UNIT 8 - SOCIO-ECONOMIC DATA

UNIT 8 - SOCIO-ECONOMIC DATA

Compiled with assistance from Hugh Calkins, State University of New York at Buffalo

---

For Information that Supplements the Contents of this Unit:

[Links to the following resources have been omitted.]

- Information in Social Context (Chrisman/U of Washington)
- Resources for Geographers -- WWW sites for geographers; sources of geographic information; social science information gateway.
- B.C. Statistics -- Quick facts; infoline; subject areas including: population, labour/social, business/economic; 1996 Census info.
- Statistics Canada -- Canadian statistics on: the land, the people, the economy, the state; internet services at Statistics Canada; other statistical web servers.
- U.S. Census Bureau -- Official statistics; access tools; subject search.
- Government of B.C. Homepage -- Links with hospitals, ministries, crown corporations, municipalities, etc.; subtree of federal and international government links.
- International Data Base (IDB) -- Summary Demographic Data (US Census Bureau).
- International Dataset Guides (CIESIN) -- World resources; US national, social and environmental databank; EuroMab ACCESS; world tables; global population database; etc.
- World data sets (CIA)
- Socio-economic Data and Application Centre (SEDAC) (CIESIN)
- Graphics Illustrating Socio-economic Data -- Business related analysis; employment; housing; population density (DDViewer -- CIESIN)
- Organizing Attribute Data (Foote and Huebner/Geographer's Craft) -- Flat files and spreadsheets, hierarchical files, relational files.
- TIGER -- Coast to coast digital map database (US Census Bureau)
- US Demography Homepage (CIESIN) -- Information gateway; applications development (e.g. integrated population, land use and emissions data; visualization and analysis of integrated assessment models of climate change); SEDAC data products and discussion lists; etc.

It may be useful to illustrate this unit with several different examples of the data products described, including examples of census products such as summary reports, maps and even digital tapes.

UNIT 8 - SOCIO-ECONOMIC DATA

Compiled with assistance from Hugh Calkins, State University of New York at Buffalo

A. INTRODUCTION

Socio-economic data

- are data about humans, human activities, and the space and/or structures used to conduct human activities

- specific classes include
  - demographics (age, sex, ethnic and marital status, education)
  - housing (quality, cost)
  - migration
  - transportation
  - economics (personal incomes, employment, occupations, industry, regional growth)
  - retailing (customer locations, store sites, mailing lists)

Aggregate and disaggregate data

- disaggregated data - data about individuals or single entities, for example:
  - a person's age, sex, level of education, income, occupation, etc.
  - gross sales, number of employees, profit, etc. for a retail store
  - registration number and type for a single vehicle

- aggregated data - describing a group of observations with the grouping made on a defined criterion
  - geographical data are often grouped by spatial units such as a census tract, traffic zone, etc.
  - aggregation can also be by time interval
    - e.g. number of persons leaving area in 5 years
  - also by socio-economic grouping
    - e.g. persons aged 5 through 14 years
  - examples of aggregated data are:
    - number of persons, average income, median housing value for a census tract
    - number of commute trips and average trip length from a suburban traffic zone to the central business district

Cross-sectional and longitudinal data

- recall from Unit 6
  - cross-sectional data gives information on many areas for the same single slice or interval of time
    - e.g. average income in census tracts of Los Angeles for 1988
    - e.g. numbers migrating out of each state in the period 1971-75
  - longitudinal data gives information on one or more areas for a series of times
    - e.g. average income for State of New York from 1970-1988 by year

B. SOCIO-ECONOMIC DATA FOR GIS

Sources of socio-economic data

- field surveys

    - much data used in marketing is gathered by door-to- door or street interview
    - field surveys require careful sampling design
        - how to obtain a representative sample
        - how to avoid bias toward certain groups in street interviews

- government statistics
    - statistics collected and reported by government as part of required activities, e.g. Bureau of the Census
    - usually based on entire population, except sampling is used for some Census questions

- government administrative records
    - records are collected by government as part of administrative functions, e.g. tax records, auto registrations, property taxes
    - these are useful sources of data provided confidentiality can be preserved
    - usually available only to government or for research purposes

- secondary data collected by another group, often for different purposes
    - e.g. the original mandated purpose of the Census was to provide data for congressional districting

- increasingly socio-economic data is available in digital form from private sector companies
    - retailers and direct-mail companies are major clients for these companies
    - includes data originally from census augmented from other sources and surveys
    - data can be customized for clients (special sets of variables, special geographical coverage or aggregation)
    - customizing justifies costs, which are often higher than for "raw" census data

