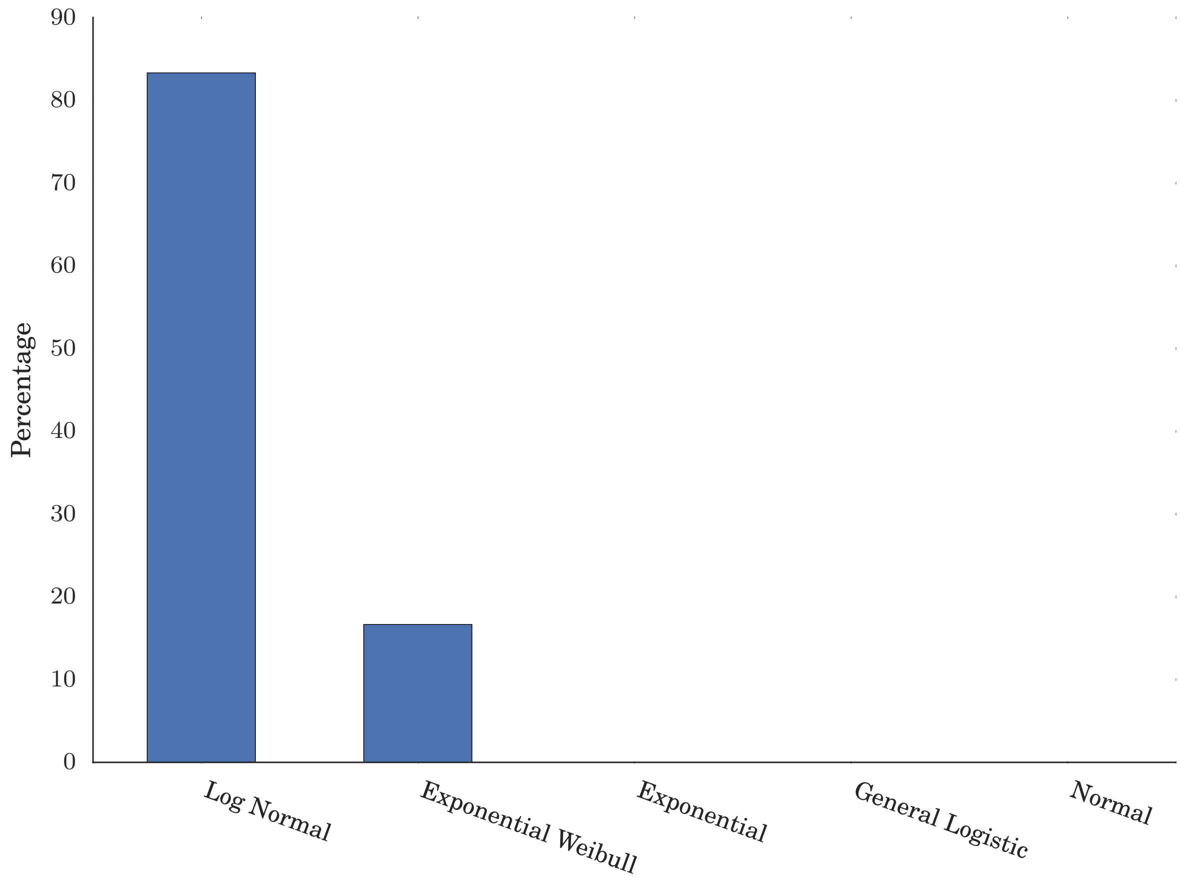

Figure S6: BIC fit test for the distribution of mean income in all CBSA's in the US, first place results.

