# The Distribution and Concentration of Population in the United States, 1900-2000

Gregory K. Ingram
John Whitehead



**Lincoln Institute of Land Policy
Working Paper**



**Lincoln Institute Product Code: WP08GI1**

## Abstract

Spatial Gini coefficients are used to analyze the distribution and concentration of population in the U.S. by decade from 1900 through 2000. The analysis first uses states as units of observation and compares the results with those obtained from several other countries also using state or provincial data. These results show that the regional distribution of population in the U.S. has become more even from 1900 to 2000, and that countries range widely in the spatial concentration of their populations. Next, the analysis uses U.S. counties as the unit of observation to analyze population density distributions for the entire country and for each state. The U.S. national population has become more spatially concentrated across counties from 1990 to 2000, and state populations have also become more concentrated at the county level in three out of four states in the U.S. since 1900. An analysis using Massachusetts data suggests that counties are sufficiently small in size to capture the underlying changes in population density relevant at the urban level. Regressions then explore the determinants of the level and change of population concentration within states for the most recent decades.

## About the Authors

**Gregory Ingram**, President and CEO of the Lincoln Institute of Land Policy since June 2005, was formerly Director-General, Operations Evaluation at the World Bank, where he also held positions in urban development and research and was Staff Director for the *World Development Report 1994, Infrastructure for Development*. His areas of expertise include urban economics, housing markets, transportation, infrastructure, evaluation, and economic development.

gkingram@lincolninst.edu

**John Whitehead** is the Program Administrator for the Lincoln Institute of Land Policy's Program on the People's Republic of China. Before joining Lincoln, he received a bachelor's degree in International Affairs and Mandarin Chinese from the University of Puget Sound. Following graduation, John worked in the mining and energy industries in China. His interests include sustainable development, urbanization, migration, and transportation development.

jwhitehead@lincolninst.edu

## Table of Contents


Introduction 1

What is the expected distribution of population by density? 1

Data on population and areas 3

National distributions of population by density using state level data 3

Measuring national population concentration using state data for the U.S. 5

Measuring national population concentration in other countries 7

Measuring national population concentration using county data 8

Measuring state population concentration using county data 12

A modest check on the validity of counties as a unit of observation 15

Relating the level of population concentration to other variables 16

Relating the change in population concentration to other variables 18

Conclusion 20

References 22

## The Distribution and Concentration of Population in the United States, 1900-2000

### 1. Introduction

The density of population within settled areas, and particularly the growth of low density urban development, has received increasing attention in the past two decades. In the US, concern about sprawl underpins growth management policies—particularly smart growth—one of whose objectives is to promote population growth at higher densities than typical for suburban areas. While the density of development is often addressed at the level of cities or metropolitan areas, the distribution of population at a larger scale—at the level of nations, regions, and states—also provides a useful context.

The analysis reported here uses Gini coefficients as a summary measure of the distribution and concentration of population density in the U.S. by decade from 1900 through 2000. The analysis first uses states as units of observation and compares the results with those obtained from several other countries also using state or provincial data. These results show that the regional distribution of population in the U.S. has become more even from 1900 to 2000. The international comparisons show that such density convergence is common but not observed in all countries, and also that countries range widely in terms of the regional concentration of their populations.

Next, the analysis is repeated using U.S. counties as the unit of observation to develop measures of population density distributions for the entire country and also for each state. These results indicate that the urbanization of the population over time has made the national population more spatially concentrated across counties. In addition, state populations have also become more concentrated at the county level in three out of four states in the U.S. since 1900. A brief analysis using data from 351 cities and towns in Massachusetts produces similar results to those based on the state's 14 counties, suggesting that counties are sufficiently small in size to capture the underlying changes in population density relevant at the urban level. Regressions are then used to analyze the determinants of the level of population concentration and changes in the distribution of population by density within states for the most recent decades.

### 2. What is the expected distribution of population by density?

Students of urban economics have for many years analyzed how population density within cities varies as distance from the center of the city increases. This analysis assumes that cities have one predominant center (are monocentric). There is a large literature reporting estimates of the values of density gradients using a gradient function of the form:

$$D = D_0 e^{-bx}, \text{ where}$$

D is population density at a distance x from the center,
$D_0$ is the density at the city center,
e is the natural log base, and
b is the density gradient.

The empirical results from the density gradient literature show that the gradient, b, has declined in magnitude over time as cities have grown, that b is similar in magnitude in large cities around the world, and that b is absolutely larger in smaller cities in developing countries relative to smaller cities in the U.S. Empirical work has also shown that large cities typically develop a polycentric pattern with the emergence of subcenters of higher densities. That is, the monocentric assumption underlying the estimation of density gradients has become less consonant with reality over time [Ingram, 1998].

It is straightforward to transform this gradient function into a frequency distribution of population by density level, but this perspective has not received much attention in the literature. As an example, consider a density gradient where $D_0$ = 30,000 persons per square kilometer and b = -0.15, values of a density gradient for a city of about 8 million. For these values, figure 1 shows the density gradient; figure 2 the number of people in each one mile ring by distance from the center; and figure 3 the cumulative distribution of the population by density. The objective of this paper is to focus on analysis related to figure 3, the distribution of the population by density within the United States.

**Figure 1. Density gradient where $D_0$ = 30,000 and b = -0.15**

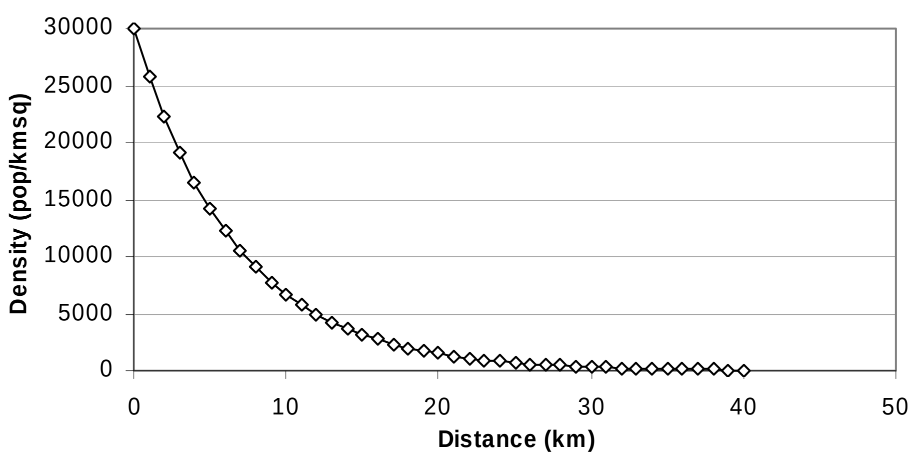
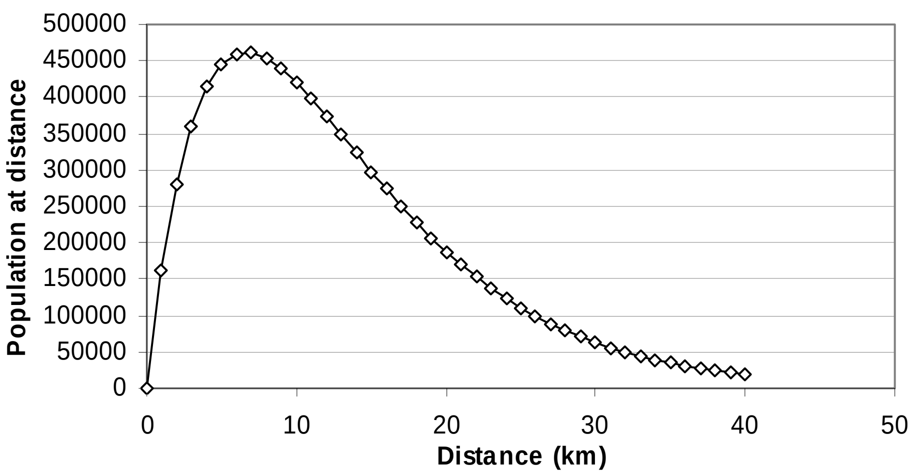



**Figure 2. Distribution of population by distance from center**



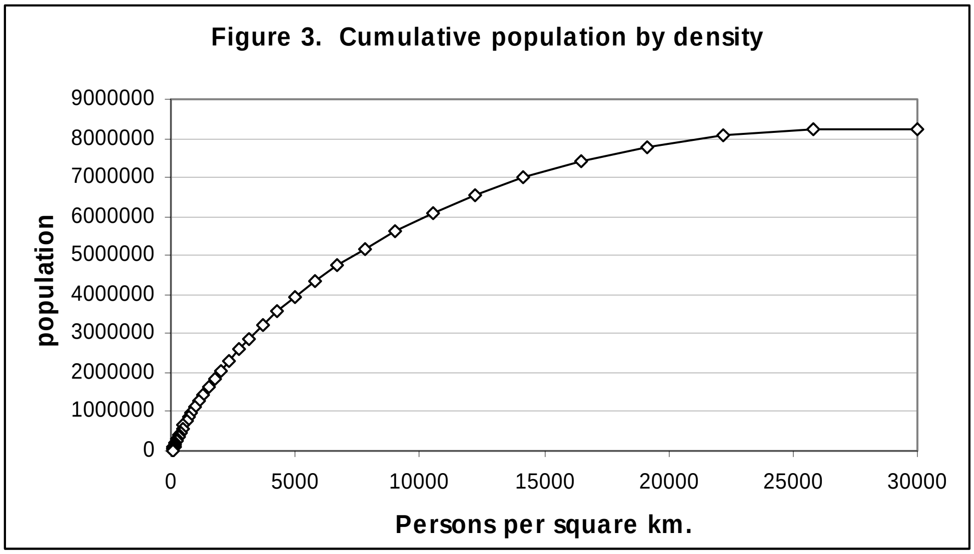



Figure 3 indicates the general shape (convex upwards) of the cumulative distribution of the population by density in a large metropolitan area. A province, state, or country is likely to have, in addition to a major city, a large share of its population living in smaller cities and rural areas with relatively low densities. Hence, the curvature of cumulative density gradients for larger areas is likely to be sharper and have a much larger share of the population at low densities than suggested in figure 3. In addition, the cumulative density distribution may not be smooth because numerous settlements at varying densities comprise the overall area.