"Geography"

- for use in GIS, socio-economic statistics are of little use without associated "geography," the term often used to describe locational data
    - e.g. data on census tracts must be supported by digital information on locations of census tract boundaries

- geography also allows data to be aggregated geographically, e.g. by merging data on individual cities into metropolitan regions

- thus, many suppliers of socio-economic data also supply digitized geography of reporting zones

- boundaries of many standard types of reporting zones change from time to time
    - e.g. changes occur occasionally in county boundaries
    - e.g. census enumeration districts are redefined for each census (see Redistricting in Unit 56)
    - difficult to assemble longitudinal data for such units due to changing geography

- data is often needed for one set of reporting zones, only available for another set

    - e.g. data available for census tracts, required for school districts which do not follow same boundaries
    - such problems of cross-area estimation are facilitated by GIS technology

- these problems are often grouped into the area of modifiable area problems (MAP)
    - considerable effort has been expended recently to develop statistically sound techniques to deal with these problems (see Openshaw, 1981)

Issues in using secondary socio-economic data

- cost
    - usually secondary data is much less expensive than field surveys
    - large expenditures by government agencies on data collection (e.g. US Census) are indirect subsidies to users, who often pay much less than real cost of data

- documentation
    - quality of documentation, supporting information (e.g. maps) is usually high for data collected by government

- data quality

    - major difficulty is undercounting - census and other social surveys tend to miss certain groups, leading to bias in results
    - undercounting in US Census may be as high as 25% for certain social groups

- data conversion
    - conversion steps may be necessary to make data useful in GIS
        - e.g. format, type of data may be incompatible

- aggregation
    - are data available with suitable level of spatial, temporal aggregation?
        - e.g. study to change elementary school district boundaries will require data at resolution of city blocks or higher
        - e.g. location for gas station will require city block level data, for regional shopping mall much lower resolution (greater aggregation of data) is adequate

- currency
    - social data changes rapidly, can be quickly out of date because of births, deaths, migration, changing economy
    - competitive edge in retailing depends on having current data
    - US has a major census only every 10 years, so its data may be 10 years old
    - often have to estimate current or future patterns based on old data

- accuracy of location
    - census locates people by place of residence - "night-time" census
    - "daytime" data would show locations during the day (place of work, school etc.) but is generally not available from standard sources

    - medical records often locate individuals by place of treatment (hospital), not residence or workplace
        - e.g. consider implications for detecting exposure to cancer-causing agents

C. SOURCES OF SOCIO-ECONOMIC DATA

Population census

- questions on age, sex, income, education, ethnicity, migration, housing quality etc.
- summary statistics used in research, planning, market research, available at high level of geographic resolution in many countries
- see detailed discussion following for US case (Census of Population and Housing)

Economic census

- enumeration and tabulation of business activity is conducted in the US by the Census Bureau in years ending in 2 and 7
- detailed information on classes of industry
- low level of geographic resolution (i.e. large reporting zones)
- data collected in many countries through annual, quarterly or monthly returns of information from companies

Agricultural census

- annual data on crops, yields, livestock etc.
- more extensive periodic surveys of farm economy
- available in spatially disaggregated form to e.g. county level in US

Labor force statistics

- enumeration of employment, unemployment
- produced from periodic (e.g. monthly) sample surveys of workforce
- other special-purpose surveys often combined with regular labor force survey - e.g. household expenditures, recreation activities
- often available for small areas, e.g. parts of city

Land records

- record of land parcel description, ownership and value for taxation purposes

- updated on a regular basis (e.g. annually) by municipality or county government
- also used for land use planning
- source of current demographic information in some countries/states (i.e. local census)
- see detailed discussion following

Transportation and infrastructure inventories

- planning, management and maintenance of facilities
- includes roads and streets, power lines, gas lines, water, sewer lines
- collected by local utilities, responsible government departments
- valuable to variety of users
    - e.g. construction companies needing information on buried pipes
    - e.g. emergency management departments needing data on hazardous facilities
- compiling agency often sees a substantial market for such data which can offset costs of collection

Administrative records

- vehicle registrations, tax returns etc.
- useful for various marketing, research purposes
- based on 100% sample so can be disaggregated spatially
    - however, disaggregation causes problems over confidentiality of records

D. US CENSUS OF POPULATION AND HOUSING

Process of taking the census

- purpose is to enumerate the population for redefining election districts
- taken every ten years (l960, l970, etc.)
- April lst is census day, although complete enumeration takes a "few" weeks
- most households receive forms in mail, some require visit by enumerator

