



# Scaling of urban amenities: generative statistics and implications for urban planning

Talia Kaufmann<sup>1,2</sup>, Laura Radaelli<sup>2</sup>, Luis M.A. Bettencourt<sup>3</sup> and Erez Shmueli<sup>2\*</sup> 

\*Correspondence: [shmueli@tau.ac.il](mailto:shmueli@tau.ac.il)

<sup>2</sup>Department of Industrial Engineering, Tel-Aviv University, Tel-Aviv, Israel

Full list of author information is available at the end of the article

## Abstract

Cities have been extensively studied as complex adaptive systems over the last 50 years. Recently, several empirical studies and emerging theory provided support for the fact that many different urban indicators follow general consistent statistical patterns across countries, cultures and times. In particular, total personal income, measures of innovation, crime rates, characteristics of the built environment and other indicators have been shown to exhibit non-linear power-law scaling with the population size of functional cities. Here, we show how to apply this type of analysis inside cities to establish universal patterns in the quantity and distribution of urban amenities such as restaurants, parks, and universities. Using a unique data set containing millions of amenities in the 50 largest US metropolitan areas, we establish general non-linear scaling patterns between each city's population and many different amenities types, the small-area statistics of their spatial abundance, and the characteristics of their mean distance to each other. We use these size-specific statistical findings to produce generative models for the expected amenity abundances of any US city. We then compute the deviations observed in given cities from this statistical many-amenity model to build a characteristic signature for each urban area. Finally, we show how urban planning can be guided by these systemic quantitative expectations in the context of new city design or the identification of local deficits in service provision in existing cities.

**Keywords:** Land use; Service provision; Spatial statistics; Urban analytics; Planning support systems

## 1 Introduction

A city is a complex and dynamic system of interactions between people and a rich ecology of organizations, mediated by physical infrastructure and built spaces [1].

Jane Jacobs was the first to define cities as problems of organized complexity [2], adopting Warren Weaver's definition of complex systems [3] as problems dealing concurrently with many variables which are interrelated and in constant simultaneous change. Jacobs' characterization of the "kind of problem a city is" marked a new start to the quest of understanding cities in more holistic and people-centered ways.

This framework for thinking about cities was then embraced by many architects and urbanists, such as Christopher Alexander [4] who started outlining cities as complex net-

© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

works of overlapping uses in space and time. More recently these general qualitative expectations have gained even more support as well as detail via empirical and theoretical studies which have successfully shown that cities manifest universal and quantifiable features, spanning across time, cultures and nations [1, 5–12].

The recent unprecedented availability of urban data is revolutionizing a number of scientific disciplines as well as the practice of policy and planning in cities. In the case of urban studies, data-driven approaches have been increasingly successful in identifying universal patterns in the behavior of cities. Specifically, several non-linear scaling laws have been observed and predicted, connecting the city's population size with a variety of urban indicators, such as economic activity [1, 5, 6], road network length [1, 7], crime rate [8, 13], traffic congestion [9, 14], shared means of transportation [15], and polycentric cities [11].

In this work, we investigate the existence of general statistical patterns, which may apply across cities, in the quantity and spatial distribution of different urban amenities. These include a wide variety of public spaces and institutions as well as businesses, which all provide different services to urban populations. Examples are restaurants, parks or universities. The diversity, location and quality of urban amenities play a crucial role in shaping urban environments as they have a critical impact on the quality of life and opportunities experienced by urban dwellers. Neighborhoods with scarce access to amenities lose their attractiveness, typically causing the selective relocation of people to more attractive locations. The correlated mobility of many amenities and households that follows, and the patterns of spatial (dis)advantage that may result, is one of the mechanisms generating racial segregation and economic inequality in American cities [16]. Since primarily poor residents who cannot afford to relocate remain in under-served neighborhoods their access to important resources often lowers their access to opportunity and potential for upward mobility [17, 18]. Amenities are also central for generating and supporting economic agglomeration effects. They do that by attracting investment to developing neighborhoods, promoting economic growth, supporting innovation clusters and facilitating businesses linkages in specific urban areas [19].

These spatial heterogeneities in cities are almost always mirrored by amenity distributions, which also constitute one of the most important instruments for public policy to both foster economic development and/or generate more equitable livelihoods and opportunities.

But these general observations have mainly stemmed from local observations and case studies, and rarely by systematic quantitative analysis with the necessary diversity and spatial detail. Until recently, data availability has been the main barrier to comparing the detailed quantities and spatial distribution of amenities in and across cities and consequently, the inequality of access for local populations. Here, we analyze a large new data set containing millions of amenities across the 50 largest metropolitan areas in US to find evidence for general scaling relations between the city's population and its amenities. In particular, we start by showing that the quantity of amenities scales as a power-law with population size across US metropolitan areas. We then move to a finer-grained level of spatial resolution to show that the way amenities are distributed among neighborhoods (proxied by census tracts) scales with the way population is distributed. Finally, we find that within a single census tract the distance to the closest amenity scales with the density of amenities. This allows us to characterize local “amenity deserts” and deficiencies in service provision.

The scaling laws identified here provide the basis for a generative model of amenities in any typical US city given its population size. Thus, such a model proves the expected abundance and composition of the amenity set a city should manifest in the absence of any other local features. While this model is interesting as a baseline expectation, it is precisely these local unique features that affect the relative attractiveness of urban areas, making a particular urban environment more desirable than others. Therefore, we use the deviations from the average amenity scaling model to generate urban signatures that uniquely highlight each actual city's characteristics. This allows us to identify interesting and unexpected abundances of special amenities in some places, and potential deficits in others.

Our findings have the potential to benefit the process of urban planning in general, and detailed land-use planning in particular. They provide objective quantitative measures to evaluate the current performance of a city, benchmark that performance to other cities in the same nation and assess the potential value of alternative planning choices. Specifically, we demonstrate how the scaling laws identified from rich local amenity data can be used in the construction of new cities, and how the deviations from the expected behavior can guide the development of growing cities with certain desirable profiles.

## 2 Materials and methods

Population data was extracted from the 2010 decennial Census, including the geographic boundaries and respective population counts for the 50 most populated metropolitan areas in the US and all census tracts included in their boundaries. We chose Metropolitan Statistical Areas (MSAs) as our main geographical unit of analysis because they are defined as coherent functional urban regions in terms of the flow of people, goods and information (also known as integrated labor markets). As a finer-grained unit of analysis inside cities, we chose census tracts due to their consistent definition across the nation, which allows for comparative analysis in terms of similar population sizes (between 1200 and 7500 residents). For this reason urban census tracts are often used as proxies for neighborhoods.

Amenities data was extracted from Google Places API (2012). The resulting data set consists of the geographical location (latitude and longitude) and land use type (e.g., bar, restaurant, park) for approximately 2.5 million amenities within the boundaries of our 50 MSAs. This data set includes all land uses specified in the original source (except residential and industrial) with a standardized index of 78 types that are globally consistent. For more details on the data extraction process see Additional file 1, Sect. 1.

Unlike the administrative data provided by local governments, which typically includes six to ten land use types that vary between cities, the Google Places data set provides a unique opportunity to study land use patterns at high spatial resolution using a consistent index across cities.

In the Figures, we color code the 78 amenity types using five main groups: commercial (red), e.g., grocery stores, restaurants, bars, shopping malls; services (orange), e.g., banks, doctors, beauty salons, real estate agencies; offices (yellow), e.g., accounting offices, lawyers; open spaces (green), e.g., parks; public facilities (blue), e.g., police stations, museums, universities, hospitals, libraries.

