New data sources and statistical methods for humanitarian action and development

Linus Bengtsson MD, Ph.D.
linus.bengtsson@flowminder.org

Hammamet 14 November 2017

FLOWMINDER.ORG

world pop

Flowminder Foundation: Non-profit working with data providers and international/government agencies to operationalize and scale applications in support of vulnerable populations and sustainable development.

WorldPop Program: Research Program led by Prof. Tatem improving the spatial demographic evidence base for low and middle income countries

60 staff focused on data science and integration in spatial demography

---

Academic partners:

University of Southampton

Harvard

Karolinska Institutet

Princeton University

Stockholm School of Economics

Collaborators and donors:

The World Bank — Working for a World Free of Poverty

UNFPA

WFP United Nations World Food Programme

United Nations Foundation

International Organization for Migration (IOM)

OCHA

ADB Asian Development Bank

Clinton Health Access Initiative

DLR

wellcome trust

Microsoft Research

Bill & Melinda Gates foundation

CIESIN Columbia University

All methods are open and published in peer reviewed journals for validation and transparency

70

Spread of yellow fev
Democratic Republic
study

by Moritz U G Kraemer, Nunc
Golding, Birgit Nikolay, Steph
Henrik Salje, Ousmane Faye,
Freya M Shearer, Sarah C Hil
Prof Nuno Taveira, Heinrich H
Nsoesie, Nicholas R Murphy,
John S Brownstein, Prof And
David L Smith, Amadou A Sa
and Simon Cauchemez

The Lancet Infectious Disea

69

Plasmodium falcipar
to China and its mor

by Shengjie Lai, Nicola A. Wa
Junling Sun, Tomas Bird, Am
Canjun Zheng, Zhongjie Li, A

Scientific Reports, Decembe

76

Housing Improvements and Malaria Risk in Sub-Saharan Africa: A Multi-Country Analysis of Survey

by Lucy S. Tusting, Christian Bottomley, Harry Gibson, Immo Kleinschmidt, Andrew J. Tatem, Steve W. Lindsay, and Peter W Gething

PloS Medicine, Feb. 2017

75

Mapping poverty using mobile phone and s

by Jessica E. Steele, Pål Roe Sundsøy, Carla Pezzulo, Victor A Alegana, Tomas J. Bird, Joshua Blumenstock, Johannes Bjella Kenth Engø-Monsen, Yves-Alexandre de Montjoye, Asif M. Iqb Khandakar N. Hadiuzzaman, Xin Lu, Erik Wetter, Andrew J. Ta and Linus Bengtsson

Journal of the Royal Society Interface, February 2017 Volume

74

Changing Epidemiology of Human Brucello 1955–2014

by Shengjie Lai, Hang Zhou, Weiyi Xiong, Marius Gilbert, Zhuo Huang, Jianxing Yu, Wenwu Yin, Liping Wang, Qiulan Chen, Yu Mu, Lingjia Zeng, Xiang Ren, Mengjie Geng, Zike Zhang, Buyun Tiefeng Li, Dali Wang, Zhongjie Li, Nicola A. Wardrop, Andrew Tatem, and Hongjie Yu

Emerging Infectious Diseases, Volume 23, Number 2 - Feb 20

80

Malaria in China, 2011–2015: an observational study

by Shengjie Lai, Zhongjie Li, Nicola A Wardrop, Junling Sun, Michael G Head, Zhuojie Huang, Sheng Zhou, Jianxing Yu, Zike Zhang, Shui-Sen Zhou, Zhigui Xia, Rubo Wang, Bin Zheng, Yao Ruan, Li Zhang, Xiao-Nong Zhou, Andrew J Tatem, and Hongjie Yu

Bulletin of the World Health Organization, May 2017

79

Treatment-seeking behaviour in low- and middle-income countries estimated using a Bayesian model

by Victor A. Alegana, Jim Wright, Carla Pezzulo, Andrew J. Tatem, and Peter M. Atkinson

BMC Medical Research Methodology, Apr. 2017

78

Virus genomes reveal factors that spread and sustained the Ebola epidemic

by Gytis Dudas, Luiz Max Carvalho, Trevor Bedford, Andrew J. Tatem, Guy Baele, Nuno R. Faria, Daniel J. Park, Jason T. Ladner, Armando Arias, Danny Asogun, Filip Bielejec, Sarah L. Caddy, Matthew Cotten, Jonathan D'Ambrozio, Simon Dellicour, Antonino Di Caro, Joseph W. Diclaro, Sophie Duraffour, Michael J. Elmore, Lawrence S. Fakoli, Ousmane Faye, Merle L. Gilbert, Sahr M. Gevao, Stephen Gire, Adrianne Gladden-Young, and et al.

Nature, Apr. 2017

Pioneered Anonymised Mobile Network Data for Infectious Disease (2008 Zanzibar, 2012 Haiti, 2013 Namibia, Indonesia) and Crisis Response (Haiti Earthquake and Cholera, Nepal Earthquake)

CLINTON HEALTH ACCESS INITIATIVE

MIT Technology Review

NEWS & ANALYSIS | FEATURES | VIEWS | MULTIMEDIA | DISCUSSIONS | TOPICS

COMMUNICATIONS NEWS

How Cell-Phone Data Could Slow the Spread of Malaria

Location data suggests a better way to fight a disease that kills a million people a year.

REPORT

Quantifying the Impact of Human Mobility on Malaria

Amy Wesolowski[1,2], Nathan Eagle[3,4], Andrew J. Tatem[5,6,7], David L. Smith[6,8], Abdisalan M. Noor[9,10], Robert W. Snow[9,10], Caroline O. Buckee[4,11,*]

Science

AAAS

Average daily numbers of sims that moved out from the communal sections surrounding Saint-Marc, Oct 15 to Oct 23, 9:00 am, 2010.

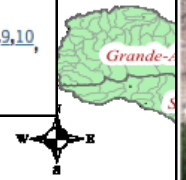
Outbreak area

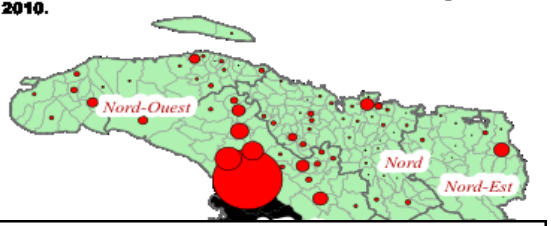
10

100

250

500





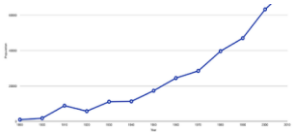
GLOBAL UPDATE

Haiti: Cellphone Tracking Helps Groups Set Up More Effective Aid Distribution, Study Says

world pop FLOWMINDER.ORG

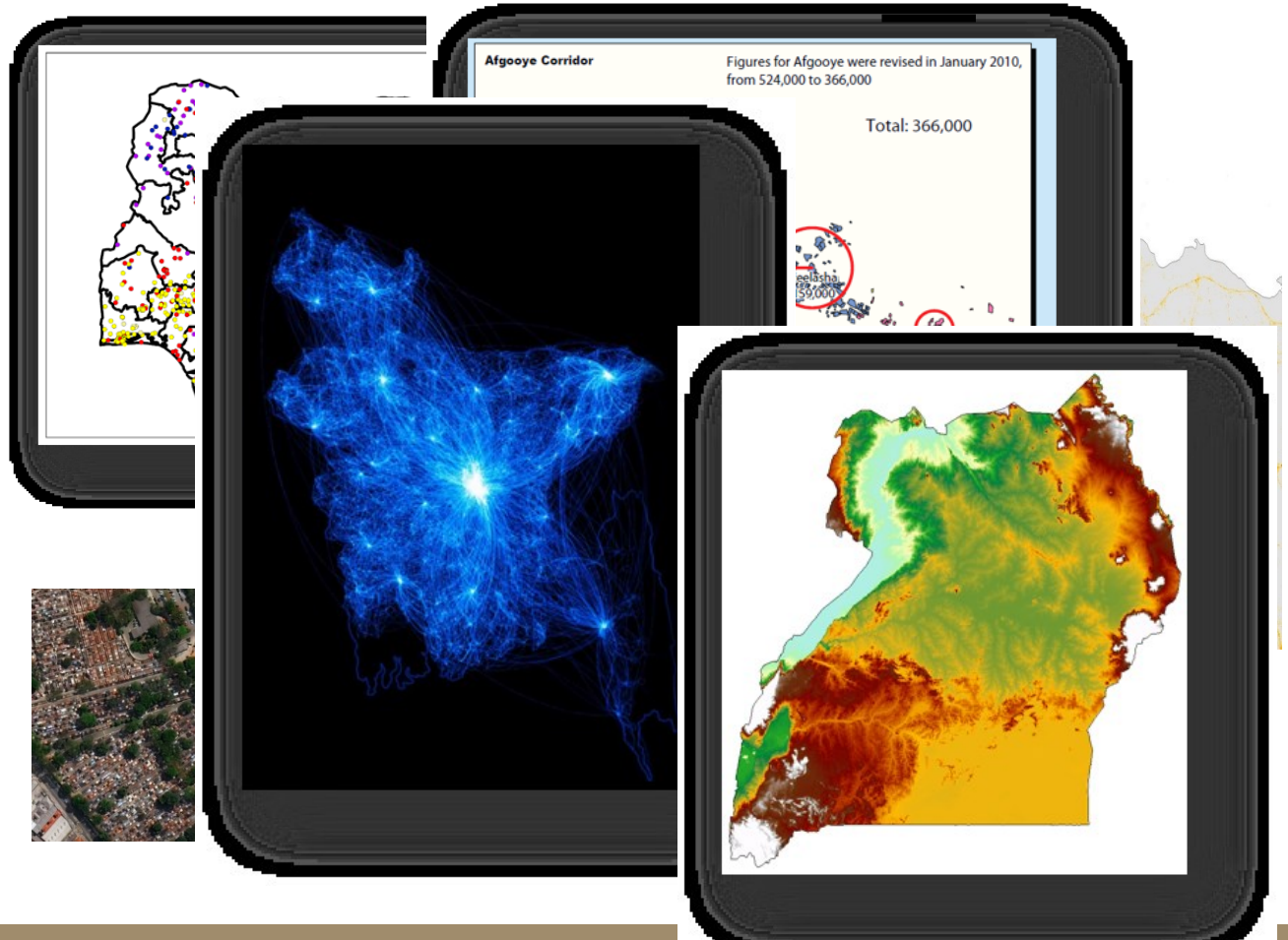
Population data: Questions and applications

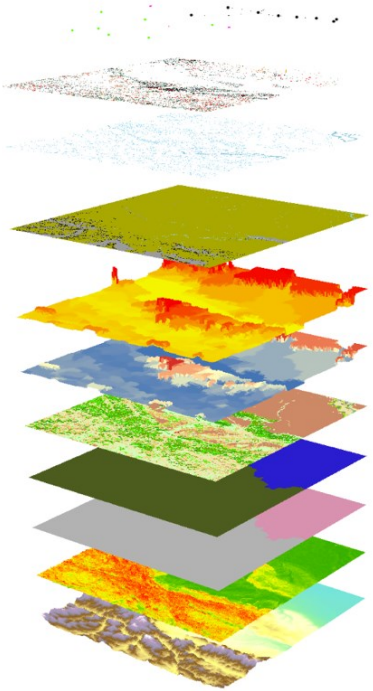
- How many people in an area?
  - What is my health facility catchment size?
  - How many people are affected by conflict?
- How is the population structured?
  - How many children under 5 yrs in this area to vaccinate?
  - How many people are eligible to vote in this village?
- How does the population change?
  - How are healthcare demands changing in my area?
  - Where are extensions/upgrades to sanitation and electrification needed?




