ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

ICPSR 38528

National Neighborhood Data Archive (NaNDA): Socioeconomic Status and Demographic Characteristics of Census Tracts and ZIP Code Tabulation Areas, United States, 2000-2020

P.I. Documentation for Socioeconomic Status and Demographic Characteristics of Census Tracts and ZIP Code Tabulation Areas, United States, 2016-2020 Data

Inter-university Consortium for Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

National Neighborhood Data Archive (NaNDA): Socioeconomic Status and Demographic Characteristics of Census Tracts and ZIP Code Tabulation Areas, United States, 2000-2020

Philippa Clarke
University of Michigan. Institute for Social Research

Grace Noppert
University of Michigan. Institute for Social Research

Robert Melendez
University of Michigan. Institute for Social Research

Megan Chenoweth
University of Michigan. Institute for Social Research

Lindsay Gypin
University of Michigan. Institute for Social Research

Terms of Use

The terms of use for this study can be found at:
http://www.icpsr.umich.edu/web/ICPSR/studies/38528/terms

Information about Copyrighted Content

Some instruments administered for studies archived with ICPSR may contain in whole or substantially in part contents from copyrighted instruments. Reproductions of the instruments are provided as documentation for the analysis of the data associated with this collection. Restrictions on "fair use" apply to all copyrighted content. More information about the reproduction of copyrighted works by educators and librarians is available from the United States Copyright Office.

National Neighborhood Data Archive (NaNDA): Socioeconomic Status and Demographic Characteristics of Census Tracts and ZIP Code Tabulation Areas, United States, 2016-2020

Overview and Data Dictionary

Documentation Version: 1.3
Last updated: 04/06/2023

Dataset Overview

Description

This dataset contains measures of socioeconomic and demographic characteristics (SES) by US census tract and zip code tabulation area (ZCTA) for the years 2016-2020. Example measures include population density; population distribution by race, ethnicity, age, and income, income inequality by race and ethnicity; and proportion of population living below the poverty level, receiving public assistance, and female-headed or single parent families with kids. The dataset also contains a set of theoretically derived measures capturing neighborhood socioeconomic disadvantage and affluence, as well as a neighborhood index of Hispanic, foreign born, and limited English.

Principal Investigators

- Philippa Clarke, University of Michigan Institute for Social Research
- Robert Melendez, University of Michigan Institute for Social Research
- Grace Noppert, University of Michigan Institute for Social Research
- Lindsay Gypin, University of Michigan Institute for Social Research

Funding Sources

- National Institutes of Health/National Institute of Nursing Research and the National Institute on Minority Health and Health Disparities (NINR/ NIMHD U01NR020556)
- National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR, 90RTHF0001)

Data Sources

Population values and proportion variables per census tract and ZCTA are from the American Community Survey (ACS) five-year estimates for 2016-2020 (United States Census Bureau, 2022a, 2022d).

Land area for each census tract or ZCTA comes from the 2020 TIGER/Line shapefiles (United States Census Bureau, 2021).

Coverage

The dataset contains one observation per census tract or ZCTA in the fifty United States including Alaska, Hawaii, Washington DC, and Puerto Rico.

Methodology

To construct this dataset, we extracted key census indicators related to race, ethnicity, age, income, employment, poverty, and home ownership from the ACS 2020 five-year estimate (covering 2016-2020). We merged the variables with each tract's or ZCTA's land area from the 2020 TIGER/Line shapefiles for census tracts and ZIP code tabulation areas.

Updates in this dataset include:

- A variable for proportion of single-parent households has been added.
- Income quartiles have been updated to be consistent with the income distribution in the United States for 2016-2020. (This was not done in previous versions of this dataset.)
    - Proportion of families with income less than $40,000
    - Proportion of families with income $40,000 to $74,999
    - Proportion of families with income $75,000 to $124,999
    - Proportion of families with income $125,000 or more

- Measures of racial/ethnic income inequality have been created based on a ratio of median household income in White and Black residents; White and Hispanic residents.

We then conducted a principal components analysis, exploratory factor analysis, and confirmatory factor analysis with census tract indicators from the 2016-2020 ACS 5-year estimates to empirically re-evaluate the neighborhood socioeconomic and demographic indices from previous versions of this dataset. Our aim was to provide a parsimonious set of theoretically-derived factors that capture the shared variance across a broad spectrum of structural socioeconomic characteristics. Results from the factor analysis indicated three separate factors:

