I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

POLICY PAPER SERIES

IZA Policy Paper No. 190

The Community Explorer: Bringing Populations' Diversity into Policy Discussions, One County at a Time

Claude Lopez
Hyeongyul Roh
Maggie Switek

SEPTEMBER 2022

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

POLICY PAPER SERIES

IZA Policy Paper No. 190

The Community Explorer: Bringing Populations' Diversity into Policy Discussions, One County at a Time

Claude Lopez
Milken Institute

Hyeongyul Roh
Milken Institute

Maggie Switek
Milken Institute and IZA

SEPTEMBER 2022

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

The Community Explorer: Bringing Populations' Diversity into Policy Discussions, One County at a Time

The Community Explorer provides new insights and data on the characteristics and diversity of the US population. Using machine learning methods, it synthesizes the information of 751 variables across 3,142 counties from the US Census Bureau's American Community Survey into 17 communities. Each one of these communities has a distinctive profile that combines demographic, socio-economic, and cultural behavioral determinants while not being geographically bounded. We encourage policy makers and researchers to make use of the results of our analysis. The Community Explorer dashboard provides the location of these profiles, allowing for targeted deployment of community interventions and, more broadly, increasing the understanding of socioeconomic gaps withing the US.

JEL Classification: D31, J08, J10,O12, R10

Keywords: diversity, communities, development, economic well-being

Corresponding author:
Maggie Switek
Milken Institute
1250 Fourth Street
Santa Monica, CA 90401
USA
E-mail: mSwitek@milkeninstitute.org

CONTENTS

INTRODUCTION

Black Lives Matter and other social justice movements have increased the general awareness of the diversity of the US population and the need for societal changes. Diversity awareness is becoming an essential element of many policy efforts, from access to health care and financial inclusion to initiatives addressing systemic racism and inequities. Yet, most of these discussions and initiatives overlook the complexity of diversity in the United States. Instead, they focus on a few essential dimensions, such as race and ethnicity, gender, and age.

Such simplification is necessary to bring attention to the urgency of changes. However, identifying the changes and related actionable solutions requires a more refined understanding of the challenges. This starts with a granular understanding of a population's characteristics, allowing tailored and more effective policies and initiatives to be designed.

While data on the multidimensionality of US diversity exists, the challenge stands in making sense of it. How can we account for race and ethnicity, gender, age, income, education, and other relevant dimensions while presenting the data in a format suited to inform decision-making?

With the Community Explorer, we synthesize the information related to the different dimensions of US diversity into a few communities. Using Census Bureau's American Community Survey (ACS) data, we apply machine-learning techniques to identify population-characteristic patterns across the 3,142 counties.[1] The county location is not part of the dimensions considered, which allows for identifying similarities across counties, regardless of their proximity. As a result, each community has a distinctive profile that combines demographic, economic, and many other determinants while not being geographically bounded.

We first presented this novel approach in *The Community Explorer: Informing Policy with County-Level Data*. Using 26 behavioral, economic, and social factors, we sorted the

[1] In 2019, Valdez-Cordova Census Area in Alaska is divided into two, making the number of counties 3,143 for the Census 2020.

3,142 US counties into eight community profiles, each grouping counties that share a common combination of behavioral determinants while not being geographically bounded.

In this report, we extend the number of dimensions considered to 751 variables for the 3,142 counties. The extra 725 variables add tremendous granularity to the analysis, resulting in 17 community profiles that emerge from the data.
The Community Explorer dashboard provides the location of these profiles, allowing for targeted deployment of community interventions and, more broadly, increasing the understanding of socioeconomic gaps withing the US. We have identified four main benefits of our approach:

- **Lets the data speak:** We use an agnostic approach to recognize the interactions among a wide range of factors at the county level. The resulting profiles provide an objective snapshot of how communities can be described based on the Census data, without imposing any assumptions or restrictions.

- **Leverages the data granularity when aggregating its information:** Our approach uses the county dimension as the aggregation unit, not as a geographic restriction. As a result, communities are defined by the core characteristics of their population. In contrast, most analyses either impose a geographic dimension and pool the data at the state or regional levels or ignore it by pooling the information at the national level.
- **Allows for peer-counties comparison and insightful benchmarking:** Counties in each profile have more in common, based on the variables considered, than with the rest of the US or the other profiles. As a result, comparing the performance of two counties within the same profile or using the profile average as a benchmark, in addition to the state and the national level, provides new insights toward actionable solutions.

- **Performs as a great visualization tool.** The Community Explorer dashboard provides an interactive map with the location of the profiles and graphs with additional

statistics for the US, the profiles, and for each county. This allows users to visually explore and download information on the profiles and to compare county-level data to the averages for each county's profile and for the US.

Five categories summarize the main feature of these profiles:

Urban America captures four community profiles that represent 74 percent of the US population across 819 urban core, suburban, and small metro counties.
Urban Core => Prosperous, ethnically and linguistically diverse large metro areas with substantial disparities between their highly educated (largely White[2]) and less educated (largely Black or African American) residents (26 percent of the population)
Lower-Middle Class => Less populous suburban and small metro counties that are not as economically prosperous as the rest of Urban America (18 percent of the population)
Affluent Suburbs => Affluent and more populous (but less diverse) suburban and small metro counties that jointly represent the profile with the highest median income (16 percent of the population)
Middle Class => Middle-class communities with a largely White population that resides in large- to medium-sized suburban and small metro counties (14 percent of the population)

Industry-Driven America captures five community profiles that include 17 percent of the US population across 1,507 counties where employment is concentrated in one industry that shapes all aspects of the population's profile.
College Towns => College towns with a relatively young, highly educated, and highly geographically mobile population (5.4 percent of the population)
Manufacturing Midwest => Counties primarily located in the Midwest that form the profile with the highest proportion of the White population working in the

[2] Here and throughout the report we refer to racial or ethnic descriptions as recorded by the US Census Bureau. All racial or ethnic groups include only the non-Hispanic population (except for the Hispanic or Latino group, which includes Hispanic population of any race).

manufacturing sector (5.2 percent of the population)

Low-Wage Manufacturing => Low-wage workers in the manufacturing and chemical industries located largely in the South and Northeast regions of the country, with an above-average proportion of the population living below the poverty line (4.9 percent of the population)

Hispanic Agriculture => Highly agricultural communities with a higher than average concentration of Hispanic or Latino population residing mostly in the West and South (1.2 percent of the population)

The Great Plains => Agricultural counties located in the Great Plains with a high proportion of the White population (0.3 percent of the population)

Graying America captures two community profiles that include 5.1 percent of the US population across 378 counties and jointly represent the profiles with the highest concentration of population of age 65 years or older.

Retiree Communities => Retiree communities with adequate household incomes and access to economic resources (4.5 percent of the population)

Isolated Seniors => Isolated seniors with high disability rates and relatively low incomes (0.6 percent of the population)

Extremely Vulnerable America captures four community profiles that include 3.5 percent of the US population across 424 counties and form the profiles with the lowest levels of income.

Hispanic Southern Border => Counties mostly located along the US southern border with a majority of a relatively young Hispanic or Latino population living in extreme poverty (1.4 percent of the population)

Black South => Southern counties with the highest proportion of Black or African American population and lowest median household income of all profiles (1.3 percent of the population)

White Appalachia => White communities in Appalachia with the third-highest level of unemployment rates and second-lowest household income of all profiles (0.7

percent of the US population)

American Indian Reservations => American Indian Reservation communities living in extreme poverty, with more than one-third of the population with income below the poverty line (0.1 percent of the population)

Noncontiguous America captures two community profiles that include 0.42 percent of the US population across 34 counties, combining all Hawaiian and nine Alaskan counties.

Hawaii => The Aloha State with high racial and ethnic diversity, high income, and relatively low income inequality (0.4 percent of the population)

Native Alaska => Alaskan communities with large economic gaps between the White and Alaska Native populations (0.02 percent of the population)

DATA

We use the US Census Bureau's ACS five-year data that pool 2015-2019 yearly estimates to include all US counties to have equally reliable information for the 3,142 counties in this report.

We obtain two types of information from the 2015-2019 data: the most frequently requested social, economic, housing, and demographic characteristics,[3] and additional microlevel information such as means of transportation to work, educational attainment, bachelor's degree field, disability characteristics, median income, employment status, characteristics of health insurance coverage, types of computers and internet subscriptions, among many others.[4] The combined data include 4,017 variables; we used the 751 pertinent to our analysis for the population profiles.[5]

[3] Table identification codes for the four tables in ACS are DP02, DP03, DP04, and DP05.
[4] Table identification codes for the 11 tables in ACS are S0802, S0804, S1501, S1502, S1810, S1903, S2301, S2701, S2801, S2802, and B19083.
[5] Pertinent variables include all information related the communities' socioeconomic characteristics. A few examples of variables that we consider as non-pertinent are population counts (as we include the percentages), detailed information on the types of household computing devices (such as having a desktop or laptop), and the number of available vehicles in a household.

METHOD

We synthesize the information of 751 variables across 3,142 counties into a few communities. The number of communities is defined endogenously from the following two-step approach that relies on machine learning techniques: First, we deal with the variables that do not add new information, ultimately reducing the number of variables, then we cluster the counties with similar characteristics.

Variable Reduction

We identify the variables that are correlated or implicitly contain the same information. Not controlling for that double counting would put too much emphasis on these dimensions and mislead the clustering outcome.

We determine the variables essential to our analysis based on the degree of their redundancy or irrelevance. First, we use a density-based spatial clustering algorithm of applications with noise (DBSCAN) to pinpoint highly correlated variables (Ester et al. 1996). DBSCAN enables the clustering of variables while preventing the outliers from influencing the main clusters' profiles. For our analysis, we keep the outliers as variables as they are poorly correlated with one another.[6] Second, based on the clusters found by DBSCAN, we address highly correlated variables in a cluster in one of three possible manners:

- Remove apparent redundancy. For example, several variables in different tables represent household/family income statistics: per capita income, mean family income, median household income, etc. We use only median household income for our analysis.

[6] The algorithm needs two parameter specifications: a search radius (ϵ) and a minimum number of samples. If the distance between two data points is below the threshold ϵ, the two points are considered neighbors. The points in the same neighborhood comprise a cluster only if the cluster has the minimum number of samples that a user defines. Otherwise, the data points are classified as outliers. We set the minimum number of samples as 3 to identify any redundant variables. One strategy for estimating a value for ϵ is to generate a k-distance graph for the input data, in which k is 3 in our case. For each point in the data, this method finds the distance to the k^{th} nearest point, and plots sorted points against this distance. The resulting graph contains a knee, at which the distance rapidly increases. Based on the knee, we chose 10 as the distance. However, for robustness, we also repeated the whole process with widely ranging ϵ, from 1 to 1000, and the minimum number of samples, ranging from 2 to 10. We found the solutions of our method are very robust over different sets of parameters.

- Combine if the details are not critical. For example, percentages of households with income less than $10,000, $10,000-$14,999, and $15,000-$24,999 are highly correlated. The same is true for percentages of households with incomes of $150,000-$199,999 and $200,000 or more. We combine the highly correlated ranges and generate two new variables: the percentage of households with income less than $25,000 and $150,000 or more.
- Keep if each of the correlated variables still gives specific information. For example, the percentage of the Hispanic or Latino population in a county is significantly correlated with overall English fluency (a -0.82 correlation coefficient) and the population speaking a language other than English at home (a 0.9 correlation). Unemployment rate, poverty rate, disability, population percentage without a high school diploma, lack of digital access, and portion of single female parents are highly correlated. Likewise, higher educational attainment is correlated with the prevalence of lucrative industries, such as finance and information, and high-income households. Despite the high correlations between these variables, they all provide valuable and distinct information. Therefore, we keep them all to develop more granular county profiles.

Using one of the above methods, we reduce 751 variables to 199 while effectively retaining all necessary information. Table 1 summarizes the variables used, sorting them under 11 main categories.[7]

Table 1: List of Variables

Category	Variables (#)	Variables (Descriptions)
Demographic	10	Sex ratio, Median age, Race (White, Black or African American, American Indian and Alaska Native, Asian, Native Hawaiian and Other Pacific Islander, Some other race, Two or more races, and Hispanic or Latino).
Social	5	Civilian veterans, Foreign-born population, Non-US citizens, language at home: not English, English fluency: not very well.

[7] See the online appendix for more details.

Income	26	Income distribution (Less than $25,000, $25,000-$34,999, $35,000-$49,999, $50,000-$74,999, $75,000-$99,999, $100,000-$149,999, $150,000 or more), Median household income, Receiving Food Stamp/SNAP benefits, Income below the poverty level (family and people), Median Income by race (White, Black or African American, Asian, Two or more races, Hispanic or Latino, White), Median Income by age (15 to 24 years, 25 to 44 years, 45 to 64 years, 65 years and over), Median Income: single male and female parents, Gini index, Gender wage gap, Racial income gap.
Employment Status	22	Armed forces, Unemployment rate, Unemployment rate by race (White, Black or African American, Asian, Two or more races, Hispanic or Latino, White), Unemployed male and female, Unemployed: below/above poverty, Unemployment with a disability, Unemployment by education (Less than high school, high school, college/associate's, bachelor's), Unemployment by age (less than 25, 25-64, 65 over), Unemployment: racial difference.
Housing	24	Residence one year ago: same/different/abroad, vacant housing units, homeowner vacancy rate, rental vacancy rate, owner-occupied, renter-occupied, no vehicles available, lacking complete plumbing facilities, lacking complete kitchen facilities, no telephone service available, housing costs (SMOCAPI with a mortgage <20%, 20-30%, 30-35%, 35% over, SMOCAPI without a mortgage <10%, 10-30%, 30-35%, 35% over, GRAPI <15%, 15-30%, 30-35%, 35% over).
Employment Sectors	22	Five occupation types and 13 different employment industries categorized by the US Census Bureau (See footnote 8 for more details), profile of workers (private wage and salary workers, government workers, self-employed, unpaid family workers).
Education	28	Educational attainments (Less than 9th grade, 9th to 12th grade, no diploma, high school graduate, some college, no degree, associate's degree, bachelor's degree, graduate or professional degree), median earnings by education levels (less than high school graduate, high school graduate, college/associate's, bachelor's, graduate/professional), bachelor's or higher by race (White, Black, Asian, two or more races, Hispanic or Latino), poverty rate by education (less than high school, high school graduate, college/associate's, bachelor's or higher), Field of Bachelor's degree: science and engineering, science and engineering related, business, education, arts, humanities and others, racial gap for higher education.
Household Type	17	Population, married-couple family, cohabiting couple, single male and female, single male and female parent, male and female householders living alone, senior male and female householders living alone, households with people under 18 years, households with people age 65 years and over, grandparents responsible for grandchildren, school enrollment: elementary school (1-8), high school (9-12), college or graduate school.
Health Insurance / Disability	22	With health insurance, disability by race (White, Black, Asian, two or more races, Hispanic or Latino), disability types (hearing, vision, cognitive, ambulatory, self-care, independent living difficulty), uninsured seniors (65 years over), uninsured people with a disability, uninsured and unemployed, disability by age (under 18, 18-64 years, 65 and over), racial gap by health insurance
Digital Access	17	With a computer, with a broadband internet subscription, no internet with a computer, no internet by age (under 18 years, 18 to 64 years, 65 years and over), no internet by education (Less than high school, high school, bachelor's or higher), no internet unemployed, no computer by age (under 18 years, 18 to 64 years, 65 years and over), no computer unemployed, no internet: racial gap, no computer: racial gap.
Commuting	6	Commuting (drove alone, carpooled, public transportation, walked), work from home, mean travel time to work (minutes)

Note: Variables (#) show how many variables are in a category. SMOCAPI is an acronym for selected monthly owner costs as a percentage of household income. GRAPI denotes gross rent as a percentage of household income.

