H-Net Reviews
in the Humanities & Social Sciences

*CensusCD*. GeoLytics.

**Reviewed by** David L. Elliott

**Published on** H-Demog (May, 1997)

The first time I used a CD-ROM to access 1990 US Census data, the CD was produced by the Census Bureau and held the STF-3A file for New York. My computer was a PCXT clone running DOS using a 2X CD-ROM drive. The Census Bureau's built-in Go program was utterly useless for my purpose which was to obtain machine readable population counts for four New York counties by age, race, and sex. To obtain machine readable counts, I downloaded the Extract program by telephone from the Census Bureau. With a couple of telephone conversations with the good folks at the Census Bureau, I learned how to use Extract. This took a number hours (and a high phone bill) to accomplish. <p> Once I had exported the files created by Extract, columns and rows needed to be transposed to be analyzed. I had exported the files in dBase format, but discovered that dBase would not transpose columns and rows. Consequently, I found it necessary to retrace my steps in Extract, the second time creating files in ascii format. The ascii files were imported into Stata which could be programmed to transpose columns and rows, and then I completed my analysis of the 1990 data using Stata and manually-entered data from the 1980 Census.[1] <p> My next project required only the 1990 STF files.[2] The data extraction was slightly more complicated but I had a 25 MHZ 486 machine with a 4X CD-ROM drive. A T-1 line and FTP allowed for faster downloading of the appropriate Extract files with telephone voice assistance, once again, from the Census Bureau. This data extraction was much faster–thanks not so much to the technology as the trial and error experience the first time around. <p> Recently, I agreed to review the new Windows software, <cite>CensusCD</cite>, created by GeoLytics, Inc. GeoLytics claimed that they had compressed the entire 67 CD-ROMs of 1990 Census data to the point where it could reside on one disc. Some skepticism on Internet lists accompanied this claim. As editor of H-<A HREF="mailto:Demog@msu.edu">Demog@msu.edu</a>, I decided not to post unsubstantiated claims for or against this product until it had been reviewed. <p> I tested the <cite>CensusCD</cite> product on a 200 MHZ Pentium Windows95 machine with an 8X CD-Rom drive. I also duplicated several tests on the 486 (upgraded to 50MHZ) and found little difference in performance. In a word, <cite>CensusCD</cite> appears to fulfill all its producer's claim, and does so quickly and efficiently. Nobody, in my opinion, with the minimum hardware and software requirements for <cite>CensusCD</cite>, should purchase the Census Bureau CD- ROMs. Moreover, I hope that GeoLytics will begin producing 1980 and earlier discs, as well as be prepared to produce the definitive CD-ROM for Census 2000! Furthermore, production of similar products for the censuses of other countries would be welcome, as well. <p> This review will focus on the overall capabilities of <cite>CensusCD</cite> and ease of use rather than a detailed list of what can be done. For the latter, and a Guided Tour of <cite>CensusCD</cite>, the reader should visit GeoLytics' first-rate home page at `http://www.censusCD.com`. <p> Loaded on the CD-ROM is a "5 Minute User Guide" that should be printed out. This is an excellent guide that includes color icons and takes the user step-by-step through the process of selecting and extracting data files. Allow more than five minutes and keep the User Guide handy, as it explains how to make use of the <cite>CensusCD</cite> program. While the researcher familiar with the 1990 Census will probably not need to refer to the User Guide in selecting the geographic area and variables, to export the data, a quick read of the User Guide may prove helpful. <p> The CD-ROM I received came with an upgrade on a 3.5 disc. Due to an apparent error on the

