LINCOLN INSTITUTE
OF LAND POLICY

The 2010 Universe of Cities: A New Perspective on Global Urbanization

Working Paper WP18NG1

Nicolás Galarza Sánchez
Marron Institute of Urban Management

Yang Liu
Marron Institute of Urban Management

Shlomo Angel
Marron Institute of Urban Management

Alejandro M. Blei, Achilles Kallergis, Patrick Lamson-Hall
Marron Institute of Urban Management

Maria Mónica Salazar
Marron Institute of Urban Management

Daniel L. Civco
University of Connecticut Center for Land Use Education and Research

Jason Parent
Connecticut Institute for Resilience and Climate Adaptation

September 2018

The findings and conclusions of this Working Paper reflect the views of the author(s) and have not been subject to a detailed review by the staff of the Lincoln Institute of Land Policy. Contact the Lincoln Institute with questions or requests for permission to reprint this paper. help@lincolninst.edu

© 2018 Lincoln Institute of Land Policy

Abstract

Using a number of data sources, scholars at New York University and the University of Connecticut, in partnership with UN-Habitat and the Lincoln Institute of Land Policy, and in collaboration with the UN Population Division, have identified the universe of all 4,231 cities and metropolitan areas that had 100,000 people or more in 2010. All these cities have identifiable, freestanding, contiguous urban extents, and all of these cities have names, coordinates, and population figures associated with them from three time periods: 1990, 2000, and 2010. This paper introduces this universe of cities and presents a number of new findings: (1) the population sizes of the universe of cities did not follow the regularities of Zipf's Law or Gibrat's Law completely; (2) the numbers of large and small cities grew proportionately to the total number of cities in the universes of cities; and (3) the variations of city population size among regions were distinguishing but dwindled over time. We present our findings from two perspectives: the trends and distributions of city population sizes, and of city population growth rates. This new universe of cities lays the foundation for a new perspective on global urbanization that focuses on cities as units of analysis, rather than on countries.

About the Authors

Nicolás Galarza Sánchez, Research Scholar
NYU Urban Expansion Program
Marron Institute of Urban Management
New York University
60, 5th Avenue
ngalarza@stern.nyu.edu

Yang Liu, Research Scholar
NYU Urban Expansion Program
Marron Institute of Urban Management
New York University

Shlomo Angel, Professor of City Planning
Director, NYU Urban Expansion Program
The Marron Institute of Urban Management and the Stern School of Business
New York University

Alejandro M. Blei, **Achilles Kallergis**, **Patrick Lamson-Hall**, Research Scholars
NYU Urban Expansion Program
Marron Institute of Urban Management
New York University

Maria Mónica Salazar, Research Assistant
NYU Urban Expansion Program
Marron Institute of Urban Management
New York University

Daniel L. Civco
Department of Natural Resources and the Environment
Director, Center for Land Use Education and Research
University of Connecticut

Jason Parent
Assistant Research Professor
Connecticut Institute for Resilience and Climate Adaptation (CIRCA)
Department of Natural Resources and the Environment
University of Connecticut

Table of Contents

The 2010 Universe of Cities: A New Perspective on Global Urbanization

1. The Context

The study of global urbanization from the perspective of countries yields important, but limited, findings and insights.

The study of global urbanization—its past trends, present patterns, and future prospects—has largely relied on country-level data, rather than on city-level data. It is safe to say that until recently we did not have a comprehensive, consistent, and comparable global data set on cities. The main source for country-level data on global urbanization is the United Nations (UN) Population Division which publishes its authoritative *World Urbanization Prospects* every two years. The most recent publication is its 2014 Revision.

World Urbanization Prospects (Hijmans 2015; *World Urbanization Prospects: The 2014 Revision* 2014) reports on three key metrics associated with countries and regions: Their urban population; the share of their urban population in the total population; and the growth rate of their urban population at 5-year intervals, from 1950 to 2050. It also provides population estimation for cities with 300,000 people or more.

These data (*World Urbanization Prospects* 2014) reveal three important trends in global urbanization. To simplify the presentation of these trends, we focus on two world mega-regions identified by the UN Population Division: 'More Developed Countries' and 'Less Developed Countries'.

The 'More Developed Countries' mega-region includes North America (U.S. and Canada), Australia and New Zealand, Europe and Japan; often regarded to as the Global North. The 'Less Developed Countries' mega-region includes all other countries; often regarded to as the Global South (see Figure 1).

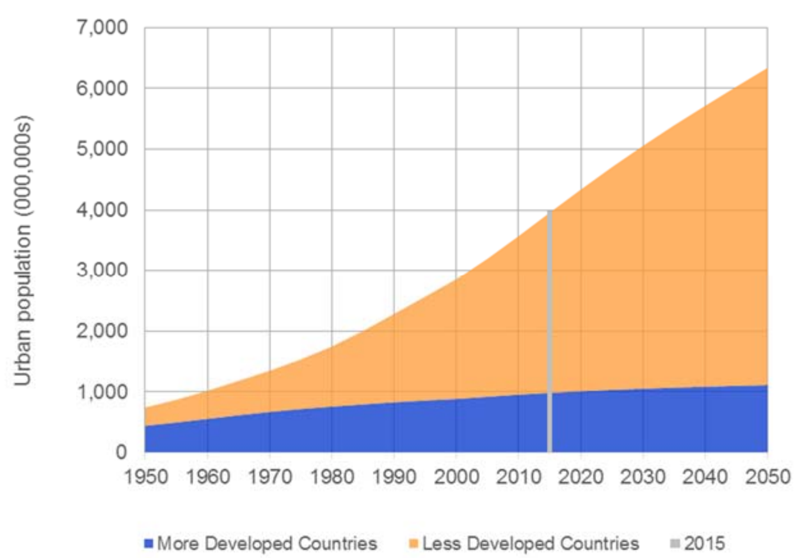
Figure 1. division of the world by the UN into two mega-regions, 'More Developed Countries' or the Global North, and 'Less Developed Countries' or the Global South.

1.1. The First Trend

Between 2015 and 2050 the urban population in Less Developed Countries is expected to increase by 2.35 billion, or 18 times the expected increase of 130 million in More Developed Countries (see figure 2).

In other words, Urbanization is now largely restricted to the Global South; 33 percent of the expected growth in the world urban population between 2015 and 2050 will be in Sub-Saharan Africa and 25percent in the Indian Subcontinent. The rest will be in China (11 percent), Southeast Asia (9 percent), Latin America and the Caribbean (9 percent) and the Middle East and North Africa (9 percent).

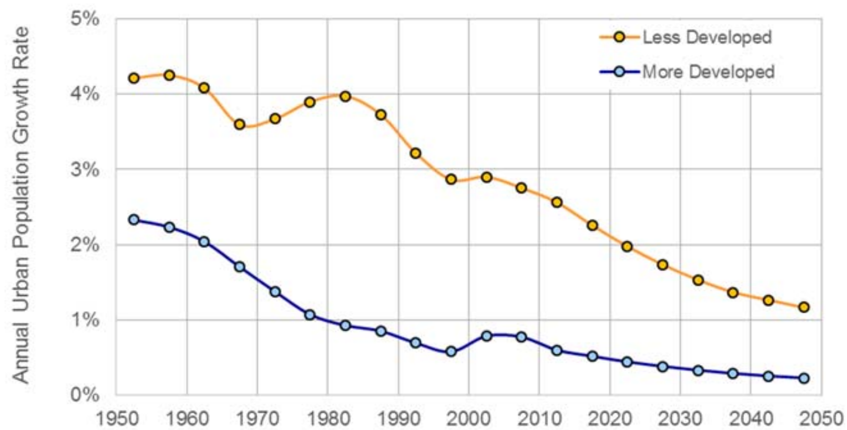
Figure 2. Between 2015 and 2050 the urban population in Less Developed Countries will increase by 2.35 billion, or 18 times the expected increase of 130 million in More Developed Countries.



1.2. The Second Trend

Between 2015 and 2050 urban population growth rates are expected to decline from 0.6 percent to 0.2 percent in more developed countries and from 2.6 percent to 1.2 percent in Less Developed Countries (see figure 3).

Figure 3. Between 2015 and 2050 urban population growth rates are expected to decline from 0.6 percent to 0.2 percent in More Developed Countries and from 2.6 percent to 1.2 percent in Less Developed Countries.



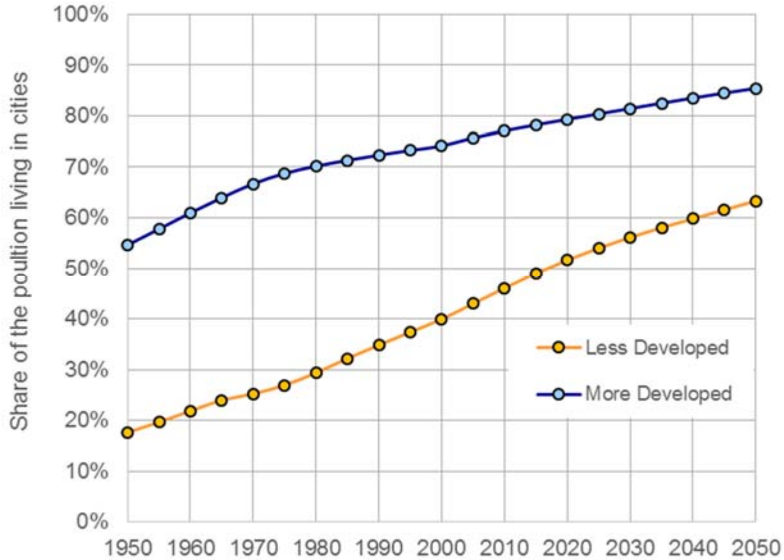
In other words, the population of cities in the Global North is now growing at very slow annual average rates, and already showing negative rates of urban population growth in Eastern Europe and Japan.

