Big Data for Development

Malarvizhi Veerappan
Senior Data Scientist

@worldbankdata
@malarv

Why is Big Data relevant for Development?

In developing countries there are large data gaps, both in quantity of available data and in data quality

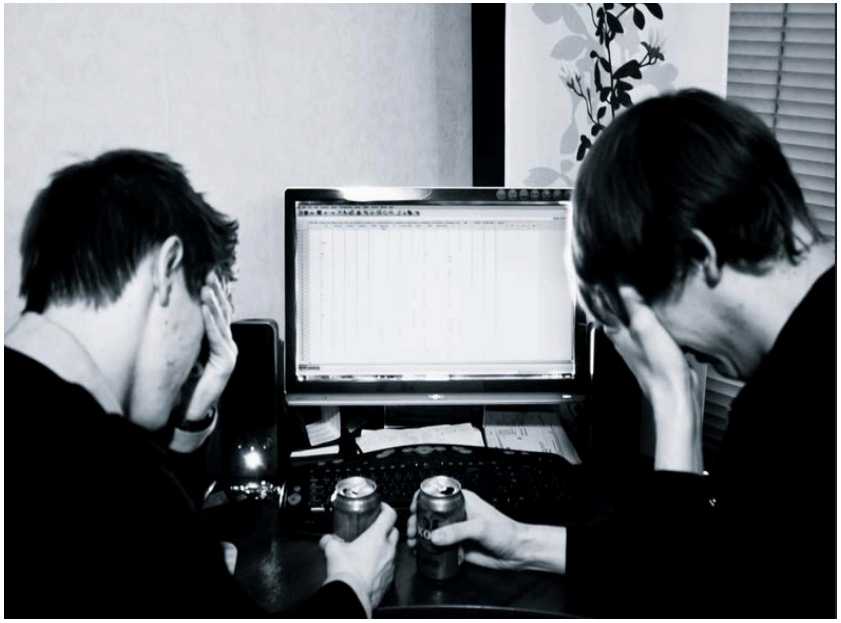
- Of the 60 countries with complete vital statistics, not one is in Africa[1]
- Of the World's 6.5 billion people, "we do not have accurate information on more than 4 billion," said Margaret Chan, director-general of the WHO

Can Big Data help close these gaps? We still do not have a conclusive answer, but the potential rewards –in terms of filling data quantity and data quality- are very large

1. Source: Delivering on the Data revolution in Sub-Saharan Africa

Opportunity: Improve data production process…

Inefficient data systems



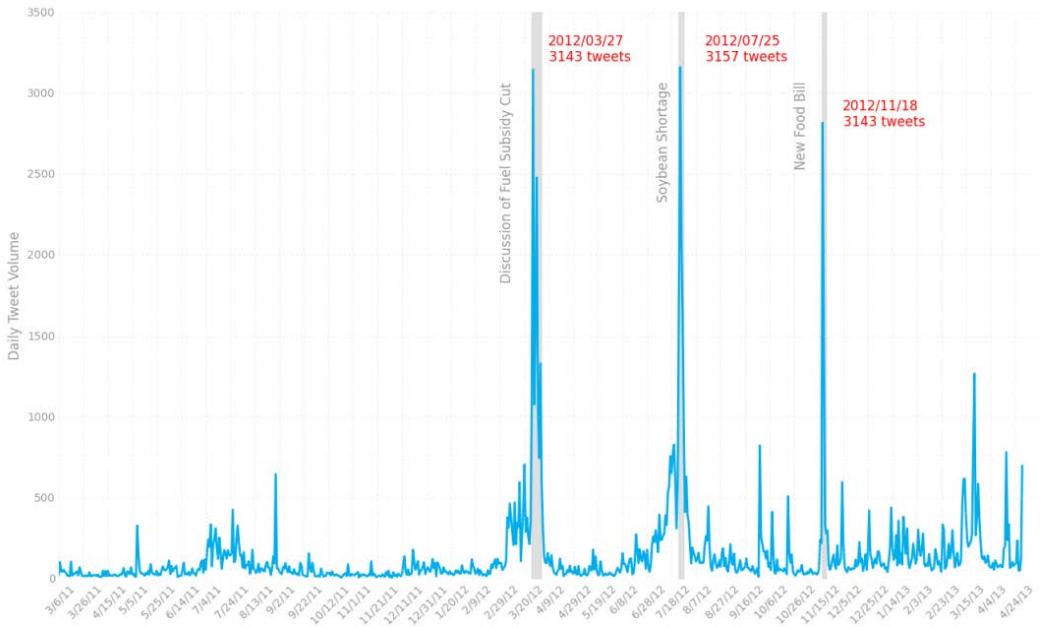
Leapfrogging – new innovations in developing and emerging economies

