Urban Patterns of the Largest Urban Areas in the United States, 1950-2020: A New Dataset

John R. Ottensmann
Indiana University-Purdue University Indianapolis
john.ottensmann@gmail.com
urbanpatternsblog.wordpress.com
February 2023

Abstract

A new dataset has been developed using housing units for census tracts to identify the 56 largest urban areas in the United States in 2020. These areas have been delineated following as closely as possible the procedures being used by the Census to define Urban Areas for 2020. Census tract counts and estimates for each census back to 1950 have likewise been used to define urban areas in the same way for each census year. The result is a dataset having urban areas defined in a consistent manner for each year with census tract data on housing units that can be used to study patterns within the urban areas and their change over time.

Introduction

The original urban patterns dataset used census tract data to delineate large urban areas from 1950 to 2010 for research on urban structure. The dataset is documented in Ottensmann (2014). These data have been used to examine a wide range of topics. The work has resulted in the production of over 30 papers which are described and are available here.

The release of data from the 2020 census raised the obvious possibility of updating the dataset to extend the range of years covered. After some consideration, the decision was made to create a new dataset rather than simply adding the 2020 data to the current urban patterns dataset. Three things led to this decision: Experience gained from using the original data over a period of years showed ways in which the creation of the dataset might be improved. The availability of new data, especially for land cover, provided opportunities for enhancement. And finally, creation of a new dataset allowed the delineation of the urban areas to be made as consistent as possible with the 2020 Urban Area definition, which was greatly changed from previous years.

This paper documents the new *Urban patterns 2* dataset. The next section describes the enhancements in the creation of the dataset compared with the original. Data sources are described next. The following sections then cover the various steps in producing the dataset. First comes the identification of the largest urban areas to be

included in the dataset. The delineation of the urban areas for each census year is the central task. Since some of the urban areas consisted of two or more separate urban areas that had grown together, identifying these urban centers follows. Distances from census tracts to the Central Business Districts of the urban centers requires the delineation of those areas and the calculation of the distances.

Enhancements compared to the original urban patterns dataset

This section describes the major ways in which the development of the new dataset was improved compared to the original. Only brief descriptions of the new approaches are provided here. Full documentation is given in the following sections. The first enhancement of course is the incorporation of data from the 2020 census, extending coverage forward by a decade.

Another major improvement is the use of 2020 census tracts as the units for the new dataset. Since most of the data for the original urban patterns dataset came from the Neighborhood Change Database, with data from earlier censuses normalized to the 2000 tract boundaries, those were the tracts used. The 2010 census population and housing units were assigned to these tracts using block data. This led to less than satisfactory situations at the fringes of some urban areas. Tracts are defined to have similar populations. As new development occurs, large undeveloped tracts are subdivided at each census to reflect new development patterns. Using the 2000 tracts meant some of the tracts at the edges of urban areas were very large, producing a crude delineation of the urban areas. Using the most recent tracts addresses this to the greatest extent possible.

The original dataset identified the areas to be included as those within the Combined Statistical Areas (CSAs) having populations over one million in 2010. But since the urban areas include varying proportions of the CSA population and housing units, some of the resulting areas were considerably smaller than most of the others and smaller than areas not included in the dataset. The new dataset includes the largest urban areas as they are being delineated for this research.

Some extensive areas of continuous urban development should reasonably be considered to include multiple urban areas. (Urban areas in the Northeast are the prime example.) The original dataset used CSAs to identify separate areas and to establish the boundaries. The CSAs are still being used for identification of urban areas that should be considered separate, but boundaries are established where the urban areas actually have grown together, which can be somewhat different.

The delineation of urban tracts used housing units with a density threshold estimated to be equivalent to the population density standard used by the Census for Urbanized Areas in 2000 and 2010. For the 2020 Census, the Urban Area definition was

changed to use housing unit densities. The new dataset uses the threshold from the new definition.

The Census has been using land cover data on impervious surface to identify nonresidential areas that could be considered urban. The availability of land cover data aggregated to census tracts makes it reasonable to add this for the delineation of tracts in the new dataset.

The original dataset used Urbanized Areas to identify those parts of urban areas that could be considered to be urban centers that had grown together. The new dataset uses areas of urban development that were separate at some earlier time as candidates for determining which should be included as urban centers. Boundaries are established based on where these areas have grown together.

Other, smaller improvements have been made to make the process more rigorous and to more closely match the 2020 Urbanized Area definition. These are noted in the detailed documentation that follows.