## 3. Data on population and areas

The empirical work presented here for the U.S. uses decennial population and area information from 1900 through 2000 for the United States, principally for states and counties. Because Alaska and Hawaii became states only in 1959, data are used for the 48 continental states and are drawn from Census sources. For counties, data from 1950 forward are from Census sources, while data for several earlier decades were drawn mainly from U.S. Department of Agriculture sources. The number of states is constant at 48, but the number of counties in those states increased by nearly 240 between 1900 and 1920—from 2792 to 3031 as large counties were subdivided in western states—and the number then changed little, reaching 3092 in 1950 and 3108 in 2000.[1] Data for other countries are drawn from national census sources.

## 4. National distributions of population by density using state level data

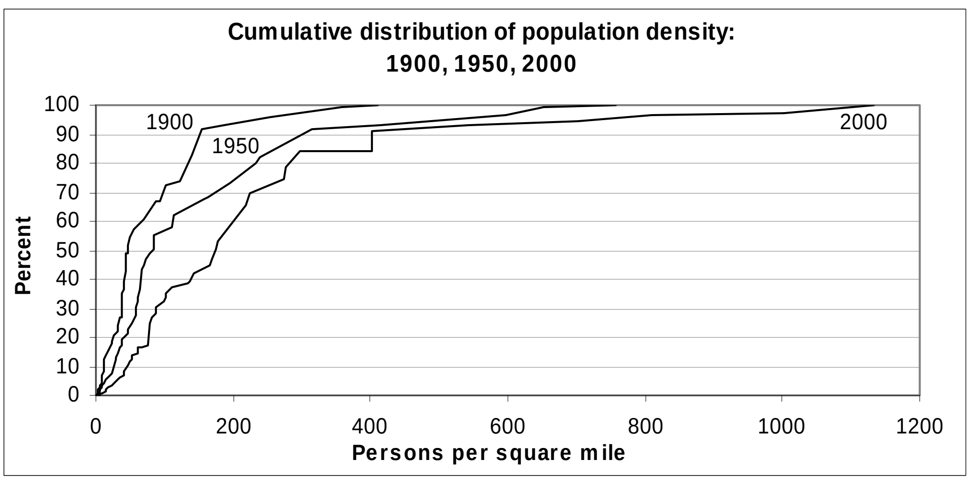
Data on area and population for each of the 48 states by decade from 1900 through 2000 were used to calculate state densities. Figure 4 summarizes the cumulative percentage distributions and frequency distributions of population by density level for 1900, 1950,

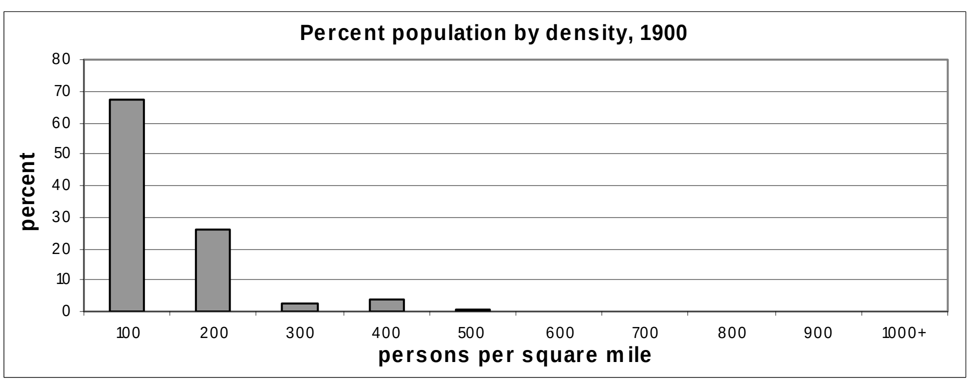
[1] See http://www.ac.wwu.edu/~stephan/Animation/us.gif for a graphic depiction of the increase in counties by decade

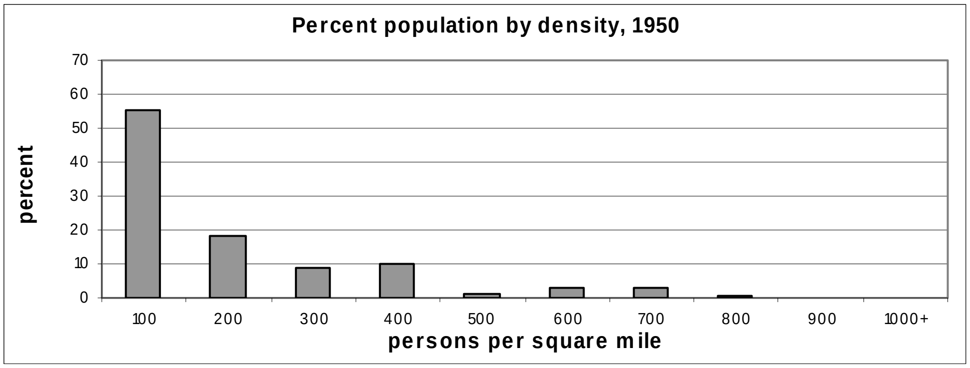
and 2000. Because population densities increased at all percentiles for each of these three periods, the three cumulative density distributions do not intersect. Figure 4 shows average population densities for states, so the population densities are low, with most states below 300 persons per square mile. Over the past 100 years maximum state densities have increased from about 400 persons per square mile (Rhode Island in 1900) to about 1100 persons per square mile (New Jersey in 2000), and the lowest densities observed have increased modestly.

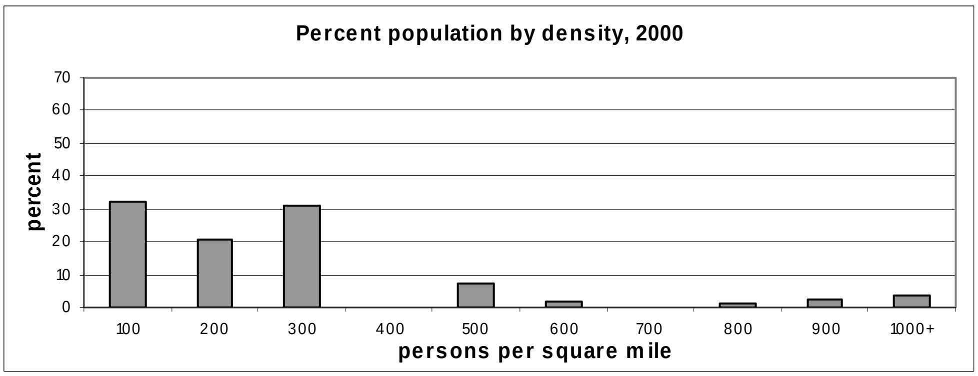
The frequency distributions of population by density correspond to the cumulative distributions and show that two-thirds of the U.S. population lived in states with average densities less than 100 persons per square mile in 1900. From 1900 to 2000, the range of average state densities doubled each 50 years. The frequency distribution in 2000 has a distinctly multimodal pattern that is typical of such distributions, and by 2000 the average density of several states is well above the highest state density observed in 1950. This increase in average density level is consistent with growth of high density development, but is not definitive evidence of that as it also reflects general population growth.

**Figure 4. Distribution of population by density using state averages**






### 5. Measuring national population concentration using state data for the U.S.

Analyzing the distribution of population in terms of its cumulative and frequency distributions conveys a sense of the underlying pattern of change, but it provides little information about the spatial concentration of population. For example, a reduction in the range of densities could occur while the population was spreading more evenly across all states or counties, or while the population was concentrating and becoming more evenly spread in only a few states or counties. What is needed is a single parameter that can measure the spatial concentration of the population in order to facilitate systematic comparisons over time and across jurisdictions.

