MICROCOMPUTER SOFTWARE for CREATING STATISTICS from POPULATION CENSUS MICRODATA for GIS PLANNING APPLICATIONS

(with a description of the REDATAM software)

International Conference on Geographic Information Systems (GIS) Applications for Urban and Regional Planning

3-6 October 1989
Jakarta, Indonesia

Decentralization tendencies in Latin America and the Caribbean

In many countries of the Latin American and Caribbean region there is a growing trend toward institutional and territorial decentralization. The predominance of the central government and capital city is lessening and regional and municipal institutions, to varying degrees, are obtaining an increasing role in the planning and implementation of local development projects and the management of their own affairs. Among the many implications is that sub-national authorities will require suitable technology and training for organizing and utilizing information from many fields.

The relevance of GIS

As many regional and local projects have aspects involving the spatial distribution of population, resources, infrastructure, etc., geographic information systems (GIS) can be employed as "information integrating machines" (Dangermond, 1988), taking advantage of the underlying geography to "relate" data from many different sources and fields, while also facilitating cartographic and spatial analysis of the data. In the developed countries, both the public and private sectors increasingly recognize the benefits of using GIS (Ibid; Tomlinson, 1987), but this is only incipient in most Latin American countries. Until recently, GIS software was very expensive and required large computers, placing GIS out of reach and out of mind for most national as well as sub-national institutions.

Population and housing information

Population and housing statistics are among the important types of information required for specific development activities at the regional and local levels, as well as at the national level. The population statistics required may be relatively conventional measures, such as the total population and density in each area of interest, or may be non-standard as a measure of poverty based on housing and population variables, or estimates of numbers of family members according to their occupational status and their household's characteristics, such as might be employed in an urban traffic model.

The census as source of small-area data

There are frequently various sources of population information in the developing countries including specialized surveys, vital statistics, etc. However, normally only the population and housing census can provide data on each small area comprising any given region, since the census is designed to be universal and the person and housing records have geographical codes for areas usually down to city blocks or smaller. The data in the census for very small areas also allows ad hoc larger zones to be constructed for specific purposes.

As the countries in Latin America and the Caribbean, and in most other developing nations, will carry out a population and housing census in the early years of the 1990 decade, it is important to ensure that the data from the 1990 censuses are suitably packaged and more

readily accessible (with appropriate safeguards) to regional and local users in the private and public sector than was true in the past.

Need for microdata when producing tailored population statistics for small-areas

Pre-conceived standardized aggregate variables, when available for census geographic entities down to the lowest level, may be adequate for some purposes, but information usually has to be specially "tailored" for detailed analyses of specific problems or projects. Researchers and planners may also desire to "interact with the data" to derive improved indicators and understanding. In these situations, or when suitable aggregate data is not available for all levels of census geography, it is imperative to be able to reprocess the census microdata. In the past this has usually meant working on mainframe or minicomputers, since census microdata files for a whole country or large region or city may involve many millions of cases. As with GIS, this places the processing of census data in the developing countries out of the reach of most subnational institutions, since they cannot easily draw on the computing power, resources or programming skills required for mainframe census data processing.

Technology to permit sub-national authorities to use GIS and process microdata

The municipal and other sub-national agencies in a country must have appropriate easy-to-use software working on low-cost equipment, if these institutions are to utilize GIS and are to be able to obtain population census and other data relevant for their normal operations. The rapid introduction of the microcomputer throughout Latin America and the Caribbean since the mid-1980's is making this possible. Furthermore, while high-end GIS software is still quite expensive, it is much less costly than that for large computers and is likely to come down in price. Hence, in the near future the purchase of a GIS should become feasible for many of the larger municipalities and regional authorities.

With respect to the processing of census microdata, most of the powerful statistical processing software available for the microcomputer, such as SPSS or SAS, are not designed to facilitate the rapid, low-cost processing of user-defined small areas, which have been selected from the complete census file containing all variables and their detailed codes, as is often likely to be needed for specific projects or detailed research.

The REDATAM system for tailoring small area census data

The Latin American Demographic Centre (CELADE), in 1983, carried out an examination of public and private sector requests for population information made to their respective national statistical offices (NSO) in each country studied. The major finding was that there was a strong unmet demand for tailored small-area population and housing information and that the NSO were unable to answer most requests within a reasonable period, because they did not have the programmers nor the mainframe time available for reprocessing the census microdata nor the tradition of providing such services.

To help alleviate this problem, CELADE began work in 1985, on the REDATAM system (REtrieval of DATa for small Areas by Microcomputer) to provide the NSO with the technology that would enable them, without using scarce programmer skills, to supply tailored population and housing census tabulations rapidly and at low cost. To permit the efficient creation of

statistics for any small areas of a country or region, the system was designed to store compressed microdata in such a fashion that processing is limited to the records and variables belonging to the area of interest, while maintaining available all records and variables to meet other user needs. REDATAM version 3.1 (CELADE, 1987a,b; 1988), available in English and Spanish, is now used in many Latin American and Caribbean countries for census and similarly structured data.