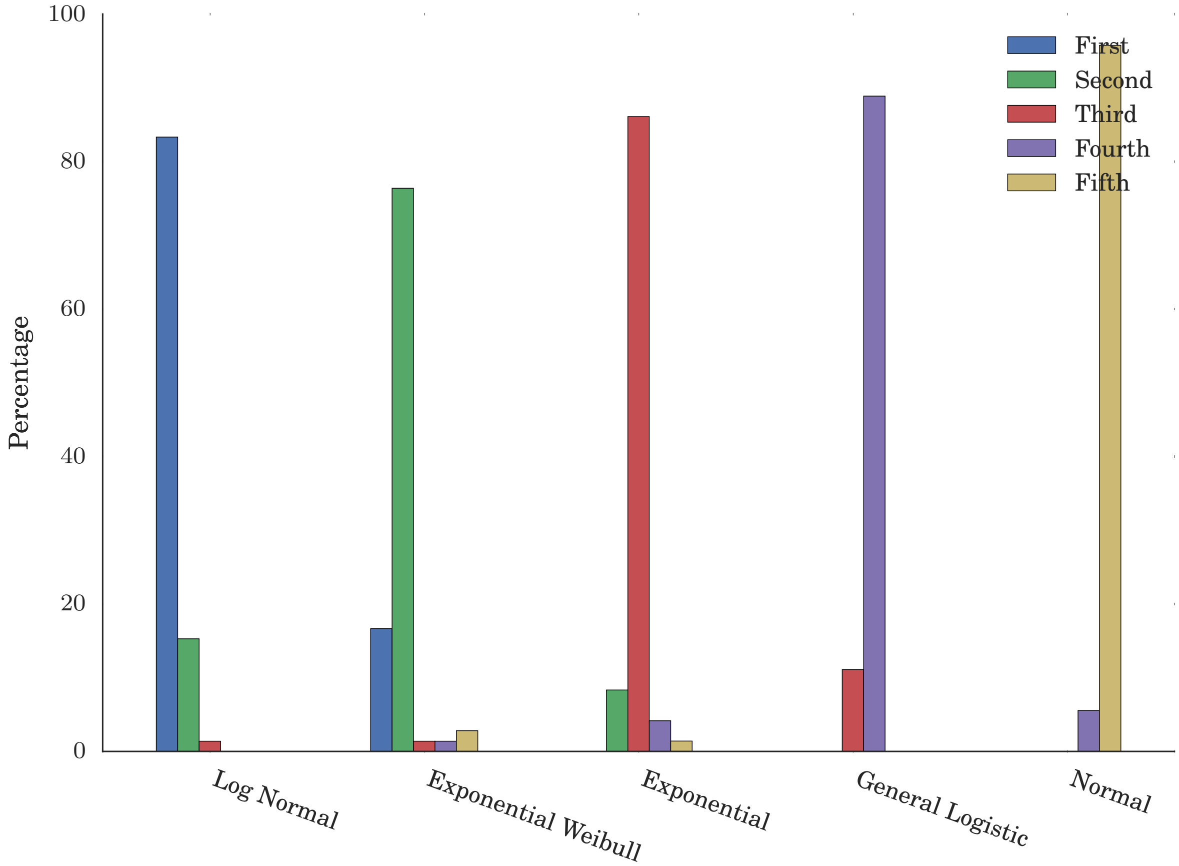

Figure S7: BIC fit test for the distribution of mean income in all CBSA's in the US

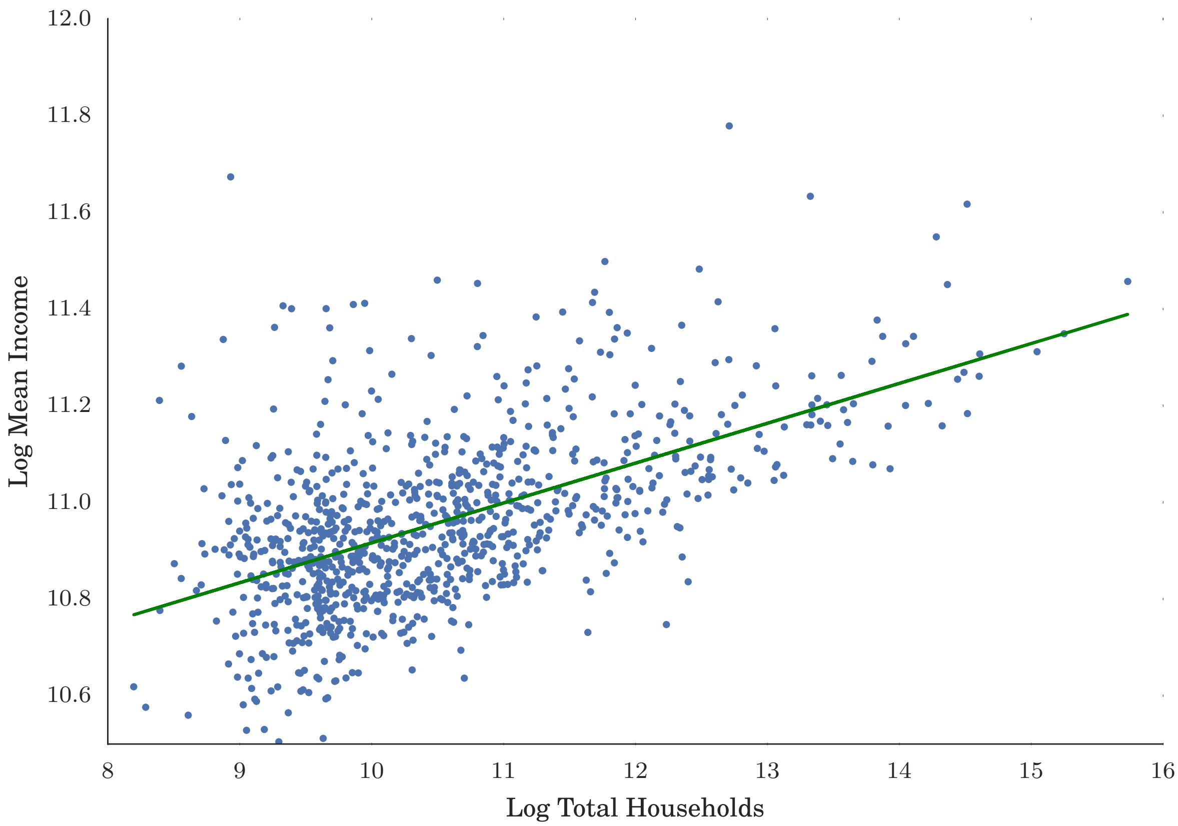

Figure S8: Scaling Mean. The mean income of a city, in relation to number of households, is characterized by an exponent of 0.0825 ($95\%CI = [0.075, 0.090]$, $R^2 = 0.317$).

