

Understanding China Census Data with GIS

By

Shuming Bao and Susan Haynie
China Data Center, University of Michigan

The Census data for China provides comprehensive demographic and business information for research, education and consulting services related to China. This presentation will describe and provide examples from two different censuses in China: The 2000 Population Census and the 2004 Economic Census. Types of data and geographic coverage will be described. The presentation will demonstrate how the demographic and business data can be integrated with GIS maps of China at the province, prefecture, county, and township levels, how to estimate the township boundaries, and how to project the demographic data into 1 sq km Grid data using multilayer information. Some innovative toolkits will be introduced for advanced spatial data analysis and mapping of China.

Introduction

The knowledge gained from a population census about the numbers of people, their characteristics, and their housing is of critical importance to governments, businesses, researchers and service providers. This paper describes the Census data that is available on China, and the advantages of using GIS and geo-spatial analysis to help analyze that data. An actual analysis of the areas affected by the May 12, 2008 Sichuan Earthquake illustrates how GIS and spatially-organized information can be used to help respond to a natural disaster.

Geo-demography, the linkage of geographic and demographic information, is a powerful tool, useful for many applications. Information about people and businesses, including where they are located, is very important in many areas of research:

- Academic Research – Social, Demographic, Economic, etc.
- Government – Where should scarce resources be allocated?
- Environmental – What are the environmental impacts of population? What are the human and economic costs from natural disasters?
- Urban Planning – What are the needs of the community for housing, transportation, and other services?
- Business – Where are my best customers? What are their characteristics? Where is a good site, and how should I advertise?

Geospatial analysis can create new information. In the past, the reporting of census information was typically limited to standard administrative units. The power of geospatial analysis allows data to be retrieved for custom geographic areas (for example, a one km radius around a store, or a buffer along a transportation corridor.) In this new era of electronic data, the internet, and GIS, geospatial tools have removed the limitations of printed reports and standard administrative boundaries. The result is a fundamental change in how Census data can be utilized.

This paper addresses information from two censuses conducted in the People's Republic of China:

- 2000 Population Census
- 2004 Economic Census

China's population censuses were previously conducted in 1953, 1964, 1982, and 1990. However, the 2000 Population Census is a "first" in several ways:

- It is the first census to be released outside of China down to the township level. Prior census data are available only at the province, prefecture or county level.
- It is the first time township locations are available in GIS format
- It is the first time approximate township boundaries have been made available outside of China (the actual boundaries are unavailable).
- All variables (approximately 1,500) were released

The 2004 Economic Census is also a "first" for the People's Republic of China. Data on over 5.1 million businesses (economic units with legal status) were collected and tabulated at the postal code level.

2000 China Population Census

A critical aspect of Census data, whether in China, the U.S., or other countries, is that the data are typically reported as geographic totals. Data about individuals are usually not publicly available (at least not in a form that can identify specific individuals). Rather, it is summarized to a variety of geographic levels. Table 1 lists the geographic levels reported in the 2000 Chinese Population Census data, as well as the number of geographies of each type.

Table 1. Chinese Census Geographies

	China	Number of geographies
Largest Census Geographies	China total	1
	Province	31
	Prefecture	345
	County	2,873
Smallest Census Geographies with complete data	Township	50,503
Smallest Geographies – very limited data	1km grid*	9.6 million (approx)

Note: 1km grid data is not reported by the China Census. It is an estimate produced by China Data Center, University of Michigan, U.S.

In the past, Census data was distributed in printed books and reports. The geographic boundaries were available as paper maps only. The U.S. delivers Census data and boundaries in GIS formats, however, the Chinese government has never released any official GIS boundary files for the Chinese Census data.