More importantly, the rate of urban population growth in the Global South has been in significant decline since the mid-1980s; by 2050, it is expected to be similar to that of the Global North in the mid-1970s.

1.3. The Third Trend

Between and 2015 and 2050 the share of the population living in cities expected to increase from 78 percent to 85 percent in more developed countries and from 49 percent to 63 percent in less developed countries.

Figure 4. Between and 2015 and 2050 the share of the population living in cities expected to increase from 78 percent to 85 percent in More Developed Countries and from 49 percent to 63 percent in Less Developed Countries.



The movement of people to cities to be closer to each other rather than be closer to the land, what can be termed humanity's *Urbanization Project*, started in earnest in 1800 and can now be expected to be largely completed by 2100.

By 2100, our cities will have been built, their streets and arterial roads constructed, their buildings erected. From then on, it will be largely a matter of caring for them, recycling them, and rebuilding them. Rapid urban population growth will be a thing of the past.

We have a window of opportunity during the remaining decades of this century to intervene in the global urbanization process, so that it proceeds in an orderly fashion and so that it renders cities more productive, more inclusive, more sustainable, and more resilient.

We can see that studying global urbanization trends from the perspective of countries yields important insights as well as important policy prescriptions. Yet these are limited. They are limited because they lump all the cities in every country.

We can advance our knowledge and understanding of global urbanization trends if we move towards studying cities, rather than countries, as units of analysis. The Universe of Cities, in our case the *2010 Universe of Cities* discussed in the next section, aims to fulfill this role.

In this paper, we want to explore two sets of simple questions that cannot be explored at the country level, one set of questions focusing on the population size of cities and one set of questions focusing on the population growth rate of cities.

The questions pertaining to the distribution and median of city population size are:

- Do they vary over time?

- Do they vary among different world regions?
- Do they vary among cities at different levels of economic development?

The questions pertaining to the median population growth rate of cities are:

- Does it vary over time?
- Does it vary among different world regions?
- Does it vary among cities at different levels of economic development?
- Does it vary among cities with different population sizes?

The third and fourth sections of the paper provide rigorous statistical evidence pertaining to our answers to these questions.

Beyond exploring these two sets of questions, the universe of cities makes it possible to study a number of additional questions by focusing on a representative sample of cities in the universe, rather than working with the universe of cities as a whole.

In separate papers now being drafted, we introduce numerous findings obtained from the study of the Global Sample of Cities, a stratified sample of 200 cities selected to represent the 2010 Universe of Cities.

The cities in the UN Sample of Cities contained 728 million people in 2010, some 29 percent of the population of the universe of cities in that year, and some 20 percent of the total world urban population in that year.

The UN Sample of Cities has now been used to estimate the worldwide expansion of urban areas between 1990 and 2015, as well as to estimate changes in urban land consumption per capita during this period in different world regions. It has also been used to study changes in the physical layouts of urban footprints in areas built before and after 1990.

Indeed, the UN Sample of Cities now provides researchers the world over with a generic platform for studying cities and drawing more and more scientifically valid inferences from these studies on the state of cities the world over. This was never possible before, greatly hampering serious progress towards a science of cities.

Three additional research programs of the tri-partite collaboration between UN Habitat, New York University, and the Lincoln Institute of Land Policy are already making use of the UN Sample of Cities to draw inferences about the 2010 Universe of Cities as a whole.

2. The 2010 Universe of Cities

Many different authors have sought to define what constitutes a city spatially. Although there is not yet a unified official definition that is universally accepted, it can be observed that there is

some consensus, not in arriving at a single definition, but regarding two broad issues. The first is that a city can be defined by observing its physical area or geographic extent, with an emphasis on the contiguity of its built-up area. The second focuses on how cities function as integrated economic units. J.B. Parr (Parr 2007) and J. Borja (Borja 2003) suggest a number of definitions that provide a complete basis for understanding different attempts to arrive at a common definition. That said, scholars apply different rules within the categories mentioned above, with the result we end up with multiple definitions that are all equally valid.
Not only do scholars apply different rules to define a city spatially, but countries and international organizations use their own independent rules to subdivide the territories they govern as well. Administrative boundaries are used to define municipalities, cities, or urban agglomerations. The European Union, for example, promotes the use of the Urban Functional Region (FUR), to define cities but there is no uniquely correct definition for what a FUR is (Cheshire & Gornostaeva 2002).

Identifying the universe of all cities in a given year requires, of course, a unified 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, by their municipal boundaries, by their labor markets, or by their geographical extent; and they can also be identified by their local newspapers or by their local sports teams.

We have chosen to identify cities by a population threshold and then by their geographical extent. We have chosen 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.

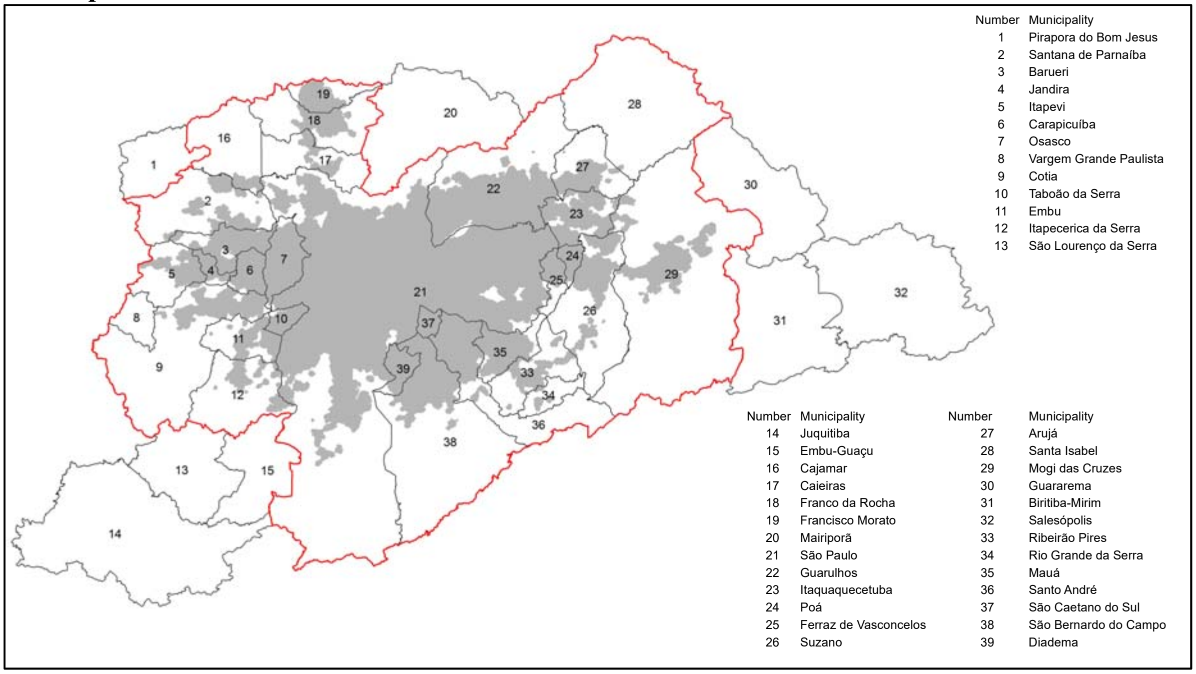
Defining cities by their geographical extent follows the Roman tradition of defining a city by the edge of its built-up area, its *extrema tectorum*. That geographical extent is typically associated with a city name associated with its historical center.

The built-up areas of municipalities often merge into each other, so 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 that may contain a large number of municipalities, but often constitute a single labor market.

For example, we consider the metropolitan region of São Paulo, Brazil to be a single city, while it contains no less than 39 municipalities[1]. We define São Paulo as a city by its urban edge, its *extrema tectorum,* which can be observed in freely available satellite imagery.

[1] Taken from the Metropolitan Planning Department from Sao Paulo. https://www.emplasa.sp.gov.br/RMSP

Figure 5. The urban extent of São Paulo, Brazil, showing the administrative boundaries of the municipalities that constitute it.



Altogether, in the 2010 Universe of Cities, there were 156 cities of 100,000 people or more in Brazil that were contiguous built-up areas, made up of one or more municipalities. In contrast, there were a total of 5,570 municipalities[2] in the country, defined as administrative subdivisions, at that time.