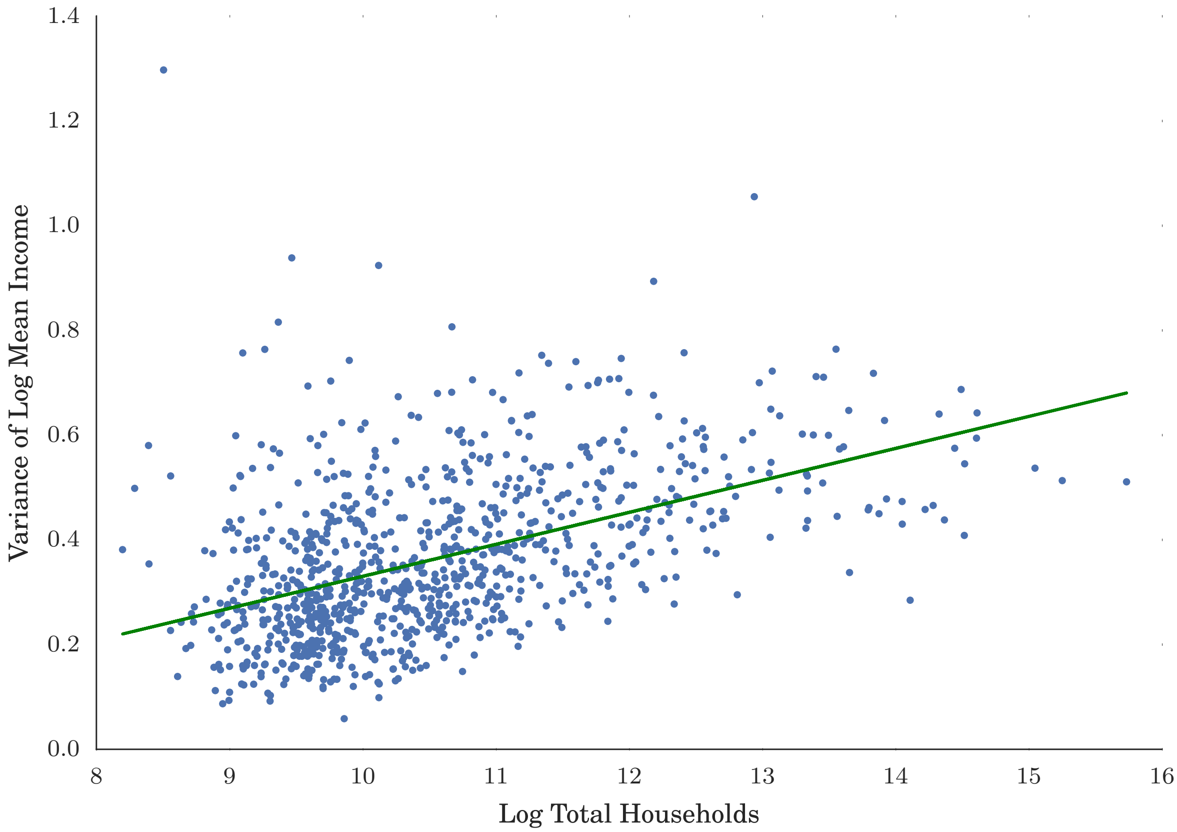

Figure S9: Scaling Variance. The variance of block group income for a city, in relation to number of households, is characterized by an exponent of 0.0611 ($95\% CI = [0.054, 0.068], R^2 = 0.235$).