Source: Milken Institute (2022)

Clustering of Counties

We use the k-means clustering algorithm that partitions data into 'k' mutually exclusive clusters (Lloyd 1982) to group the counties using the information of the 199 variables. While this method is one of the most popular machine learning algorithms, it (as any statistical method) has some drawbacks and assumptions. We tackle three relevant limitations of this method by adjusting the algorithm and transforming the data.

- **Data-specific number of clusters**: The k-means method entails a predetermined number of clusters k. The wrong choice of k could yield poor clustering results. We let the data dictate k by comparing the clustering solutions for different values of k, ranging from 2 to 50, based on four widely used clustering evaluation metrics: silhouette values, gap statistics, the Calinski-Harabasz index (also known as the Variance Ratio Criterion), and the Davies-Bouldin index (Rousseeuw 1987; Tibshirani, Walther, and Hastie 2001; Caliński and Harabasz 1974; Davies and Bouldin 1979). The four methods use different algorithms to approximate scores, indicating the quality of clusters, and complement each other's pitfalls. We choose the best-performing k over those four evaluating algorithms.
- **Clusters robust to initial data points:** The k-means method begins the clustering process using a randomly selected set of initial values and finds a solution, thereby offering a chance to converge to a local minimum solution. To mitigate the dependence on the initial values, we repeat the clustering process with 30,000 different randomly selected initial values and choose the best results.
- **Data standardization**: The k-clustering method uses distance-based measurements to determine the similarity between data points and is sensitive to large numbers and variables with large variance. To deal with this, we

standardize the data, so the variables range from 0 to 100 and rescale them by their standard deviations to ensure a unit variance.

Finally, given the nature of the datasets, a few variables are missing in some counties. For example, the median income for Asians in a county without any Asian population is missing. Replacing the missing values with manipulated values is likely to create unintended bias. Thus, we modify the distance function to calculate a distance based only on a complete set of variables. Specifically, for a county missing any Asian population, a distance metric measures the distance from this county to others without considering Asians' median income, even if the other counties have the value.

SEVENTEEN COMMUNITY PROFILES IN THE US

The machine learning clustering algorithm identifies seventeen communities with a distinctive profile that combines demographic, economic, and many other determinants while not being geographically bounded.

Table 2 summarizes population density, the number of counties, and the average county-level population for each profile. The online appendix further discusses the outstanding features of each profile.

Table 2: Clustering Result

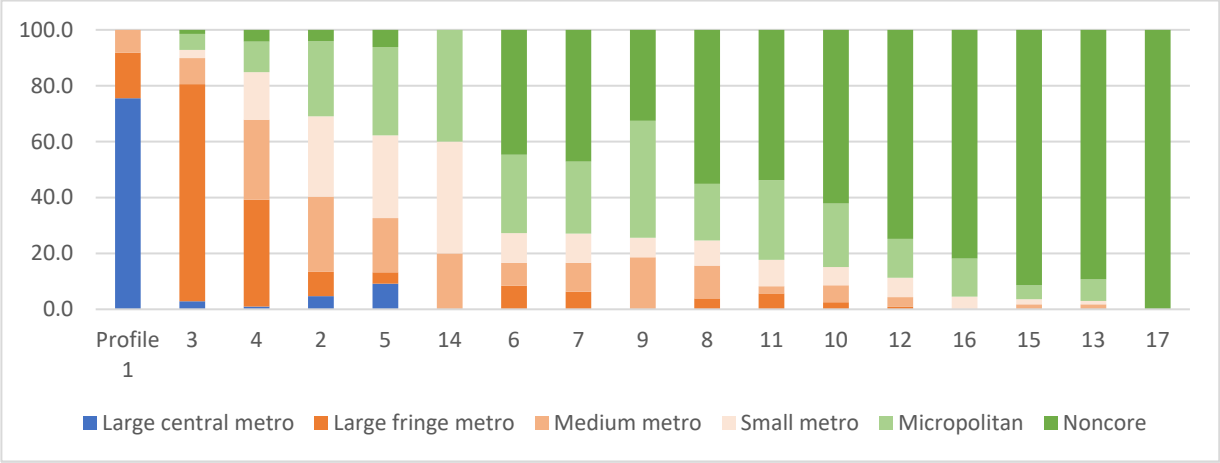
Profile	Population (%)	Number of Counties	Average Population for Counties (thousands)	Group
1/ Urban Core	25.9	49	1,719	Urban America
2/ Lower-Middle Income	18.2	320	185	Urban America
3/ Affluent Suburbs	16.1	139	375	Urban America
4/ Middle Income	13.8	311	144	Urban America
5/ College Towns	5.4	98	178	Industry-Driven America
6/ Manufacturing Midwest	5.2	506	33	Industry-Driven America
7/Low-Wage Manufacturing	4.9	524	30	Industry-Driven America
8/Retiree Community	4.5	256	56	Graying America
9/ Hispanic Southern Border	1.4	43	103	Extremely Vulnerable America
10/Black South	1.3	198	21	Extremely Vulnerable America
11/ Hispanic Agriculture	1.2	158	25	Industry-Driven America
12/ White Appalachia	0.7	115	20	Extremely Vulnerable America
13/ Isolated Seniors	0.6	168	12	Graying America
14/Hawaii	0.4	5	284	Noncontiguous America
15/The Great Plains	0.3	221	4	Industry-Driven America
16/ American Indian Reservations	0.1	22	18	Extremely Vulnerable America
17/Native Alaska	0.02	9	8	Noncontiguous America

Notes: The table shows population density by profile, the number of counties clustered in each profile, and an average of the county-level population. Different color themes of the shades categorize profile by a group.

Source: Milken Institute (2022)

Box 1. How Urban or Rural Are the Profiles?

Percentage of Urban, Suburban, and Rural Counties Per Profile



We use the National Center for Health Statistics' Urban-Rural Classification Scheme to assess each profile's urban profile, using the six classifications of Ingram and Franco (2014):

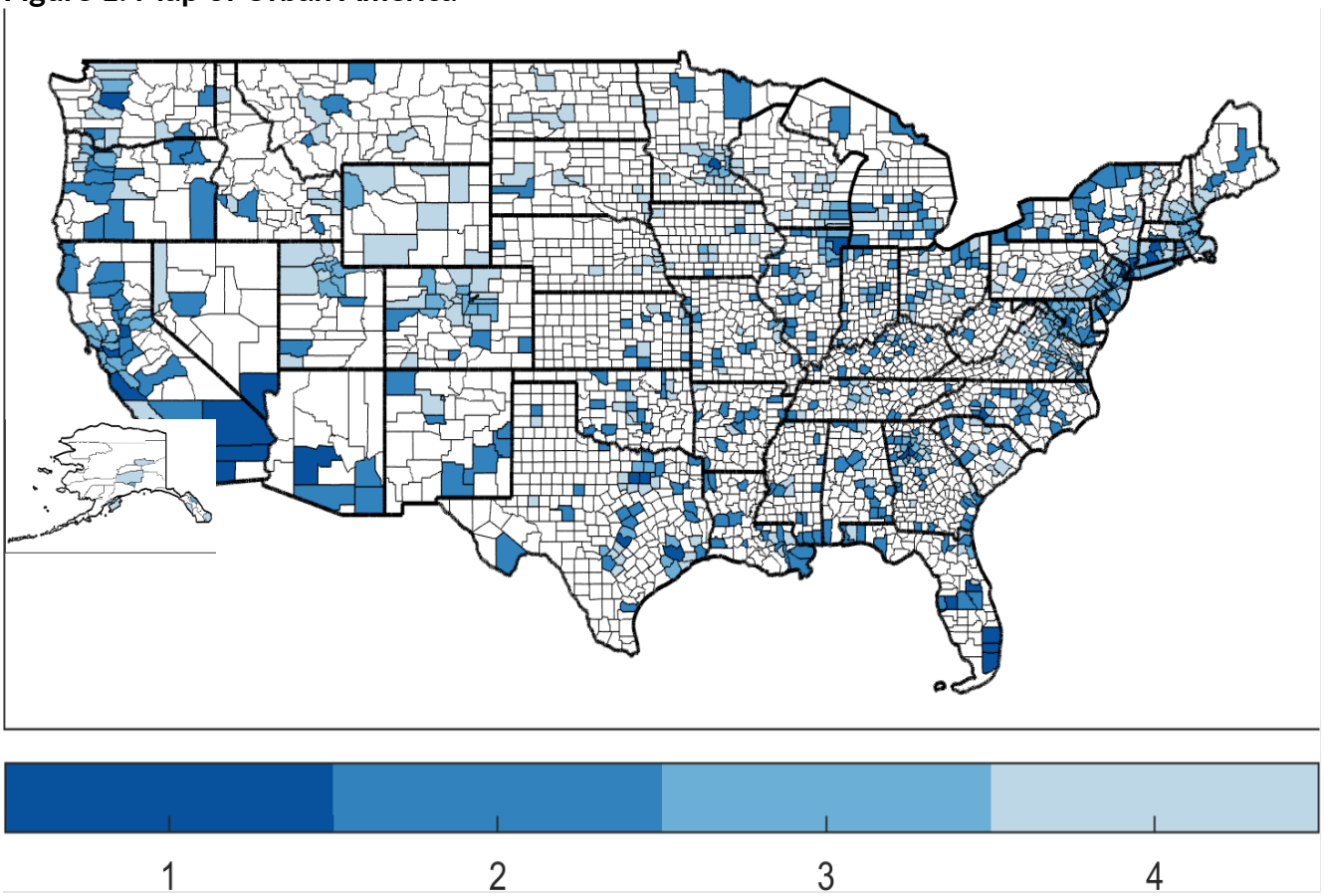
1. Large central metro counties—Counties in a metropolitan statistical area (MSA) of 1 million population that: 1) contain the entire population of the largest principal city of the MSA, or 2) are entirely contained within the largest principal city of the MSA, or 3) contain at least 250,000 residents of any principal city in the MSA.
2. Large fringe metro—Counties in MSAs of 1 million or more population that did not qualify as large central metro counties.
3. Medium metro—Counties in MSAs with populations of 250,000 to 999,999.
4. Small metro—Counties in MSAs of populations less than 250,000.
5. Micropolitan—Counties in micropolitan statistical areas. Each micropolitan statistical area must have at least one urban cluster of at least 10,000 but with less than 50,000 population.
6. Noncore—Nonmetropolitan counties that did not qualify as micropolitan. The Noncore can be thought of as the most rural areas.

Parker et al. (2018) sort these six categories into three main groups: "Urban Core" counties as the 53 US metropolitan areas including 68 counties in Large central metro; "Suburban and Small Metro" counties as 1,098 counties in Large fringe metro, Medium metro, and Small metro; and "Rural" counties as 1,976 counties in Micropolitan and Noncore.

Source: Milken Institute (2022) and the National Center for Health Statistics' Urban-Rural Classification Scheme (2014)

Urban America

Figure 1: Map of Urban America



Source: Milken Institute (2022)

Two-thirds of the American population live in the Urban Core and the surrounding metropolitan counties. As shown in Box 1, the Urban Core profile groups the largest central metro counties while the Affluent Suburbs profile comprises the large fringe metro counties. The Middle-Class profile is a mix of large to medium metro counties, while the Lower-Middle Class profile predominantly comprises medium and small metro and micropolitan counties.

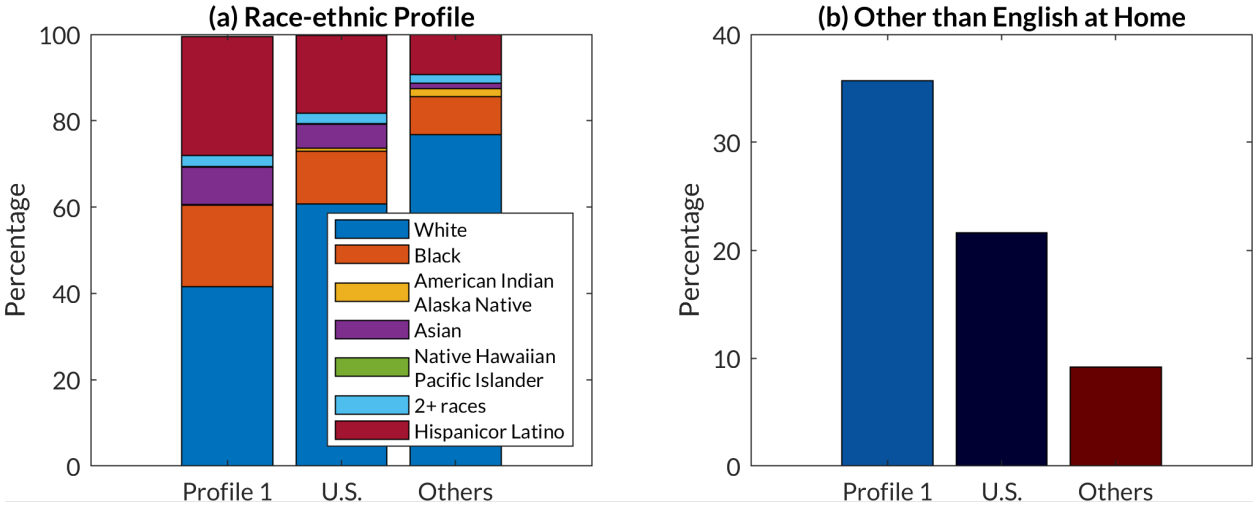
Urban Core: The Large Metropolitan Areas

Accounting for the 49 most populous counties and home to 25.9 percent of the population, the Urban Core is one of the most racially and linguistically diverse profiles, with the highest proportion of foreign-born population. Its population is more educated

than the rest of the US, with the exception of the Hispanics and Latinos . Yet, the higher education benefits mostly the White population, with Whites being the only racial or ethnic group earning a significantly higher income than the national average for their racial or ethnic category, and more than the other racial or ethnic groups in this profile. The Urban Core's higher-paying jobs also coincide with higher housing costs, more renter-occupied units, and better digital access than most profiles.

The Urban Core's racial and linguistic diversity is a key factor of differentiation from the rest of the US. Only 41.5 percent of the Urban-Core's population is White, which is 19 percentage points less than the nationwide average and 35 percentage points less than the average of counties in the other profiles (see Figure 2 (a)). In contrast, the proportions of Asian, Hispanic or Latino, and Black or African American populations in the Urban Core are markedly larger than the other profiles' average. Figure 2 (b) shows the linguistic diversity of the Urban Core: 35.7 percent of the population uses a language other than English at home, which is 14 percentage points more than the national average and 26.5 percentage points more than the average for counties in the other profiles. Furthermore, 14.8 percent of people in this profile report they speak English less than very well, which is 11.6 percentage points more than in the other profiles.