Data sources

Housing units by census tract for 2020 come from the redistricting data (U.S. Census Bureau 2022a). Population is from the National Historical GIS (NHGIS) website (Manson, *et al.* 2022). Housing units and population for 2010 are from the the 2010 census tract relationship file (U.S. Census Bureau 2022b). Housing units and population for 1970 to 2000 are from the Neighborhood Change Database, a unique source with census tract data from those censuses normalized to 2000 census tract boundaries (Urban Institute and Geolytics 2003). Data using 2010 and 2000 tract boundaries are estimated for 2020 census tracts using the tract relationship files for 2010 to 2020 and for 2000 to 2010 (U.S. Census Bureau 2022b,c). Counts for the source tracts are allocated to the target year tracts using the proportions of the areas of the source tracts in the target tracts.[1]

Housing units for 1940 to 1960 are estimated using the housing-units-by-year built data from the Neighborhood Change Database, assuming the reported numbers of housing units built in each decade reflect the numbers present in the earlier years. Several prior studies have used year-built data to make estimates for prior years in this

[1] The 2000 to 2010 relationship file includes the percentage of the 2000 housing units (and population) in each part of the 2010 tracts. (The 2010 to 2020 file includes only areas and not this information.) Thus it would have been possible to use this information to assign the 2000 housing units to the 2010 tracts. This was not done for two reasons: First, it seemed somewhat inconsistent to use housing units for the allocation for the first estimation to 2010 and then to use area for the estimation for the 2020 tracts from these values. Second, while this would have been better for housing units, other data available for the 2000 tracts will have to be estimated for the 2020 tracts, and housing units would not necessarily be most appropriate for those data. Using area provides for consistent estimation for all data in the future.

manner (Radeloff, *et al.* 2001; Theobald 2001; Hammer, *et al.* 2004; Radeloff, Hammer, and Stewart 2005).

The housing unit estimates for earlier years using the year-built data are imperfect for two reasons: First, error can result from a lack of accurate knowledge of the year in which the structure was actually constructed. Second, there can be differences in the number of units reported to have been built prior to a given year and the number actually existing in that earlier year. Residential units can be demolished or converted to other uses, reducing the number, which is likely to be the greatest source of error. However, residential units can also be subdivided and nonresidential structures can be converted to residential use, increasing the number.

The amount of error is expected to be greater for estimates made for times further in the past. For this reason the estimates are made using the 1970 year-built data where available. However, not all of the United States was tracted in 1970 and 1980, so data are missing for some tracts, generally in the outer portions of the areas included. In those cases, the year-built data from 1980 or 1990 are used to make the estimates.

An analysis was undertaken to estimate the magnitude of the error introduced when using year-built data to estimate numbers of housing units for prior census years. The 2000 census year-built data for the tracts in the selected areas were used to estimate numbers of housing units for the years 1970 through 1990. The analysis was restricted to those tracts with no changes in tract boundaries over the period to eliminate any error associated with the Neighborhood Change Database normalization of earlier census data to 2000 census tract boundaries. The correlations of the actual housing unit counts with the values estimated from the 2000 year-built data were very high for 1990 and 1980, 0.98 and 0.97 respectively. The correlation dropped off to 0.91 for the 1970 estimates. The mean number of housing units estimated was lower than the actual counts, with a difference of 1.9 percent for the 1990 estimates, increasing to 12.0 percent for 1970.

So as expected, the error associated with using the year-built data to estimate numbers of housing units in prior years increases with the number of decades back for which the estimates are being made. The estimates for 10 years in the past (1990) had relatively little error, while estimates 30 years back (1970) had quite substantial error. Assuming a similar pattern would hold for estimates made using the 1970 year-built data, the 1960 estimates should be very good, 1950 not as good, and the 1940 estimates might be expected to have significant error. For this reason, the estimates of housing units for 1940 are not used for analysis.

The presence of significant area of impervious surface is used as an indicator of a tract being urban even when housing unit density is low. The National Land Cover Database provides information from classified satellite imagery (Multi-Resolution Land Characteristics (MRLC) Consortium. 2023). The NHGIS (Manson, *et al.* 2022) has

aggregated the data for census tracts, providing the areas in the 16 land cover classes. Data for 2001 and 2011, the earliest and latest years available, are used.

The geographic boundary file (shapefile) for 2020 census tracts from NHGIS is used for mapping the data. Creation of the dataset also used the shapefile for 2020 Core-Based Statistical Areas, also from NHGIS (Manson, *et al.* 2022).

Census tracts containing significant airports adjacent to urban areas are added to those urban areas. A point shapefile of all airports includes location and enplanements (U.S. Federal Aviation Administration 2022). A second polygon shapefile gives the areas of the airports and their extent, used to examine the relationships to the census tracts (Esri 2022).

The census tracts included in the Central Business Districts (CBDs) for the urban centers come from the 1982 Economic Censuses, the last year the Census reported such information (U.S. Census Bureau 1983). As these were for 1980 census tracts, the tract shapefile for that year was obtained from the NHGIS to maps these tracts (Manson, *et al.* 2022) .

Numbers of housing units in 2020 for metropolitan and micropolitan areas and estimates for Urbanized Areas from the American Community Survey for 2016 to 2020 are likewise from the NHGIS (Manson, *et al.* 2022)

Identifying the largest urban areas

The first step in the development of the dataset is the identification of the largest urban areas in 2020 to be included. As the urban areas are being delineated using housing unit densities, numbers of housing units in the urban areas seemed to be the appropriate measure of size rather than population. Obviously housing units and population tend to vary together. However, there is enough variation in the population per housing unit across the urban areas that the choice can affect the selection of the largest areas at the margin. A number of areas having lower incomes and large Latino populations had significantly higher populations per housing unit. El Paso did make the list of the largest areas included while McAllen, Texas, did not, though it would have if population had been used. In the other direction, some areas with large numbers of retirees had low populations per housing unit, obviously resulting from the numbers of one- and two-person households. For example, Sarasota-Bradenton greatly exceeded the minimum housing unit threshold used but would have been near the margin based on population.