- The first factor, which we interpret as Neighborhood Disadvantage, is characterized by high levels of poverty, low family income, and households receiving public assistance income in the neighborhood. Neighborhoods that concentrate the socioeconomically disadvantaged have fewer resources (e.g., healthy food stores, well-maintained parks, good schools, quality medical care) to promote good health (Ross et al., 2001; Ross & Mirowsky, 2001) and are often vulnerable to disinvestment and environmental hazards (Mohai et al., 2009).
- The second factor, which we interpret as Neighborhood Affluence, is characterized by high levels of people with a college education, families with high income, and people employed in professional/managerial occupations in the neighborhood. Affluent neighborhoods are likely to attract a set of institutions (e.g., food stores, places to exercise, well-maintained buildings and parks) that foster a set of norms (e.g., an emphasis on exercise and healthy diets) conducive to good health (Clarke et al., 2014). Distinct from simply being the absence of neighborhood disadvantage, neighborhood affluence is associated with higher levels of social control and leverage over local institutions that can foster social environments that facilitate health (Browning & Cagney, 2003).
- The third factor represents a higher proportion of Hispanic, foreign born, and people with limited English proficiency in the neighborhood.

Results from the factor analysis were used to inform the creation of the neighborhood SES mean indices in the dataset:

- The **neighborhood disadvantage** index variable is a mean of the proportion of people with income below federal poverty level, proportion of households receiving public assistance income, and proportion of families with annual income less than $40,000.
- The **neighborhood affluence** index variable is a mean of the proportion of adults with a college degree, proportion of people employed in professional or

managerial occupations, and proportion of families with annual income of $125,000 or more.
- The **Hispanic Foreign Born Limited English** index variable is a mean of the proportion of people self-identifying as Hispanic, proportion of people who are foreign born, and proportion who don't speak English well.

Further details on the statistical methodology and results of the factor analyses are provided in the technical documentation (Clarke & Melendez, 2023) for these measures.

Usage Notes

Data users should be aware of the differences in the Neighborhood Disadvantage and Affluence measures in the 2016-2020 dataset compared to previous years. While they capture similar constructs of concentrated disadvantage or concentrated affluence and may be conceptually comparable, the items included in these measures have changed as the socioeconomic structure of US neighborhoods has evolved over time.

Users wanting to use this dataset alongside measures from NaNDA's 2000-2017 SES datasets should be aware of key differences between how measures are calculated across datasets. For the year 2000, the indicators necessary to build these neighborhood disadvantage and affluence measures came from the decennial census. These variables were moved to the American Community Survey (ACS) in 2005 and removed from the decennial census in 2010 with the introduction of the American Community Survey. In more recent years, they can be found only in ACS five-year estimates.

When constructing measures for 2000-2010, we used data from the 2000 decennial census to represent 2000, data from the 2008-2012 American Community Survey five-year estimate to represent 2010 (the midpoint of the range). Because no data was available for the intervening years, we imputed values for 2001-2009 using a linear interpolation between the two endpoints. For years 2008-2012, 2013-2017, and 2016-2020 the values represent all five years of the ACS estimate range, because they represent the Census Bureau's best estimate for the entire period. The Census Bureau recommends against combining data from overlapping five-year estimates (United States Census Bureau, 2022c).

Additionally, this dataset uses United States Census Bureau geography files from 2020 (United States Census Bureau, 2021), while previous versions use United States Census Bureau geography files from 2010 (United States Census Bureau, 2012). Census tracts and ZCTAs have changed from 2010 to 2020. Detailed descriptions of the changes to census geography, as well as documentation that may help with longitudinal analysis can be found at https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.2020.html (United States Census Bureau, n.d.).

Zip Codes and ZIP Code Tabulation Areas

Users should be aware that ZCTAs are not equivalent to ZIP codes. ZIP codes are linear mail delivery routes created by the US Postal Service. ZIP code tabulation areas are spatial features consisting of census blocks grouped by the predominant ZIP code found on the block (United States Census Bureau, 2020).

In some cases, a location's address is not the same as its ZCTA. For example, some ZIP codes represent single-point addresses such as large post offices or office buildings. Also, the ZIP code for an address may not match its ZCTA if the ZIP code is not the most common ZIP code on the block. See the Census Bureau's ZCTA overview at https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html (United States Census Bureau, 2022b) for more information on how ZCTA boundaries are calculated.

Users wanting to combine this dataset with ZIP code geocoded data must use a ZIP code to ZCTA crosswalk. Such a crosswalk is available on the UDS Mapper website at https://udsmapper.org/zip-code-to-zcta-crosswalk/ (John Snow, Inc, 2022). Sample code for merging the UDS Mapper crosswalk with NaNDA datasets is available in the ICPSR Linkage Library at http://doi.org/10.3886/E124461V2 (Chenoweth & Khan, 2021)

Related Datasets

Additional related datasets (listed below) are available through the National Neighborhood Data Archive (NaNDA) at ICPSR.