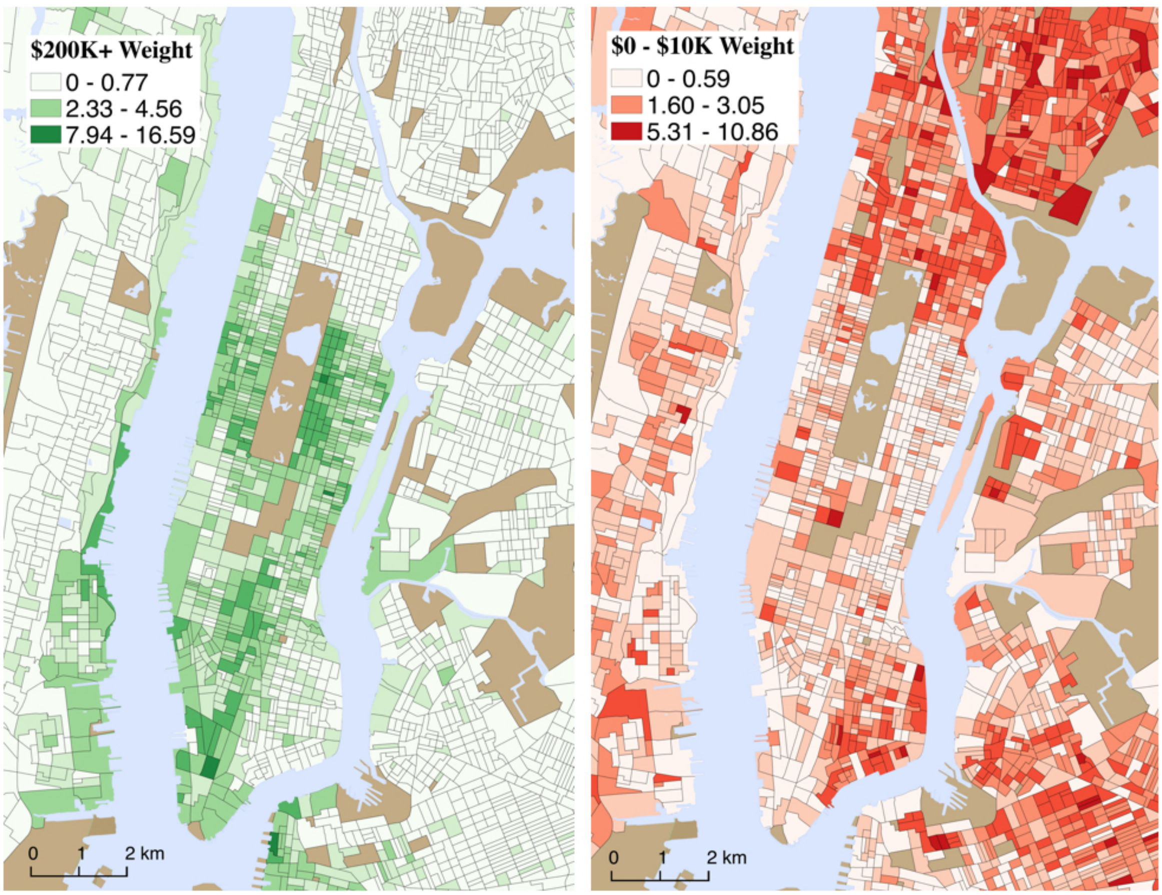

Figure S10: The weights $w_{\ell j}$ for the richest (left, in green) and poorest (right, in red) income bins for New York City.

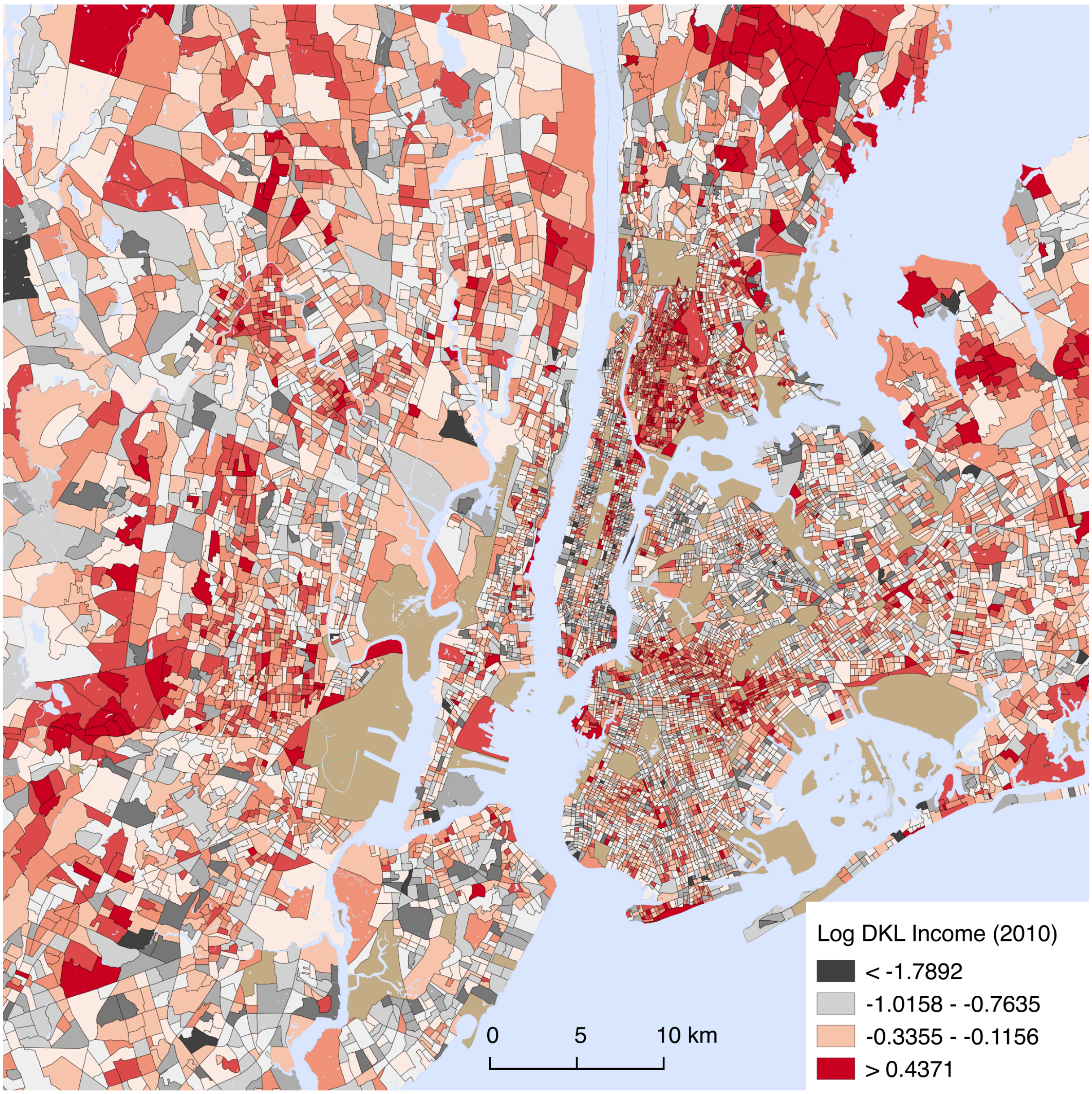

Figure S11: New York, NY $\langle \log w_j \rangle$ (2010)