The largest urban areas were selected based on the number of housing units in those areas as they were delineated for this dataset. As this of course could not be known until urban areas had been specified, it was necessary to delineate a greater number of areas to assure that the largest were indeed included. With the expected goal of selection perhaps 50 to 60 areas, longer lists of the largest metropolitan areas and

Urbanized Areas with respect to both housing units and population were identified and combined to produce a list of 107 areas within which urban areas would be delineated for 2020 using the procedures described in the following section.

Urban areas are identified as areas of contiguous contiguous census tracts that meet the criteria for being considered urban. This raises the issue of large areas of continuous urban development that should reasonably be considered as consisting of two or more urban areas. Urban development in the Northeast is the major example of this. The Core-Based Statistical Areas consist of groups of counties having significant commuting interchange and provide a basis for the identification of separate urban areas.

Metropolitan Statistical Areas (MSAs) are the most familiar and widely used of these areas. But the decision was made to use Combined Statistical Areas (CSAs) where they have been identified (and MSAs otherwise). The CSAs are combinations of one or more MSAs and the smaller Micropolitan Statistical Areas that meet a lower commuting interchange threshold (applied somewhat differently) than used for identifying counties for inclusion in MSAs. The CSAs were selected as it was believed that they better represent the full extent of urban areas. Three examples are discussed to make the point. First, the New York MSA only includes areas in the states of New York and New Jersey. It does not include areas in Connecticut that are connected to New York City by commuter rail, that are recognized as suburbs of New York, and that have been considered part of the New York region by the Regional Plan Association (2017) since at least the 1920s. The New York CSA extends far into Connecticut, including those areas. Raleigh and Durham, North Carolina are currently separate MSAs, though they had been in a combined MSA at the 2000 Census and earlier. They have long been recognized by residents as being part of a single area and are served by a single airport, Raleigh-Durham International. The Riverside-San Bernardino-Ontario MSA is separate from the Los Angeles MSA despite the areas being linked by three commuter rail lines and four freeways, each at least eight lanes wide and congested at rush hours. For many years the Ontario airport was literally owned and operated by the Los Angeles airport authority.

The CSAs and MSAs are only used to identify separate areas within which the urban areas evolved. Thus Philadelphia, New York, and Hartford are considered separate urban areas. The CSA boundaries are not used to draw dividing lines between the areas. Rather, the division is made at the points where the areas grew together. Tracts are assigned to the area seen as growing more rapidly towards the other and are assigned to provide more compact, less irregular boundaries.

After the 2020 urban areas had been delineated and the number of housing units in each area had been determined, a line had to be drawn to set the threshold for inclusion among the largest urban areas. Selecting a minimum size threshold is of course arbitrary. To at least make the choice not dependent on any biases with respect to

particular urban areas, a round number of housing units was to be used as the threshold. A minimum of 400,000 housing units yielded 48 areas, and somewhat more were desired. Going down to 300,000 gave 56 urban areas. El Paso, Rochester, and Birmingham were the smallest areas included. Albany-Schenectady-Troy, Dayton, and Harrisburg-York were the next three, below the line. Drawing the line at 250,000 gave 64 areas, including Baton Rouge, Colorado Springs, and Columbia, South Carolina. These were not areas I would consider to be among the largest urban areas in the United States. The decision was made to include those 56 urban areas having more than 300,000 housing units in 2020. Since population is the more familiar metric for urban size, the smallest urban area in that respect was Birmingham, with a population of 668,000. It was joined by Rochester as the only other area with a population less than 700,000, and all but five areas had populations exceeding 800,000.

Delineating the urban areas

The major task in developing the dataset is the delineation of the urban areas, the identification of the census tracts included in the urban areas for each census year from 2020 back to 1950. In any year, a tract could be assigned to the urban area only if it had been considered urban in the following year. The goal was to have a set of urban areas that represented the cumulative growth of the urban areas over time. To have a tract considered urban in one year and not urban in a later year would not have been consistent and would have created problems for some of the analyses contemplated. The rule has been imposed in this direction—if not urban in any year, then not urban earlier—rather than in the opposite direction—if urban, then urban later—because the more recent data are considered to be more accurate.

The delineation of the urban areas follows as closely as possible the standards used for by the Census for defining Urban Areas for 2020 (U.S. Census Bureau 2022d). Modifications are made to reflect the use of census tracts rather than census blocks for the creation of the urban areas, the availability of data, and practical considerations.