Using the geographical extent definition of a city makes it possible to construct an entire universe of cities for the world at large. Other definitions that use information on commuting patterns or on small-area population densities cannot be used to create a universe of cities because those data are not universally available for all cities.

Ideally, the population of a city, using our geographical extent definition is the share of the population of all the administrative districts that encompass it, excluding the population of villages and towns within those districts that are not part of its contiguous built-up area.

These population estimates can be constructed from available population data for administrative districts for dates roughly corresponding to 1990, 2000, and 2010, but they need to be estimated carefully and this requires considerable time and effort. We used this method of obtaining population estimates for the built-up areas of all 200 cities in the global sample of cities.

In the remaining cities in the universe of cities, we used a number of data sources that provide information on the populations of cities, associating population with city names and coordinates without associating a specific administrative boundary with those names.

---

[2] Taken from the Brasilian Institute of Geography and Statistics Website. https://www.ibge.gov.br/geociencias-novoportal/organizacao-do-territorio/estrutura-territorial/15761-areas-dos-municipios.html?&t=sobre

Notably, the most useful sources of information on city populations are the United Nations Population Division (for cities of 300,000 or more) and the website City Population ("City Population," 2016). Both sources have been consulted extensively to construct the 2010 universe of cities.

However, both sources could not provide precise data on Chinese cities. According to the official definition of a city, China has only 662 cities. We have identified a total of 1,029 cities in China that had substantial geographical extents and populations of 100,000 or more in 2010. Their populations were estimated from data we obtained from the Chinese Academy of Sciences.

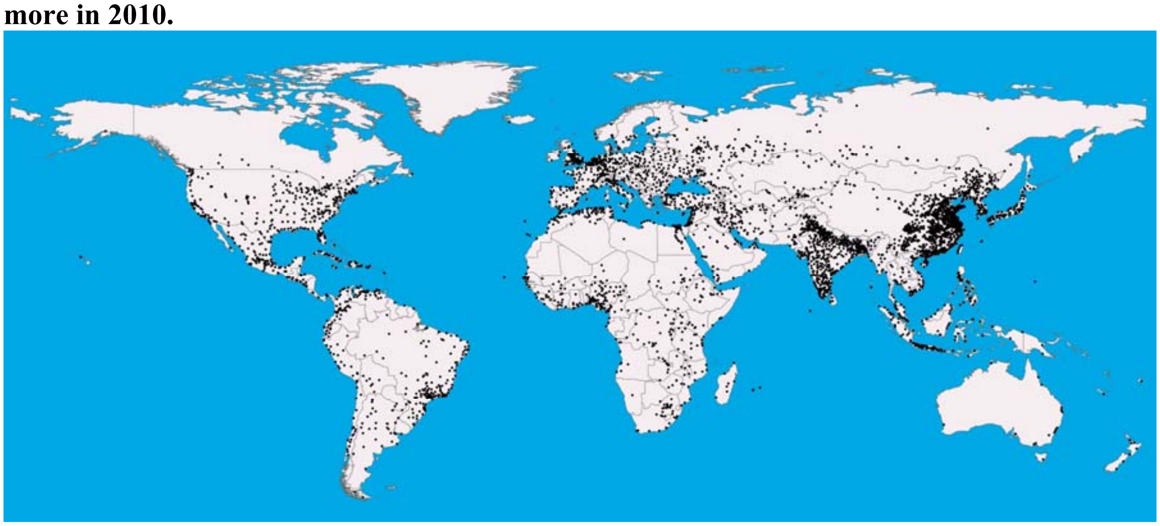
All the cities that were listed by these sources to contain 100,000 people or more in 2010 were identified on Google Earth to determine whether they were or were not part of larger urban agglomerations.

Urban agglomerations were identified and listed in the universe by a single city name, and 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 2000 Universe of Cities, described in Angel et al's *The Dynamics of Global Urban Expansion* (Angel et al 2005) identified a total of 3,943 cities that had 100,000 or more in 2000. The second 2000 Universe of Cities, described in Angel et al's *Atlas of Urban Expansion* (Shlomo Angel, 2012) identified a total of 3,646 cities that had 100,000 people or more in 2000.

The 2010 Universe of cities is shown in figure 6 below. It contains a total of 4,231 self-standing cities that had 100,000 people or more in that year in 172 Countries or territories.

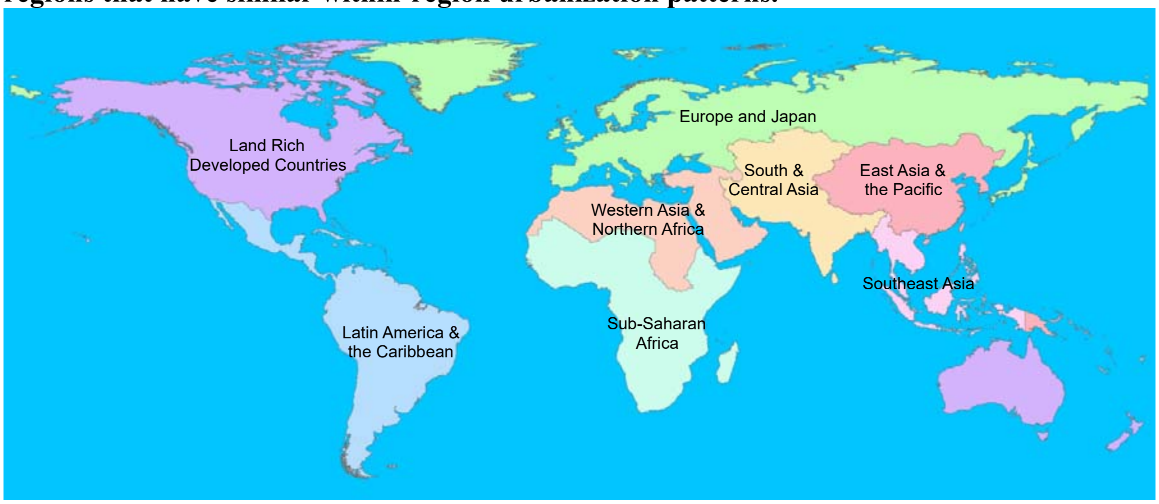
Figure 6. The 2010 Universe of Cities, comprising a total of 4,231 cities that had 100,000 people or more in 2010.



For purposes of analysis, we further divided the two mega-regions shown in figure 1 into eight world regions, shown in figure 7 below. The number of cities and their populations in each of the regions in 2010 is shown in table 1 below.

The More Developed Countries mega-region was divided in two, largely with urban population densities in mind: Land-Rich Developed Countries (The U.S., Canada, Australia and New Zealand) and Europe and Japan. Cities in the former region have considerably lower population densities than cities in the latter.

Figure 7. The division of the world into eight world regions, all composed of UN regions and sub-regions that have similar within-region urbanization patterns.



The Less Developed Countries mega-region was divided into six regions: East Asia and the Pacific, Southeast Asia, South and Central Asia, Western Asia and North Africa, Sub-Saharan Africa, and Latin America and the Caribbean, following the UN classification.

Table 1. The universe of cities in 2010 divided into eight world regions.

Region	Number of Cities in Region	Percent of Number of Cities	Total Population in Region (Millions)	Percent of Total Population
East Asia and the Pacific (EAP)	1,082	26%	639	26%
Europe and Japan (E&J)	780	18%	396	16%
Land-Rich Developed Countries (LRDC)	334	8%	241	10%
Latin America and the Caribbean (LAC)	483	11%	320	13%
South and Central Asia (SCA)	693	16%	393	16%
Southeast Asia (SEA)	229	5%	128	5%
Sub-Saharan Africa (SSA)	329	8%	186	7%
Western Asia and North Africa (WANA)	301	7%	181	7%
Total (World)	4,231	100%	2,483	100%

Cities in the universe were also divided into four city population size categories and four levels of economic development categories.

City population data for the universe of cities were available from various sources and interpolated for three periods: 1990, 2000, and 2010. In each time period, cities were ranked by their population size and divided into four ranges by their population sizes, so that the total population in each group was roughly equal. The division of the 2010 universe of cities into four population size ranges is shown in table 2 below.

Table 2. The universe of cities in 2010 divided into four city population size ranges.

City Population Size Range	Number of Cities in Range	Share of Cities in Range	Total Population in Range (Millions)	Share of Range's Population in Universe
100,000 - 427,800	3,140	74%	621	25%
427,801 - 1,570,000	814	19%	625	25%
1,570,001 - 5,750,000	225	5%	618	25%
> 5,750,000	52	1%	619	25%
Total (World)	4,231	100%	2,483	100%

City GDP for every city in the universe was obtained by projecting information for 2,910 cities contained in the McKinsey Global Institute's (MGI) Cityscope database, a proprietary product, onto our universe of 4,231 cities. The data used to create city GDP estimates for the universe of cities included GDP and population data from city-level observations in the MGI dataset, population data from the universe, and national average GDP per capita from the World Bank (*International Comparison Program database* 2016).