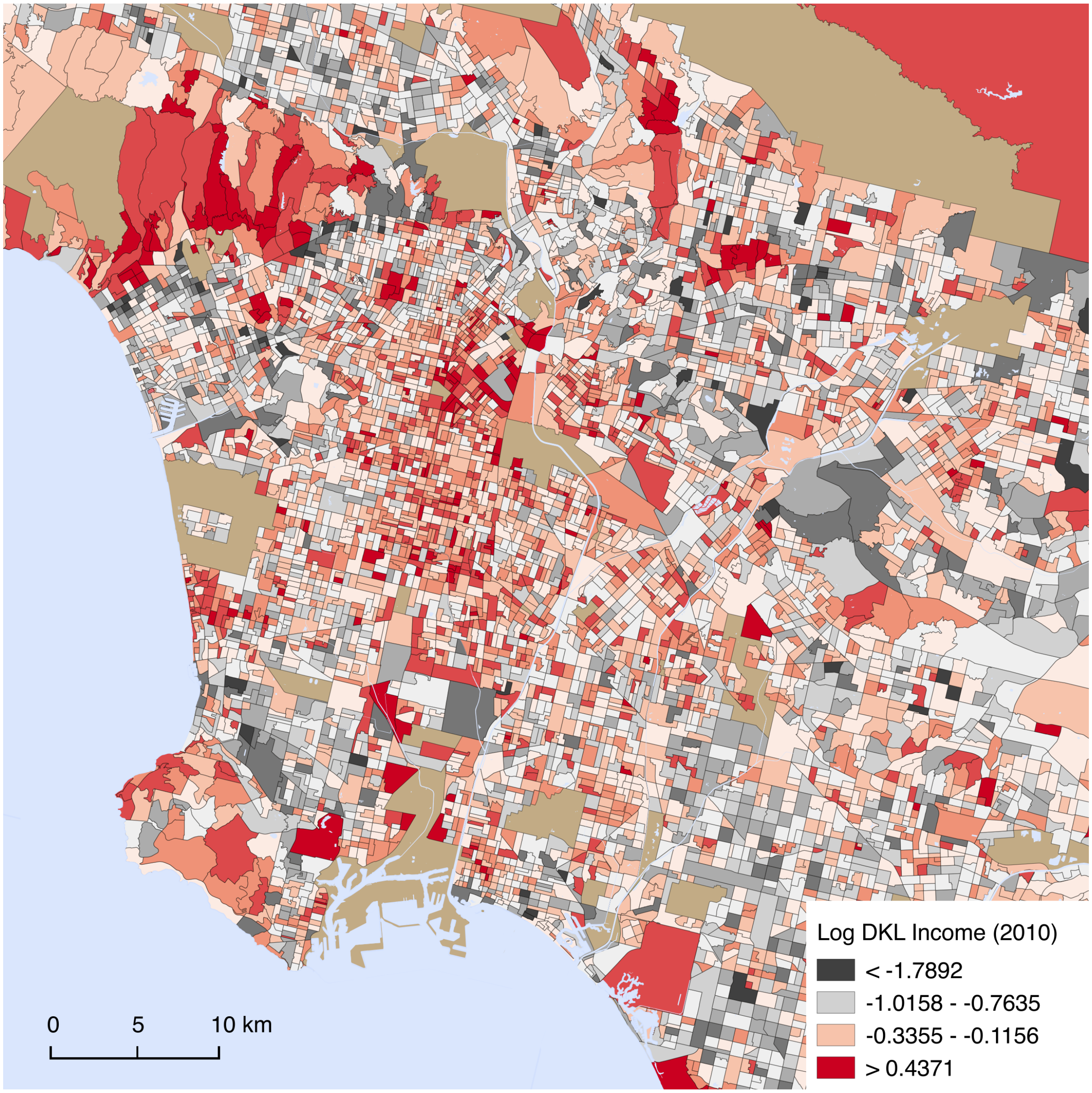

Figure S12: Los Angeles, CA $\langle \log w_j \rangle$ (2010)