### 3 Results

#### 3.1 Scaling relations for urban amenities

##### 3.1.1 Quantity of amenities

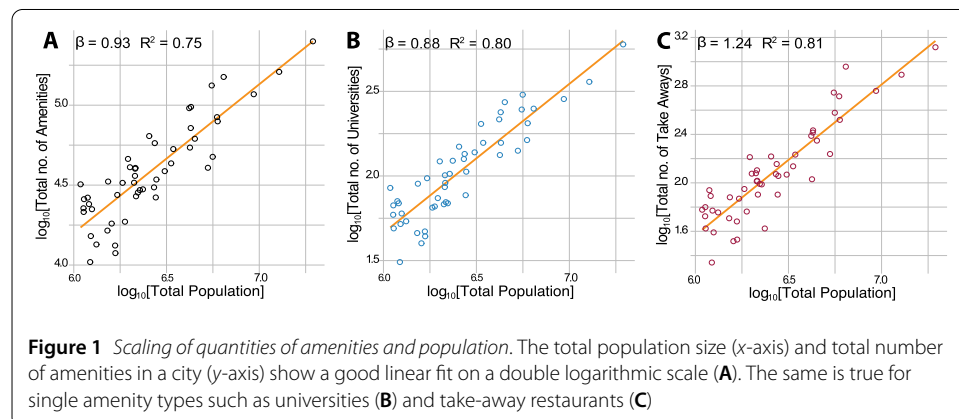
In their work, Bettencourt et al. [1, 5, 8] show that important demographic, socioeconomic, and behavioral urban indicators are, on average, well described by non-linear scaling functions of a city's population size, expressing both increasing returns and economies of scale. The existence of such scaling has also been shown to be quantitatively consistent across a large number of different nations [13, 20–24] and times [25–27].

In this work we investigate whether similar general scaling relations applies to urban amenities. If so, given the population size of a city, we would be able to predict the expected quantities for different amenity types. We recognize from the outset that amenity type specifications are, to some extent, dependent on local and contextual factors such as a nation's level of economic development, its technologies and culture. Moreover, the various types and quality of amenities (as restaurants, museums, etc.) differ between countries. The present study is focused on empirical information for US cities from Google Places. Because this type of technology is being expanded globally, there is a real prospect that the present analysis can soon be expanded globally to many different national urban systems and that these apparent differences can be better assessed in the near future.

To set the stage, we apply standard scaling analysis [5] to find that the total number of amenities is indeed well described on average by power-law scaling with city population size (adjusted  $R^2 = 0.75$ , Fig. 1A). More formally, this power-law scaling function can be written as a linear equation on a double logarithmic scale:

$$\log Y_c(t) = \log Y_0(t) + \beta \cdot \log(n_c(t)) + \bar{e}_c(t), \quad (1)$$

where  $c$  labels the city (MSA),  $Y_c(t)$  is the total number of amenities in that city, observed at time  $t$ .  $n_c$  is city  $c$ 's population size,  $\log Y_0(t)$  is the overall intercept (in logarithmic scale) for the fit across cities and  $\beta$  is the scaling exponent. The quantities  $\bar{e}_c(t)$  are residuals or deviation from the general scaling fit; their average over cities is zero:  $\sum_c \bar{e}_c(t) = 0$ . These residuals capture local factors in each city, beyond the general tendency for amenities to vary with city size across urban areas. All these quantities are in general time dependent, so that the scaling fit applies cross-sectionally across cities at the same time (same year). The scaling exponent  $\beta$  represents the average elasticity of amenities to city population



size,  $\beta = \frac{d \log Y_c}{d \log n_c}$ , is expected to not vary much in time (see [14]). In general, we will be concerned with analysis at a single time, so we will drop the time dependence of scaling parameters for simplicity of notation.

The above equation can be written as a power-law function, where  $Y_0(t)$  is the normalization constant or prefactor and  $\beta$  is the scaling exponent, reflecting the pattern common to all cities (the choice of the log base is arbitrary, but here we use  $\log_{10}$  as the standard):

$$Y_c = Y_0(t) 10^{\tilde{e}_c} \cdot n_c^\beta = Y(n_c) 10^{\tilde{e}_c}, \quad (2)$$

where  $Y(n_c) = Y_0 n_c^\beta$  is the scaling law, shown as the solid line in Fig. 1.

Next, we disaggregate amenities by type and location in order to characterize their detailed spatial statistics. When considering each amenity type separately, only about 70% show good power-law scaling patterns ( $R^2 \geq 0.6$ , see full regression results in Additional file 1, Table S1). The remaining 30%, which include types such as embassies, RV parks, cemeteries and airports do not show a good fit with population size. The majority of these non-scaling amenity types are public facilities (with the exception of health and education), which are controlled by government agencies. Their development is costly in terms of both resources and time, they are mostly land-intensive, and their demand is relatively indifferent to market forces, as they provide essential services to the urban community. Additionally, large retail facilities, such as shopping malls, department stores, lodging and car services, also do not show a good fit with population size, possibly due to being very elastic and heterogeneous in terms of their size and quality, which is not captured by simply counting their numbers.

More specifically, we observe that the total number of amenities in a city shows a sub-linear scaling ( $\beta = 0.93$ , 95% CI = [0.89, 0.97], where the 95% CI of  $\beta$  is entirely below 1) with population size. This kind of sub-linear scaling suggests a phenomenon known as *economies of scale*. For amenities, it means that as the population grows, a city requires fewer new amenities per capita because the existing amenities can be shared to some extent. This may not be typical of all amenity types, and will depend on their size and capacity to serve larger populations with similar resources.

These variations in use per capita with city population size can be assessed by the type specific scaling exponents, which are the elasticities of demand relative to population. When examining amenity types separately we find that only 22% show sub-linear scaling with population size, for example universities ( $\beta = 0.88$ , 95% CI = [0.78, 0.98], Fig. 1B), movie theaters ( $\beta = 0.88$ , 95% CI = [0.79, 0.97]), and libraries ( $\beta = 0.91$ , 95% CI = [0.85, 0.97]). In contrast, approximately 74% of amenity types show linear scaling (95% CI of  $\beta$  intersects with 1) and another 4% show super-linear scaling (95% CI of  $\beta$  is entirely above 1). Amenity types with linear scaling usually provide services related to individual needs that grow proportionally with population size, as grocery stores ( $\beta = 1.01$ , 95% CI = [0.85, 1.17]) and restaurants ( $\beta = 1.05$ , 95% CI = [0.90, 1.20]). On the other hand, super-linear scaling behavior may be the result of network effects and related services that support changes in lifestyles in terms of constraints on the use of time and higher incomes. This may also correspond to situations when the city is better established and entrepreneurs and developers can invest in amenities that are not daily necessities. Examples of amenities with such super-linear behavior include, among others, take-away restaurants ( $\beta = 1.24$ , 95% CI = [1.07, 1.41], Fig. 1C), gyms ( $\beta = 1.18$ , 95% CI = [1.02, 1.34]) and travel

agencies ( $\beta = 1.29$ , 95% CI = [1.07, 1.51]). For the full regression values see Additional file 1, Table S1.

These results constitute a general multidimensional baseline for the quantity of amenities in US urban areas that can now be tracked over time and better understood in terms of their uses and dynamics, as previously observed for example by using a census of business types [28]. However, the distribution of amenities inside cities is very heterogeneous spatially, an issue to which we now turn.

### 3.1.2 Spatial distribution of amenities

We now leverage the rich location information in our data sets to analyze the spatial distribution of amenities within cities. We will want to relate this information to demographic variables, which are available from the US Census for small areas. To do this, we will work at the level of Census tracts, which are small areas that tile the entire territory of the US, with an average population of 4000 people. These units are often considered reasonable proxies for neighborhoods. In large urban areas, there will be typically several hundred or even thousands of such tracts. For reference, the New York City MSA (the largest in the country) has 4784 census tracts.