Figure 2: Race-Ethnicity and Linguistic Diversity in the Urban Core

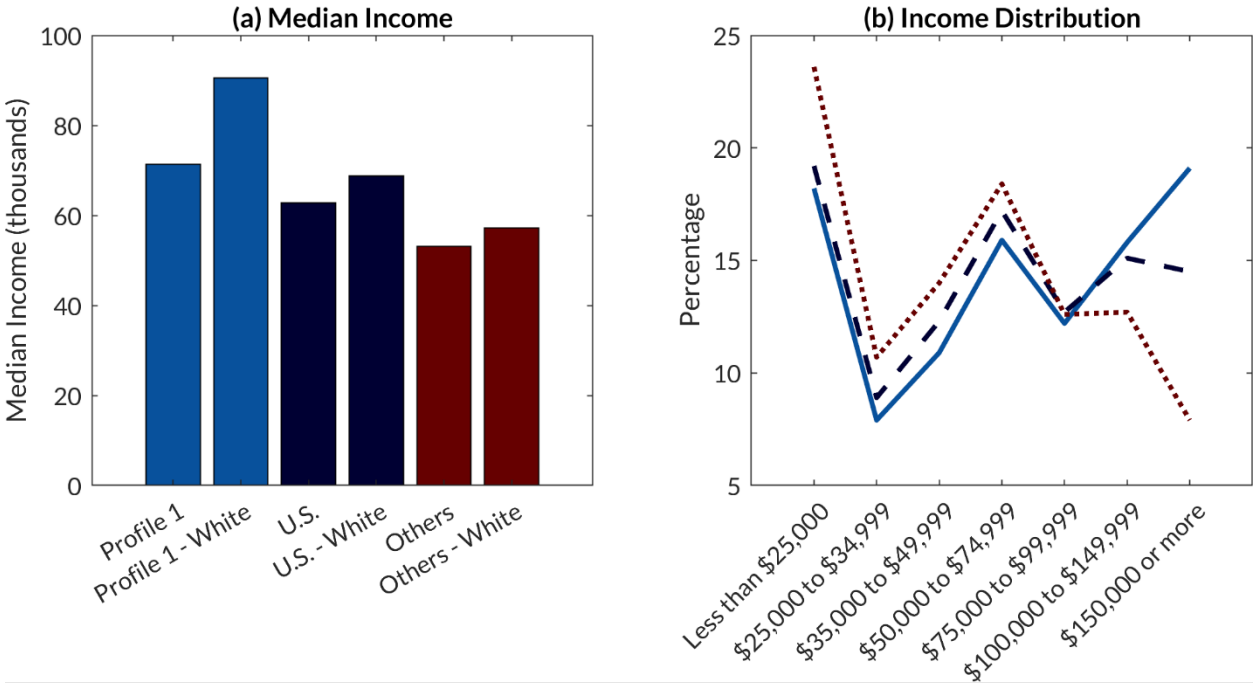


Notes: Panel (a) shows the racial and ethnic profile for the Urban Core, the US, and the average of the counties in all profiles excluding the Urban-Core. The percentage counts members of a race-ethnicity who do not identify as Hispanic or Latino so

that the total can be 100 percent. Panel (b) indicates the percentage of the population that uses a language other than English at home.
Source: Milken Institute (2022)

The economic advantages of the Urban Core areas benefit mainly the highly educated White population. Figures 3 (a) and (b) show that the White population's income drives the overall higher income in the Urban Core. At $90,540, the White population's income is the third-largest across all profiles, falling below only the Affluent Suburbs ($98,659) and Native Alaska ($100,900) profiles. Most (51.6 percent) of the White population in the Urban Core has a bachelor's degree or higher, and (as discussed later) this higher-than-average education is correlated with the higher income for this population.

Figure 3: Income and Income Distribution in the Urban Core

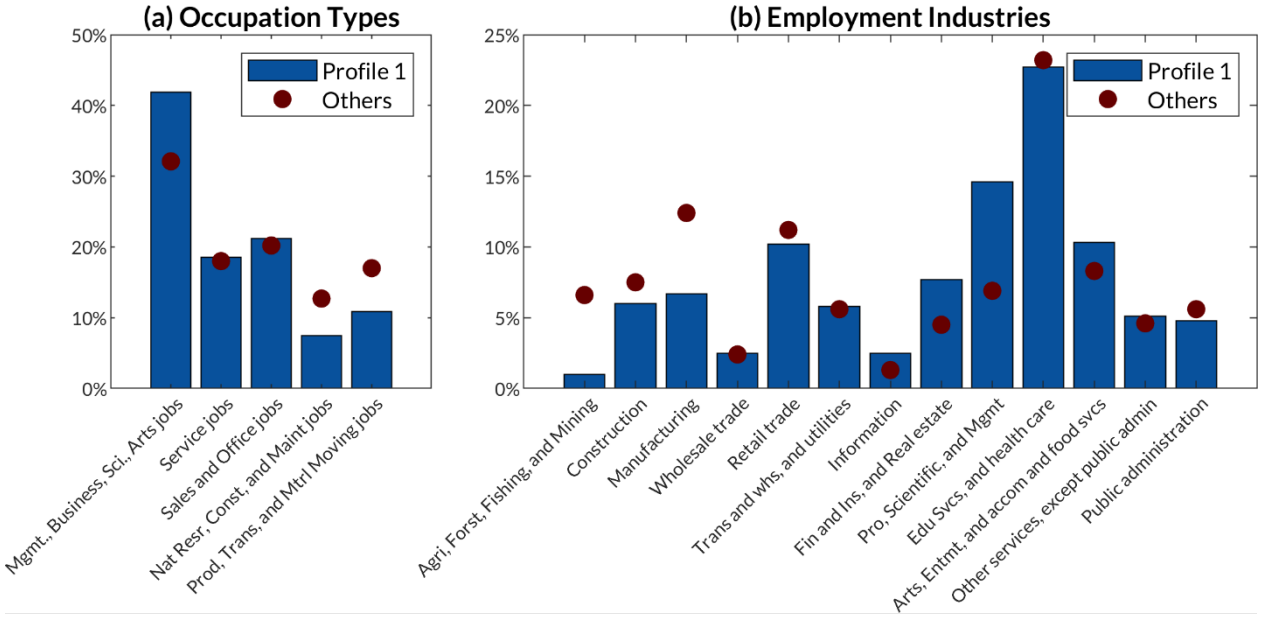


Notes: Panel (a) shows the median income of the Urban Core, the US, and the average of the counties in all profiles excluding the Urban-Core. The category "White" shows the median incomes for the white population. Panel (b) reports the population percentages of the Urban Core, the US, and other profiles for each income bracket. The same colors for bars and lines report information for the same group.

Source: Milken Institute (2022)

These counties offer more jobs in high-paying industries. Among all profiles, the Urban Core has the second-largest (after the Affluent Suburbs) portion of employment in white-collar jobs.[8] This is especially true for "Management, business, science, and arts" jobs (see Figure 4 (a)). These jobs are more concentrated in the top three best-paying industries: "Professional, scientific & management, and administrative & waste management services," "Information," "Finance & insurance, and real estate, rental & leasing" (see Figure 4 (b) and Table 4).

Figure 4: Jobs and Employment Industry in the Urban Core[9]



Notes: Panel (a) shows percentages of the population per job category in the Urban Core and Others (all profiles excluding the Urban-Core). Panel (b) indicates employment per type of industry in the Urban Core and Others.

Source: Milken Institute (2022)

[8] We define white-collar jobs as including the" Management, business, science, and arts" and the "Sales and office jobs" occupations as classified by the US Census Bureau.

[9] The US Census Bureau divides occupations into five categories: Management, business, science, and arts occupations; Service occupations; Sales and office occupations; Natural resources, construction, and maintenance occupations; and Production, transportation, and material moving occupations. Also, employment industries are divided into 13 categories: Agriculture, forestry, fishing and hunting, and mining; Construction; Manufacturing; Wholesale trade; Retail trade; Transportation and warehousing, and utilities; Information; Finance and insurance, and real estate and rental leasing; Professional, scientific, management, and administrative and waste management services; Educational services, and health care and social assistance; Arts, entertainment, and recreation, and accommodation and food services; Other services, except public administration; and Public administration.

Table 4: Average Salary by Industry

Industry Sector	Average Wage
Agriculture, Forestry, Fishing and Hunting, and Mining	$54,998
Arts, Entertainment & Recreation and Accommodation & Food Services	$26,814
Construction	$54,951
Educational Services, Health Care & Social Assistance	$52,666
Finance & Insurance, and Real Estate, Rental & Leasing	*$84,499*
Information	*$79,359*
Manufacturing	$64,861
Other Services, Except Public Administration	$38,552
Professional, Scientific & Management, and Administrative & Waste Management Services	*$75,119*
Public Administration	$66,232
Retail Trade	$37,040
Transportation & Warehousing, and Utilities	$56,463
Wholesale Trade	$66,275

Notes: National average salary for 13 industries in 2019. The top three best-paying industries are italicized.

Source: American Community Survey Public Use Microdata Sample 5-Year Estimate (2019)

The Urban Core has more college graduates than the rest of the US, and they are better compensated for their degrees. However, they also face some of the highest costs of living. Table 5 highlights the higher (relative to other profiles) educational attainments for all races and ethnicities except Hispanics and Latinos in the Urban Core and the gains in income resulting from these post-secondary degrees. It also shows that housing in the Urban Core relies more on renter-occupied units than in the rest of the US, and the related costs are noticeably higher.

The Urban Core has one of the best digital access rates, one of the lowest disability rates, and the longest commutes of all profiles. It has the second-highest rate of access to computers and broadband internet subscriptions and the second-lowest percentage of people with disabilities, all after the Affluent Suburbs

Table 5: Education, Housing, and Infrastructures in the Urban Core

Category	Variable	The Urban-Core	*US*	*Other Profiles*
Education	White with bachelor's or higher (%)	51.6***	*35.8*	*24*
	Black or African American with bachelor's or higher (%)	24.7**	*21.6*	*15.2*
	Asian with bachelor's or higher (%)	56.1**	*54.3*	*41.1*

	Hispanic or Latino with bachelor's or higher (%)	19.5	*16.4*	*14.3*
	Median earnings for college/associate's ($)	39,309**	*37,471*	*34,730*
	Median Earnings for bachelor's ($)	60,272**	*54,925*	*46,474*
	Median Earnings for graduate/professional ($)	80,514**	*74,253*	*58,461*
Housing	Owner-occupied (%)	52.7***	*64*	*71.9*
	Renter-occupied (%)	47.3***	*36*	*28.1*
	SMOCAPI with a mortgage 35% or over (%)	26.2**	*20.9*	*19.1*
	SMOCAPI without a mortgage 35% or over (%)	14.4**	*10.6*	*9*
	GRAPI 35% or over (%)	42.6**	*40.5*	*34.7*
Disability, Computer /Internet, Commuting	Disability (%)	10.6***	*12.6*	*16*
	With a computer (%)	91.9**	*90.3*	*85.3*
	With a broadband Internet subscription (%)	84.7**	*82.7*	*75.3*
	Mean travel time to work (minutes)	30.7**	*26.9*	*23.7*

*Notes: The table compares the average of selected variables with the US average and other profile average. Different race-ethnicity categories count members of a race-ethnicity who do not identify as Hispanic or Latino. SMOCAPI is an acronym for selected monthly owner costs as a percentage of household income. GRAPI denotes gross rent as a percentage of household income. The asterisks indicate that a profile average is statistically different from the US average (denoted as one asterisk, *), from the other profile average (**), and both (***). All values are shown as percentage of the population except the median earnings ($).*

Source: Milken Institute (2022)

US Metropolitan Areas

These three profiles represent the higher, middle, and lower-middle class living mostly in the suburban, medium, and small metropolitan areas of the US.

The Lower-Middle Class, accounts for 320 counties, primarily in medium, small metropolitan, and micropolitan areas. Less populated and less wealthy than counties in the two other U.S. metropolitan areas (Profiles 3 and 4), the Lower-Middle Class counties are home to 18 percent of the US population. While the overall demographic and housing characteristics of the Lower-Middle Class profile are similar to the national average, its median income is lower as there are fewer jobs in high-paying industries and fewer individuals with bachelor's degrees or higher.

Affluent Suburbs groups the 139 counties with the wealthiest neighborhoods of the large suburban and small metro counties with at least 1 million residents. Home to 16 percent of the US population, these counties are the most affluent in the US,

concentrating the population with the highest median income and the highest proportion of university degrees. This population often consists of families who live in an owned house with one adult staying at home and one adult working a white-collar job in a high-paying industry. This profile also has the best digital and health insurance access and the lowest percentage of people with disabilities.

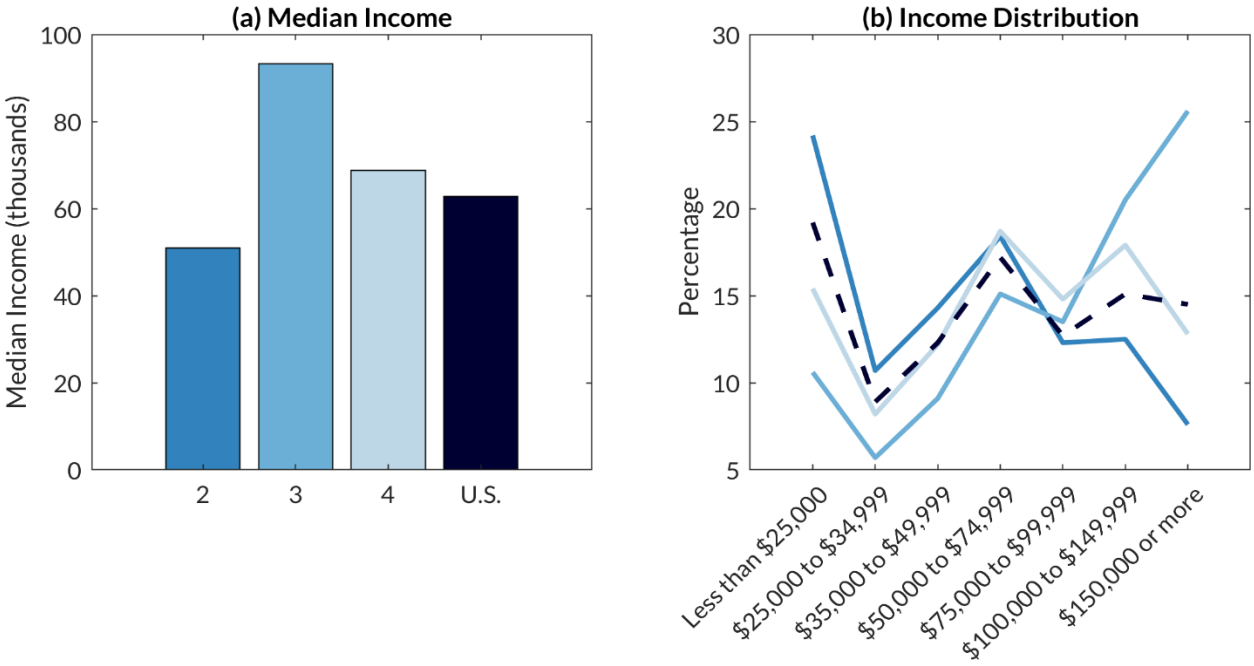
The Middle Class clusters the 311 least racially and ethnically diverse counties of the Mainstream America profiles. Primarily located in large- to medium-size suburban and small metro counties next to the other US metropolitan areas, they are home to 14 percent of the US population. The Middle Class profile's household income structure is similar to the national average, with lower poverty rates and lower income inequality. More people in this profile own their houses and are married than in the rest of the US.

Six variables explain the difference between these metropolitan profiles: income, jobs and employment industries, educational attainment, health insurance coverage, disability, and digital access.

These counties have levels of income that are at and around the national average. Figure 5 (a) shows that the median household incomes for the US metropolitan areas (Lower-Middle Class, Affluent Suburbs, and Middle Class) are below, higher, and at the national median level, respectively. The Affluent Suburbs have the highest median income among all 17 profiles, at $30,447 more than the national median of $62,843.

The income distribution, reported in Figure 5 (b), confirms that income distribution in the Affluent Suburbs is more concentrated in the range greater than $100,000. In contrast, the Lower-Middle Class counties have a greater percentage of households with an income of less than $50,000. The Middle Class counties have a similar income range as the national values.

Figure 5: Income and Income Distribution in US Metropolitan Areas



Notes: Similar colors for bars and lines report the same profile information. Panel (b) indicates what percentages of each profile's population have income falling into specified ranges.

