CHAPTER 2

The Global Sample of Cities

THE 2010 UNIVERSE OF 4,231 CITIES

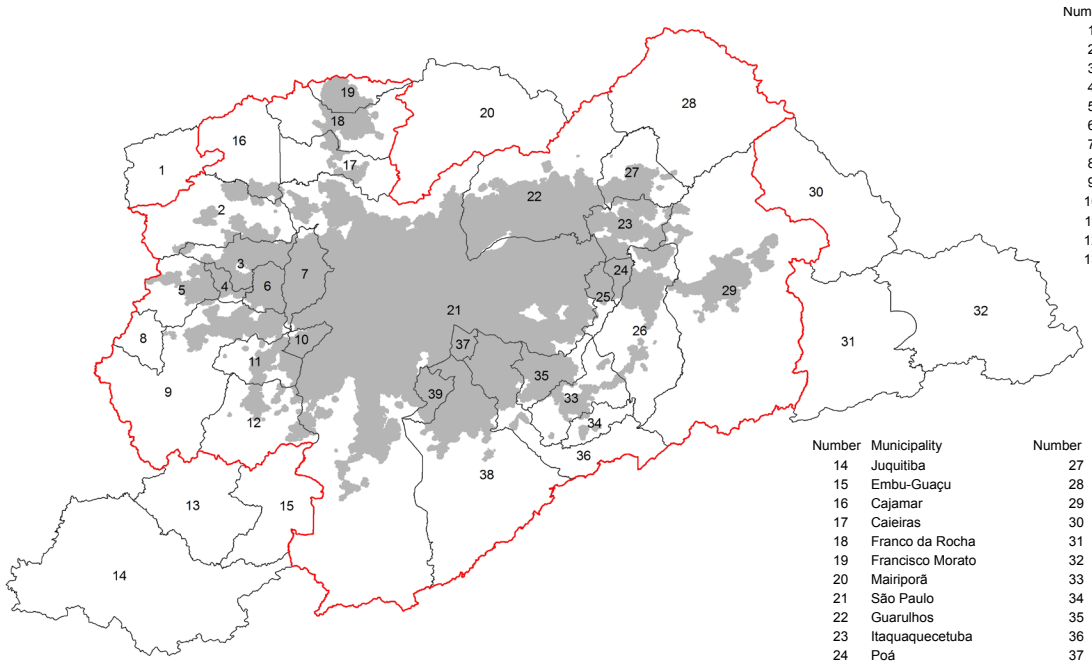
The study of global urbanization trends from the perspective of countries in which national censuses differentiate between urban and rural populations yields important insights and policy prescriptions as we have seen in the previous chapter. Yet these results are limited because national urban population statistics lump all cities, large and small, together. We can advance our knowledge and understanding of global urbanization attributes and trends if we focus our attention on all the cities in the world, rather than on all countries, as units of analysis.

Identifying the universe of all cities in a given year requires a definition of what constitutes a city. Since cities have been defined along many different dimensions, any such definition involves a choice, or rather a number of choices. Cities can be distinguished from hamlets, villages, or towns by population thresholds; they can be identified by their historical centers, their municipal boundaries, the commuting patterns of their workers, or their geographical extent (Parr, 2007). They can also be identified by their local newspapers or by their local sports teams. We chose to identify cities first by a population threshold and then by their geographical extent. To ensure that the settlements we defined were indeed cities, we

chose a population threshold of 100,000, a threshold that is above the thresholds used to define what constitutes a city in all countries except China.

Identifying cities by their geographical extent follows the Roman tradition of defining a city by the edge of its built-up area, its *extrema tectorum*. That geographical extent is typically associated with a city name, the name of its largest and most prominent historical center. The built-up areas of municipalities—the governmental units associated with well-defined administrative boundaries—often merge into each other over time, as do their labor markets, as more and more people live in one municipality and commute to work in another. We define cities as agglomerations of contiguous built-up areas (and the open spaces in and around them) that may contain a large number of municipalities but, more often than not, constitute a single labor market. We consider the metropolitan region of São Paulo, Brazil, for example, to be a single city even though it contains no fewer than 39 municipalities (figure 2.1). We define São Paulo as a city by its urban edge, its *extrema tectorum*, which can be derived from freely available satellite imagery. In 2010, there were 156 free standing cities of 100,000 people or more in Brazil that had their own contiguous built-up areas made up of one or more municipalities. In contrast, there were no fewer than 5,570 municipalities in the country at that time, defined as administrative subdivisions of its national territory.