We first examine the correlation between the spatial density of amenities (number of amenities in a tract divided by its area) and the corresponding population spatial density. We use densities instead of bare counts because the construction of census tracts varies their area (and thus their density), while attempting to keep their population within a narrow range. This analysis results in a relatively weak correlation ( $R^2 \sim 0.41$ ). This finding suggests that an explicitly statistical approach to the variation of these densities is necessary, which can take into account the differences in abundance across tracts.

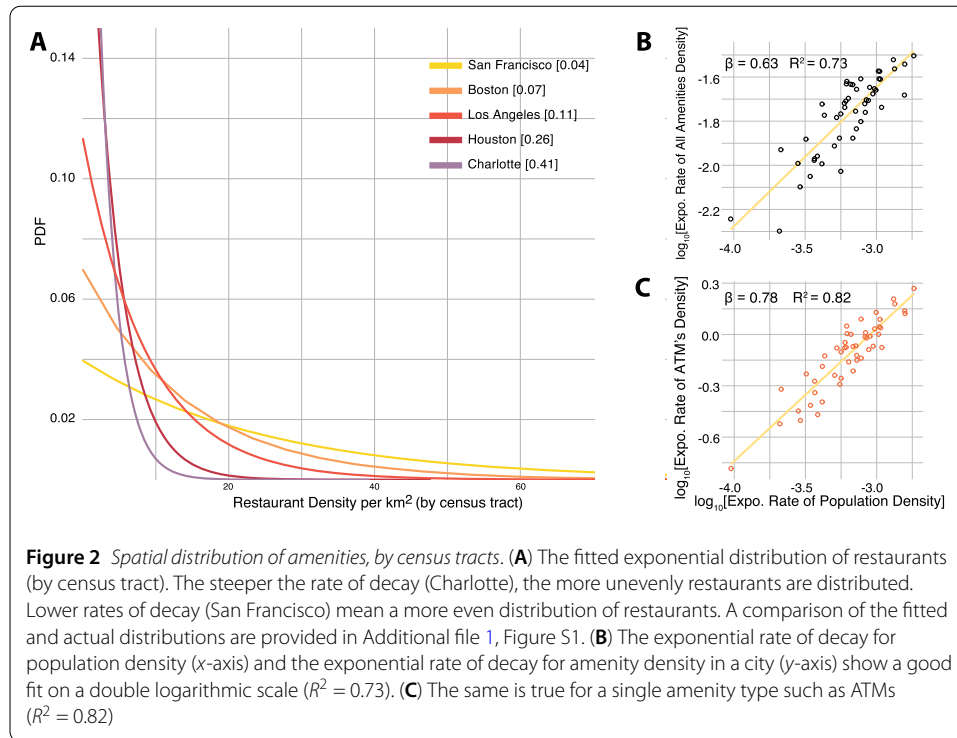
The most general distribution (maximum entropy) consistent with a city's average amenity density per tract  $\langle x \rangle$  (in units of amenities/km<sup>2</sup>) is an exponential distribution of the form [29]:

$$f(x) = \lambda e^{-\lambda x}, \quad (3)$$

where  $x$  is the spatial density of amenities (or people) across census tracts, calculated by dividing the number of amenities (or people) within a tract by the tract's land area. Thus, up to an overall normalization,  $f(x)$  is proportional to the number of census tracts with density  $x$ . The quantity  $\lambda$ , estimated through the maximum entropy procedure as  $\lambda = 1/\langle x \rangle$ , is the exponential rate quantifying the decay in probability with higher amenity densities. It has units of inverse  $x$  (area per amenity), giving us the average territory size of each amenity type in each city. The higher this rate (larger territory), the steeper the decay, meaning that there will be many more tracts with low amenity density  $x$  and that those with larger amenity density  $x$  will be exponentially rare.

Different cities are characterized by different values of  $\lambda$ , meaning that in some cities amenities are more evenly distributed (lower  $\lambda$ ) than in others (higher  $\lambda$ ). For example, Fig. 2A shows that if you were to walk around Houston, you would have a hard time finding a restaurant in most areas, unless you are in the city center or in a few other very specific areas where they are distributed more densely. Hence, the distribution of restaurants in Houston is highly uneven (higher rate of decay,  $\lambda = 0.26$  km<sup>2</sup>/restaurant). In contrast, if you were to walk around the neighborhoods of San Francisco, you would have a better chance

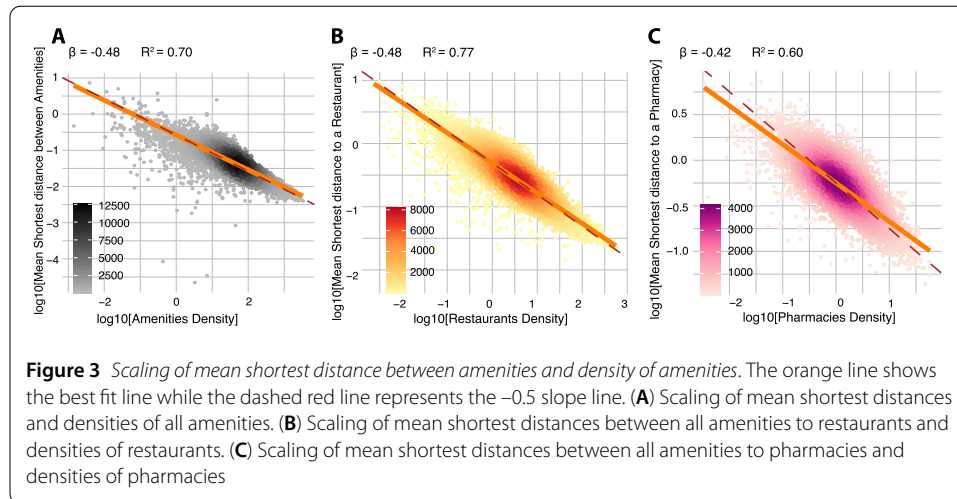




of finding a restaurant in more areas of the city (smaller territory  $\lambda = 0.04 \text{ km}^2/\text{restaurant}$ ). Thus, a higher density of amenities also means that statistically there will be more places (tracts) in which to find a substantial number of services, whereas a lower density intensifies *amenity deserts*, i.e. many areas in which there are none or very few.

Next, we test whether the spatial distribution of population in a given city can explain its spatial distribution of amenities. To do so, we fit the two rates for all 50 metropolitan areas. We find a clear power-law scaling between the two rates of decay (adjusted  $R^2 = 0.73$ , Fig. 2B). The strong correlation between the spatial distributions of population and amenities at the city level, put together with the weak correlation at the tract level, suggest that if a city has a highly populated district, it will also have a highly dense commercial district, but they will not necessarily coincide in space (many US cities have Central Business Districts, with high amenity density but low resident population). In other words, areas that are packed with activities, such as city centers, are usually not the highly dense dwelling neighborhoods, but these commercial areas serve multiple urban neighborhoods including the highly dense ones. This finding suggests that even after years of promoting mixed-use urban development across the US, cities can still show high rates of segregation between residential and commercial districts.

When examining each amenity type separately, we find that the spatial distribution of approximately 80% of the 78 amenity types shows a good fit with population distribution ( $R^2 > 0.5$ ), for example ATMs ( $R^2 = 0.82$ , Fig. 2C). The remaining 20% that do not scale ( $R^2 < 0.5$ ) mostly include public facilities and open spaces such as airports, museums, hospitals and parks (for full regression values, see Additional file 1, Table S2). Indeed, some of these public facilities can be few and concentrated in specific areas of the city, and therefore not track at all the spatial distribution of the city's population.