disc, I could not copy the upgrade to my hard drive. A quick (toll) call for technical support revealed that the upgrade could be downloaded from the home page. This was successfully accomplished in a minute or two. Telephone voice help was prompt, friendly, and informative. <p> The <cite>CensusCD</cite> includes all data collected by the Census Bureau in 1990 and published in the STF-3 files: 1.3 billion data for over 375 thousand geographic areas. These geographic areas include 226,398 block groups; 61,258 Census tracts; 29,676 MCDs; 29,467 Zip Codes (as of 1992); 23,435 places (cities and towns); 3,141 counties; 583 American Indian Reservations and trust lands; 428 urban areas; 438 Congressional Distracts (104th Congress); 327 MSA/CNSAs; 80 PMSAs; 51 states (which includes Washington, D.C. as a state for statistical purposes); tweleve Alaska Native Regional Councils; and two sets of groupings of states by the Census Bureau: nine "divisions" and four "regions." Regions outlying the United States that the Census Bureau treats as the statistical equivalent of states–American Samoa, Guam, the Northern Mariana Islands, Palau, Puerto Rico, and the Virgin Islands–are not included with the <cite>CensusCD</cite> product. <p> Also supplied on <cite>CensusCD</cite> is a Windows-based program with extensive help files, including hundreds of pages of Census Bureau STF-3 documentation in hypertext format. While I personally prefer the Census Bureau's loose-leaf documentation for initial browsing and familiarization, searching for specific topics is accomplished far more quickly by using the <cite>CensusCD</cite> help facility. <p> The <cite>CensusCD</cite> program may be run either by mouse or by keyboard. People who cannot (or prefer not to) use the mouse can accomplish everything using the keyboard. I found that with experience, using keyboard and mouse together was fastest. Other than speed, the only drawback to using the keyboard exclusively is that keystrokes open dialog boxes while mouse clicks sometimes open more informative icon-based screens. The latter is particularly useful for the inexperienced <cite>CensusCD</cite> user. I would note that persons unable to use either keyboard or mouse should be able to access <cite>CensusCD</cite> if they have an appropriately programmed voice recognition program–an important asset for the disabled. <p> <cite>CensusCD</cite> terminology generally follows that of the Census Bureau. Researchers familiar with US Census Bureau geographic terms (such as "block group," "tract/bna," and so forth) will have no trouble finding the data they need even though they are sometimes accessed through menus of non-Census terms. Researchers not familiar with Census terms, likewise, will have little trouble selecting appropriate geographical areas as the data are accessed through such intuitively appealing CensusCD terms as "subarea" (leading one from "nation" all the way to "block groups"). Once again, the help facility clearly defines each term–using Census Bureau definitions for Census Bureau terms. In addition to the ability to select geographic area by Census Bureau designation, the researcher may select by Zip Code (STF-3D). One can enter a Zip code, the longitude and latitude, or a street address to obtain the characteristics of the people living in the corresponding Zip code. <p> Even more remarkable, one can enter a street address or longitude and latitude to find characteristics of the population residing in a user-defined radius from the location so defined. The radial area covered follows boundaries of Census-defined areas (block groups, Zip codes, etc.). If the user selects summary statistics, the CensusCD program summarizes the data from all block groups whose centroid falls within the radial area. Or, the user can select any other census area and generate detailed reports for each such area with a centroid falling within the radial area. This was the only feature for which I found the online help and documentation inadequate. Email messages and a telephone call to the GeoLytics help desk, however, clarified how the program worked. <p> All variables reported by the Census Bureau can be accessed for any geographic category down to the block group level. CensusCD groups variables into the following categories: persons, race, education, housing, families, language, income, renter housing, households, occupation, poverty, and owner housing. Also available are imputed values for 178 variables, headers, and a "general profile" of 355 variables under the categories of "social characteristics, "labor force, commuting," "income, poverty," and "housing characteristics." Additionally, a "snapshot" of some 200 characteristics can be generated. The user can define, with the maximum degree of specificity desired, the data count that will be selected. <p> Users may also employ the search engine to select variables or geographic areas. The search engine finds the character string entered and produces all variables containing that exact character string, but does not recognize Boolean operators. Entering simply "age," for example, results in hundreds of variables–good for browsing but if the user has a specific variable in mind, it is best to enter a character string unique to the variable. For example, entering "age under 1" lists thirteen variables: age under 1 year for total population and broken down by sex and race/Hispanic origin. The search engine also recognizes field names (such as P0010001, total population).

Users possessing some familiarity with Census Bureau variables will find this especially expeditious in selecting geographic areas and counts. <p> Once the geographic area(s) and data count(s) have been selected, there are several ways of viewing and using the results. By selecting "view," the database that has been constructed will appear on the screen. Variables appear in columns and records of geographical areas appear in rows. One of the shortcomings is that if the number of records exceeds fourty-six, the column heads scroll off the screen. This disadvantage may be overcome by running the program with dBase selected and viewing the output on-screen; CensusCD also exports ASCII files and a "list" file (suitable for printing), but when viewed on the monitor, these output forms also scroll the column heads off the screen. These column heads will appear as either 8 character field names (such as P0010001) or as a short descriptive name (e.g. TotalPer). <p> A particularly useful feature of <cite>CensusCD</cite> is the ability to format output in several ways. In addition to the capacity to select list, ascii, or dBase formats, in most cases, records may be sorted in ascending or descending order according to a user-selected column variable. The latitude and longitude of the centroid of each geographic area can be selected to appear on each record. Additionally, a geographic key (such as a state abbreviation) can be added to each record. When the area is defined by a radius and small areas appear within the radial area, the bearing of the smaller area's centroid from the user-defined centroid of the radial area may displayed, along with the distance between the two centroids. According to information on the home page, both ascii and dBase files can be imported into mapping programs, but I lack the software to verify this capability. Lastly, a special document file can be generated that provides complete documentation for the short column heads, and other information. <p> One drawback of <cite>CensusCD</cite> is that if the user makes a mistake in selecting an area or count, it is often necessary to abandon the file and start again, unlike the fine help file that allows the user to back up to the prior screen. This is an inconvenience when working fast (and more subject to error), and I found starting all over again frustrating. For that reason, I came to ignore the User Guide's suggestion that the first step should be to name the file; I waited to name the files until after I had successfully selected my variables. <p> In summary, <cite>CensusCD</cite> is a first-rate resource for anyone interested in making use of the 1990 US Census. <cite>CensusCD</cite> is a resource that will be used by professional demographers, students, and business researchers and should be in every college and university library in the United States, and in many public libraries. <p> Notes: <p> [1]. David L. Elliott, "Projecting the Size and Composition of the Elderly Population of Substate Regions for the Year 2000 Using STF3A File on CD-ROM: An Example from the New York State Capital District." <cite>Proceedings of the 1992 International Conference on Applied Demography</cite>, Bowling Green, Ohio, 1994. <p> [2]. David L. Elliott, "Locating Potential Adult Learners in New York State: An Intercensal Test of the 1990 STF3A File on CD-ROM," presented at the Fifth International Conference on Applied Demography, Bowing Green, Ohio, 1994. <p>