FIGURE 2.1:
The urban extent of São Paulo, Brazil (grey), showing the administrative boundaries of the 39 municipalities that constituted its metropolitan region. The urban extent of São Paulo is contained in 31 municipalities (bounded in a red line).



Number	Municipality
1	Pirapora do Bom Jesus
2	Santana de Parnaíba
3	Barueri
4	Jandira
5	Itapevi
6	Carapicuíba
7	Osasco
8	Vargem Grande Paulista
9	Cotia
10	Taboão da Serra
11	Embu
12	Itapecerica da Serra
13	São Lourenço da Serra

Number	Municipality
14	Juquitiba
15	Embu-Guaçu
16	Cajamar
17	Caieiras
18	Franco da Rocha
19	Francisco Morato
20	Mairiporã
21	São Paulo
22	Guarulhos
23	Itaquaquecetuba
24	Poá
25	Ferraz de Vasconcelos
26	Suzano

Number	Municipality
27	Arujá
28	Santa Isabel
29	Mogi das Cruzes
30	Guararema
31	Biritiba-Mirim
32	Salesópolis
33	Ribeirão Pires
34	Rio Grande da Serra
35	Mauá
36	Santo André
37	São Caetano do Sul
38	São Bernardo do Campo
39	Diadema

Using the population threshold and geographical extent definition of a city enables us to construct an entire universe of cities for the world at large. Other, possibly more precise, definitions that use information on commuting patterns or on small-area population densities cannot be used to create such a universe of cities because those data are not universally available for all cities in all countries.

Ideally, the population of a city in the universe, using our definition of population threshold and geographical extent, is the share of the population within the geographical area of the city in all the administrative (or census enumeration) zones that encompass that extent—identifiable in satellite imagery—excluding the population of villages and towns within those zones that are not part of its extent. These population estimates can, in principle, be constructed from available population data for census enumeration zones for dates roughly corresponding to 1990, 2000, and 2010. They require population data for well-defined enumeration zones, as well as rules for allocating the population of a given zone among its urban and rural built-up areas. We used this more demanding method of obtaining population estimates for the urban extent of all 200 cities in the global sample of cities described here.

For the remaining cities in the universe, we used a number of data sources that provide information on their populations, associating population with city names and coordinates without associating a specific set of enumeration zones with those names. Notably, the most useful sources on information on city populations were the United Nations Population Division (for cities of 300,000 or more) and the website www.citypopulation.de (Brinkhoff, 2016). Both sources had been consulted extensively to construct the 2010 universe of cities. That said, neither source could provide precise data on Chinese cities. According to the official definition of a city in China, the country had no more than 662 cities in 2010. We have identified a total of 1,029 settlements in China that had contiguous geographical extents of substantial area as well as populations of 100,000 or more in 2010. Their populations were estimated from data we obtained from the Chinese Academy of Sciences.