We could identify matches between 2,617 cities in the MGI list and the universe of cities. Linear regressions were constructed within each country using city population and GDP data from MGI. Then we obtained GDP estimates for 966 (23 percent) cities using these models with city population size as predictor. In an additional 648 (15 percent) cities for which regression was deemed inappropriate to estimate city GDP, national average GDP per capita calculated from MGI data was used with population data from the universe to estimate city GDP. For 20 cities (0.5 percent) whose countries do not appear in the MGI table, national GDP per capita data from World Bank were used together with population size from the universe to estimate city GDP.

Then city GDP per capita was calculated from city GDP for every city. Cities were ranked by their GDP per capita and divided into four GDP per capita ranges. The division of the 2010 universe of cities into four GDP per Capita ranges is shown in table 2 below.

Table 3. The universe of cities in 2010 divided into four GDP per Capita ranges.

City GDP per Capita Range	Number of Cities in Range	Share of Cities in Range	Total Population in Range (Millions)	Share of Range's Population in Universe
$427 to $7,776	1,293	31%	619	25%
$7,776 to $15,690	1,337	32%	662	27%
$15,690 to $27,522	772	18%	574	23%
$27,522 to $108,398	829	20%	628	25%
Total (World)	4,231	100%	2,483	100%

Using the divisions of the universe of cities into 8 world regions, four city population size ranges, and four GDP per capita ranges, we addressed the two sets of questions presented earlier, one pertaining to variations in the average population size of cities and the other pertaining to variations in the average population growth rates of cities. We examine these questions more carefully through our analysis in the following section.

3. Population Size of the Universe of Cities: Trend and Discovery

It has been well established that city sizes follow two empirical regularities, which are internally connected. The first regularity is that large cities follow the rank-size rule (Zipf's Law) (Gabaix 1999). When cities are ranked according to their population sizes in descending order, their population sizes, if above certain baseline, are inversely proportional to their ranks. J. Eeckhout (2004) examined the US data on 25,359 cities, towns and villages ranging in population from a single person to over 8 million people, and showed that the power-law, or Pareto distribution only fit for cities over certain lower bound, while the lognormal distribution would fit the entire population.

The second observed regularity is that the city growth is a proportionate growth process (Gibrat's Law) (Eeckhout 2004). Proportionate growth was coined by biologists to describe an interesting feature in animals. Different organs of the body grow at approximately the same rate (Dhar, 2013). In our context, although the growth rate varies considerably among cities, it does not depend on city population size. Large cities do not necessarily grower faster or slower than small cities. This has been observed in many country studies including the United States and China (Eaton & Eckstein 1997; Wang & Zhu 2013). Gabaix (1999) showed that if cities grow randomly according to Gibrat's Law, the limit distribution of city size converges to obey Zipf' Law. But so far, there have been no empirical studies on a global level rather than on country level, reporting on all the cities in the world.

3.1. Large Cities in the Universe do not Comply with the Rank-Size Rule

We first explore the regularity of the rank-size rule (Zipf's Law), which predicts that the city population sizes plotted against their ranks on the log-log scale would fall on a straight line. This is because any distribution that follows the rank-size rule is a power-law function. Let *X* be the random variable that represents city size. The probability density function (PDF) and cumulative density function (CDF) are:

$$f_X(x) = Cx^{-a}, \quad (1)$$

$$P(X < x) = 1 - \frac{C}{a-1}x^{-(a-1)}, \quad (2)$$

where

C = constant,

a = power-law exponent.

And let N be the total number of cities with population sizes above 100,000. If we rank cities from the largest (rank 1) to smallest (rank N), the rank for a city of size x is:

$$Rank(X = x) = \frac{NC}{a - 1} x^{-(a-1)}, \quad (3)$$

These functions are straight lines if plotted on a log-log scale. For example, the Zipf CDF, which is used to show the rank-size rule, is the log-log form of formula (3):

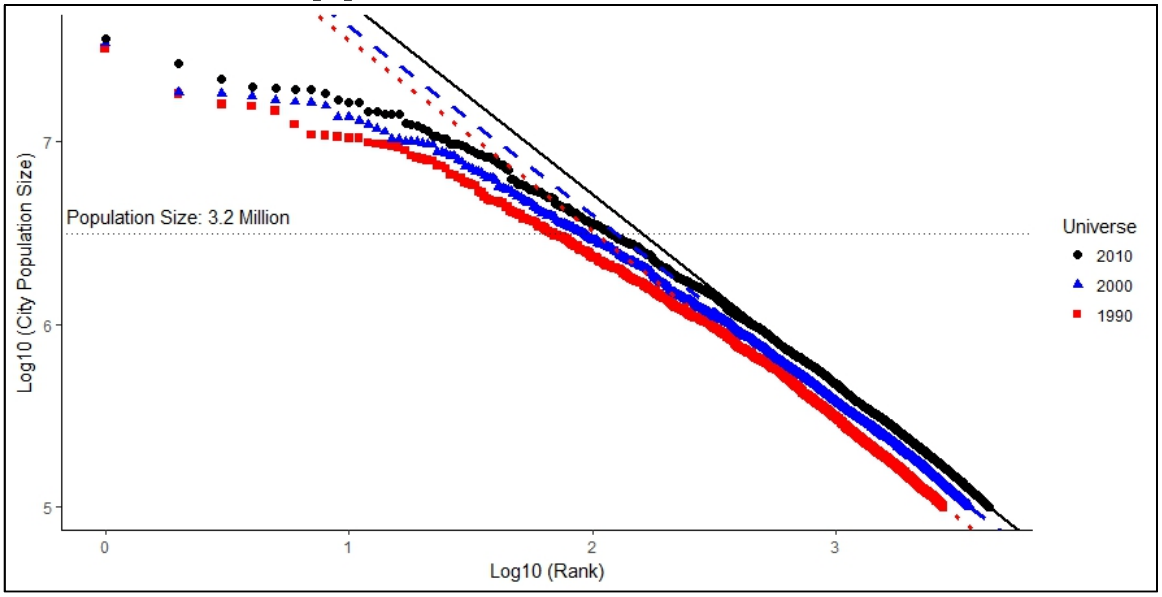
$$\log(X) = -\frac{1}{a - 1}\log(Rank(X)) + \frac{1}{a - 1}\left(\log\frac{C}{a - 1} + \log N\right), \quad (4)$$

in which $\frac{1}{a-1}$ is the Zipf exponent. The slope of the line is $-\frac{1}{a-1}$.

Data points for the three universes of cities, corresponding to 1990, 2000, and 2010, are plotted in Figure 8 with city rank on the x-axis and population size on the y-axis on a log-log scale. The three lines represent 1990, 2000 and 2010 from bottom to top. They move upwards with roughly equal distance everywhere on the line, indicating the similarity of the distributions. The three straight lines are the fitted Zipf CDFs, which would fit perfectly if the data followed a power-law distribution. The power-law exponents were obtained by maximum likelihood estimation, and then the slopes and intercepts for Zipf CDFs were calculated. The slopes of the lines for 1990, 2000, and 2010 are -1.048, -1.035 and -1.054 respectively, which are close to each other, and are similar to what was reported in the literature. (Dobkins & Ioannides 1999; Rosen & Resnick 1980).

We observe in Figure 8 that the fitted lines do not fit the universe if the population size is above around 3.2 million ($10^{6.5}$), which is illustrated by a horizontal dotted line. It is worth noticing that the power-law distribution still fits most of the cities since only 117 cities (3 percent) in the 2010 universe have population sizes larger than 3.2 million. As the intercept of the black fitted line for 2010 is 8.82, the first rank city would have to have a population of an unthinkable 665 million ($10^{8.82}$), the second city a population of 310 million, and so on, for the population size distribution to follow Zipf's Law. At present, the largest city in the 2010 universe is Tokyo, and it had a population of 35 million. In short, the power-law function that can fit most cities cannot fit the whole universe of cities because the largest cities are not large enough.

Figure 8. Zipf CDF which plots city population sizes against their ranks on the log-log scale for the 3 universes of cities with population sizes over 100,000 in 1990, 2000, and 2010.



Notes: for every universe of cities, city population size is plotted against the rank of the city. The largest city is plotted with rank 1. From outside to inside are the universes of cities corresponding to 2010, 2000, and 1990. The black solid line is the Zipf CDF fitted for the universe 2010, the blue dash line is for the universe 2000, the red dotted line is for the universe 1990. The horizontal dotted line shows where the population size equals 1E6.5, 97 percent of the cities are below this line.

The universe of cities still manifests the general trend of a power-law distribution: it is highly skewed to the right and fat-tailed. As a result, the median is a more appropriate statistic to analyze the data. We will adopt non-parametric tests to compare the medians in the following sections. For all the tests mentioned in this paper, if the non-parametric test rejects the null hypothesis, so will the test of means. If the distributions are very close to each other, the means would not be significantly different either.