Most access has been:

- In printed form
- As summaries for relatively large geographic areas
- Without geographic location information (GIS)

An important step was to develop a spatial GIS (Geographic Information System) database for China's population distribution. A portion of this work has been accomplished by China Data Center, producing:

- Geographic coordinates for the townships (allowing for GIS and geospatial analysis)
- Approximate township boundaries (the official boundaries may not be released outside of China)
- Estimates of population for 1 km² grids throughout China, based on the township data, land area calculations, population density, and elevation and slope data.

This ground-breaking work was needed before any of the township-level Census information could be displayed in GIS.

The extensive Population Census data includes information on:

- Total, agricultural, family and collective population
- Age by gender
- Births and Deaths
- Occupation and Industry
- Education and Literacy
- Housing Characteristics (Price paid, Rent amount, facilities, etc.)
- Registered Population
- Race and Ethnicity
- Migration
- Marital Status

Similar to the U.S. Census, two types of questionnaires (long form and short form) were used for the 2000 China population census. The short form questions were collected on the entire population, while the long form was collected on a sample of approximately 9.5% of the population. The undercount is estimated at about 1.8%, based on a post-enumeration survey.

2004 China Economic Census

Not only can censuses be used to collect information about people, they can also be used to collect information about other things – including businesses. According to the “Communique on Major Data of the First National Economic Census of China:

“China conducted its first economic census in 2004 with the aim to keep abreast of the development of the secondary and tertiary industries of China in terms of their size, structure and economic results, to establish a sound system of the registers and databases of the basic units, and to provide the foundation for preparing plans for national socio-economic development and improving policy-making and management. The reference time for the economic census was December 31st of 2004, and the flow data covered the whole year of 2004. The economic census covered all legal person units, establishments and self-employed individuals who were engaged in the secondary and tertiary industries within the territory of China. The main contents of the census include the basic characteristics of the units, the employees, the financial situations, the production and business operation situations, the production capacity, the consumption of raw materials and energy, as well as the scientific and technological activities.”

(http://english.gov.cn/2006-03/03/content_217134.htm):

Not all data from the 2004 Economic Census were released internationally. The available data include:

- Industry classification (863 classifications)
- Employee size (10 categories)
- Revenue Range in Yuan (15 categories)
- Ownership status (23 categories)

This business information is available at a variety of geographic levels, the smallest being postal code (approximately 40,000 in China). The postal codes are also summarized (aggregated) to a variety of geographic levels, similar to those reported for the 2000 population census. China Data Center has facilitated the acquisition of spatial location data for the postal codes, allowing for display in GIS.

Business Census data can provide valuable insight into:

- Level of competition – How many competitors exist in a trade area?
- Availability of services – For example, convenient access to an office supply store or access to shipping and transportation services.
- Total employment in an area. – Often referred to as “daytime” population, people sometimes shop where they work, rather than where they live. Persons employed in a store’s trade area can be an important source of customers.

Data Examples

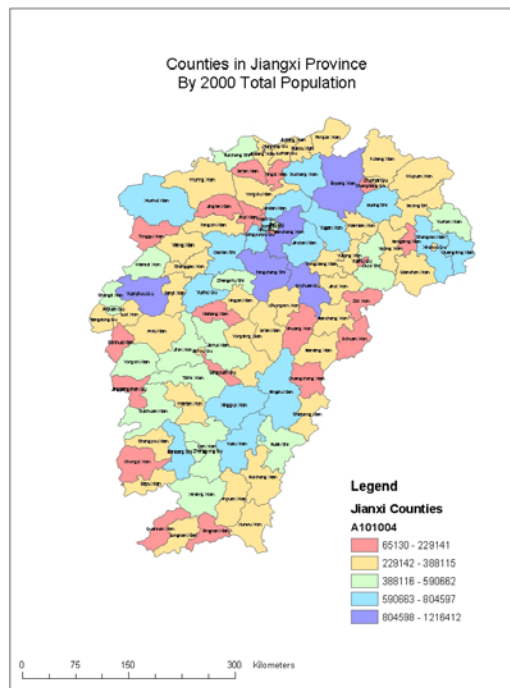


Figure 1. Map of 2000 Population Totals - Counties in Jiangxi Province

China Data Center's work to provide geographically-oriented township-level data is a huge step forward in gaining a more detailed understanding of the spatial distribution of population in China. Most demographic information currently available for China is at the county level or larger. Figure 1 shows a map of total population by county in Jiangxi Province. Townships are much smaller in land area.