Tracts are considered to be urban if they exceed minimum thresholds for either housing unit density or impervious surface. Density has been the primary standard for defining Urbanized Areas, though this has mainly been population density. The 2020 Urban Area definition uses housing unit density, applying three different thresholds. The highest level is used for identification of an urban nucleus. This is generally relevant only for very low-density development in more rural areas and will not be an issue for the large urban areas. The next level of 425 housing units per square mile includes contiguous blocks in in what is termed the "initial urban core." This is the starting point for adding additional noncontiguous urban blocks along roads extending from the core via what are termed hops and jumps. After this step has been completed, tracts meeting a lower "fill" density threshold of 200 housing units per square mile that

are contiguous are added. Because the current dataset is using the much larger census tracts rather than blocks, it does not make sense to generally allow hops and jumps. Therefore the "fill" density of 200 units per square mile is used as the minimum threshold for tracts to be considered urban and eligible for inclusion in an urban area.

It is useful to consider how the Census may have arrived at these housing unit density thresholds. In the proposed urban area criteria for 2020, a single density minimum of 385 housing units per square mile was to be used (U.S. Bureau of the Census 2021). This value was chosen as being equivalent to the population density threshold of 1,000 persons per square mile using the value of 2.6 persons per household for the nation. (Why the Census chose equivalence with the level of 1,000 persons per square mile used for Urbanized Areas prior to 2000 rather than the value of 500 used for 2000 and 2010 was not explained.) They also proposed that when areas were added via hops and jumps, the lower-density intervening territory would no longer be added to keep the areas contiguous as had been done in previous years. The comments on the proposed criteria included significant criticism of the decision to produce urban areas consisting of noncontiguous parts. In the final criteria, the Census did not back down from not arbitrarily including all intervening low-density areas for hops and jumps but greatly reduced the number of the gaps by establishing the lower "fill" density for adding blocks after the hops and jumps were completed (U.S. Census Bureau 2022d). They also inexplicably raised the threshold for including tracts earlier from the proposed value of 385 to 425 housing units per square mile. It is interesting to consider the selection 200 as the "fill" density threshold. The original proposed density of 385 persons per square mile was offered as being equivalent to a population density of 1,000 persons per square mile. Half of that, 192.5, would then be equivalent to 500 persons per square mile, the standard used in 2000 and 2010. And 200 housing units per square mile is a nice round number that is very close to that. The original urban patterns dataset used a standard of 213.3 units per square mile, calculated to be equivalent to the 500 persons per square mile using the standard of population per housing units for the nation rather than persons per household used in the Census proposal (reasonable arguments can be made for either choice). The final level of 200 housing units per square mile is right between the values of 192.5 and 213.3.[2]

---

[2] In proposing to use housing unit density rather than population density to define urban areas, the Census said, "Housing unit density provides a more direct measure of the densely developed landscape than population density." (U.S. Census Bureau 2021) This is not dissimilar from one of the reasons I have offered for using housing unit density: "Housing units better represent the physical pattern of urban development as they are relatively fixed, while the population of an area can change without any changes in the stock of housing." (Ottensmann 2015) Other researchers have made similar arguments for using housing units rather than population, for example Galster, *et al.* (2001), Theobald (2001), Radeloff, Hammer, and Stewart (2005); and Paulsen (2014).

The Census urban area criteria also provide for the inclusion of blocks having primarily nonresidential development that do not meet the housing unit density threshold but that reasonably can be considered to be urban. A block is considered to be urban if at least one-third of the area of the block consists of areas having least 20 percent impervious surface using the National Land Cover Database. Exactly the same standard is used for census tracts. The land cover data aggregated to the census tracts includes the area of the tract in three classes of developed land cover of low, medium, and high intensity, each having minimum impervious surface of 20 percent or more. The proportion of such land area in tracts is calculated and tracts having over 33.33 percent impervious surface over 20 percent are considered to be urban. The land cover data aggregated to tracts are available for 2001 and 2011. The 2011 impervious surface data are used for identifying urban areas for both 2010 and 2020. (Unfortunately later data are not available). And the 2001 land cover data are used for the 2000 urban area delineation.

This raises the issue of what to do for the identification of urban areas for earlier years. Obviously tracts with sufficient impervious surface beyond the edge of the urban area as delineated using housing unit density cannot be assumed to have been developed and urban in earlier years. Some may have been but many are likely to have been developed as the urban area expanded over time. However, it may not be unreasonable to assume that tracts embedded within the delineated urban area in earlier years had been developed around the same time as the adjacent residential areas that have qualified as urban. The rule will be established as follows: Using the 2001 impervious surface data, if a tract meeting the standard is within an area completely surrounded by the tracts qualifying as urban using housing unit density, the tract is also considered to be urban and is added to the urban area. This still leaves impervious surface tracts out at the edge that are almost completely surrounded by urban tracts. These are also likely to have developed around the same time as those tracts. A stringent standard is used that if 85 percent of the perimeter of the impervious surface tract or tracts is contiguous to the housing-unit-density urban tracts, the tract will be considered to be urban.