Mapping Populations:
Distributions

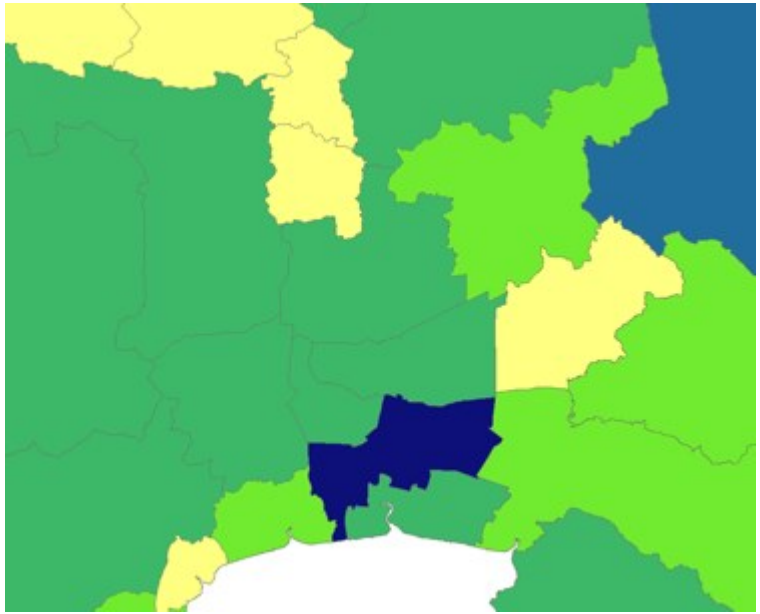
200+ geospatial layers

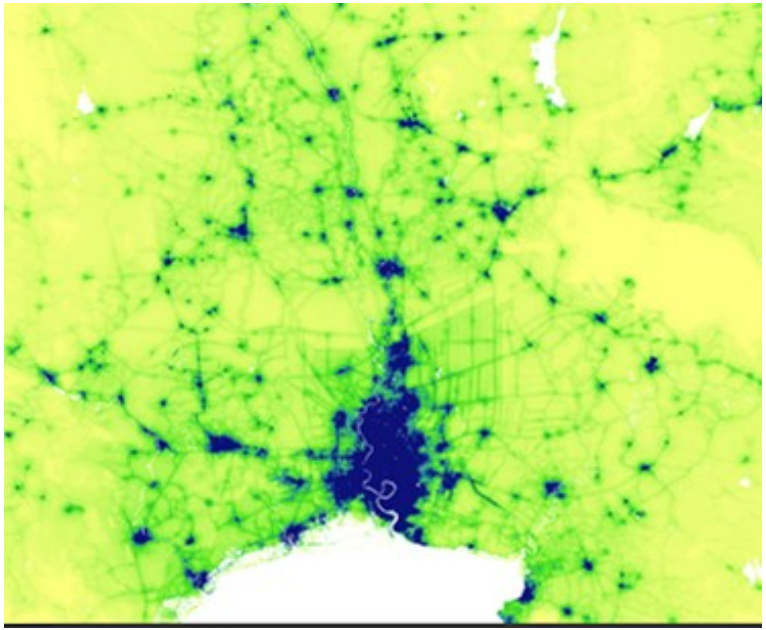




world pop
FLOWMINDER.ORG

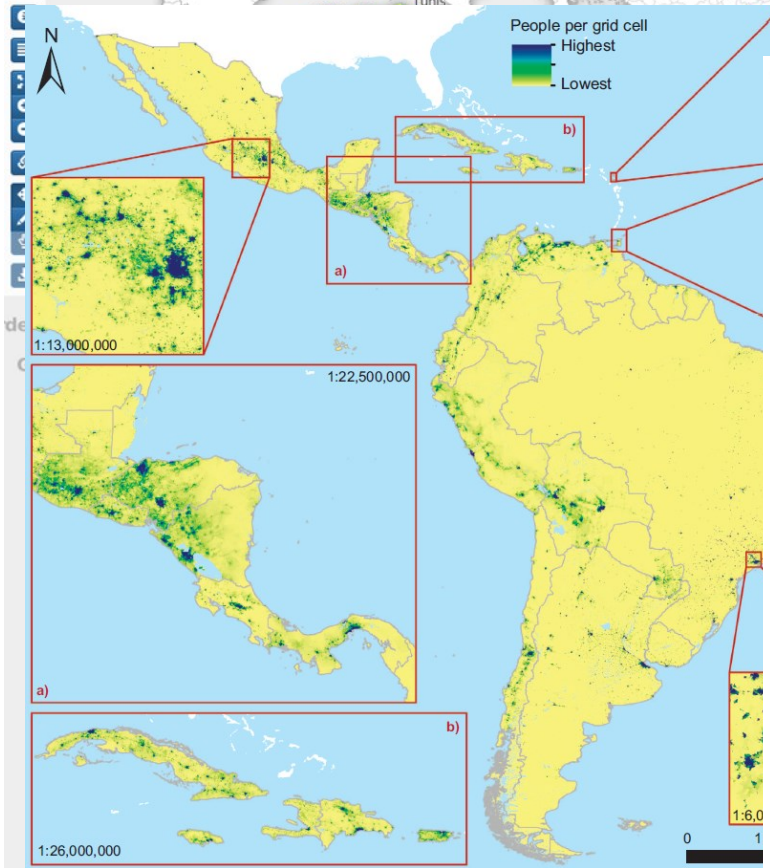
Disaggregation of public census data using Random Forrest approach: 100x100m grid globally

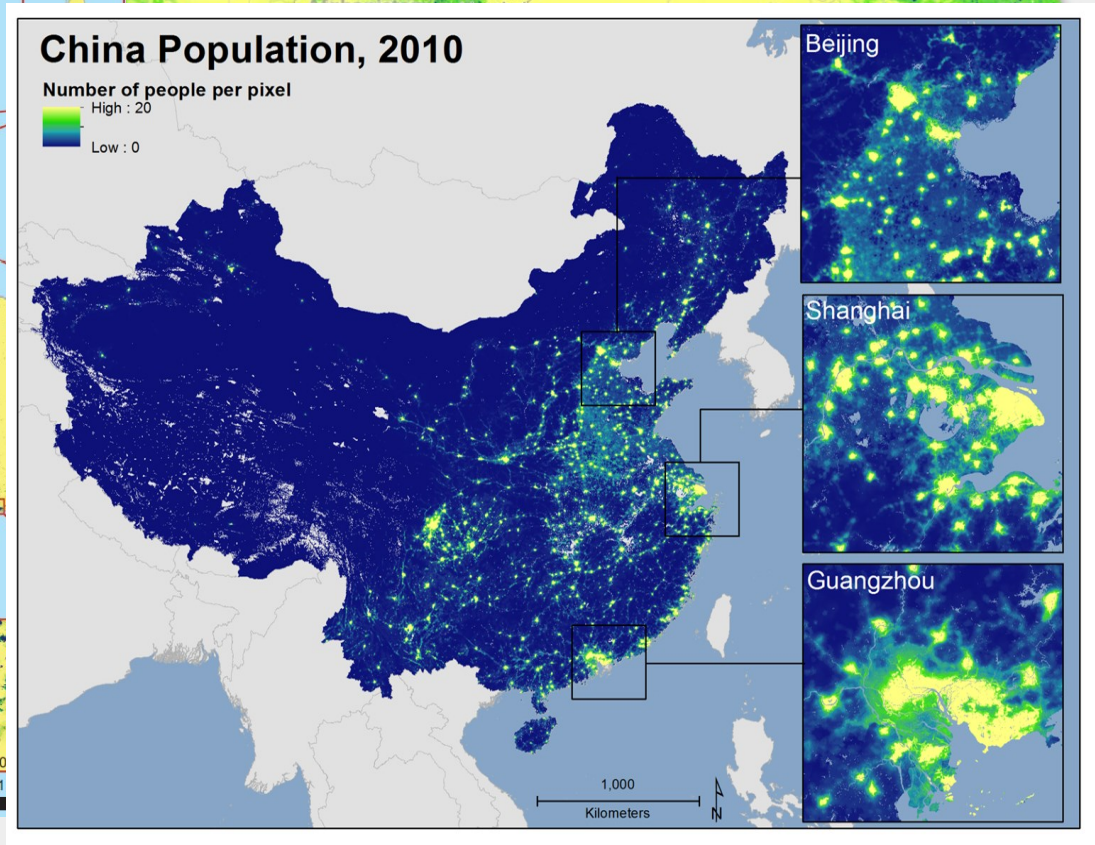




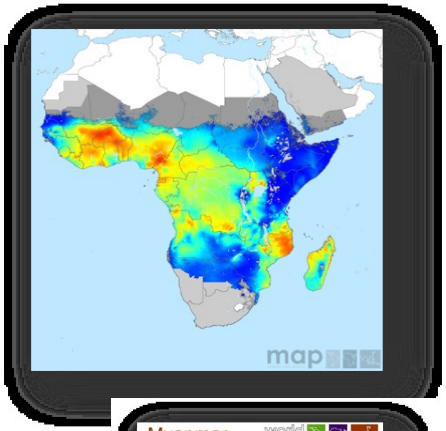
BILL & MELINDA GATES foundation

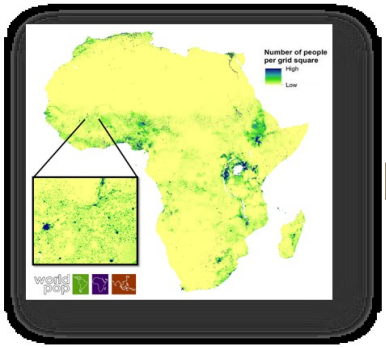
world pop FLOWMINDER.ORG

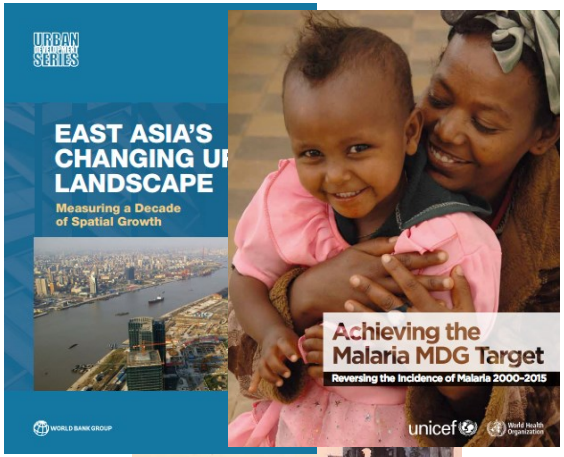


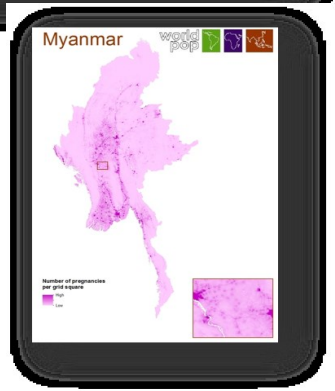


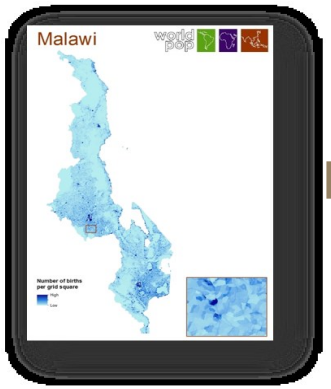
Large-scale use of WorldPop data

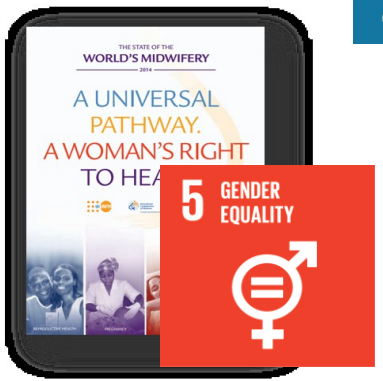








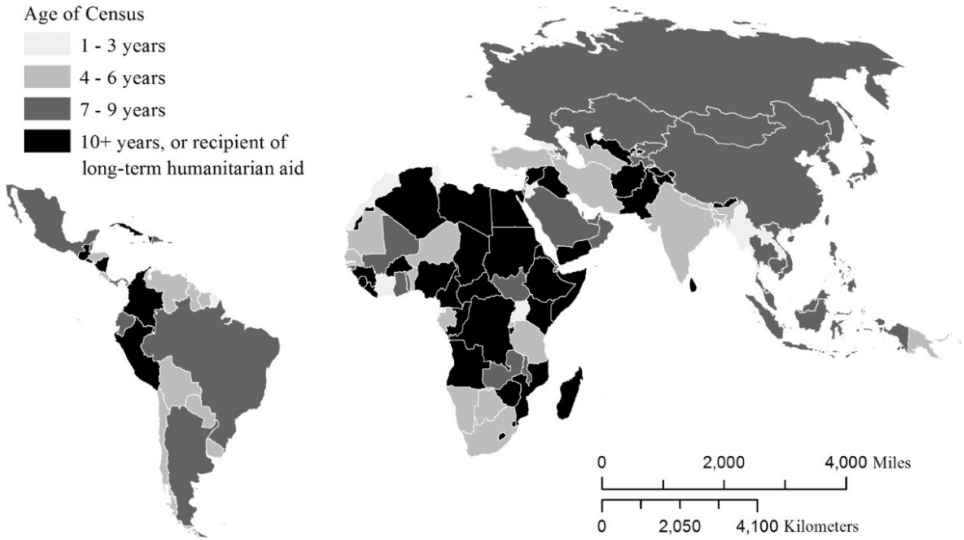




world pop
FLOWMINDER.ORG

But census data have problems

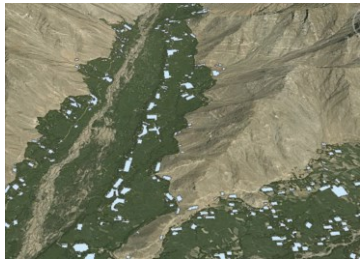
- Collected once every ten years
- Released with delays
- Sometimes much older
- Sometimes manipulated
- Reality changes: migration, displacements, birth, deaths

