

ICPSR BULLETIN

Volume XXIX, No. 2

Fall 2008



ICPSR's New Resource for Multi-Level Analyses: *County Characteristics, 2000–2007*

Alon Axelrod

Social science and health researchers often want to investigate the extent to which the characteristics of geographic locales — the environments in which experiences, states of mind, and behavior occur — influence individual-level outcomes such as obesity, depression, criminal sentences, health and access to health care, voting behavior, reproductive decisions, social trust, or job satisfaction. To that end, researchers analyze macro-level contextual variables and individual-level variables together in multi-level analyses of individual-level outcomes. To facilitate such investigations, ICPSR created *County*

Characteristics, 2000–2007 (ICPSR 20660), a county-level contextual data file covering a wide array of county attributes, which can be used to study county-level contextual influences. *County Characteristics* can also be used in research where groups or institutions are the lower (or intermediate) unit of analysis and in research where the county is the only unit of analysis, for example, to study county attributes that foster migratory population growth. Moreover, *County Characteristics* can be used as a comprehensive data source to inform policy decisions.

Overview

County Characteristics comprises 470 variables: population size and the components of population change during 2000–2005 as well as numerous environmental, demographic, economic, political, and health-related attributes. As shown on the next page,



Alon Axelrod is a research associate at ICPSR with a background in demography and more than 25 years

experience in data management and programming. He designed and produced *County Characteristics, 2000–2007* (ICPSR 20660).

the variables are organized into 15 groups, each derived from different data sources. Most of the variables were derived from data files prepared by various federal agencies, hundreds of data files with millions of discrete records. About one half of the contextual file's variables were copied as is from the data sources, while the rest were computed. Most of the computed variables were generated by combining information across records

See *County Characteristics*, p. 2

In this issue

ICPSR's New Resource for Multi-Level Analyses	1
The Scientific Value and Disclosure Risk of Contextual Data	3

County Characteristics (cont. from page 1)

or variables in the original data files. In addition, some rates, percentages, and sums were computed from other variables in the contextual data file. The contextual data file contains 3,141 cases, one for each county and county equivalent in the United States. The District of Columbia, parishes in Louisiana, boroughs and census areas in Alaska, and independent cities, mostly in Virginia, are the county equivalents.

An Example: The Oldest Old

Since the unit of observation is the county, the variables can be easily displayed in maps using GIS software. For example, Figure 1 shows a map of the variable Pct85plus05, the percent of the 2005 population aged 85 and over (only the 48 contiguous states are displayed). This is a continuous variable that was aggregated into six categories for the map. The red counties have the highest percentages aged 85 and over while the dark green counties have the lowest percentages. Looking at the map, one can see that some of the oldest counties (i.e., those with the highest percentages 85+) are obvious retirement destinations, while others, especially in the central and upper Great Plains, are not.

Data Formats and Documentation

County Characteristics is provided in four data formats: column-delimited ASCII, SAS, SPSS, and Stata. In addition, SAS, SPSS, and Stata setups accompany the ASCII data file. The documentation consists of a codebook, which describes the variables and data sources and explains key concepts; a record layout file which shows the column locations of the variables in the ASCII data file; and a separate volume with frequencies and descriptive statistics.

Dissemination

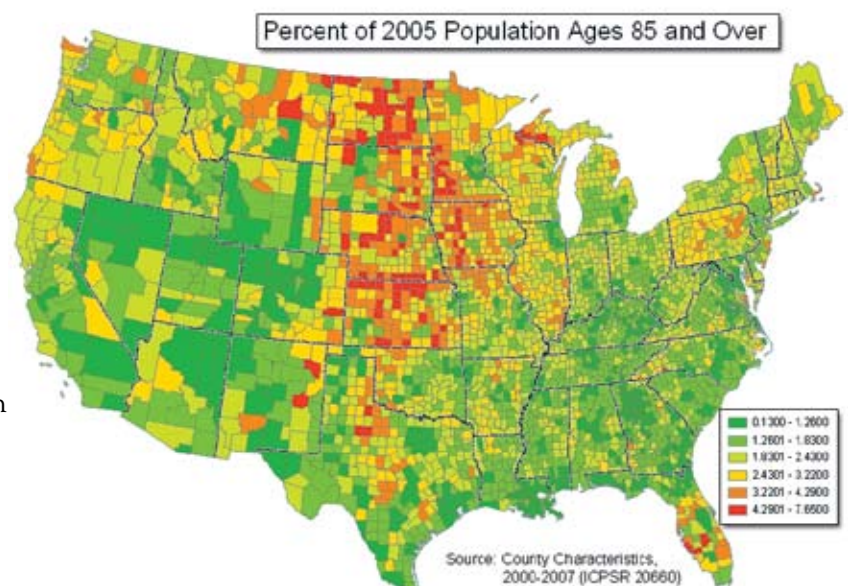
Disseminated free of charge by ICPSR, *County Characteristics* is one of our “best sellers.” Since its release on October 24, 2007, *County Characteristics* has been downloaded from ICPSR’s server more than 570 times — involving some 3,300 copies of data, documentation, and setup files. Thus, it ranks ninth among the most frequently downloaded ICPSR datasets.