For more information on REDATAM software,
please refer to the attached description

---

While REDATAM was not originally conceived as a source of population statistics for a GIS, it soon became evident that the next version, to be known as REDATAM-Plus, must have the capability of creating attribute data of the type used by a GIS and must have a convenient means of exporting on demand the aggregate variables to selected GIS. As many smaller municipal offices and researchers in universities, etc., are unlikely to have a GIS in the near future, the database of the new version also will have a multidisciplinary database that stores aggregate data describing geographic entities as well as compressed microdata.

All work involving REDATAM is carried out with the generous support of the International Development Research Centre (IDRC) and with assistance from the United Nations Fund Population Fund (UNFPA) and the Canadian International Development Agency (CIDA).

BIBLIOGRAPHY

Dangermond, Jack, 1988. Trends in GIS and comments. Comput. Environment and Urban Systems, Vol 11, pp.137-159.

CELADE, 1987a, REDATAM Version 2.00 User's Manual. CELADE, Santiago (June 1987). Series A-0173. LC/DEM/G.50. [Use with CELADE, 1988 for version 3.1; also in Spanish].

____, 1987b, REDATAM Database Generation Manual. CELADE, Santiago (October 1987). Series A-0178, LC/DEM/G.53. [also in Spanish].

CELADE, 1988, Supplementary Manual for REDATAM Version 3.1: Supplement to the User's Manual and the Database Generation Manual. CELADE, Santiago (May 1988). Series A-181. LC/DEM/G.50/Add.1. [also in Spanish].

Tomlinson, R.F., 1987. Current and Potential Uses of Geographic Information Systems: The North American Experience. Paper based on a report published within the report: Committee of Enquiry Chaired by Lord Chorley, 1987, Handling of Geographic Information (London: HMSO), Appendix 6.
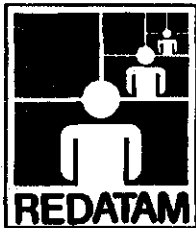
REDATAM 3.1

REtrieval of census DATa for small Areas by Microcomputer

Software, sample database and manuals in English and Spanish

What is REDATAM?

REDATAM is a user-friendly interactive software system designed to make it convenient for population information users to obtain small-area information required for decision making, modelling and related purposes from large census datasets and similarly structured data. REDATAM stores compressed microdata of an entire population and housing census that may involve millions of cases, or the data of a region or city, on an IBM or compatible microcomputer and produces tabulations, usually within minutes, for any variables and for any small geographical areas specified by the user down to city blocks or groupings of such areas.

Planners, researchers and other substantive users can work with REDATAM without assistance from programmers, once the database of interest has been created.

So that time is not lost processing the large amount of data that is always readily available in a REDATAM database, the system processes only the data in the areas selected by the user and only the variables of interest.

What does REDATAM do?

The following outputs can be generated rapidly from a large REDATAM population and housing census or other database for any geographical areas defined by the user:

- Cross-tabulations with up to four variables;
- Averages with up to four classification variables;
- Marginal frequencies;
- Export of the data for the user-defined areas as a DOS flat file or with complete dictionary parameters for direct input into SPSS-PC or SL-MICRO;
- Export of self-contained REDATAM sub-databases for areas defined by the user, permitting work on smaller machines or decentralization of data to sub-regional or municipal offices.

For any output:

- Results can be obtained for sub-areas as well as for the entire user-defined area;
- Cases can be weighted;
- Variables can be recoded;
- New variables can be derived;
- Subpopulations can be selected;
- Processing can be hierarchical, that is, within each household, specific persons can be selected or new household variables derived, such as the number of males attending secondary school by household income category;
- Command sets and geographical area definitions, can be saved for later use or edited to introduce changes;
- The number of cases can be limited or may be sampled;
- The database can be protected by passwords;
- Processing can be executed interactively or in batch.

Extensive help is available on-line. There are also extensive management features to facilitate the construction and editing of REDATAM databases and the addition of new variables or geographical areas to existing databases.

Inputs required

The census or other dataset of interest must first be converted into a REDATAM database. While users can obtain results from an existing database without assistance, a programmer normally is required to make a large census database. The one-time creation of a REDATAM database requires a very clean set of microdata, codebook information on each variable and the geographical hierarchy and associated names. Use of the database may require maps to locate the geographic subdivisions of interest.

For demonstration purposes, the REDATAM software package comes with a ready-made population and housing database with around 7000 persons for the hypothetical country, Miranda, which is also used for all examples and a tutorial.

Salient strengths

REDATAM makes it practical for a user with a microcomputer to have readily available the microdata of an entire hierarchically structured population and housing census, to define any geographic areas of interest and to produce rapidly any tabulations for the selected areas, which may have a few hundred records to hundreds of thousands or more.