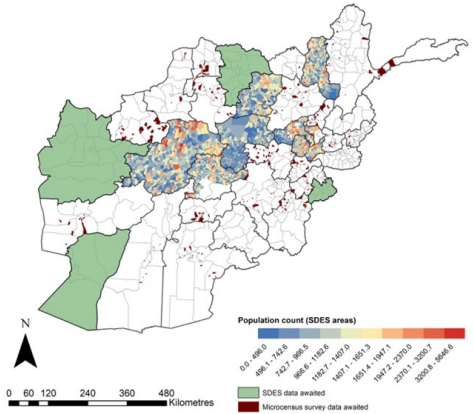


Mapping Afghanistan

Problem

- Last census conducted in 1979
- Exponential projections since then





Solutions

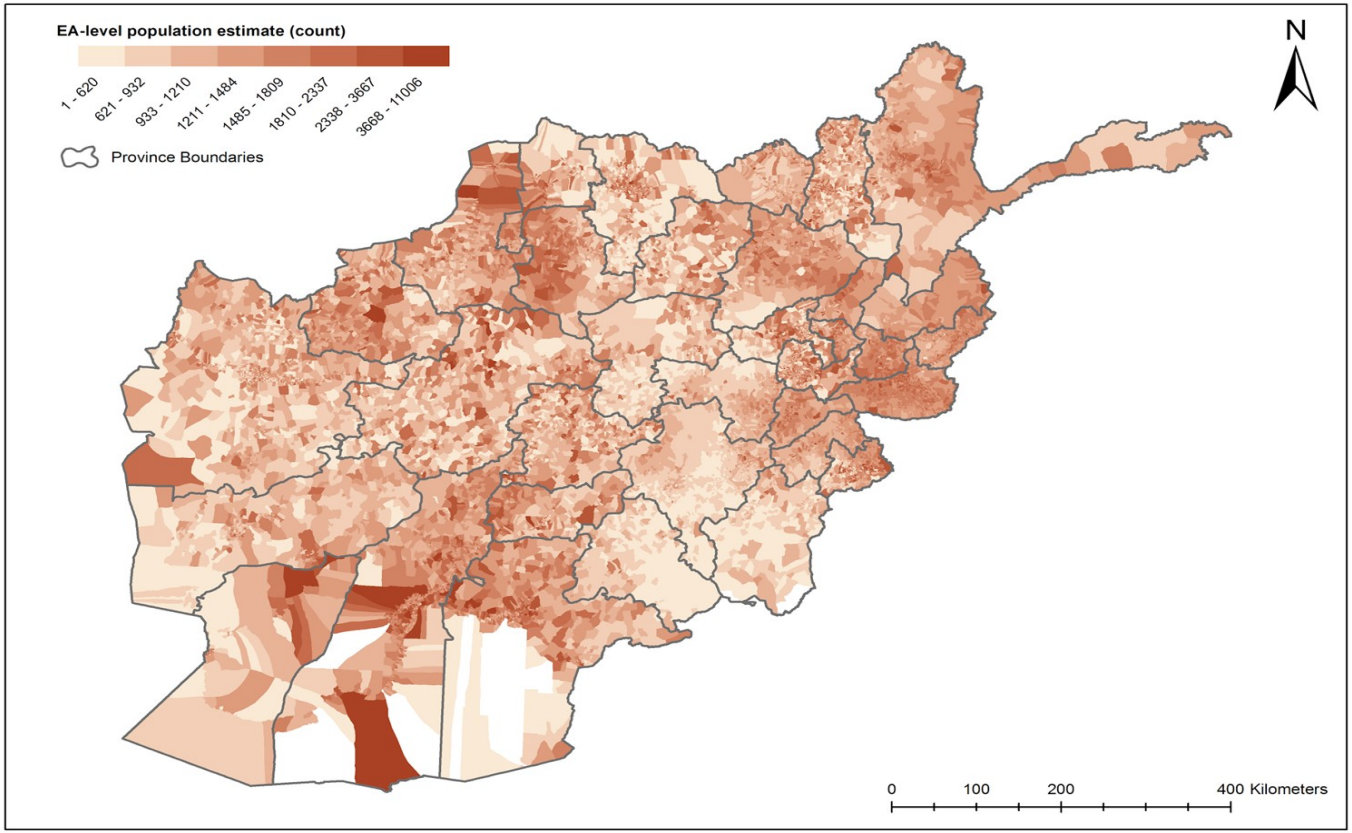
- 99% cost saving compared with a census
- Can be used between censuses
- Close collaboration needed with the government: Critical for acceptance of estimates and understanding local conditions which can influence estimates (e.g. nomads)

UNFPA

BILL & MELINDA GATES foundation

world pop FLOWMINDER.ORG

Predicting population distribution

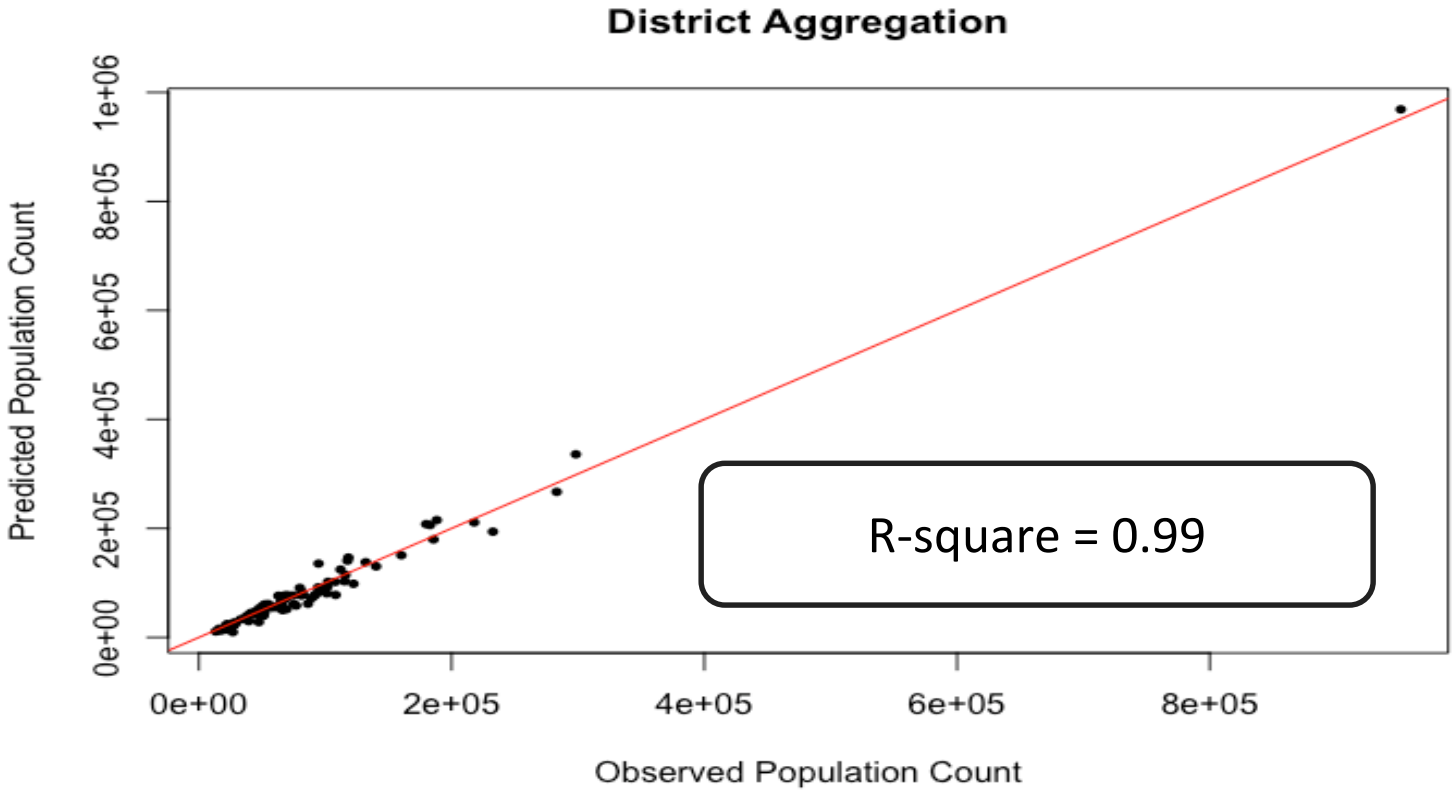


UNFPA

BILL & MELINDA GATES foundation

world pop
FLOWMINDER.ORG

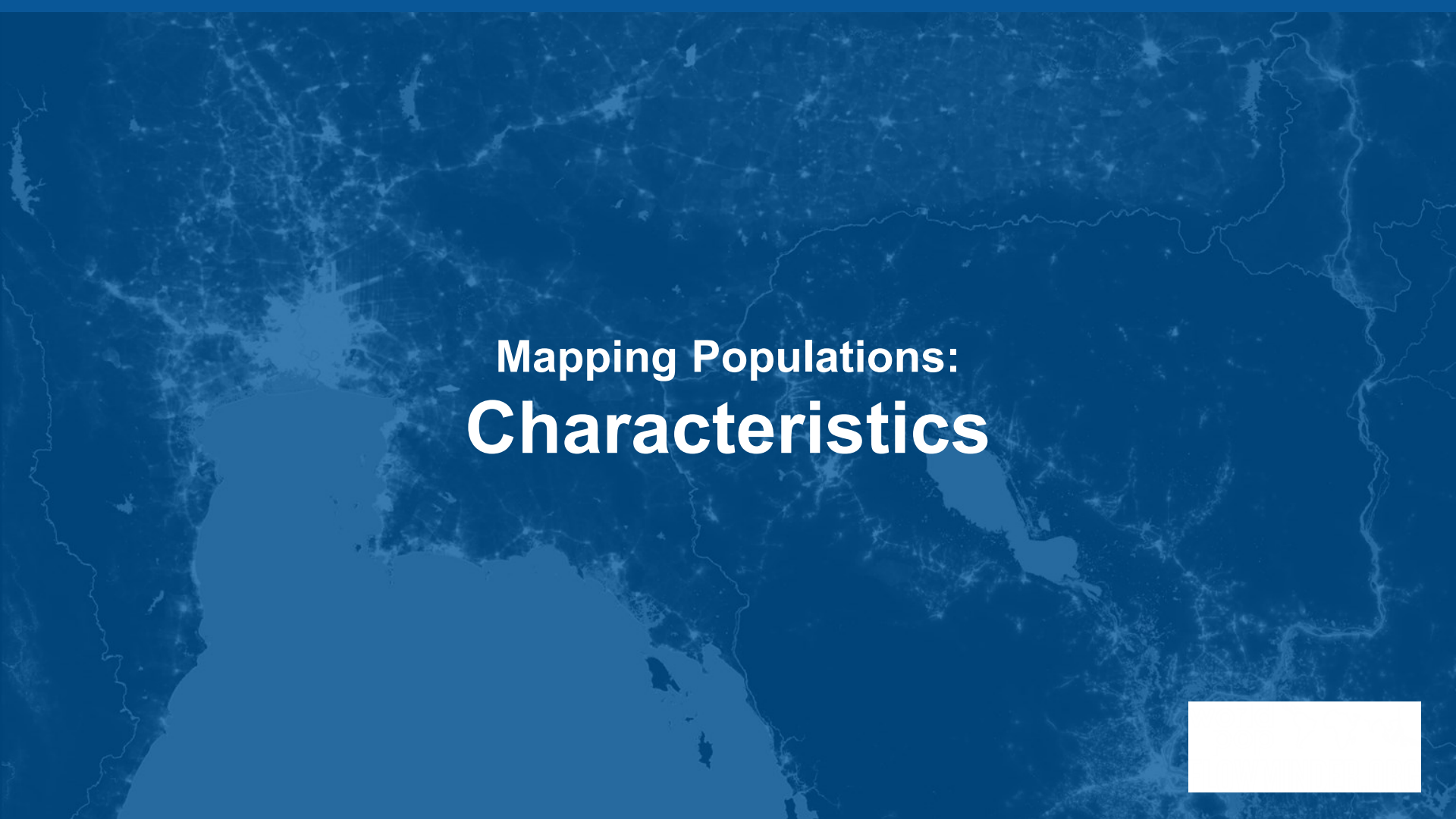
Accuracy assessment



world pop
FLOWMINDER.ORG

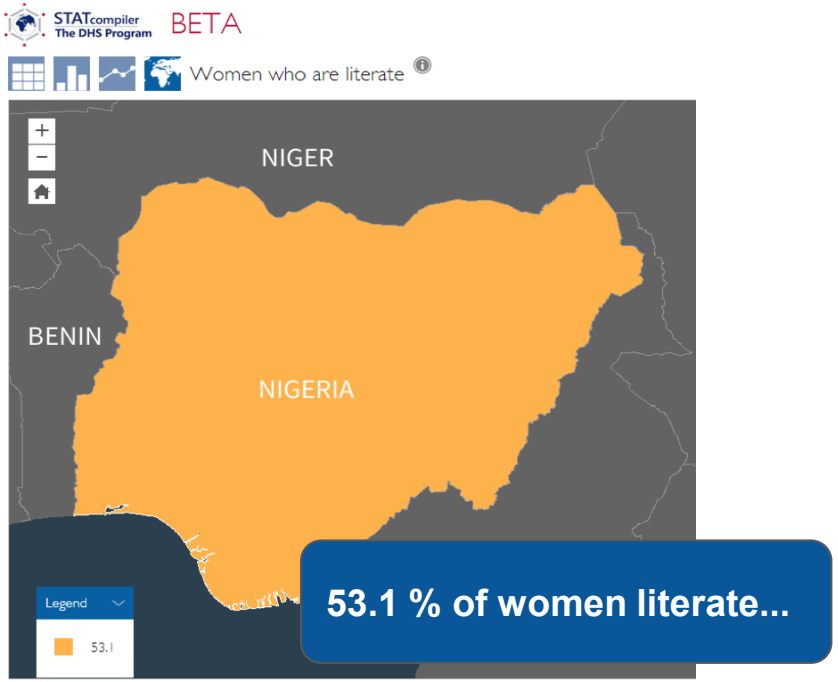
Commissioned by the President: Country-led process

world pop FLOWMINDER.ORG

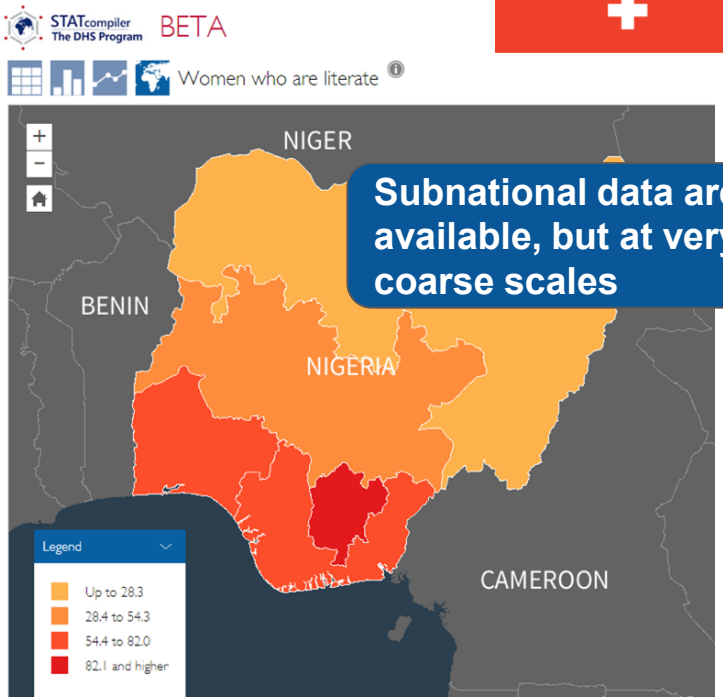

Mapping Populations:
Characteristics

Do all women have the same access to opportunities?