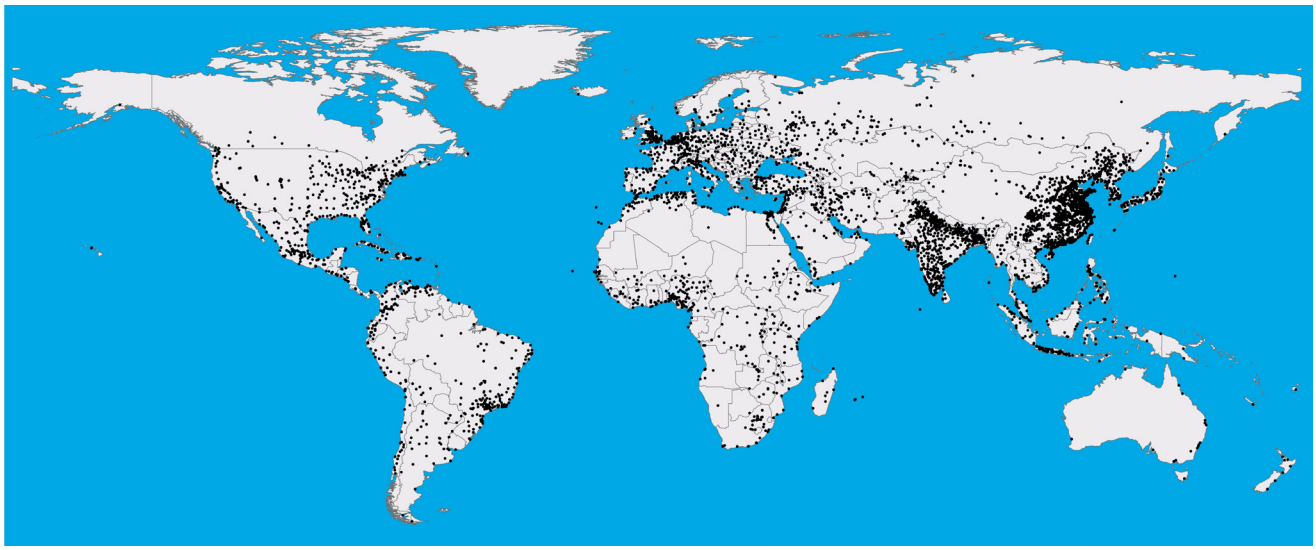
All the cities that were found to contain 100,000 or more people in 2010 were identified on Google Earth to determine whether they were part of larger urban agglomerations. Urban agglomerations were identified and listed in the universe by a single city name. Only cities that were not part of larger, named urban agglomerations were listed as cities in the universe.

The 2010 universe of cities is the third universe of cities constructed by the authors and their colleagues. The first universe of cities, described in *The Dynamics of Global Urban Expansion* (Angel, S. et al., 2005), identified a total of 3,943 cities with 100,000 or more residents in 2000. The second universe of cities, described in the *Atlas of Urban Expansion* (Angel, S., et al., 2012), identified a total of 3,646 cities that had 100,000 or more people in 2000. The 2010 universe of cities shown in figure 2.2 contains a total

of 4,231 free-standing cities in 172 countries or territories that had 100,000 or more people that year.

The universe of cities provides us with a new and powerful tool for analyzing urbanization patterns, attributes, and trends on a global scale. It makes it possible for us to assign individual values to cities in the universe—such as populations or population growth rates, for example—and then to study variations in these values among regions, income groups, or population sizes. However, the greatest and most promising value of having a universe of cities is in taking a stratified sample of cities from this universe and obtaining rigorous results from this sample and generalizing these results to the universe of cities as a whole. The global sample of 200 cities, drawn from the 2010 universe of cities, is the focus of this Atlas.

FIGURE 2.2:
The 2010 Universe of Cities, comprising a total of 4,231 cities that had 100,000 people or more in 2010.



THE GLOBAL SAMPLE OF CITIES

Beyond the names of cities, their locations, and their estimated populations at several points in time, no quantitative information pertaining to the universe of cities is available at this time. We can learn more about these cities by studying a carefully constructed sample from this universe selected with the goal of obtaining quantitative measures that can be generalized to the entire universe. For this edition of the Atlas, we selected a global sample of 200 cities (see figure 2.1). The sample was stratified so as to be more representative of this universe—namely, to ensure that cities of all sizes, from all regions, and from large and small countries were well-represented. The sample was constructed with three strata in mind:

World Regions: Cities were selected at random from eight world regions in proportion to the urban population in each region. The eight regions were:

(1) East Asia and the Pacific;

(2) Southeast Asia;

(3) South and Central Asia;

(4) Western Asia and North Africa;

(5) Sub-Saharan Africa;

(6) Latin America and the Caribbean;

(7) Europe and Japan; and

(8) Land-Rich Developed Countries.

City Population Size: An approximately equal number of cities were selected at random from four ranges of population size, each range containing one-quarter of the total population of the cities in the universe. The population ranges were:

(1) 100,000 - 427,000;

(2) 427,001 - 1,570,000;

(3) 1,570,001 - 5,715,000; and

(4) 5,715,001 and above.