Examples of big data analysis relevant to development

Real-time monitoring

Mining Tweets in Indonesia to Understand Food Price Crises

Daily Tweet Volume Related to Food Price Rise (March 2011 – April 2013)



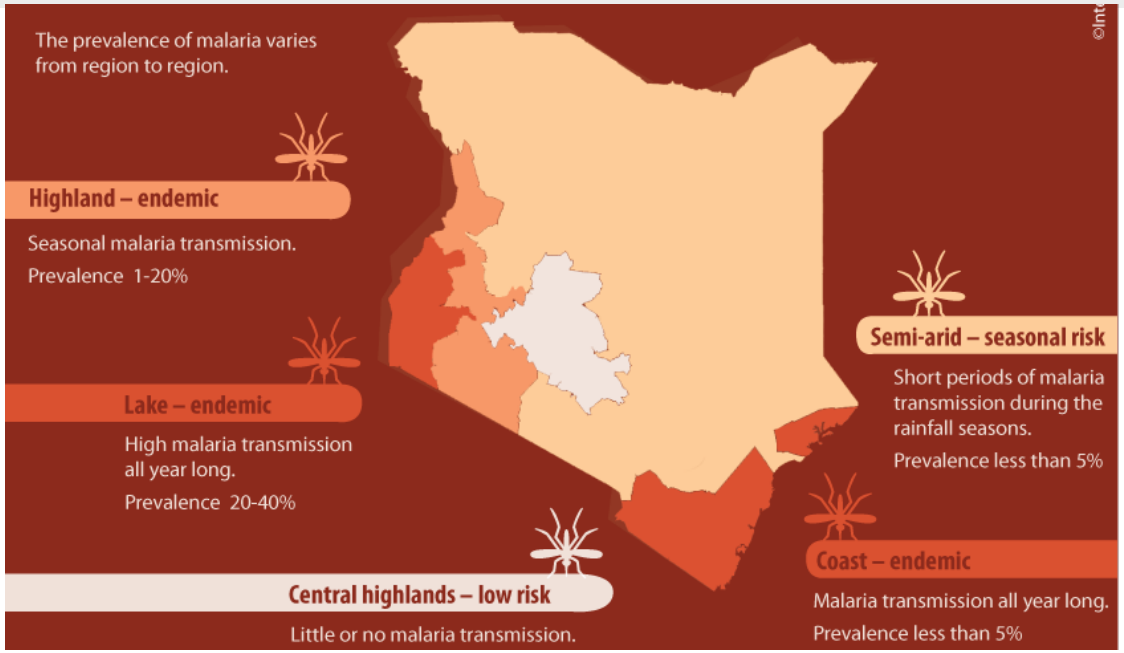
Monitor food-price related tweets between January 2011 and December 2012 to see if variations in tweet volumes could be connected with food and fuel price inflation.

Results: Conversations related to food prices spiked dramatically among Indonesian Twitter users, corresponding to events such as global soybean price rise, proposed cut in food subsidy, etc. **This illustrates the potential value of employing regular social media analysis for early warning and impact monitoring.**

Source: UN Global Pulse (2012)

Local disaggregation

Tracking Malaria in Kenya using Cellphones



Track texts and calls from nearly 15 million cellphones in Kenya for an entire year and use the data to make a map for how malaria spreads around the country.

Results: The results were unexpected. The roads to and from the capital city, Nairobi, are the most heavily traveled, yet they are not the most important for spreading the disease throughout the country. Instead, regional routes around Lake Victoria serve as the major disease corridors for the parasite. And, towns along the routes are hot spots for transmitting malaria to the rest of the country.

Source: Harvard School of Public Health: Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO. Science. 2012 Oct 12;338(6104):267-70.