3.1. The Distribution of Cities Above 100,000 has Remained the Same

Most cities grow larger over time. But if we only look at the universe of cities, how would the distribution evolve? In 1990, only 2,758 cities had population sizes over 100,000, which is 65 percent of the 4,231 cities in 2010. The number of cities grew by 28 percent to 3,535 in 2000, and by another 20 percent to 4,231 in 2010. Examining city population sizes from these three periods, we have an interesting finding that the distribution of cities having population sizes above 100,000 has remained unchanged. If we group them by population size range in Table 4, the percentages of cities in each range are almost the same. For example, in 1990, there were only 33 cities with population size over 5.75 million, their share in the 2,758 cities was 1.2 percent. In 2010, we had more large-sized cities. But because the total number of cities increased to 4,231, the share of the 52 large-sized cities in 2010 was still only 1.2 percent. Applying the Kruskal-Wallis rank test, the non-parametric test for distributions, we find that there is not enough evidence to reject the null hypothesis that the distributions are the same ($p>0.05$). In other words, the percentages of cities in each population size range are close enough that we cannot say they are significantly different, no matter how the ranges are divided. The largest difference of the three medians are within 2 percent, the medians are not significantly different from each other either.

Table 4. The number and percentage of cities with population sizes over 100,000 in each population group in 1990, 2000, and 2010. Although the total count of cities increased year by year, the row percentage in each range remained roughly the same.

Year	Count (Row Percent) in Population Size Range				Total Count (%)	Median Population
	100,000 - 427,800	427,801 - 1,570,000	1,570,001 - 5,750,000	Over 5,750,000		
Year 1990	2,033 (74%)	555 (20%)	137 (5.0%)	33 (1.2%)	2,758 (100%)	216,160
Year 2000	2,647 (75%)	682 (19%)	166 (4.7%)	40 (1.1%)	3,535 (100%)	219,656
Year 2010	3,140 (74%)	814 (19%)	225 (5.3%)	52 (1.2%)	4,231 (100%)	220,090

This may sound counter-intuitive as it seems the numbers of cities in each range were scaling up proportionately. A second thought reveals that this observation echoes the second regularity that city population size growth rate does not depend on the population size. It illustrates the proportionate growth of city population size from another perspective.

3.2. The Difference in City Size Between More Developed and Less Developed Countries Has Disappeared

The more developed countries mega region includes the North America region (United States and Canada), Australia, New Zealand, and Europe and Japan as shown in **Error! Reference source not found.**. The remaining regions form the group of less developed countries. If we divide the cities in these two groups by population size ranges in 2010 (Table 5), the shares are almost the same within each range. Wilcoxon rank test cannot reject the null hypothesis and claim a difference between more developed countries and less developed countries. The more sensitive Kolmogorov-Smirnov tests for two samples cannot reject the null hypothesis either ($p>0.05$). The 1,114 cities in more developed countries have a median population only 3.6 percent higher than the 3,117 cities in less developed countries (226,096 versus 218,183), which is not statistically significant.

Table 5. Counts and proportions of cities in 2010 in more and less developed regions by four city population size range. Each row adds up to 100 percent.

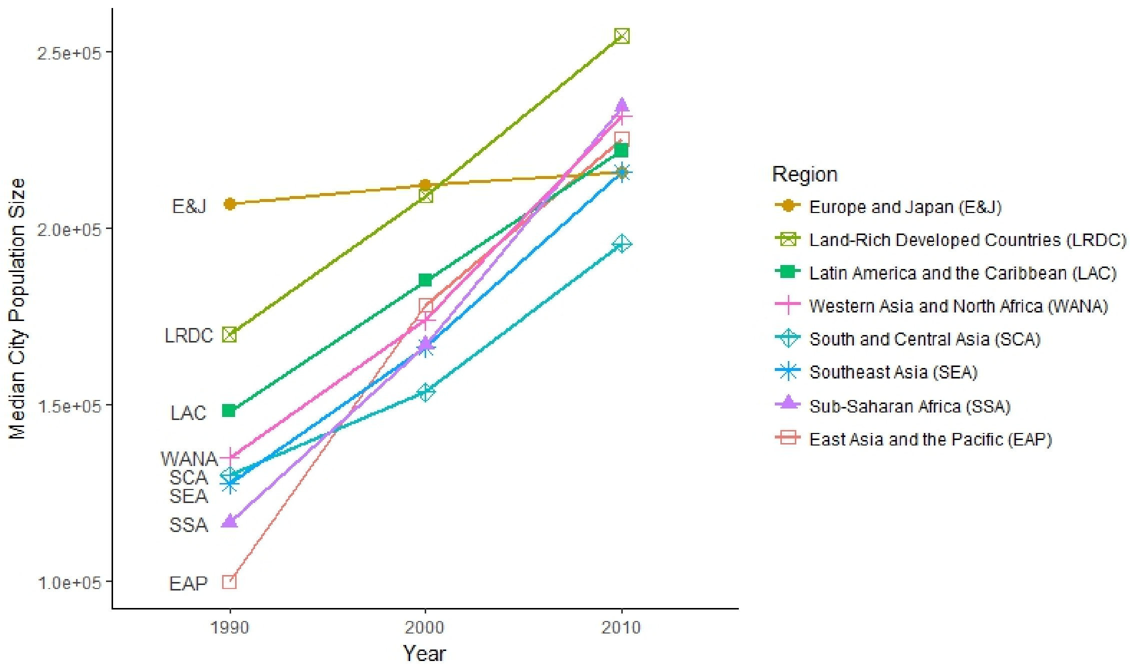
2010	Count (Row Percent) in Population Size Range				Total Count (%)
	100,000 - 427,800	427,801 - 1,570,000	1,570,001 - 5,750,000	Over 5,750,000	
Less Developed Countries	2,301 (74%)	605 (19%)	169 (5.4%)	42 (1.4%)	3,117 (100%)
More Developed Countries	839 (75%)	209 (19%)	56 (5.0%)	10 (0.9%)	1,114 (100%)

Such similarity between more developed and less developed counties is not observed for these 4,231 cities if trace back to 2000 or 1990. In 2000, the median of cities in more developed countries was 25 percent higher the median of cities in less developed countries. In 1990 this difference was 62 percent. This surprising convergence of more developed countries and less developed ones leads us to further investigate how the city populations changed within each region.

3.3. The Relative Ranks of Regions in Terms of Median City Population Size Changed Over Time, but the Differences Among Them Became Smaller.

Among the 4,231 cities in the 2010 universe, 690 of them had populations less than 100,000 in 2000 and 1,323 of them had populations less than 100,00 in 1990. We show how the regional medians of the 2010 universe of cities developed over the twenty years between 1990 and 2010 in **Error! Reference source not found.**9 and Table 6. There are two interesting observations. First, in all regions except Europe and Japan (E&J) the median city populations increased over time and the slopes of the median population trend line varied across regions. As a result, the ranks of each region in terms of median city size changed over time. Second, the medians converged over time. In 2010, they were considerably more clustered than in 1990.

Figure 9. Change of median city population sizes from 1990 to 2010, the legend is sorted by median population from high to low in 1990.



Regions have been ranked from high to low by their median city population values in 1990 in Table 6. It is interesting to note how the ranks changed for the 8 regions as time passed. In 1990, Cities in Europe and Japan (E&J) had the largest median population size, exceeding 206,000 people, but as that median population size remained roughly the same, the region ranked penultimate in 2010. Sub-Saharan Africa (SSA) was penultimate in 1990 but ranked second in 2010, indicating a strong growth. East Asia and the Pacific (EAP) had the lowest median in 1990 because many cities just started to grow, in 2010 it ranked in the middle. Compared 2010 to 1990, E&J, SSA, and EAP merged into the majority, SCA dropped down to have the lowest median.

Table 6. Ranking of city population size medians by year and region

Median by Region	1990	Ranking	2000	Ranking	2010	Ranking
Europe and Japan (E&J)	206,759	1	212,272	1	215,717	7
Land-Rich Developed Countries (LRDC)	169,875	2	209,034	2	254,413	1
Latin America and the Caribbean (LAC)	147,984	3	184,974	3	221,817	5
Western Asia and North Africa (WANA)	134,999	4	173,804	5	231,701	3
South and Central Asia (SCA)	129,872	5	153,487	8	195,596	8
Southeast Asia (SEA)	127,640	6	166,248	7	215,779	6
Sub-Saharan Africa (SSA)	116,430	7	166,651	6	234,225	2
East Asia and the Pacific (EAP)	99,817	8	178,056	4	225,142	4
Total (World)	216,160		219,656		220,090	

Notes: Table is sorted by medians in descending order in 1990. The top-ranked E&J in 1990 only ranked at number 7 in 2010. Based on 4231 observations in 2010, 4225 observations in 2000, and 4081 observations in 1990. Note that in 2000, 690 cities (16 percent) had less than 100,000 people. In 1990, 1,323 cities (32 percent) that had less than 100,000 people.