The main tools used here to measure the level and change in the concentration of population over space are Lorenz curves and Gini coefficients.[2] Both measures have been used widely to monitor income concentration and income inequality. Instead of analyzing what percent of the population receives what percent of income (as is done to measure income inequality), we examine what percent of the land holds what percent of the population. In the analysis of population distributions across zones, the Gini coefficient has a value of 1 if population is completely concentrated (everyone resides in a single zone) and a value of zero if population is evenly distributed across all zones. Hence, rising Gini coefficients indicate a population's spatial concentration, and falling Gini coefficients indicate a population's deconcentration.

[2] Spatial Gini coefficients have been used to measure population and employment distributions in the literature, but not in the fashion that they are used here. For population examples see Eidlin 2005; Henderson and Wang, 2004; and Ying 1987. For employment, see He, Wei, and Pan 2007. For a discussion of Ginis versus other measures, see Tsai 2005.

Lorenz curves constructed using land areas and decennial census data for the 48 contiguous states for 1900, 1950, and 2000 are shown in figure 5. Their movement closer to the diagonal over time indicates that the national population became more evenly distributed across states during the two 50 year intervals. Figure 6, showing Gini coefficients for each decennial year, indicates that the deconcentration of national population across states occurred consistently every decade throughout the twentieth century. It also indicates that deconcentration occurred in two waves: one from 1900 attenuating through 1970, and another rapid reduction from 1970 on. The absolute reduction in population concentration across states is roughly the same from 1900 to 1970 as from 1970 to 2000.

**Figure 5. U.S. density Lorenz curves for 1900, 1950, 2000 using state data**

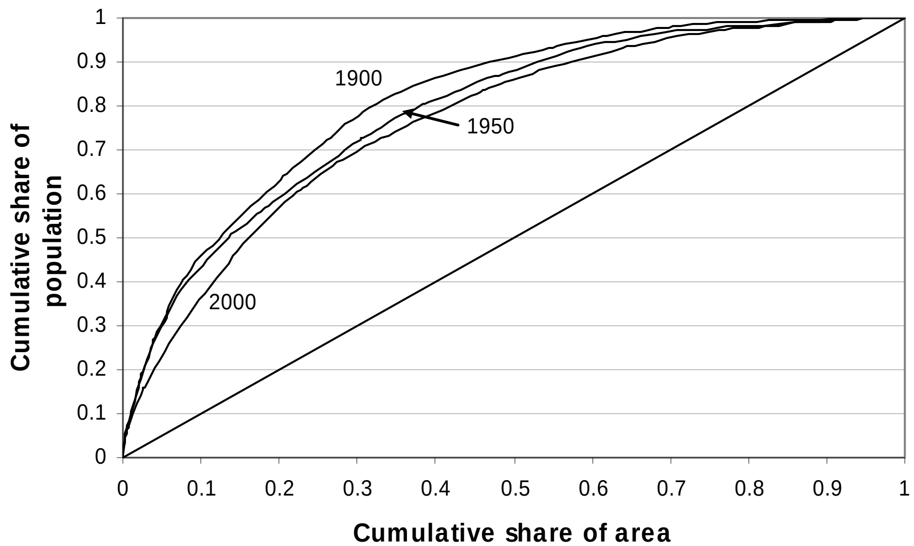
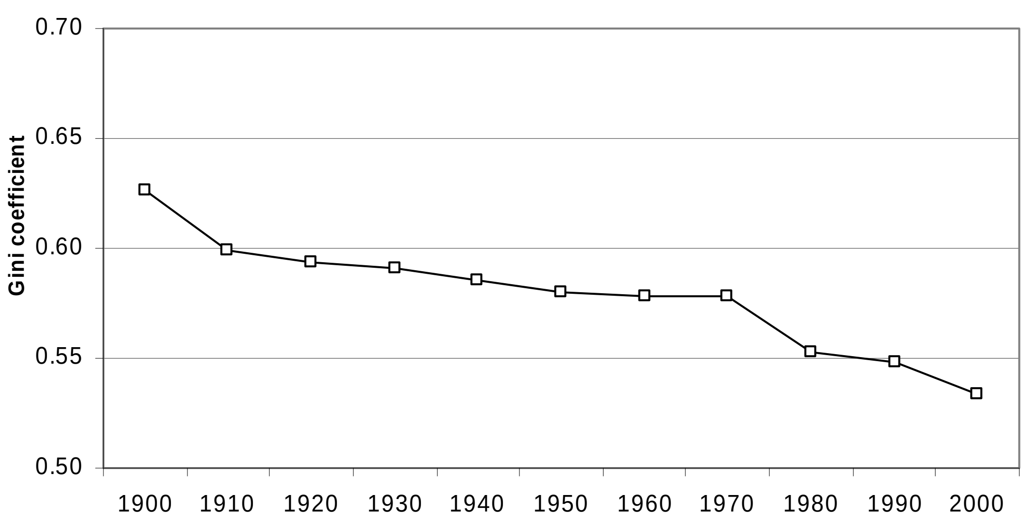
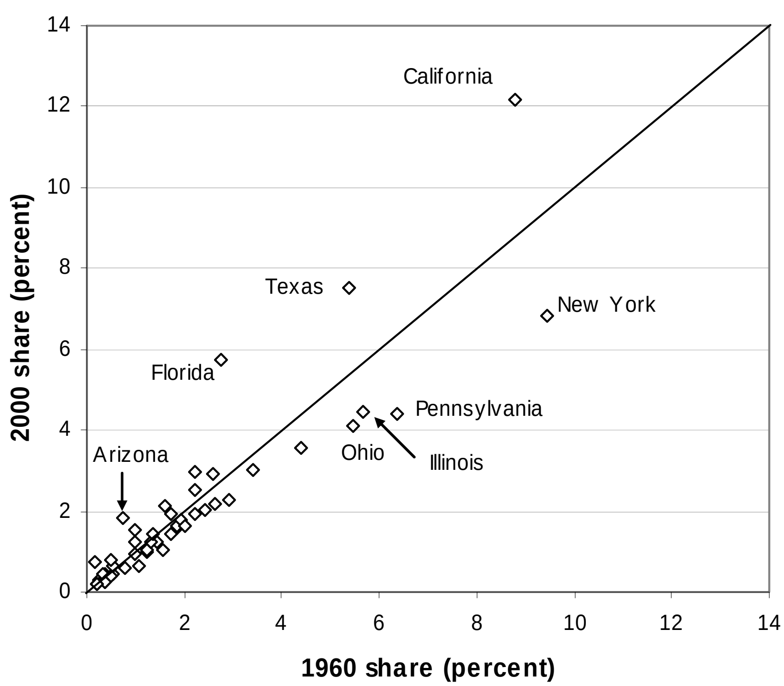



**Figure 6. National population distribution Gini, state data**



The first wave appears to be associated with the ongoing settlement of the country along with flows of migrants from the south seeking economic opportunity in the north, and flows of migrants from the north seeking more hospitable climates in the south and southwest. After 1970, the net flows were decidedly from the northeast and mid-Atlantic states to the south and west. From 1960 to 2000, four states (Illinois, New York, Ohio, Pennsylvania) experienced losses in their population that exceeded one percent of the nation's population, and four states (Arizona, California, Florida, Texas) increased their populations by more than one percent of the nation's population (figure 7). Accordingly, this inter-regional migration produced a more uniform distribution of population across states and a more even regional distribution of the U.S. population.

**Figure 7. Population share by state, 1960-2000**




## 6. Measuring national population concentration in other countries

Does the finding that the U.S. population became more equally distributed at the state level represent a universal trend or a tendency for populations to converge to a more equal distribution across national regions? To explore this question, data were assembled for several large countries using spatial areas analogous to states in the U.S. Many countries have major subdivisions such as states or provinces for which national census data are often available. The sample of countries presented here includes those with data readily available for at least a 50 year period. Areas and populations were assembled and used to create Lorenz curves and to calculate Gini coefficients for population density. Results are displayed in figure 8 for Argentina, Brazil, China, India, and the U.S.