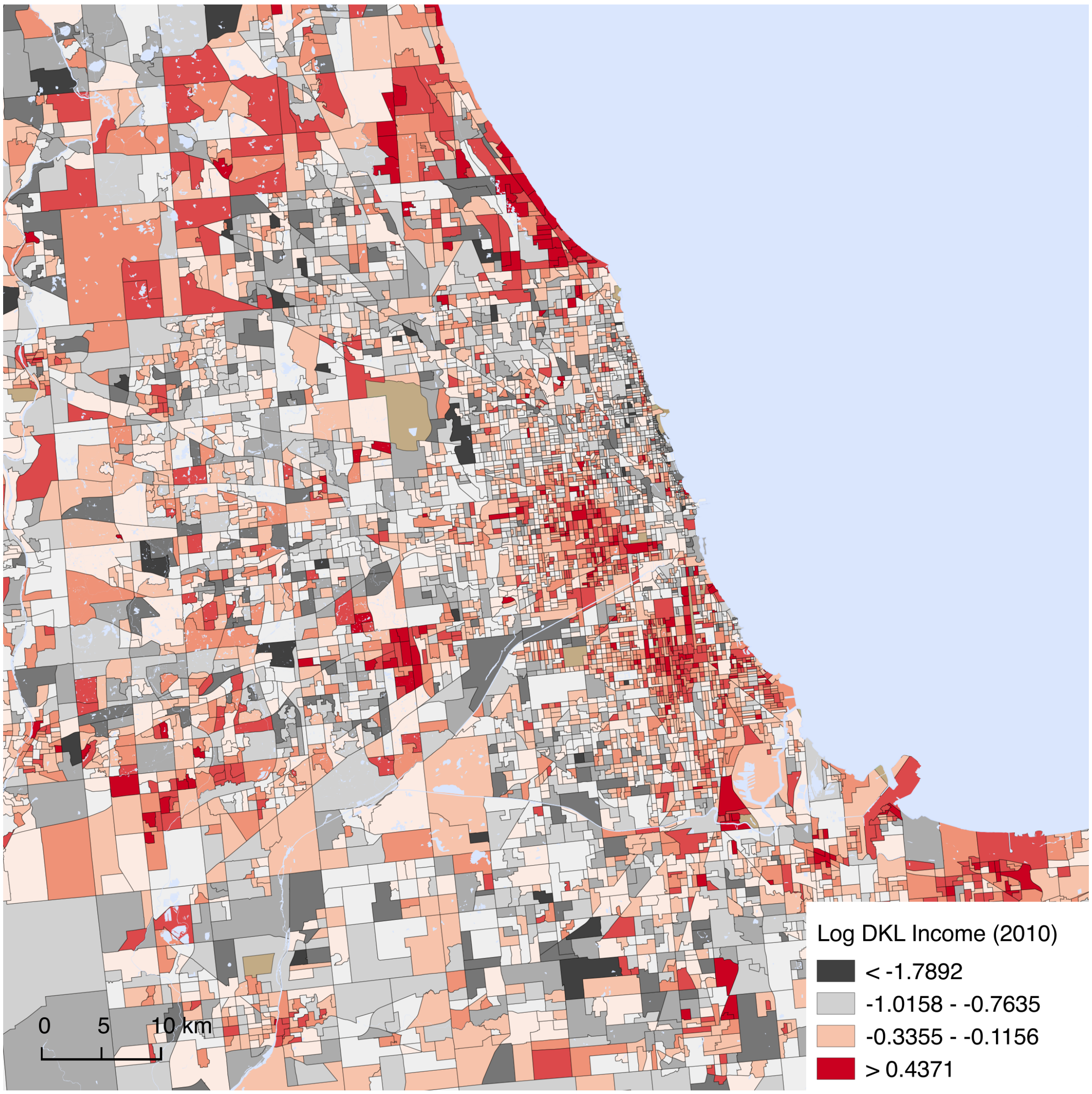

Figure S13: Chicago, IL $\langle \log w_j \rangle$ (2010)