### 3.1.3 Distance between amenities

We conclude our empirical analysis by examining how amenities are located with respect to each other within census tracts. As a measure of amenities co-location, we calculate the average shortest distance from one amenity to another in that tract.<sup>1</sup> We then compute the mean shortest distance by iterating over all amenities in the tract, finding the distance to the closest amenity, and computing the mean.

On purely dimensional grounds we expect the mean shortest distance,  $d_{\min}$ , between amenities to relate to the spatial density as  $d_{\min} = x^{-1/2}$ , so that it should scale with a negative sub-linear exponent of the density of amenities. Indeed, as shown in Fig. 3A, when analyzing our real-world dataset, we observe a very strong correlation between the density of amenities (of any type) in a census tract and the mean shortest distance between amenities ( $R^2 = 0.70$ , on log log scale). Moreover, we observe an estimated exponent of  $-0.48$ , which is very close to the general  $-1/2$  expectation.

This sort of expectation generalizes to specific amenity types. Specifically, when we examine the correlation between the density of amenities of type  $T$  and the mean shortest distance between an amenity of any type and an amenity of type  $T$ , we again find a clear scaling relationship. For example, the mean shortest distance from amenities of any type to restaurants scales with the density of restaurants ( $R^2 = 0.77$ ) as shown in Fig. 3B. Such scaling holds for almost all amenity types (adjusted  $R^2 > 0.5$ , see Additional file 1, Table S3).

However, in all considered cases, the observed exponents deviate somewhat from the theoretical exponent of  $-0.5$ , and vary between amenity types (ranging between  $-0.37$  to  $-0.48$ ). For example, the estimated exponent for amenities of any type is  $-0.48$  (99% CI  $[-0.49, -0.48]$ ), the exponent for restaurants is  $-0.48$  (99% CI  $[-0.48, -0.47]$ ), and the exponent for pharmacies is  $-0.42$  (99% CI  $[-0.43, -0.42]$ ). This suggests that for tracts with lower density, the mean shortest distance is shorter than expected from a random placement at fixed density, possibly suggesting spatial clustering, and for tracts with high density the mean shortest distance is longer than expected in the same sense. This phenomenon is more extreme for certain amenity types, such as pharmacies. These exponent

<sup>1</sup>Geographic distance between two points was computed in kilometers as great-circle distance, or surface distance, representing the shortest path between the two points on the surface of a sphere. For that purpose we assumed a static radius of the earth of 6367 km.



deviations therefore express how the location of different amenity types is determined by considerations beyond overall density and composition, including zoning, road network layout limitations and agglomeration effects.

### 3.2 Deviations from scaling as urban signatures

We have now described three different scaling patterns for urban land uses relating to quantities of amenities, their spatial distribution across census tracts, and the distances to the closest amenity within a census tract. These scaling relations represent the average behavior a city would manifest if it were to follow the common pattern shown in urban systems across the US. This provides a generative statistical model for amenity allocation for any US city with a given population and their spatial location. We can compare the local characteristics of each actual city by measuring how they deviate from these scaling patterns. This provides a multidimensional benchmark in “amenity space” to assess the performance of a city versus the expected average behavior across the nation.

Many urban indicators are expressed as *per capita* measures, implicitly assuming that the captured phenomenon grows linearly with population. However, a per capita measure fails in capturing the deficiencies as well as strengths of a city with respect to a phenomenon that has a non-linear dependence on population size or density, making it an unfit measure to compare between cities with different population size. A more appropriate measure of this type of performance is the deviation from the expected behavior as described by non-linear scaling relation across many places. This leads to the concept of Scale-Adjusted Metropolitan Indicators, SAMI [8, 13, 30], which we now extend to local amenities in cities.

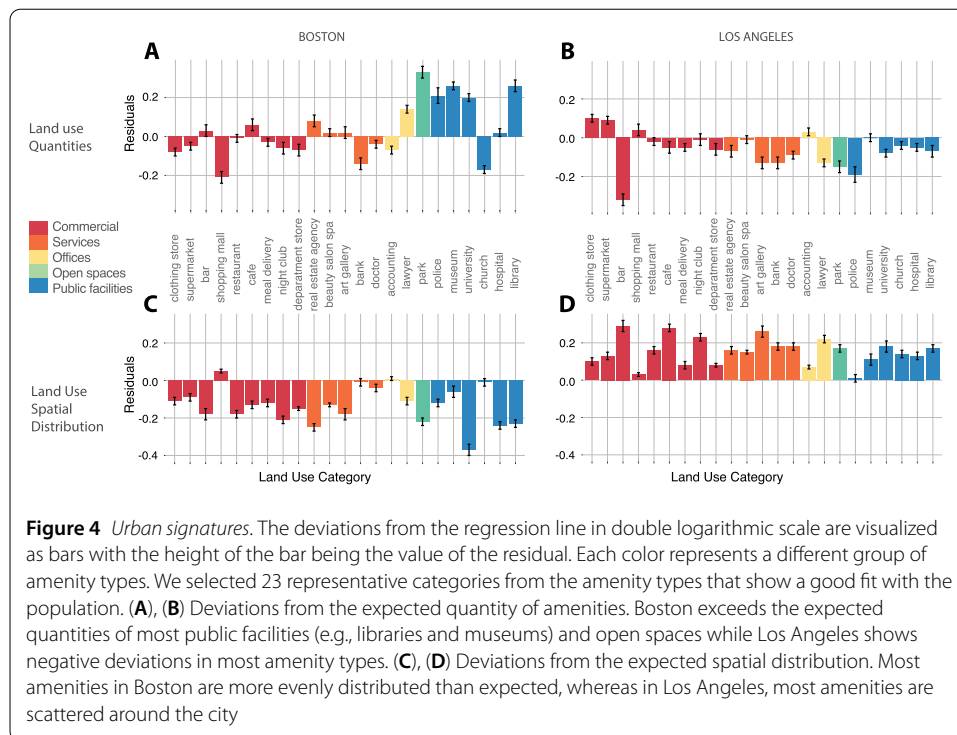
More formally, the deviation of a city is computed as its residual from the fitted regression line

$$\bar{e}_c = \log_{10}(Y_c) - \log_{10}(Y(n_c)) = \log_{10}\left(\frac{Y_c}{Y_0 n_c^\beta}\right). \quad (4)$$

Computing a city’s deviations across all amenity types results in its own “urban signature” (see for example, Fig. 4) in the form of a vector of performance indices. Such an urban signature provides, in our view, meaningful insights about the characteristics of that city with respect to the way it is serving its citizens, independent of its population size.

For example, the first scaling law we identified describes the relationship between population size and the quantity of amenities in the city. The deviations for quantity of amenities are positive when the number of amenities of a given type is higher than expected with respect to a city’s population size and negative otherwise. Examining Boston’s evidence (Fig. 4A and Additional file 1, Table S4), we see positive deviations from the expected average in parks, museums and universities, showing that Boston has more open spaces and public facilities than the average US city of its size. By contrast, a sprawling city such as Los Angeles (Fig. 4B and Additional file 1, Table S4) shows positive deviations in supermarkets and clothing stores, while showing negative deviations in all other amenity types. Thus, the analysis suggests that Los Angeles shows wide deficits in terms of many public facilities and services.