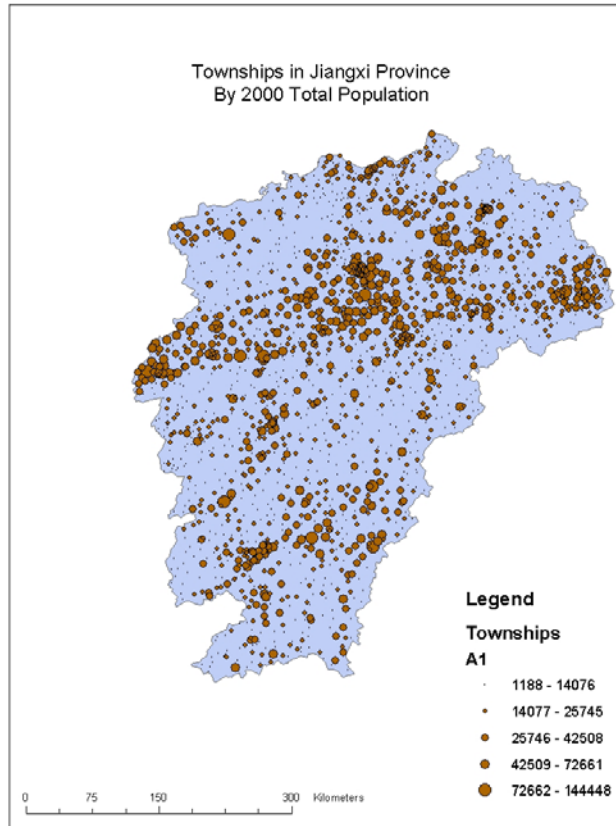


Figure 2. The 2000 Population by Townships of Jiangxi Province, China

Figure 2 displays townships as symbols on the map of Jiangxi province. The larger symbols represent townships with larger populations. Township level data provides a much more precise picture of the 2000 population distribution in Jiangxi province than county level. This map by township can help identify areas with larger concentrations of people – and greater numbers of potential customers.



Figure 3. Potential new site in XuZhou, China

Figure 3 shows a closer view of a potential site for a new business location. Buffers have been drawn around the location (one-half, one, and two km radii). Townships are shown with varying sizes, representing the 2000 population. Demographics can be summarized to the total trade area (radius), allowing for more objective comparisons among multiple sites.

Analysis of Sichuan Earthquake - May 12, 2008

Recently, following the devastating Sichuan earthquake of May 12, 2008, China Data Center facilitated demographic and economic analysis of the affected geographic regions. Using GIS and geospatial analysis, information and maps from the 2000 Chinese Population Census and 2004 Economic Census were delivered to key organizations in an effort to assist in the relief efforts. Initial data for a 60 mile radius around the epicenter was distributed by China Data Center on the same day as the earthquake, in part due to the speed and efficiency of the software.

China Data Center released an initial assessment of the potential population and local economy affected by the earthquake (see <http://chinadatacenter.org/presentation/earthquake/wenchuan.htm>). The report helped people understand the impact of the earthquake on the population. The

information was widely cited by the news media, including AP, USNews, YahooNews, NewsWeek, NewYork Times and many others. IFRC (International Federation of Red Cross and Red Crescent) and other relief agencies in China have been using the information to help evaluate the capacity of local health facilities, along with estimating the extent of the population and businesses impacted within the quake zones.

On May 12, 2008 the following map and report were created and posted online by the China Data Center. Initial studies were for a 60 mile radius around the epicenter. The radius is illustrated on the map in Figure 4, along with a map of total population by province from the 2000 Census.

