

Big Data and Gender Analysis

Summary of Pilot Projects April 2017

Data2X has supported a variety of pilot projects exploring how promising sources of big data – cell phone call detail records, credit card records, satellite datasets, and social media – can help close global gender gaps. The work was exploratory and aimed to demonstrate what big data can and cannot achieve. Data2X has synthesized the outcomes of the pilots into a report, “Big Data and the Well-Being of Women and Girls: Applications on the Social Scientific Frontier,” which explains what has been learned and ways forward for using big data to analyze gender questions.

Pilot Projects

Digital Exhaust: Call Detail Records and Credit Card Records

Cell phones and other digital records, such as credit card transactions, are potentially a major source of information about women’s lives and well-being. Unlocking this potential depends on either having information on a user’s sex made available to researchers, or constructing models for automatically predicting the sex of cell phone users while preserving anonymity. Data2X supported researchers to utilize a combination of digital call and credit card records, a subset of which included sex information, to analyze women’s economic lifestyles and mobility patterns. These methodologies can be used to close significant gender data gaps, by illuminating critical information on access to markets, health facilities, and government services.

Satellite Datasets

Improvements in computing power are enabling analysis of the tremendous amount of increasingly high-resolution satellite data. Data2X supported the Flowminder Foundation to create models that utilize satellite data to increase the spatial resolution of existing information from standard surveys, such as the DHS, on key indicators of women’s welfare, mortality, morbidity, malnutrition, contraceptive access, freedom of movement, and other topics. Such data is often only available at district or higher geographical levels; geospatial information from satellites can help generate highly localized information throughout a country. Flowminder’s work covers a range of countries, including Bangladesh, Haiti, Kenya, Nigeria and Tanzania.

Social Media

Data2X also supported several projects using social media data. In 2016, our partner United Nations Global Pulse analyzed ideas and opinions expressed on Twitter from women all across the world on topics related to the Post-2015 Development Agenda, developing an [open-source toolkit](#) for automatically identifying the sex of Twitter users with 74% accuracy. Data2X also supported Dr. Munmun de Choudhury of Georgia Tech University to construct Twitter-based models to estimate regional prevalence of depression among adolescent girls in South Africa, India, the United States and the United Kingdom. Given that depression is one of the leading global causes of girls’ illnesses and little information is available on the topic, these methodologies hold great promise for filling a key gender data gap.