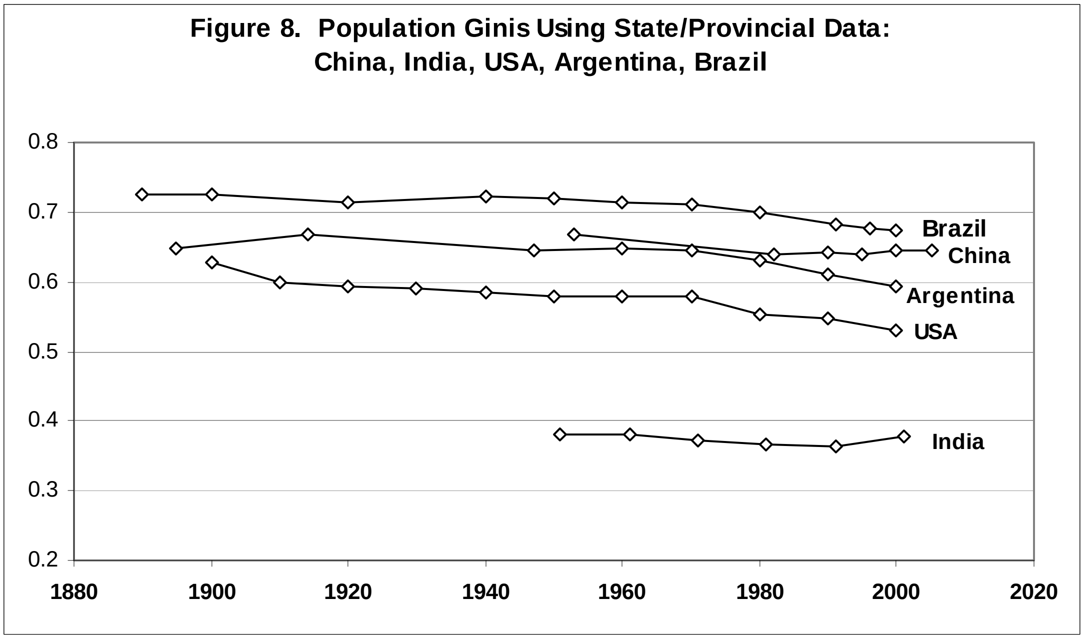



Figure 8 suggests that there is a general tendency for population density Gini coefficients to decline over time for countries that have relatively concentrated populations as indicated by relatively high Gini coefficients. India has a much lower population Gini coefficient, indicating that its population is much more equally distributed across provinces than is the case for the other countries. Moreover, in India the Gini coefficient has increased from 1991 to 2001, essentially returning to its 1961 level, so that overall there is essentially no change in the concentration of India's population at the provincial level. The regional distribution of population across countries is obviously affected by specific country geography, such as the presence of mountains and deserts, and is a topic worthy of additional attention.

## 7. Measuring national population concentration using county data

Using data at the state or provincial level provides a useful perspective on regional population distributions at the national level, but does not address metropolitan development or the distribution of population at a scale relevant for assessing the impacts of state-level growth management policies. Metropolitan areas are normally comprised of counties, and county level data provide information about the density patterns within metropolitan areas, about impacts of growth management policies, and about how population densities are affected by population movements from rural to urban areas. Figures 9, 10, and 11 present cumulative distributions of the U.S. population by density at the national level for 1900, 1950, and 2000. These figures parallel the cumulative density distributions in figure 4 that use state level data.

These cumulative distributions are similar in shape to those based on state level data, but their curvature is sharper and the range of density is much greater, rising nearly to 90,000 persons per square mile in 1950. This is not surprising since there are 3108 counties in the U.S. 48 states in 2000, and some counties are completely urbanized and have very high densities.  For example, in 1900, 1950, and 2000 New York County had the highest recorded density.  It comprised Manhattan and the Bronx in 1900 but only Manhattan in 1950 and 2000, so its density increased from 1900 to 1950 because of its re-definition. Its density decline from 1950 to 2000 reflects Manhattan's population reduction from 1.96 million in 1950 to 1.54 million in 2000.

**Figure 9.  National population by density, 1900: county data**

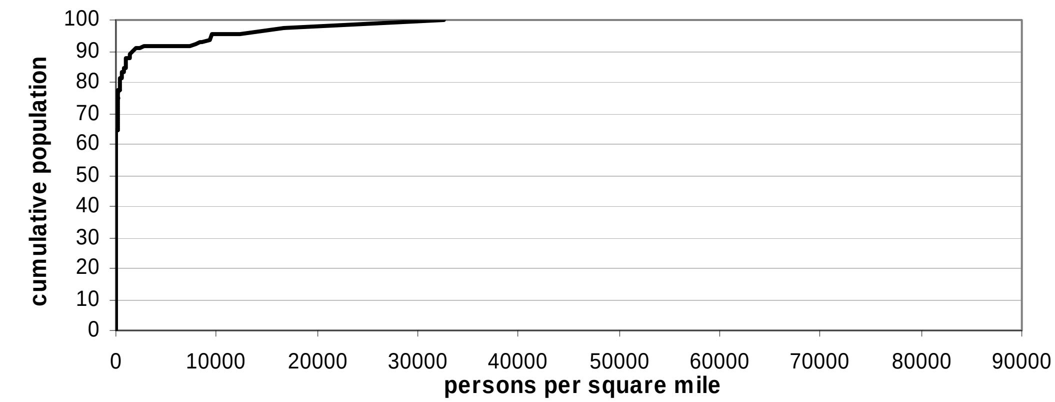



**Figure10.  National population by density, 1950: county data**

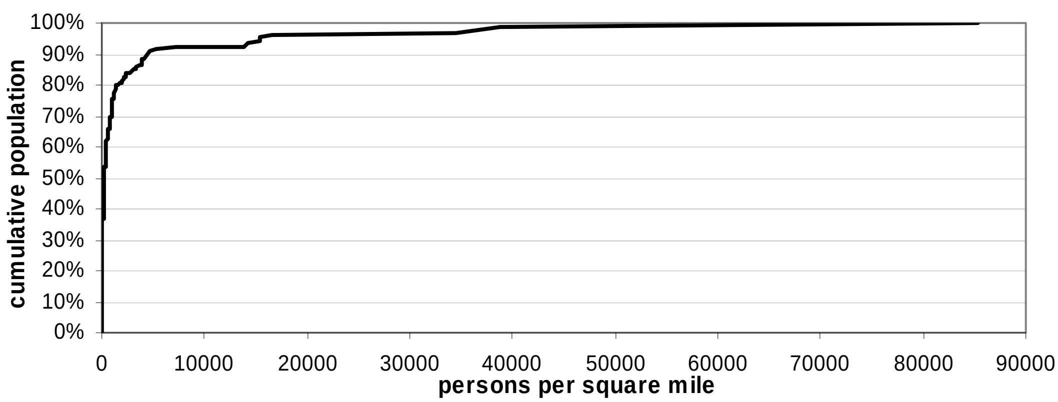
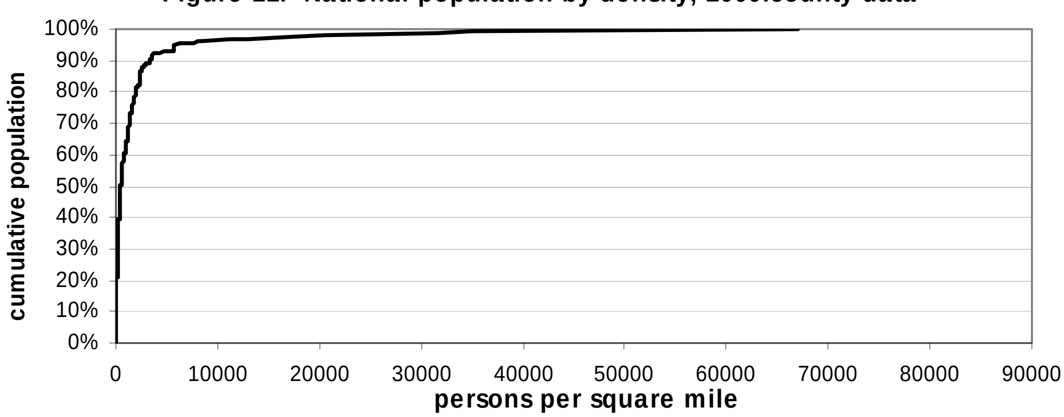



**Figure 11.  National population by density, 2000:county data**



Using the county data to construct density Lorenz curves at the national level, shown in figure 12, produce different results from those shown in figure 5 that used state level data. The Lorenz curves relating national population to area using county data for 1900, 1950, and 2000 indicate that at the county level the U.S. population became more concentrated during the twentieth century. Figure 12 shows that the population in 2000 is consistently more spatially concentrated than in 1900 because the Lorenz curve for 2000 lies northwest of, or coincides with, that for 1900. The Lorenz curve for 1950 intersects both the 1900 and 2000 Lorenz curves. Thus in 1950 the population was slightly more concentrated at the very highest density levels than in 2000, while at low density levels in 1950 the population was somewhat more dispersed than in 1900.

