Demographic Data

Jennifer Ferreira

Accepted manuscript PDF deposited in Coventry University's Repository

Demographic Data

Demography, formed from two Greek words broadly means 'description of the people', and in general refers to the study of populations, processes and characteristics including, population growth, fertility, mortality, migration, and population ageing while the characteristics examined are as varied as age, sex, birthplace, family structure, health, education and occupation. Demographic data refers to the information that is gained about these characteristics which can be used to examine changes and behaviours of population, and in turn be used to generate population predictions and models. Demographic data can be used to explore population dynamics, analytical approaches to population change, the demographic transition, demographic models, spatial patterns, as well as planning, policy making and commercial applications especially projecting and estimating population composition and behaviour.

 Applied demography seeks to emphasize the potential for the practical application of demographic data to examine present and future demographic characteristics, across both time and space. Where demographic data is available over time this allows for historical changes to populations to be examined in order to make predictions and develop models about how populations may behave in the future. Traditionally the principal sources for the study of population are censuses and population surveys which are often infrequent and not always comprehensive. The generation of demographic data is a key element of many big data sets. It is the knowledge about people which can be gleaned from big data which has the potential to make these data sets even more useful not only for researchers but also policy makers and commercial enterprises.

Understanding demographic data and population patterns has useful applications in many areas, indulging planning, policy making and in commercial enterprises. In planning, estimates and projections are important in terms of ensuring accurate allocation of resources according to population size, determining the level of investment needed for particular places. Planning requires reliable demographic data in order to make decisions about future requirements, and so will impact on many major planning decisions, and therefore many large financial decisions are made on the basis of demographic data in combination with other information. Population statistics reveal much about the nature of society, the changes that take place within it, and the issues that are relevant for government and policy. Therefore, demographic projections and models can also stimulate actions in policy making,

in terms of how to meet the needs of present and future populations. A key example of this relates to ageing populations in some developed countries where projections of how this is likely to continue have informed policies around funding pensions and healthcare provision for the elderly. For businesses, demographic data is a vital source of information about their consumer base (or potential consumer base); understanding how consumers behave can inform their activities, or how they target particular cohorts or segments of the population. The wide relevance of demographic data for policy, planning, research and commerce means that this particular aspect of big data has experienced much attention.

Many of the traditional data sets which generate demographic data, such as the census, are not conducted frequently (every 10 years in the UK) and are slow to release the results, so any analysis conducted on a particular population will often be significantly out of date given the constant changing dynamics of human populations. Furthermore, other population surveys draw on a relatively small sample of the population and so may not be truly representative of the range of situations experienced in a population.

Demographic data refers to data which relates to a particular population which is used to identify particular characteristics or features. There are a wide range of demographic variables which could be included in this category, although these commonly used include: age, gender, ethnicity, health, income, employment status. Demographic data has been a key area which has been the focus for big data research, primarily because much of the data generated relates to individuals and therefore has the potential to provide insights into the characteristics and behaviours of population beyond what is possible from traditional demographic data sources. Demographic data is also vital for many commercial enterprises as they seek to explore the demographic profile of their customer base, or the customer base whom they wish to target for their products.

This vast new trove of big data generated by new technological advancements (mobile phone, computers, satellites and other electronic devices) has the potential to transform spatial-temporal analyses of demographic behaviour, particularly related to economic activity. As a result of the technological advancements which have led to the generation of many big data sets the quantity of demographic data available for population research is increasing exponentially. Where there are consistent large scale data sets that now extend over many years sometimes crossing national boundaries with fine geographic detail this collectively creates a unique laboratory for studying demographic processes and for examining social and

economic scenarios. These models are then used in order to explore population changes including fertility, mortality and depopulation.

The growth in the use of big data in demographic research is reflected in its growing presence in discussions at academic conferences. In the Population Association for America in 2014 a session entitled 'Big Data for Demographic Research' demonstrates some of the ways big data has been used.

Mobile phone data in Estonia was used to examine ethnic segregation. This data set included information about the ethnicity of individuals (Russian/Estonian), the history of locations visited by the individuals, and their phone-based interactions. This study found evidence to suggest that ethnic composition of an individual's geographic neighbourhood influenced the structure of an individual's geographic network. It also found that patterns of segregation were evident where migrants were more likely to interact with other individuals of their ethnicity. A further study also used mobile phone data to explore human mobility. This project highlighted the potential that large scale data sets like this, have for studying human behaviour on a scale not previously possible. It argued that some measures of mobility using mobile data are contaminated by infrastructure, and demographic and social characteristics of a population. The authors also highlight problems with using mobile phone data to explore mobility and outline potential new methods to measure mobility in ways which respond to these concerns. The measures developed were designed to address the spatial and social nature of human mobility, to remain independent of social, economic, political, or demographic characteristics of context, and to be comparable across geographic regions and time.

