National Equity Atlas Data & Methods: Technical Documentation

Prepared by PolicyLink and the USC Program for Environmental and Regional Equity

Last updated: June 22, 2017.

This document provides more detailed information about the indicators and methods used to produce the data in the National Equity Atlas. Since the Atlas is a living resource to which we will continue to add new indicators regularly, this document will also be updated on a regular basis. If you have additional questions about the data or methods, please contact Justin Scoggins, Data Manager at PERE, at info@nationalequityatlas.com.

Contents

Data source summary and regional geography

The National Equity Atlas draws upon a regional equity indicators database assembled using a broad array of data sources and methodologies. Unless otherwise noted, all of the data presented on this website are based on analysis by PolicyLink and the USC Program for Environmental and Regional Equity (PERE). The database contains information for 301 geographies including the United States as a whole, all 50 states, the 150 largest metropolitan areas (based on 2010 population), and the 100 largest cities. Metropolitan areas are defined based on the U.S. Office of Management and Budget's (OMB's)

December 2003 Core Based Statistical Area (CBSA) definitions while cities are defined to follow the designation of incorporated places used in the 2010 Census. The only the exception to this rule is the City and County of Honolulu, which is defined to include Honolulu County in its entirety to be consistent with the jurisdiction of the consolidated city-county government of Honolulu. Unless otherwise noted, data reported for all years have been assembled to follow these geographic definitions, and are thus geographically consistent over time.

While specific data sources and notes accompany each indicator displayed in the Atlas, here we provide more detail on the methods used in developing the indicators, adjustments that were made to some of the underlying datasets, and clarification on some of the key terms that are used. For each section below describing analyses and adjustments performed to develop the indicators and underlying database, we include a list of the particular indicators for which they are relevant.

The user should bear in mind that many of the analytical choices in generating the underlying regional equity indicators database were made with an eye toward replicating the analyses in multiple regions and the ability to update them over time. Thus, while more regionally specific and/or recent data may be available for some indicators, the data in the Atlas are drawn from our regional equity indicators database, which provides data that are comparable and replicable over time.

The specific data sources are listed in Table 1 below.

Table 1. Data Sources

Source	Dataset
Integrated Public Use Microdata Series (IPUMS)[1]	1980 5% State Sample 1990 5% Sample 2000 5% Sample 2014 American Community Survey (ACS), 5-year microdata sample
U.S. Census Bureau, accessed via the National Historic Geographic Information System (NHGIS)[2]	1980 Summary Tape File 1 (STF1) 1980 Summary Tape File 2 (STF2) 1980 Summary Tape File 3 (STF3) 1990 Summary Tape File 1 (STF1) 1990 Summary Tape File 2A (STF2A) 1990 Summary Tape File 3 (STF3) 1990 Modified Age/Race, Sex and Hispanic Origin File (MARS) 2000 Summary File 1 (SF1) 2010 Summary File 1 (SF1)
U.S. Census Bureau	2015 Population Estimates 2015 ACS 1-year Summary File 2014 ACS 5-year Summary File 2014 National Population Projections 2010 TIGER/Line Shapefiles, 2010 Counties/States/Metro Areas
U.S. Bureau of Economic Analysis	Gross Domestic Product by State Gross Domestic Product by Metropolitan Area Local Area Personal Income Accounts, CA30: regional economic profile
U.S. Bureau of Labor Statistics	Quarterly Census of Employment and Wages (QCEW)
Woods & Poole Economics, Inc.	2016 Complete Economic and Demographic Data Source
Geolytics, Inc.	1980 Census in 2010 Boundaries 1990 Long Form in 2010 Boundaries 2000 Long Form in 2010 Boundaries

[1] Steven Ruggles, Katie Genadek, Ronald Goeken, Josiah Grover, and Matthew Sobek. Integrated Public Use Microdata Series: Version 6.0 [Machine-readable database]. Minneapolis: University of Minnesota, 2015. https://usa.ipums.org/usa/

[2] Minnesota Population Center. National Historical Geographic Information System: Version 2.0. Minneapolis, MN: University of Minnesota 2011. https://www.nhgis.org/

Table 1 (Continued).

Centers for Disease Control and Prevention	Behavioral Risk Factor Surveillance System (BRFSS)
Georgetown University Center on Education and the Workforce	Updated projections of education requirements of jobs in 2020, originally appearing in: Recovery: Job Growth And Education Requirements Through 2020; State Report
National Center for Education Statistics (NCES) Common Core of Data (CCD) Public Elementary/ Secondary School Universe Survey	School Year 1999-00 School Year 2009-10 School Year 2010-11 School Year 2011-12 School Year 2012-13 School Year 2013-14
U.S. Environmental Protection Agency	2011 National-Scale Air Toxics Assessment (NATA)

Selected terms and general notes

Broad racial/ethnic origin

In the Atlas, categorization of people by race/ethnicity is generally based on individual responses to various census surveys. Unless otherwise noted, people are categorized into six mutually exclusive groups based on their response to two separate questions on race and Hispanic origin as follows:

- "White" and "non-Hispanic White" are used to refer to all people who identify as White alone and do not identify as being of Hispanic origin.
- "Black" and "African American" are used to refer to all people who identify as Black or African American alone and do not identify as being of Hispanic origin.
- "Latino" is used to refer to all people who identify as being of Hispanic origin, regardless of racial identification.
- "Asian or Pacific Islander" and "API" are used to refer to all people who identify as Asian American, Native Hawaiian, or Pacific Islander alone and do not identify as being of Hispanic origin.
- "Native American" is used to refer to all people who identify as Native American or Alaskan Native alone and do not identify as being of Hispanic origin.
- "Other" or "Mixed/other" are used to refer to all people who identify with a single racial category not included above, or who identify with multiple racial categories, and do not identify as being of Hispanic origin. Importantly, prior to the 2000 Census the questionnaire did not allow for multiple responses to the race question, causing some degree of inconsistency in data for this racial/ethnic category before and after 2000.
- "People of color" is used to refer to all people who do not identify as non-Hispanic White.

Ancestry

In the Atlas, some indicators based on the IPUMS microdata provide data for people grouped by ancestry, with both broad and detailed ancestral groups identified. The purpose of the data by ancestry is to provide further information on equity indicators and population diversity for distinct subgroups *within* each of the mutually exclusive broad racial/ethnic groups described above (except for the Mixed/other group).[3] For this reason, the ancestral groupings were defined based on examining each broad racial/ethnic group separately, and selecting the ancestries within each group that captured a reasonably large number of people identified nationwide. Therefore, some ancestral groups are included in more than one broad racial/ethnic group. For example, data for those of Panamanian ancestry is broken out and reported on for both the Black and Latino populations, while data for those of Irish ancestry is available for both the Black and White populations (subject to sample size limitations in the IPUMS data).

The ancestral groups broken out for each broad racial/ethnic group other than Native Americans are based on the first response to the census question on ancestry, recorded in the IPUMS variable "ANCESTR1." For Native Americans, they are based on the detailed responses to the census question on race, recorded in the IPUMS variable "RACED." The reason for this that the vast majority of responses for Native Americans to the ancestry question (about 75 percent in most years) are coded in the ANCESTR1 variable as simply "American Indian (all tribes)" while the responses reflected in the RACED variable identify a variety of detailed Native American tribes.

For each broad racial/ethnic group, the responses to ANCESTR1 (or RACED, for Native Americans) were examined and the most common responses were identified nationwide. These became the detailed ancestral groups reported in the Atlas, and tend to reflect specific countries or ethnicities (or tribes, for Native Americans). The detailed ancestral groups were then organized into broader groups based on geography. These broader ancestral groups are reported in the Atlas. They tend to reflect continents of the world (or U.S. regions, for Native Americans), and were defined to include, in addition to their constituent detailed ancestries, all other relevant detailed ancestries not listed on the Atlas. Finally, a residual "Other" broad ancestral category was defined for each race/ethnicity to include all people for which the response to the ancestry (or race) question was too ambiguous to assign to any of the broad ancestries identified.[4] Thus, while the detailed ancestral groups only account for a portion of each racial/ethnic population, the broad ancestral groups (including the residual "Other" broad ancestral category that appears for each race/ethnicity) account for the entire racial/ethnic population. Finally, in an effort to strike some balance between the numbers of detailed ancestries reported within each broad ancestral group, only the eight detailed ancestral groups with the largest populations in in the United States were broken out within each broad group.

As an example of how this classification scheme works, consider the Asian or Pacific Islander population. Within this broad racial/ethnic group, Vietnamese is a detailed ancestral group within the broader Southeast Asian ancestral group. The Southeast Asian ancestral group also includes those of Malaysian ancestry, even though they are not included among the detailed groups that are broken out.

---

[3] We did not disaggregate the Mixed/other racial/ethnic group by ancestry mainly because, for this population, a more appropriate/interesting disaggregation would be by the various racial/ethnic groups people identify with rather than by ancestry—and that would be inconsistent with the way we disaggregate the other five major racial/ethnic groups.

[4] For the Black population, this residual "Other" broad category is mostly comprised of people who identified as being of "African-American" or "Afro-American" ancestry, and thus the category is labelled "African American/Other Black."

Both detailed and broad sets of ancestral groups were created within each broad race/ethnicity so we could maximize the amount of racial/ethnic subgroup data reported in the Atlas. For example, because we do not report data for Atlas indicators and geographies that are based upon a small underlying sample size (see "A note on sample size," below), many regions would have very little or no data by ancestry reported on the Atlas if we were to only designate the detailed ancestral groups. Our strategy to break out both broad and detailed ancestral groups means that Atlas indicator data for the broader ancestral groups can often still be reported for geographies where the detailed ancestral group populations are not large enough to report on; in geographies with large and diverse populations, both broad and detailed data are likely to be reported.

While most of the broader geographic groupings should be intuitive and are largely based on the IPUMS documentation for the ANCESTR1 variable, the way we defined those for Native Americans, and for the Middle East/North African group that is broken out for the White population, deserve a bit of explanation. Due to the lack of a generally accepted geographic categorization of Native American tribes in the United States, we relied upon a variety of maps of Native American regions found on the internet and applied what appeared to be the most common broad geographic groupings (e.g., Northwest Coast, Great Plains, Eastern Woodlands) into which the detailed Native American ancestries were assigned. The latter assignment was guided by maps of Native American tribes in the continental United States created by Aaron Carapella, among other online sources.[5] To define the Middle East/North African broad ancestral group that is broken out for the White population, we included all ancestries classified under "North African and Southwest Asia" in the IPUMS documentation for the ANCESTR1 variable.

