Introduction to U.S. Census Data at CSSCR

Written by Grace N. Gu

Center for Social Science Computation & Research
145 Savery Hall
University of Washington
Seattle WA 98195 U.S.A.
(206)543-8110

September 2004

http://julius.csscr.washington.edu/pdf/census02.pdf

INTRODUCTION TO U.S. CENSUS DATA AT CSSCR

Required by the U.S. Constitution, the U.S. decennial Census has been conducted every ten years since 1790 to count the population and housing units for the entire nation. The Census provides population counts that determine how seats in the U.S. House of Representatives are apportioned. Census data are also widely used in drawing congressional and state legislative boundaries, allocating federal and state funds, formulating public policy, and assisting with planning and decision-making in the public and private sectors.

Census Files Available at CSSCR

As one of the distribution centers for the Census Bureau's electronic products, CSSCR houses Census data from 1980, 1990 and 2000, including Summary Tape Files (STF), Summary Files (SF) and Public Use Micro Samples (PUMS). On request, we also help retrieve historical U.S. Census data from the IPUMS-USA database available at http://www.ipums.org. IPUMS-USA is a comprehensive national census database containing integrated microdata from 1850 to 2000. We host a website for the Washington State Census 2000 data files and documentation at http://julius.csscr.washington.edu/Census2000.htm. Like the 1980 and 1990 Census, most of the Census 2000 data that CSSCR provides are SF and PUMS. In addition, CSSCR also distributes TIGER/Line files, a Census geographic product.

STF/SF and PUMS are two major types of Census data products. STF/SF contain tabulated summary statistics for different levels of Census geography from the nation, the state, the county, down to the city block. STF/SF 1 and 2 present tabulated data from the Census short-form (100%) questionnaire. STF/SF 3 and 4 present cross-tabulations of information from the long-form (sample) questionnaire. Tables in STF/SF 2 and 4 are iterated for many detailed racial groups, as well as American Indian and Alaska Native tribes. In SF 4, many data are also tabulated by detailed ancestry groups.

PUMS files present a sample of individual records of responses to the long-form questionnaire with unique identifiers (such as addresses, names, etc.) removed to protect individual confidentiality. Different from the STF/SF which provides summary data information for a specific geographic entity, in PUMS, the basic unit of analysis is a housing unit or a person. PUMS files allow users to prepare their own customized tabulations of most population and housing subjects. There are two sets of PUMS: 5-percent sample file (PUMS-A file) and 1-percent sample file (PUMS-B file). Besides the obvious difference in sampling size, the 5-percent and 1-percent files differ in the geography around which the files are constructed. The 5-percent sample is basically a county level file. The 1-percent sample is basically a metropolitan file. The Public Use Microdata Area (PUMA) is the lowest level of geography identified on any PUMS file. For the 5-percent file, the PUMA can be a single county (or county equivalent), a group of counties, a place, or county/place parts if that county had more than 200,000 persons; for the 1-percent file, the PUMA is an MSA, groups of MSAs, parts of MSAs when the MA is larger than 200,000 persons, and groups of non-metropolitan areas. At CSSCR, PUMS data files are available in SPSS format for all states.

TIGER is the acronym of Topological Integrated Geographic Encoding & Referencing. TIGER/Line files are extracts of geographic and cartographic information from the TIGER

system, describing line segments that represent physical features of the entire U.S. such as roads, railroads, rivers, lakes, political boundaries, census statistical boundaries. The database contains information about these features such as their location in latitude & longitude, name, type of feature, address ranges for most streets, the geographic relationship to other features, and other related information. These files are not graphic images of maps, but rather digital data describing geographic features. With appropriate mapping or GIS software that can import TIGER/Line data, users can produce maps ranging in detail from a neighborhood street to the entire U.S.. The latest version of Census 2000 TIGER/Line files are the 108th CD Census 2000 TIGER/Line files which contain the 108th Congressional Districts. The 108th CD Census 2000 TIGER/Line files also contain the corrected Census 2000 Urban Areas and the 1990 Urban Areas, redefined based on the Census 2000 Urban and Rural criteria. CSSCR website hosts the 108th Census 2000 TIGER/Line files for the State of Washington.

Data Access Media

Most of our Census data are available on CDROM/DVD or through our Web at http://julius.csscr.washington.edu/Census2000.htm. We also have free FTP access to the Census files archived by the Inter-University Consortium for Political & Social Research (ICPSR) (http://www.icpsr.umich.edu/access/subject.html#I). Most online Census data are flat ASCII files which require programming to become usable. For some files, SPSS or SAS data definition statement files are provided along with the data.

CensusCDs from Geolytics

Geolytics publishes detailed Census demographic, housing and geographic data on CDROM with innovative and easy-to-use data extract and mapping software. It allows users to choose geographic entities and data items and easily create reports, dBase files, tab- or comma-delimited ASCII files, thematic maps, and Arcview files.

CSSCR currently has four sets of CensusCDs from Geolytics and plans to purchase more when relevant Census 2000 data become available in the future. The four Geolytics products are: CensusCD+Map, CensusCD Blocks, CensusCD 1980 and StreetCD.

CensusCD+Map contains data from U.S. 1990 Census STF 3 A, B, C & D (3400 variables), 1999 Population Estimates and 2004 Projections, 1999 Consumer Expenditure data, County Time Series Data back to 1970 for crime, industry, federal spending, vital statistics, etc., historical county level population since 1790, and boundaries for all geographies.

CensusCD Blocks contains demographic and housing data and map boundaries for 7 million blocks from the 1990 Census STF 1B and PL94-171 files, the TIGER boundaries, and over 50 geographic identifiers including the 1980 FIPS codes.

CensusCD 1980 contains 1980 Census STF 1 & 3 data and boundary files from the block group level and up.

StreetCD contains all layers of map data from TIGER/Line 1999. Its built-in Window software let users easily select the layer and geography, view output data and maps, and export the selected data into documents or a database. Users can generate layer data in two formats: Arcview shape files and MapInfo MID/MIF files. This integrated data product

significantly reduces the difficulty of acquiring and converting TIGER 1999 street and boundary data.

All Geolytics products are accessible through the computers in the CSSCR lab. To access the Geolytics CensusCDs at CSSCR, follow the procedures below:

- Log in as *public*
- Click on *Start* and select *Run*
- To use CensusCD+Map, type *o:census3.bat*
- To use CensusCD Blocks, type *o:ccdblock.bat*
- To use CensusCD 1980, type *o:ccd1980*
- To use StreetCD, type *o:tiger99.bat*

If you cannot launch any of the programs above, see the consultants on duty.

LandView 6

CSSCR also archives a TIGER related product on DVD called LandView, which is a federal geographic data viewer. LandView 6 contains database management software and mapping software that allows users to create thematic maps of Census 2000 data, browse and query the Census, EPA and USGS databases and show the query results on the map. It also provides the capability to locate a street address or intersection on a map based on Census 2000 TIGER/Line road features and address ranges. With its mapping software, users can create large scale maps showing Census 2000 legal and statistical entities, EPA regulated site locations and USGS GNIS features. It allows users to customize the maps by varying the scale and controlling which map layers are shown. The product also provides search capability for map objects based on radius or map layer and offers tools to users to add information to the maps.

To access Census CDROM/DVDs or Census files provided by ICPSR, please contact CSSCR Archivist at (206) 543-8110 or email gunning@u.washington.edu.

Grace N. Gu
July 7, 2004