**Figure 12. U.S. density Lorenz curves 1900, 1950, 2000: county data**

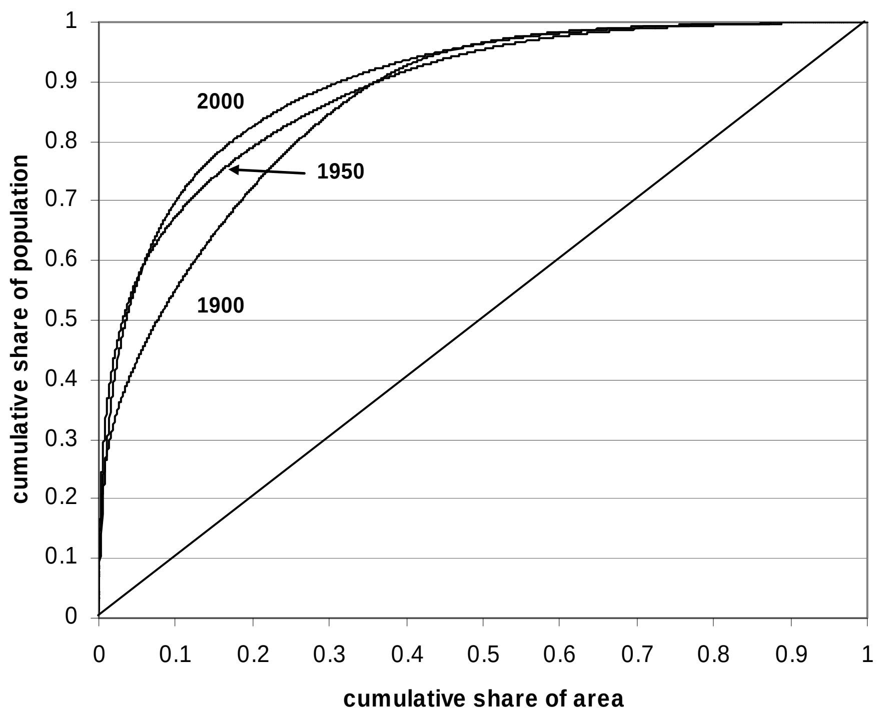
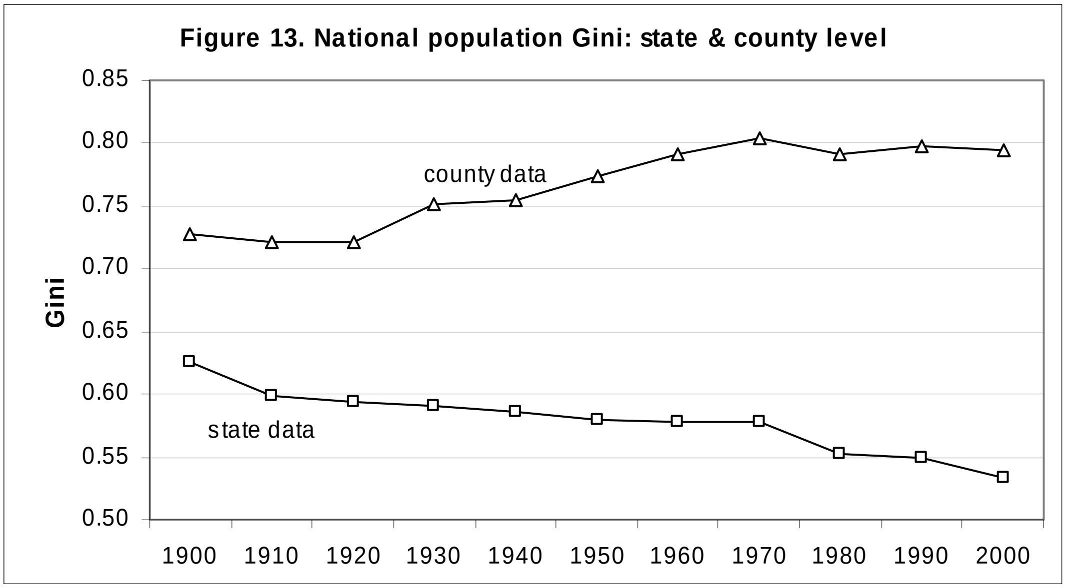
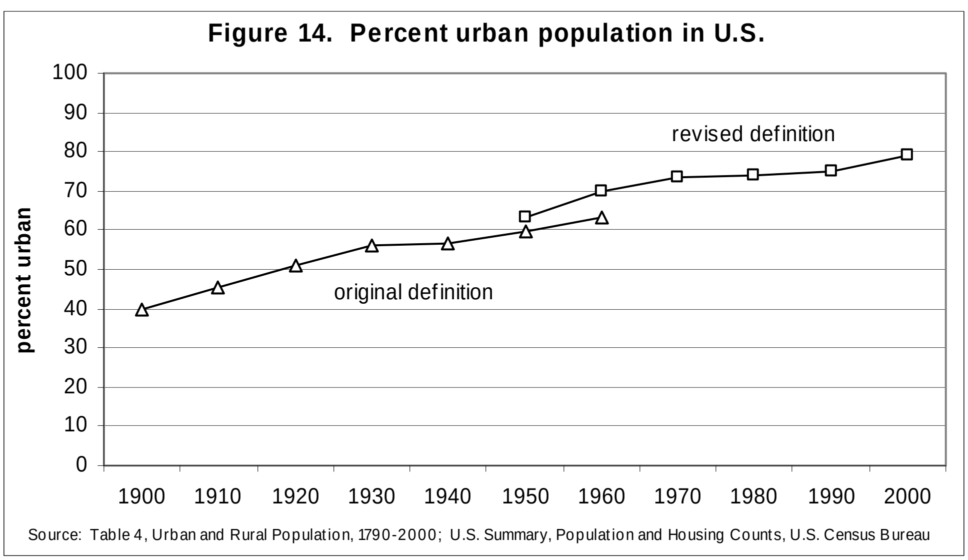



Figure 13 displays the national population density Gini coefficients for each decade based on the county data, and also reproduces the state level Gini coefficients from figure 6 for comparison. The county level Gini coefficients increase steadily from 1920 to 1970 (with a pause from 1930 to 1940) and then change little in the remaining decades. The change in trend in 1970 coincides with that observed with the state level data. A notable empirical regularity with the state- and county-based Gini coefficients is that they seem to change in relay fashion. That is, when the county-level Gini coefficient is stable, the state-level Gini changes, and vice-versa.




One of the major shifts contributing to the concentration of the nation's population observable at the county level is urbanization—the movement of population from rural to urban areas. This phenomenon is readily discernible with county level data because most counties are predominantly urban or rural. State level analysis is generally at too coarse a scale to measure the effects of urbanization. This difference in scale between states and counties accounts for the deconcentration found using state level data (reflecting inter-regional migration) and the concentration found using county level data (reflecting rural to urban migration).

Figure 14 shows how urbanization has changed in the U.S. from 1900 to 2000. The Census Bureau revised the definition of urbanization in the 1960s, and figure 14 shows the time series using both definitions. Over the 100 year period, the population share living in urbanized areas doubled from 40 to 80 percent. While there are two definitions



for 1950, both are close to 60 percent—midway between the 1900 and 2000 share. Figure 14 also shows a slowdown in urbanization after 1970, consistent with the stability of the county-based population Gini coefficient after 1970.

The 40, 60 and 80 percent of the population that were urbanized in 1900, 1950, and 2000 occupy changing shares of land area in each of these years. While county areas do not correspond to urbanized areas, the county areas that match specific population shares can be obtained from the data underlying figure 12. These areas are shown as shares of the national area in figure 16 for 40, 60 and 80 percent of the population living in the densest counties in each of the three years: 1900, 1950, and 2000. Figure 15 shows that populations became much more concentrated among counties between 1900 and 1950 at all three population shares because for each population percentage living in the densest counties, the share of land occupied by those densest counties declined sharply.