5 GENDER EQUALITY

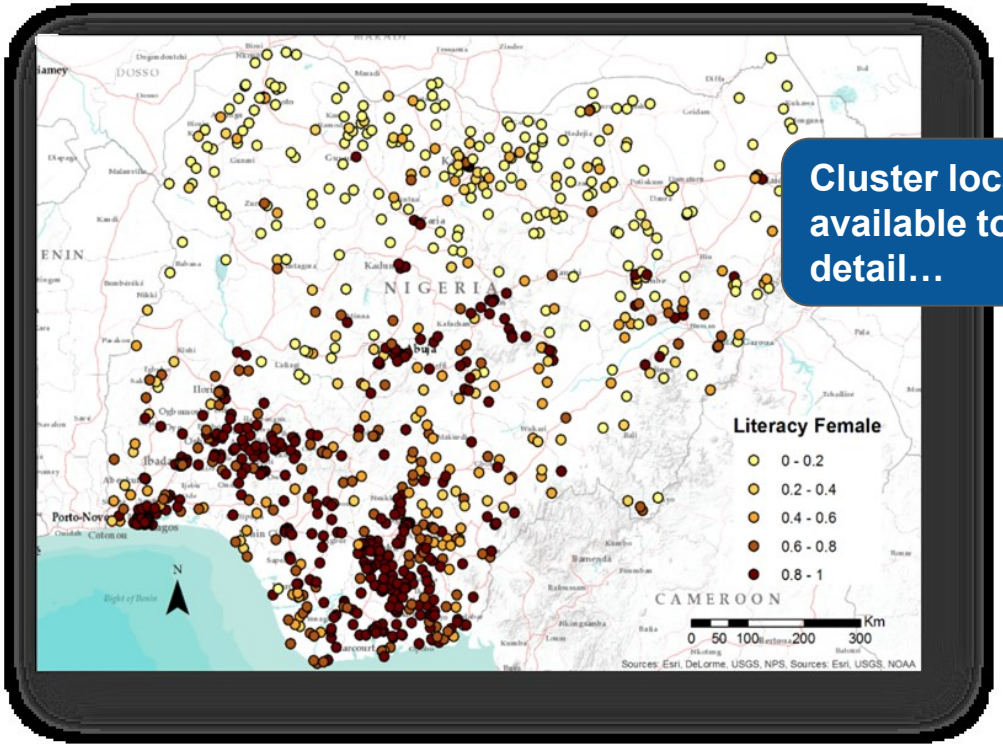


53.1 % of women literate...



Subnational data are available, but at very coarse scales

GPS locations of survey clusters available



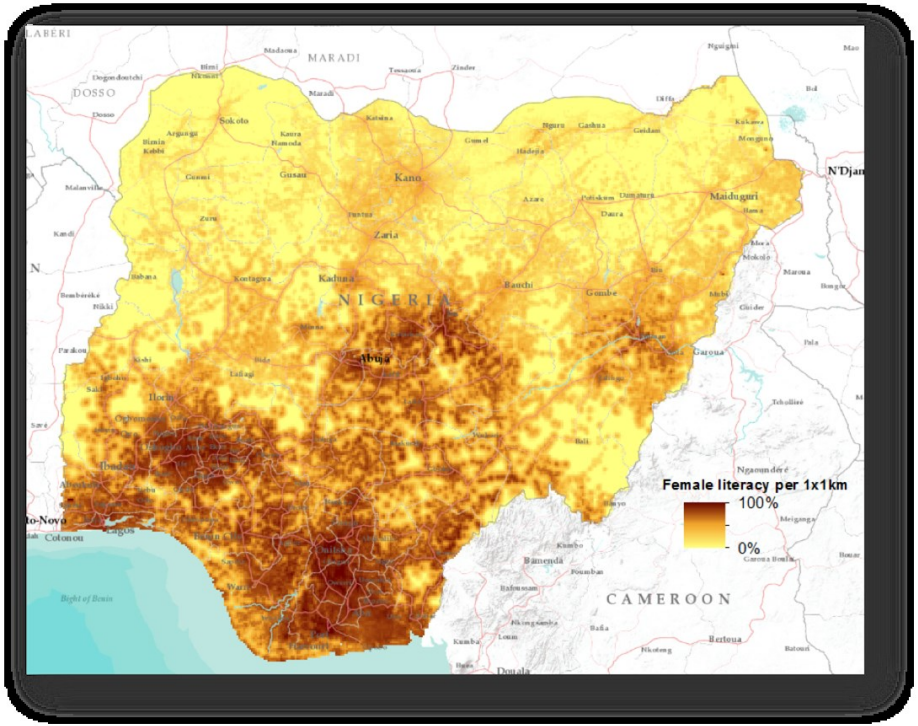
Cluster location data are available to provide relevant detail…

….but no measurements in the unsampled locations – what can we do?

Bosco et. al. Royal Soc. Int. 2017


Satellite imagery can be processed to map factors known to correlated with human welfare

High resolution map of female literacy



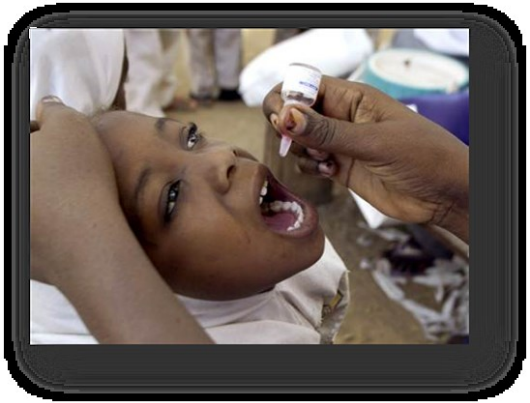
Nigeria: Proportion of females who are literate

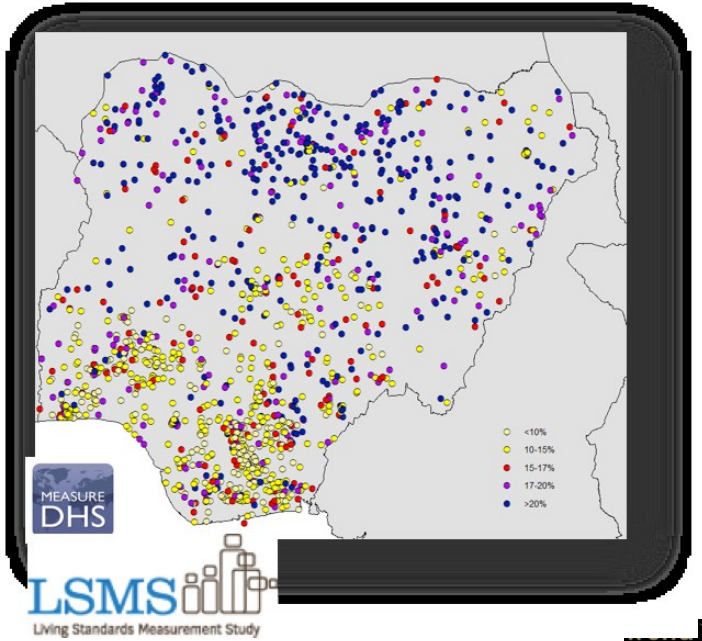
Bosco et. al. Royal Soc. Int. 2017

How to Vaccinate Close to All Children Under 5?

Problem

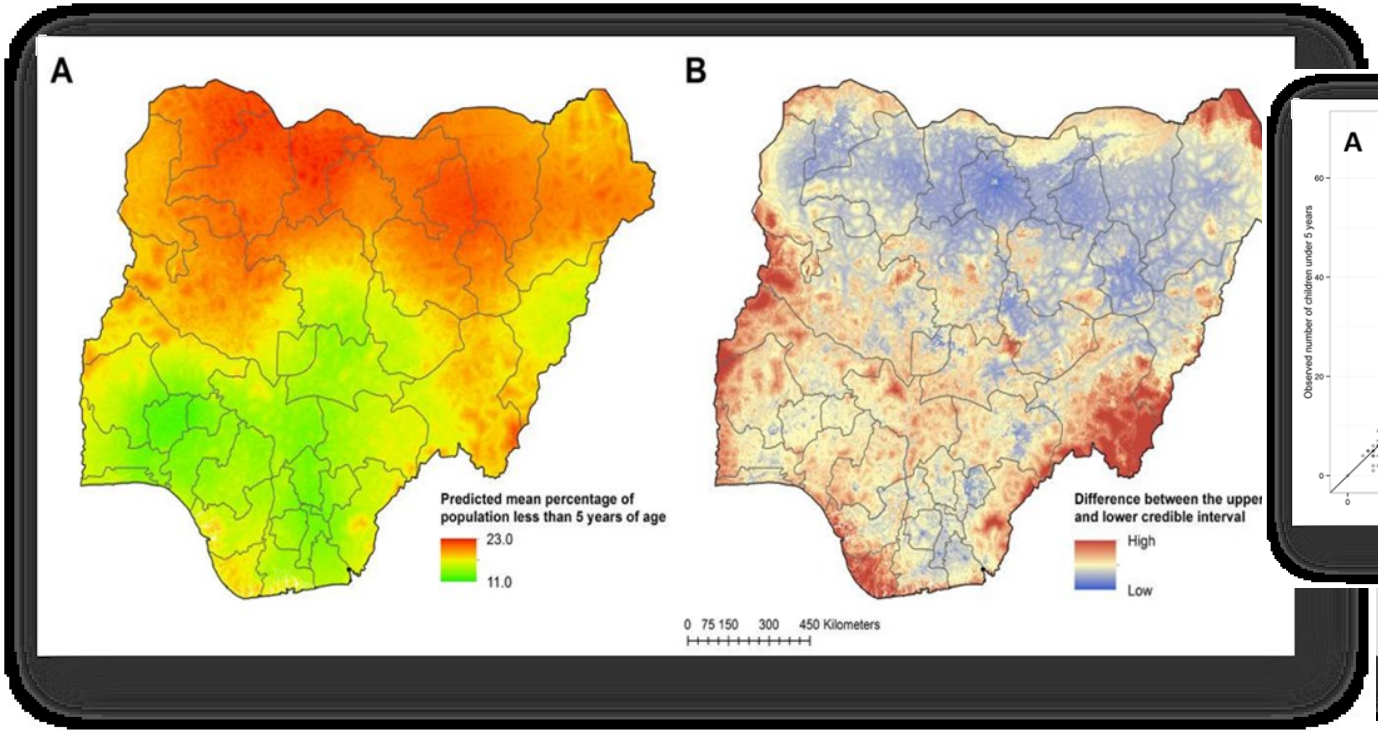
- Vaccine availability
- Local vaccination needs, routes and logistics
- Outdated census

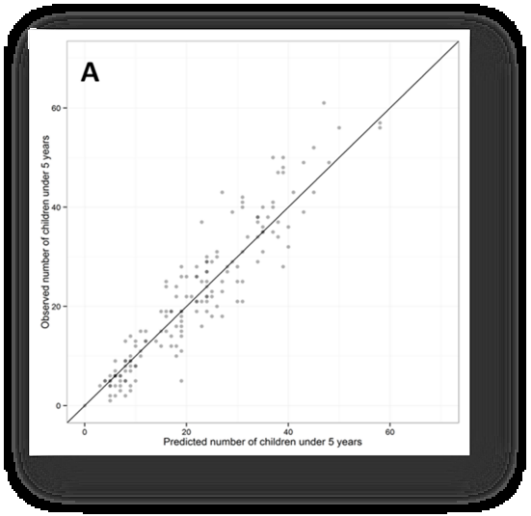


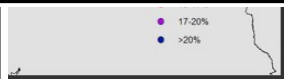


pop
FLOWMINDER.ORG

How to Vaccinate Close to all Children under 5?