- Neighborhood Socioeconomic and Demographic Characteristics by Tract, United States, 2000-2010
- Socioeconomic Status and Demographic Characteristics of Census Tracts, United States, 2008-2017
- Socioeconomic Status and Demographic Characteristics of ZIP Code Tabulation Areas, United States, 2008-2017

National Neighborhood Data Archive (NaNDA)
Socioeconomic Status and Demographics Characteristics of Census Tracts and ZIP Code Tabulation Areas, United States, 2016-2020

Census Tract Variables

Variable	Label
tract_fips20	Census Tract FIPS code, 2020
aland20	Census land area, square miles
totpop20	Total population
popden16_20	Persons per square mile
phispanic16_20	Proportion of people of Hispanic origin
pnhwhite16_20	Proportion of people non-Hispanic White
pnhblack16_20	Proportion of people non-Hispanic Black
pfborn16_20	Proportion of people who are foreign born
plimeng16_20	Proportion of persons (age 5+) who don't speak English well
ped1_16_20	Proportion with Less than High School Diploma
ped2_16_20	Proportion of persons with High School Diploma and/or Some College
ped3_16_20	Proportion of persons with Bachelors Degree or Higher
pfaminclt40k16_20	Proportion of families with income less than $40,000
pfamincge40lt75k16_20	Proportion of families with income $40,000 to $74,999
pfamincge75lt125k16_20	Proportion of families with income $75,000 to $124,999
pfamincge125k16_20	Proportion of families with income $125,000 or more
pnvmar16_20	Proportion of People Age 15+ Never Married
p18yr_16_20	Proportion of population under 18 yrs
p18_29_16_20	Proportion of population 18-29
p30_39_16_20	Proportion of population 30-39 yrs
p40_49_16_20	Proportion of population 40-49 yrs
p50_69_16_20	Proportion of population 50-69 yrs
pge70_16_20	Proportion of population 70+ yrs
punemp16_20	Proportion of age 16+ civ labor force unemployed
pprof16_20	Prop. of civ employed in Management, business, science, and arts occupations
ppov16_20	Proportion people w/ income in the past 12 months below poverty level
ppubas16_20	Proportion of households with public assistance income or food stamps
pfhfam16_20	Proportion female-headed families w/ kids
psngpnt16_20	Proportion single parent families w/ kids
pownoc16_20	Proportion Owner Occupied Housing Units
affluence16_20	mean of ped3_16_20, pfamincge125k16_20, and pprof16_20
disadvantage16_20	mean of pfaminclt40k16_20, ppov16_20, and ppubas16_20
hispan_forborn_limeng16_20	mean of phispanic16_20, pfborn16_20, and plimeng16_20
medfaminc16_20	Median Family Income
medfaminc_nhwhite16_20	Median Family Income (White Alone, Not Hispanic Or Latino Householder)
medfaminc_black16_20	Median Family Income (Black Or African American Alone Householder)
medfaminc_hispanic16_20	Median Family Income (Hispanic Or Latino Householder)
ratio_medfaminc_nhwtob16_20	Ratio Median Family Income, NH-White / Black
ratio_medfaminc_nhwtohisp16_20	Ratio Median Family Income, NH-White / Hispanic

National Neighborhood Data Archive (NaNDA)
Socioeconomic Status and Demographics Characteristics of Census Tracts and ZIP Code Tabulation Areas, United States, 2016-2020

ZCTA Variables

Variable	Label
zcta20	zip code tabulation area (5-digit)
aland20	Census land area, square miles
totpop20	Total population
popden16_20	Persons per square mile
phispanic16_20	Proportion of people of Hispanic origin
pnhwhite16_20	Proportion of people non-Hispanic White
pnhblack16_20	Proportion of people non-Hispanic Black
pfborn16_20	Proportion of people who are foreign born
plimeng16_20	Proportion of persons (age 5+) who don't speak English well
ped1_16_20	Proportion with Less than High School Diploma
ped2_16_20	Proportion of persons with High School Diploma and/or Some College
ped3_16_20	Proportion of persons with Bachelors Degree or Higher
pfaminclt40k16_20	Proportion of families with income less than $40,000
pfamincge40lt75k16_20	Proportion of families with income $40,000 to $74,999
pfamincge75lt125k16_20	Proportion of families with income $75,000 to $124,999
pfamincge125k16_20	Proportion of families with income $125,000 or more
pnvmar16_20	Proportion of People Age 15+ Never Married
p18yr_16_20	Proportion of population under 18 yrs
p18_29_16_20	Proportion of population 18-29
p30_39_16_20	Proportion of population 30-39 yrs
p40_49_16_20	Proportion of population 40-49 yrs
p50_69_16_20	Proportion of population 50-69 yrs
pge70_16_20	Proportion of population 70+ yrs
punemp16_20	Proportion of age 16+ civ labor force unemployed
pprof16_20	Prop. of civ employed in Management, business, science, and arts occupations
ppov16_20	Proportion people w/ income in the past 12 months below poverty level
ppubas16_20	Proportion of households with public assistance income or food stamps
pfhfam16_20	Proportion female-headed families w/ kids
psngpnt16_20	Proportion single parent families w/ kids
pownoc16_20	Proportion Owner Occupied Housing Units
affluence16_20	mean of ped3_16_20, pfamincge125k16_20, and pprof16_20
disadvantage16_20	mean of pfaminclt40k16_20, ppov16_20, and ppubas16_20
hispan_forborn_limeng16_20	mean of phispanic16_20, pfborn16_20, and plimeng16_20
medfaminc16_20	Median Family Income
medfaminc_nhwhite16_20	Median Family Income (White Alone, Not Hispanic Or Latino Householder)
medfaminc_black16_20	Median Family Income (Black Or African American Alone Householder)
medfaminc_hispanic16_20	Median Family Income (Hispanic Or Latino Householder)
ratio_medfaminc_nhwtob16_20	Ratio Median Family Income, NH-White / Black
ratio_medfaminc_nhwtohisp16_20	Ratio Median Family Income, NH-White / Hispanic