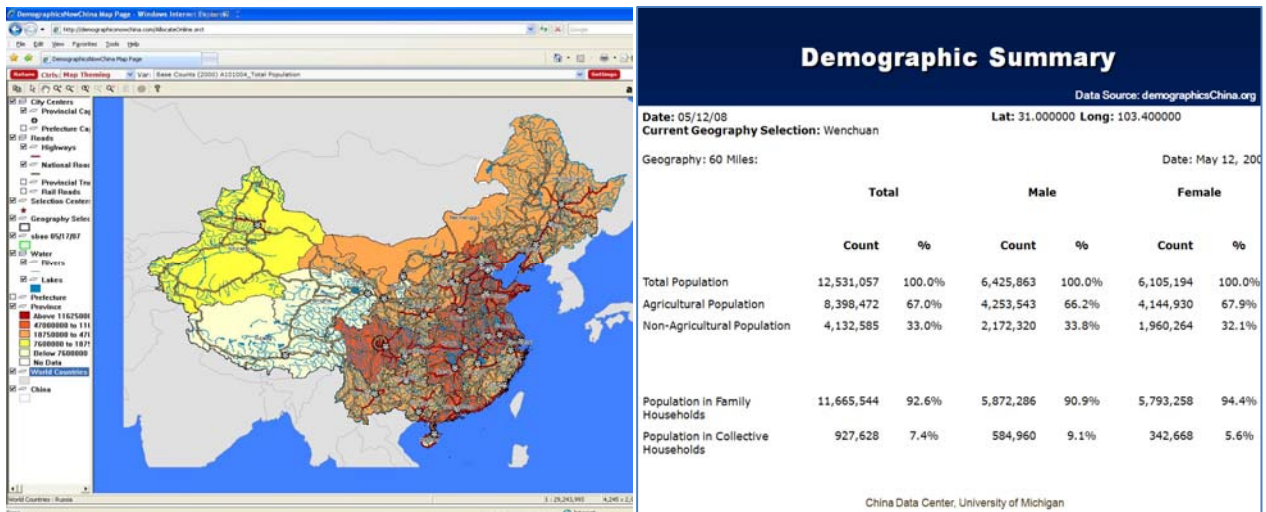


Figure 4. 2000 Population by Province and Demographic summary for 60 mile radius around 5-12-08 earthquake epicenter.

The report in Figure 4 shows counts of the 2000 population by gender, agricultural status, and type of household (family or collective). This information was precisely retrieved for the 60 mile radius, using underlying population estimates for 1 km² grids (produced by the China Data Center). The spatial distribution of the population at this small level is used to apportion township level data, and aggregate it for the 60 mile radius. Once the latitude and longitude of the epicenter was known, summary reports of the area were generated in seconds.

As more precise delineations of the affected regions were released by the USGS (United States Geological Survey Earthquake Hazards Program), the demographic analyses were updated to provide estimates of the population affected by varying

levels of intensity. The “Quakeshake” areas were loaded and precise estimates of the population within each “Ring” were retrieved. Ring 9 was the area of most intense shaking with “Extreme Perceived Shaking”, and “Very Heavy Potential Damage.” <http://earthquake.usgs.gov/eqcenter/shakemap/global/shake/2008ryan/>