Nativity

The term "U.S.-born" refers to all people who identify as being born in the United States (including U.S. territories and outlying areas), or born abroad of at least one U.S. citizen parent. The term "immigrant" refers to all people who identify as being born abroad, outside of the United States, of non-U.S. citizen parents.

Other selected terms

Below we provide definitions and clarification around some of the terms used in the Atlas.

- The term "**communities of color**" generally refers to distinct groups defined by race/ethnicity among people of color.
- The term "**full-time**" workers refers to all persons who reported working at least 45 or 50 weeks (depending on the year of the data) and usually worked at least 35 hours per week during the year prior to the survey. A change in the "weeks worked" question in the 2008 American Community Survey (ACS) caused a dramatic rise in the share of respondents indicating that they worked at least 50 weeks during the year prior to the survey, as compared with prior years of the ACS and the long form of the decennial census. To make our data on full-time workers more comparable over time, we applied a slightly different definition in 2008 and later than in earlier years: in 2008 and later, the cutoff applied to identify full-time workers is at least 50 weeks while in 2007 and earlier it is 45 weeks per year. The 45-week cutoff was found to produce a national trend in the incidence of full-time work over the 2005–2010 period that was most

[5] The maps created by Aaron Carapella can be viewed, and purchased, here: http://www.tribalnationsmaps.com/.

consistent with that found using data from the March Supplement of the Current Population Survey, which did not experience a change to the relevant survey questions. For more information, see http://www.census.gov/acs/www/Downloads/methodology/content_test/P6b_Weeks_Worked_Final_Report.pdf.

- The terms "**region**," "**metropolitan area**," "**metro area**," and "**metro**" all refer to the geographic areas defined as metropolitan statistical areas by the OMB under the December 2003 definitions.

- The term "**housing unit**" refers to the underlying physical sampling unit for the Decennial Census and the ACS. There are three types of housing units: households, group quarters, and vacant units.

- The term "**group quarters**" refers to residences that are institutions or other group-living arrangements that are owned or managed by an entity or organization providing housing and/or services for the residents.

- The term "**household**" refers to residences that are not group quarters.

- The term "**civilian noninstitutional**" refers to all persons who do not report employment in the armed forces and do not report living in an institution.

- The term "**wage and salary workers**" refers to all persons who report working during the year prior to the survey and report receiving wage and salary income but no self-employment income (e.g., income from a business, professional practice, or farm).

- The term "earned income" refers to all pre-tax wage and salary income received by employees.

Summary measures from Integrated Public Use Microdata Series (IPUMS) microdata

Relevant indicators:

- Race/ethnicity/nativity
- Wages: Median
- Wages: $15/hr
- Income inequality: Gini
- Income inequality: 95/20 ratio
- Income growth
- Unemployment
- Home ownership
- Education levels and job requirements
- Disconnected youth
- Housing burden
- Car access
- Income gains with racial equity
- Contribution to growth: Immigrants
- Median age

About IPUMS microdata

Although a variety of data sources were used, much of our analysis is based on a unique dataset created using microdata samples (i.e., “individual-level” data) from IPUMS for four points in time: 1980, 1990, 2000, and 2008 through 2012 pooled together. While the 1980 through 2000 files are based on the decennial census and cover about 5 percent of the U.S. population each, the 2008 through 2012 files are from the ACS and cover only about 1 percent of the U.S. population each. The pooled five-year ACS microdata sample was used to improve the statistical reliability and to achieve a sample size that is comparable to that available in previous years.

Compared with the more commonly used census “summary files,” which include a limited set of summary tabulations of population and housing characteristics, use of the microdata samples allows for the flexibility to create more illuminating metrics of equity and inclusion, and provide a more nuanced view of groups defined by age, race/ethnicity, and nativity in each region of the United States.

A note on sample size

While the IPUMS microdata allows for the tabulation of detailed population characteristics, it is important to keep in mind that because such tabulations are based on samples, they are subject to a margin of error and should be regarded as estimates—particularly in smaller regions/cities and for smaller demographic subgroups. In an effort to avoid reporting highly unreliable estimates, we do not report any estimated ratios or measures of central tendency (e.g., means and medians) that are based on a universe of fewer than 100 individual survey respondents. For example, the universe for the unemployment rates reported for Latinos by the Unemployment indicator is the Latino civilian noninstitutional population ages 25-64, so unemployment rates are only reported for Atlas geographies that have at least 100 actual survey respondents who fall in this particular universe. However, even with this restriction in place, users should not assume that small differences in indicator values between demographic subgroups are statistically significant.

Geography of IPUMS microdata

A key limitation of the IPUMS microdata is geographic detail. Each year of the data has a particular lowest level of geography associated with the individuals included, known as the Public Use Microdata Area (PUMA) for years 1990 and later, or the County Group in 1980. PUMAs are generally drawn to contain a population of about 100,000, and vary greatly in geographic size from being fairly small in densely populated urban areas, to very large in rural areas, often with one or more counties contained in a single PUMA.

While not a challenge for producing state-level data (as PUMAs do not cross state boundaries), summarizing IPUMS data at the city and regional levels is complicated by the fact that PUMAs do not neatly align with the boundaries of cities and metropolitan areas. Rather, large cities and metropolitan areas have several PUMAs entirely contained within the core of the city (or metropolitan area) but several other, more peripheral PUMAs straddling the metropolitan area boundary.

PUMA-to-region and PUMA-to-city crosswalks

To create a geographic crosswalk between PUMAs and metropolitan areas, and between PUMAs and cities for the 1980, 1990, 2000, and 2010–2014 microdata, we used the approach described below. For simplicity, the description refers only to the PUMA-to-region crosswalk but the same procedure was used to generate the PUMA-to-city crosswalk as well. We first estimated the share of each PUMA's population that fell inside each metro area using population information specific to each year from Geolytics, Inc. at the 2000 census block group level of geography (2010 population information was used for the 2010–2014 geographic crosswalk). If the share was at least 50 percent, then the PUMAs were assigned to the metro area and included in generating our regional summary measures. For most PUMAs assigned to a region, the share was 100 percent. For the remaining PUMAs, however, the share was somewhere between 50 and 100 percent, and this share was used as the "PUMA adjustment factor" to adjust downward the survey weights for individuals included in such PUMAs when estimating regional summary measures. Finally, we made one final adjustment to the individual survey weights in all PUMAs assigned to a metro area: we applied a "regional adjustment factor" to ensure that the weighted sum of the population from the PUMAs assigned to each metro area matched the total population reported in the official census summary files for each year/period. The final adjusted survey weight used to make all metro-area estimates was thus equal to the product of the original survey weight in the IPUMS microdata, the PUMA adjustment factor, and the regional adjustment factor.

Because the PUMAs used to generate summary measures for metro areas and cities do not always correspond perfectly to their geographic boundaries, all such summary measures are subject to some degree of geographic—and therefore statistical—inaccuracy in describing metro area/city social and demographic characteristics. In order to quantify the degree of any such inaccuracies, we developed measures of both the "geographic fit"—that is, how well the PUMAs assigned to a metro area/city line up with the actual metro area/city boundaries—and "demographic fit"—that is, how well demographic measures summarized from the IPUMS microdata line up with the same measures derived from sources for which there is no geographic mismatch (such as census summary data). These measures of geographic and demographic fit were examined and used to withhold the reporting of IPUMS-based data in the Atlas for a handful of cities in 1990 and 2000, and to flag other cities and metro areas for which data in the Atlas should be used with extra caution.

To measure geographic fit, we calculated three measures: the share of the metro area/city population in each year that was derived from PUMAs that were 80 percent, 90 percent, and 100 percent contained in the metro area/city (based on population). For example, a metro area/city with perfect geographic fit would be one in which 100 percent of the population was derived from PUMAs for which 100 percent of the PUMA population was contained in that metro area/city. For brevity in the discussion below, we refer to such as metro area/city as having 100 percent of its population from 100-percent-contained PUMAs. A metro area/city of dubious geographic fit thus might be one in which zero percent of its population was from 80-percent-contained PUMAs (indicating that all of the PUMAs assigned to it were somewhere between 50 and 80 percent contained since a PUMA must be at least 50 percent to be assigned to the metro area/city in the first place). For most cities and metros in each year, the population shares from 80-, 90-, and 100-percent contained PUMAs are near 100 percent, with the vast majority at or above 80 percent. For others, however, these measures of geographic fit fell below 80 percent, and these cities and metros are regarded as having "poor geographic fit."

To measure demographic fit, we then summarized the IPUMS microdata for the 150 largest metro areas and the 100 largest cities to calculate three demographic measures for each year that are related to the sorts of demographic measures included in the Atlas, but far simpler: the percentage people of color, the poverty rate, and the percentage immigrant. We then calculated the same three measures for all

250 geographies in each year using summary data from Geolytics, Inc. for 1980 through 2000 and from the 2014 5-year ACS summary file for 2010–2014, and compared the estimates for each measure derived from the microdata to the corresponding measure derived from the summary data. Because we would expect some degree of variation between the two estimates even for metro areas/cities with perfect geographic fit due to sampling differences between the census microdata and census summary data for all years, as well as any additional variation caused by the estimation procedures applied by Geolytics, Inc. to re-allocate data to 2010 census geographies for 1980 through 2000, we calculated the maximum absolute difference (and percentage difference) between the two estimates for each of the three measures across the set of metro areas/cities with perfect geographic fit and used these as benchmarks to assess the degree of demographic fit for the remaining metro areas/cities. If the absolute difference (or percentage difference) in any of the three demographic measures for a metro area/city was well beyond the benchmark value, we identified the metro area/city as having "poor demographic fit."