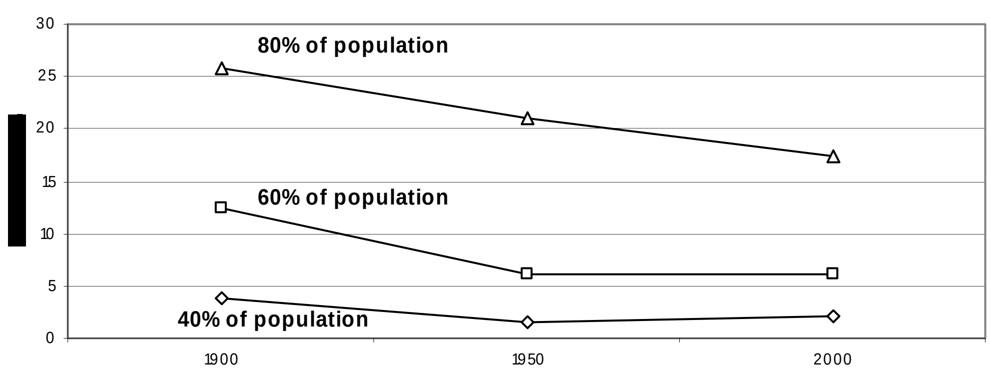
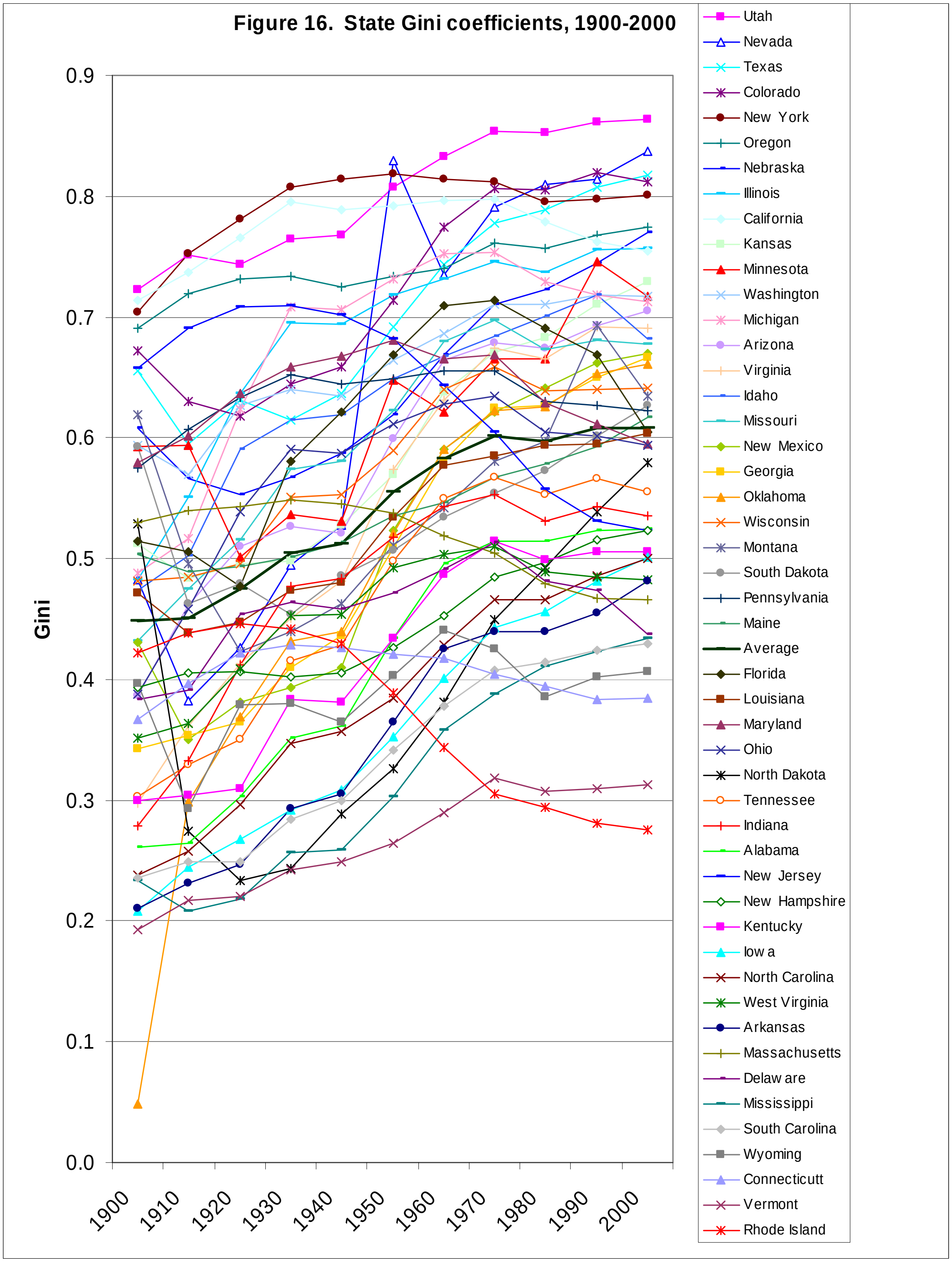



**Figure 15. Share of land area occupied by counties containing 40, 60, and 80 percent of national population**

To give a sense of the increased concentration over time, figure 15 shows that 80 percent of the population in 1900 lived in counties covering 26 percent of the land, in 1950 this same share lived in counties covering 21 percent of the land, and in 2000 this same share lived in counties covering 17 percent of the land. From 1950 to 2000, the 48 state U.S. population increased from 150 million to 278 million while the area of the 48 states remained constant. Eighty percent of the population (220 million in 2000 and 120 million in 1950) were living in counties covering a smaller area in 2000 than in 1950.

## 8. Measuring state population concentration using county data

As at the national level, county data have also been used to estimate Gini coefficients for each state and for each decennial year from 1900 through 2000. These are displayed in figure 16, which also shows the arithmetic average of the Gini coefficients across states for each decennial year. The state names in the key are in the same order as the state Gini coefficients in 2000. Evident from the average and for most states is a pattern of increasing Gini coefficients from 1900 to 2000, indicating an increasing concentration of population similar to the change observed at the national level with county data. Also notable is the wide range of Gini coefficients, reflecting the markedly different distributions of population across states.  This wide range is consistent with the wide range of Gini coefficient values shown in figure 8 for countries.

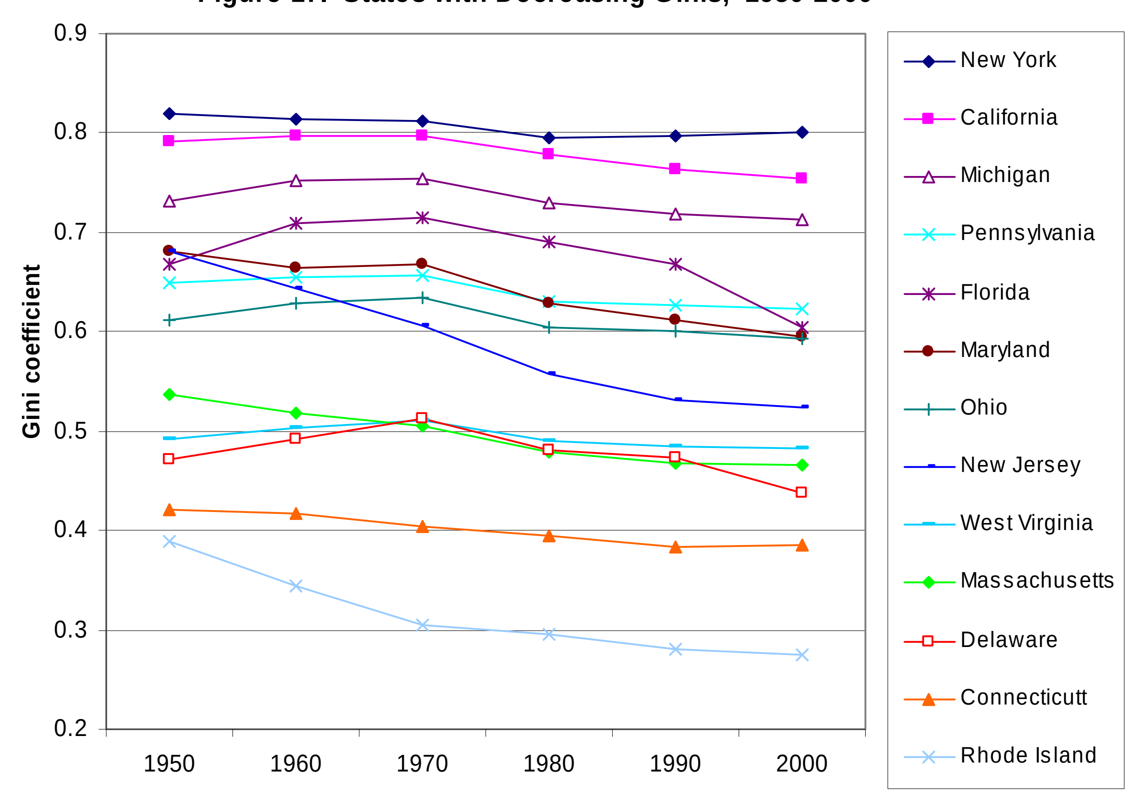


In 2000 the four states with the lowest Gini coefficients (most dispersed) are Rhode Island, Vermont, Connecticut, and Wyoming. Rhode Island and Connecticut are heavily urbanized states, and Vermont and Wyoming are states with dispersed populations and no large cities. At the other extreme, the five states with the highest Gini coefficients (most concentrated) are Utah, Nevada, Texas, Colorado, and New York. These are states with one or more large cities that contain a large fraction of each state's population.

With respect to changes over time only three states—New Jersey, Massachusetts, and Rhode Island—have Gini coefficients in 2000 that are smaller than in 1900. For each of these states the Gini increased from 1900 through 1920 or 1930 and then declined. These were also the three densest states in every decade from 1900 through 2000. The fourth densest state every decade, Connecticut, has an overall Gini coefficient pattern of increase and decline similar to the other three densest states, but Connecticut's Gini in 2000 is still slightly above its value in 1990. Both the state and county level analysis so far has revealed that 1970 seemed to mark a change in trends of density and of Gini coefficients. Such a change is evident in the average of the state Ginis in figure 16. It is also evident for some individual states, but not for all.