Content

- two types of items - those completed by "100%" of the population, those by random sample

Processing of returns

- automated encoding to digital form
- automated editing to correct obvious inconsistencies
- some missing items can be assigned automatically using simple rules
- other missing items are assigned based on probabilities
- data assembled into master database
- sample surveys processed to produce statistical summaries

Geographic referencing

- initially returns are identified by street address
- address is converted into geographic location using a digital referencing system
  - for the 1980 census, DIME (Dual Independent Map Encoding) files were used for digital geographic referencing of urbanized portions of the US
  - for the 1990 census, TIGER files covering every county will be used
- since TIGER files will have a major impact on GIS databases in the next decade, they are discussed in detail in the next section

Census reporting zones

- range from blocks to states
- as noted previously, the geographic boundaries and definitions of these areas may change from one census to the next

Availability of Census data

- tabulation of statistics by reporting zones, e.g. population by county, population by age by county
- crosstabulation, e.g. population by age and sex by county
- special tabulations, e.g. for unusual combinations of characteristics, or for unusual or custom reporting zones
- number of possible tabulations and crosstabulations is infinite, volume of census products vastly exceeds volume of data collected
- alternative formats for products
  - printed reports
  - magnetic media - tapes, disks
  - microfiche, microfilm, now CDs

sources of census data
  - state data centers distribute Census data
  - private firms repackage and customize data, produce custom reports (e.g. tabulation of population by distance from proposed mall location)

- geography products available
  - base maps showing reporting zones
  - atlases produced for urban areas
  - digital products - boundary files, TIGER

E. TIGER

- reference: beginning of this Unit (TIGER)

Development

- TIGER stands for Topologically Integrated Geographic Encoding and Referencing
- designed to:
  - support pre-census geographic and cartographic functions in preparation for the 1990 Census
  - to complete and evaluate the data collection operations of the census
  - to assist in the analysis of the data as well as to produce new cartographic products
- TIGER files were created by the Bureau of the Census with the assistance of the US Geological Survey

Content

- TIGER/line files are organized by county
- they contain:
  - map features such as roads, railroads and rivers
  - census statistical area boundaries
  - political boundaries
  - in metropolitan areas, address ranges and ZIP codes for streets

Marketing TIGER files

- Census Bureau
  - 1990 Census versions of TIGER/Line files will be available from the Census Bureau in early 1991
    - cost for prototype and precensus TIGER/Line files on magnetic tape are $200 (US) for the first county and $25 for each additional county in that state ordered at the same time
    - the 50 states plus DC on tape cost $87,450
  - precensus files are also available on CD-ROM for $250 per disk, 40 disks are

required for coverage of the entire country (all prices as of Jan. 1990)

- Third party vendors
    - as of December 1989, 25 vendors had notified the Census Bureau that they will market repackaged versions of TIGER/Line files, in many cases with software which will enable users to access this data easily and quickly
    - many of these products are being designed for use on micro-computers

Non-census uses for TIGER

- TIGER files are valuable for other purposes
    - e.g. locating customers from address lists
    - e.g. planning vehicle routes through city streets, for parcel delivery, cab dispatching
    - for these purposes TIGER files need to be kept current at all times, but Bureau of the Census only requires them to be current every 10 years

- see Unit 29 for technical details of TIGER files

F. LAND RECORDS

- many systems have been developed by local governments in the US to manage land, particularly in urban areas

- in other countries there has been more effective coordination at provincial and national levels, e.g. Australia
    - practices in different countries depend on the system of land tenure

- the basic entity in land records systems is the land parcel, i.e. the basic unit of ownership

- traditionally, land records have been managed by hand using methods which often date back 200 years

- land records are the basis of the system of local taxation, administration, as well as transfer of ownership and subdivision

Issues in land records modernization

- accurate land records systems require accurate base mapping at a large enough scale, e.g. 1:1,000
    - such base mapping is not normally available in the US, only the wealthiest governments can afford to create it, e.g. from air photos
    - the term cadaster is used for mapping of land ownership

- the cost of building land records systems can often be recovered, at least partially, from sales of data (e.g.

to utilities, real estate developers) and use in other departments

- the term multi-purpose cadaster (MPC) describes the idea of using the cadaster for many purposes

- because land records systems are being developed independently by many different jurisdictions, there is little standardization of approach, software, etc.

