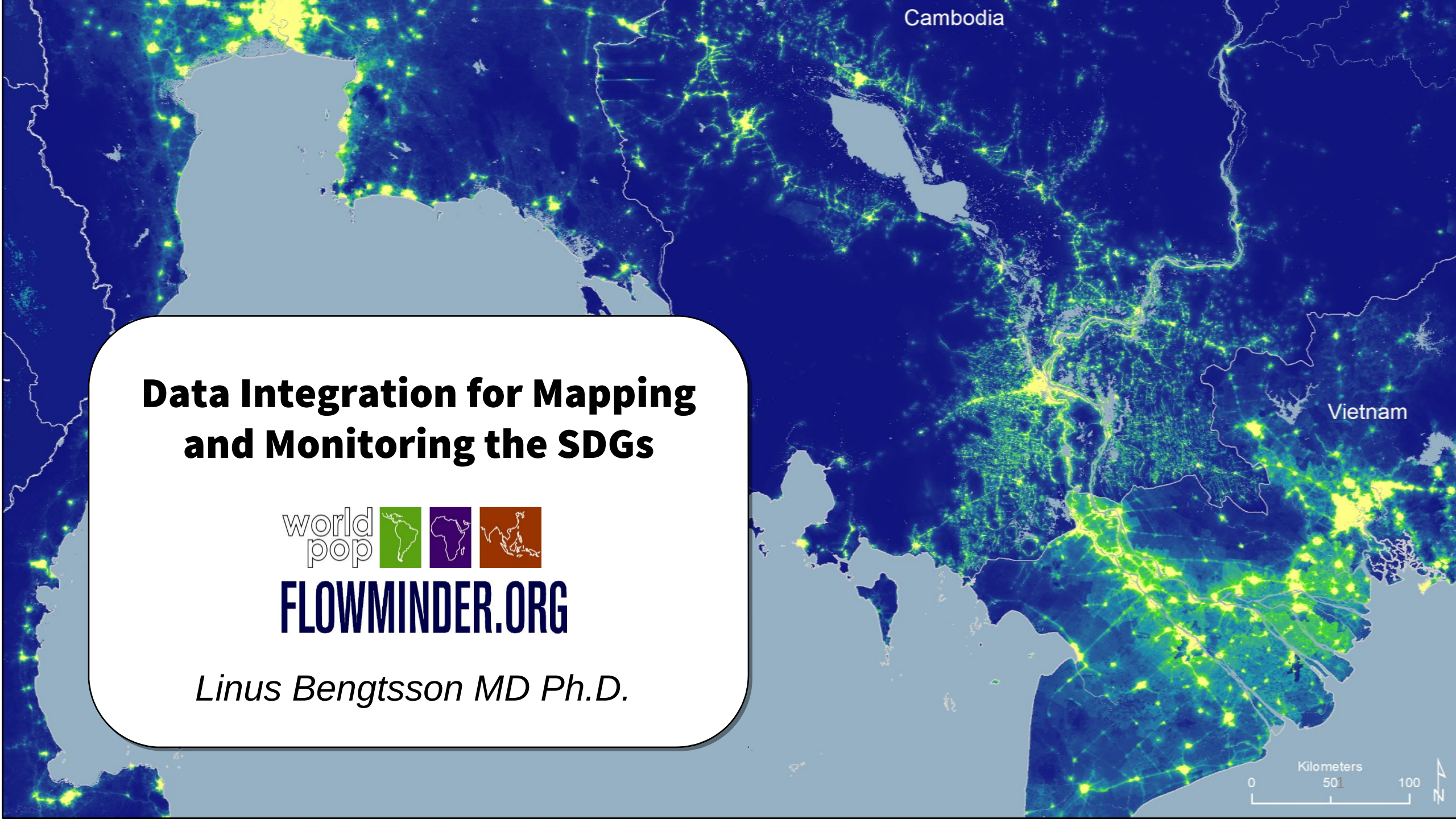


Data Integration for Mapping and Monitoring the SDGs

worldpop

FLOWMINDER.ORG

Linus Bengtsson MD Ph.D.