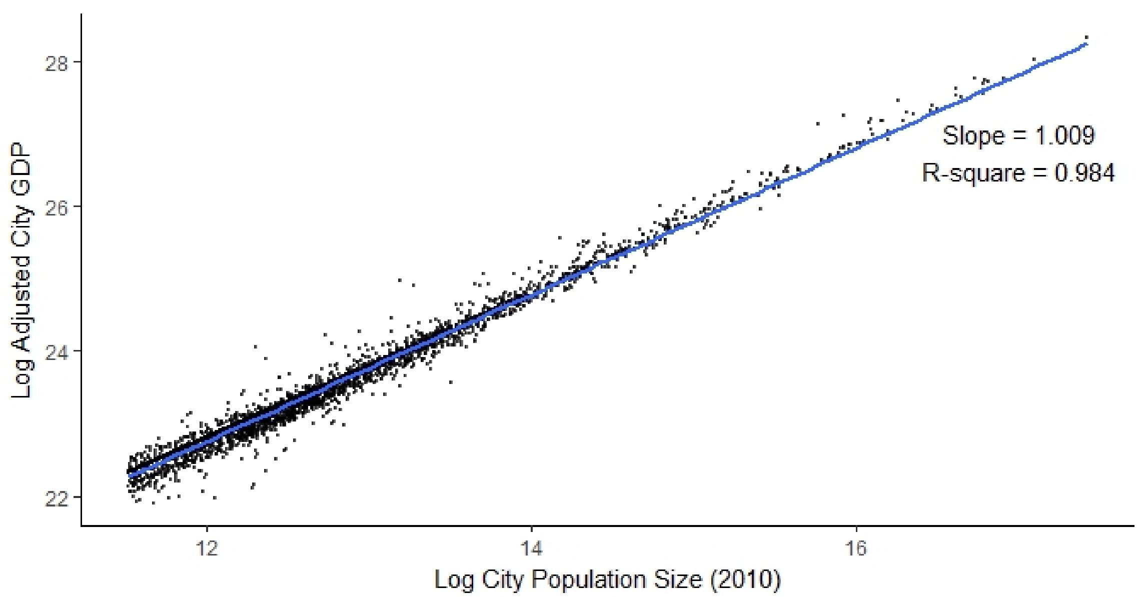
We apply a multiple comparison post hoc test for the Kruskal-Wallis analysis to show that there were indeed more significant differences in pairwise comparisons of regions during earlier periods. In 2010 the median population size of LRDC was significantly higher only than the median city size of SCA ($p<0.001$). The median city size of LRDC was not significantly higher than the median of any of the 6 regions clustered in the middle. Among the 6 regions excluding LRDC and SCA, there was no significant difference between their medians. In 2000, the medians of E&J and LRDC were significantly higher than every region except LAC. After removing E&J, LRDC, and SCA, the remaining 5 regions have medians that are not significantly different from each other. In 1990, the median city size of E&J was significantly higher than the medians of all other regions. LRDC was higher than all other regions except WANA. EAP was lower than all the other regions. After removing E&J, LRDC, SSA, and EAP from the top and bottom, only the remaining 4 regions had medians that are not significantly different from each other ($p = 0.051$). We can conclude that the differences among regions, especially between the more developed regions (E&J, LRDC) and others in terms of median city population size were much larger in the earlier years.

3.4. GDP Adjusted for National Average Correlates Best with Population Size

It is well established that city GDP correlates positively with city population size within a country. (Lobo, 2013). When we examine the whole universe of cities, the trend is strong after adjusting for national average GDP as shown in Figure 10. Pooled national average GDP per capita is subtracted from the city GDP per capita to adjust for the national average. Then the city GDP per capita difference from national average multiplied by the city population sizes yields the adjusted city GDP. The linear correlation on the log scale fits very well with an R-square of 0.98. In contrast, the R-square from a regression on log scale in the universe of cities using the unadjusted city GDP and population size is only 0.47.

The slope of the regression is 1.009, which is larger than 1, but smaller than the power law scaling exponents measured from a single country. Bettencourt et al. (2007) found that Germany had the lowest scaling exponent of 1.13 in 2003, while the European Union had a scaling exponent of 1.26 during 1999 to 2003 period. A scaling exponent higher than 1 indicates the effect of economies of scales. Other things being equal, larger cities are more productive than smaller ones. For the 2010 universe of cities as a whole, a city that doubles the population of another one can be expected to have a GDP per capita adjusted for national average that is more than doubled (2.013 times).

Figure 10. GDP adjusted for national average associates with population size on a log-log scale.



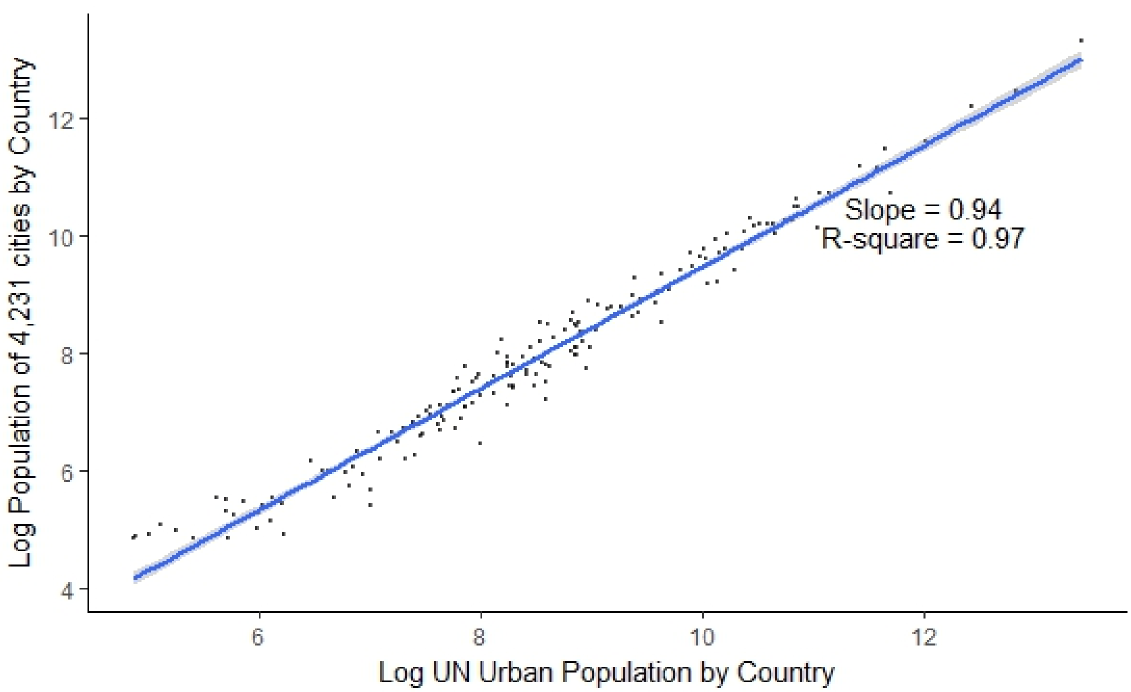
3.5. The Difference in Median City Population Between More Urbanized Countries and Less Urbanized Countries is Becoming Smaller

There are 173 countries in the 2010 universe of cities, among which 33 have more than 75 percent of their population living in cities according to UN. We define these 33 counties to be more urbanized countries, whereas the remainder of countries are less urbanized. We tested the median of the population size of the 685 cities in more urbanized countries against the median of the 3,546 cities in less urbanized countries. In 2010, the median of the city population size in more urbanized countries was 7 percent higher than in less-urbanized countries, which is not a statistically significant difference. In 2000, the difference was 13 percent and in 1990, the difference was 34 percent. In terms of the distributions of city population and the median city population, differences have become smaller over tiem between more urbanized countries and less urbanized countries.

3.6. The Universe of Cities Data Compared with UN World Urbanization Prospects Data

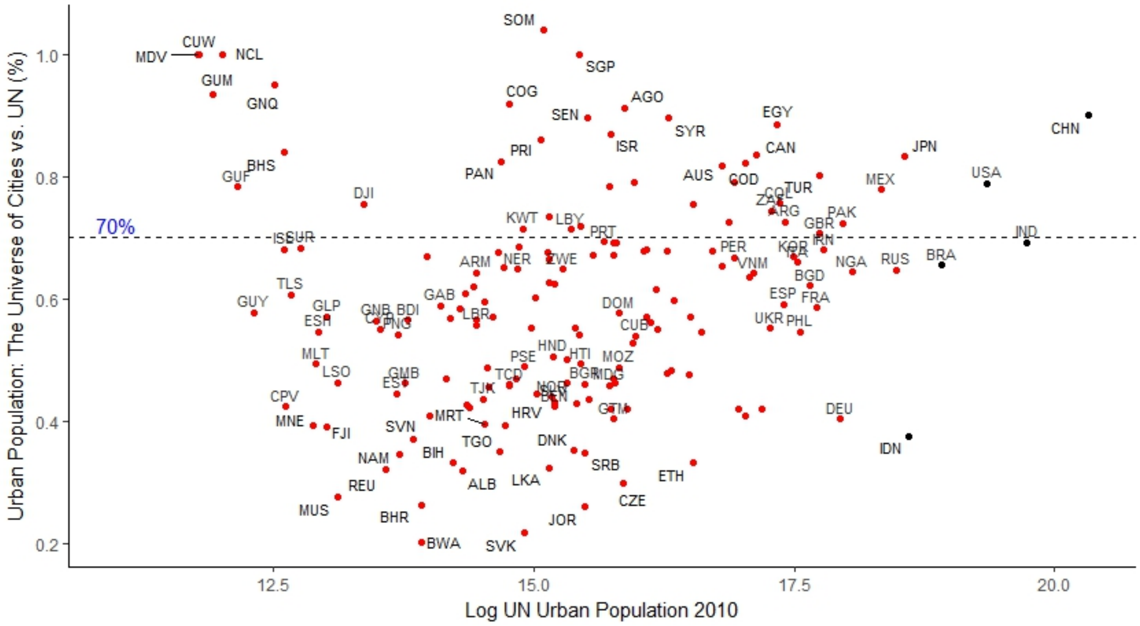
The World Urbanization Prospects report published by the Population Division of UN provides urban population at the country level. We cross-checked their data with ours. The total population in cities having sizes above 100,000 in 2010 was 2.48 billion, while the total UN urban population was 3.54 billion. The share of the total population in the universe of cities compared to UN estimate was about the same during the three periods: 70 percent for 2010 and 71 percent for both 2000 and 1990. A linear regression on the log-log scale between the population in the universe of cities by country and UN urban population by country has a slope of 0.94 and R-square of 0.97 (Figure 11). A more urbanized country in the universe data is also a more urbanized one according to UN. And the universe of cities tends to contain fewer population than the UN estimate for each country.