We welcome your comments about how you were able to put *County Characteristics* to use. Send your comments to netmail@icpsr.umich.edu. ■

Variables in *County Characteristics*, 2000–2007

- Geographic Identification Variables, 2005
- Geographic Coordinates, Land Area, and Water Area, 2000
- Total Population, Births, Deaths, Migration, Group Quarters Population, and Housing Units, 2000–2005
- Population by Selected Age Groups and Sex; Sex Ratio; Median Age by Sex; Population by Sex, Race, and Hispanic Origin, 2005
- Labor Force Size, Employment, and Unemployment, 2005
- County Typology Codes, 2004; Urban Influence and Rural-Urban Continuum Codes, 2003
- Personal Income by Major Source; Earnings, Compensation, and Employment by Industry, 2005
- Land Surface Form Typography, Climate, and Natural Amenity Scale
- Federal Government Expenditures, Fiscal Year 2004
- Local Government Revenue and Expenditures, 2001–2002 Fiscal Year
- Crimes Reported to Police, 2004
- 2004 Presidential Election Results
- New Privately-Owned Residential Housing Units Authorized by Building Permits, 2005
- Health Profession Shortage Areas (HPSA), 2007
- Medicare Enrollment, 2003

Figure 1





The Scientific Value and Disclosure Risk of Contextual Data

Kristine Witkowski

Where we reside reflects the quality of our lives and determines our behavior, health, and socioeconomic outcomes as well as the functioning of organizations in our environment. The attributes of neighborhoods and larger surrounding geographic areas represent contextual data that are spatial in nature. Such data are a rich source of information to social science researchers who seek to understand (1) the determinants of the spatial distribution of populations, institutions, resources, and barriers, and (2) how these geographic settings affect individuals and establishments.

Researchers can analyze contextual data in two ways. First, studies can be conducted at the macro-level where neighborhoods are the only unit of analysis. Second, studies can utilize data that is multi-level in nature, whereby the outcomes for persons and establishments are viewed as being determined by both individual and contextual characteristics. In constructing data for such a multi-level study, one must link geographic-level variables to records within a microdata file.

An Application From Public Health Research

Scientists have examined neighborhood attributes to assess patterns of residential segregation and the health implications for different populations.¹ Health-promoting environments are characterized by sustainable

employment, high-performing schools, low crime rates, the absence of environmental hazards, and the availability of healthy food outlets and high-quality health care. To broadly capture the degree to which an environment is healthful, researchers frequently use such contextual measures as the neighborhood poverty rate, its unemployment rate, the proportion of households headed by single females, and the proportion of adults without a high school diploma.

Substantial racial/ethnic disparities exist in children's access to neighborhoods that support healthy development. It may be argued that minority families are less affluent than whites, and therefore they are unable

to locate housing in healthier environments. However, the typical neighborhood environment of black and Hispanic children is much worse than for the poorest of white children. This finding indicates that racial discrimination tends to confine these minorities into relatively disadvantaged neighborhoods, regardless of their families' purchasing power.

Scientific evidence consistently shows that, after taking into account individual- and family-level factors, disadvantaged environments are associated with detrimental health

outcomes and negative health behaviors of children that affect their well-being throughout their lives. Not only are black and Latino children more likely to live in poor families than other children, they are also more likely to live in neighborhoods with unfavorable socioeconomic environments. Hence minority children often face a "double jeopardy" to their health that stems from determinants that are multi-level in nature and are best assessed with contextualized microdata.