FLOWMINDER.ORG

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

Karolinska Institutet

University of Southampton

Harvard T.H. Chan School of Public Health — Powerful ideas for a healthier world

Stockholm School of Economics

DATA POP ALLIANCE

DATA-POP ALLIANCE

Harvard Humanitarian Initiative

MIT Media Lab

FLOWMINDER.ORG

Key partners and donors:

OCHA

UNFPA

WFP United Nations World Food Programme

International Organization for Migration (IOM)

ADB Asian Development Bank

Clinton Health Access Initiative

Bill & Melinda Gates foundation

The World Bank — Working for a World Free of Poverty

The Rockefeller Foundation

United Nations Foundation

Pioneered Anonymized Mobile Network Data for Infectious Disease: (2008 Zanzibar, Kenya, 2012 Haiti, 2013- Namibia, Indonesia) and Crisis Response (Haiti 2010 earthquake and cholera. Nepal 2015)

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.

Science

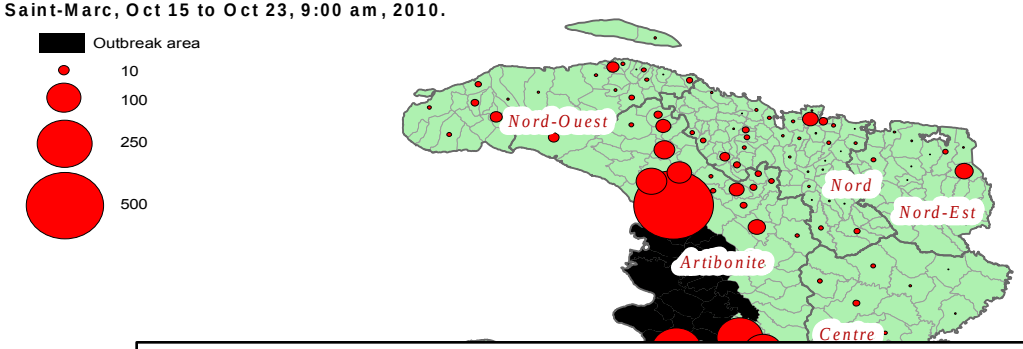
AAAS

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,*]

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



GLOBAL UPDATE

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

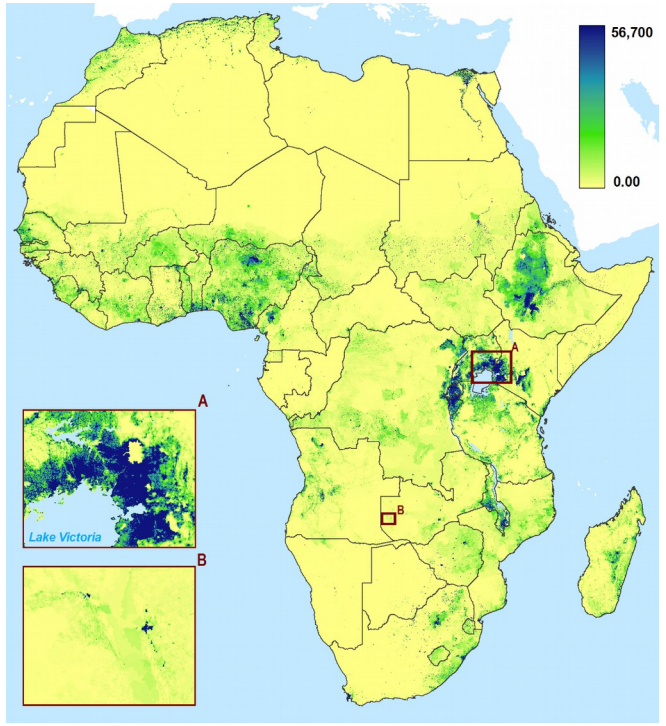
Population distributions

Population characteristics

Population dynamics

WorldPop Project: Improving the spatial demographic evidence base for low and middle income countries