Behavioral Insights

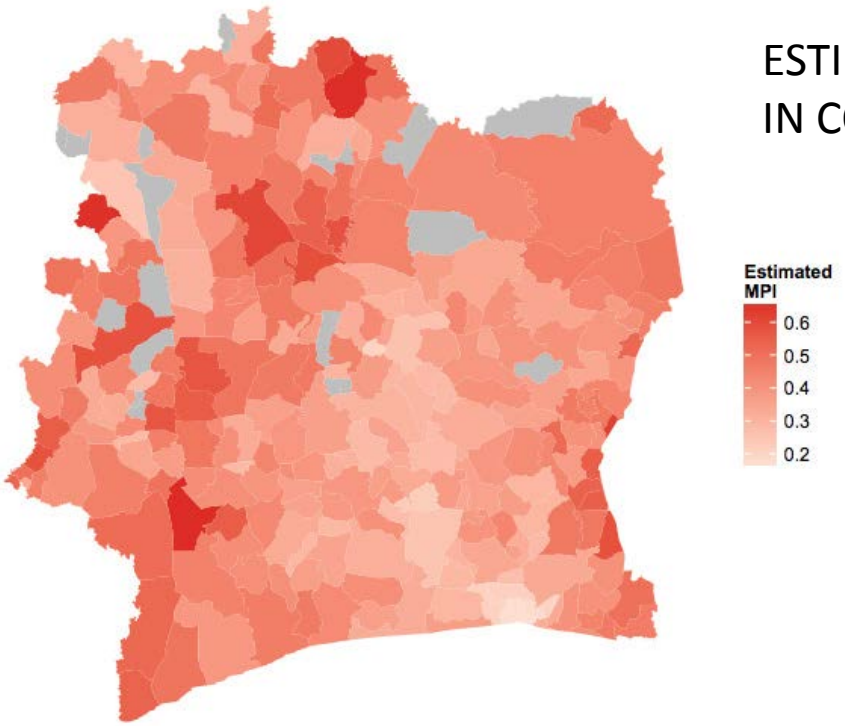
Transport: Exploring Urban Mobility and Public Transport Optimization Using Cellphone Data

Researchers at IBM, using movement data collected from millions of cell-phone users in Ivory Coast in West Africa, have developed a new model for optimizing an urban transportation system.

Results: By comparing existing bus routes to end-to-end journey requirements, **the analysis identified four new bus routes and led to changes in many others. As a result, 22 routes now show increased ridership, and city-wide journey times have decreased by 10%.**

Source: IBM Research (2013)

Proxy Measures



ESTIMATING POVERTY LEVELS IN CÔTE D'IVOIRE

Use anonymized call detail records (CDRs) of 5 million Orange telecommunications customers between December 2011 and April 2012 to assess (1) level of activity among subscribers and (2) locations where calls were made. **Higher levels of mobile communication and wider range of calls are a proxy indicator for prosperity.**

Results: Department poverty levels as approximated by the model used on regional level indicate the finer granularity possible when using CDRs.

Source: Wesolowski A, Eagle N , Tatem A, Smith D, Noor A, Snow R, Buckee C (2012) Quantifying the impact of human mobility on Malaria, Science 338(6104) 267-270.

What is The World Bank doing?

2012- 2013: Exploration via Data Dives

1. Predicting Small-Scale Poverty Measures from Night Illumination
2. Scraping Websites to Collect Consumption and Price Data
3. Measuring Socioeconomic Indicators in Arabic Tweets
4. Test Data Quality of House Hold Surveys using cell phones
5. Analyzing World Bank Data for Signs of Fraud and Corruption

Summary of World Bank work

- Community of Practice (staff in several sectors): Build internal capacity through exchange of results and joint “discovery” of best practices
- Program for Innovations in Big Data Analytics for Development
- Partnerships

2013-2014 Bottom-up Big Data initiatives

- No formal World Bank policy or strategy on Big Data agreed or mandated by senior management or the Board
- Instead, several staff in sector-specific Bank units independently started to explore Big Data to solve specific problems in their sectors
- This contrasts the “solution looking for a problem” paradigm