References

Browning, C. R., & Cagney, K. A. (2003). Moving beyond Poverty: Neighborhood Structure, Social Processes, and Health. *Journal of Health and Social Behavior*, *44*(4), 552–571. https://doi.org/10.2307/1519799

Chenoweth, M., & Khan, A. (2021). *Code for merging National Neighborhood Data Archive ZCTA level datasets with the UDS Mapper ZIP code to ZCTA crosswalk* (Version v2) [Data set]. ICPSR - Interuniversity Consortium for Political and Social Research. https://doi.org/10.3886/E124461V2

Clarke, P. J., & Melendez, R. (2023). *The National Neighborhood Data Archive (NaNDA): 2016-2020 Neighborhood Sociodemographic Measures Technical Documentation on Index Creation*. https://doi.org/10.7302/7033

Clarke, P. J., Melendez, R., Debbink, M., Golberstein, E., Elliott, M. R., & Lantz, P. M. (2014). Cumulative Exposure to Neighborhood Context: Consequences for Health Transitions Over the Adult Life Course. *Research on Aging*, *26*, 113–140. https://doi.org/doi.org/10.1177/0164027512470702

John Snow, Inc. (2022). *ZIP Code to ZCTA Crosswalk: UDS Mapper*. American Academy of Family Physicians. https://udsmapper.org/zip-code-to-zcta-crosswalk/

Mohai, P., Lantz, P. M., Morenoff, J., House, J. S., & Mero, R. P. (2009). Racial and socioeconomic disparities in residential proximity to polluting industrial facilities: Evidence from the Americans' Changing Lives Study. *American Journal of Public Health*, *99 Suppl 3*(Suppl 3), S649-656. https://doi.org/10.2105/AJPH.2007.131383

Ross, C. E., & Mirowsky, J. (2001). Neighborhood disadvantage, disorder, and health. *Journal of Health and Social Behavior*, *42*(3), 258–276. https://doi.org/10.2307/3090214

Ross, C. E., Mirowsky, J., & Pribesh, S. (2001). Powerlessness and the amplification of threat: Neighborhood disadvantage, disorder, and mistrust. *American Sociological Review*, *66*(4), 568–591. https://doi.org/10.2307/3088923

United States Census Bureau. (n.d.). *Geographic Relationship Files*. Census.Gov. Retrieved April 3, 2023, from https://www.census.gov/geographies/reference-files/time-series/geo/relationship-files.html

United States Census Bureau. (2012, June 13). *Index of Tiger/Line shapefiles: All lines*. https://www2.census.gov/geo/tiger/TIGER2010/EDGES

United States Census Bureau. (2021). *Index of TIGER/Line shapefiles: All lines*. https://www2.census.gov/geo/tiger/TIGER2020/EDGES/

United States Census Bureau. (2022a). *Index of American Community Survey 2016-2020 Five Year Estimate Summary Files*. https://www2.census.gov/programs-surveys/acs/summary_file/2020/data/5_year_entire_sf/

United States Census Bureau. (2022b, May 16). *ZIP Code Tabulation Areas (ZCTAs)*. Census.Gov. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/zctas.html

United States Census Bureau. (2022c, July 8). *Comparing ACS Data*. Census.Gov. https://www.census.gov/programs-surveys/acs/guidance/comparing-acs-data.html

United States Census Bureau. (2022d, November 30). *American Community Survey 5-Year Data (2009-2021)*. Census.Gov. https://www.census.gov/data/developers/data-sets/acs-5year.html