While the 100 year pattern of Gini coefficients provides a useful historic context, the policy interest in influencing patterns of population distribution is recent. Accordingly, we note those states that have experienced a reduction in Gini coefficients from 1950 to 2000 in figure 17. Again, the state names in the key are in the same order as the state

**Figure 17. States with Decreasing Ginis, 1950-2000**



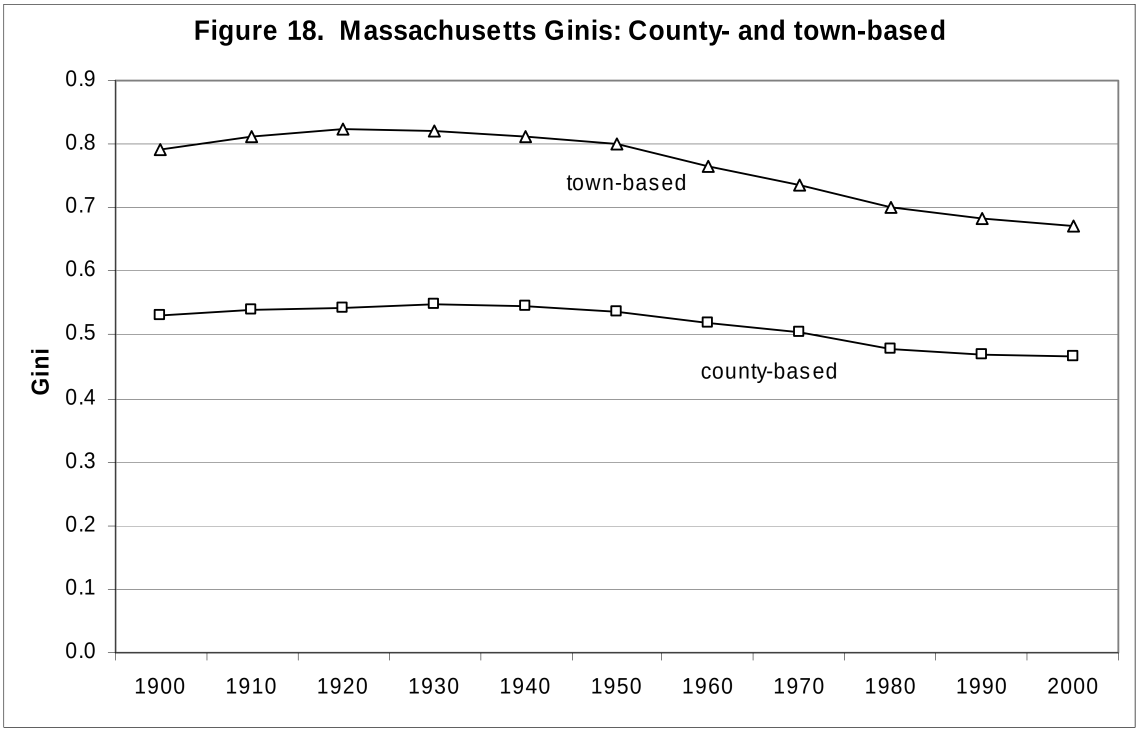
Gini coefficients in 2000. These 13 states experienced deconcentration of their populations over this period while the populations in the other 35 states became more concentrated. These thirteen states have both common and diverse characteristics. Ten of the states experiencing population deconcentration are in the middle-Atlantic region in a cluster stretching from Massachusetts to Maryland. Three states, New York, Ohio, and Pennsylvania, lost more population than other states over this period, while two other states, California and Florida, gained more population than any other states over this period. Michigan is a 'rust-belt' state like Pennsylvania and Ohio. Twelve of these thirteen states are among the 15 most densely populated states in 2000, with West Virginia the exception. Three have state-level growth management policies. The range of values for the Gini coefficient for these deconcentrating states is very broad, similar to that for all states.

### 9. A modest check on the validity of counties as a unit of observation

The analysis of the distribution and concentration of population by density presented so far has made clear that the size of the unit of observation can have a dramatic impact on results over time. At the national level, analysis shows that the U.S. population is becoming less concentrated when states are the unit of observation and more concentrated when counties are the unit of observation. Results also indicate that populations are more concentrated (Gini coefficients are larger) when the unit of observation is smaller. An obvious question is whether analysis at the county level is capturing appropriately the changes over time in population concentration that are relevant to the evaluation of growth management policies. One way of assessing the adequacy of the county as a unit of observation is to compare county-level results with results from a smaller spatial unit.

Census totals are typically reported for different levels of government. Below the county level, some states have smaller units of governments that cover the entire area of each county. Massachusetts is one such state where the entire area of its counties is divided into cities and towns, and where data on the population and area of the cities and towns are available from 1900 through 2000. The number of cities and towns increased somewhat over the past century and totaled 351 in 2000, whereas Massachusetts had 14 counties in 2000.

Population Gini coefficients using town- and county-based data were calculated for each decennial year from census data and are displayed in figure 18. As expected, the town-based Gini coefficients are larger in magnitude than the county-based coefficients because their areas are smaller. However, the trend over time of the two sets of Gini coefficients is very similar, with an increase in concentration in the early part of the century followed by a decrease after 1930 and then a subsequent slowing in the rate of decrease. The simple correlation coefficient between the two time series is 0.995. While this is far from definitive evidence that county-based data will capture all effects of growth management policies in all states, this single result provides some reassurance that the analysis of county-level data is worth pursuing.




## 10. Relating the level of population concentration to other variables

One striking aspect of the pattern of Gini coefficients is that states with similar Gini levels or changes seem very diverse. The most concentrated states include western states with large areas that are uninhabited—as well as New York where the concentration stems from the large population located in New York City. Similarly, at the deconcentrated end of the scale, one finds Rhode Island, Vermont, Connecticut, and Wyoming—another disparate grouping of states. The challenge is to see what state characteristics are associated with concentrated or deconcentrated populations. Such an analysis may provide some insights about possible entry points for policies that could influence the distribution of population within states and also a framework that can be used to see if growth management policies are having any impact on population distributions and concentrations within states.

The first step is to determine what state characteristics are associated with the Gini coefficient's level—which varies dramatically across states. The variables that are explored here are characteristics of the states (geography, development, economy, and infrastructure) and of the population. Geographic characteristics include the size of the state and the average size of its counties. Large states are likely to have more concentrated populations, while states with larger (fewer) counties are likely to be less concentrated. Higher densities are likely to promote de-concentration, while states with larger shares of their land areas in urban use are also likely to be more deconcentrated. The only infrastructure variable is miles of urban roads, assumed to be associated with concentration. The economic variables that are readily available are household median income and housing vacancy rates; the latter is likely to be inversely related to housing prices. Income might reduce concentration given evidence that the income elasticity of demand for residential lot size is high in the U.S., but high incomes may also signal a high level of urbanization—consistent with concentration. Higher vacancy rates (suggesting lower housing prices) would tend to be associated with deconcentration.

The statistical analysis takes the level of the Gini coefficient for the state in 2000 as the dependent variable, and the independent variables are "level" variables for the attributes described in the previous paragraph. The unit of observation is the state, so all variables are defined at the state level. The results from seven regressions are shown in Table 1. The first four regressions are bivariate for state area, median household income, the share of state land classified as urban by the Census, and the housing vacancy rate. The fifth regression combines these four variables; all are significant, and together they explain two-thirds of the variation in the levels of state Gini coefficients in 2000. The remaining

**Table 1. Regression results: Gini level in 2000**
Dependent variable is 2000 state Gini

| Independent variable | Regression coefficients; [t value] [1] | [2\ | [3] | [4] | [5] | 6] | [7] |
|---|---|---|---|---|---|---|---|
| Intercept | 0.494<br>[18.6] | 0.485<br>[3.53] | 0.652<br>[29.5] | 0.732<br>[10.6] | 0.294<br>[2.24] | 0.336<br>[2.40] | 0.298<br>[1.66] |
| State area Sq. miles | 0.0000019<br>[5.39] | | | | 0.0000012<br>[3.73] | 0.00000088<br>[2.07] | 0.00000095<br>[1.60] |
| Median HH Income, 1999 | | 0.000003<br>[0.91] | | | 0.00001<br>[3.95] | 0.0000093<br>[3.30] | 0.0000095<br>[3.21] |
| Urban land share | | | -0.64<br>[-3.45] | | -0.87<br>[-4.81] | -0.879<br>[-4.78] | -0.814<br>[-3.28] |
| Vacancy rate 2000% | | | | -0.0124<br>[-1.87] | -0.012<br>[-2.57] | -0.0129<br>[-2.48] | -0.0117<br>[-2.08] |
| Population 2000 | | | | | | 2.50E-09<br>[0.98] | 7.08E-09<br>[0.61] |
| Avg. county area, sq. mi. | | | | | | 0.0000096<br>[ 0.87] | 0.000011<br>[0.298] |
| Density popn/sq.mi. | | | | | | | 0.00021<br>[0.29] |
| Urban road miles | | | | | | | -1.90E-06<br>[-0.41] |
| R squared | 0.39 | 0.02 | 0.21 | 0.07 | 0.66 | 0.67 | 0.67 |

two regressions add four more variables:  state population, average county area, population density, and miles of urban road.  All are insignificant and they add little to the explanatory power of the fifth regression.