- We develop scalable methods and models for integrating ancillary datasources to complement and fill data gaps in census
- We integrate new technologies, including high resolution satellite imagery and cellphone data
- We run training for government agencies in these methods
- We publish fully documented, peer-reviewed methods and make outputs open access

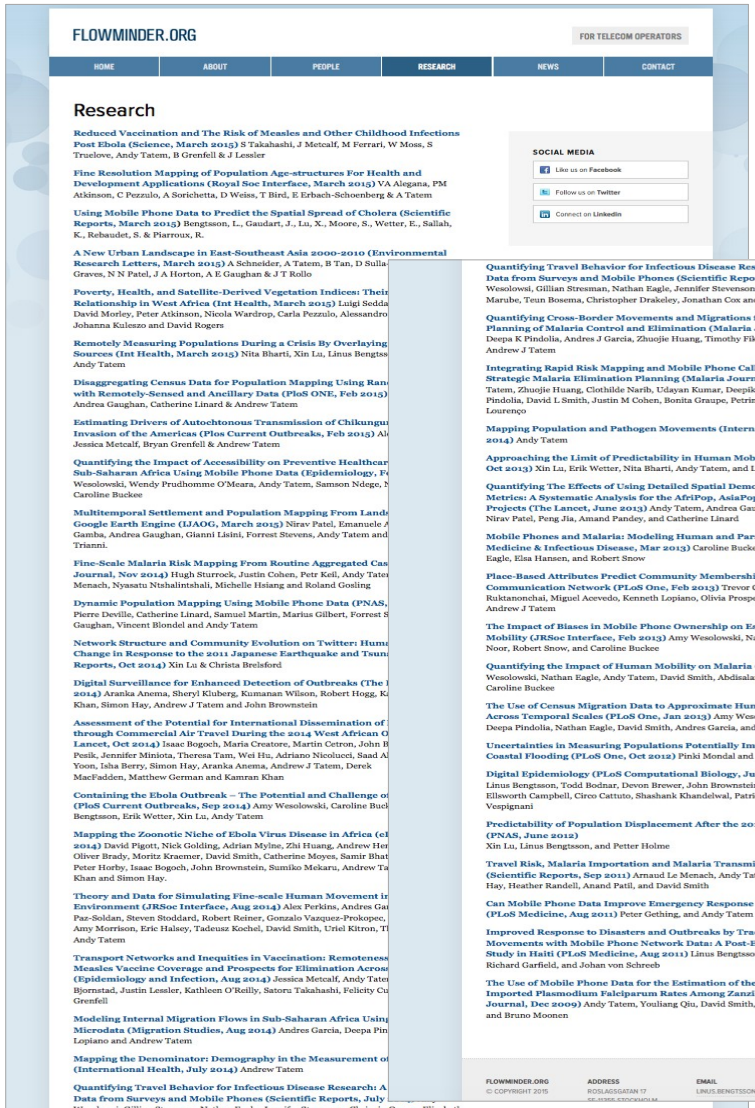


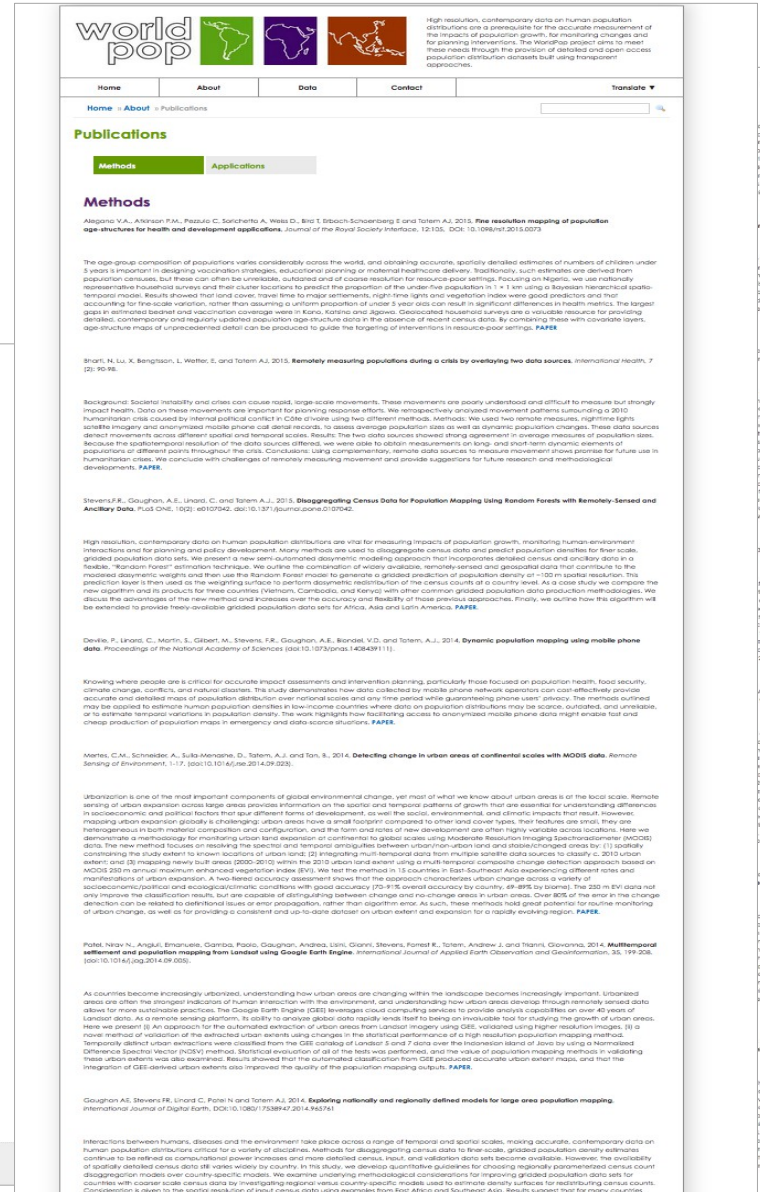
Distribution of live births, 2015

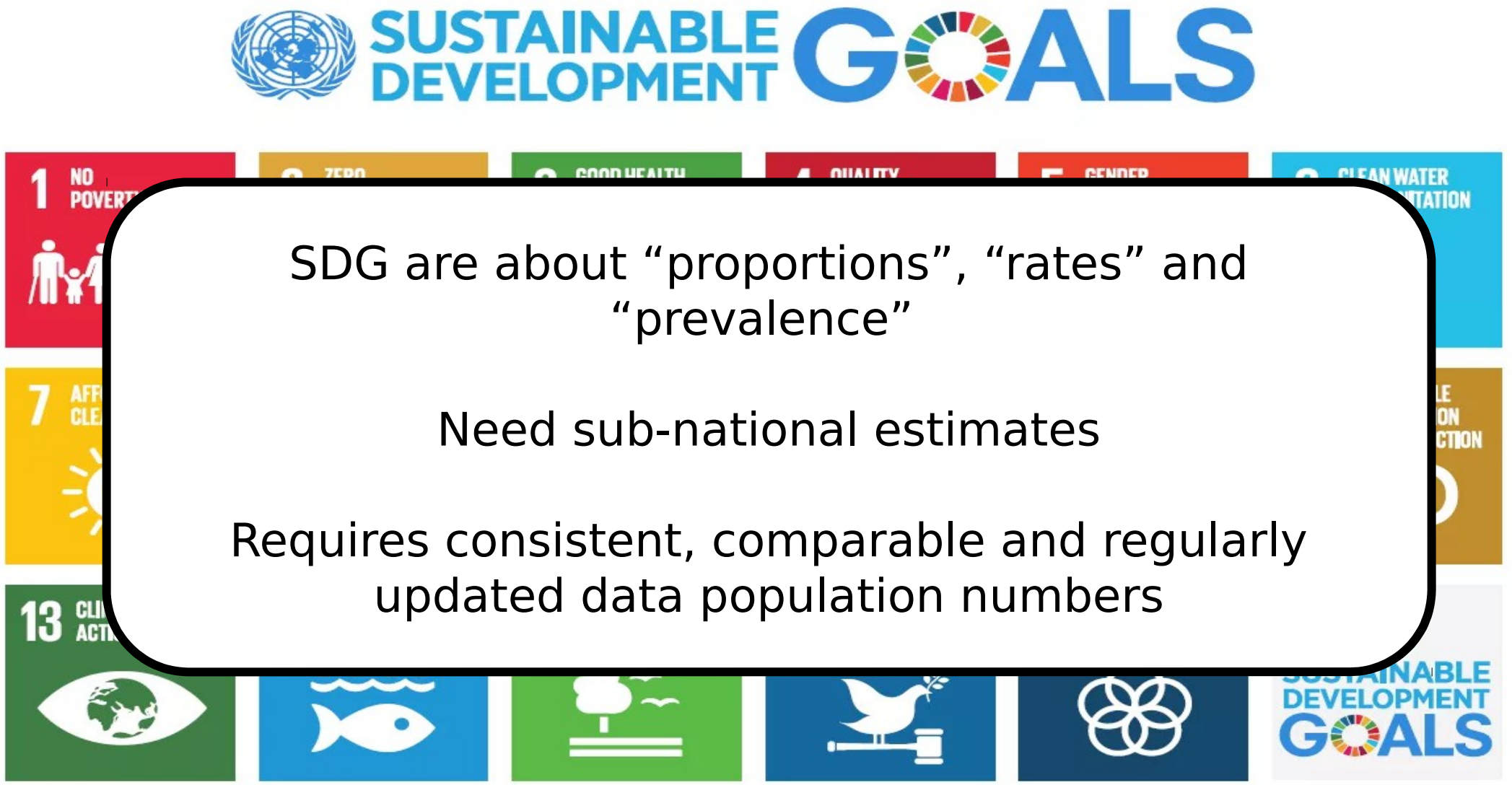
Partners:

University of Southampton
Clinton Health Access Initiative
JRC European Commission
UNFPA
WFP
Norad
CIESIN
The World Bank Working for a World Free of Poverty
Microsoft Research
Bill & Melinda Gates foundation
wellcome trust
OCHA

All Methods are Open and Published in Peer Reviewed Journals for Validation and Transparency