Tracts are designated as part of the urban area if they meet the urban criteria for either housing unit density or impervious surface and are contiguous to the urban area. Tracts are considered to be contiguous if they share a common vertex—queen contiguity —the same standard used by the Census for Urban Areas. And while the hops and jumps are generally not being used to qualify tracts, one of the criteria for those additions is relevant. Hops and jumps must be separated by road distances less than certain maxima. But road distance along what is termed “exempted territory” where urban development is not possible is not included in the distance. Water and wetlands are considered to be exempted territory. Therefore tracts separated by a bridge over water are separated by zero distance and are considered to be contiguous.

A suggestion was made in response to the proposal for the 2000 urban area criteria that ferries be considered in addition to roads for establishing links to noncontiguous areas for hops and jumps (U.S. Bureau of the Census 2002). The Census demurred, saying there was no consistent database of ferry connections. For the large urban areas being considered here, ferry service is especially important for the Seattle-Tacoma area, connecting the main urban area to areas across Puget Sound that are considered part of the CSA based on commuting interchange. So those tracts linked to Seattle and Tacoma by ferry are considered contiguous.

Wetlands are relevant for one connection in the New Orleans area. A road extends from the main portion of the New Orleans urban area across the Bayou Sauvage National Wildlife Refuge, wetlands, and over the eastern portion of Lake Ponchatrain, connecting to tracts in the vicinity of Slidell. The first tract is considered to be contiguous to the New Orleans urban area.

A few cases of connections between urban tracts over both water and a tract that is not urban warrant considering the urban tracts to be contiguous. This would be equivalent to adding to the urban area via the hops and jumps used in the Census urban area definition. Four of the cases are along major rivers in four urban areas—Memphis, Kansas City, Sacramento, and Portland. In most areas, the regular pattern of tracts extends to the river and if they meet the urban criteria they are included in the urban area. If the tract is connected by a bridge to a tract on the other side of the river, that tract will then be considered to be contiguous to the urban area. But in the areas noted, elongated tracts have been delineated along a river and those tracts include the river and some land, presumably low-lying, along the river. The tracts may be undeveloped floodplain or may include some nonresidential development qualifying as urban based on impervious surface in the later years, but the tracts do not have substantial residential development such that they qualify as urban based on housing unit density. Tracts connected by road across these tracts and the river are considered to be contiguous. The fifth case is in the San Francisco-Oakland-San Jose urban area. Marin County is north of San Francisco across the Golden Gate Bridge and includes areas of urban development that should reasonably by considered part of the larger urban area. From the north end of the bridge, the road extends about a mile through more rugged territory that is not urban before coming to the first urban tract. This tract will be considered to be contiguous to the urban tracts in San Francisco at the south end of the Golden Gate Bridge. (Note that this provision is actually only significant for the 1950 urban area. In the later years, the bridge across the northern part of San Francisco Bay connecting Marin County with Richmond in the East Bay ended in tracts that were urban on both ends, making these areas contiguous under the standard criterion. Only in 1950 did the tracts on both ends drop out of the urban area, making the Golden Gate connection necessary to maintain contiguity of Marin urban development with the remainder of the urban area.)

Special treatment is given to barrier island for some urban areas in Florida. Various barrier islands are connected to the mainland by bridges, of course. And strings of barrier island are connected by roadways and bridges as well, making them contiguous using the standard criteria. This produces several instances where census tracts on the barrier islands meet the density criterion and could be added to the urban area creating a narrow portion of the urban area extending well beyond the area considered to be urban on the mainland. This can become even more anomalous when that barrier island is then connected back to the mainland making a tract meeting the density threshold contiguous and part of the urban area via the link through the barrier islands even though it is not contiguous to the rest of the urban area on the mainland. To address these issues, tracts on barrier islands are included in the urban area only when they not only meet the standard conditions but lie offshore from the the urban area on the mainland.

The Raleigh-Durham urban area includes another unique situation. Research Triangle Park was developed between those cities in 1959 as a site for high-tech research and development facilities. (The Triangle refers to the presence of major research universities in the cities of Raleigh, Durham, and Chapel Hill.) Two census tracts are coincident with the development. It is a very large site, much of it wooded, with very low-intensity development. The tracts have little or no residential development, and the spread-out nature of the research facilities means that they also fail to meet the impervious surface standard. However, these tracts obviously should be considered urban and are included in the urban area when contiguous.

After urban areas have been delineated using the housing unit density and impervious surface criteria, areas of nonurban land can be included within the urban areas. Following the approach taken by the Census in defining Urban Areas, those enclaves totally surrounded by the urban area are added if they have an area less than five square miles. Areas surrounded by the urban area and water are likewise added if the areas is less than five square miles and if over half of the boundary is shared with the urban area.

The Census also includes as urban the blocks approximating airports that fall within one-half mile of the urban area. Because of the use of the larger census tracts, the treatment of airports must be somewhat different. Airports are identified for possible inclusion if they have had over 2,500 enplanements in a year, the same standard used by the Census. While a few tracts have been drawn to generally coincide with airports, many tracts are much larger than the airports, while in some instances the airports extend over more than one tract. If the area of an airport is greater than 50 percent of the area of the tract in which is it primarily located, and if that tract is contiguous to the urban area, the tract is added to the urban area. This is done as the last step for each year. No additional tracts can be added that have become contiguous due to the addition of the airport tract.