- see Unit 54 for a discussion of MPC applications

REFERENCES

The Bureau of the Census, US Department of Commerce produces numerous documents on the Census and its products, including TIGER. Factfinder for the Nation describes data available from the Census Bureau. Census '90 Basics describe the content, geographic areas and products of the census. Similar material is available from appropriate organizations in other countries, e.g. Statistics Canada.

Marx, R. W., ed, 1990. "The Census Bureau's TIGER System," a special issue of Cartography and Geographic Information Systems Vol 17(1). Contains several articles providing details on the contents and database structure of TIGER.

Kaplan, C.P. and T.L. van Valey, 1980. CENSUS '80: Continuing the Factfinder Tradition, US Department of Commerce, Bureau of the Census. A good review of Census applications.

Richards, D. and P.M. Jones, 1984. "General sources of information," in R.L. Davies and D.S. Rogers, eds., Store Location and Store Assessment Research, John Wiley and Sons, New York, Chapter 4. This chapter reviews sources of socio-economic data in both the US and the UK.

Marx, R.W., 1986. "The TIGER System: Automating the Geographic Structure of the United States Census," Government Publications Review 13:181-201. Discusses the development of the TIGER system

Openshaw, S., 1977. "A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modelling," Institute of British Geographers, Transactions 2(NS):459-72.

Openshaw, S., and P.J. Taylor, 1981. "The modifiable areal unit problem," in N. Wrigley and R.J. Bennett, editors, Quantitative Geography: A British View, Routledge, London.

EXAM AND DISCUSSION QUESTIONS

1. Confidentiality is a major issue in the US Census, and the need to preserve privacy conflicts directly with the need for disaggregated data for numerous purposes. What are the factors to be considered in trying to reconcile these conflicting needs? Is the balance affected by use of GIS?

2. Devise a scheme for creating and maintaining a constantly updated digital file of all streets

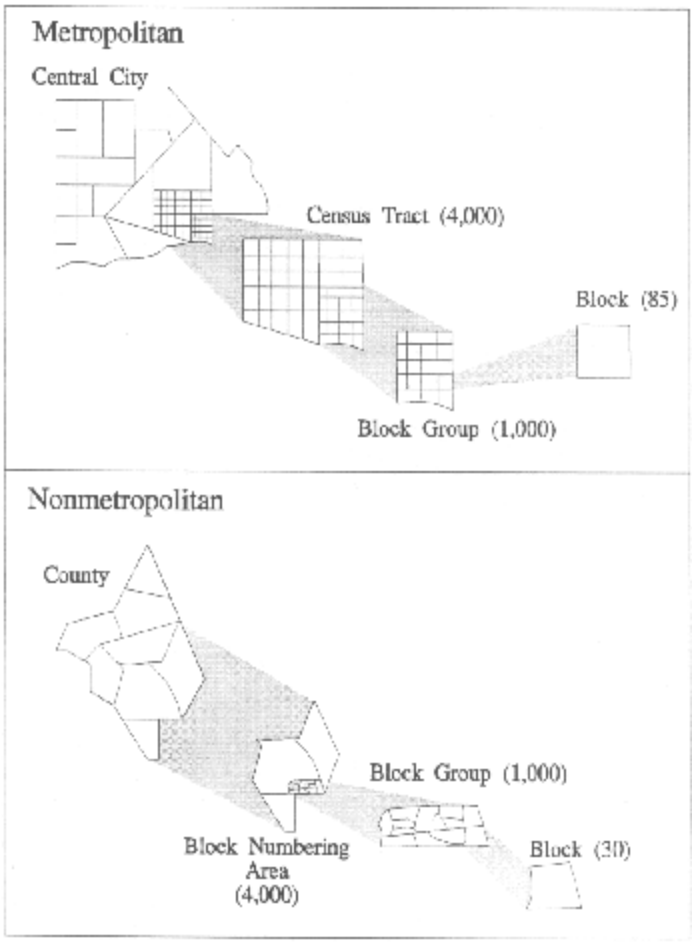
and associated address ranges etc., i.e. a perpetually current TIGER. What would be the costs of the scheme, and what advantages would it have over the current situation?

3. "The concept of a decennial census was devised almost two hundred years ago and has become increasingly inappropriate to the modern age". Discuss.

4. A spreadsheet (such as Lotus 1-2-3) allows the user to perform a variety of functions on tabular data. Discuss the possibility of a "geographical spreadsheet" - what would it do, and what applications would it have it?

Last Updated: August 30, 1997.

UNIT 8 IMAGES



Heirarchy of Census Areas, 1990
(Values in brackets indicate average population)