Figure 11. Linear regression on the log-log scale between the population in the universe of cities by country and UN urban population by country. The three countries with largest urban populations according to UN are China (669 million), India (373 million), and United States (252 million).



However, when we look into each country and compare the ratio of the total population in the universe of cities with the urban population estimated by UN, the variations are distinct (Figure 12). From the right of the plot shows the countries with the largest urban population. For example, China (CHN) had 669 million urban population in 2010 according to UN, among which 600 million people are living in cities with population sizes over 100,000, so the proportion of Chinese population living in the universe of cities compared to UN urban population is 90 percent. There is no pattern between this proportion and the UN urban population, indicating that different standards were adopted at the national level to define the urban population. The observation suggests the potential benefit in aligning the definition of the urban population for each country.

Figure 12. The share of the population in the universe of 4,231 cities versus the UN urban population (y-axis), plotted against the UN urban population (x-axis); Shown by country.



Notes: The urban population by country estimated by UN is plotted on x-axis on log-scale, the urban population in cities with size above 100,000 is smaller than the estimates of UN for almost all the cities. The shares for each country is plotted on the y-axis. Some countries clustered in the middle of the plot are not labeled randomly to avoid overlap. The 5 countries (China, India, United States, Brazil, Indonesia) with the largest urban populations according to UN are plotted in black near the right end of the figure.

4. The Annual Population Growth Rate of Cities: Understanding Population Growth from Another Perspective

We can obtain a better understanding of the trends discussed in the previous sections when we examine the growth rate of city populations. The world regions have become more similar in terms of population size distributions because city population growth rates were faster for less developed regions. Sometimes the population growth rate of cities shows a weak but significant association between with cities' populations.

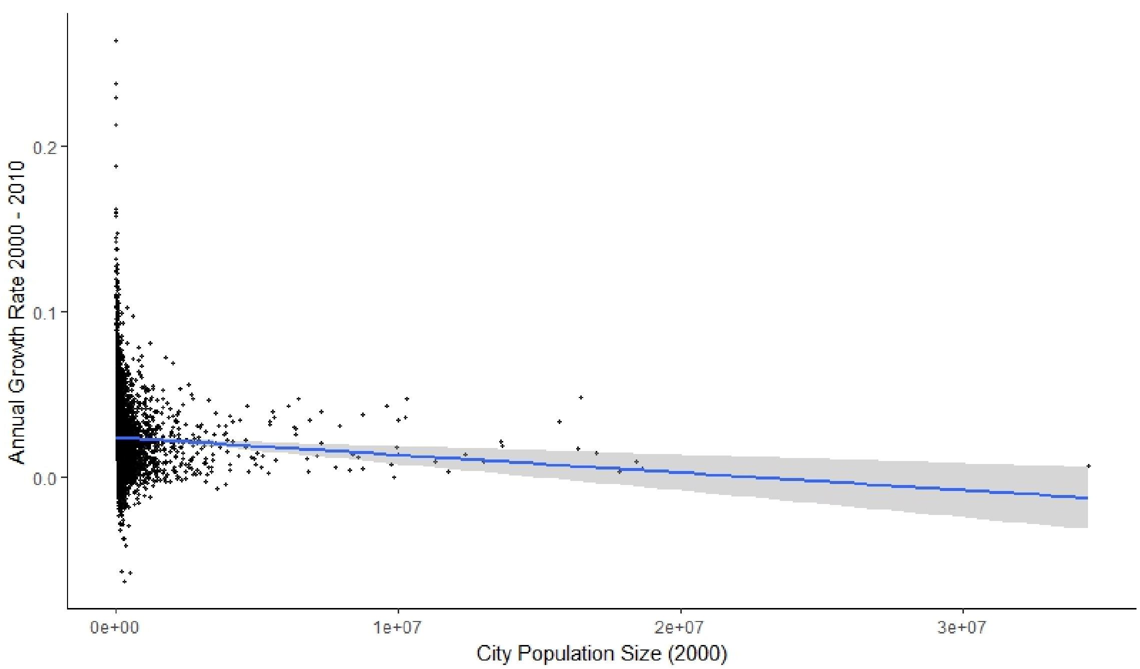
The annual population growth rate, r, for a given city is calculated from three variables: The population at the beginning of the period, P_0, the population at the end of the period, P_t, and the number of years between the periods, t, assuming that the population in the city is growing exponentially according to the formula $P_t = P_0 \times e^{rt}$.

4.1. Population Growth Rates Depend on Population Sjze in Some Regions

The second regularity of proportionate growth claims that city population growth rates do not depend on city population size. In figure 13 we plot t annual city-level growth rates over the 2000 - 2010 period against the city population size in 2000 on a log-scale. The slope is slightly negative (-0.006) and the R-square is only 0.07. If the population size in 2010 is used as the independent variable, the slope is not significantly different than zero. For the annual city-population growth rate over the 1990 – 2000 period, the slope of the regression line is -0.015 if plotted against the city population sizes in 1990, and is -0.006 if plotted against the city population size in 2000. Comparing growth rate to the baseline generates more negative slopes.

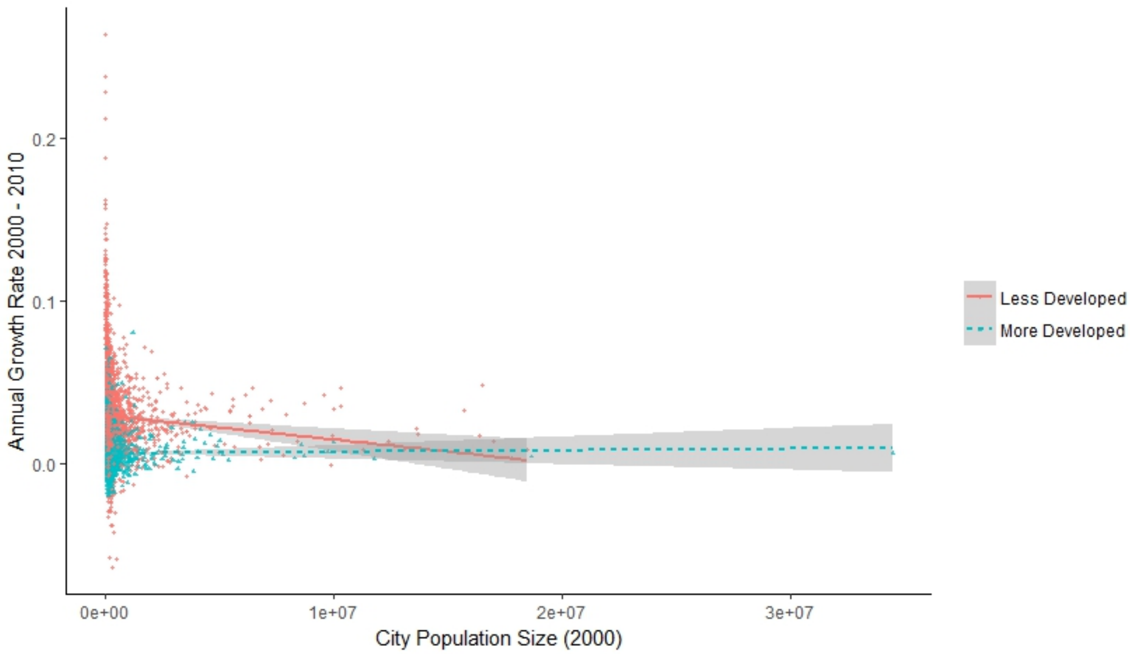
Overall, the proportionate growth conjecture, which posits that the population growth rate doesn't depend on city population size, is only approximately true in the whole universe of cities since cities that were smaller in the beginning of the period did grow a little bit faster. What is clear, however, is that in the universe of cities as a whole large cities are not growing any faster than smaller ones.