The second type of scaling relation describes the relationship between the spatial distribution of population and amenities across neighborhoods (census tracts). In this case, a negative deviation (e.g., universities in Boston, Fig. 4C and Additional file 1, Table S5)

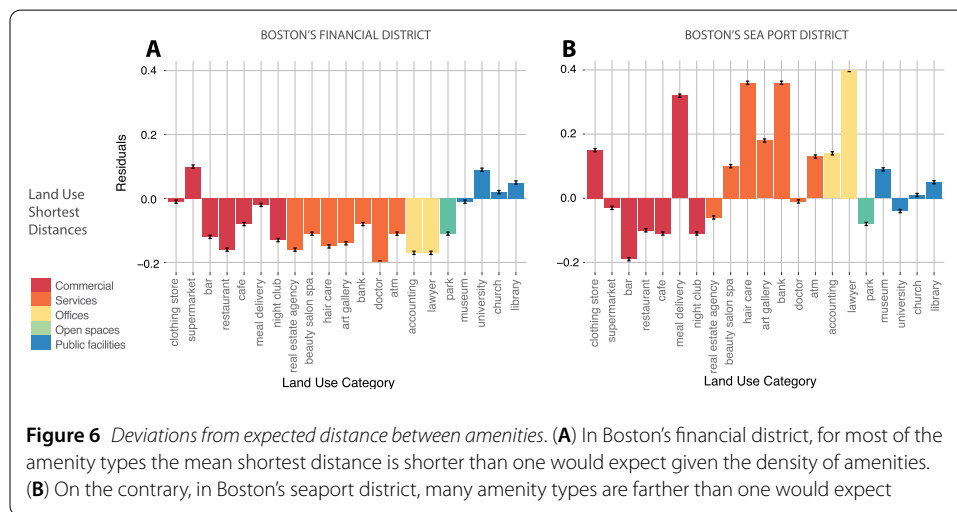
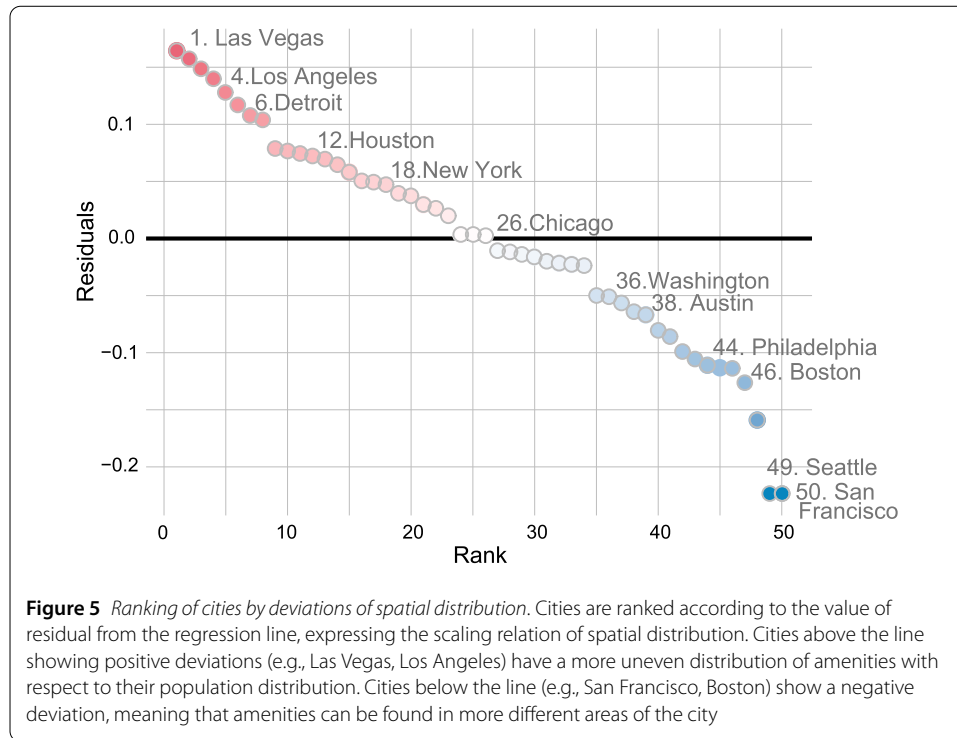


indicates a spatial distribution of amenities with a lower rate of decay than expected, suggesting that an amenity is more evenly distributed across a city than expected by the average generative model. By contrast, a positive deviation (e.g., cafes in Los Angeles, Fig. 4D and Additional file 1, Table S5) indicates a higher rate of decay than expected, suggesting that an amenity is less evenly located and is likely to be observed only in particular locations. These examples also convey some of the zeitgeist of each of these two cities, in out view, a feature long sought after by social psychologists and urbanists [31].

Proceeding in this way to observe the deviations of spatial distribution for all cities (Fig. 5 and Additional file 1, Table S6), we find that in general, sprawling cities such as Las Vegas and Detroit tend to demonstrate positive deviations in the distribution rate for most amenity types, while denser cities such as San Francisco and Seattle tend to demonstrate negative deviations for most types. These findings support the idea that amenities are typically distributed more unevenly in sprawling cities, usually only found within city centers from where they serve the entire urban area. This of course creates more load on transportation infrastructures and exposes urban communities of such cities to issues of differential spatial disadvantage. By contrast, in denser cities many amenities can also be found in small commercial centers outside the core, serving small communities in the periphery, often closer to their places of residence.

It is worth noting that deviations in spatial distribution are in general statistically independent of amenity quantities. Hence, cities that show negative deviations in quantities may show positive deviations in spatial distributions and vice versa (see Fig. 4).

The third scaling law we identified describes the relationship between the density of a given amenity type and the mean shortest distance to that amenity. In this case, the deviations are positive when the mean shortest distance is larger than expected, and negative when the distance is smaller than expected. As an example, we plotted the deviations



of two specific neighborhoods in Boston: the financial district and the seaport district, which are comparable in their day-to-day functioning. Figure 6 (full deviation values in Additional file 1, Table S7) reveals that most amenity types in the financial district are located at a shorter distance than expected by the density of the region (Fig. 6A). On the other hand, most of the amenities in the seaport district (a region currently under development) are farther away than expected, suggesting that there is still room for further development (Fig. 6B) and synergy. In particular, we find that take-away restaurants—an important amenity in a business district such as the seaport district—are not highly accessible and remain located farther away from each other than one would expect.

### 3.3 Data-driven land use planning

Scaling relations across and within cities hold great potential as general statistical models that can benchmark and guide development processes and assist developers, policy makers and city planners in assessing urban areas in detail. From real estate development, to zoning and land use choices, to environmental policies and well-being metrics, our findings can be applied in diverse planning and policy processes to measure, assess and benchmark the performance of cities and their neighborhoods.

The most immediate and obvious implications of our findings is for the process of land-use planning, one of the core functions of urban planning and design. Land-use planning determines how many amenities will be built where, with the goal of servicing residents, businesses, daily commuters and population groups that the city wishes to retain or attract. The amenity quantities, spatial distribution and mix of types determined as part of a land use plan play a key role in shaping the character of urban areas and cities as a whole.

The traditional land use planning process is typically structured in two main stages. In the first phase, expected population and economic growth are estimated based on local data and analysis of past trends. Then, the quantities and total area required for each coarse land use category (e.g., residential, commercial) are assessed with the goal of sufficiently servicing the current and expected population and businesses. In the second phase, a zoning ordinance divides the city into relatively large zones, controlling the permitted land use categories, their densities, and their coarse locations within each zone [32].

This procedure has two major limitations. First, quantities and spatial allocation are determined at the macro-level using typically 6–10 coarse land use categories and relatively large zones. Therefore, fine-grained choices are left entirely in the hands of market forces [33], without giving a chance to policy makers to fulfill their objective of correcting market failures and providing resources to under served communities [34]. Second, in both planning phases, assessing the current performance of urban areas is a challenging task without the support of proper quantitative tools [35]. Consequently, the planning process is left susceptible and exposed to arbitrary influences, often failing to reflect issues of equity and general public interest.