Table 2 provides information on the quality of the demographic and geographic fit of the IPUMS microdata for all of the 100 largest cities and 150 metro areas included in the Atlas. The share of the city/metro area population from 80-percent-contained PUMAs in each year is reported in the table, and the quality of the demographic fit is indicated by the shading of the cells in the table. Also indicated in the table is whether or not IPUMS-based data is included in the Atlas for each geography/year. While data is included in the Atlas for all of the 150 largest metro areas in all years, for some of the 100 largest cities in some years IPUMS-based data is excluded from the Atlas—either because there was a poor demographic fit or because there were no PUMAs with at least 50 percent of their population residing in the city (and thus no PUMAs were assigned to the city).

Table 2. Measures of Demographic and Geographic Fit for the 100 Largest Cities and 150 Largest Metro Areas

	1980	1990	2000	2010-2014
100 Largest Cities				
Albuquerque City, NM	0.99	0.68	0.71	0.71
Anaheim City, CA	0.98	0.92	0.99	0.93
Anchorage Municipality, AK	1.00	1.00	1.00	1.00
Arlington City, TX	1.00	0.96	1.00	0.58
Atlanta City, GA	1.00	0.90	0.74	0.81
Aurora City, CO	0.81	0.94	0.78	0.69
Austin City, TX	0.98	0.87	0.95	0.87
Bakersfield City, CA	0.88	0.81	0.94	0.68

IPUMS-based data is:

Included

Quality of Demographic Fit:

- Excellent
- Good
- Reasonable

Excluded

- Poor demographic fit
- No PUMAs assigned

Table 2 (Continued).

	1980	1990	2000	2010-2014
Baltimore City, MD	1.00	1.00	1.00	1.00
Baton Rouge City, LA	0.99	0.86	1.00	0.84
Birmingham City, AL	1.00	0.75	0.42	0.45
Boston City, MA	1.00	1.00	1.00	1.00
Buffalo City, NY	1.00	1.00	1.00	1.00
Chandler City, AZ	N/A	0.00	0.99	0.90
Charlotte City, NC	0.92	0.59	0.93	0.81
Chesapeake City, VA	1.00	1.00	1.00	1.00
Chicago City, IL	1.00	1.00	1.00	0.99
Chula Vista City, CA	N/A	N/A	0.93	0.65
Cincinnati City, OH	1.00	0.70	1.00	0.84
City and County of Honolulu, HI	1.00	1.00	1.00	1.00
Cleveland City, OH	1.00	1.00	1.00	1.00
Colorado Springs City, CO	0.00	0.96	0.78	0.72
Columbus City, OH	0.97	0.60	0.34	0.38
Corpus Christi City, TX	1.00	1.00	1.00	0.84
Dallas City, TX	0.99	0.95	0.59	0.61
Denver City, CO	1.00	1.00	1.00	1.00
Detroit City, MI	1.00	1.00	1.00	0.98
Durham City, NC	1.00	1.00	0.56	0.56
El Paso City, TX	1.00	0.93	0.87	0.93
Fort Wayne City, IN	0.76	0.66	0.79	0.86
Fort Worth City, TX	0.99	0.98	0.98	0.56
Fremont City, CA	N/A	0.89	0.89	0.89
Fresno City, CA	0.98	0.88	0.99	0.88
Garland City, TX	#N/A	0.91	1.00	0.65
Glendale City, AZ	N/A	0.94	0.97	0.78
Greensboro City, NC	0.94	0.83	0.93	0.92
Henderson City, NV	N/A	N/A	1.00	0.85
Hialeah City, FL	#N/A	0.88	0.54	0.73
Houston City, TX	0.96	0.91	0.91	0.75
Indianapolis City (balance), IN	1.00	0.89	0.90	0.90
Irvine City, CA	N/A	0.00	0.97	0.90
Irving City, TX	1.00	1.00	1.00	0.69
Jacksonville City, FL	1.00	1.00	0.82	0.85
Jersey City City, NJ	1.00	1.00	1.00	0.99
Kansas City City, MO	1.00	0.58	0.99	0.78
Laredo City, TX	0.00	0.00	1.00	0.77
Las Vegas City, NV	0.99	0.96	1.00	0.72
Lexington-Fayette urban county, KY	#N/A	1.00	1.00	1.00
Lincoln City, NE	1.00	1.00	1.00	1.00
Long Beach City, CA	1.00	1.00	1.00	1.00
Los Angeles City, CA	1.00	1.00	1.00	0.97
Louisville/Jefferson County metro government (balance), KY	0.52	0.87	0.68	0.64
Lubbock City, TX	1.00	0.59	0.54	0.48
Madison City, WI	0.98	0.00	0.99	0.67
Memphis City, TN	0.95	0.85	0.94	0.91
Mesa City, AZ	N/A	0.69	1.00	0.85
Miami City, FL	1.00	0.99	0.99	0.98
Milwaukee City, WI	1.00	1.00	1.00	0.96
Minneapolis City, MN	1.00	1.00	1.00	1.00

IPUMS-based data is:

Included

Quality of Demographic Fit:

- Excellent
- Good
- Reasonable

Excluded

- Poor demographic fit
- No PUMAs assigned

Table 2 (Continued).

	1980	1990	2000	2010-2014
Nashville-Davidson (balance), TN	1.00	1.00	1.00	1.00
New Orleans City, LA	1.00	1.00	1.00	0.83
New York City, NY	1.00	1.00	1.00	1.00
Newark City, NJ	1.00	0.99	1.00	1.00
Norfolk City, VA	1.00	1.00	1.00	0.91
North Las Vegas City, NV	N/A	N/A	1.00	0.92
Oakland City, CA	1.00	0.97	0.99	0.99
Oklahoma City City, OK	0.99	0.00	1.00	0.71
Omaha City, NE	1.00	0.79	0.53	0.52
Orlando City, FL	0.90	0.84	0.00	0.26
Philadelphia City, PA	1.00	1.00	1.00	1.00
Phoenix City, AZ	1.00	0.99	1.00	0.93
Pittsburgh City, PA	1.00	1.00	1.00	1.00
Plano City, TX	N/A	1.00	1.00	0.90
Portland City, OR	0.90	0.90	0.98	0.98
Raleigh City, NC	0.93	0.68	0.95	0.76
Reno City, NV	0.00	0.86	0.92	0.78
Riverside City, CA	0.98	0.95	0.98	0.98
Rochester City, NY	0.99	1.00	1.00	1.00
Sacramento City, CA	1.00	0.96	1.00	0.84
San Antonio City, TX	0.96	0.90	1.00	0.90
San Bernardino City, CA	0.94	0.94	0.98	0.84
San Diego City, CA	0.99	0.75	0.72	0.75
San Francisco City, CA	1.00	1.00	1.00	1.00
San Jose City, CA	0.97	0.75	0.72	0.72
Santa Ana City, CA	1.00	0.97	1.00	0.98
Scottsdale City, AZ	N/A	1.00	0.45	0.46
Seattle City, WA	1.00	1.00	1.00	1.00
Spokane City, WA	0.98	0.97	1.00	0.71
St. Louis City, MO	1.00	1.00	1.00	1.00
St. Paul City, MN	#N/A	1.00	1.00	1.00
St. Petersburg City, FL	#N/A	N/A	0.52	0.57
Stockton City, CA	0.99	0.55	0.98	0.73
Tampa City, FL	1.00	N/A	0.92	0.87
Toledo City, OH	1.00	0.34	1.00	0.98
Tucson City, AZ	0.96	0.89	0.88	0.70
Tulsa City, OK	1.00	0.00	1.00	0.71
Virginia Beach City, VA	1.00	0.99	1.00	1.00
Washington City, DC	1.00	1.00	1.00	1.00
Wichita City, KS	1.00	0.00	0.98	0.60
Winston-Salem City, NC	0.78	0.67	0.90	0.90
150 Largest Metro Areas				
Akron, OH	1.00	1.00	1.00	1.00
Albany-Schenectady-Troy, NY	0.96	0.96	0.96	0.96
Albuquerque, NM	0.92	1.00	0.98	0.98
Allentown-Bethlehem-Easton, PA-NJ	0.78	0.78	1.00	1.00
Anchorage, AK	0.91	0.85	0.81	0.77
Ann Arbor, MI	1.00	1.00	1.00	1.00
Asheville, NC	0.84	0.57	0.61	0.58
Atlanta-Sandy Springs-Marietta, GA	0.84	0.87	0.97	0.97

IPUMS-based data is:

Included

Quality of Demographic Fit:

- Excellent
- Good
- Reasonable

Excluded

- Poor demographic fit
- No PUMAs assigned

Table 2 (Continued).