Number of Cities in the Country: Cities were selected at random from three country groups identified by the number of cities in the country in proportion to the urban population in each group. The three groups were:

(1) 1 - 9 cities;

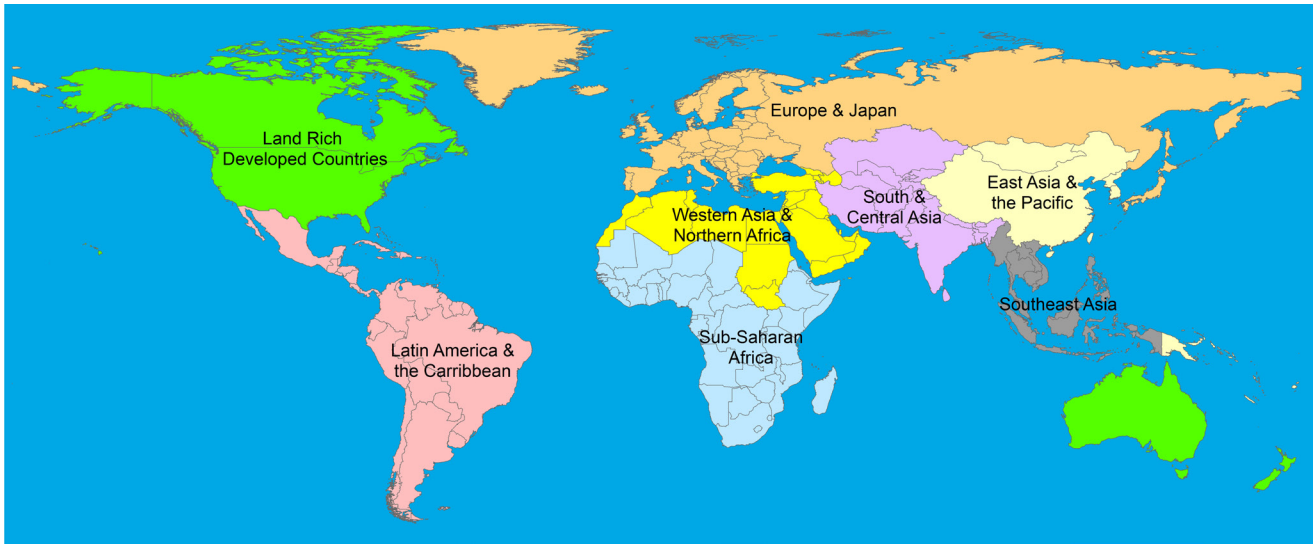
(2) 10 - 19 cities; and

(3) 20 or more cities.

The eight world regions largely followed the divisions presented in the United Nations' *World Urbanization Prospects* (U.N. Population Division, 2014), with minor changes. The United Nations divided countries into two mega-regions: more developed countries and less developed countries. The more developed countries mega-region included North America (U.S. and Canada), Australia and New Zealand, Europe, and Japan. The less developed countries mega-region included all other countries, even though some of them, (e.g. Singapore), had higher per capita income than many more developed countries. The more developed countries mega-region was divided in two to reflect different patterns of urban expansion: (1) Europe and Japan, with lower levels of arable land per person and typically higher urban densities; and (2) land-rich developed countries (U.S., Canada, Australia, and New Zealand) with higher levels of arable land per person and typically lower urban densities. The less developed countries

mega-region was divided into six regions: (1) East Asia and the Pacific, (2) Southeast Asia, (3) South and Central Asia, (4) Western Asia and North Africa, (5) Sub-Saharan Africa, and (6) Latin America and the Caribbean (see figure 2.3). To ensure that there were a minimum number of cities representing each of the eight world regions, we over-sampled cities from the smaller regions—Southeast Asia and Western Asia and North Africa—and under-sampled cities from the largest region, East Asia and the Pacific.

FIGURE 2.3:
The assignment of countries into eight world regions.