Source: Milken Institute (2022)

Differences exist in employment and education levels across the US Metropolitan Areas. Table 6 shows variables related to education and employment (such as unemployment rates, employment industry and occupation, and educational attainments) for the US metropolitan areas.

The Affluent Suburbs have the lowest unemployment rate among these three profiles and have the highest percentage of "Management, Business, Science, and Arts" jobs among all 17 profiles. The top three best-paying industries—"Professional, scientific & management, and administrative & waste management services," "Information," and "Finance & insurance, and real estate, rental & leasing"—also occupy a larger share of the labor market in the Affluent Suburbs (Tables 4 and 6). People in this profile are highly educated, with the percentage of the population holding a bachelor's degree or higher being 12 percentage points above the national average.

The Lower-Middle Class profile has significantly fewer jobs in high-paying industries than the rest of the US. Compared to the national average, the Lower-Middle Class counties also have a lower education level, with a smaller proportion (by 8

percentage points) of the population with a bachelor's degree or higher. Finally, the Middle Class counties are the most similar to the national average, with none of the variables (except for one, lowest educational attainment) in Table 6 being statistically significantly different from the national averages, and all of them ranging between the values of the other US metropolitan areas (Lower-Middle Class and Affluent Suburbs).

Table 6: Employment and Education in US Metropolitan Areas

Category	Variable	Lower-Middle Class	Affluent Suburbs	Middle Class	*US*
Employment	Unemployment Rate	6.2	3.9*	4.2	*5.3*
	Occupations: Management, Business, Science, Arts	33*	46.3*	37.1	*38.5*
	Industry: Information	1.4*	2.2	1.5	*2*
	Industry: Finance and Insurance, and Real estate	4.9*	7.6	*6*	*6.6*
	Industry: Professional, Scientific, and Management	8.2*	14.4	9.3	*11.6*
Education	Less than 9th grade	4.2	3.1	2.9*	*5.1*
	9th to 12th grade, no diploma	8	4.1*	5.7	*6.9*
	High school graduate	31.3	21.3*	29.5	*27*
	Bachelor's degree	15.3*	26.7*	19.2	*19.8*
	Graduate or professional degree	8.9*	17.5	10.6	*12.4*
	White, not Hispanic or Latino, Bachelor's or higher	27.1*	47.1*	31.3	*35.8*
	Field of bachelor's degree: Science and Engineering	29.9*	37.5	31.6	*35.1*
	Field of bachelor's degree: Education	17.7*	10.7	16.1	*12.2*

Notes: The table shows averages for selected variables that distinguish the US Metropolitan Areas profiles. The asterisk indicates that a profile average is statistically different from the US average. All values are in percentage of the population.

Source: Milken Institute (2022)

Differences also exist in health insurance, disability, and digital access. Similar patterns emerge by looking at the distributions of health insurance, disability, and computer access across the metropolitan areas (see Table 7). The fraction of people with disabilities is lowest in the Affluent Suburbs and highest in the Lower-Middle Class counties. Similarly, the ratio of households having access to computers, high-speed services, and health insurance is highest in the Affluent Suburbs and lowest in the Lower-Middle Class profile, illustrating the respective affluence (and lack thereof) of

these profiles. Again, none of the statistics reported in Table 7 for the Middle Class profile are significantly different from the national averages, and all are within the range of the other US metropolitan profiles.

Table 7: Health Care and Digital Access in US Metropolitan Areas

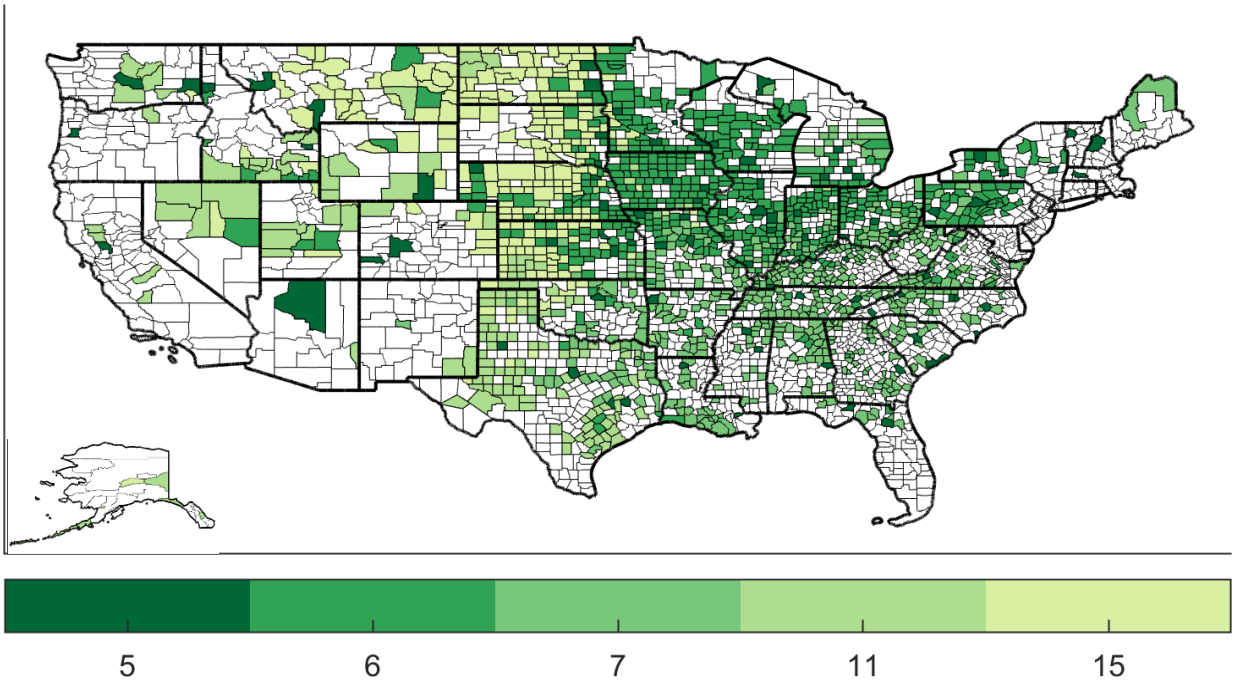
Category	Variable	Lower-Middle Class	Affluent Suburbs	Middle Class	*US Average*
Health	With health insurance	90.8	93.9	92.8	*91.2*
	Disability	15.6*	9.5*	12.4	*12.6*
Computer /Internet	With a computer	87.9	94.5*	91.5	*90.3*
	With a broadband internet subscription	78.8	89.8*	84.1	*82.7*

Notes: The table shows averages for selected variables that distinguish the US Metropolitan Areas profiles. The asterisk indicates that a profile average is statistically different from the US average. All values are shown as percentage of the population.

Source: Milken Institute (2022)

Industry-Driven America

Figure 6: Map of Industry-Driven America



Source: Milken Institute (2022)

Figure 6 highlights the 1,507 counties, home to 17.6 percent of the US population, whose industrial concentration shapes their population profiles. Specifically, the occupations driving these profiles are education for College Towns, manufacturing for Manufacturing Midwest and Low-Wage Manufacturing, and agriculture for Hispanic Agriculture and the Great Plains.[10]

College Towns groups 98 counties, 5.4 percent of the population, located mostly in suburban and metro areas that are home to the most sizeable universities in the country. Almost one-third of the labor force in this profile works in the educational sector, representing the largest concentration of labor in a single employment sector in the US.

[10] Helper et al (2012) identify six broad groups defined by common patterns of manufacturing industry employment composition. Each group is defined by an anchor industry or combination of industries, in which all metropolitan areas in the group are relatively strongly (usually highly) specialized, and by another industry in which all metropolitan areas in the group are less specialized. The six anchor manufacturing industries are computers and electronics (West in general; California, Colorado, New England), transportation equipment (including motor vehicles and parts, aerospace, and other transportation equipment), low-wage manufacturing industries (a broad category that combines food, textile mills, textile product mills, apparel, leather, wood, and furniture), chemicals, machinery, and food.

Due to the large student population, the residents of this profile are generally young; they come from another county, state, or country; and have a median household income lower than the national median (significantly lower for Asians). This profile has the highest level of enrollment in post-secondary education and the second-highest educational attainments of all profiles. More of the population in this profile rents their houses than on average in the US.

Manufacturing Midwest includes 506 mostly Midwestern counties, 5.2 percent of the US population, that represent some of the least diverse areas, with over 91 percent of their population being White. Population in the Manufacturing Midwest is primarily employed in manufacturing industries, specialized in manufacturing transportation equipment (including motor vehicles and parts, aerospace, and other transportation equipment) and machinery. Residents have more access to job-related benefits, such as health insurance, than on average in the US, while the level of qualification and resulting income are lower. These communities maintain low unemployment rates (especially for high school graduates), low housing costs, and less income inequality compared to the average of the country.

Low-Wage Manufacturing clusters 524 counties, 4.9 percent of the US population, with the second-highest concentration of manufacturing jobs, after the Manufacturing Midwest (Profile 6). These communities are primarily located in the South with more challenging overall conditions, ranging from lower income and education levels to higher poverty rates and worse access to digital infrastructure relative to other Industry-Driven America profiles.

Hispanic Agriculture groups 158 counties, 1.2 percent of the US population, that have the second-largest concentration of jobs in the "Agriculture, forestry, fishing and hunting, and mining" industries. These communities have a prominent Hispanic or Latino population, which represents more than 30 percent of the population. They report

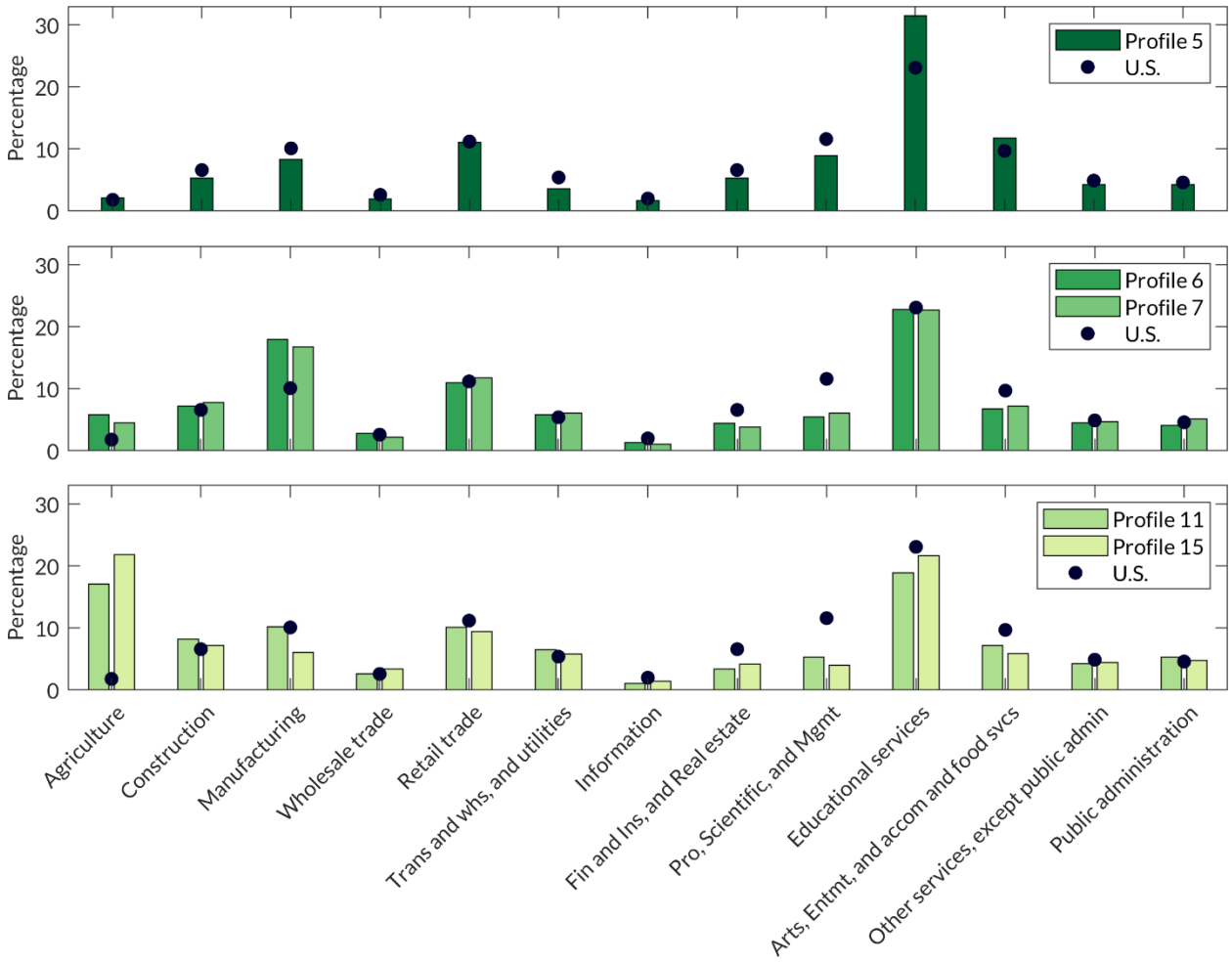
below-average levels of education and access to health insurance and the internet compared to the average of the US.

The Great Plains includes 221 counties, 0.3 percent of the US population, that are rural and primarily located in the Great Plains. These communities have the highest concentration of jobs in wheat production (21.8 percent) and among the highest percentage of jobs in natural resources, construction, and maintenance (46.4 percent). With the second-largest concentration of White population (90.8 percent of the population), these communities have the lowest unemployment rate, the second-lowest ratio of people receiving Supplemental Nutrition and Assistance Program (SNAP) (after the Affluent Suburbs), and the third-lowest poverty rate of all profiles.

Four variables can articulate the differences between these five Industry-Driven America profiles: employment industry, race/ethnicity, income, and education. These profiles are also distinguished by other social and digital components, such as the proportion of the foreign-born population and access to a computer.

One industry stands out from the 13 employment industries defined by the US Census Bureau for each profile. Figure 7 summarizes the percentages of workers in a specific industry in each profile and compares them to the national average. As shown in the top panel, College Towns have the highest percentage of the population (31.5 percent) working in education, with sizable universities in its counties. The distribution of the other industries is in line with the national one. The second panel shows that approximately 18 percent of the population in the Manufacturing Midwest and 17 percent in the Low-Wage Manufacturing profiles work in manufacturing industries, the largest ratios among all profiles. These profiles have a relatively low ratio of workers in the professional, scientific, and management industries, with employment ratios in these industries being about 6 percentage points below the national average. Finally, the bottom panel indicates that jobs in the Hispanic Agriculture and the Great Plains profiles are concentrated in the agricultural industry.

Figure 7: Employment Industries[9] in Industry-Driven America



Notes: These figures show the types of industries that employ people in the Industry-Driven America profiles and the total population in the US. Bars denote percentages of the adults who work for a specific industry. The 13 employment industries denote Agriculture, forestry, fishing and hunting, and mining; Construction; Manufacturing; Wholesale trade; Retail trade; Transportation and warehousing, and utilities; Information; Finance and insurance, and real estate and rental and leasing; Professional, scientific, and management, and administrative and waste management services; Educational services, and health care and social assistance; Arts, entertainment, and recreation, and accommodation and food services; Other services, except public administration; and Public administration.