The generation of big data via social media has also led to the development of new research methods Pablo Mareos and Jorge Durand explore the potential value of netnographic methods in social media for migration studies. To do this the researchers explore data obtained from internet discussion forums on migration and citizenship. This research uses a combination of classification methods to analyse discussion themes of migration and citizenship. The study revealed results which identified key migrating practices which were absent from migration and citizenship literature; suggesting that analyses of big data may provide new avenues for research, with the potential for this to revolutionize traditional population research methods.

Capturing demographic patterns from big data is a key activity for researchers, political teams, marketing teams. Being able to examine the behaviour of populations, but in particular specific segments of population is a key activity. While many of the techniques and technologies which harness big data may have been developed by commercial enterprises or for commercial gain, there are an increasing number of examples of where this data is being used for public benefit, as seen in Chicago.

In Chicago, health officials employed a data analytics firms to conduct data mining to explore ethnic minority women who were not getting breast screenings even though try were offered them for free at a hospital in a particular area. The analytics firm helped Chicago health department to refine its city outreach for breast cancer screening programme by using big data to identify the uninsured women aged 40 and older living in the south side of the city. This project indicates the potential impact that big data could have on public services.

There are of course challenges associated with using demographic data collected in data sets. The use of twitter data for example, to examine population patterns or trends though is problematic in that population of twitter users is not necessarily representatives of the wider population, or the population being studied. The massive popularity of social media, and the ability to extract the data about communication behaviours has made them a valuable data source. However, a study which compared the twitter population to the US population along three axes (geography, gender and race/ethnicity) found that the twitter population is a highly non-uniform sample of the population. Ideally when comparing the twitter population to society as a whole we would compare properties including socio-economic status, education level and type of employment. However, it is only possible to obtain characteristics which are self-reported and made visible by the user in the twitter profile (usually the name, location and text included in the tweet). Research has indicated that twitter users are more likely to live within densely populated area, and that sparsely population regions are underrepresented. Furthermore, research suggested that there is a male bias in twitter users, again making the results gained from this big data source unrepresentative of the wider population.

Despite the challenges and limitations associated with big data, the study of populations, their characteristics and behaviours is a growing area for big data researchers. The applications for demographic data analysis are being adopted and explored by data scientists in both the public and private sector in an effort to explore the past, present and future patterns of populations across the world.

Jennifer Ferreira

Coventry University

See Also: Education; Epidemiology; Geography; International Labor Organization World Health Organization

Further Readings

Blumenstock, J. and Toomet, Ott. *Segregation and 'Silent Separation': Using Large-Scale Network Data to Model the Determinants of Ethnic Segregation*. Paper presented at the Population Association of America 2014 Annual Meeting. Boston May 1-3. (2014).

Girosi, Federico and King, Gary. *Demographic Forecasting*. Princeton University Press. (2008).

Mateos, Pablo, and Durand, Jorge. *Netnography and Demography: Mining Internet Discussion Forums on Migration and Citizenship*. Paper presented at the Population Association of America 2014 Annual Meeting. Boston May 1-3. (2014).

Mislove, Alan., Lehmann, Sune., Ahn, Yong-Yeol., Onnela, Jukka-Pekka., Rosenquist, Niels. *Understanding the Demographics of Twitter Users* Proceedings of the Fifth International AAAI Conference on Weblogs on and Social Media. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewFile/2816/3234

Ruggles, Steven. Big Microdata for Population Research. *Demography*. 51(1): 287-297 (2014).

Rowland, Donald. *Demographic Methods and Concepts*. Oxford University Press: Oxford. (2003).

Sobek, Matthew., Cleveland, Lara., Flood, Sarah., Hall, Patricia. King, Miriam., Ruggles, Steven and Shroeder, M. Big Data: Large Historical Infrastructure from the Minnesota Population Center. *Historical Methods* 44(2):61-68. (2011).

Williams, Nathalie., Thomas, Timothy., Dunbar, Matt., Eagle, Nathan., Dobra, Adrian. *Measurement of Human Mobility Using Cell Phone Data: Developing Big Data for Demographic Science*. Paper presented at the Population Association of America 2014 Annual Meeting. Boston May 1-3. (2014).