Work can be done interactively and results are normally ready promptly depending on the number of records, complexity of the requested data manipulation and tables and the microcomputer speed. Batch processing is also available to make various different runs or process millions of cases without attention.

REDATAM is in English and Spanish and comes with the corresponding manuals, which include a tutorial for persons without previous experience with the system. The context-sensitive help and screens are programmed for easy translation to other languages.

Salient limitations

With the microcomputers now widely available, REDATAM is designed to facilitate small-area access to large microdata files, rather than to be used for regularly producing national-level tables of an entire census with many millions of cases. Processing an entire census can usually be done much more rapidly on a large computer. However, a small country or a city with 500,000 or a 1,000,000 inhabitants and without easy access to a large computer will find REDATAM practical for ad hoc processing its entire file.

As REDATAM is not a survey analysis package like SPSS-PC, it provides only limited statistics: marginal frequencies, cross-tabulations and averages. If more complex statistical output is required, data for the user-defined areas can be exported to SPSS-PC and other survey analysis systems.

The one-time creation of a large REDATAM database is time consuming and complex and often will require a skilled programmer/ analyst or external technical assistance. However, for a dataset such as a household survey already on a microcomputer, it is only necessary to define the data dictionary for the variables and geographical entities and execute the autoload feature.

Known applications of REDATAM in Latin American and Caribbean countries

REDATAM version 3.1 is presently being used with the population and housing censuses in Barbados, Bolivia, Chile, Colombia, Costa Rica, Ecuador, Jamaica, Uruguay, Dominica, Trinidad & Tobago, and Saint Lucia; pre-1990-census experiments in Argentina, Brazil and Mexico; and with survey data for Cuba, the Dominican Republic, Nicaragua, Guyana and Belize. Chile, with a census of 12 million persons and 4 million households uses "WORM" laser disks to store its 300 megabytes, while hard disks are the database storage medium of the other countries, such as Colombia which has separate regional databases totaling 70 mb for its reduced census questionnaire for 27 million persons and 5 million households.

Particular attention is now being directed towards planning the future use of REDATAM with the 1990 censuses to facilitate and improve utilization of the data by the public and private sectors.

Hardware required for REDATAM 3.1

- IBM PC, XT, AT, PS/2 or fully compatible with 640K RAM memory.
- 1 floppy disk.
- Monochrome or graphics monitor.
- Printer with a paper width of at least 80 characters.
- IBM PC-DOS operating system, Version 2.0 or higher.
- Hard disk with around 1.4mb for the system and demonstration database. The storage for an actual database will depend on the record size and number. A 20 megabyte disk will probably be sufficient for the compressed microdata of nearly 1 million census records. Very large databases can use 5 1/4 inch WORM (Write-Once, Read Many) laser disks or CD-ROM.
- REDATAM is written in C; a compiled version is provided.

REDATAM-PLUS: GIS and multidisciplinary and multi-level databases

An entirely new system, REDATAM-Plus is being programmed. With all the previous features, plus many additional capabilities, REDATAM-Plus is directed toward facilitating the utilization of the 1990 population and other censuses by the public and private sectors. Major new features include:

- **Multidisciplinary multi-level database capabilities.** Various hierarchical levels of microdata will be permitted along with aggregate variables (generated from the microdata or imported from other sources) describing the geographical entities.

- **Geographic information system (GIS) interfaces,** including export of REDATAM outputs to pcARC/INFO and MAP-for-the-PC for cartographic display and spatial analysis.

- **Network operation** to allow multiple use of a database.

- **Camera-ready output** as an option.

REDATAM-Plus is being developed with a grant from the International Development Research Centre (IDRC) of Canada, with additional assistance from the United Nations Population Fund (UNFPA) and the Canadian International Development Agency (CIDA). IDRC, with assistance from UNFPA and CIDA, also made possible the development of REDATAM 3.1.

Availability and cost

The REDATAM 3.1 software in English and Spanish is presently available free on request to government offices of developing countries and for US$20 to non-profit institutions. Please specify the language for the manuals.

REDATAM-Plus will be released for general use in early 1990.

Additional information and requests for the REDATAM software should be directed to:

Latin American Demographic Centre (CELADE)
Casilla 91
Santiago, Chile
FAX: 480252 TEL: 485051 TELEX: 340295 TransRadio

Other software available

CELADE also has available related microcomputer software (IBM or compatible) for working with population data:

- **PRODEM** Demographic projections for national and sub-national areas. (Spanish version)

- **PANDEM** Demographic calculations and indirect estimation of fertility and mortality. (English and Spanish versions)

- **LRPM/PC** Long-range planning model for demographic and social sector projections. (English and Spanish versions)

r2descri.res CELamc50 (06/09/1989 <10:03> |)