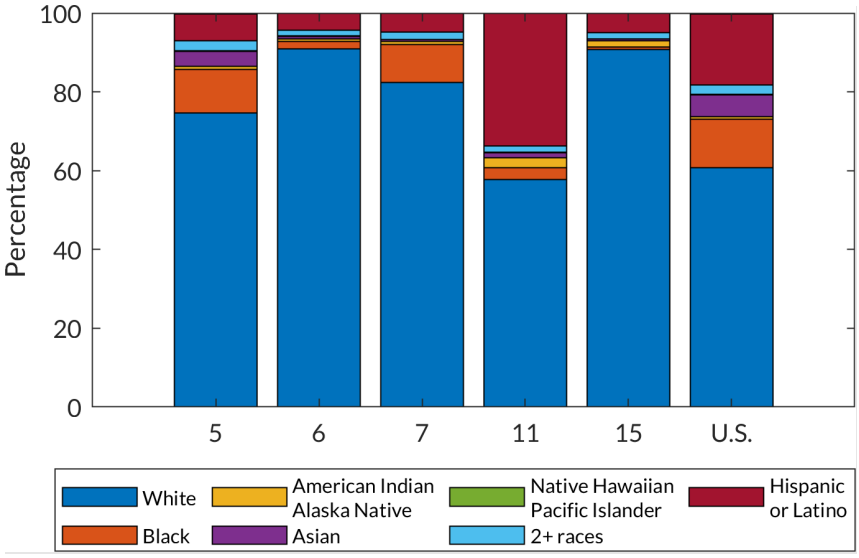
Source: Milken Institute (2022)

The Industry-Driven America profiles are characterized by differences in race and ethnicity. Figure 8 shows the racial and ethnic differences across these profiles. Among the manufacturing-driven profiles, the Manufacturing Midwest has a large ratio of the White population (the highest of all profiles), while the Low-Wage Manufacturing profile has a larger percentage of Black or African American population (relative to the

Manufacturing Midwest profile) as it encompasses the South. The ratio of Hispanic or Latino population is significantly lower in both profiles than in the average of the US.

Similarly, the population distribution strongly differs among the agricultural profiles. Communities in the Hispanic Agriculture profile have the second-largest Hispanic or Latino population ratio (33.6 percent) after the Hispanic Southern Border (73.2 percent). In contrast, communities in the Great Plains have the second-largest percentage of the White population (90.8 percent) after the Manufacturing Midwest (91 percent). The racial makeup of communities in the College Towns is similar to the national average, except for an 11 percentage points lower ratio of the Hispanic or Latino population and a higher proportion of the White population.

Figure 8: Race-Ethnicity in Industry-Driven America



Notes: The percentage counts members of a race ethnicity who do not identify as Hispanic or Latino so that the total can be 100 percent.

Source: Milken Institute (2022)

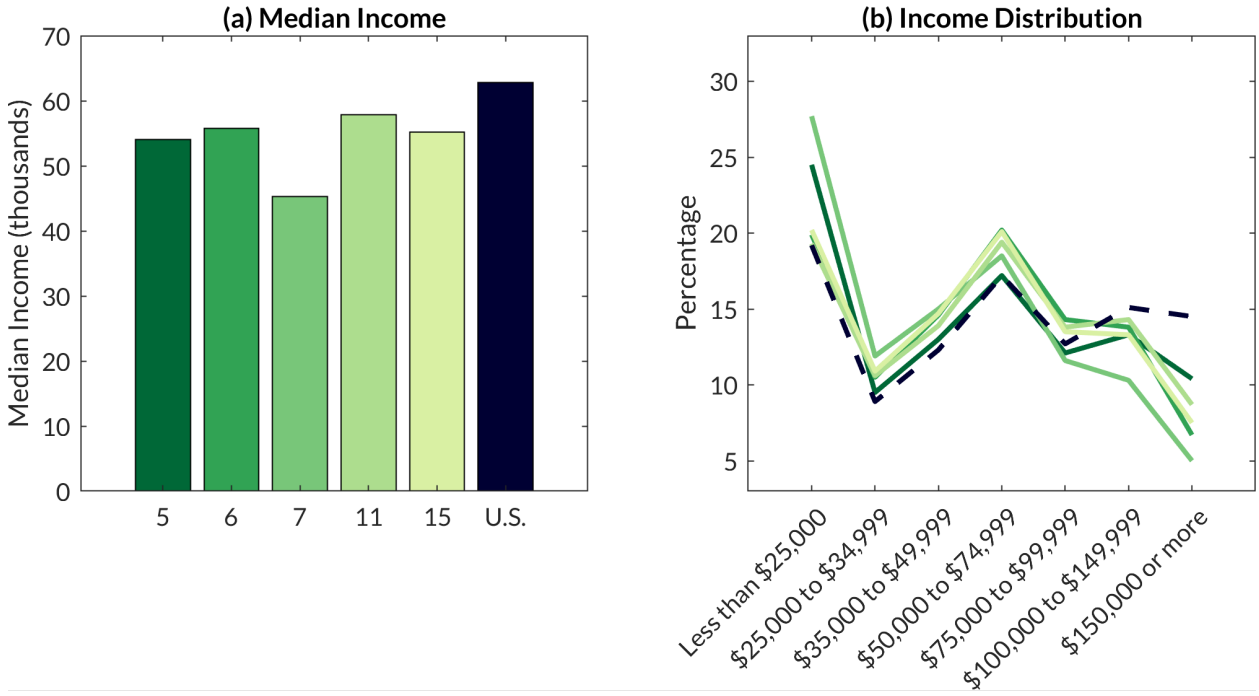
Differences also exist in income levels that correspond to the different industries that drive these profiles. The household median incomes for the College Towns, Hispanic Agriculture, and the Great Plains profiles in Figure 9 (a) are in line with the average industry salaries reported in Table 4. The College Towns' median income is close to the $52,666 for educational services and health care and social assistance, and

median incomes in the Great Plains and the Hispanic Agriculture profiles are close to the $54,998 for agriculture, forestry, fishing and hunting, and mining.

Differences in manufacturing specializations lead to significantly different income levels for the Manufacturing Midwest and Low-Wage Manufacturing profiles, which also differ from the national average (Helper, Krueger, and Wial 2012). The average national salary reported in Table 4 (at $64,861) accounts for high-technology manufacturing jobs in computers and electronics that are not part of the Manufacturing Midwest and Low-Wage Manufacturing profiles. The Manufacturing Midwest specializes in the manufacturing of transportation equipment[11] and machinery, resulting in a lower median income for this profile at $55,748, or $9,113 less than the national average for the manufacturing sector. Similarly, the Low-Wage Manufacturing profile has a median income of $45,249, or $17,594 lower than the national manufacturing average, reflecting its counties' specialization in low-wage manufacturing industries[12] and chemicals other than pharmaceuticals. Figure 9 (b) confirms that the income distribution of the Low-Wage Manufacturing profile is more concentrated in the ranges below $50,000 and much less so in the ranges greater than $100,000 compared to the national distribution. This profile also has higher poverty rates compared to the other Industry-Driven America profiles and the rest of the country.

[11] This includes manufacturing of motor vehicles and parts, aerospace, and other transportation equipment.
[12] This is a broad category that combines manufacturing of food, textile mills, textile product mills, apparel, leather, wood, and furniture.

Figure 9: Income and Income Distribution in Industry-Driven America



Notes: Panel (a) shows the median income of the Industry-Driven America profiles and the national median. Panel (b) indicates what percentages of the population in each profile have income falling into specified ranges. The line color in panel (b) matches the one on panel (a).

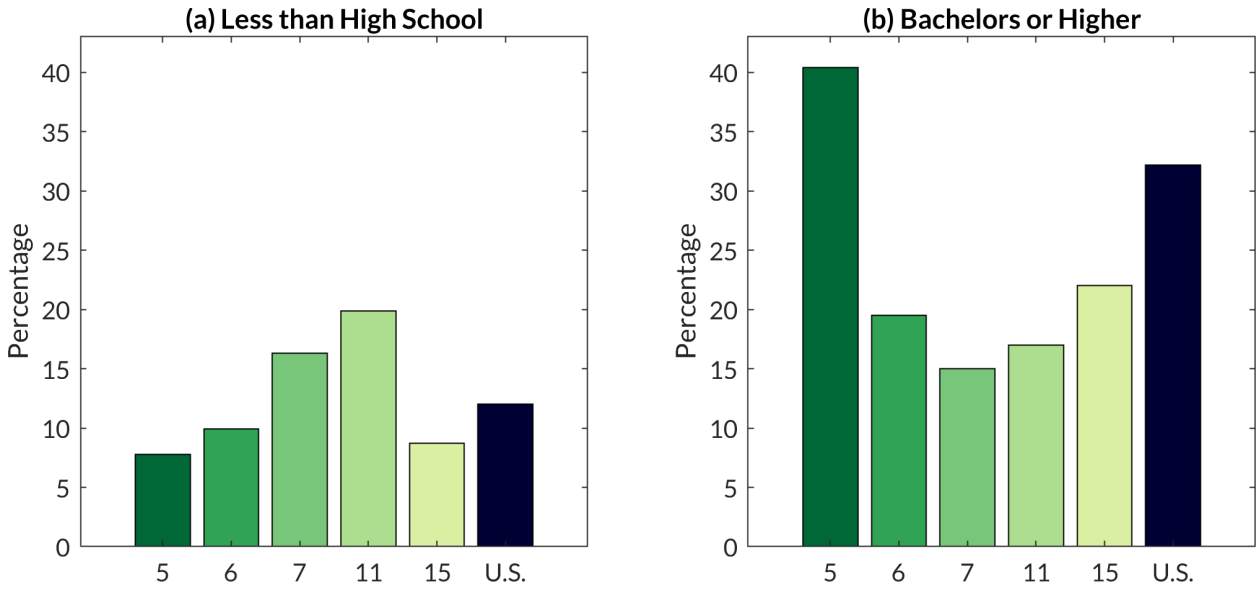
Source: Milken Institute (2022)

Levels of educational attainment differ across the Industry-Driven profiles. Figure 10 highlights the relatively high percentage of the population with a bachelor's degree or higher in College Towns, exceeding the national average. Yet, the population with post-secondary degrees in College Towns is less compensated for its high education: The median income in this profile is $44,474 for bachelor's degree holders ($10,451 less than the national average) and $60,134 for graduate degree holders ($14,119 less than the national average).

The Manufacturing Midwest and Low-Wage Manufacturing profiles have relatively high ratios of the population whose highest degree is a high school diploma: 38.2 and 39.8 for the Manufacturing Midwest and Low-Wage Manufacturing, respectively, as compared to 27 for the country on average. The Hispanic Agriculture profile, with a larger Hispanic or Latino population, has the lowest educational achievements among the Industry-Driven America profiles: Almost 20 percent of its

residents do not have a high school diploma (8 percentage points more than the country average), and only 17 percent hold a bachelor's degree or higher (15 percentage points less than the national average).

Figure 10: Education in Industry-Driven America



Notes: Panel (a) shows ratios of the population who did not complete high school. Panel (b) indicates the fraction of the people who hold a bachelor's degree or higher.

Source: Milken Institute (2022)

These profiles are distinguished by other noteworthy characteristics. The Hispanic Agriculture and College Towns profiles have the first and third-largest ratios among all profiles of foreign-born residents who are not US citizens. However, the College Towns frequently use English at home, with 89.3 percent of English-speaking homes. This is much higher than the percentage of homes primarily using English in the Hispanic Agriculture profile (71.7 percent) and the national average (78.4 percent). Finally, all manufacturing and farming communities have limited digital access.

Table 8: Other Characteristics of Industry-Driven America

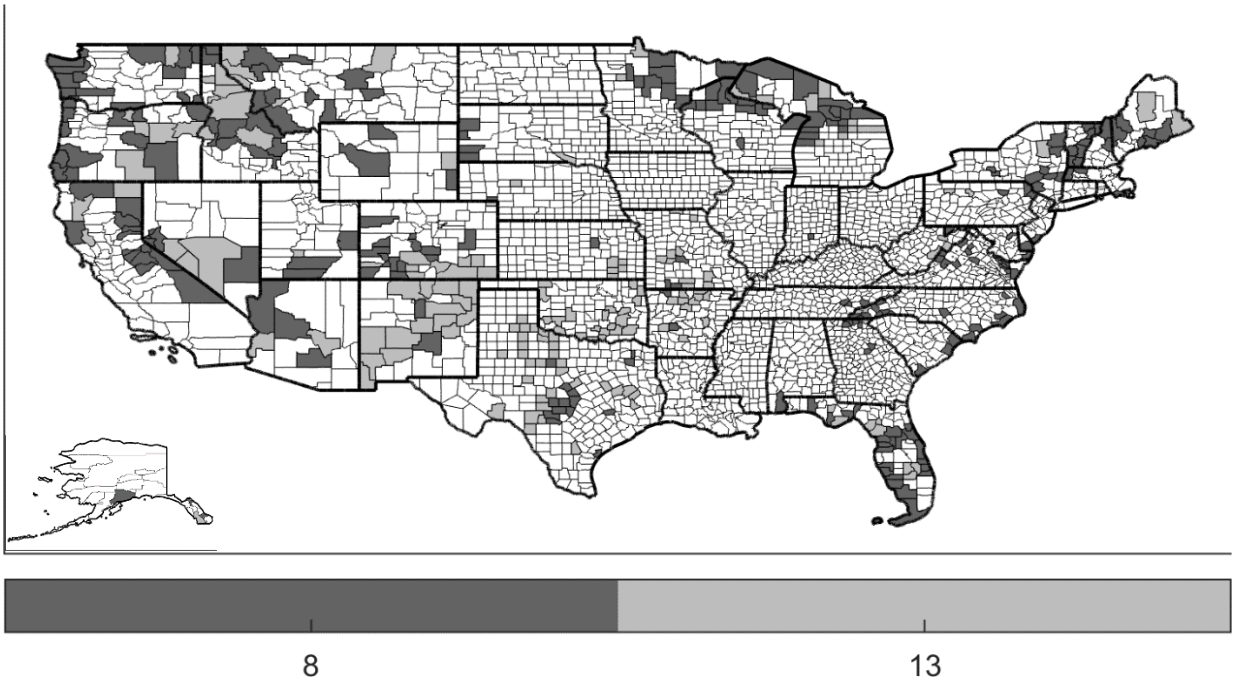
Category	Variable	College Towns	Manuf. Midwest	Low-Wage Manuf.	Hispanic Agric.	Great Plains	*US*
Social	Foreign-born population, Not a US citizen	62.7*	52.5	56.5	67.8*	55	*50.4*
	Language at home not English	10.7*	4.8*	4.8*	28.3	5.4*	*21.6*
Household	Married-couple family	42.4	53*	50.9	55.5*	55.4*	*48.2*
Health	With health insurance	92.6	93.4	89.7	83.3*	91.8	*91.2*
	Disability	11.5	14.4	19.2*	13.2	14.1	*12.6*
Computer /Internet	With a computer	91.7	86.1*	82*	87.1	85.6*	*90.3*
	With a broadband internet subscription	82.7	77.5*	70.7*	75.2*	75.6*	*82.7*

Notes: The asterisk indicates that a profile average is statistically different from the US average. All values are shown as percentage of the population.

Source: Milken Institute (2022)

Graying America

Figure 11: Map of Graying America



Source: Milken Institute (2022)

These 424 counties, home to 5.1 percent of the US population, have more than 40 percent of households with people age 65 and older. The Retiree Communities and Isolated Seniors profiles group these graying communities based on income level and living conditions.

Retiree Communities includes 256 counties, 4.5 percent of the US population, where the primarily White middle-class retiree communities drive part of the local economy. These communities have the highest ratio of civilians who formerly served in the military of all profiles. While the youngest and oldest residents in this profile (those 15 to 24 years old and 65 years and older) have incomes in line with the US average for those age groups, the rest of its population (those 25 to 64 years old) is worse off.