Identifying urban centers

Some urban areas encompass the area of two or more separate urban areas that have grown together. Dallas and Fort Worth are an obvious example. This raises the issue of how such multiple urban centers are identified and how tracts are assigned to the centers after they have coalesced. As the areas of urban development of urban centers become contiguous, tracts are assigned to one of the areas in a manner similar to that used is assigning tracts to urban areas that were contiguous. Tracts are assigned to the area seen as growing more rapidly towards the other and are assigned to provide more compact, less irregular boundaries.

An area is considered to be an urban center if the number of housing units in the tracts assigned to that center in 2020 exceeds a minimum percentage of all housing units in the urban area. As urban areas expand, they encounter existing urban areas of varying sizes that are incorporated into the urban area. Very small areas should certainly not be considered urban centers. When they are not contiguous with the main urban area, they are no longer considered parts of the area. At the other extreme, if two areas are both reasonably large when separate, such as Dallas and Fort Worth, they can be assumed to be separate urban centers. The development of both areas should be identified going back to the earliest year and the tracts becoming urban in succeeding years need to be assigned to the areas.

The issue is at the margin. One cannot know whether the number of housing units in an area exceeds some percentage of the total units in the urban area in 2020 without having treated it as a center, delineating its extent from the earlier years, and assigning the tracts to the centers after the areas have become contiguous. The solution is to identify candidates for urban centers whenever the area that became separate seems to have any possibility, however remote, of being considered an urban center at the end. To accomplish this, 52 urban center candidates were identified in 26 urban areas. After the decision was made on the areas to include as second or third centers, 20 areas were added in 15 urban areas.

In most cases, the assignment of the tracts to the candidate areas in the years after the centers merged was relatively straightforward. Judgement is obviously involved in applying the rate of growth and compact area criteria. But I expect that most doing so would come up with similar results with differences that would not substantially affect the outcome. Assignment of tracts in three of the areas was less clear.

The easiest case was the Minneapolis-St Paul urban area. The area obviously has two centers. But they had grown together long before the period for which data are being used to delineate urban areas. In this case, tracts were assigned to the center with the closer Central Business District (CBD). CBD identification and distance calculation are discussed in the next section.

The Norfolk-Virginia Beach urban area has those two centers but their growth was more unusual. Norfolk was a fairly large urban area from the start in 1950. As it grew, much of that growth occurred to the south and west, with limited expansion towards Virginia Beach to the east. Virginia Beach started out as a small beach community on the Atlantic coast, over 10 miles from Norfolk. But between 1950 and 1960, a strip of urban tracts extended west from the 1950 Virginia Beach urban area about halfway to Norfolk and then expanded into a much larger urban area contiguous to Norfolk. Nothing in that pattern provided a clue as to where a dividing line might be drawn and which tracts should be considered part of the Norfolk urban area versus the Virginia Beach area. Other information was required. Virginia Beach was established as an independent city in 1952 and merged with the the remainder of the county in which it had been located in 1963. It includes essentially all of the area newly urban in 1960, up to the Norfolk urban area. In addition, the location of what is currently considered the Virginia Beach CBD is within that larger area added as urban in 1960 adjacent to Norfolk. This makes it reasonable to consider all of these tracts added from 1950 to 1960 as belonging to the Virginia Beach urban center. Then as development continued in later years, the standard growth and compact area criteria could be applied.

San Francisco-Oakland-San Jose was the remaining difficult area. The Oakland and San Jose urban areas became contiguous in 1970 and establishing a boundary was not especially problematic. The San Francisco and San Jose urban areas were contiguous in 1960. They first were separate in 1950, but in a very unusual way. A long stretch of urban development extended down the peninsula from around San Mateo to Mountain View, separated from both the San Francisco and San Jose urban areas. To gain some insight, the areas urban in 1940 were delineated. This did not provide much clarity, with that area in the middle fragmenting into five separate areas of urban development. Two of the larger clusters of urban development were around Redwood City and Palo Alto-Menlo Park in the middle, separated from each other by several larger tracts. Again, some other information was needed. The line between the San Francisco and San Jose MSAs and Urbanized Areas follows the county boundary between Palo Alto and Menlo Park and is more-or-less in the middle between San Francisco and San Jose. Santa Clara county, which includes San Jose, certainly is the heart of the Silicon Valley and extends up to Palo Alto and Stanford University. But Menlo Park, which is to the north of Palo Alto, is the location of Facebook/Meta and many of the venture capitalists and should be seen as part of Silicon Valley as well. Therefore drawing the division north of Menlo Park, including that in with the area for the San Jose urban center, makes sense. The nonurban tracts separating the Palo Alto-Menlo Park urban area from the Redwood City area include Facebook headquarters and therefore are logically included with the San Jose center as well.