Understanding how to deploy quantitative methods in the process of land use planning is a task that has ignited the minds of scientists, urban planners and economic geographers for most of the last century [36–40]. Most of these efforts formally modelled the urban environment given a set of constraints, and developed algorithms to find the optimal location for land uses in cities. However, when city data became available, it was clear that such models suffered from limited empirical validation and they could not be calibrated to represent the complexity of real cities [41]. Moreover, these tools did not capture or support the iterative nature of the planning process [42]. Recent results which identified universal patterns of cities can highly benefit the urban planning process by providing guidelines, baselines for reference, and performance estimates of proposed plans. However, the process of land use planning has yet to capitalize on the predictive power emerging from these universal patterns of cities as complex systems.

Our work can bring several significant advantages to the land use planning process.

- Quantitative metrics: the identified scaling laws and deviations can support the definition of quantitative metrics to evaluate the performance of cities, promoting a more objective discussion around the urban development process. In the following

subsections we explain how these scaling laws can be used to support the two phases of land use planning.

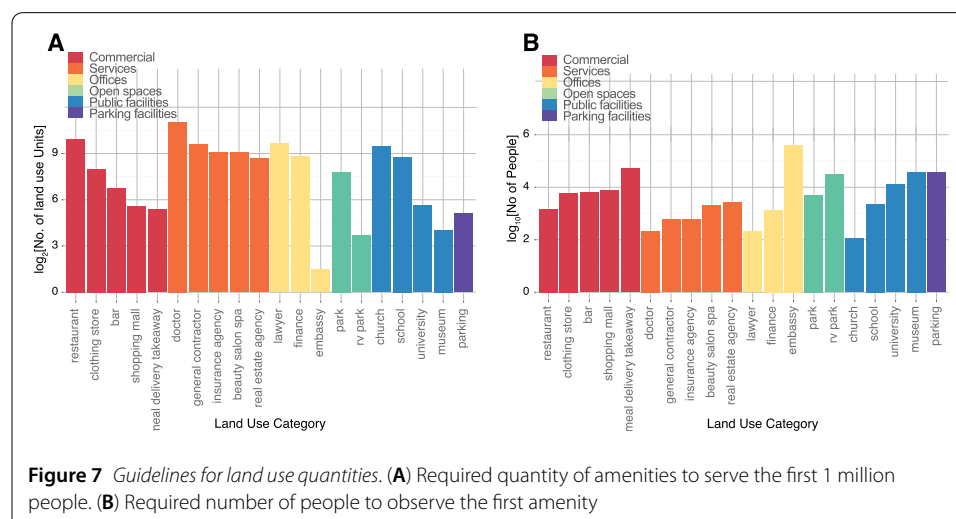
- Granularity of analysis: the unique data set we use allows us to study land uses in much finer grain levels. The data set contains 78 land use types, offering a richer semantic than the usual 6–10 types offered by local government data. Moreover, it captures the fine-grained location of amenities in urban areas, in contrast with current coarse-grain zoning.
- Global comparison: the index of land use types used in our data set is uniform for all cities, allowing a consistent analysis and comparison of cities across the world, where data availability permits.

### 3.3.1 Phase 1: land use quantities

The first stage of each land use plan requires estimating the number of units of each land use type essential to support planning goals. The identified scaling law of quantity of amenities can support the planners in this task. More specifically they can be used when constructing a new city, planning development of a growing city, or promoting changes in a city's performance levels through a new master plan. These cases are somewhat different so that we discuss them separately:

(i) New cities. In new cities and new development projects, scaling laws can be useful for calculating average population thresholds and type sequences, which are the minimum number of people required to support the introduction of new urban activities (see thresholds examples in Fig. 7B). For example, in a scenario when a new city is constructed in a location where there are no accessible services or commercial establishments essential for everyday life, such as supermarkets, banks, gas stations or hospitals these will have to be introduced. All these urban establishments need a sufficient amount of potential customers to support their financial activity in a sustainable manner for the developers to justify their initial investment. In this scenario, developers need a recommended measure of population thresholds to assess how many people their development project has to house in order for it to sustain a type of urban activity.

(ii) Growing cities. Scaling laws can serve as a consistent planning guideline for growing cities, as a tool to assess the rate of development: for the quantity and quality of amenities



and their spatial distributions viz a viz demographic growth trends; or to maintain their current characteristics. More specifically, if a growing city wants to maintain the current level of services, the required quantity of amenities for the new population size can be computed as follows:

$$\log_{10}(Y_c(n_c + g)) = \log_{10}(Y(n_c + g)) + \bar{e}_c, \quad (5)$$

where  $c$  is the city,  $n_c$  is its current population size,  $n_c + g$  is the new population size,  $\bar{e}_c$  is  $c$ 's deviation from the scaling law and  $Y_c(n_c + g)$  is the required quantity of amenities for the new population size.

For a concrete example, consider Boston's recent efforts to develop a 2030 master plan [43]. Assuming planners are satisfied with the city's performance in terms of parks (Boston is ranked 4th in the US), as part of the planning process planners will need to calculate how many additional parks are needed to support the expected population growth. Using the scaling law found for parks ( $\alpha = -3.88$  and  $\beta = 1.03$ ) and the equation above (Eq. (5)), considering the population projection for 2030 of 4,547,611 people ( $n_{\text{Boston}} = 4,494,611$  current inhabitants,  $g = 53,000$  additional people) and the current quantity of parks of  $y_{\text{Boston}} = 2010$  (with a deviation in parks of  $\bar{e}_{\text{Boston}} = 0.33$ ), the city of Boston will need to add only 25 new parks to maintain its current performance and "feel".

(iii) Adjusting a city's character. Communities and planners may want to take actions in order to change a city's character in a way that would best serve its population or a vision of future developments. In this case, the catalog of 'urban signatures' can support planners by providing them a diverse set of city characters, they could use as role models. Once the role model city is chosen, planners can use that city's deviations to calculate the quantity of amenities the city in question needs to have in order to adjust its character. For example, if Los Angeles is interested in creating a new master plan for parks, planners may decide to use Boston's level of performance as their development goal. Los Angeles planners can then quantify how many parks they would need to add to their city in order to reach Boston's performance standard by using the following equation:

$$Y_{\text{target}}(n_c) = Y(n_c) \cdot 10^{\bar{e}_{\text{target}}}, \quad (6)$$

where target is the target city,  $c$  is the city in question,  $n_c$  is the population size of  $c$ ,  $Y(n_c)$  is the fitted quantity of amenities for population size  $n_c$  and  $\bar{e}_{\text{target}}$  is the deviation of the target city. The metropolitan area of Los Angeles has a population of  $n_{\text{LA}} = 12,801,183$ , Boston's performance in parks is represented by a deviation of  $\bar{e}_{\text{Boston}} = 0.33$  and the fitted quantity of parks required for the population size of Los Angeles is  $Y(n_{\text{LA}}) = 2758$ . Plugging these numbers into the scaling relation indicates that Los Angeles needs to add a total of 5895 parks to reach Boston's performance, which translates into 3961 additional new parks!

### 3.3.2 Phase 2: land use allocation

The second phase of land use planning addresses the spatial distribution of land use units across the city. After land use quantities have been assessed and the number of units to be added (or possibly subtracted) to a city has been set, planners need to tackle the issue of land use allocation. Using our findings, given the spatial distribution of a city's population scaling can be used to estimate the spatial distribution of amenities. In our running



example of Boston's master plan, one of the plan's stated goals is to improve accessibility to parks. Indeed, our analysis shows that Boston's rate of decay for parks is relatively high ( $2.6 \cdot 10^5$ ), supporting the plan's claim of a current uneven distribution of parks. To reach the plan's goal, Boston will need to reduce the spatial concentration of its parks, which can be done, for example, by adding new parks only to tracts that currently do not contain any.