The assignment of cities in the universe of cities to four population-size categories entailed ranking the cities in the universe in increasing order of their populations and then dividing them into four ranges—small, medium, large, and very large cities—so that each of the four ranges contained approximately the same total population. The universe of cities had a total population of 2.49 billion in 2010. The four population-size ranges had approximately 622 million people in each range. This division into ranges resulted in a highly skewed distribution of the number of cities in each range: there were 3,150 small cities in the first range, 814 medium-sized cities in the second, 227 large cities in the third, and only 54 very large cities in the fourth. Each range contained approximately one-quarter of the number of cities in the preceding range, yet each range contained the same population total. Sampling at random from the universe as a whole would have resulted in three-quarters of the cities in the sample being small cities. Instead, we opted to under-sample small cities and to over-sample larger ones, drawing approximately the same number of cities from each city-size range. More specifically, we drew 56 small cities, 50 medium-sized ones, 54 large ones, and 40 very large ones from the universe. As a result of this decision, the 200 cities in the sample—while constituting only 4.7% of the total number of cities in the universe—contained 29% of the population of the universe.

Finally, the assignment of cities to one of three groups, each pertaining to the number of cities in the country, was important to ensure that countries with fewer cities were adequately represented in the sample. Indeed, less than 7% of the population of the universe of cities was found to be in countries with 1–9 cities and less than 6% in countries with 10–19 cities. Almost 88% were in countries with 20 or more cities. Cities in the first two groups would be under-represented if the sample were drawn at random from the universe as a whole. To correct this bias, we sampled cities from countries with fewer cities in slightly higher proportion than the share of their population in the universe of cities. As a result, the sample contains cities from as many as 79 countries.

Summary values for the three strata comparing the cities in the universe and the cities in the sample are given in table 2.1. The location of cities in the sample is shown in figure 2.4.

TABLE 2.1:
A comparison of the universe of cities and the sample of cities, stratified according to world regions, city population ranges, and number-of-cities-in-the-country groups.

Categories in the Three Strata		Universe of Cities				Sample of Cities				Sample/Universe Ratios	
Category ID Number	Categories	Number of Cities in this Category in Universe	Share of Cities in this Category in Universe	Population in this Category in Universe	Share of Population in this Category in Universe	Number of Cities in this Category in Sample	Share of Cities in this Category in Sample	Population in this Category in Sample	Share of Population in this Category in Sample	Ratio of Cities in this Catgory in Sample and Universe	Ratio of Population in this Catgory in Sample and Universe
World Regions											
1	East Asia and the Pacific (EAP)	1,081	26%	652,310,754	26%	42	21%	174,414,516	24%	4%	27%
2	Southeast Asia (SEA)	229	5%	143,551,770	6%	15	8%	53,516,916	7%	7%	37%
3	South and Central Asia (SCA)	693	16%	387,180,823	16%	32	16%	115,807,394	16%	5%	30%
4	Western Asia and North Africa (WANA)	301	7%	176,496,133	7%	15	8%	57,446,118	8%	5%	33%
5	Sub-Saharan Africa (SSA)	329	8%	186,702,647	8%	18	9%	51,003,826	7%	5%	27%
6	Latin America and the Caribbean (LAC)	483	11%	310,444,386	12%	26	13%	89,709,870	12%	5%	29%
7	Europe and Japan (E&J)	781	18%	389,298,026	16%	34	17%	119,848,657	16%	4%	31%
8	Land-Rich Developed Countries (LRDC)	334	8%	242,563,694	10%	18	9%	70,259,700	10%	5%	29%
	Grand Total	**4,231**	**100%**	**2,488,548,233**	**100%**	**200**	**100%**	**732,006,997**	**100%**	**5%**	**29%**
City Population Ranges											
1	100,000 - 427,000	3,143	74%	622,020,086	25%	59	30%	14,185,408	2%	2%	2%
2	427,001 - 1,570,000	811	19%	621,981,767	25%	47	24%	38,611,298	5%	6%	6%
3	1,570,001 - 5,715,000	225	5%	617,006,284	25%	54	27%	173,340,491	24%	24%	28%
4	5,715,001+	52	1%	627,540,096	25%	40	20%	505,869,800	69%	77%	81%
	Grand Total	**4,231**	**100%**	**2,488,548,233**	**100%**	**200**	**100%**	**732,006,997**	**100%**	**5%**	**29%**
Number-of-Cities-in-the-Country Groups											
1	1-9	368	9%	183,410,690	7%	24	12%	38,599,273	5%	7%	21%
2	10-19	306	7%	160,113,938	6%	17	9%	41,477,283	6%	6%	26%
3	20 +	3,557	84%	2,145,023,605	86%	159	80%	651,930,441	89%	4%	30%
	Grand Total	**4,231**	**100%**	**2,488,548,233**	**100%**	**200**	**100%**	**732,006,997**	**100%**	**5%**	**29%**