Isolated Seniors consists of 168 counties, 0.6 percent of the US population (2 million people), with a large portion of older households with lower incomes than the rest of the

US. These communities have lower levels of education and more low-skilled agricultural jobs compared with the national average. Older people (65 years and older) are most likely to live alone in this than in any other profile. At the same time, the percentage of people living with disabilities is the second-largest (after White Appalachia). Finally, access to digital infrastructure is a concern among the population living in the counties covered in the Isolated Seniors profile.

The Isolated Seniors profile has a higher percentage of rural counties (97 percent) than the Retiree Communities (75.4 percent). Income levels, disability rates, and the percentage of seniors living alone also differentiate these two profiles.

It is not all about Florida. Florida has long attracted retirees and has been one of the nation's grayest states, as Figure 11 confirms. However, our two profiles of Graying America tell more profound stories about retiree havens and pinpoint where the 65 years and older population is retiring. Table 3 lists the counties with the largest percentage of population ages 65 and above that compose the Retiree Communities and Isolated Seniors profiles.

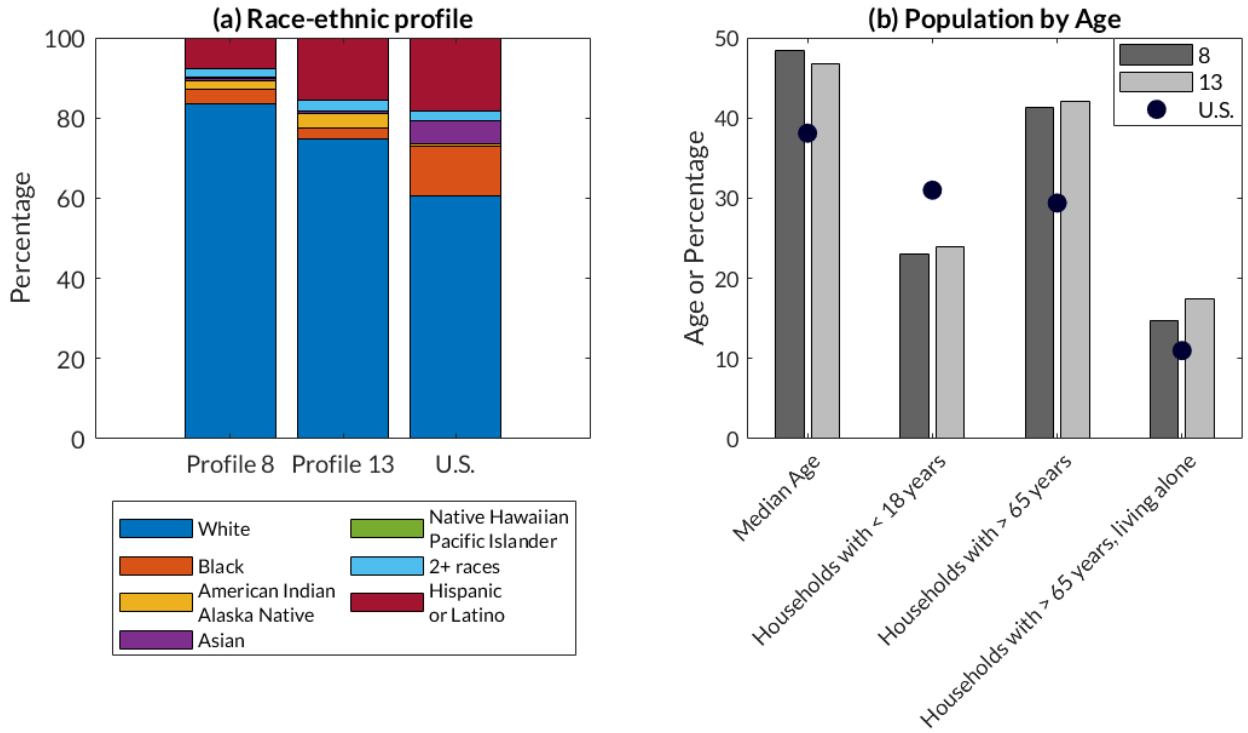
Table 3: Counties with Largest Percentage of Population Age 65+ in Graying America

County	State	Total Population	Ages 65+ (%)	Profile
Sumter	FL	125,044	56.7	Retiree Communities
Charlotte	FL	181,067	39.6	Retiree Communities
Harding	NM	441	39	Isolated Seniors
Highland	VA	2,204	38.9	Retiree Communities
La Paz	AZ	20,793	38.6	Isolated Seniors
Catron	NM	3,526	37	Isolated Seniors
Northumberland	VA	12,190	36.7	Retiree Communities
Llano	TX	21,047	36.4	Retiree Communities
Citrus	FL	145,169	36.3	Retiree Communities
Lancaster	VA	10,724	36.2	Retiree Communities

Source: Milken Institute (2022)

Differences exist in race, ethnicity, and age distribution among the two Graying America profiles. Figure 12 (a) shows that both profiles are predominantly White. Yet, the Isolated Seniors profile has a higher percentage of the Hispanic and Latino population, leading to a relatively lower White representation (74.7 percent, compared to 83.6 percent for the Retiree Communities). The Retiree Communities and Isolated Seniors profiles also have the two oldest median ages at 46.8 and 48.4, respectively, which are more than seven years higher than the median age of the total US population. The high median age of these profiles impacts their entire age distributions: These profiles have at least 7 percentage points fewer households with people 18 years and younger, and 12 percentage points more households with people 65 years and older than the national average.

Figure 12: Race-Ethnicity, and Age in Graying America



Notes: Panel (a) shows the race-ethnicity for the Graying America profiles and the US. The percentage counts members of a race-ethnicity who do not identify as Hispanic or Latino so that the total equals 100 percent. Panel (b) indicates the median age and the ratio of households aged under 18 years or 65 and over. It also denotes the percentage of households with seniors living alone.

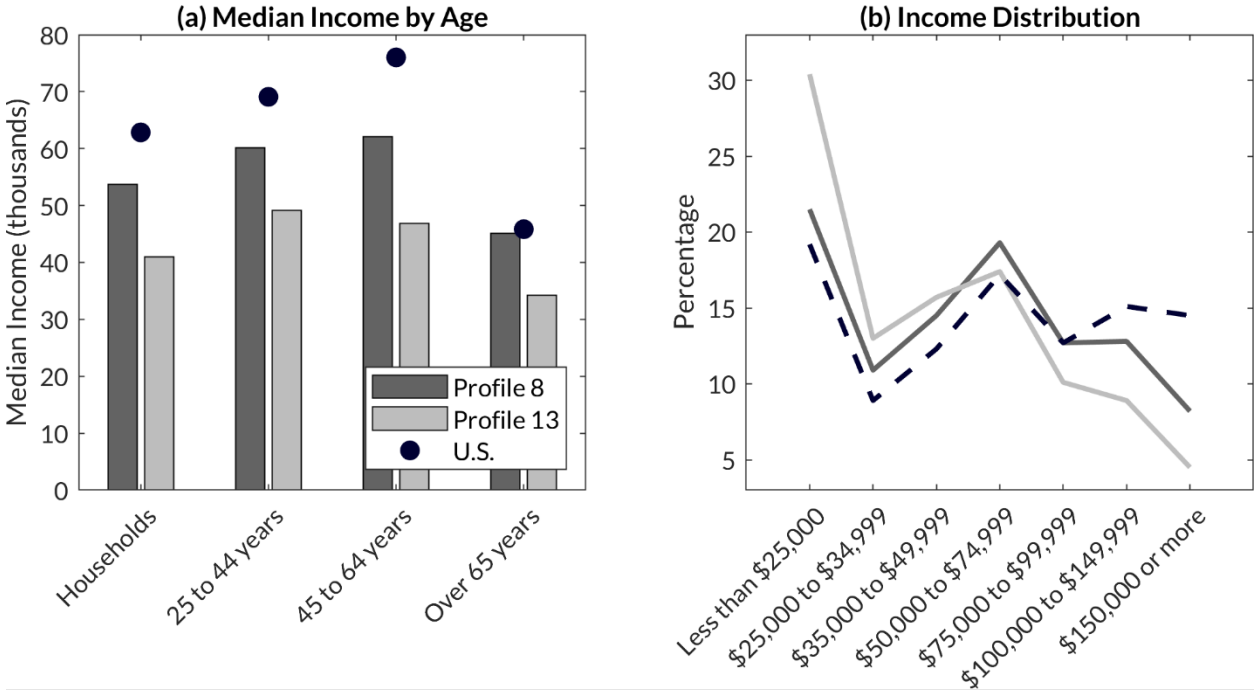
Source: Milken Institute (2022)

Income differences also exist across these two profiles. Figure 13 (a) indicates that, compared to the national average, the median income for the working-age group (ages 25 to 64) in the Retiree Communities is below average, while the median income for people aged 65 years or above is in line with the national average. In contrast, the median household income in the Isolated Seniors profile is lower than the national median income for all age categories, with a higher concentration of incomes below the poverty line.

Figure 13 (b) highlights the difference in income distributions between the two profiles: 30.4 percent of the Isolated Seniors households have an income below $25,000, which is 11 percentage points more than the national average and 9 percentage points more than the average of the Retiree Communities. The flip side of the same pattern emerges for the higher income range: 13.4 percent of the Isolated Seniors' population

has an income higher than $100,000, which is 16 percentage points less than the national average, and 8 percentage points less than the average of the Retiree Communities.

Figure 13: Demography and Income in Graying America



Notes: Panel (a) shows median household income overall and by age. Panel (b) indicates what percentages of the population in each profile have income falling into specified ranges. The line color in panel (b) matches each profile's bar color on panel (a).

Source: Milken Institute (2022)

More people are self-employed in these profiles. The Isolated Seniors profile has fewer private wage and salary workers than the Retiree Communities since more government employees and self-employed workers reside in the counties in the Isolated Seniors profile. Both profiles have significantly fewer (relative to the national average) jobs in the top three high-paying industries: "Information," "Finance & insurance, and real estate, rental & leasing," and "Professional, scientific & management, and administrative & waste management services" (see Tables 4 and 9).

Post-secondary degrees are less common and less rewarded than in the rest of the country. The Isolated Seniors profile has 8.6 percentage points fewer bachelor's

degree holders and 6.9 percentage points fewer graduate degree holders than the US average. Both profiles' compensation for higher degrees is significantly less than the national median: Earnings with a bachelor's degree are $11,605 and $14,504 below the national median in the Retiree Communities and Isolated Seniors profiles, respectively. Holders of graduate or professional degrees earn significantly less that the corresponding national median, with earnings of $17,625 and $24,897 below the national median, respectively. The median earnings for all higher education levels of people living in the Isolated Seniors profile are the lowest among all profiles.

Table 9: Industry and Education in Graying America

Category	Variable	Retiree Communities	Isolated Seniors	*US*
Employment	Information	1.5	1*	*2*
	Finance and Insurance and Real estate	4.8*	3.7*	*6.6*
	Professional, scientific, and management	8.3*	5.5*	*11.6*
	Private wage and salary workers	72.8*	*67.7**	*80.2*
	Government workers	17.2	21.7*	*13.7*
	Self-employed	9.7*	10.2*	*5.9*
Education	Bachelor's degree	16.2	11.2*	*19.8*
	Graduate or professional degree	9.4	5.5*	*12.4*
	Median earnings with some college/associate's	32,835*	29,223*	*37,471*
	Median earnings with bachelor's	43,320*	40,421*	*54,925*
	Median earnings with graduate/professional	56,628*	49,356*	*74,253*

Notes. The asterisks indicate that a profile average is statistically different from the US average. All values are shown as percentage of the population except the median earnings ($).

Source: Milken Institute (2022)

More veterans live in these profiles and disability rates are high. Table 10 shows that both profiles have a significantly larger veteran population than the rest of the country: The Retiree Communities have the highest percentage of the veteran population of all the profiles.

The communities in the Isolated Seniors profile are more likely to live with disabilities as the rates, overall and for four of the six different types of disabilities that ACS surveyed, are the second-highest after the White Appalachia (which represents

predominantly White communities with high poverty levels). They also have significantly less access to computers and quality internet services than the rest of the country.

Finally, housing vacancy rates for both profiles are among the highest of all profiles.

Table 10: Other Characteristics of Graying America

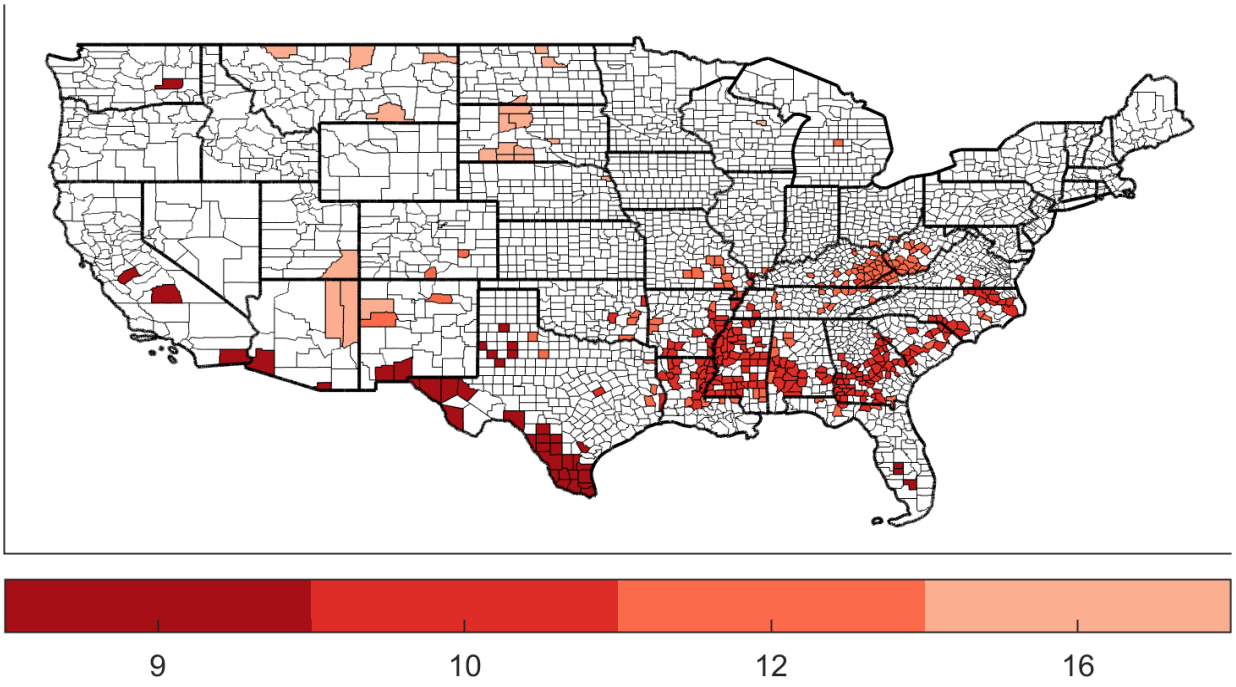
Category	Variable	Retiree Communities	Isolated Seniors	*US Average*
Social	Civilian veterans	11.2*	10.7*	*7.3*
Housing	Vacant housing units	31.3*	34.4*	*12.1*
	Owner-occupied	75.6*	74.7*	*64*
Household Type	Grandparents responsible for grandchildren	42.4	55.8*	*34.1*
	Enrollment, Elementary school (1-8)	44.8	47.6*	*40.4*
Disability, Computer /Internet	With health insurance	90.6	87.3	*91.2*
	Disability	17*	21.8*	*12.6*
	With a hearing difficulty	6.1*	8*	*3.6*
	With a vision difficulty	2.9	4.6*	*2.3*
	With a cognitive difficulty	6.1	7.9*	*5.1*
	With an ambulatory difficulty	9.2*	12.9*	*6.9*
	With a self-care difficulty	3.2	4.2*	*2.6*
	With an independent living difficulty	6.8	9*	*5.8*
	With a computer	88.1	80.9*	*90.3*
	With a broadband internet subscription	78.8	66.8*	*82.7*
	No computer, 65 years and over	17.1	27.2*	*18.1*

Notes: The table shows averages of selected variables that distinguish the Graying America profiles. The asterisks indicate that a profile average is statistically different from the US average. All values are shown as percentage of the population.