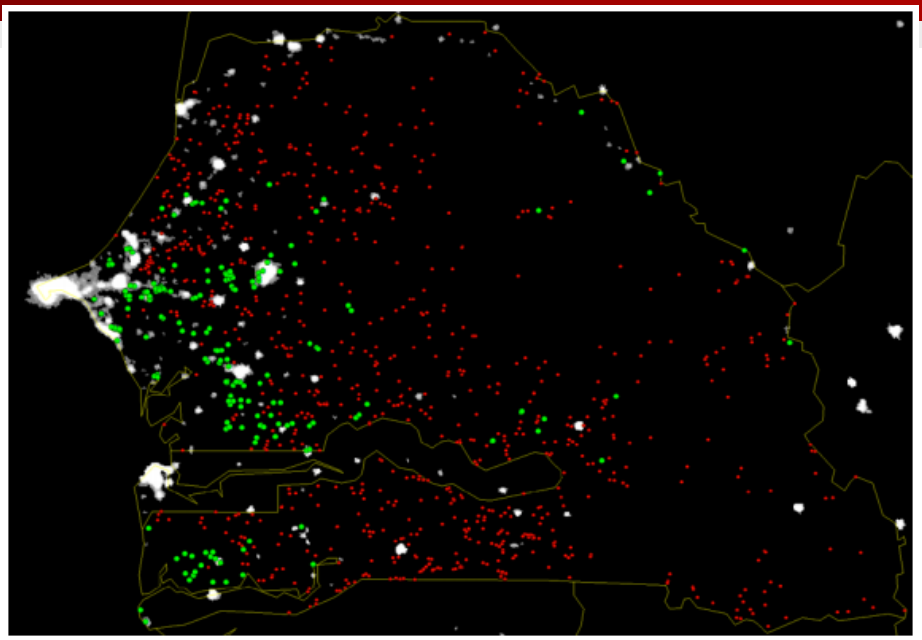
On-going research projects

Title	Country	Sector	Data Source
Understanding relationships between urban infrastructure and crime	Bogota, Columbia	Urban	Crime data, Bus Transit data, urban characteristics
Predicting vulnerability to flooding and enhancing resilience	India	Disaster Risk management	Satellite imagery, census data
Provide Measures for Socio-Economic Indicators	Columbia	Poverty	CDR
Improve Freight transportation flows and environment sustainability of supply chains	Indonesia	Transport	CDR
Monitoring Rural Electrification using satellites	Senegal, Indonesia	Energy	Satellite imagery
Provide transparency and accountability using Open transport data	Sao Paulo, Brazil	Transport	Automatic Vehicle location

On-going research projects

Title	Country	Sector	Data Source
Provide transparency and accountability using Open transport data	Sao Paulo, Brazil	Transport	Automatic Vehicle location
Bus performance dashboard to monitor	Sao Paulo, Brazil	Transport	Automatic Vehicle and fare card data
Improve transportation planning and optimizing public systems	Mexico City and Rio de Janeiro	Transport	Cell phone data
Analyzing data to identify indicators of corruption in Bank-financed projects or public spending	The World Bank	Fraud detection	Bank projects, contracts, financing data
Creating a big data “just in time” analysis platform for disaster risk management	Caribbean	Disaster Risk management	Multiple sources

Monitoring rural electrification using satellite imagery in Senegal and Indonesia



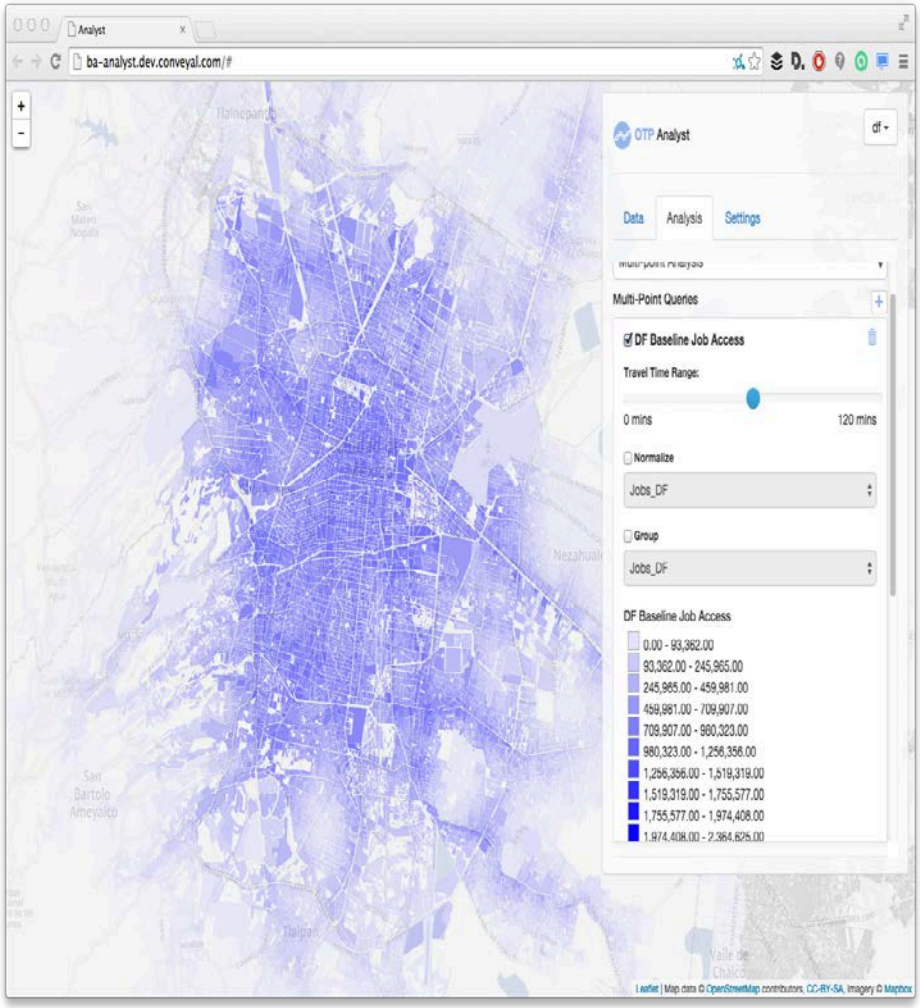
Location of villages in study, Senegal. Surveyed electrified villages in green . Control group of un-electrified villages in red.