	1980	1990	2000	2010-2014
Augusta-Richmond County, GA-SC	0.75	0.91	0.91	0.92
Austin-Round Rock, TX	0.92	0.85	0.93	0.93
Bakersfield, CA	1.00	1.00	1.00	1.00
Baltimore-Towson, MD	0.99	0.99	0.98	0.98
Baton Rouge, LA	1.00	1.00	1.00	1.00
Beaumont-Port Arthur, TX	1.00	1.00	1.00	1.00
Birmingham-Hoover, AL	0.95	0.84	0.94	0.91
Boise City-Nampa, ID	0.62	0.64	0.93	0.90
Boston-Cambridge-Quincy, MA-NH	0.97	0.95	0.92	0.93
Bridgeport-Stamford-Norwalk, CT	0.90	1.00	1.00	1.00
Brownsville-Harlingen, TX	1.00	1.00	1.00	1.00
Buffalo-Niagara Falls, NY	1.00	1.00	1.00	1.00
Canton-Massillon, OH	0.94	0.93	1.00	1.00
Cape Coral-Fort Myers, FL	1.00	1.00	1.00	1.00
Charleston-North Charleston, SC	1.00	1.00	0.82	0.92
Charlotte-Gastonia-Concord, NC-SC	0.87	0.80	0.84	0.93
Chattanooga, TN-GA	0.92	0.92	0.92	0.92
Chicago-Naperville-Joliet, IL-IN-WI	0.98	1.00	0.99	0.99
Cincinnati-Middletown, OH-KY-IN	0.94	0.94	0.96	0.96
Cleveland-Elyria-Mentor, OH	1.00	0.96	1.00	1.00
Colorado Springs, CO	0.97	0.97	0.96	0.99
Columbia, SC	0.82	0.83	0.83	0.94
Columbus, OH	0.96	0.95	0.98	0.98
Corpus Christi, TX	0.79	0.79	0.94	0.86
Dallas-Fort Worth-Arlington, TX	0.98	0.98	0.97	0.98
Davenport-Moline-Rock Island, IA-IL	0.81	0.81	0.98	0.89
Dayton, OH	1.00	0.84	0.95	0.95
Deltona-Daytona Beach-Ormond Beach, FL	1.00	1.00	1.00	0.99
Denver-Aurora, CO	0.97	1.00	0.96	0.93
Des Moines, IA	0.77	0.79	0.78	0.80
Detroit-Warren-Livonia, MI	0.98	0.98	1.00	0.99
Durham, NC	0.90	0.91	0.92	0.85
El Paso, TX	1.00	1.00	1.00	1.00
Eugene-Springfield, OR	1.00	1.00	1.00	1.00
Evansville, IN-KY	0.83	0.65	0.73	0.80
Fayetteville, NC	0.92	0.92	0.90	0.87
Fayetteville-Springdale-Rogers, AR-MO	0.50	0.93	0.90	0.92
Flint, MI	1.00	1.00	0.55	0.49
Fort Wayne, IN	0.85	0.85	0.85	0.85
Fresno, CA	1.00	1.00	1.00	1.00
Grand Rapids-Wyoming, MI	0.77	0.78	0.78	0.78
Greensboro-High Point, NC	0.64	1.00	0.86	0.87
Greenville, SC	0.68	0.68	0.68	0.77
Harrisburg-Carlisle, PA	1.00	1.00	1.00	1.00
Hartford-West Hartford-East Hartford, CT	0.95	1.00	1.00	1.00
Hickory-Lenoir-Morganton, NC	1.00	1.00	1.00	0.85
Honolulu, HI	1.00	1.00	1.00	1.00
Houston-Baytown-Sugar Land, TX	0.99	0.96	0.99	0.99
Huntsville, AL	0.81	1.00	1.00	0.96
Indianapolis, IN	0.97	0.97	0.97	0.99
Jackson, MS	0.90	0.88	0.89	0.81

IPUMS-based data is:

Included

Quality of Demographic Fit:

- Excellent
- Good
- Reasonable

Excluded

- Poor demographic fit
- No PUMAs assigned

Table 2 (Continued).

	1980	1990	2000	2010-2014
Jacksonville, FL	0.93	0.91	0.98	0.98
Kalamazoo-Portage, MI	0.76	0.76	1.00	0.86
Kansas City, MO-KS	0.88	0.90	0.91	0.90
Killeen-Temple-Fort Hood, TX	0.95	0.71	0.95	0.84
Knoxville, TN	0.94	0.63	0.94	0.93
Lakeland, FL	1.00	1.00	1.00	1.00
Lancaster, PA	1.00	1.00	1.00	1.00
Lansing-East Lansing, MI	1.00	1.00	1.00	1.00
Las Vegas-Paradise, NV	1.00	1.00	1.00	1.00
Lexington-Fayette, KY	0.64	0.65	0.64	0.63
Little Rock-North Little Rock, AR	0.79	0.96	0.96	0.96
Los Angeles-Long Beach-Santa Ana, CA	1.00	1.00	1.00	1.00
Louisville, KY-IN	0.95	0.90	0.92	0.94
Madison, WI	0.84	0.85	0.85	0.86
Manchester-Nashua, NH	0.43	1.00	0.88	0.87
McAllen-Edinburg-Pharr, TX	1.00	1.00	1.00	1.00
Memphis, TN-MS-AR	0.78	0.77	0.83	0.83
Miami-Fort Lauderdale-Miami Beach, FL	0.93	1.00	0.99	0.99
Milwaukee-Waukesha-West Allis, WI	1.00	1.00	1.00	1.00
Minneapolis-St. Paul-Bloomington, MN-WI	0.93	0.97	0.97	0.93
Mobile, AL	1.00	1.00	1.00	1.00
Modesto, CA	1.00	1.00	1.00	1.00
Montgomery, AL	0.95	0.69	0.96	0.99
Naples-Marco Island, FL	0.00	0.00	1.00	1.00
Nashville-Davidson--Murfreesboro, TN	0.93	0.94	0.94	0.95
New Haven-Milford, CT	0.81	1.00	1.00	1.00
New Orleans-Metairie-Kenner, LA	0.95	0.93	0.93	0.96
New York-Northern New Jersey-Long Island, NY-NJ-PA	0.99	0.99	1.00	1.00
Ocala, FL	1.00	1.00	1.00	1.00
Ogden-Clearfield, UT	0.98	1.00	0.98	0.98
Oklahoma City, OK	0.82	0.90	0.94	0.94
Omaha-Council Bluffs, NE-IA	0.75	0.78	0.76	0.78
Orlando, FL	1.00	1.00	1.00	0.98
Oxnard-Thousand Oaks-Ventura, CA	1.00	1.00	1.00	1.00
Palm Bay-Melbourne-Titusville, FL	1.00	1.00	1.00	1.00
Pensacola-Ferry Pass-Brent, FL	1.00	1.00	1.00	1.00
Peoria, IL	0.86	0.85	0.95	0.89
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	0.98	0.97	1.00	0.99
Phoenix-Mesa-Scottsdale, AZ	0.94	0.95	0.94	1.00
Pittsburgh, PA	0.91	0.84	0.94	0.94
Port St. Lucie-Fort Pierce, FL	0.00	1.00	1.00	1.00
Portland-South Portland-Biddeford, ME	0.92	0.92	0.86	0.95
Portland-Vancouver-Beaverton, OR-WA	0.93	0.93	0.93	0.93
Poughkeepsie-Newburgh-Middletown, NY	0.51	0.54	1.00	1.00
Providence-New Bedford-Fall River, RI-MA	0.95	0.88	0.98	0.96
Provo-Orem, UT	0.97	0.98	0.98	0.98
Raleigh-Cary, NC	0.75	0.78	1.00	0.97
Reading, PA	1.00	1.00	1.00	1.00
Reno-Sparks, NV	0.99	0.49	0.99	0.99
Richmond, VA	0.91	0.92	0.91	0.92
Riverside-San Bernardino-Ontario, CA	1.00	1.00	1.00	1.00

IPUMS-based data is:

Included

Quality of Demographic Fit:

Excellent

Good

Reasonable

Excluded

Poor demographic fit

No PUMAs assigned

Table 2 (Continued).

	1980	1990	2000	2010-2014
Rochester, NY	0.96	0.96	0.90	0.84
Rockford, IL	0.90	1.00	1.00	1.00
Sacramento--Arden-Arcade--Roseville, CA	0.92	1.00	1.00	1.00
Salem, OR	0.82	0.82	0.82	0.81
Salinas, CA	1.00	1.00	0.69	0.75
Salt Lake City, UT	0.94	0.95	0.93	0.92
San Antonio, TX	0.94	0.94	0.91	0.94
San Diego-Carlsbad-San Marcos, CA	1.00	1.00	1.00	1.00
San Francisco-Oakland-Fremont, CA	1.00	1.00	1.00	1.00
San Jose-Sunnyvale-Santa Clara, CA	0.98	0.98	0.97	0.97
Santa Barbara-Santa Maria-Goleta, CA	1.00	1.00	1.00	1.00
Santa Rosa-Petaluma, CA	1.00	1.00	1.00	1.00
Sarasota-Bradenton-Venice, FL	1.00	1.00	1.00	1.00
Savannah, GA	0.88	0.84	0.79	0.76
Scranton--Wilkes-Barre, PA	0.96	0.95	0.83	0.86
Seattle-Tacoma-Bellevue, WA	1.00	1.00	1.00	1.00
Shreveport-Bossier City, LA	0.69	0.93	0.67	0.65
South Bend-Mishawaka, IN-MI	0.83	0.83	0.84	0.83
Spokane, WA	1.00	1.00	1.00	1.00
Springfield, MA	1.00	1.00	0.87	0.95
Springfield, MO	0.71	0.70	0.88	0.89
St. Louis, MO-IL	0.89	0.91	0.96	0.92
Stockton, CA	1.00	1.00	1.00	1.00
Syracuse, NY	0.90	0.90	0.97	0.89
Tallahassee, FL	0.70	0.84	0.89	0.80
Tampa-St. Petersburg-Clearwater, FL	0.97	1.00	1.00	1.00
Toledo, OH	0.88	0.88	0.94	0.97
Trenton-Ewing, NJ	1.00	1.00	1.00	1.00
Tucson, AZ	1.00	1.00	1.00	1.00
Tulsa, OK	0.80	0.80	0.81	0.83
Vallejo-Fairfield, CA	0.00	0.99	1.00	1.00
Virginia Beach-Norfolk-Newport News, VA-NC	0.96	0.83	0.98	0.97
Visalia-Porterville, CA	1.00	1.00	1.00	1.00
Washington-Arlington-Alexandria, DC-VA-MD-WV	0.91	0.96	0.95	0.94
Wichita, KS	0.79	0.79	0.95	0.96
Wilmington, NC	0.67	0.63	0.84	0.94
Winston-Salem, NC	0.74	0.74	1.00	0.84
Worcester, MA	0.95	0.83	0.86	0.86
York-Hanover, PA	0.33	1.00	1.00	1.00
Youngstown-Warren-Boardman, OH-PA	1.00	1.00	1.00	1.00

IPUMS-based data is:
Included
Quality of Demographic Fit:
Excellent
Good
Reasonable
Excluded
Poor demographic fit
No PUMAs assigned

Estimates and adjustments made using census summary data

Estimated data on race/ethnicity by age

Relevant indicators:

- Racial generation gap

Demographic change and what is referred to as the “racial generation gap” are important elements of the Atlas. Due to their centrality, care was taken to generate consistent estimates of people by race/ethnicity and age group (under 18, 18–64, and over 64 years of age) for the years 1980, 1990, 2000, and 2010, at both the city and county levels. The county-level estimates were then aggregated to generate estimates at the regional, state, and U.S. levels.