Source: Milken Institute (2022)

Extremely Vulnerable America

Figure 14: Map of Extremely Vulnerable America



Source: Milken Institute (2022)

These 378 counties, where 3.5 percent of the US population resides, are primarily rural (85 percent of their population), with widespread poverty. The Extremely Vulnerable America profiles (Hispanic Southern Border, Black South, White Appalachia, and American Indian Reservations) significantly lag the rest of the US regarding income, education, employment, and essential infrastructures. These communities are in regions with above-average percentages of disadvantaged groups of diverse ethnic backgrounds. Such racial or ethnic regional differences distinguish these four profiles.

Hispanic Southern Border includes 43 counties, 1.4 percent of the US population, primarily located close to the US southern border. These young, mostly Hispanic or Latino communities have the lowest English proficiency, among the lowest income levels, and the lowest attainments of compulsory education of all profiles. Compared to other profiles, more workers in these communities have low-skilled jobs in the service

and agricultural industry. These communities have low access to digital infrastructure and health insurance.

Black South clusters 198 counties, 1.3 percent of the US population, located mostly in the South, encompassing a stretch of counties from Virginia down through the Deep South and including parts of Arkansas. These largely Black or African American communities (46.3 percent on average) are historically poor. They remain extremely vulnerable, with lower education levels and the lowest income and highest income inequality of all profiles. Compared to other profiles, more workers in these communities have low-skilled jobs in the manufacturing industry. These communities have diminished access to digital infrastructure and health insurance. Finally, the Black South has the second-lowest ratio of married-couple families and the highest ratio of single female parents of all profiles.

White Appalachia groups 115 counties, 0.7 percent of the US population, populated by primarily White communities (84.7 percent on average). These communities have the second-lowest median income, a high poverty rate, and a very high unemployment rate, which is the third-largest after the American Indian Reservations and Native Alaska profiles. White Appalachia has the highest unemployment rates among the White population of all profiles. More people have blue-collar jobs in the agriculture and manufacturing industries and lower educational attainments than the national average. The percentage of people living with disabilities is the highest among all profiles, while the access to digital infrastructure is very limited.

American Indian Reservations comprises 22 counties, 0.1 percent of the US population, where the majority (67 percent) of the population belongs to the American Indian and Alaska Natives racial or ethnic categories. These communities have the highest poverty rate among all the profiles (36 percent for individuals and 29 percent for families) and the second-highest percentage of households receiving SNAP benefits (26 percent). The

unemployment rate is the second-highest at 13 percent, falling below only that of the Native Alaska profile (with an unemployment rate of 16 percent). An abnormally large percentage of the population in the American Indian Reservations profile works for the government (43.4 percent). These communities have the US's lowest health insurance coverage and digital access.

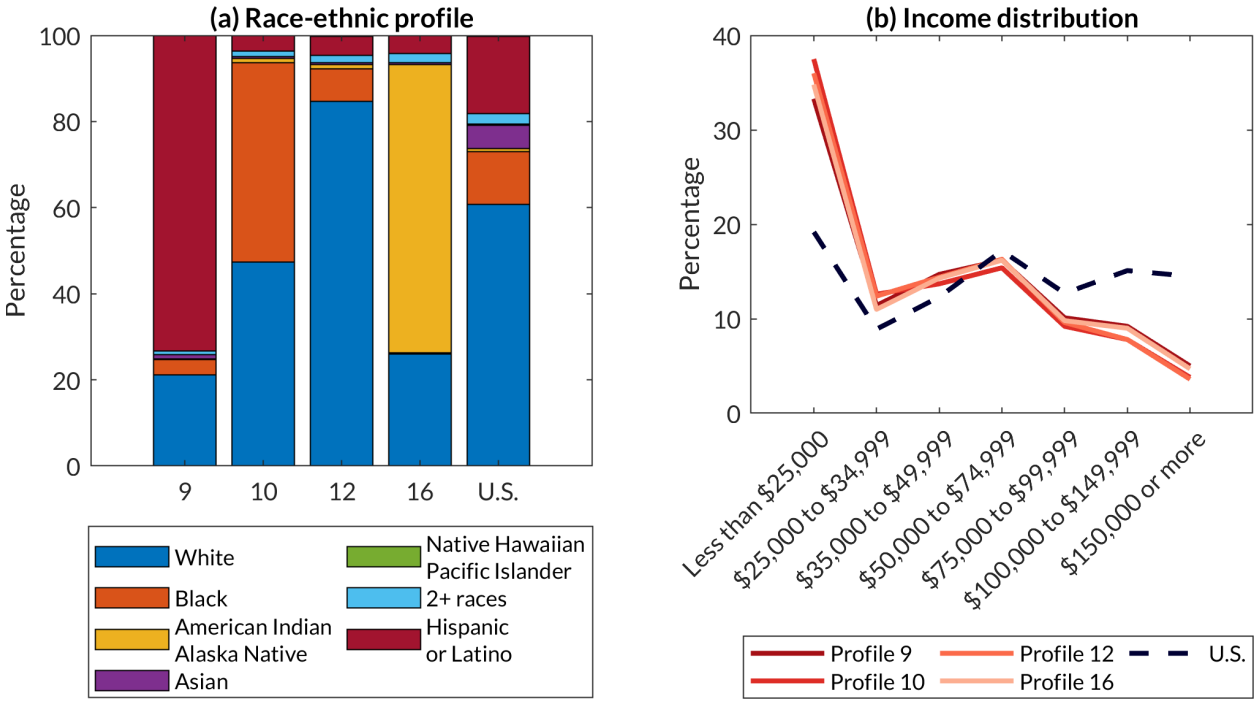
Extreme poverty is a common factor, while the racial and ethnic profiles differ. Each of these profiles is characterized by the prominence of a racial or ethnic group: The Hispanic or Latino population represents 73.2 percent of the Hispanic Southern Border, the Black and African American population represents 46.3 percent of the Black South; the White population represents 84.7 percent of the White Appalachia, and the American Indian or Alaska Native population represents 67 percent of the American Indian Reservations (see Figure 15 (a)). Except for the White Appalachia, these profiles comprise the most congregated levels of racial or ethnic minorities in the contiguous US. White Appalachia's relatively large White population ratio also stands out compared to the average US racial composition.

Figure 12 (b) shows the similarity of income distributions across these four profiles: About 35 percent of the population has an income below $25,000, and close to half the population has an income below $35,000 in all four profiles. The Extremely Vulnerable America profiles also have the highest poverty rates (in terms of both families and individuals) among all profiles. Table 11 shows that the percentages of households receiving SNAP benefits in these four profiles are the highest after the Native Alaska communities.

Low-wage jobs and high unemployment are at the core of poverty. Unemployment rates for the Native American Reservations and White Appalachia profiles (13 percent and 9.6 percent, respectively) are the second- and third-highest among all profiles (after the Native Alaska profile). Unemployment rates in the Hispanic Southern Border and Black South profiles are larger but not significantly different from the national average unemployment rate. In addition, these four profiles rely more on

blue-collar jobs in relatively low-paying industries. For example, workers in these profiles are among the least likely (across all profiles) to find jobs in the top three high-paying industries: "Information," "Finance & insurance, and real estate, rental & leasing," and "Professional, scientific & management, and administrative & waste management services" (see Tables 4 and 11). Government workers occupy about 40 percent of the labor market in the Native American Reservations profile.

Figure 15: Race-Ethnicity and Income in Extremely Vulnerable America



Notes: The percentage in panel (a) counts members of a race-ethnicity who do not identify as Hispanic or Latino so that the total can be 100 percent. Panel (b) indicates what percentages of the population have income falling into specified ranges.

Source: Milken Institute (2022)

Table 11: Poverty Rates and Employment[9]

Category	Variable	Hispanic Southern Border	Black South	White Appalachia	Native American Reservations	*US Average*
Income	With Food Stamp/SNAP benefits	23.4*	22.4*	23.3*	26.1*	*11.7*
	Below poverty level - family	20.9*	20.4*	18.9*	28.9*	*9.5*
	Below poverty level - individuals	25.4*	26*	24.1*	35.9*	*13.4*
Employment Status	Unemployment rate	8.1	8.5	9.6*	13.3*	*5.3*

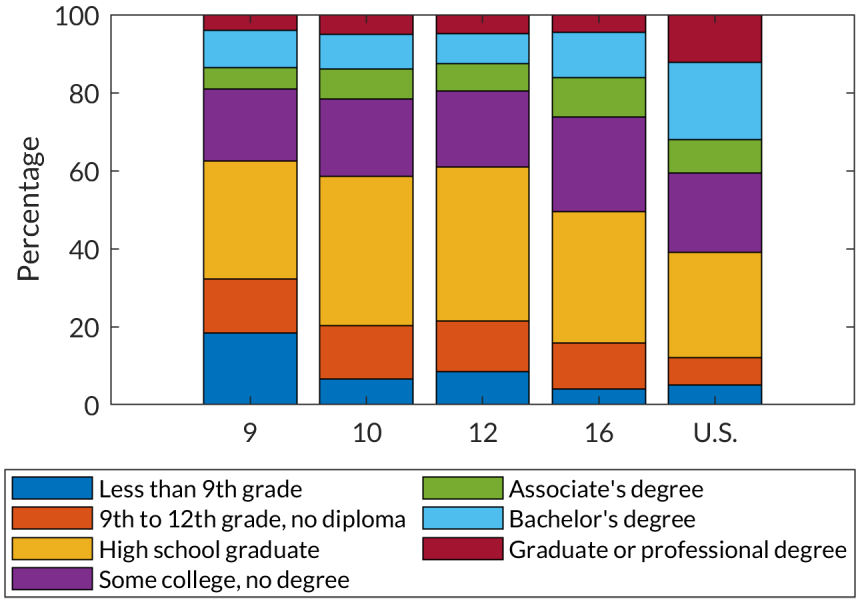
Employment	Management, Business, Science, Arts jobs	24.1*	26.6*	27.9*	36.7	*38.5*
	Service jobs	23.1*	19.8	19.5	22.6	*17.8*
	Natural resources, Construction, and Maintenance jobs	18.2*	12.2	14*	12*	*8.9*
	Production, Transportation, and Material moving jobs	15	21.7*	18.5*	9.9*	*13.2*
	Agriculture, Forestry, Fishing, and Mining	14.1*	5.8*	6.4*	11.6*	*1.8*
	Manufacturing	5.3*	15.5	11.7	2.7*	*10.1*
	Information	0.9*	0.8*	1.2	0.9*	*2*
	Finance and insurance, and Real estate and rental and leasing	3.3*	3.4*	3.5*	3.5*	*6.6*
	Professional, Scientific, and Management	5.4*	5.5*	5.8*	3.2*	*11.6*
	Private wage and salary workers	72.7*	74*	73.4*	46.3*	*80.2*
	Government workers	19.7*	19.8*	19.4*	43.4*	*13.7*

Notes: The table shows averages of selected variables that distinguish these profiles from the rest. The asterisk indicates that a profile average is statistically different from the US average. All values are a percentage of the population.

Source: Milken Institute (2022)

These profiles are characterized by low educational attainments and deep poverty. Figure 16 highlights the prevalence of educational inequality in these profiles, which correlates with the considerably lower incomes relative to the national average. Compared to the national average, the ratio of the population without compulsory education (all grades through high school) is notably high. The Hispanic Southern Border has the lowest educational attainments of these profiles across all categories, which aligns with almost a quarter of its population not speaking English very well. In addition, the percentages of the population holding post-secondary degrees in the Hispanic Southern Border, Black South, and White Appalachia profiles are the lowest among all profiles.

Figure 16: Educational Attainments in Extremely Vulnerable Communities



Notes: This figure shows ratios of the population that attains a certain level of educational degree in the Extremely Vulnerable America profiles and compares them to the US ratios.

Source: Milken Institute (2022)

High disability rates, low health insurance coverage, and lack of digital access are worrisome. Significantly more residents in the group of counties that comprise the Extremely Vulnerable America profiles live with disabilities than in the rest of the US. White Appalachia has the highest disability rate among all profiles, which correlates with a relatively older population compared to the other Extremely Vulnerable America profiles (see Table 12). The Hispanic Southern Border and Native American Reservations profiles have among the lowest health insurance coverages. In contrast, the Black South deviates less from the US average, and the White Appalachia has coverage that is close to the national average. Finally, access to digital services, from owning a computer to having access to quality internet, is a significant concern for all these profiles.

Several other prominent characteristics are correlated to deep poverty in these profiles. Female single-parent households are relatively prevalent in Extremely Vulnerable America: The Black South, American Indian Reservations, and Hispanic Southern Border profiles have the first- to third-highest percentages of single-mother

households, respectively. Lack of English proficiency is an issue: More than half of the Hispanic Southern Border profile's population does not use English at home, almost a quarter of the population does not speak English very well, and foreign-born non-US citizens comprise a large fraction of the population (66 percent). All four profiles have a high vacancy rate of housing units, around 10 percentage points higher than the national vacancy rate.

Table 12: Other Characteristics in Extremely Vulnerable America

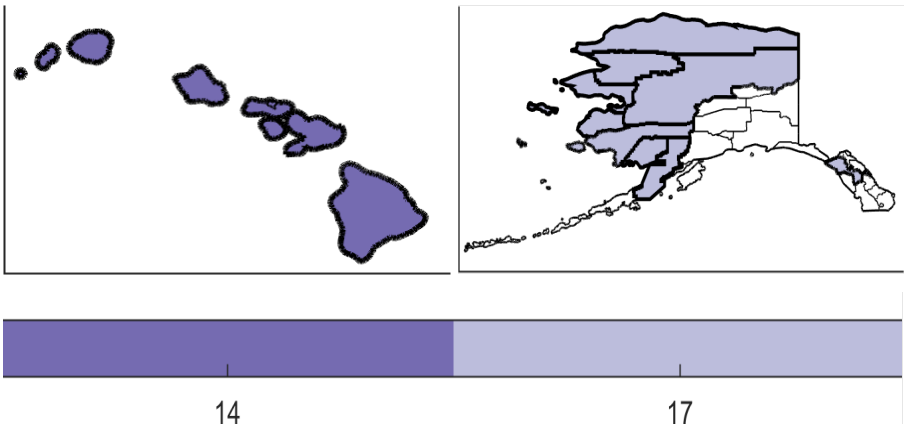
Category	Variable	Hispanic Southern Border	Black South	White Appalachia	Native American Reservations	*US Average/ Median*
Demography	Median age	33.9*	40.6	43.2*	30.5*	*38.1*
Social	Foreign-born population, Not a US citizen	*66**	59.9	56.9	58.3	*50.4*
	Language at home, not English	59.1*	4*	4.4*	17.2*	*21.6*
	Language at home, not English Speak English less than very well	23.1*	1.6*	1.3*	3.1*	*8.4*
Housing	Vacant housing units	21.5*	22.7*	22.4*	23.4*	*12.1*
	No telephone service available	2.5	3.2	2.9	8.3*	*1.9*
Household Type	Married-couple family	48.9	40.1*	48.3	39.7*	*48.2*
	Female householders, no spouse, with children	7.9*	8.3*	4.9	8.3*	*5.3*
	Households with people under 18 years	37.6	*29*	29	42.2*	*31*
	Households with people 65 years and over	*33.9*	34.9*	35.7*	29.1	*29.4*
	Grandparents responsible for grandchildren	42.2	54*	59*	61.8*	*34.1*
Health Insurance / Disability	With health insurance	80*	87.1*	90.1	73.6*	*91.2*
	Disability	16.6*	18.4*	25.6*	13.2	*12.6*
	Uninsured with a disability	10.7	7.9	6.3	15.6*	*5.5*
	Uninsured among unemployed	52.3*	45.5*	36.2	58.4*	*27.6*
Computer /Internet	With a computer	78.9*	74.6*	77.9*	70.7*	*90.3*
	With a broadband Internet subscription	*64.9**	59.7*	66.8*	57.2*	*82.7*

Notes: The table shows an average of selected variables that distinguish these profiles from the rest. The asterisk indicates that a profile average is statistically different from the US average. All values except median age are a percentage of the population.