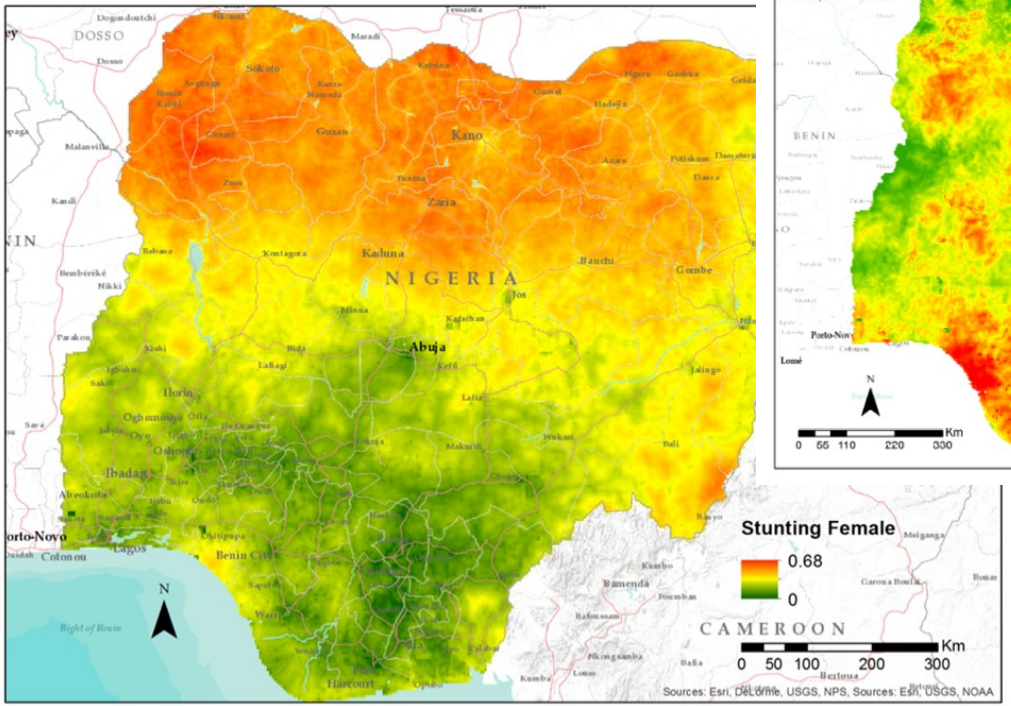


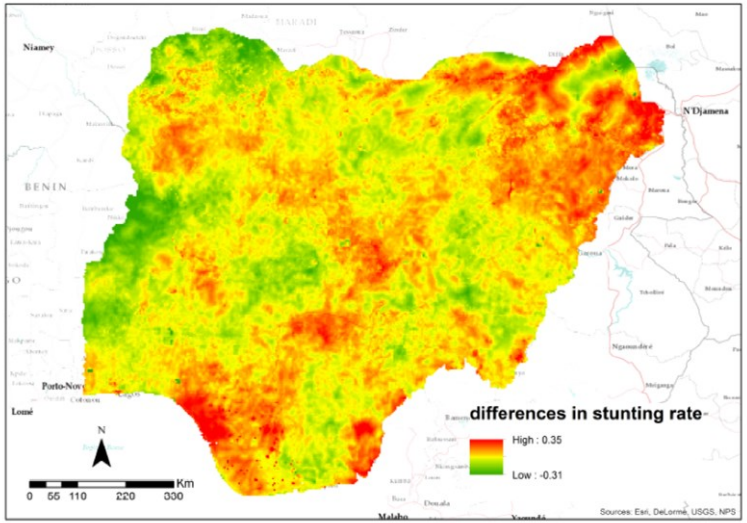




Alegana et. al. Royal Soc. Int. 2015

Stunting Among Girls in Nigeria





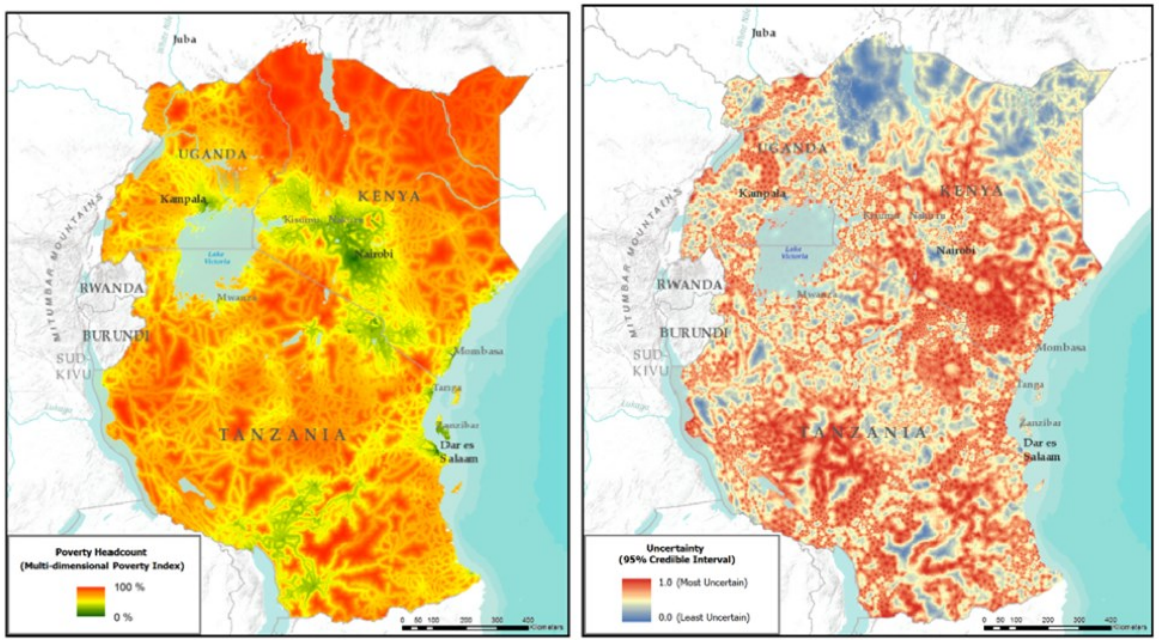
Bosco et. al. Forthcoming Royal Soc. Int.

5 GENDER EQUALITY

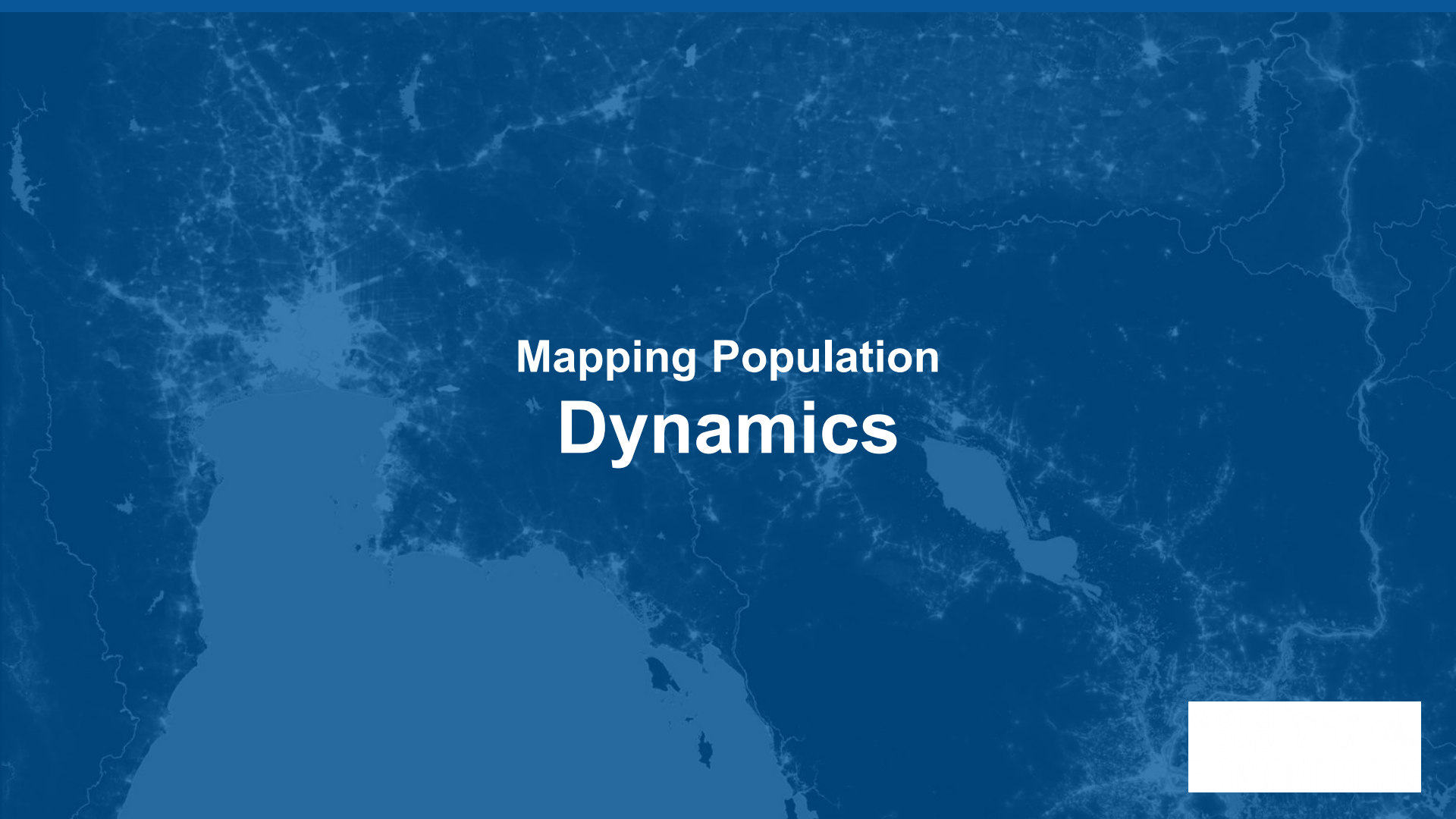
High-Resolution Poverty Maps

1 NO POVERTY

Multi-Dimensional Poverty Index (MPI)

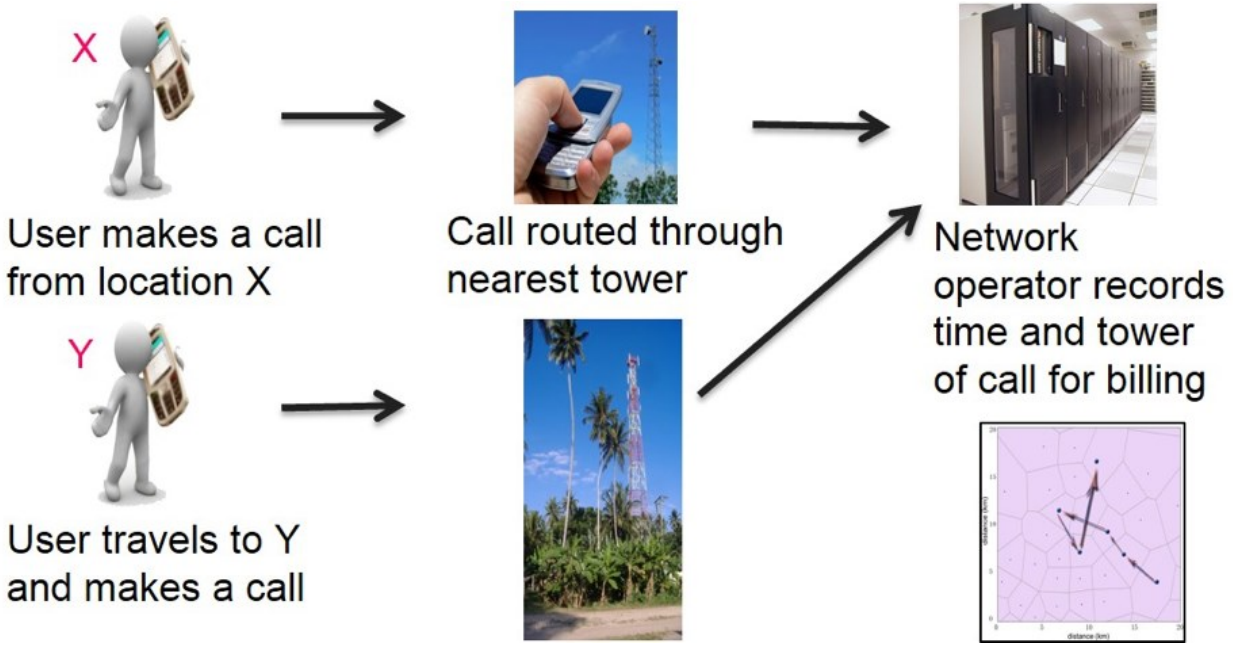


world pop FLOWMINDER.ORG


Mapping Population
Dynamics

Mapping Denominators: CDR Enhanced Datasets

Aggregated call detail records (CDRs) + survey data + satellite/GIS data to model characteristics and mobility