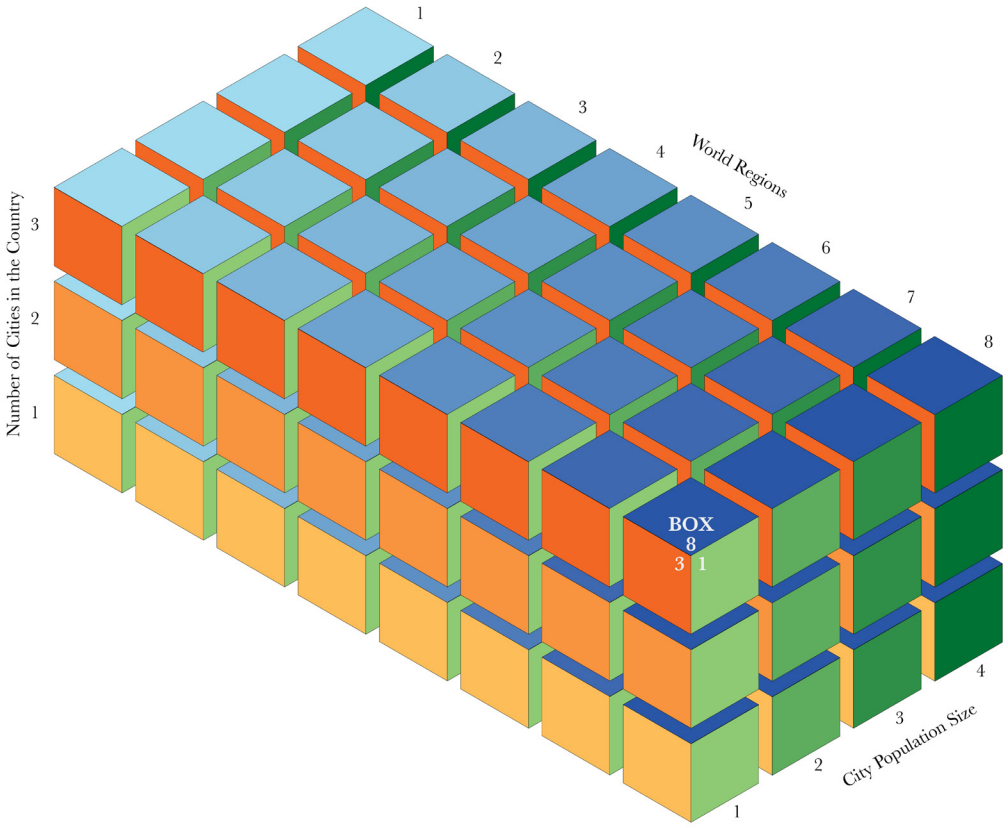
The new global sample of 200 cities is different in some respects from the sample of 120 cities used in the two earlier publications, *The Dynamics of Global Urban Expansion* (Angel et al., 2005) and *Atlas of Urban Expansion* (Angel et al., 2012). The first two strata, eight world regions, and four city population size ranges used in the earlier sample were maintained. However, the earlier sample used countries' Gross Domestic Product (GDP) per capita as a stratum. This was abandoned because of the strong correlation between the regional affiliation of cities in the sample and their countries' GDP per capita. The number

of cities in the country was introduced instead as a third stratum for the reasons explained here. Cities in the earlier sample that fit into the new sampling framework were retained in the new sample. Other cities were dropped because they were parts of larger metropolitan agglomerations, they had less than 100,000 people in 2010, or they did not represent enough similar cities in the universe. Altogether, 96 cities from the earlier sample of 120 cities are in the new sample. The earlier classifications of the satellite imagery of these cities were revisited, completed, and corrected where necessary. New metrics were derived for them as well, in line with the revised definitions of the metrics in this edition of the Atlas described in detail in the following chapter.

FIGURE 2.4:
The global sample of 200 cities.