In terms of signs, most are as predicted.  Household median income is positively associated with concentration, suggesting that higher incomes are associated with urbanization and population concentration.  A higher urban share of land is associated with deconcentration, as is the vacancy rate, both as expected.  Surprisingly, population density has no explanatory power.  Given the apparent heterogeneity of states with similar levels of population concentration, it is surprising that a few independent variables explain two thirds of the variance in concentration.

## 11.  Relating the change in population concentration to other variables

While the level of the Gini coefficient is of interest, especially given its wide range across states, most policy interest relates to the change in the distribution and in the Gini coefficient over time.  Interventions such as smart growth policies have as one objective increasing the density and concentration of the population, and therefore increasing the Gini coefficient.

The variables used to explain the change in the Gini coefficient are closely related to those used to explain the level of the Gini coefficient, with many used in change form. That is, the dependent variable is the ratio of the Gini coefficient in one decade to that in the previous decade, and several independent variables are similar ratios, so their coefficients are elasticities.  Thus, the income variable is the ratio of median income at the end of a decade to its value at the beginning, and similarly for state population and state urban road miles.  "Level" variables are used to control for state geography (average county area and number of counties), and for population density and housing vacancy rates (both at the beginning of the respective decade).  An additional variable, each state's Gini coefficient at the beginning of the decade minus the corresponding average Gini, tests for any tendency to revert to the mean.  A lagged dependent variable tests for perseverance of trend.  The regressions results are first pooled across two decades, 1980 to 2000, giving 96 observations, and then presented for each decade separately.  Results are in Table 2.

A first observation about the results in Table 2 is that the explanatory power of the variables is much greater in the decade of the 1980s than the 1990s.  Yet the pattern of signs and the magnitudes of coefficients are similar across the two decades, with the exceptions of urban road miles (always insignificant) and the lagged dependent variable. The lagged dependent variable is highly significant in the 1980s and becomes completely insignificant in the 1990s.  This indicates that in the 1980s many states continued their trends in concentration change from the 1970s, whereas the 1990s trends changed and the 1990s seem to mark a turning point for many states.  The state Gini minus the cross-state mean Gini has consistently negative and significant or nearly significant coefficients, suggesting a consistent tendency to move toward the mean value of the Gini.  High density states are likely to see their populations deconcentrate, although the significance of this variable diminishes greatly in the 1990s.  Changes in income and population are

**Table 2. Regression results: Change in Gini coefficient**

Dependent variable is ratio of Gini at end of decade to Gini at beginning of each decade
Regression coefficients; [t value]

| Independent variable | 1980-1990 & 1990-2000 {1} | 1980-1990 {2} | 1990-2000 {3} |
|---|---|---|---|
| Intercept | 0.936 [8.15] | 0.627 [4.44] | 1.16 [5.36] |
| Median HH income ratio for decade | -0.038 [-0.93] | 0.081 [1.25] | 0.02 [[0.18] |
| Population ratio for decade | -0.053 [-0.99] | -0.12 [-1.55] | -0.089 [-1.12] |
| Urban road miles ratio for decade | -0.00011 [-0.39] | -0.00045 [-1.25] | 0.000083 [0.18] |
| Previous decade's Gini ratio (lagged dep. Var.) | 0.19 [1.82] | 0.43 [3.38] | -0.07 [-0.44] |
| Gini - mean Gini at decade start | -0.073 [-2.40] | -0.066 [-1.62] | -0.086 [-1.96] |
| Population density at decade start | -0.000053 [-2.74] | -0.000073 [-2.75] | -0.000052 [-1.49] |
| Housing vacancy rate at decade start | -0.0021 [-1.76] | -0.00084 [-0.335] | -0.0021 [-1.30] |
| Average county area (sq. miles) | 0.0000074 [2.11] | 0.000013 [2.58] | 0.0000097 [1.83] |
| No. of counties | 0.00015 [1.57] | 0.00027 [2.03] | 0.00014 [0.97] |
| 1990s decade dummy | -0.011 [-1.41] | | |
| R square | 0.396 | 0.613 | 0.239 |

not significant, although income has a positive sign in the decade regressions. The measures of county number and size are often significant—showing that more and larger counties are associated with concentration. Looking at the effect of these two county coefficients together while holding state size constant (where an increase in the number of counties by ten percent would decrease the average county area by ten percent), indicates that the net effect would be to increase concentration, as expected. In the pooled regression across the two decades, the dummy variable for the 1990s is negative, indicating an overall deconcentration of the population in the 1990s.

The surprising outcome of the regressions displayed in Tables 1 and 2 is the relatively high proportion of the variance in the level and change in state Gini coefficients that is associated with a few variables. This is surprising because of the seeming diversity of states with low and high Gini coefficient levels, and of states with changes in their Gini coefficients. Moreover, the pattern of coefficients accords with expectations. A driving force behind population concentration has been urbanization, because urbanization increases the difference in population densities between urban and rural areas. However, states that are highly urbanized or that have high overall population densities have smaller differences in population densities between urban and rural areas, a smaller share of rural area, and higher rural densities. The signs of coefficients were generally similar in the level (table 1) and change (table 2) regressions although their significance varied. In particular, density had no association with Gini levels, but a strong association with the change in Gini levels over time.

## 12. Conclusion

This paper demonstrates that spatial Gini coefficients are a useful measure of population concentration at different spatial scales. Using state or provincial data, it provides a continuous and robust measure of regional population distribution across countries that is free of distortions introduced by differing national-level definitions of urbanized land or urban areas. Using county-level data, the spatial Gini coefficient measures population concentration at a scale that is relevant for urban analysis. The preliminary evidence presented here suggests that county-level data produces results similar to that yielded from municipal-level data. Further testing of this correspondence is necessary. While not explored here, county-level data are likely to be available for many countries over time, facilitating more comparable cross-national analysis over time.

The results using state and provincial data for the U.S. and other countries show modest but quite evident changes in the regional distribution of population over time. In the U.S., the population has become more evenly distributed across states in the past century as people have moved from the mid-Atlantic states to the south and west. Comparisons across countries (and across states) show very wide ranges in the spatial Gini coefficient, indicating that populations are distributed very differently within countries and states. From 1900 to 2000, the measured change in the concentration of population in the U.S. is larger than for the few other countries analyzed so far. While population has become more evenly distributed in the U.S., this is not a universal pattern across countries.

Using county data for the U.S., the results indicate that population has become more concentrated at the county level. This is mainly the result of the urbanization of the population, as the urbanized share of the U.S. population doubled from 40 percent in 1900 to 80 percent in 2000.

Using county data to measure the concentration of population by state reveals very wide differences in population concentrations across states, and large changes in state concentrations from 1900 to 2000. The predominant change at the state level echoes the national trend:  population became more concentrated at the county level in three out of four states in the twentieth century.  Of the thirteen states whose populations became less concentrated from 1950 to 2000, all are mid-Atlantic states except for Michigan and Florida.  High population density is strongly associated with a decline in population concentration over time.

While those states that are at the extremes of population concentration or dispersal seem like odd bedfellows, the level of population concentration as measured by the spatial Gini coefficient is strongly associated with a few variables—state area, income, housing vacancy rates, and urban land share—that together explain two thirds of its variance in 2000.  The statistical analysis is less successful explaining the change in the spatial Gini coefficient over time.  While population density is not correlated with the level of the spatial Gini, it is strongly associated with its change.  The analysis also suggests that the changes in the spatial Gini had different associations with other explanatory variables in the 1980s relative to the 1990s.

## References


Eidlin, Eric (2005), "The Worst of All Worlds: Los Angeles, California, and the Emerging Reality of Dense Sprawl," *Transportation Research Record*, Vol. 1902, pp. 1-9. Calculates spatial Ginis using census tracts for LA, SF, NY

He, Canfei, Y.H. Dennis Wei, and Fenghua Pan (2007), "Geographical Concentration of Manufacturing Industries in China: The Importance of Spatial and Industrial Scales," *Eurasian Geography and Economics*, Vol 48, No. 5, pp. 1-23.

Henderson, V., and H.G. Wang (2004), Urbanization and City Growth. *Brown University Working Paper* (http://www.econ.brown.edu/faculty/henderson/papers.html). Ginis across countries based on city populations.

Howell-Moroney, Michael (2007), "Studying the Effects of the Intensity of US State Growth Management Approaches on Land Development Outcomes," Urban StudiesVol 44, No. 11, pp. 2163-2178, October.

Ingram, Gregory K. (1998), "Patterns of Metropolitan Development: What Have We Learned?" *Urban Studies*, Vol. 35, No. 7 (June) pp. 1019-1035.

Tsai, Yu-Hsin (2005), "Quantifying Urban Form: Compactness versus 'Sprawl'", *Urban Studies*, Vol. 42, No. 1, 141–161, January 2005. Uses spatial ginis and other indices to measure distribution and concentration of population.

Ying, Killian Ping-Hung (1987***).*** "District population change in Hong Kong, 1961-1981: some policy implications". *Urban Geography*, Vol. 8, No. 1, Jan-Feb. 31-54 pp. Uses spatial Gini coefficient at the census district level in Hong Kong to analyze the trend in population concentration.