Mapping Denominators: CDR Enhanced Datasets

	A_NU...	B_NU...	CELL_ID	LOCATION	TIER3_NAME	MAIN_GROUP	CALLI...	CALLI...	EVENT_...					DUCT
423	20923...	23214...	(null)	Norge	SMS i Norge	SMS	(null)	242013...	15.02.201					
424	20923...	23214...	(null)	Norge	SMS i Norge	SMS	(null)	242013...	15.02.201					
425	20923...	23214...	(null)	Norge	SMS i Norge	SMS	(null)	242013...	15.02.201					
426	24115...	46230...	242012040141314	Utland	NR - Utgående tale	Tale	35526...	242050...	15.02.201					
427	46240...	3...	2...		fastnett	Tale	12164...	242010...	15.02.201					
428	46240...	4...	2...		operatører hos Netcom	Tale	12164...	242010...	15.02.201					
429	46240...	3...	2...		fastnett	Tale	12164...	242010...	15.02.201					
430	46240...	24002...	242012070143948	Norge	Til operatører hos Netcom	Tale	12164...	242010...	15.02.201					
431	46240...	24002...	242012070143948	Norge	Til operatører hos Netcom	Tale	12164...	242010...	15.02.201					
432	46240...	11168...	242012070143948	Norge	Til fastnett	Tale	12164...	242010...	15.02.201					
433	48311...	45754...	2...	Norge	Viderekobling i Telenors Nett	Tale	(null)	242010...	15.02.201					
434	48311...	(null)	2...			Internett	(null)	242010...	14.02.201					
435	48311...	(null)	2...			Internett	(null)	242010...	14.02.201					
436	48311...	(null)	2...			Internett	(null)	242010...	14.02.201					
437	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.201					
438	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.201					
439	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	14.02.201					
440	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.201					
441	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.2011	144	(null)	(null)	T3-VOT	04577
442	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	14.02.2011	3711	(null)	(null)	T3-VOT	04577
443	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.2011	3202	(null)	(null)	T3-VOT	04577
444	48311...	(null)	(null)	Norge	Internett	Internett	(null)	242010...	15.02.2011	152	(null)	(null)	T3-VOT	04577
445	45047...	49639...	(null)	Norge	SMS i Norge	SMS	(null)	242010...	15.02.2011	(null)	(null)	(null)	(null)	04747
446	22637...	(null)	(null)	Utland	GPRS i utlandet	Internett	(null)	238209...	15.02.2011	341	(null)	(null)	T3-VOT	(null)
447	46835...	23959...	242012080150195	Norge	NR - Utgående SMS	SMS	35821...	242050...	14.02.2011	(null)	(null)	(null)	-MO	(null)
448	23170...	(null)	2420115... 0118026	Utland	NR - Inngående tale/data/fax	Tale	35492...	242059...	14.02.2011	(null)	(null)	(null)	-MTC	(null)
449	23698...	(null)	2420115... 0112573	Utland	NR - Inngående tale/data/fax	Tale	35489...	242050...	15.02.2011	(null)	(null)	(null)	-MTC	(null)
450	29651...	(null)	2...		...kedata	Internett	(null)	242059...	14.02.2011	28754	(null)	(null)	T3-VOT	(null)
451	48376...	23959...	2...		...ngående SMS	SMS	35536...	242059...	15.02.2011	(null)	(null)	(null)	T3-TA2...	(null)

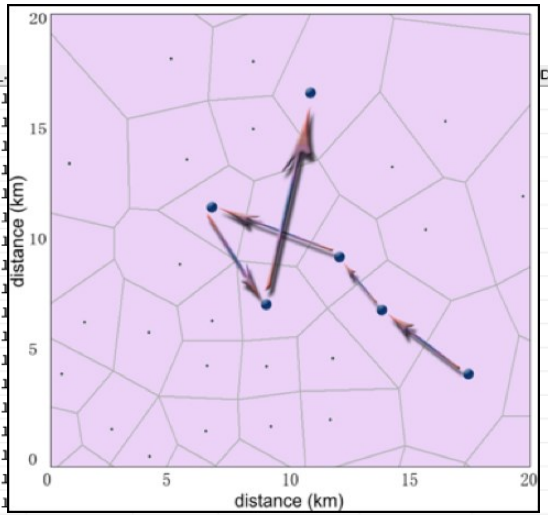
(A) Number Caller - Anonymised

(B) Number Receiving Party - Anonymised

Type Call: SMS, Call, Data, etc

Data Volume

Cell ID Location

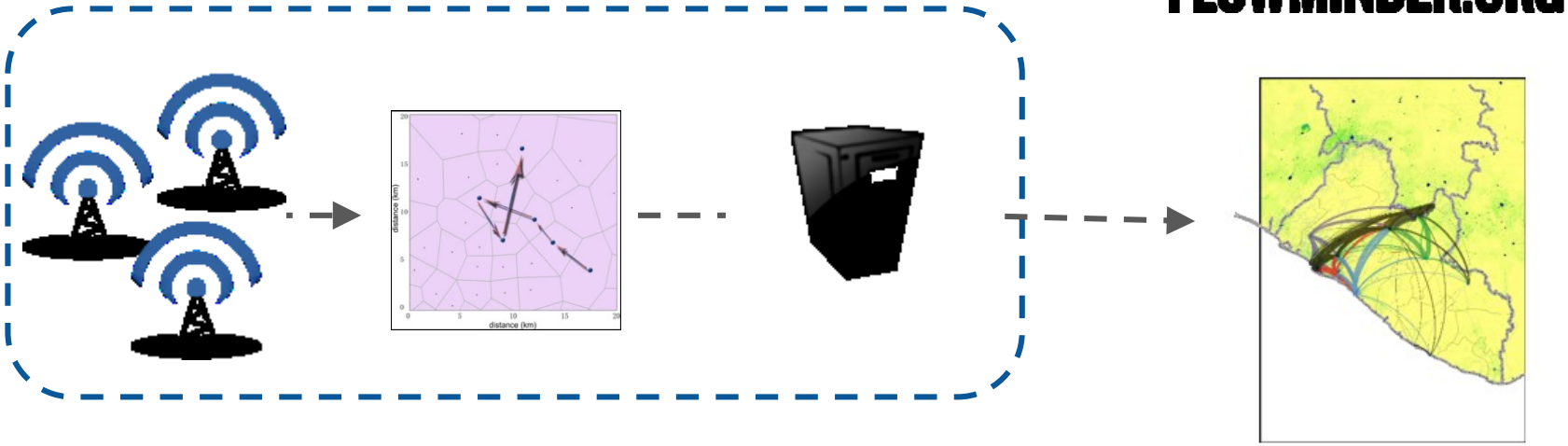


world pop FLOWMINDER.ORG

Preserving User's Privacy

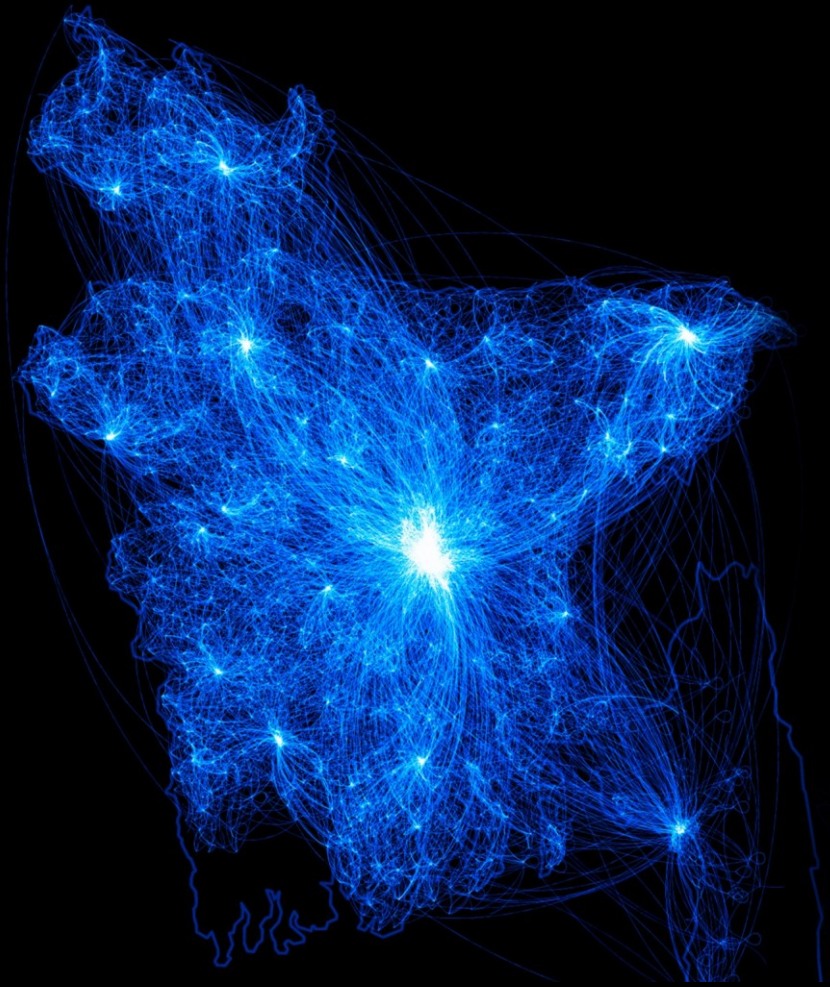
Compliance with GSMA data integrity guidelines: Data never leaves mobile operator's system to avoid any privacy, commercial concerns.

FLOWMINDER.ORG



Mobile operator firewall

world pop FLOWMINDER.ORG

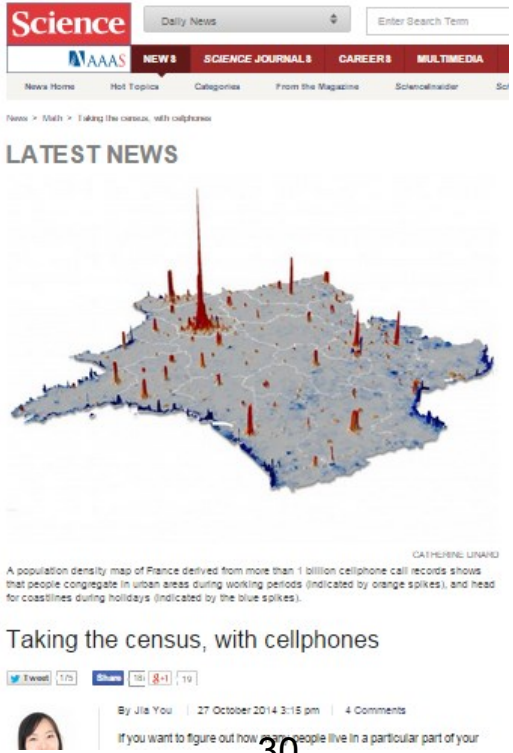


Dynamic population mapping using mobile phone data

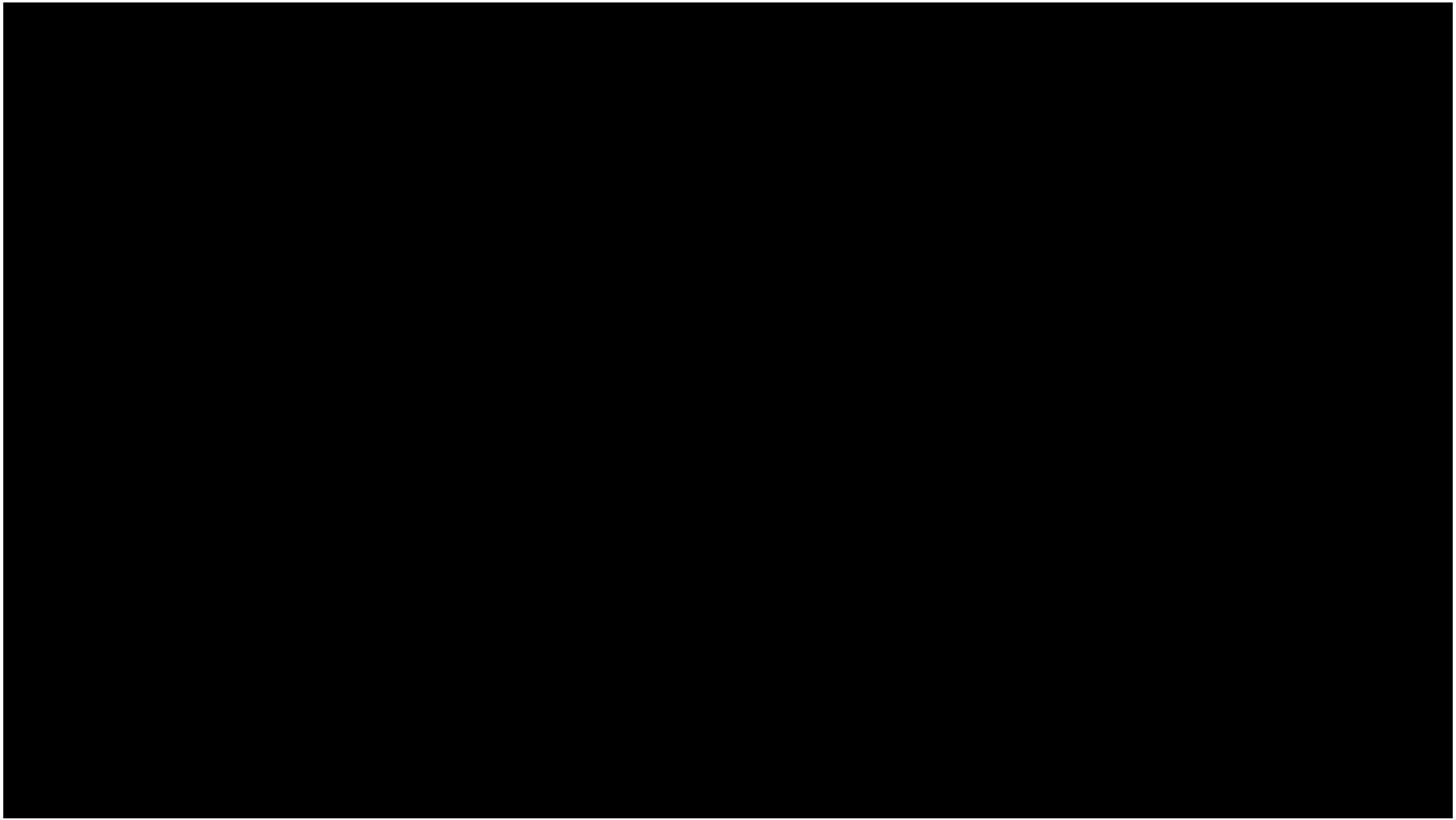
Pierre Deville[a,b,c,1], Catherine Linard[c,d,1,2], Samuel Martin[e], Marius Gilbert[c,d], Forrest R. Stevens[f], Andrea E. Gaughan[f], Vincent D. Blondel[a], and Andrew J. Tatem[g,h,i]