The simplest way to envision the stratified sampling process, given the three strata chosen in this edition of the Atlas, is to envision these strata as dimensions: world regions along the x-axis, city population size ranges along the y-axis, and the number of cities in the country groups along the z-axis. Each city in the universe (or in the sample, for that matter) could then be seen as belonging to a box in three-dimensional space, identified by a three-digit number, its world region (1–8), its city population size range (1–4), and its number of cities in the country group (1–3) (see figure 2.5). Halifax, Canada, for example, belongs to box 813. It is located in Region 8 (land-rich developed countries); it had 390,000 people in 2010, assigning it to city population size range 1; and Canada had 34 cities in the 2010 universe of cities, assigning Halifax to number of cities in the country group 3 (20+ cities in the country). Box 813 contains all 210 cities in the universe that were located in land-rich developed countries, that had less than 427,000 people in 2010, and that were in countries with 20 cities or more.

FIGURE 2.5:
The sampling framework comprising 96 boxes, each box corresponding to one of eight world regions, one of four city-population-size ranges, and one of three number-of-cities-in-the-country groups (8 x 4 x 3 = 96).



Of the 96 boxes (8 × 4 × 3 = 96) shown in figure 2.5, only 76 had cities in the universe of cities. The rest were empty. Of these, 61 boxes had cities in the sample. The remaining 15 boxes that are not represented by cities in the sample contain 114 cities in the universe with a total population of 63.2 million, comprising 3% of the cities and 3% the population of the universe in 2010. These cities were assigned to "nearby" boxes, boxes in the same region with cities with similar population size and similar number of cities in the country assignments, to be represented by the sample as well. In this manner, all the cities in the universe were represented by cities in the sample.

The process of selecting cities in this framework consisted of picking cities at random from each box in rough proportion to the total population in each box. For example, four cities were selected at random to represent box 813: Victoria, British Columbia in Canada, and Gainesville FL, Killeen, TX, and Modesto, CA, in the United States. As there were 210 cities in the universe in this box, one city in the sample represented some 50 cities in the box 813. In parallel, as there were 44.9 million people in the cities in the universe in this box and 1.1 million people in the four sample cities in the box, every urban dweller in the cities in the sample in box 813 represented 40 urban dwellers in the universe of

cities in this box.

The values 50 and 40 in this example can be thought of as city-based and population-based weights respectively. They can be used to obtain weighted averages for the universe from values obtained for the sample. If a city in a given box represents 50 cities, then any value associated with it—say, its population growth rate between 2000 and 2010—is given a city-based weight of 50, while another city in the sample representing, say, only 27 cities is given a city-based weight of 27. Similarly, if the population of a sampled city in a given box represents a population 40 times as large, then each resident in this city is given a population-based weight of 40.

The population growth rates for these cities were not used as a stratum in the creation of the sample. The universe of cities contains data on the population of each city for three time periods, 1990, 2000, and 2010. We could use this information to test whether the sample was representative of the universe. Indeed, when we compared the average population growth rates between 2000 and 2010 in all the cities in the universe with both the city-based and population-based weighted averages of the cities in the sample, we found that they were not different from each other at the 95% confidence level. This assured us that the global sample of cities was indeed representative of the universe of cities.

