

# Web 人口統計による小規模人口の発見

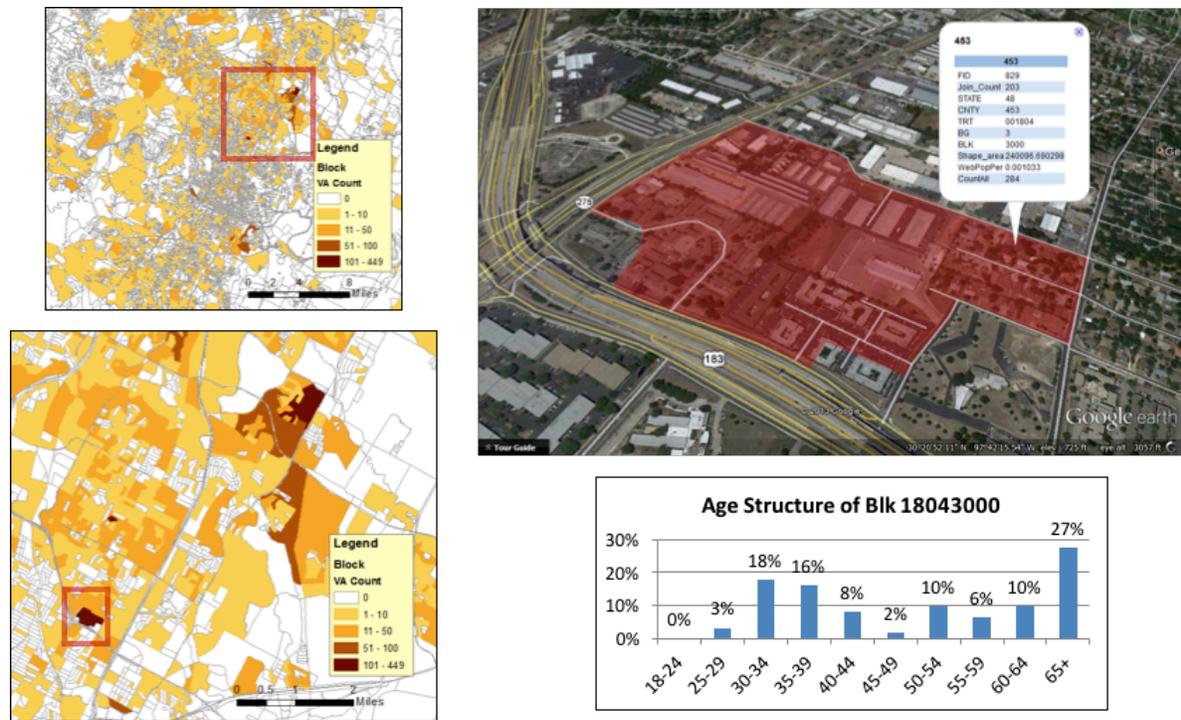
## Discovery of Small Area Population through Web Demographics

T. Edwin Chow  
Department of Geography  
Texas State University  
[chow@txstate.edu](mailto:chow@txstate.edu)

Geographic information provided by crowdsourcing (e.g. open street map), social media, and data from Information and Communication Technology are useful sources of big data which can contribute to the development and understanding of a specific spatiotemporal landscape concerning an unknown population. However, the inherent anonymity of these crowd-sourced or crowd-harvested data sources lack the socioeconomic and demographic attributes to examine human dynamics and patterns revealed from the data. In this paper, we investigate an Internet-based demographic data source, personal microdata databases on the World Wide Web (hereafter web demographics), as potential sources of aspatial and spatiotemporal information regarding the landscape of human dynamics. The objective of this paper is to explore their geographic and demographic patterns (Figure 1).

People search engines, such as WhitePages ([www.whitepages.com](http://www.whitepages.com)), gather personal-level microdata (hereafter web demographics), including full name, address, age, phone number and household members. This research highlights the emergence of web demographics as a new form of demographic data that are made searchable on the Internet by people finder sites like WhitePages. Based on surname analysis, it is possible to use surnames that are highly predictive to an ethnic group to solicit their associated web demographics and conduct a “web census” (Chow et al., 2011). Comparing the spatial pattern of web demographics collected in 2009 and Census 2000 data revealed demographic trends that conform to general understanding of spatial demography, including the urban-to-suburban and rural-to-urban migration in major metropolitan areas of Texas. As web demographics contain personal level microdata, the spatial and temporal resolutions can be flexible for space-time demographic research.

Using web demographics of Vietnamese-Americans in Texas collected in 2010 as a case study, this paper: 1) identifies small area population through the application of a Cost-Sensitive Alternative Decision Tree (CS-ADT) algorithm, 2) analyze the demographic characteristics of small area population and the functional relationship with travel distance, and 3) investigates migration pathways and clusters to include both short- and long-distance patterns. By linking the physical space at the individual level, this unique methodology attempts to enhance the understanding of the spatial distribution and dynamic pattern of small area population at multiple spatial scales (Chow, 2013).



**Figure 1.** The spatial distribution of Vietnamese-American in Austin, Texas at the block level and their demographic structure.

## References

- Chow, T. E., 2013. "We know who you are and we know where you live": A research agenda of web demographics, *Crowdsourcing Geographic Knowledge. Volunteered Geographic Information (VGI) in Theory and Practice*. Edited by Sui D., S. Elmwood and M. Goodchild, (Springer). pp. 265-285.
- Chow, T. E., Y. Lin, & W. D. Chan. 2011, The Development of a Web-based Demographic Data Extraction Tool for Population Monitoring, *Transactions in GIS*. 15(4): 479-494.