[a]Department of Applied Mathematics, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium; [b]Center for Complex Network Research and Physics Department, Northeastern University, Boston, MA 02115; [c]Fonds National de la Recherche Scientifique, B-1000 Brussels, Belgium; [d]Biological Control and Spatial Ecology, Université Libre de Bruxelles, B-1050 Brussels, Belgium; [e]Université de Lorraine CNRS, Centre de Recherche en Automatique de Nancy, UMR 7039, 54518 Vandoeuvre-lès-Nancy, France [f]Department of Geography and Geosciences, University of Louisville, Louisville, KY 40292; [g]Department of Geography and Environment, University of Southampton, Southampton SO17 1BJ, United Kingdom; [h]Fogarty Internati… MD 20892; and [i]Flowminder Foundation, 17177 Stockholm, Sweden

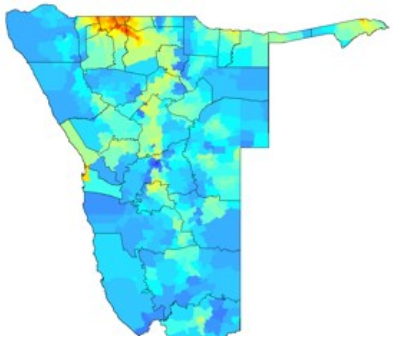
Edited by Michael F. Goodchild, University of California, Santa Barbara, CA, and approved September 15

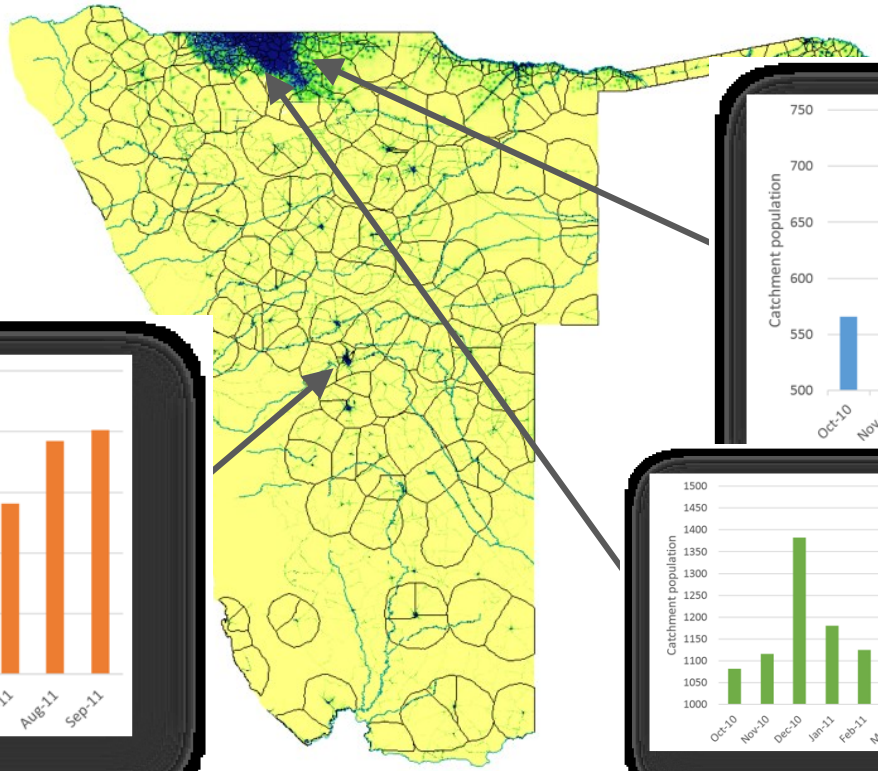


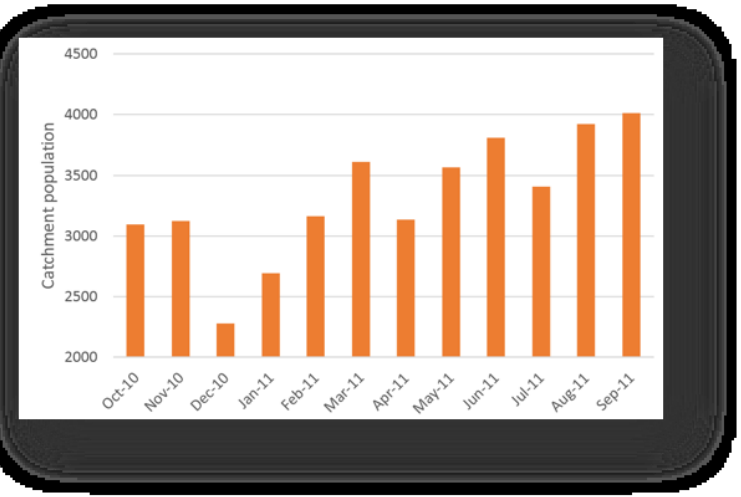
world pop FLOWMINDER.ORG

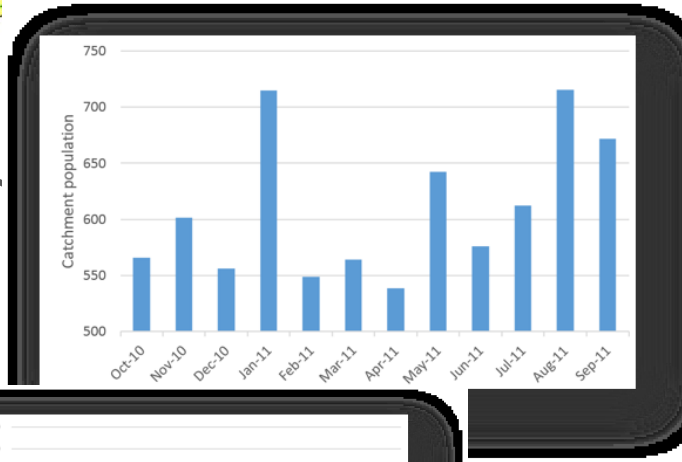


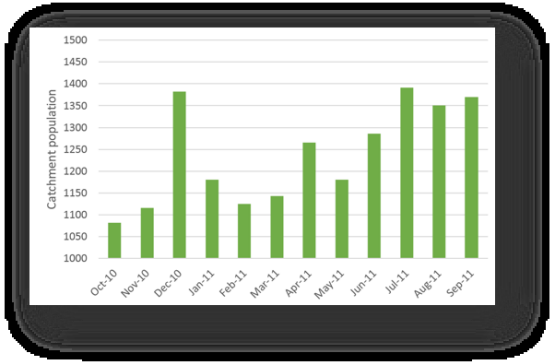
Dynamic Facility Catchment Populations







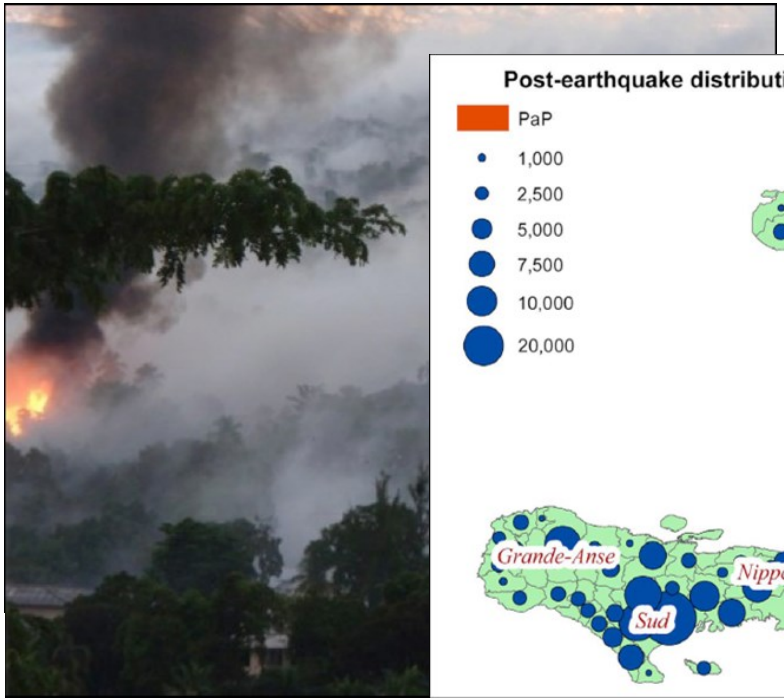


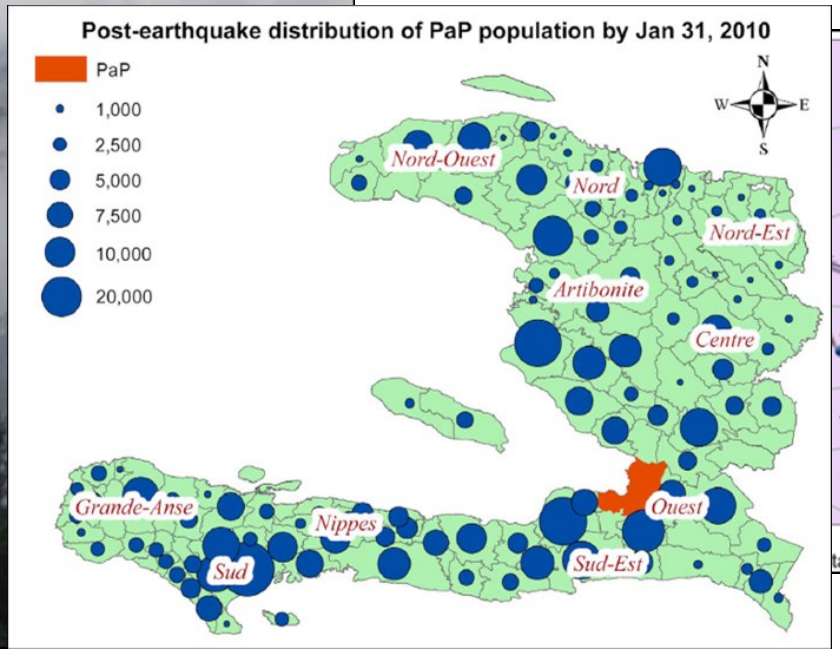


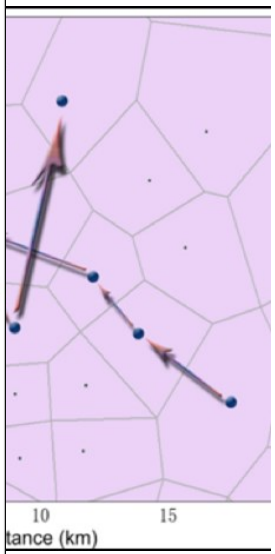
Population Health Metrics. Zu

Disaster Causes Large-Scale Population Movements

Haiti Earthquake Response







Bengtsson et al. PLoS Medicine 2011

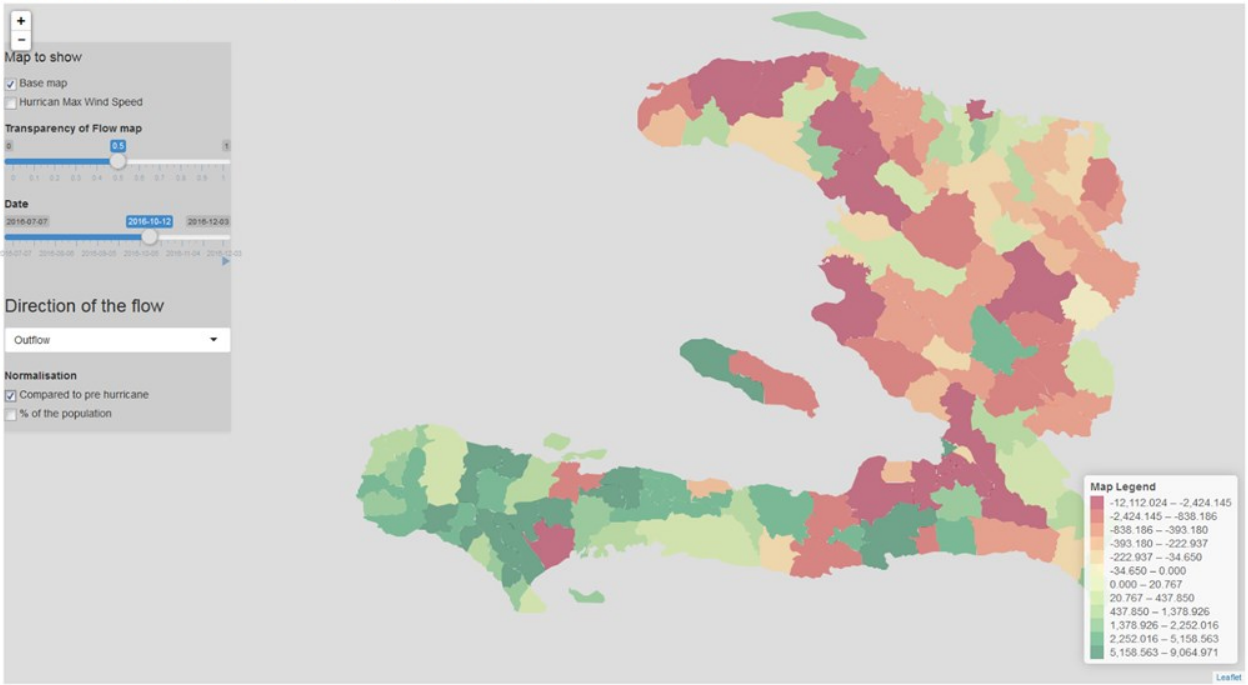
Haiti Hurricane Matthew Response

13 CLIMATE ACTION

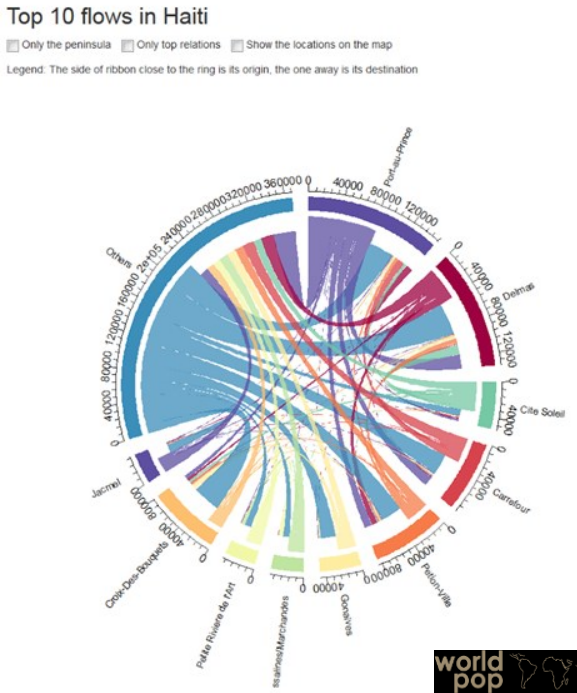
Mobile phone data and population displacement following hurricane Matthew

tab-9316-1 tab-pane active tab-9316-2

Outflow of people : compared to pre-hurricane



Top 10 flows in Haiti



world pop FLOWMINDER.ORG

Nepal Earthquake 2015

First insights within 14 days

Nepal Population Estimates as of May 1, 2015

2.8m

+390,000 (246,000~540,000)