Safely Providing Contextual Information

As the above discussion illustrates, information about geographic settings offers important insights into complex processes that affect society. Consequently, researchers are increasingly calling for spatial data. Data producers



Kristine Witkowski is a Research Investigator at ICPSR whose work focuses on disclosure risk associated with contextual data and experimental traits

of public-use collections. Her research is part of a larger program at ICPSR that aims to answer the call for knowledge and innovation in the dissemination of confidential social science data.

¹ My discussion draws heavily upon the work of: Acevedo-Garci, Dolores, Theresa L. Osypuk, Nancy McArdle, and David R. Williams. 2008. "Toward A Policy-Relevant Analysis of Geographic and Racial/Ethnic Disparities in Child Health." *Health Affairs* 27(2): 321–333. (<http://content.healthaffairs.org/cgi/reprint/27/2/321>)

INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

Myron Gutmann, Director
George Alter, Associate Director
Mary Vardigan, Assistant Director

COUNCIL MEMBERS

Aletha C. Huston, Chair
University of Texas at Austin

Francine Berman
University of California, San Diego

Michael F. Goodchild
University of California, Santa Barbara

Michael R. Haines
Colgate University

Kathleen Mullan Harris
University of North Carolina at
Chapel Hill

Thomas LaVeist
Johns Hopkins University

Jeffrey Moon
Queen's University

Samuel L. Myers Jr.
University of Minnesota

C. Matthew Snipp
Stanford University

Lori M. Weber
California State University, Chico

Ann Wolpert
Massachusetts Institute of Technology

Christopher Zorn
Pennsylvania State University

Ruth Peterson, Past Chair
The Ohio State University

BULLETIN STAFF

Editor: Ruth Shamraj

have responded to this demand by providing tables that present statistics derived from persons and institutions nested within identified geographies. *County Characteristics, 2000–2007* (ICPSR 20660) gathers an abundance of such tabulations. Facilitating both macro- and multi-level analyses, this study offers a wide array of contextual attributes of U.S. counties.

Government agencies that have provided the original sources of data for *County Characteristics* have utilized a variety of methods to guarantee the anonymity of individuals whose personal characteristics were used to derive geographically-specific estimates.² Consequently, these published tabular data do not pose a disclosure risk when researchers analyze county-level records alone.

However, disclosure risk becomes a concern when researchers attach these contextual measures to their own microdata files. If researchers plan to distribute these enriched data to their colleagues or to the public, they should first conduct a risk assessment that considers both the

personal characteristics of individuals and the contextual characteristics of unidentified geographies. Once the investigator locates a respondent who can be easily identified, the researcher should modify their personal and contextual information by choosing from a variety of disclosure limitation techniques, so that an intruder is unlikely to be able to pinpoint these respondents' identities. In so doing, the researcher must strike a delicate balance between providing easy access to scientifically useful data while ensuring that the identities of survey respondents remain confidential.³ ■

2 Massell, Paul B. 2003. "Statistical Disclosure Control for Tables: Determining which Method to Use." Proceedings of Statistics Canada Symposium: Catalogue no. 11-522-XIE. (<http://www.census.gov/srd/sdc/Massell%20StatCan%20Meth%20Symp%20english.pdf>)

3 For additional discussion about the disclosure risk of contextualized microdata and research informing the design of these files, see the following ICPSR Working Papers by Kristine Witkowski: "Finding a Needle in a Haystack: The Theoretical and Empirical Foundations of Assessing Disclosure Risk for Contextualized Microdata" (<http://deepblue.lib.umich.edu/handle/2027.42/58628>); "Disclosure Risk Components of Contextualized Microdata: Identifying Unique Geographic Units and the Implications for Pinpointing Survey Respondents" (<http://deepblue.lib.umich.edu/handle/2027.42/58627>); "Disclosure Risk of Contextual Data: The Role of Spatial Scale, Identified Geography, and Measurement Detail in Public-Use Files" (<http://deepblue.lib.umich.edu/handle/2027.42/58626>)