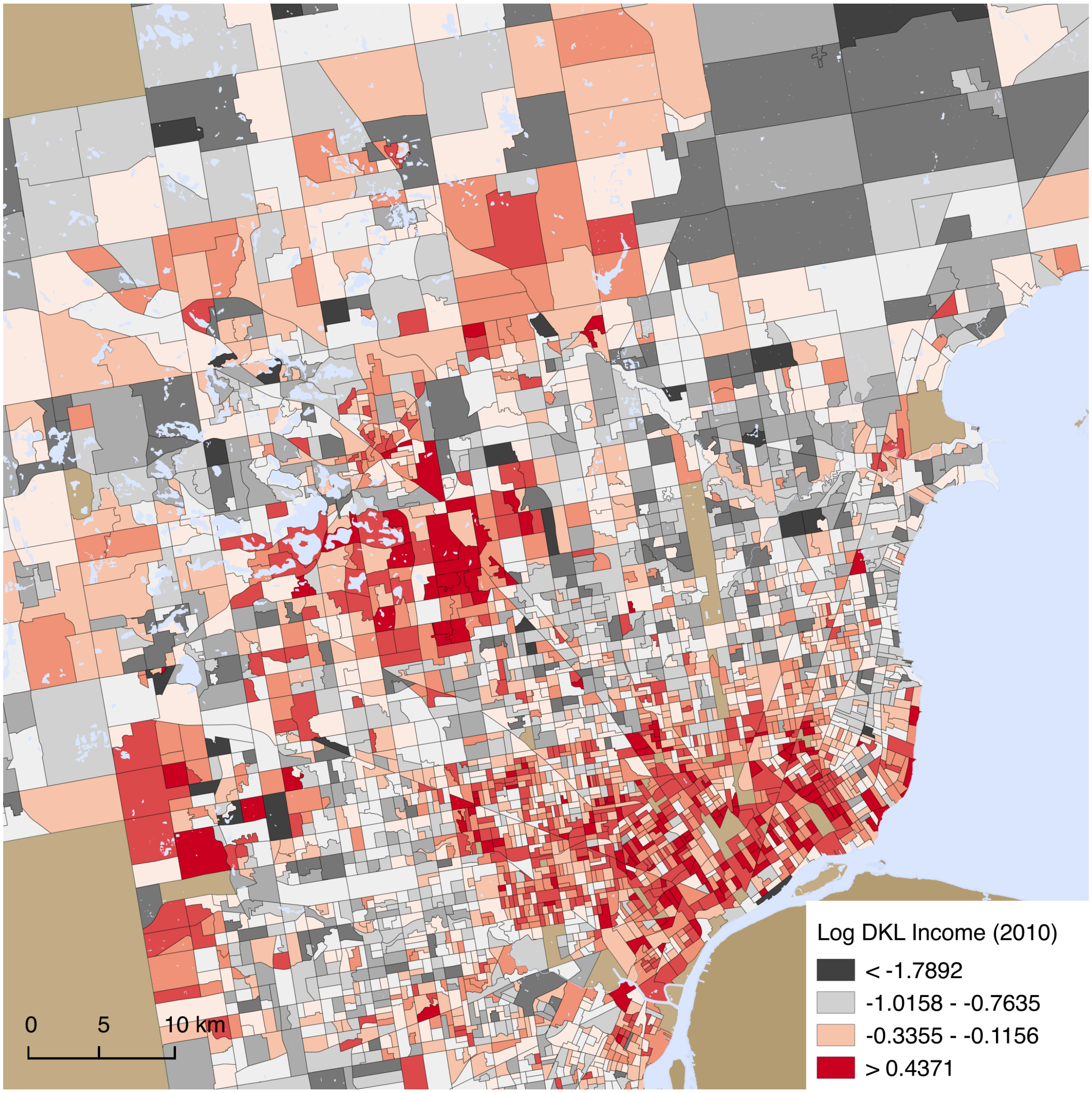

Figure S14: Detroit, MI $\langle \log w_j \rangle$ (2010)

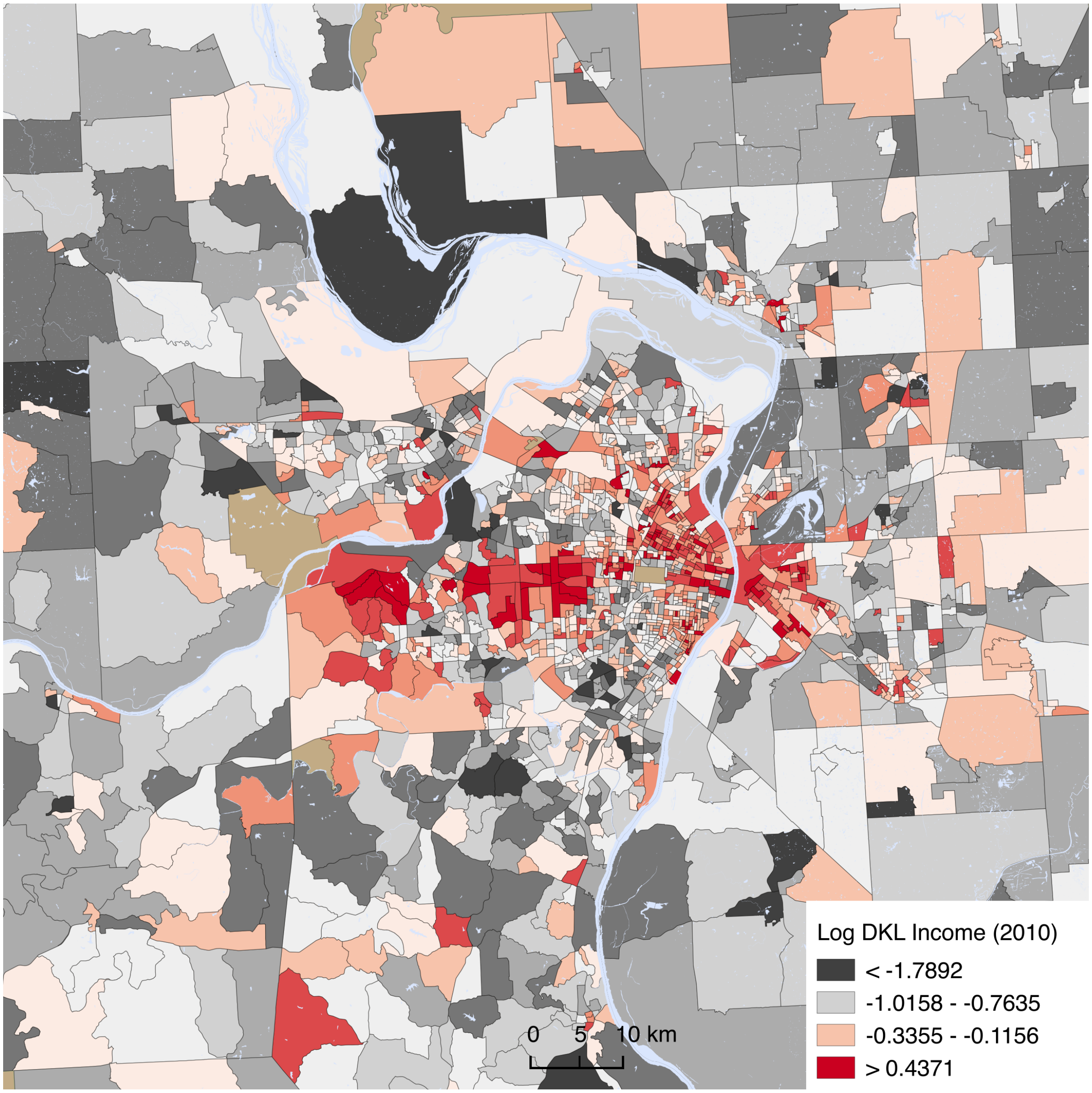

Figure S15: St Louis, MO $\langle \log w_j \rangle$ (2010)