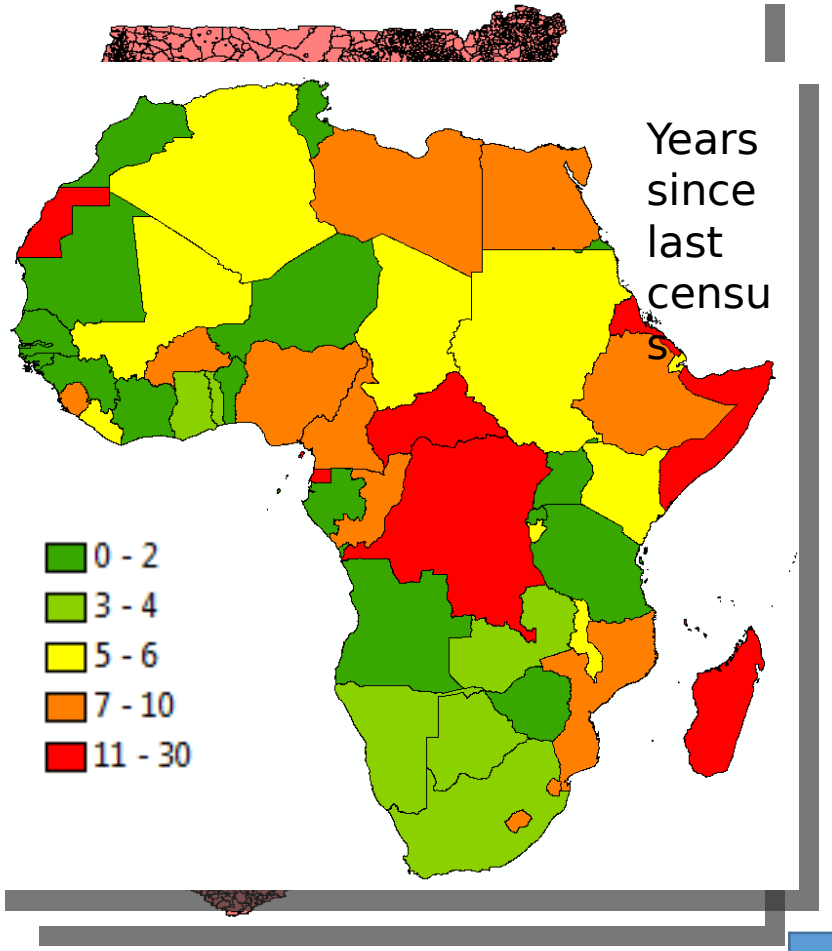


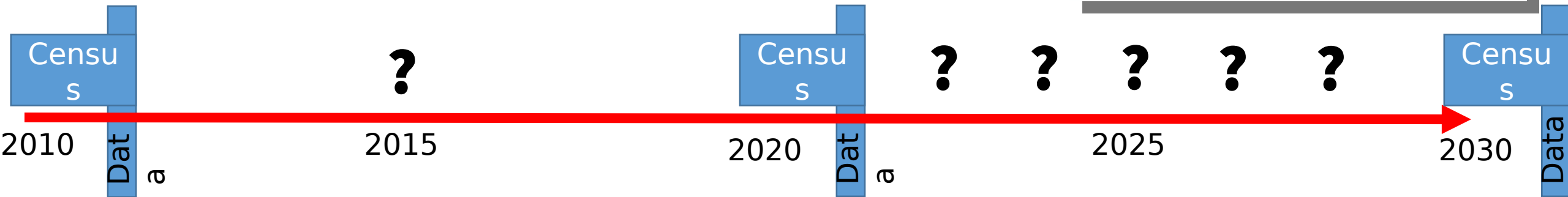



SUSTAINABLE DEVELOPMENT GOALS
SDG are about “proportions”, “rates” and “prevalence”
Need sub-national estimates
Requires consistent, comparable and regularly updated data population numbers

The SDG Challenge: Measuring the Denominator

- National census data will continue to be our most important datasource
- Provides denominators and some numerators for the SDGs, and requisite subnational detail
- But, the 2015-2030 SDG period typically includes just one census datapoint
- And in some settings the situation is more challenging



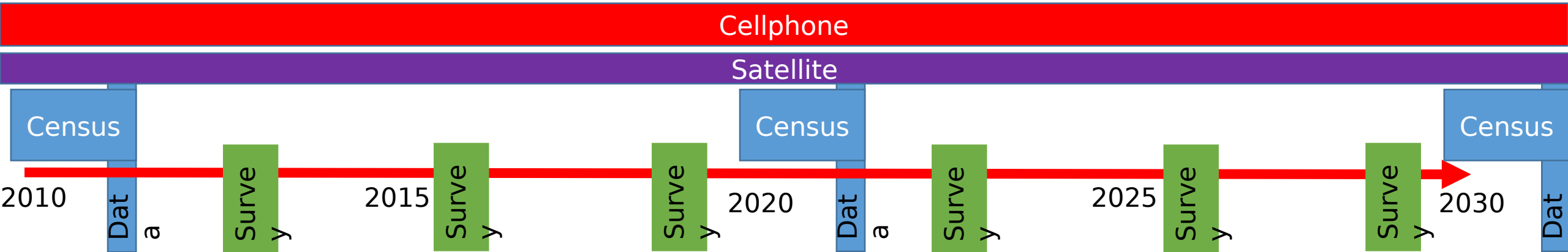


What Do We Have to Help Us?