Source: Milken Institute (2022)

Noncontiguous America

Figure 17: Map of Noncontiguous America



Source: Milken Institute (2022)

These 14 counties, where 0.46 percent of the US population resides, are located in the two noncontiguous areas of the country: The Hawaii profile accounts for the five counties in Hawaii, and the Native Alaska profile accounts for nine of Alaska's 29 counties (the other 22 counties are widely spread across Affluent Suburbs, Middle Class, Retiree Communities, Hispanic Agriculture, Isolated Senior, and The Great Plains).

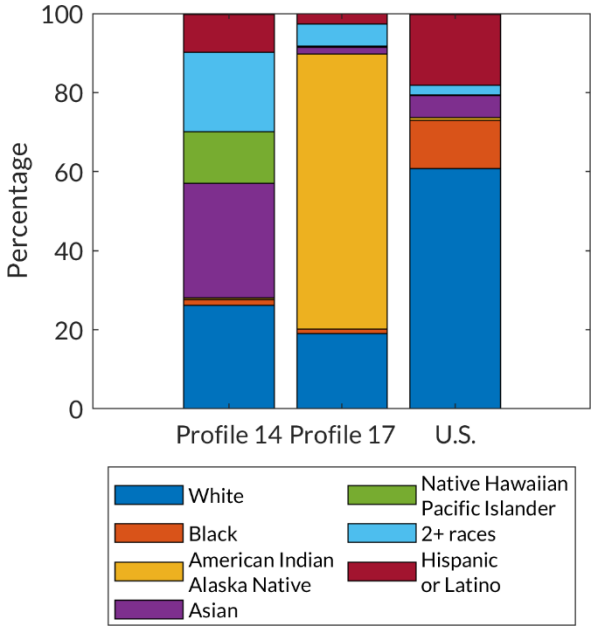
Hawaii comprises 0.4 percent of the US population. Hawaii has the second-smallest White population (26.2 percent), higher only than that of the Native Alaska profile. This profile also has the highest proportions of Asian, Native Hawaiian and Pacific Islander, and Two or More Races populations of all profiles. Household incomes are higher in Hawaii than the national average, and the median income for people aged 25 to 44 is among the highest. But the residents face expensive housing markets. Compared to other profiles, this profile has the largest portion of jobs in the hospitality industry ("Arts, entertainment, and accommodation and food services"). There is a gap in the average

education levels between the White and the Asian populations: Compared to their racial or ethnic groups in other profiles, the percent of the population with a post-secondary degree is the second-highest for the White population, while for Asians, it is the second-lowest.

Native Alaska accounts for 0.02 percent of the US population that lives in counties where a majority of the population (69.6 percent) belong to the American Indians and Alaska Natives racial or ethnic category. Yet, the White population in this profile is the local minority that is better off, with the highest median income for its racial or ethnic category ($100,900) and one of the lowest unemployment rates (2.4 percent) among all profiles. In contrast, Alaska Natives have low median incomes ($43,049) and suffer the most considerable unemployment rate (23.2 percent) among all profiles. Similarly, the Native Alaska profile also has the highest percentage of households receiving SNAP benefits. Finally, its access to quality internet and health insurance coverage is one of the lowest in the country.

One or two national minorities comprise the largest racial or ethnic groups in the Noncontiguous America profiles. In Hawaii, the Asian population (29 percent) exceeds the White population by 3 percentage points. For two other racial or ethnic groups, Two or More Races and Native Hawaiian and Pacific Islander, their shares are the highest in this than in any other profiles. The Native Alaska profile's population belongs predominantly to the American Indian and Alaska Natives racial or ethnic group (69.6 percent of the population), though we suspect most of this population is Alaska Native, since all these counties are located in Alaska.

Figure 18: Race-Ethnicity in Noncontiguous America

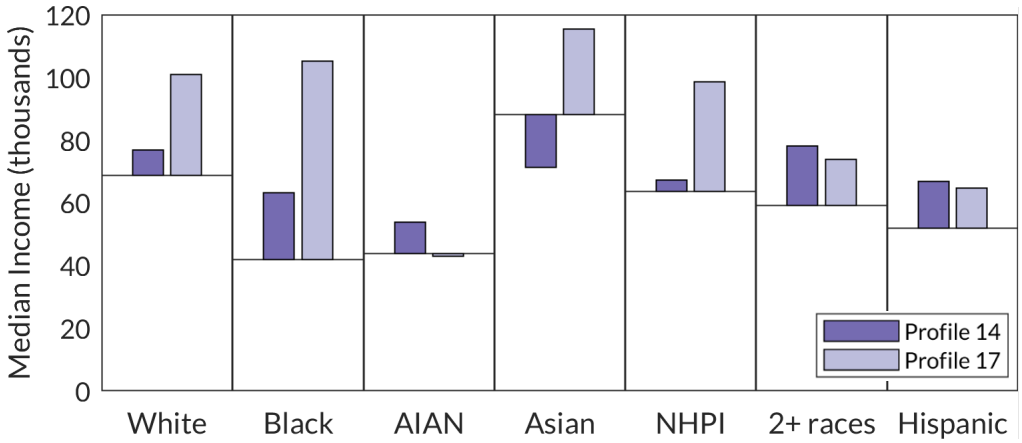


Note: The percentage counts members of a race-ethnicity who do not identify as Hispanic or Latino so that the total can be 100 percent.
Source: Milken Institute (2022)

The White population may be a smaller fraction of the overall population, but it keeps its economic advantages. Figure 19 indicates the difference between the median income for each racial or ethnic group in the Noncontiguous America profiles compared to their national medians. For the Native Alaska profile, White, Black or African American, and Asian populations have the highest median incomes within their racial or ethnic groups among all profiles ($100,900, $105,267, and $115,372, respectively). These relatively high incomes also show clear departures from the overall national median income. The median incomes for American Indians and Alaska Natives, at $43,049, remain in line with the national median for this group.

Hawaii's median incomes for all seven racial or ethnic groups are not statistically significantly different from their respective national median incomes.

Figure 19: Income by Race-Ethnicity in Noncontiguous America

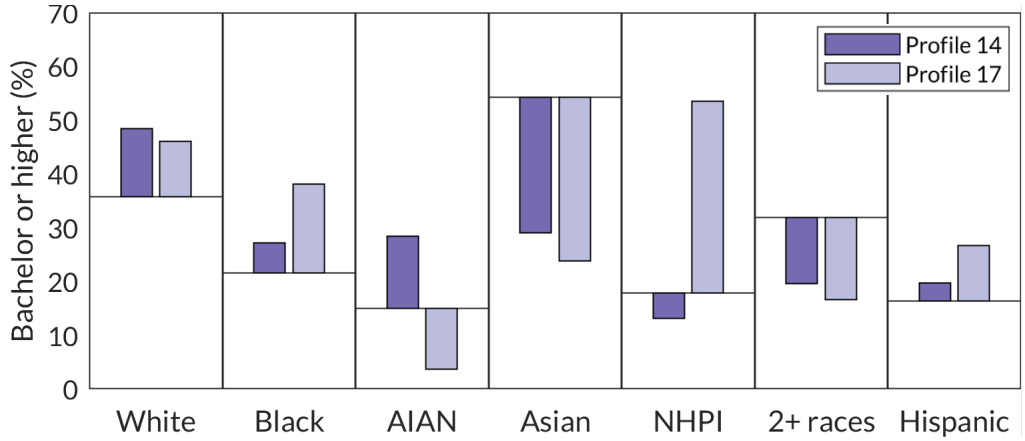


Notes: The bars show, for each race-ethnicity, how the profile's median income deviates from the US median income for from the median income for its respective group. Bar baseline denotes the US median. AIAN is an abbreviation for American Indian and Alaska Natives, and NHPI denotes Native Hawaiian and Other Pacific Islander.

Source: Milken Institute (2022)

A significant gap exists in educational achievements between the White and non-White prominent racial groups compared to their national averages. The White populations in the Hawaii and Native Alaska profiles have higher than national average educational achievements. In contrast, Hawaii's other than White most prominent populations—Asian and Two or More Races—and the Native Alaska profile's American Indians and Alaska Natives significantly lag in educational attainment relative to the national averages for their racial or ethnic groups (see Figure 20).

Figure 20: Education by Race-Ethnicity in Noncontiguous America



Notes: The bars show, for each race-ethnicity, how the profile's percentage of population with bachelor's or higher degrees deviates from the US average and from the average for the respective group. Bar baseline denotes the US average. AIAN is an abbreviation for American Indian and Alaska Natives, and NHPI denotes Native Hawaiian and Other Pacific Islander.

Source: Milken Institute (2022)

These profiles have similar labor markets but different social infrastructures. Table 13 shows that the "Arts, entertainment, recreation and accommodation, and food services" (linked to "Service" jobs) and "Government workers" industries lead Hawaii's labor market. Likewise, the government employs about half (45.6 percent) of the Native Alaska profile's residents, the most significant percentage among all profiles.

Health insurance coverage in the Hawaii and Native Alaska profiles are at the opposite ends of the spectrum—Hawaii has the highest and the Native Alaska profile has the second-lowest coverage. Finally, significantly fewer residents of the Native Alaska profile have access to quality internet services compared to the national average.

Table 13: Other Characteristics of Noncontiguous America

Category	Variable	Hawaii	Native Alaska	*US*
Employment	Service jobs	29.5*	19.6	17.8
	Arts, Entertainment, Recreation, and Accommodation and Food services	19.9*	5.9*	*9.7*
	Private wage and salary workers	67.5	50.2*	*80.2*
	Government workers	24.8*	45.6*	*13.7*
Health Insurance, Computer /Internet	With health insurance	96.4*	76.5*	*91.2*
	With a computer	88.1	84.1	*90.3*
	With a broadband internet subscription	79.9	67.5*	*82.7*

Notes: The table shows averages of selected variables that distinguish these profiles from the rest. The asterisks indicate that a profile average is statistically different from the US average. All values are shown as percentage of the population.

Source: Milken Institute (2022)

CONCLUDING REMARKS

The latest census confirms that the US population will continue to change in many dimensions. Just to name a couple, the population will get older, and the White population will shrink to less than 40 percent of the entire population by the year 2060, while the Hispanic and Latino population will continue to rise with all the other minorities except for the Black or African American population. In light of these changes, the multidimensionality of diversity cannot be ignored when tackling issues related to inequalities.

With the Community Explorer, we propose a new approach to policy that effectively leverages county-level data produced by the Census to inform decisions related to equity across America.

We have already explained in this report the benefits of clustering the information into communities defined by the populations' characteristics and not their location. It allows for insightful benchmarking when determining or assessing the impact of an initiative, permitting comparison across peer counties even if they are not within the same state or region. It also identifies the main factors that differentiate one community from another one.

We would like to share some final remarks on this novel and informative policy and visualization tool.

- **It identifies correlations:** The combination of information related to a specific topic with the community profiles highlights patterns across the US but does not provide causal insights.
- **It allows states to leverage the complexity of their population's diversity to produce tailored and flexible policies.** The 254 counties of Texas are spread across 14 profiles, while the 58 counties of California are spread across 9. The Community Explorer can help states align their policies with their diversity while allowing for economies of scale or scalability in policy implementation. The same reasoning applies to policy within a region or on a national level.

- **It goes beyond the rural versus urban differentiation.** The profiles provide informative nuances beyond the rural versus urban dimensions. To illustrate, let us compare the Great Plains with the Black South. Both are highly rural, with 96 percent and 85 percent of their respective populations living in non-core and micropolitan areas. Yet, the Great Plains profile represents a group of relatively well-off, middle-class counties with a dominant non-Hispanic White population (90.8 percent) working largely in agriculture. In contrast, the Black South profile groups economically vulnerable counties with a large Black or African American population (47.3 percent) working in low-skilled manufacturing jobs. Similar contrasts can be drawn across urban counties, some of which represent the Affluent Suburbs with a mostly non-Hispanic population, while others belong to the ethnically diverse Urban-Core.

As an accompaniment to this report, the Community Explorer dashboard provides an appealing and intuitive visual tool that allows for user exploration of the wealth of information discussed here. Users can also download the graphs and information provided in the dashboard to use it in their own research and analyses.

REFERENCES

Caliński, T, and J Harabasz. 1974. "A Dendrite Method for Cluster Analysis." *Communications in Statistics* 3 (1): 1–27. https://doi.org/10.1080/03610927408827101.

Davies, D L, and D W Bouldin. 1979. "A Cluster Separation Measure." *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1 (2): 224–27. https://doi.org/10.1109/TPAMI.1979.4766909.

Ester, Martin, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. 1996. "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." *KDD-Association for the Advancement of Artificial Intelligence* 95 (34): 226–31. https://doi.org/10.1016/B978-044452701-1.00067-3.

Helper, Susan, Timothy Krueger, and Howard Wial. 2012. "Locating American Manufacturing: Trends in the Geography of Production." *SSRN Electronic Journal*, Metropolitan Policy Program at Brookings: 1–60. https://doi.org/10.2139/ssrn.3798078.

Ingram, Deborah D, and Sheila J Franco. 2014. "2013 NCHS Urban-Rural Profileification Scheme for Counties. National Center for Health Statistics." *Vital Health Statistics* 2 (166): 1–73.

Jensen, Eric, Nicholas Jones, Megan Rabe, Beverly Pratt, Lauren Medina, Kimberly Orozco, and Lindsey Spell. August 12, 2021. "The Chance That Two People Chosen at Random Are of Different Race or Ethnicity Groups Has Increased Since 2010." *US Census Bureau*. https://www.census.gov/library/stories/2021/08/2020-united-states-population-more-racially-ethnically-diverse-than-2010.html.

Lloyd, S. 1982. "Least Squares Quantization in PCM." *IEEE Transactions on Information Theory* 28 (2): 129–37. https://doi.org/10.1109/TIT.1982.1056489.

Parker, Kim, Juliana Horowitz, Anna Brown, Richard Fry, D'Vera Cohn, and Ruth Igielnik. 2018. "What Unites and Divides Urban, Suburban and Rural Communities." *Pew Research Center*, May: 1–90. https://www.pewsocialtrends.org/2018/05/22/what-unites-and-divides-urban-suburban-and-rural-communities/.

Rousseeuw, Peter J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20 (C): 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Tibshirani, Robert, Guenther Walther, and Trevor Hastie. 2001. "Estimating the Number of Clusters in a Data Set via the Gap Statistic." *Journal of the Royal Statistical Society: Series B*

(Statistical Methodology) 63 (2): 411–23. https://doi.org/https://doi.org/10.1111/1467-9868.00293.

ACKNOWLEDGMENTS

The authors would like to thank Keith Savard, Perry Wong, and participants of the Global Conference 2022 for insightful discussions about this topic. They also thank participants at the Milken Institute Research Department brown bag meetings for useful discussions.