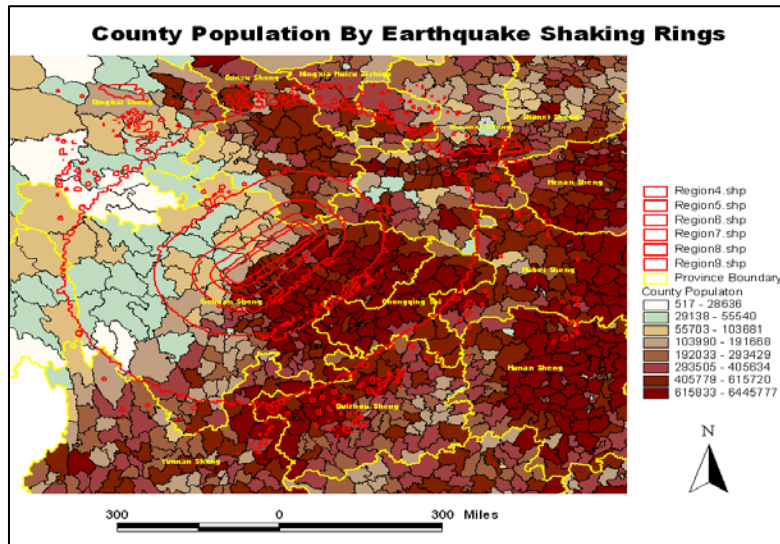


Figure 5. 2000 County Population, USGS “Quakeshake” areas

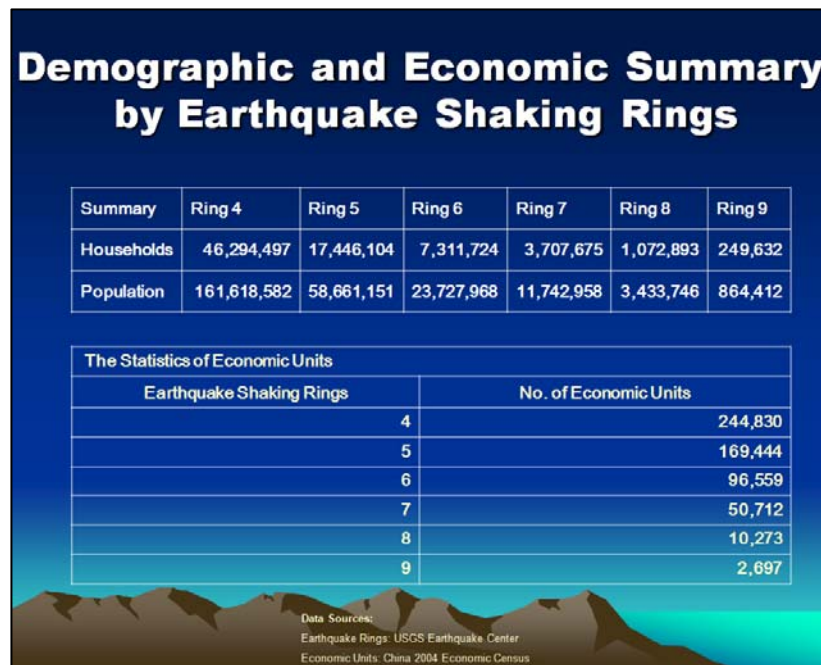


Figure 6. 2000 Demographic and 2004 Economic Summaries for regions of 6 earthquake intensity levels.

China Data Center's ability to quickly analyze the impact of the Sichuan earthquake has demonstrated how critical spatial information can be in response to emerging events. This real life and death application of the technology demonstrates that timely distribution of data customized to the area of need is very useful in the event of a major disaster.

Conclusion

Census data for large geographic areas in China historically has been relatively accessible, for instance, at the country and province level. Getting access to the smallest census geographies has been a different challenge.

China Data Center (University of Michigan) has collaborated with the All China Marketing Research Company (Beijing) to make the 2000 Chinese Population Census and 2004 Chinese Economic Census data available internationally. Chinese 2000 census data (at province, county and township levels) and 2004 Economic Census data is available for purchase through China Data Center of the University of Michigan at <http://chinadatacenter.org/>.

Contact Information:

Susan Haynie, Ph.D.
Demographic Consulting, Inc.
Tel: 714-558-1529
Email: shaynie@geodemo.com

Shuming Bao, Ph.D.
China Data Center, the University of Michigan
Ann Arbor, MI 48109-1106
Tel: 734-647-9610
Fax: 734-763-0335
Email: chinadata@umich.edu