−247,000 (−155,000~−339,000)

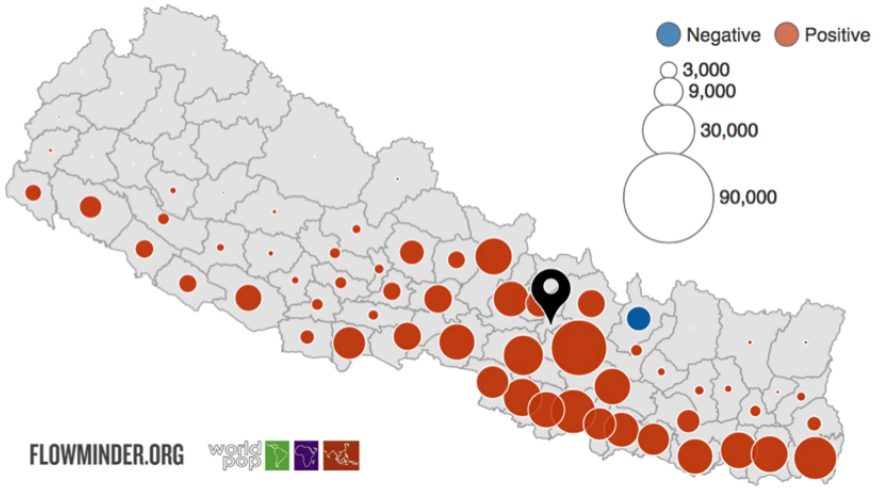
2. Kathmandu Valley

Kathmandu Valley is here defined as the districts Kathmandu, Bhaktapur and Lalitpur. Kathmandu Valley is one of the most densely populated areas in Nepal and home to ca 2.8 m people [1].

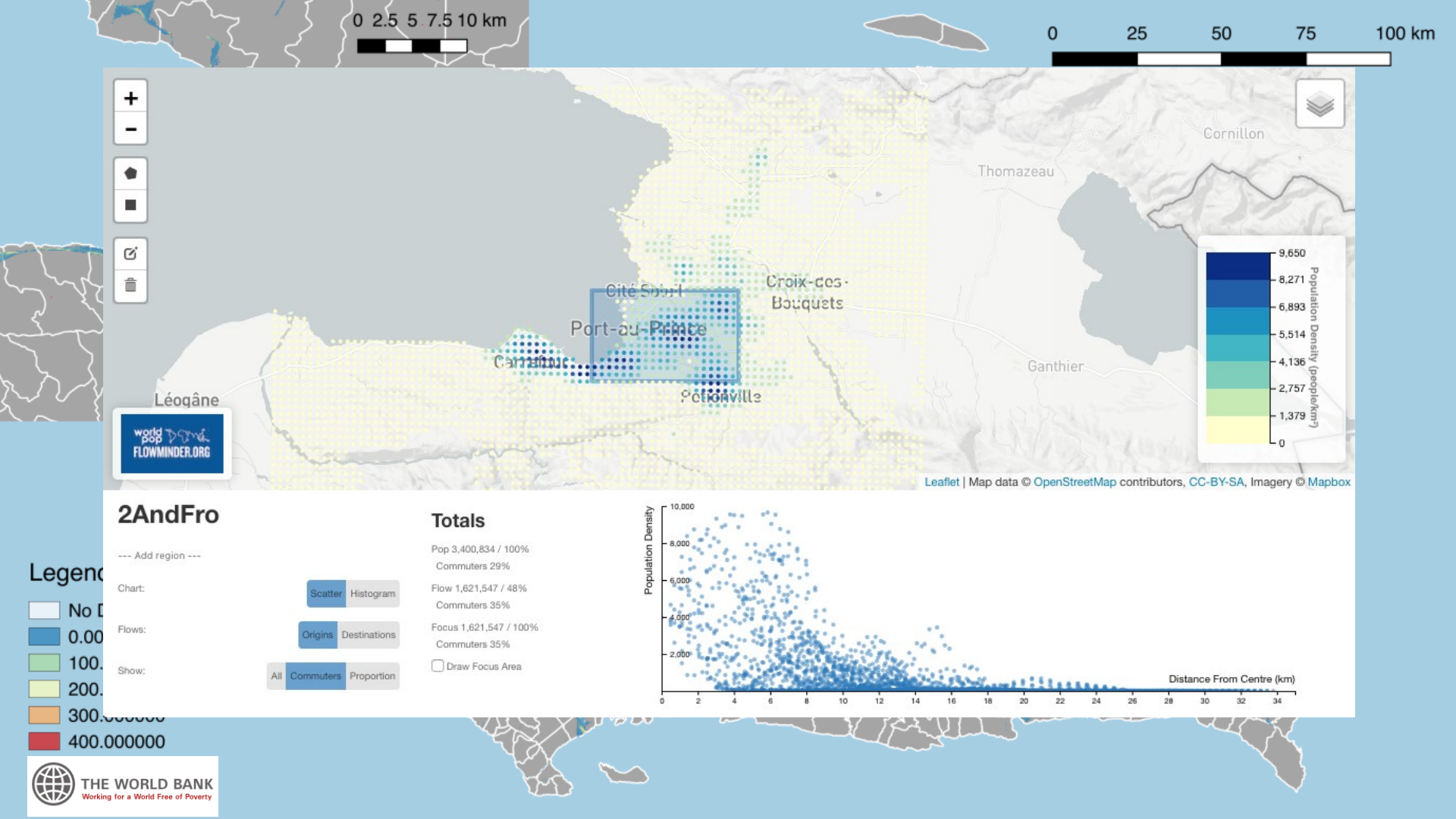
Key findings:

- An estimated 390,000 people more than normal had left the Kathmandu valley - comparing May 1 with the day before the earthquake April 24 (ratio to the population: 14%).
- An estimated 247,000 persons less than normal had come into the area during the same period(ratio to the population: 8.8%)
- People leaving Kathmandu Valley went to a large number of areas, notably the populous areas in the south and the Central and West Development Regions.

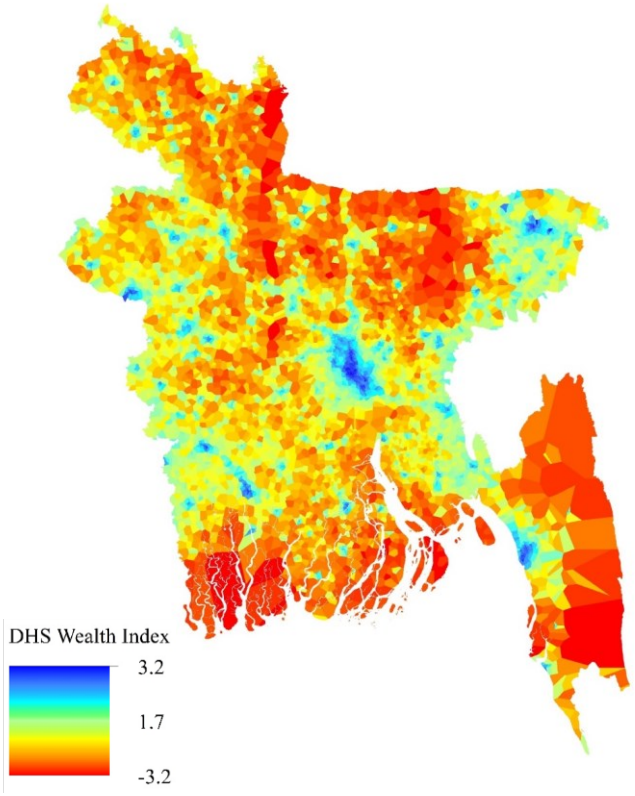
Above normal flows from Kathmandu Valley to other districts

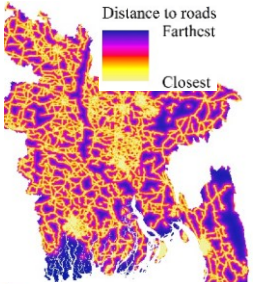


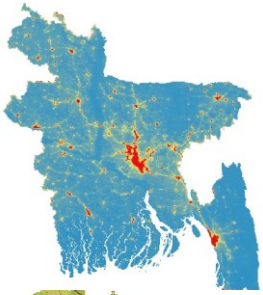
Bengtsson et al. Sci Rep. 2015; Wesolowski et al. PLoS Currents 2014; Wilson et al. PLoS Current 2016.

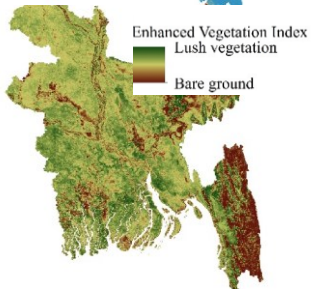

0 2.5 5 7.5 10 km
0 25 50 75 100 km
Cornillon
Thomazeau
Cité Soleil
Croix-des-Bouquets
Port-au-Prince
Carrefour
Pétionville
Ganthier
Léogâne
9,650
8,271
6,893
5,514
4,136
2,757
1,379
0
Population Density (people/km²)
Leaflet | Map data © OpenStreetMap contributors, CC-BY-SA, Imagery © Mapbox
2AndFro
--- Add region ---
Chart:
Scatter
Histogram
Flows:
Origins
Destinations
Show:
All
Commuters
Proportion
Totals
Pop 3,400,834 / 100%
Commuters 29%
Flow 1,621,547 / 48%
Commuters 35%
Focus 1,621,547 / 100%
Commuters 35%
Draw Focus Area
Population Density
10,000
8,000
6,000
4,000
2,000
Distance From Centre (km)
0
2
4
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
Legend
No D
0.00
100.
200.
300.000000
400.000000
THE WORLD BANK
Working for a World Free of Poverty

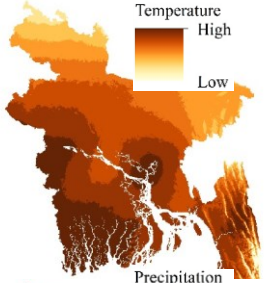
Mobile data for mapping poverty: potential for ongoing monitoring

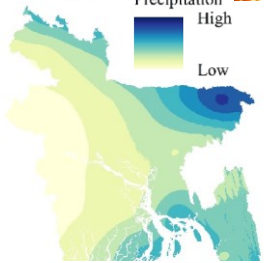


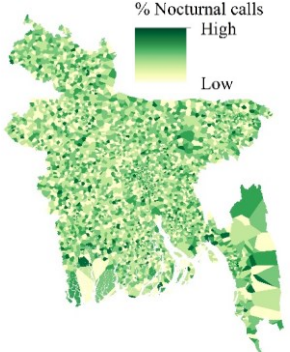


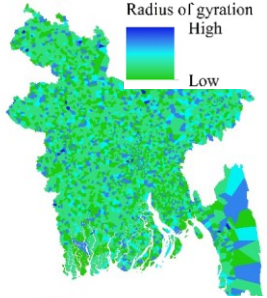


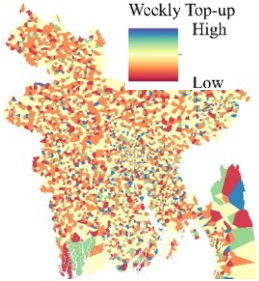


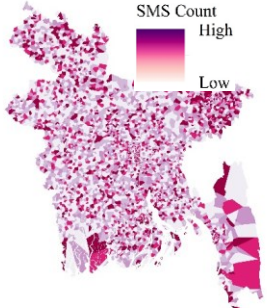












Steele et. al. Royal Soc. Int. 2017


STAR TREK

The Star Trek Fallacy

1. Data **is the tool, not the solution** – issue-driven vs. data-driven problem solving

2. Remote sensing data and analytics **can augment but not replace** traditional data (”ground truth”), eg. surveys

3. Few studies of bias

4. Bias include:
   - Mobile data is **heterogeneous** – market/operators.
   - Fundamental characteristics (subscribers) **constantly changing**
   - **Representativeness** – what does a SIM card represent?
   - **Real-time mobile data without validation = real-time mistakes**

Summary

- New data sources and statistical data have a profound impact on our ability to map **population distributions, characteristics and dynamics**
- We are seeing an explosion of high quality geospatial data, but every dataset has its biases and gaps: ***Data integration is key***
- Potential to rapidly scale the use of mobile operator data, but **accounting for bias is crucial**