For 2000 and 2010, data on the number of people by race/ethnicity and age is readily available in SF1 of the Census in categories that are consistent with the six broad racial/ethnic groups detailed in the Atlas; however, this is not entirely the case for 1980 and 1990. Estimates for these years had to be made to ensure consistency over time, utilizing two different Census summary files from each year. The estimates were necessary because while all data reported in the Atlas (unless otherwise noted) treat “Hispanic or Latino” as one of six broad mutually exclusive racial/ethnic groups (with all other groups excluding people of Hispanic or Latino origin), the U.S. Census Bureau considers “Hispanic or Latino” an ethnicity and not a race, and often reports data only for groups defined by single race alone (e.g., “White alone,” “Black alone”) which includes people of Hispanic or Latino origin.

For 1980, after combining data from STF1 and STF2, information on total population by race/ethnicity for all ages combined was available at the city and county levels for all of the six requisite groups, but race/ethnicity by age was only available for non-Hispanic White, non-Hispanic Black, Latino, and the remainder of the population. To estimate the number non-Hispanic Asian/Pacific Islanders, non-Hispanic Native Americans, and non-Hispanic Others among the remainder for each age group, we applied the distribution of these three groups from the overall city and county populations (across all ages) to that remainder. So, for example, if non-Hispanic Asian/Pacific Islanders were 20 percent of the combined non-Hispanic Asian/Pacific Islander, non-Hispanic Native American, and non-Hispanic Other populations of all ages in a particular city or county, we assumed that the same was true within each of the three age categories (under 18, 18–64, and over 64).

For 1990, the level of detail available in the underlying data differed at the city and county levels, calling for different estimation strategies. At the county level, data by race/ethnicity was taken from STF2A, while data by race/ethnicity and age was taken from the 1990 MARS file—a special tabulation of people by age, race, sex, and Hispanic origin. However, to be consistent with the way race is categorized by the OMB’s Directive 15, the MARS file allocates all persons identifying as “other race alone” or multiracial to a specific race. After confirming that population totals by county (across all ages) were consistent between the MARS file and STF2A, we calculated the number of “other race alone” or multiracial people who had been added to each racial/ethnic group in each county by subtracting the number who were reported in STF2A for the corresponding group. We then derived the share of each racial/ethnic group in the MARS file (across all ages) that was made up of “other race alone” or multiracial people and applied it to estimate the number of people by race/ethnicity and age group exclusive of “other race alone” or multiracial people and the total number of “other race alone” or multiracial people in each age group.

For the 1990 city-level estimates, all data were from STF1, which provided counts of the total population for the six broad racial/ethnic groups detailed in the Atlas but not counts by age. Rather, age counts were only available for people by single race alone (including those of Hispanic origin) as well as for all people of Hispanic origin combined. To estimate the number of people by race/ethnicity and age for the six broad racial/ethnic groups that are detailed in the Atlas, we first calculated the share of each single race alone group that was Hispanic based on the overall population (across all ages). We then applied it to the population counts by age and race alone to generate an initial estimate of the number of Hispanic and non-Hispanic people in each age/race alone category. This initial estimate was multiplied an adjustment factor (specific to each age group) to ensure that the sum of the estimated number of Hispanic people across the race alone categories within each age group equated to the "actual" number of Hispanic origin by age as reported in STF1. Finally, an Iterative Proportional Fitting (IPF) procedure was applied to ensure that our final estimate of the number people by race/ethnicity and age was consistent with the total population by race/ethnicity (across all ages) and total population by age group (across all racial/ethnic categories) as reported in STF1.

Assembling census summary data that is geographically consistent over time

Relevant indicators (city and metro area data only):

- People of color
- Race/ethnicity
- Population growth rates
- Contribution to growth: People of color
- Racial generation gap
- Diversity index

One unique feature of the Atlas is that all data presented are *geographically consistent over time*—meaning that the data for any geography are assembled to reflect the same geographic boundary in all years reported. While data from the decennial census summary files for 1980 through 2010 (and subsequent ACS summary files) are already geographically consistent for all states and for the U.S. as a whole, the same is not true for regions (i.e., metropolitan areas) and cities. As noted above, the definition of regions used in the Atlas reflects the OMB's December 2003 CBSA definitions while cities are defined by the incorporated places used in the 2010 Census (with the exception of City and County of Honolulu, which covers all of Honolulu County).

Assembling geographically-consistent census summary data for regions was relatively straightforward given that they are defined as one or more counties grouped together. While there have been a several county name and Federal Information Processing Standard (FIPS) code changes since 1980, there have been few changes to county boundaries themselves, with most of them occurring in Alaska and outside the sole Alaskan metro area included in the Atlas (Anchorage, AK). Thus, an underlying database of county-level summary files along with county-to-metro-area geographic crosswalks (based on the OMB's December 2003 definitions) was sufficient to summarize data for regions in each year.

Assembling geographically-consistent census summary data for the 100 largest cities was a bit more complicated given that there have been more changes to the geographic boundaries of the 100 largest cities since 1980 and population counts by race/ethnicity are generally not available at a very detailed level of geography in 1980 and 1990. However, we were able to find sufficient data on people by race/ethnicity and age at the census "place" level of geography from the decennial census summary files for 1980 through 2010 (and subsequent ACS summary files). The census place level of geography consists mostly of incorporated cities and unincorporated areas referred to as Census Designated Places (CDPs), and while they do not provide full geographic coverage of U.S. as do other census geographies such as census tracts, they are generally areas of population clustering and thus tend to be geographically compact areas. They also represent the vast majority of the population in urban areas making them a suitable choice of geography to use as a “building block” to assemble historical data for the 2010 boundaries of the 100 largest cities in the U.S.

The specific approach we took was to identify all of the 100 largest cities that had significant boundary changes since 1980 by examining their boundaries in each decade in a Geographic Information System (GIS) and using shapefiles for the census place level of geography for 1980, 1990, 2000, and 2010, from the National Historical Geographic Information System (NHGIS). For any of the 100 largest cities that had expanded since 1980 (through annexation), we identified all places in 1980, 1990, and 2000 for which most of land area the area was covered by the 2010 boundaries of the city in question. Such places were assumed to have been annexed, and this was confirmed as possible through media reporting and public documents found in web searches. In each decadal year (1980 through 2000), data for the annexed places was combined with data for the city (among the 100 largest) that annexed them to produce historical data for the 100 largest cities based on their 2010 boundaries.

The one city that required special treatment was Louisville/Jefferson County metro government (balance), KY. It this case, the city of Louisville annexed all of Jefferson County except for other incorporated places (i.e., the “balance” of the county after removing incorporated places). Because much of the area that was annexed was not covered by census places that existed in 1980, 1990, and 2000, we determined that a more accurate approach to estimating the historical (pre-2010) data for the city based on its 2010 boundaries was to start with data for all of Jefferson County in each year and subtract out the data for all existing census places in 1980, 1990, and 2000 whose boundaries were found to be inside Jefferson County but mostly outside the 2010 boundaries of Louisville/Jefferson County metro government (balance). This was done using GIS software and the same NHGIS shapefiles noted above, identifying historical census places as being “outside” of the 2010 boundaries of Louisville/Jefferson County metro government (balance) if their centroids (i.e., a point representing their geographic center) fell outside of those boundaries.

Adjustments made to demographic projections

Relevant indicators:

- People of color
- Race/ethnicity
- Population growth rates
- Contribution to growth: People of color

Projections of the racial/ethnic composition are based on a combination of initial county-level projections from Woods & Poole Economics, Inc., and national projections from the U.S. Census Bureau.

The national projections we present are based on the U.S. Census Bureau’s 2014 National Population Projections. However, because these projections follow the OMB 1997 guidelines on racial classification and essentially distribute the other single-race alone group across the other defined racial/ethnic categories, adjustments were made to be consistent with the six broad racial/ethnic groups included in the Atlas (which are based on the 2000 Census classification). Specifically, we compared the percentage of the total population composed of each racial/ethnic group from the Census Bureau’s Population Estimates program for 2015 (which follows the OMB 1997 guidelines) to the percentage reported in the 2015 ACS 1-year Summary File (which follows the 2000 Census classification). We subtracted the percentage derived using the 2015 Population Estimates program from the percentage derived using the

2015 ACS to obtain an adjustment factor for each group (all of which were negative, except that for the mixed/other group) and carried this adjustment factor forward by adding it to the projected percentage for each group in each projection year. Finally, we applied the resulting adjusted projected population distribution by race/ethnicity to the total projected population from the 2014 National Population Projections to get the projected number of people by race/ethnicity in each projection year.

Similar adjustments were made to the initial county-level projections from Woods & Poole Economics, Inc. Like the 1990 MARS file described above, the Woods & Poole projections follow the OMB Directive 15-race categorization, assigning all persons identifying as other or multiracial to one of five mutually exclusive race categories: White, Black, Latino, Asian/Pacific Islander, or Native American. Thus, we first generated an adjusted version of the county-level Woods & Poole projections that removed the other or multiracial group from each of these five categories. This was done by comparing the Woods & Poole projections for 2010 to the actual results from SF1 of the 2010 Census, figuring out the share of each racial/ethnic group in the Woods & Poole data that was composed of other or multiracial persons in 2010, and applying it forward to later projection years. From these projections, we calculated the county-level distribution by race/ethnicity in each projection year for five groups (White, Black, Latino, Asian/Pacific Islander, and Native American), exclusive of other and mixed-race people.

To estimate the county-level share of population for those classified as other or multiracial in each projection year, we then generated a simple straight-line projection of this share using information from SF1 of the 2000 and 2010 Census. Keeping the projected other or multiracial share fixed, we allocated the remaining population share to each of the other five racial/ethnic groups by applying the racial/ethnic distribution implied by our adjusted Woods & Poole projections for each county and projection year.

The result was a set of adjusted projections at the county level for the six broad racial/ethnic groups included in the Atlas, which were then applied to projections of the total population by county from Woods & Poole to get projections of the number of people for each of the six racial/ethnic groups. Finally, an IPF procedure was applied to bring the county level results into alignment with our adjusted national projections by race/ethnicity described above. The final adjusted county-level results were then aggregated to produce a final set of projections at the metro area and state levels.