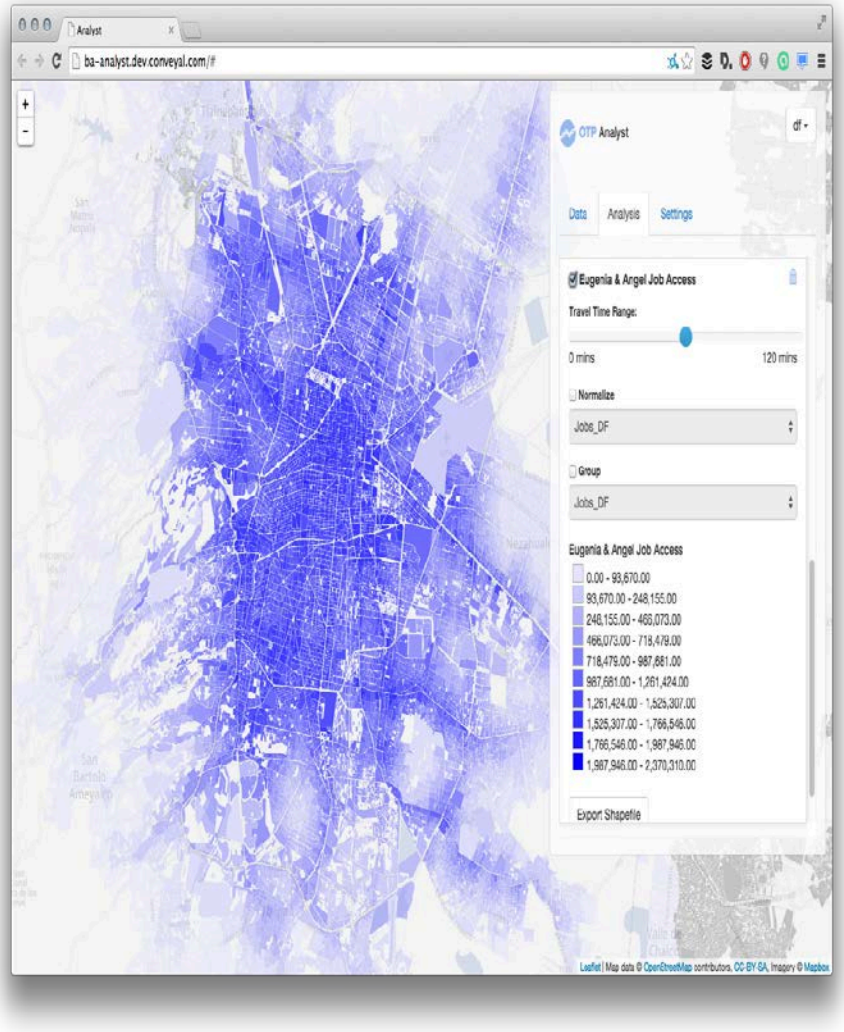
Satellite imagery of the earth at night can detect concentrations of outdoor lighting down to the village level. Satellite-based methods allow **independent tracking of project implementation and impact**, as well as **better selection of new project sites**, while **enhancing transparency and communicating the outcomes of electrification projects.**

Results: This research highlights the potential to use night lights imagery for the planning and monitoring of ongoing efforts to connect the 1.4 billion people who lack electricity around the world.

Source: International Journal of Remote Sensing (FY11 Innovation Challenge Winner)

Effects of travel time and travel cost on job accessibility - Mexico





2014-2015: Innovation challenge

2014-2015: Partnerships

- Workshop with private sector companies (April 2014)
- DataPop alliance
- Sector-specific or project-specific partnerships
  - Forest sector with University of Maryland & Hewlett Packard
  - Freight transportation (Indonesia) with MIT Lab
  - Government agencies depending on sector/country

Big Data: Constraints for developing countries

- Privacy
- Methodological issues
- Capacity
  - computing
  - analytical
  - human
- Data access and availability
- Lack of standards
- Results' replicability

github
SOCIAL CODING

Gentle reminder:

The goal is not about improving data systems, but
IMPROVING PEOPLE'S LIVES