Having identified the 52 urban center candidates and determining the urban tracts associated with them in each year, the final task is setting the cutoff for candidates

to be included as urban centers. The number of housing units in 2020 was determined for the tracts associated with each of the center candidates and the percentage of the total housing units in the urban area was calculated. Providence, with 21.3 percent of the housing units in the Boston urban area should certainly be considered to be an urban center. The next lower areas was Tacoma with 17.3 percent of the Seattle urban area and High Point with 16.0 percent of the Greensboro—Winston-Salem area (Winston-Salem having already been identified as the second center. Both had long been considered central cities of their metropolitan areas (when those were still being identified as such). And the names of the major airports serving the areas suggest inclusion: Seattle-Tacoma International and Piedmont Triad International, referring to the triad of Greensboro, Winston-Salem, and High Point. The next two candidates at about 11 percent are Port Charlotte in the Sarasota-Bradenton area and Winter Haven in the Orlando area. Neither would seem to necessarily require inclusion as additional urban centers, nor do any of the remaining candidates with even smaller percentages.

The final names given to the urban areas include the names of the additional urban centers that have been added.

Identifying Central Business Districts and adding distances

Distances from each of the urban area tracts to the Central Business District (CBD) of its urban center will be used in some analyses and are added to the dataset. The last time the Census identified CBDs was for the 1982 economic censuses. Many researchers have continued to use this information as the best available for designating CBDs.

The Census report lists the census tract or tracts constituting the CBD for a large number of cities (U.S. Census Bureau 1983). To identify point locations for the CBDs, the tracts were identified in a shapefile of the 1980 census tracts. The report did not list CBD tracts for five urban centers—Indianapolis, Jacksonville, Nashville, Naples, and Virginia Beach. Identifying CBD census tracts for the first four was relatively straightforward. Google Map searches for the downtowns of those areas indicated the area. The reasonableness of the location and the extent of the CBDs was determined by looking at the satellite imagery and the road network, as major roads typically converge on CBDs.

The Virginia Beach CBD presented more of a problem. Google Maps does not provide a location for Virginia Beach downtown or central business district. The original CBD would have been at the small settlement on the shore of the Atlantic. The City of Virginia Beach identifies the Pembroke Strategic Growth Area in the center of the current Virginia Beach urban area as encompassing the CBD and having the largest concentration of office and retail space in the city. It also links to the private Central Business District Association, which represents businesses in the same area. It is reasonable to consider this as the contemporary Virginia Beach CBD.

The census tracts constituting each CBD were dissolved into a single geographic feature and the centroid of the features was taken as the point location for the CBD. The centroids of the 2020 census tracts were used as their point locations as well. This is not always the same as the internal points for the tracts provided by the Census. For highly irregular tracts, the centroid can be located outside the boundaries of the tract. In those cases, the Census moves this to a location within the tract. For purposes of calculating distances to the center, the centroids, even if outside the tract boundaries, are better for calculating the average distance to the CBD from the various parts of a tract.

Distances were calculated directly using the projected map coordinates in meters for the CBD point and census tract centroids. (The map used the U.S. Albers Equal Area Conic projection.) Checks showed that error introduced by calculating distances in this way rather than the more elaborate procedure of calculating distances on the surface of the ellipsoid were small compared to other sources of error such as uncertainties in the point locations of the CBDs, especially since these were based on 40-year-old data provided by local sources.

Future research

Used alone, the new dataset can be used to investigate a wide range of interesting questions associated with the development of the large urban areas over time. Four such studies are planned initially. The first will address the growth of these urban areas and the changes in the distribution of urban area sizes that have occurred over the 70-year period. Changes in the densities of the urban areas comes next, with special emphasis on the effects of barriers to development on densities and change. This is extended to examine densities and change in the older, inner urban cores and the more recently developed urban peripheries, with implications for the consideration of urban sprawl. Because the negative exponential model of density decline with distance from the center has been widely used to describe urban patterns, its performance over time is the topic for the fourth paper.

Additional issues can be addressed using this dataset. Some examples of questions to be addressed: To what extent are housing units centralized, clustered more closely to the center in some areas and how has this changed over time? Weighted density has been proposed as an alternative to the usual measure of density. What are the implications? Areas are not developed all at once. How do densities of areas near the periphery change over time, both before and after they become part of the urban area?

The addition of other data at the census tract level allow more topics to be examined. Just two possibilities using data from the decennial census and the American Community Survey: Data on multifamily housing gives further insight into the nature

and evolution of the housing stock. Information on the race and ethnicity of the population allows detailed examinations of diversity and change.

The Longitudinal Employer-Household Dynamics program provides information on the location of jobs at the *block* level for every year since 2002. These data can be aggregated to the tracts, providing the important addition of the location of economic activity to study of patterns in the largest urban areas.

References

Esri. 2022. *U.S. Airport Areas*. Downloaded from *https://www.arcgis.com/home/item.html?id=2706fbe2d7c74b488a609938df8f9578#overview* on October 30, 2022.

Galster, George, and five others. 2001. Wrestling sprawl to the ground: defining and measuring an elusive concept. *Housing Policy Debate* 12, 4: 681-717.

Hammer, Roger B., and four others. 2004. Characterizing dynamic spatial and temporal residential density patterns from 1940-1990 across the North Central United States. *Landscape and Urban Planning* 69, 2-3: 183-199.