Estimates and adjustments made to U.S. Bureau of Economic Analysis data on gross domestic product

Relevant indicators:

- Job and GDP growth
- GDP gains with racial equity

Data presented on GDP is from the U.S. Bureau of Economic Analysis (BEA). However, due to changes in the estimation procedure used for the national (and state-level) data in 1997, a lack of metropolitan area estimates prior to 2001, and no available county-level estimates for any year, a variety of adjustments and estimates were made to produce a consistent series at the national, state, metropolitan areas, and county levels from 1969 forward. While the county data are not currently included in the Atlas, they were used to build a consistent set of metro-area estimates over time.

Adjustments at the state and national levels

It was necessary to generate an adjusted series of state GDP because of a change in BEA's estimation procedure from a Standard Industrial Classification (SIC) basis to a North American Industry Classification System (NAICS) basis in 1997. Data prior to 1997 were adjusted to avoid any erratic shifts in GDP that year. While the change to NAICS basis occurred in 1997, BEA also provides estimates under a SIC basis in that year. Our adjustment involved calculating the 1997 ratio of NAICS-based GDP to SIC-based GDP for each state, and multiplying it by SIC-based GDP in all years prior to 1997 to obtain our adjusted series of state-level GDP.

The adjusted series of state-level GDP was then used to derive national GDP and to estimate GDP at the county and metro-area levels as necessary (as described below). To maintain consistency with the state data, GDP for the nation was calculated as the sum of GDP by state, and may differ from national GDP reported elsewhere for the following reasons: GDP by state excludes federal expenditures on personnel stationed abroad and on military structures and military equipment located abroad (except office equipment), while these are typically included in national GDP; GDP by state and national GDP have different revision schedules.

County and metropolitan area estimates

To generate county-level estimates for all years and metropolitan-area estimates prior to 2001, a more complicated estimation procedure was followed. First, an initial set of county estimates for each year was generated by taking our adjusted series of state-level GDP and allocating it to the counties in each state in proportion to the total earnings of employees working those counties—a BEA variable that is available for all counties and years. Next, the initial county estimates were aggregated to metropolitan area level, and were compared with BEA's official metropolitan area estimates for 2001 and later (which follow the same December 2003 metro-area definitions used in the Atlas). They were found to be very close, with a correlation coefficient very close to one (0.9997). Despite the near-perfect correlation, we still used the official BEA metro-area data in our final data series for 2001 and later. However, to avoid any erratic shifts in gross product during the years leading up to 2001, we made the same sort of adjustment to our estimates of gross product at the metro-area level that was made to the state and national data pre-1997—we figured the 2001 ratio of the official BEA estimate to our initial estimate, and multiplied it by our initial estimates for 2000 and earlier to get our final estimate of gross product at the metro-area level.

We then generated a second iteration of county-level estimates—only for counties included in metro areas—by taking the final metro-area–level estimates and allocating gross product to the counties in each metro area in proportion to the total earnings of employees working in those counties. Next, we calculated the difference between our final estimate of gross product for each state and the sum of our second-iteration county-level gross product estimates for counties contained within metro areas in the state. This difference, total nonmetropolitan gross product by state, was then allocated to the nonmetropolitan counties in each state, once again using the total earnings of employees working in each county as the basis for allocation. Finally, because some metro areas cross state boundaries, one last set of adjustments was made to all county-level estimates to ensure that the sum of gross product across the counties contained in each metropolitan area agreed with our final estimate of gross product by metropolitan area, and that the sum of gross product across the counties contained in a state agreed with our final estimate of gross product by state. This was done using an IPF procedure.

Assembling a complete dataset on employment and wages by industry

Relevant indicators:

- Job and wage growth

Analyses of jobs and wages by industry are based on an industry-level dataset constructed using two-digit NAICS industry data from the Quarterly Census of Employment and Wages (QCEW) of the Bureau of Labor Statistics (BLS). Due to some missing (or nondisclosed) data at the county and regional levels, we supplemented our dataset using information from Woods & Poole Economics, Inc., which contains complete jobs and wages data for broad, two-digit NAICS industries at multiple geographic levels. (Proprietary issues barred us from using the Woods & Poole data directly, so we instead used it to complete the QCEW dataset.) While we refer to counties in describing the process for "filling in" missing QCEW data below, the same process was used for the metro area and state levels of geography.

Given differences in the methodology underlying the two data sources, it would not be appropriate to simply "plug in" corresponding Woods & Poole data directly to fill in the QCEW data for nondisclosed industries. Therefore, our approach was to first calculate the number of jobs and total wages from nondisclosed industries in each county, and then distribute those amounts across the nondisclosed industries in proportion to their reported numbers in the Woods & Poole data.

To make for a more consistent application of the Woods & Poole data, we made some adjustments to it to better align it with the QCEW. One of the challenges of using the Woods & Poole data as a "filler dataset" is that it includes all workers, while QCEW includes only wage and salary workers. To normalize the Woods & Poole data universe, we applied both a national and regional wage and salary adjustment factor; given the strong regional variation in the share of workers who are wage and salary, both adjustments were necessary. Another adjustment made was to aggregate data for some Woods & Poole industry codes to match the NAICS codes used in the QCEW.

It is important to note that not all counties and regions were missing data at the two-digit NAICS level in the QCEW, and the majority of larger counties and regions with missing data were only missing data for a small number of industries and only in certain years. Moreover, when data are missing it is often for smaller industries. Thus, the estimation procedure described is not likely to greatly affect our analysis of industries, particularly for larger counties and regions.

We applied the procedure described above to the county and state levels (though very few data points were missing the state-level QCEW data). To assemble data for metro areas, we aggregated the county-level results.

Growth in jobs and earnings by wage level, 1990 to 2015

Relevant indicators:

- Job and wage growth

The growth in jobs and earnings by wage level indicator uses our filled-in QCEW dataset described above, and seeks to track shifts in regional industrial job composition and wage growth over time by industry wage level. Using 1990 as the base year, we classified broad industries (at the two-digit NAICS level) into three wage categories: low-, medium-, and high-wage industries. An industry's wage category was based on its average annual wage, and each of the three categories contained approximately onethird of all private two-digit NAICS industries in the region. We applied the 1990 industry wage-category classification across all the years in the dataset, so that the industries within each category remained the same over time. This way, we could track the broad trajectory of jobs and wages in low-, medium-, and high-wage industries.

This approach was adapted from a method used in a Brookings Institution report, *Building From Strength: Creating Opportunity in Greater Baltimore's Next Economy*. For more information, see http://www.brookings.edu/research/reports/2012/04 /26-baltimore-economy-vey. While we initially sought to conduct the analysis at a more detailed NAICS level, the large amount of missing data at the three- to six-digit NAICS levels (which could not be resolved with the method that was applied to generate our filled-in two-digit QCEW dataset) prevented us from doing so.

Health data and analysis

Relevant indicators:

- Overweight and obese
- Asthma
- Diabetes

Health data presented in the Atlas are derived from the Behavioral Risk Factor Surveillance System (BRFSS) database, housed in the Centers for Disease Control and Prevention. The BRFSS database is created from randomized telephone surveys conducted by states, which then incorporate their results into the database on a monthly basis. The results of this survey are self-reported and the population includes all related adults (age 18 or older), unrelated adults, roomers, and domestic workers who live at the residence. The survey does not include adult family members who are currently living elsewhere, such as at college, a military base, a nursing home, or a correctional facility. The most detailed level of geography associated with individuals in the BRFSS data is the county. Using the county-level data as building blocks, we created estimates for the metro areas, states, the District of Columbia, and the nation.

While the data allow for the tabulation of personal health characteristics, it is important to keep in mind that because such tabulations are based on samples, they are subject to a margin of error and should be regarded as estimates—particularly in smaller regions and for smaller demographic subgroups. To increase statistical reliability, we combined five years of survey data, pooling together the years 2008 through 2012. As an additional effort to avoid reporting potentially misleading estimates, we do not report any estimates that are based on a universe of fewer than 100 individual survey respondents. This is similar to, but more stringent than, a rule indicated in the documentation for the 2012 BRFSS data of not reporting (or interpreting) percentages based on a denominator of fewer than 50 respondents. Even with this sample-size restriction, regional estimates for smaller demographic subgroups should be regarded with particular care.

Finally, because not all counties are represented in the BRFSS data, we only report estimates for metro areas with good coverage at the county level. In particular, we do not report estimates for metro areas in which counties identified in the BRFSS account for less than 95 percent of the total adult (age 18 or older) metro area population, based on analysis of county-level adult population counts from the 2012 ACS 5-year summary file.

For more information and access to the BRFSS database, see http://www.cdc.gov/brfss/index.htm.

Estimates of GDP gains with racial equity

Relevant indicators:

- GDP gains with racial equity

Estimates of the gains in average annual income and GDP under a hypothetical scenario in which there is no income inequality by race/ethnicity are based on the IPUMS 2014 5-Year ACS microdata and 2014 GDP data from BEA. To develop our estimates, we applied a methodology similar to that used by Robert Lynch and Patrick Oakford in Chapter Two of *All-in Nation* with some modifications to expand the analysis and to apply the analysis to multiple geographic areas. The expansions made were done to include gains from increased employment rates and to enable the decomposition of total income gains into the portions attributable to increased work efforts (figured as average annual hours of work) versus increased wages (figured as average annual income per hour of work). As in the Lynch and Oakford analysis, once the percentage increase in overall average annual income was estimated, 2012 GDP was assumed to rise by the same percentage. A more detailed description of the methodology is provided below.

We first organized individuals aged 16 or older in the IPUMS ACS into the six mutually exclusive racial/ethnic groups used in the Atlas: non-Hispanic White, non-Hispanic Black, Latino, non-Hispanic Asian/Pacific Islander, non-Hispanic Native American, and non-Hispanic Other or multiracial. Following the approach of Lynch and Oakford in *All-In Nation*, we excluded from the non-Hispanic Asian/Pacific Islander category subgroups whose average incomes were higher than the average for non-Hispanic Whites, with the particular subgroups to be excluded determined separately for each of the geographic areas for which we report data for this indicator. Also, to avoid excluding subgroups based on unreliable average income estimates due to small sample sizes, we added the restriction that a subgroup had to have at least 100 individual survey respondents in order to be excluded.