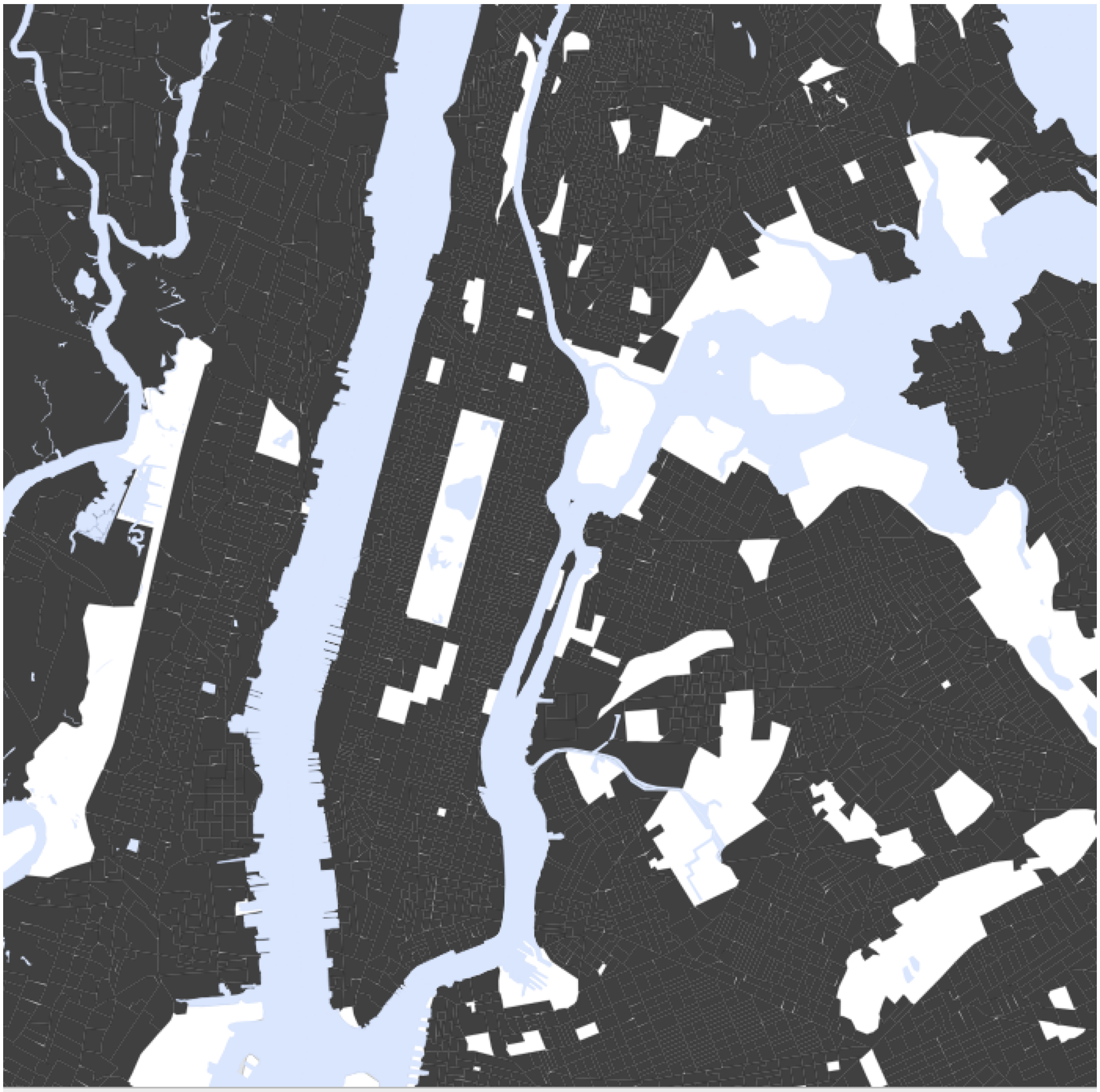

Figure S16: The $\langle \log w_j \rangle$ for different neighborhoods in New York City, obtained from random sampling of the metropolitan income distribution, c. f. Fig. 2A. We clearly observe that most observations of the strength of local selection in this random model are very small ($\langle \log w_j \rangle < 0.03$). This means that the vast majority of observed local selection cannot have arisen by chance, and specifically as a result of variation of inference due to BK-GP population size differences.

References and Notes

1. 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).

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

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

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

5. 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`.

6. 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`.

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

8. 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`.

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

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

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

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

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

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

15. 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).

16. 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).

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

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

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

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

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

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

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

24. 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).

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

26. 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).

27. 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).

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

29. A. Walks, Income Inequality and Polarization in Canada’s Cities: An Examination and New Form of Measurement (Research Paper 227, Cities Centre, University of Toronto, 2013).

30. 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).

31. 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).

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

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

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

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

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

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

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

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