Finally, the question of determining a fine-grained location for an amenity might be the hardest and most influential of all in the process of city planning. Location is everything with regard to cities, even in the globalization era [44]. The third scaling law, characterizing the distance between amenities, can provide some guidelines. As a concrete example, looking at the case of the Boston's seaport district (Fig. 6B), we already noted that take-away restaurants have a very high positive deviation in their spatial rate, meaning that they are not highly accessible and located farther away than one would expect. If planners wish to improve the accessibility to such a high-demand amenity in a business district, it would require reducing this positive deviation. This can be achieved by encouraging additional take-away restaurants at a distance smaller than 1.25 km (the current mean shortest distance) from other amenities in the district, or facilitating relocation to more convenient locations. A city should also benefit from such strategies as they can increase not only the quality of services to its inhabitants but also higher tax revenues resulting from more successful businesses. We like to think of this as *amenity oriented development* in analogy to the virtuous cycles between land uses and transportation created by well known strategies of *transit oriented development*.

#### 4 Discussion

We have used a large new data source of urban amenities in US cities to establish a number of new statistical patterns quantifying the abundance, relative composition and location of services in cities and their non-linear dependence on city size. Our analysis reveals the existence of three different types of scaling relations. First, echoing previous research with different data sets [28], the quantity of amenities in a city shows a power-law scaling with population size. Second, the distribution of amenities in a city scales with the distribution of population across census tracts in a city, indicating that if residential areas are distributed unevenly throughout a city, so are the amenity clusters and commercial districts in that city. Third, the mean distance to the closest amenity scales with the density of amenities.

These three scaling laws and associated deviations characterizing each city allow us to build a generative multidimensional model of urban signatures and evaluate cities in terms of their unique local characteristics. Such measures can be used to highlight the current relative strengths and deficiencies of service provision in any given city and its neighborhoods as judged by the same, city-size specific, quantitative standard. It is worth noting that positive and negative deviations do not hold by themselves a judgment on the performance of a city and that context and history are important in any value statement. For example, in Boston, the positive deviation in the number of universities means that they are serving more people than just the residents of Boston, whereas in Detroit, the positive deviations in all amenity types are a manifestation of a shrinking city where people are abandoning their homes faster than businesses and public facilities are closing down.

Our analysis is based on a unique new Google Places dataset, offering a uniformly defined index of 78 amenity types for all US cities. This index allows for a more consistent

analysis and comparison of amenity types than the usual 6–10 types offered by typical local government land-use classifications. However, it also presents some major limitations since the data are partly user generated and partly automatically extracted from Google Street View [45–47]. While these methods are improving fast towards a universal census of amenities and businesses, there are still potential inaccuracies and biases in the data that may create partial misrepresentations of reality. For example, previous studies found Google to over-represent health institutions and lawyers [48]. Nonetheless, when comparing Google Places API to other large scale POI datasets (e.g., Facebook, OSM, Foursquare), Google shows the most consistent point density across regions, the highest accuracy in geographical locations and significantly less falsified POIs, suggesting higher quality data in Google as a result of its data curation efforts [48–50] (Google has a team dedicated to the verification of the data and has set procedures in place to detect fake listings [51]). Future work should repeat our analyses with other datasets of amenities in addition to Google API (both user generated and official) to strengthen our findings' validity and clarify their limitations in light of potential biases in the data.

The data used in this study was extracted from Google Places API back in 2012 (i.e., almost 10 years ago). While we believe that most of our findings would be similar today, re-extracting the data and re-conducting our analyses is not trivial. In 2018 Google started charging for access to Places API (when exceeding a relatively small number of free queries), making a data extraction process like ours costly. Nevertheless, we are aware of steps Google is taking in order to support the scientific and policy communities' work, so we hope our methodology can be utilized to support such future work at scale.

In a similar context, datasets that can follow amenities over space and time, are expected to grow in quality and scope and expand to more cities across the globe. These developments hold much promise for a future extension of the studies developed here, in terms of their further validation and the observation of amenity dynamics, e.g. during crises and recessions and in cities all over the world.

In this context we expect that our strategy will apply generally, but that new features will become available for analysis. For example, scaling prefactors are typically time and urban system dependent, so that baseline values for amenities will need to be reevaluated in different nations and times. In this sense, levels of economic development and consumption habits in US cities are likely to be found to be quite different from those of cities of the Global South, such as in South America or Africa. Moreover, American cities are relatively young compared to cities in Europe and generally less dense, resulting in different physical morphologies that should affect amenity location patterns and, of course, their spatial concentration.

We used MSAs as the base unit of analysis. MSAs are the gold standard definition of functional urban areas as integrated labor markets and have been defined consistently by the US Census Bureau since the 1960s. The majority of work on urban scaling uses these units of analysis as they more closely approximate theory describing cities as networks that include places of residence and work. Previous research has also shown that scaling results applied to other units of analysis may vary more widely [52, 53]. However, the literature has shown that although scaling signals may be sensitive to urban definitions, they tend to change in a consistent and similar way [53], supporting their significance for policy decisions. Future research should continue to investigate these issues for amenity

abundance and location, using urban units of analysis that are both relevant for theory and policy applications.

Our findings hold, in our view, great potential to more systematically support development and planning practices in the construction of new cities, in growing cities and in urban areas interested in rezoning to promote change. In particular, the task of land use planning should benefit from the quantitative measures produced here, which enable city planners to assess the functional configuration of an urban area as well as to better understand the implications of their planning choices. Such a quantitative approach has the potential to promote a more objective, open and flexible discussion around urban development processes, reducing their exposure to the influence of stakeholders' narrow interests, power imbalances and inequitable market forces.

## Supplementary information

**Supplementary information** accompanies this paper at <https://doi.org/10.1140/epjds/s13688-022-00362-6>.

**Additional file 1.** Supplementary Material. We report additional details on computations and results. (PDF 174 kB)

## Acknowledgements

Not applicable.

## Funding

Not applicable.

## Abbreviations

US, United States; API, Application Programming Interface; RV, Recreational Vehicle; SAMI, Scale-Adjusted Metropolitan Indicators.

## Availability of data and materials

Data sufficient to reproduce all results in this paper will be made available upon request.

## Declarations

### Competing interests

The authors declare that they have no competing interests.

### Author contribution

TK, LR and ES conceived and designed the study, conducted the preliminary analyses and wrote the first draft of the paper. TK, LB and ES reviewed the manuscript, raised comments, performed additional analyses and edited its final version. All authors approved the final version of the manuscript.

### Author details

<sup>1</sup>School of Public Policy and Urban Affairs, Northeastern University, Boston, MA, USA. <sup>2</sup>Department of Industrial Engineering, Tel-Aviv University, Tel-Aviv, Israel. <sup>3</sup>Mansueto Institute for Urban Innovation, Department of Ecology & Evolution, Department of Sociology, University of Chicago, Chicago, IL, USA.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 18 November 2021 Accepted: 8 August 2022 Published online: 03 September 2022

## References

1. Bettencourt LMA (2013) The origins of scaling in cities. *Science* 340(6139):1438–1441. <https://science.sciencemag.org/content/340/6139/1438>
2. Jacobs J (1961) *The life and death of great American cities*. Random House, New York
3. Weaver W (1958) *Annual report*. Rockefeller Foundation, New York
4. Alexander C (1964) A city is not a tree. *Architectural Forum*
5. Bettencourt LM, Lobo J, Helbing D, Kuhnert C, West GB (2007) Growth, innovation, scaling, and the pace of life in cities. *Proc Natl Acad Sci* 104(18):7301–7306
6. Batty M (2008) The size, scale, and shape of cities. *Science* 319(5864):769–771