Manson, Steven, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. 2022. *IPUMS National Historical Geographic Information System: Version 17.0* [dataset]. Minneapolis MN: IPUMS. *http://doi.org/10.18128/D050.V17.0.* Data downloaded from *https://www.nhgis.org* on various dates, August to October 2022.

Multi-Resolution Land Characteristics (MRLC) Consortium. 2023. *National Land Cover Databas*e. Accessed at *https://www.mrlc.gov* on January 8, 2023.

Ottensmann, John R. 2014. Urban patterns dataset description. Available at *https://urbanpatternsblog.files.wordpress.com/2016/12/urban-patterns-dataset-description.pdf*.

Ottensmann, John R. 2015. Density of large urban areas in the U.S., 1950-2010. Available at *https://urbanpatternsblog.files.wordpress.com/2016/12/density-of-large-urban-areas.pdf*.

Paulsen, Kurt. 2014. Geography, policy or market? New evidence on the measurement and causes of sprawl (and infill) in US metropolitan regions. *Urban Studies* 41, 12: 2629-2645.

Radeloff, Volker C., and five others. 2001. Human demographic trends and landscape level forest management in the northwest Wisconsin Pine Barrens. *Forest Science* 47, 2: 229-241.

Radeloff, Volker C., Roger B. Hammer, and Susan L. Stewart. 2005. Rural and suburban sprawl in the U.S. midwest from 1940 to 2000 and its relation to forest fragmentation. *Conservation Biology* 19, 3: 793-805.

Regional Plan Association. 2017. *Regional Plan Association*. *History*. Website accessed at *http://www.rpa.org/about/history* on March 3, 2013.

Theobald, David M. 2001. Land-use dynamics beyond the urban fringe. *Geographical Review* 91, 3: 544-564.

U.S. Census Bureau. 1983. *1982 Economic Censuses. Reference Series. Geographic Reference Manual.* EC82-R-1. U.S. Government Printing Office.

U.S. Census Bureau. 2002. Urban area criteria for the 2000 Census. *Federal Register* 67, 51 (Friday, March 15): 11663-11670.

U.S. Census Bureau. 2021. Urban Area criteria for the 2020 Census—proposed criteria. *Federal Register* 86, 32 (February 19): 10237-10243.

U.S. Census Bureau. 2022a. *2020 Census Redistricting Data* (Public Law 94-171). Downloaded from *https://data.census.gov* on October 28, 2022.
U.S. Census Bureau. 2022b. *2010 Census Tract Relationship File*. Downloaded from *https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.2010.html* on September 13, 2022.
U.S. Census Bureau. 2022c. *2020 Census Tract to 2010 Census Tract Relationship File*. Downloaded from *https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.html* on September 13, 2022.
U.S. Census Bureau. 2022d. Urban Area criteria for the 2020 Census—final criteria. *Federal Register* 97, 57 (March 24): 16706-16715.
U.S. Federal Aviation Administration. 2022. *USA Airports*. *National Airspace System Resource Aeronautical Data Product*. Downloaded from *https://hub.arcgis.com/maps/5d93352406744d658d9c1f43f12b560c/about* on September 15, 2022.
Urban Institute and Geolytics. 2003. *Census CD Neighborhood Change Database (NCDB): 1970-2000 Tract Data*. Geolytics.

Appendix: List of Variables

GISJOIN	Tract identifier in NHGIS format
geoid	Tract identifier—state, county, and tract FIPS codes
geoidx	Tract identifier with "x" appended for import as text in some software
areanum	Urban area number
areaname	Urban area name
centernum	Urban center number (1-3)
cbdmiles	Distance to CBD in miles
areasqmi	Land area in square miles
imperv8	Percent land area impervious surface greater than 20 percent in 2011
imperv7	Percent land area impervious surface greater than 20 percent in 2001
pop9	Population in 2020
hu9	Housing units in 2020
huden9	Housing unit density in 2020
urbarea9	Urban area in 2020
pop8	Population in 2010
hu8	Housing units in 2010
huden8	Housing unit density in 2010
urbarea8	Urban area in 2010
pop7	Population in 2000

hu7	Housing units in 2000
huden7	Housing unit density in 2000
urbarea7	Urban area in 2000
pop6	Population in 1990
hu6	Housing units in 1990
huden6	Housing unit density in 1990
urbarea6	Urban area in 1990
pop5	Population in 1980
hu5	Housing units in 1980
huden5	Housing unit density in 1980
urbarea5	Urban area in 1980
pop4	Population in 1970
hu4	Housing units in 1970
huden4	Housing unit density in 1970
urbarea4	Urban area in 1970
hu3	Housing units in 1960
huden3	Housing unit density in 1960
urbarea3	Urban area in 1960
hu2	Housing units in 1950
huden2	Housing unit density in 1950
urbarea2	Urban area in 1950
urbancum	Tract added to urban area in year 2 to 9, 1950 to 2020
missing5	Missing data in 1980
missing4	Missing data in 1970