We then assumed that all racial/ethnic groups had the same average annual income and hours of work, by income percentile and age group, as non-Hispanic Whites, and took those values as the new "projected" income and hours of work for each individual. For example, a 54-year-old non-Hispanic Black person falling between the 85th and 86th percentiles of the non-Hispanic Black income distribution was assigned the average annual income and hours of work values found for non-Hispanic White persons in the corresponding age bracket (51 to 55 years old) and "slice" of the non-Hispanic White income distribution (between the 85th and 86th percentiles), regardless of whether that individual was working or not. The projected individual annual incomes and work hours were then averaged for each racial/ethnic group (other than non-Hispanic Whites) to get projected average incomes and work hours for each group as a whole, and for all groups combined. The income gains for each group (and for all groups combined) were then decomposed into the portions attributable to increased hours of work and income per hour using the following formula:

$$\underbrace{\ln(Yp_i)-\ln(Ya_i)}_{\substack{\text{Total percent increase in}\\ \text{average annual income}}}=\underbrace{\ln(Wp_i)-\ln(Wa_i)}_{\substack{\text{Portion attributable to}\\ \text{increase in average annual}\\ \text{income per hour of work}}}+\underbrace{\ln(Hp_i)-\ln(Ha_i)}_{\substack{\text{Portion attributable to}\\ \text{increase in average annual}\\ \text{hours of work*}}}$$

**Includes both an increase in employment rates and increased hours for workers.*

Where: Y = average annual income
H = average annual hours of work
W = average annual income per hour (Y/H)
i represents each racial/ethnic group (or all groups combined)
a represents actual (current) values
p represents projected (hypothetical) values

Once decomposed, the portions of the income gain attributable to increased wages (increased average annual income per hour of work) and increased employment (average annual hours of work) were restricted to range between zero and 100 percent.

One difference between our approach and that of Lynch and Oakford is that we include all individuals ages 16 years and older, rather than just those with positive income. Those with income values of zero are largely non-working, and were included so that income gains attributable to increased hours of work would reflect both more hours for the those currently working and an increased share of workers – an important factor to consider given differences in employment rates by race/ethnicity. One result of this choice is that the average annual income values we estimate are analogous to measures of per capita income for the age 16- and-older population and are thus notably lower than those reported in Lynch and Oakford. Another is that our estimated income gains are relatively larger as they presume increased employment rates.

Note that because no GDP data is available at the city level (partly because economies tend to operate at well beyond city boundaries), our estimates of gains in GDP with racial equity are only reported for metro areas, states, and the U.S. as a whole. Estimated income gains and the source of gains by race/ethnicity, however, are reported for all Atlas geographies, including cities.

School poverty data and analysis

Relevant indicators:

- School poverty

The school poverty data presented in the Atlas are derived from the National Center for Education Statistics (NCES) Common Core of Data (CCD) Public Elementary/Secondary School Universe Survey. Survey responses are submitted annually to NCES by state education agencies in the 50 states, the District of Columbia, and other U.S. territories and outlying areas. The data is then cleaned and standardized by CCD survey staff and made available to the public. All public elementary and secondary schools from pre-kindergarten through 12th grade with a positive total student count (based on the NCES variable *MEMBER*) in each year were included in our analysis of school poverty. This includes both regular schools as well as special education, vocational education, alternative, charter, magnet, and Title 1-eligible schools.

The share of students eligible for free or reduced price lunch (FRPL) was calculated at the school level by dividing the count of students eligible for FRPL (NCES variable *TOTFRL*) by the total student count (NCES variable *MEMBER*). Schools were then classified into four groups—school poverty level categories—based on this share (low, mid-low, mid-high, and high), and the number and shares of students by school poverty level category were aggregated to the various Atlas geographies for each racial/ethnic group. For the vast majority of schools, the total student count is consistent with the sum of the counts by race/ethnicity. For a small number of schools, however, it is slightly higher given that the latter excludes any students belonging to an unknown or non-CCD race category. For this reason, data for all racial/ethnic groups combined (the "All" category) reported in the Atlas is based on the sum of student counts by race/ethnicity. School classification by type (e.g., “Primary schools,” “Middle schools,” and “High schools”) is based on the highest and lowest grades offered at each school, following the categorization found in NCES variable *LEVEL*: [6]

1 = Primary (low grade: PK through 03; high grade: PK through 08)
2 = Middle (low grade: 04 through 07; high grade: 04 through 11)
3 = High (low grade: 07 through 12; high grade: 12 only)
4 = Other (any other configuration not falling within the above three categories, including ungraded and operational schools with nonapplicable grade spans)

While data for “Other” schools are not broken out separately, they are included in the data reported for “All public schools.”

Attaching accurate county and city tags (for the 100 largest cities) was done with a great deal of care. Our approach was based on the general observation that geographic information for schools in the NCES data appeared to be more accurate in later years than in earlier years, and that schools could be consistently linked across survey years by the NCES school ID (NCES variable *NCESSCH*).

The first step was to apply county, city name, state abbreviation, and zip code information (based on school location, not mailing address) from the latest survey year available backward to fill in missing or potentially erroneous information in all earlier years for each school. Given that county tags are available in the NCES data beginning with the 2009–2010 school year (at least among the years included in our analysis), this “backward casting” of geographic information filled in county tags for the majority of schools in each year, with the primary exception being schools that were included only in the 1999–2000 survey but not in any later surveys where county information was collected. To fill in county tags for remaining schools, we used a variety of geographic crosswalks (e.g., between census places and counties, and Zip Code Tabulation Areas [ZCTAs] and counties).[7] For about 100 schools for which county

---

[6] *Documentation to the NCES Common Core of Data Public Elementary/ Secondary School Universe Survey: School Year 2013–14, Provisional Version 1a*, p. 14, available at: https://nces.ed.gov/ccd/pdf/2015150_2013-14_School_documentation_v1a.pdf.

[7] Matches based on the census place-to-county crosswalks were prioritized over matches based on the ZCTA-to-county crosswalks given that places are generally more likely to be fully contained within a single county. While ZCTAs (defined by areal polygons) do not correspond perfectly with the postal zip codes provided in the NCES data

tags were still missing, or for which there was any significant doubt around the accuracy of a county code that was attached using one of the aforementioned geographic crosswalks, we conducted internet searches and used GIS software and Google Earth to confirm the correct county.[8]

The procedure for attaching accurate city tags (for the 100 largest cities included in the Atlas, which are based on 2010 jurisdictional boundaries) to schools was far more complicated. The primary challenge was the fact that we only had the city name to match by (rather than a unique ID), and for large cities, the city name indicated in the NCES data is often not a real city but rather a "neighborhood" or other area within a city that has a colloquial name. For example, a total of 41 unique city names were reported by schools located within the City of Los Angeles, including names from Arleta to Woodland Hills, while 51 unique city names were reported by schools located within New York City. In some cases, city names reported by schools reflected former independent cities or CDPs that had been annexed by one of the 100 largest cities prior to the survey year, but the outdated city names were reported nonetheless. Another challenge was that the city names in the NCES data were not always consistently reported across schools, with inconsistencies related to abbreviations, hyphens, and misspellings/typos often found in the reported city names. For example, a total of six unique city names were reported for schools located in Oklahoma City, including "OK CITY," "OKC," "OKLAHOMA CITY," "OKLAHOMA CIT," and "OKLA CITY."

To address these challenges, we mapped all schools in the 2013–2014 survey (which included latitude and longitude coordinates) using GIS software. We then intersected the resulting set of points with a polygon shapefile for the 100 largest cities, used in the Atlas. This allowed us to create a crosswalk between all of the various city names reported by schools and our internal codes for the 100 largest cities. In cases where only a portion of the schools reporting a particular city name fell within one of the 100 largest cities, all schools reporting that city name were considered either inside or outside that city based on where the largest share of all students were located, and this student share was preserved in the crosswalk. In the end, the crosswalk included a set of city names reported by schools (including both real "neighborhood" and other colloquial names as well as misspellings) that corresponded with each of the 100 largest cities.[9]

Given that most schools reported the same city name in earlier years of the NCES survey (even if it differs from how other schools in the same city report it), application of this city names–to–100 largest cities crosswalk to schools in earlier years of the survey captured the vast majority of schools in each of the 100 largest cities in each year. Prior to application, however, the crosswalk was augmented by

---

(which are defined by a set of mail delivery routes), they correspond well enough to use for our matching purposes.

[8] Uncertainty around the county code attached using one of the aforementioned geographic crosswalks can arise in cases where a census place or ZCTA intersects more than one county. In such instances, the census place or zip code was assigned to the county containing the largest share of its population, and that population share was preserved in the crosswalk. "Significant doubt" around the accuracy of crosswalk-based matches is defined as a population share associated with a match that falls below 80 percent.

[9] For two of the 100 largest cities (Las Vegas, NV, and Miami, FL), a majority of students attending schools that reported an address in these two cities actually attended schools located outside the official 2010 city boundaries. However, given that we had no viable way of distinguishing schools falling inside or outside the boundaries of the 100 largest cities in earlier years of the survey (given that no geographic coordinates are reported), data for all schools reporting location in Las Vegas or Miami are included in the data found in the Atlas for these two cities, for the sake of consistency over time.

comparing the list of 297 city names found in an initial draft of the crosswalk to a list of the full set of unique city names reported in any year of the NCES data included in our analysis. The latter list was scanned for both apparent misspellings of the city names found on the first list and the names of cities that we knew had been annexed by one of the 100 largest cities prior to 2010. This final scan resulted in as additional 21 city names being added to the crosswalk, for a total of 318 unique city names that were associated with the 100 largest cities found in the Atlas.