7. Samaniego H, Moses ME (2008) Cities as organisms: allometric scaling of urban road networks. *J Transp Land Use* 1(1):21–39
8. Bettencourt LM, Lobo J, Strumsky D, West GB (2010) Urban scaling and its deviations: revealing the structure of wealth. *PLoS ONE* 5(11):e13541
9. Louf R, Barthelemy M (2008) Modeling the polycentric transition of cities. *Phys Rev Lett* 11(19):198702
10. Pan W et al (2013) Urban characteristics attributable to density-driven tie formation. *Nat Commun* 4:1961
11. Louail T, Lenormand M, Ros OGC, Picornell M, Herranz R, Frias-Martinez E et al (2014) From mobile phone data to the spatial structure of cities. *Sci Rep* 4:5276
12. Arcaute E, Hatna E, Ferguson P, Youn H, Johansson A, Batty M (2015) Constructing cities, deconstructing scaling laws. *J R Soc Interface* 12(102):20140745
13. Alves LG, Ribeiro HV, Lenzi EK, Mendes RS (2013) Distance to the scaling law: a useful approach for unveiling relationships between crime and urban metrics. *PLoS ONE* 8(8):e69580
14. Bettencourt LMA, Yang VC, Lobo J, Kempes CP, Rybski D, Hamilton MJ (2020) The interpretation of urban scaling analysis in time. *J R Soc Interface* 17(163):20190846. <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2019.0846>
15. Tachet R, Sagarra O, Santi P, Resta G, Szell M, Strogatz S et al (2016) Scaling law of urban ride sharing. arXiv preprint. [arXiv:1610.09921](https://arxiv.org/abs/1610.09921)
16. Wilson WJ (1987) *The truly disadvantaged: the inner city, the underclass, and public policy*. University of Chicago Press, Chicago
17. Small ML, McDermott M (2006) The presence of organizational resources in poor urban neighborhoods: an analysis of average and contextual effects. *Soc Forces* 84(3):1697–1724
18. Chetty R, Hendren N, Kline P, Saez E (2014) Where is the land of opportunity? The geography of intergenerational mobility in the United States. *Q J Econ* 129(4):1553–1623
19. Krugman P (1991) Increasing returns and economic geography. *J Polit Econ* 99(3):483–499
20. van Raan AFJ, van der Meulen G, Goedhart W (2016) Urban scaling of cities in the Netherlands. *PLoS ONE* 11(1):e0146775. <https://doi.org/10.1371/journal.pone.0146775>
21. Bettencourt LMA, Lobo J (2016) Urban scaling in Europe. *J R Soc Interface* 13(116):20160005. <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2016.0005>
22. Brelsford C, Lobo J, Hand J, Bettencourt LMA (2017) Heterogeneity and scale of sustainable development in cities. *Proc Natl Acad Sci* 114(34):8963–8968. <https://www.pnas.org/content/114/34/8963>
23. Zuend D, Bettencourt LMA (2019) Growth and development in prefecture-level cities in China. *PLoS ONE* 14(9):e0221017. <https://doi.org/10.1371/journal.pone.0221017>
24. Sahasranaman A, Bettencourt LMA (2019) Urban geography and scaling of contemporary Indian cities. *J R Soc Interface* 16(152):20180758. <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2018.0758>
25. Ortman SG, Cabaniss AHF, Sturm JO, Bettencourt LMA (2015) Settlement scaling and increasing returns in an ancient society. *Sci Adv* 1(1):e1400066. <https://advances.sciencemag.org/content/1/1/e1400066>
26. Cesaretti R, Lobo J, Bettencourt LMA, Ortman SG, Population-Area SME (2016) Relationship for medieval European cities. *PLoS ONE* 11(10):e0162678. <https://doi.org/10.1371/journal.pone.0162678>
27. Hanson JW, Ortman SG, Bettencourt LMA, Mazur LC (2019) Urban form, infrastructure and spatial organisation in the Roman Empire. *Antiquity* 93(369):702–718
28. Youn H, Bettencourt LMA, Lobo J, Strumsky D, Samaniego H, West GB (2016) Scaling and universality in urban economic diversification. *J R Soc Interface* 13(114):20150937. <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2015.0937>
29. Jaynes ET (2003) *Probability theory: the logic of science*. Cambridge University Press, Cambridge
30. Alves LG, Mendes RS, Lenzi EK, Ribeiro HV (2015) Scale-adjusted metrics for predicting the evolution of urban indicators and quantifying the performance of cities. *PLoS ONE* 10(9):e0134862
31. Milgram S (1970) The experience of living in cities. *Science* 167(3924):1461–1468. <https://science.sciencemag.org/content/167/3924/1461>
32. Hoch C, Dalton L, So F (2000) *The practice of local government planning*
33. Alexander C (1987) *A new theory of urban design*, vol 6. Oxford University Press, London
34. Bator FM (1958) The anatomy of market failure. *Q J Econ* 72:351–379
35. Geertman S, Stillwell J (2004) Planning support systems: an inventory of current practice. *Comput Environ Urban Syst* 28(4):291–310
36. Von Thünen JH (1826) *The isolated state*. Perthes, Hamburg
37. Alonso W et al (1964) *Location and land use: toward a general theory of land rent*
38. Fujita M (1989) *Urban economic theory: land use and city size*. Cambridge University Press, Cambridge
39. Christaller W (1966) *Die Zentralen Orte in Suddeutschland*, Gustav Fischer (1933). Prentice Hall, New York. English translation by CW Baskin, *Central Places in Southern Germany*
40. Lösch A et al (1954) *Economics of location*. Yale University Press, New Haven
41. Klosterman RE (2008) A new tool for a new planning: the what if?™ Planning support system. In: *Planning support systems for cities and regions*. Lincoln Inst of Land Policy
42. Harris B (1989) Beyond geographic information systems. *J Am Plan Assoc* 55(1):85–90
43. City of Boston (2016) *Imagine Boston 2030*. <http://imagine.boston.gov/>
44. Fujita M, Thisse JF (2013) *Economics of agglomeration: cities, industrial location, and globalization*. Cambridge University Press, Cambridge
45. Yu Q, Szegedy C, Stumpe MC, Yatziv L, Shet V, Ibarz J et al (2015) Large scale business discovery from street level imagery
46. Wojna Z, Gorban A, Lee DS, Murphy K, Yu Q, Li Y et al (2017) Attention-based extraction of structured information from street view imagery
47. Movshovitz-Attias Y, Yu Q, Stumpe MC, Shet V, Arnoud S, Yatziv L (2015) Ontological supervision for fine grained classification of street view storefronts. In: *IEEE conference on computer vision and pattern recognition (CVPR)*
48. Hochmair HH, Juhász L, Cvetojevic S (2018) Data quality of points of interest in selected mapping and social media platforms. In: *LBS 2018: 14th international conference on location based services*. Springer, Berlin, pp 293–313

49. Deng X, Quantitative NS (2017) Comparison of open-source data for fine-grain mapping of land use. In: Proceedings of the 3rd ACM SIGSPATIAL workshop on smart cities and urban analytics—UrbanGIS'17. ACM Press, New York, pp 1–8. <http://dl.acm.org/citation.cfm?doi=3152178.3152182>
50. Kaufmann T, Schumann A, Ahrend R (2019) Measuring accessibility to services across cities. OECD working paper
51. Huang DY, Grundman D, Thomas K, Kumar A, Bursztein E, Levchenko K et al (2017) Pinning down abuse on Google Maps. In: International World Wide Web conference committee
52. Arcaute E, Hatna E, Ferguson P, Youn H, Johansson A, Batty M (2015) Constructing cities, deconstructing scaling laws. *J R Soc Interface* 12(102):20140745
53. Cottineau C, Hatna E, Arcaute E, Batty M (2017) Diverse cities or the systematic paradox of urban scaling laws. *Comput Environ Urban Syst* 63:80–94

**Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)