Figure 13. Average growth rates of the universe of cities almost do not correlate with city size. The slope is slightly different from zero.



The slopes are very different for cities in more developed counties and cities in less developed countries as shown in Figure 14. In period 2000 to 2010, the slope estimate from a linear regression is indeed not different from zero for cities in more developed countries. But for less developed countries, the slope is -0.007. Proportionate growth is not universally the same but depends on the region.

Figure 14. Average growth rates of the universe of cities by developed and less developed countries.



Among the eight regions, only SCA and E&J have slope estimates that are not significantly different from zero when the growth rate in 2000 to 2010 was plotted against population size in 2000 on the log scale. The slope is -0.002 for LRDC, -0.003 for SEA and LAC, -0.006 for SSA, and -0.007 for WANA. For the 1,080 cities in EAP (1,021 of which are in China), the slope is -0.013 with R-square of 0.19. Cities in China show the most distinguishing trend against the proportionate growth. Smaller cities in China were growing considerably faster than larger ones.

4.2 The Population Growth Rate was Fastest in Sub-Saharan Africa

When we examined the median city population growth rates of regions, the differences among regions were getting smaller. This observation indicates that population size growth rates must vary across regions. During period 2000 to 2010, the median growth rate was 2.4 percent for less developed countries and 0.5 percent for more developed ones. Moreover, growth rates vary significantly across regions. We notice from Table 7 that Sub-Saharan Africa (SSA) had the fastest growth rate for both periods, followed by EAP, while Europe and Japan (E&J) had the slowest growth rate. The differences in median growth rates among regions are statistically significant for every region. The ranking did not change much from the first period to the second except that EAP and SSA, as well as SCA and SEA, exchanged their positions, indicating a consistent trend over time.

Table 7. Median regional growth rates by region during period 1990 to 2000 and period 2000 to 2010. The table is sorted by medians during 2000 to 2010 period in descending order.

Median of Growth Rate by Region	1990 to 2000	Ranking	2000 to 2010	Ranking
Sub-Saharan Africa (SSA)	3.6%	2	3.5%	1
East Asia and the Pacific (EAP)	4.7%	1	3.0%	2
Western Asia and North Africa (WANA)	2.9%	3	2.4%	3
Southeast Asia (SEA)	2.5%	5	2.4%	4
South and Central Asia (SCA)	2.5%	4	2.1%	5
Latin America and the Caribbean (LAC)	2.4%	6	1.8%	6
Land-Rich Developed Countries (LRDC)	1.8%	7	1.4%	7
Europe and Japan (E&J)	0.2%	8	0.2%	8
Total (Global)	**2.4%**		**1.9%**	

5. Conclusion

The well-established rank-size rule describing city population size has its limit. The power-law distribution cannot fit the whole distribution if not truncated at the lower end, as shown by Jan Eeckhout (2004). It cannot fit very large cities in the universe of cities, where cities in countries are pooled together, as shown by our study. It is also true that the power-law exponent is sensitive to the definition of the city and to sample size, as observed by Rosen and Resnick (1980). The second regularity of proportionate growth also depends on the region under study. During the 2000 to 2010 period, it only strictly applied to the two of the eight regions and certainly not to the whole universe of cities. Although the overall dependence of the city growth rate on city population size is extremely weak, the association exist significantly in less developed countries.

Moreover, contrary to some prevalent arguments that there are more large cities and that the small cities are diminishing in population, our findings based on the population data of all the cities in the world since 1990 indicate that such announcements are at best, incomplete. More large cities do not necessarily imply fewer small cities. When there are more large cities, there are also more small cities proportionally. And as the total number of cities with size above 100,000 increases, the share of small or large cities in the world remains the same.

On the other hand, the difference in city population sizes among regions are shrinking. In 2010 only LRDC had a median population size significantly larger than the other regions, and only SCA had a median population size significantly smaller than the rest. The difference between the median city population size of more developed countries and less developed countries were much larger in 1990 and 2000 and eventually disappeared in 2010. There is also no difference in median city population sizes between more urbanized countries and less urbanized ones in 2010.

Beyond these new findings, the promising value of having a universe of cities is in taking a stratified sample of cities from this universe, obtaining rigorous results for the sample of cities and then, using the weights to make inferences on more general results for the universe of cities as a whole.

The authors have now engaged in a number of studies of a global sample of 200 cities, nearly 5 percent of the 4,231 cities in the 2010 Universe of cities. The underlying idea of this sample, now adopted by the United Nations as the UN Sample of Cities, is to study different attributes of cities and to have these attributes form layers of data for the cities in the sample.

We now have layers of data for three time periods: 1990, 2000, and 2014 for the urban extent of cities, for their urban extent densities as well as for their built-up area densities, for the levels of saturation of urban extents with built-up areas, for the shares of new development in infill, extension, leapfrog and inclusion, and for the shape compactness of urban extents.

We also have layers of data for the quality of urban layouts in the global sample of cities in areas built before 1990 and in areas built between 1990 and 2015: Access to arterial roads, the share of the land in streets, the share of areas that are laid out before they are occupied, the average block size, and the density of 4-way street intersections.

In addition, we have layers of data for the affordability of residential land and housing in 2015 in four housing submarkets: informal housing, public housing, formal multi-family private housing and formal single-family private housing.

We have data on the regulatory regime governing land development on the periphery of cities in the global sample of cities, focusing on the levels of stringency of land use regulations as well as on the levels of stringency of containment policies seeking to limit urban expansion. We also have data on particulate matter pollution in 2000 and 2010 in the global sample of cities.

Future studies of the sample of cities may include: (1) measuring the five attributes that make up urban density: Crowding, building height, plot coverage, the residential land use share, and the saturation of urban extent by built-up areas; (2) measuring the quality of drinking water; (3) measuring access to public open space; and (4) mapping the coverage, frequency, and speed of formal and informal public transit.

Studying the different layers of data in the global sample of cities allows us to investigate connections between different attributes of cities. For example, we can investigate whether stringent containment policies influence housing affordability or whether urban densities affect particulate matter pollution.

It is hoped that the shift in focus from urbanization levels in countries to collecting and analyzing data in a universe of cities and to collecting and analyzing data in a global sample of cities representing this universe may advance the science of cities, leading to better and more evidence-based urban plans, projects, programs, and policies.

References

Angel, S., Chabaeva, A., Gitlin, L., Kraley, A., Parent, J., & Perlin, M. 2005. The dynamics of global urban expansion. *1*(35563), 1-205.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., & West, G. B. 2007. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences, 104*(17), 7301-7306. doi:10.1073/pnas.0610172104

Borja, J. (Ed.) 2003. *La ciudad conquistada*: La Ciudad Aventura de Libertad.

José Lobo, L. M. A., Bettencourt, Deborah Strumsky, Geoffrey B. West. 2013. Urban Scaling and the Production Function for Cities. doi:10.1371/journal.pone.0058407

Cheshire, P., & Gornostaeva, G. 2002. Cities and regions: comparable measures require comparable territories. doi:Cheshire, Paul and Gornostaeva, Galina (2002) Cities and regions: comparable measures require comparable territories. Les Cahiers de L'institut d'amenagement et d'urbanisme de la Region d'ile de France, 135. pp. 13-32. ISSN 0153-6184

City Population. 2016. Retrieved from http://www.gadm.org/country

Dhar, D. S., Tridib. 2013. A sandpile model for proportionate growth. *Journal of Statistical Mechanics: Theory and Experiment*(11), article id. 11006.

Dobkins, L. H., & Ioannides, Y. 1999. *Dynamic Evolution of the U.S. City Size Distribution*. Retrieved from https://EconPapers.repec.org/RePEc:tuf:tuftec:9916

Eaton, J., & Eckstein, Z. 1997. Cities and growth: Theory and evidence from France and Japan. *Regional Science and Urban Economics, 27*(4), 443-474. doi:https://doi.org/10.1016/S0166-0462(97)80005-1

Eeckhout, J. 2004. Gibrat's Law for (All) Cities. *American Economic Review, 94*(5), 1429-1451.

Gabaix, X. 1999. Zipf's Law for Cities: An Explanation. *The Quarterly Journal of Economics, 114*(3), 739-767.

Hijmans, R. 2015. *GADM database of Global Administrative Areas*. Retrieved from: http://www.gadm.org/country

International Comparison Program database. (2016). Retrieved from: https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD

Parr, J. B. 2007. Spatial Definitions of the City: Four Perspectives. *Urban Studies, 44*(2), 381-392.

Rosen, K. T., & Resnick, M. 1980. The size distribution of cities: An examination of the Pareto law and primacy. *Journal of Urban Economics, 8*(2), 165-186.

Shlomo Angel, J. P. D. L. C. A. M. B. 2012. *Atlas of Urban Expansion* (Vol. Chapter 1): Lincoln Institute of Land Policy. Cambridge, Massachusetts.

Wang, Z., & Zhu, J. 2013. Evolution of China's City-Size Distribution. *The Chinese Economy, 46*(1), 38-54. doi:10.2753/CES1097-1475460103

World Urbanization Prospects. 2014. Retrieved from: https://esa.un.org/unpd/wup/