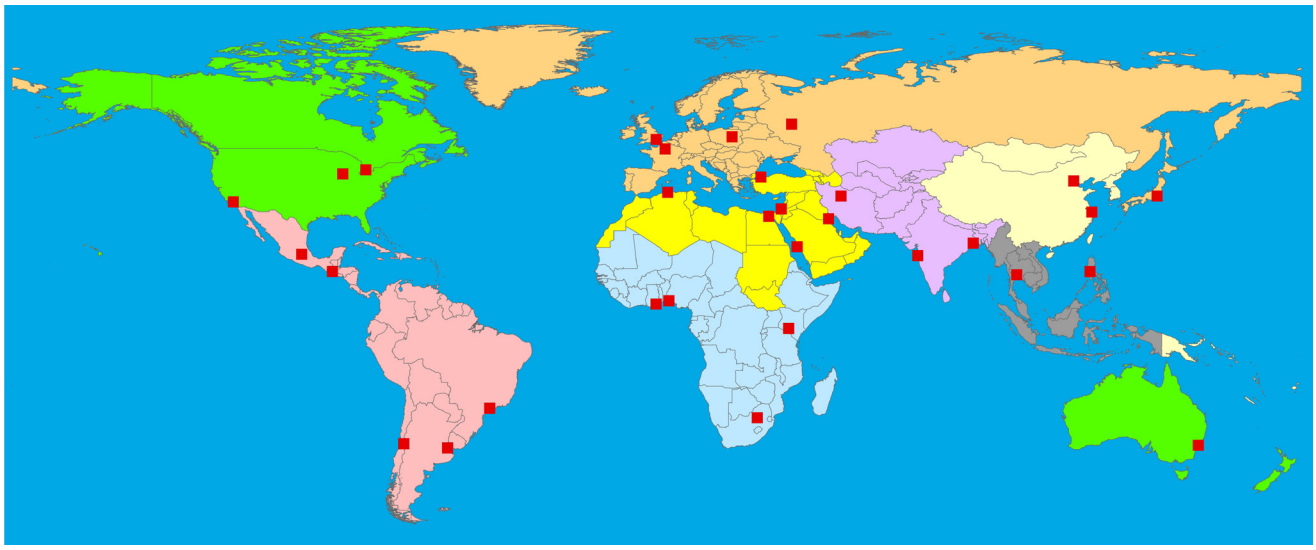
Using these city-based and population-based weights, we can now answer new questions about the universe of cities as a whole. For example, we determined that average densities in the universe declined significantly between 1990 and 2000, and continued to decline—albeit at a significantly lower rate—between 2000 and 2014. We also determined, for example, that the average share of area that was laid out before it was occupied in the expansion zones of cities in the universe—areas converted to urban use between 1990 and 2014—was significantly lower than it was in areas developed before 1990. In other words, the global sample of cities makes it possible, for the first time, to monitor global urban expansion in a consistent and rigorous manner. Needless to say, it can also be used to monitor other urban attributes of interest, from housing affordability to air quality, from Internet use to access to public open spaces, and from the quality of drinking water to the availability of public transport in the sample of cities to obtain valid, rigorous, comparative data—data that was never available before—on the universe of cities as a whole.

THE REPRESENTATIVE GROUP OF 30 CITIES

A representative group of 30 cities, including 27 from the global sample of 200 cities, was created to explore long-term changes in urban expansion, urban population density, and the attributes of urban

layouts from circa 1800 until circa 2014. The selection of cities for this historical analysis was guided by two factors: their regional distribution and the availability of historic maps depicting their built-up areas at 20- to 25-year intervals. Three cities—Jeddah, Saudi Arabia, Nairobi, Kenya, and Kuwait City, Kuwait—were added to the 27 representative cities from the global sample to ensure a balanced sub-regional distribution of cities (figure 2.6).

FIGURE 2.6:
The location of the 30 cities in the representative group of cities where urban expansion was mapped and animated between 1800 and 2014.



To be included in this representative group, the relevant maps of a given city needed to depict the totality of the urban extent of the city for time periods some 20–25 years apart and have sufficiently clear landmarks to be georeferenced to Google Earth imagery. This geo-referencing process aligned the maps to a common coordinate system, thereby allowing them to be accurately compared to each other. A complete list of the map references containing the original maps used to construct the composite maps for each city is available in the earlier *Atlas of Urban Expansion* (Angel, S. et al., 2012).

The maps are digitized composite maps of the urban extent of a given city on different dates. A total of 261 maps were used to create the composite maps for the 30 cities in this representative sample, an average of 8.7 maps per city approximately 19±1 years apart. The composite maps for each city with their associated populations, densities, and changes over time appear in the 2012 edition of the *Atlas of Urban Expansion*. They were subsequently animated to show the long-term expansion of these cities. These animations can be seen on the Atlas website at www.atlasofurbanexpansion.org.

These maps were also used in Volume 2 of the Atlas to study the changes in the attributes of urban extents over time. We divided the urban extents of the 30 cities in this representative sample to areas

that were built-up in five time periods: (1) Before ~1990, (2) between ~1900 and ~1930, (3) between ~1930 and ~1960, (3) between ~1960 and ~1990, and (5) between ~1990 and 2014. We then studied the attributes of the urban fabric and calculated the metrics associated with them in each one of these areas for each city. We used these metrics to calculate average values of each attribute—say, the share of the built-up area in streets or the average block size—in each one of the five time periods, so as to observe their changes over a century or more.