It is important to note that the measure of school poverty used, the share of students eligible for FRPL, is not always reported and is subject to some degree of error at the school level. The reasons for this include the fact that the count of students deemed FRPL-eligible may be taken at a different time than the total student count, and in some states, a single school may administer the free lunch program for a group of schools (in which case its count and share of FRPL-eligible students would be overstated). However, it is likely that any bias caused by these inconsistencies in reporting at the school level are largely mitigated once the data is aggregated across the many schools in a given Atlas geography. It is also important to note that the Healthy, Hunger-Free Kids Act of 2010 changed eligibility requirements and this can impact comparability of the school poverty data over time. In particular, the Act introduced the Community Eligibility Option (CEO), available in 11 states (including the District of Columbia) by the 2013–14 school year and in all states in the 2014–15 school year, which allows more children to be eligible for FRPL.[10]

Given the prevalence of missing data for some schools and changes to eligibility requirements in recent years, we took precautions to avoid reporting data that are inaccurate or misleading. First, we do not report school poverty information if ten percent or more of the relevant student population attends schools that do not report valid (non-missing) FRPL eligibility data. Second, after making an initial calculation of the overall share of students eligible for FRPL based on available data for the 2009–10 through 2013–14 schools years, we examined changes in this measure over time for all 301 Atlas geographies and noted any dramatic year-to-year changes. School poverty data for a handful of Atlas geographies in certain years were set to missing based on this examination (a total of 30 Atlas geographies in one or more years were affected).

Air pollution data and analysis

Relevant indicators:

- Air pollution: Exposure index
- Air pollution: Unequal burden

The air pollution data presented in the Atlas are derived from exposure to air toxics data from the 2011 National-Scale Air Toxics Assessment (NATA) developed by the U.S. Environmental Protection Agency. The NATA uses general information about emissions sources to develop risk estimates and *does not* incorporate more refined information about emissions sources, which suggests that the impacts of risks may be overestimated. Note, however, that because the Atlas indicators we developed using this data are relative—relative to the U.S. overall in the case of exposure index, and relative to each Atlas geography in the case of the unequal burden indicator—the fact that the underlying risk estimates

[10] See this NCES blog post for more information: http://nces.ed.gov/blogs/nces/post/free-or-reduced-price-lunch-a-proxy-for-poverty.

themselves may be overstated is far less problematic. The NATA data include estimates of cancer risk and respiratory hazards (non-cancer risk) at the census tract level based on exposure to outdoor sources. It is important to note that while diesel particulate matter (PM) exposure is included in the NATA non-cancer risk estimates, it is not included in the cancer risk estimates (even though PM is a known carcinogen). Emissions source types were grouped into four categories for use in the Atlas: *on-road mobile* (e.g., cars, trucks, buses), *off-road mobile* (e.g., airplanes, trains, lawn mowers, construction vehicles, farm machinery, boats), *major stationary* (e.g., larger commercial/industrial facilities such as power plants, refineries, and factories), and *area and other*. The last of these categories is a residual category and includes smaller commercial/industrial facilities (e.g., dry cleaners and small manufacturers) along with fires and other natural sources and secondary and background sources.

For more information about the NATA data, see http://www.epa.gov/national-air-toxics-assessment.

Air pollution: Exposure index

The index of exposure to air pollution presented in the Atlas is based on calculations of three exposure indices at the census tract level, using the 2011 NATA: cancer risk, respiratory (or non-cancer) risk, and the two risk types combined.[11] For the cancer and non-cancer risk indices, the two NATA risk measures were simply ranked at the census tract level across the entire United States, from 1 to 100 (in ascending order, so that “100” includes the top one percent of all U.S. tracts in terms of risk). For the combined risk index, we followed the approach used by the U.S. Department of Housing and Urban Development (HUD) in developing its Environmental Health Index.[12] The cancer and non-cancer estimates were combined by calculating tract-level z-scores for each and adding them together as indicated in the formula below:

$$COMBINED_i = \left(\frac{c_i - \mu_c}{\sigma_c}\right) + \left(\frac{r_i - \mu_r}{c_r}\right)$$

Where *c* indicates cancer risk, *r* indicates respiratory risk, *i* indexes census tracts, and μ and σ represent the means and standard deviations, respectively, of the risk estimates across all census tracts in the United States. As with the individual risk indices, the combined tract level index, $COMBINED_i$, was ranked in ascending order across the entire United States, from 1 to 100.

With the three tract-level rankings in place (for cancer, non-cancer, and combined risk), we then estimated the contribution to each ranked value of four source categories. As noted above, these include on-road mobile, off-road mobile, major stationary, and area and other. To estimate their relative contributions to each ranked risk index, we multiplied each tract-level ranking by the tract-level ratio of the risk estimate for each source category to the total risk estimate for the tract (such that the sum of contributions across source categories was equal to the ranked risk value in each tract). For the combined cancer and non-cancer ranking, the tract-level ratio for each source category was calculated

[11] While the 2011 NATA includes estimates for Puerto Rico and the U.S. Virgin Islands, our analysis for this indicator (and all indicators in the Atlas) excludes all U.S. territories.

[12] See the HUD methodology here: https://www.hudexchange.info/resources/documents/AFFH-Data-Documentation.pdf. The HUD Environmental Health Index was developed using the 2005 NATA, and includes neurological risk in addition to cancer and respiratory risk. However, given that neurological risk was is not included in the nationwide results for the 2011 NATA, it is excluded from our analysis.

by taking a weighted average of the ratios for the two risk types, using their aforementioned z-scores as weight.

Finally, the tract-level rankings (and their contributions by source category) were summarized for each Atlas geography and demographic group (i.e., by race/ethnicity and poverty status) by taking a population-weighted average using the group population as weight, with group population data drawn from the 2014 5-year ACS summary file. The result of that population-weighted average is the index of exposure to air pollution presented in the Atlas. The index value reported reflects the national percentile ranking in terms of pollution burden for each group and geography for which it is reported. For example, an exposure index value of 25 for all people suggests that the average person, in the indicated Atlas geography and for the indicated risk type, lives in a neighborhood (census tract) that ranks at the 25th percentile nationally (that is, it has greater pollution exposure than 24 percent of all census tracts in the U.S. but less exposure than 75 percent of all tracts). An exposure index value of 56 for Blacks suggests that the average Black person in the indicated Atlas geography and for the indicated risk type lives in a neighborhood that ranks at the 56th percentile nationally in terms of pollution exposure.

Air pollution: Unequal burden

The analysis of unequal burden in air pollution is based a measure called "minority discrepancy" found in a 2013 academic article by Michael Ash et al., "Is Environmental Justice Good for White Folks? Industrial Air Toxics Exposure in Urban America."[13] In that article, minority discrepancy was calculated at the regional (CBSA) level as the minority share of burden from pollution exposure minus the minority share of the population (with minorities defined in the same way as we define people of color—that is, all people except for non-Hispanic Whites). We make similar calculations, but do so for a wide variety of demographic groups, for different risk types and source categories, and for each Atlas geography. Given that we derive our measure of discrepancy in pollution burden for a variety of demographic groups (rather than just for minorities), we call our measure of the difference between a group's share of pollution burden and their share of the overall population the unequal burden index. Our calculations utilize the same census tract-level 2011 NATA data described above and the 2014 5-year ACS summary file for population data.

While calculating each demographic group's share of the overall population was straightforward, the process of calculating its share of pollution burden requires some explanation. Essentially, we calculate burden in each census tract as the product of the NATA risk estimate and the total population, and sum up the results across all tracts in each Atlas geography to measure the total burden in that geography. We then repeat the process to calculate burdens for each demographic group (using the group's tract-level population instead of the total population), and divide by total burden to get each demographic group's share of the total. The unequal burden index for each group is then calculated by subtracting their share of the population from their share of the burden in each Atlas geography. We carry out this procedure separately for each NATA risk type (cancer and non-cancer) and each source category (on-road mobile, off-road mobile, major stationary, and area and other) as well as for all sources combined.

It is important to note that this measure is calculated relative to each Atlas geography and does not reflect the actual amount of exposure that each group faces. As such, a geography might show high

[13] Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6237.2012.00874.x/epdf.

levels of unequal burden even if exposure is low; therefore, it is most useful to consider this indicator in conjunction with the Air pollution: Exposure index indicator. It is also useful to note that negative values for the unequal burden index are indeed possible and arise when a demographic group's share of pollution burden is less than its share of the overall population. If exposures were evenly distributed across all demographic groups in a given geography, the unequal burden index would be zero for all groups; positive values of the index indicate disproportionately high pollution burden, while negative values indicate disproportionately low pollution burden.

Additional notes by indicator

Below, we provide addition information that is specific to individual indicators as deemed necessary.

- **Neighborhood poverty**
  The census tract geography changes with each decennial census, which can be problematic for analyzing changes in neighborhood poverty over time. In order to insure a consistent geographic basis for our calculations, we used data from GeoLytics, Inc. to derive neighborhood poverty in 2000. While this data originates from the 2000 Census (SF3), it has been "re-shaped" to be expressed in 2010 tract boundaries, which is the geographic bases of the 2012 5-year ACS summary file (which is used for neighborhood poverty in 2012).

- **Diversity index**
  The formula used to calculate the diversity score was drawn from a 2004 report by John Iceland of the University of Maryland, *The Multigroup Entropy Index (Also Known as Theil's H or the Information Theory Index)*, available at: http://www.census.gov/housing/patterns/about/multigroup_entropy.pdf. In that report, the measure is referred to as the "entropy score" and its derivation can be found on page 7.

- **School poverty**
  While it is generally the case that all data found in the Atlas reflect consistent geographic boundaries, school poverty data for two of the 100 largest cities—Las Vegas City, NV, and Miami City, FL—include all schools reporting location in each of these respective cities regardless of whether they actually fall within the city's 2010 boundaries. See the above discussion under "School poverty data and analysis" for more information.

- **Detailed race/ethnicity**
  In the "By ancestry" breakdown, to avoid reporting highly unreliable estimates, we do not report estimated numbers of people by race/ethnicity and ancestry that are based on a universe of fewer than 50 individual survey respondents. This threshold for reporting is less restrictive than the 100-survey-respondent threshold that we apply when reporting measures of central tendency. The reason for this is that there is less risk of reporting a misleading value when reporting population estimates than when reporting measures of central tendency. For example, a population estimate that is based on a small sample size (e.g., between 50 and 99 survey respondents) will still show up as a small number while an estimated mean value could show up as a very high or very low value (relatively). By relaxing the threshold to 50 survey

respondents, we are able to more fully show the extent of the diversity that exists within Atlas geographies.