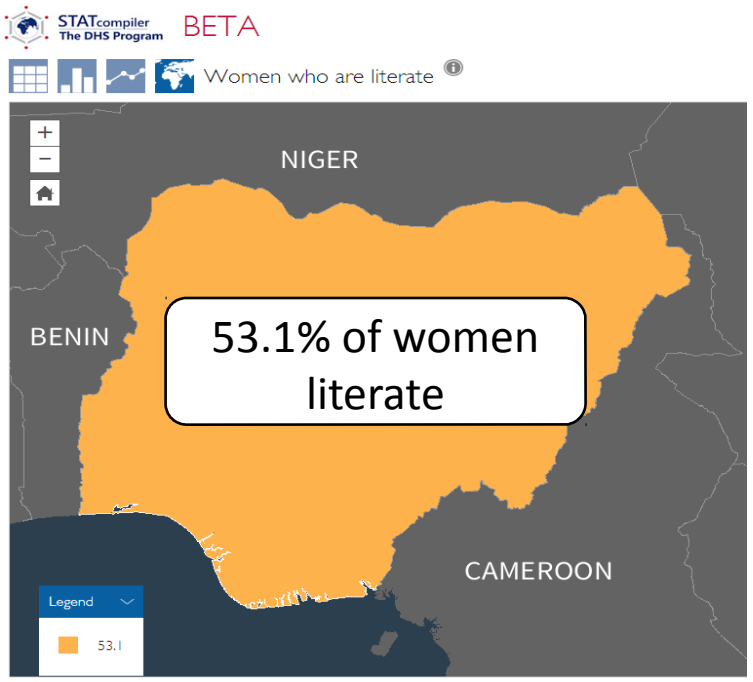
Geolocated household surveys

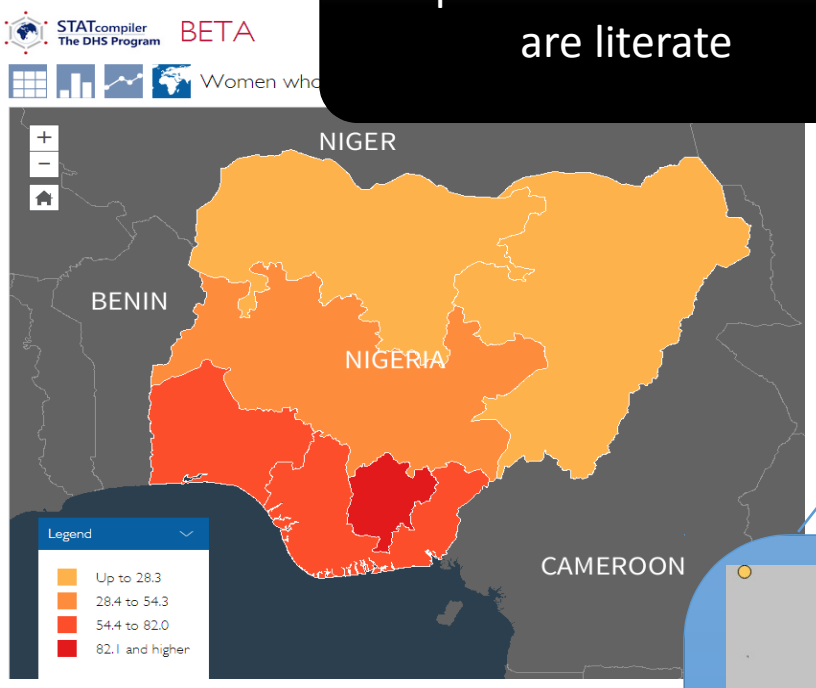
Satellite and GIS data

Cellphone data

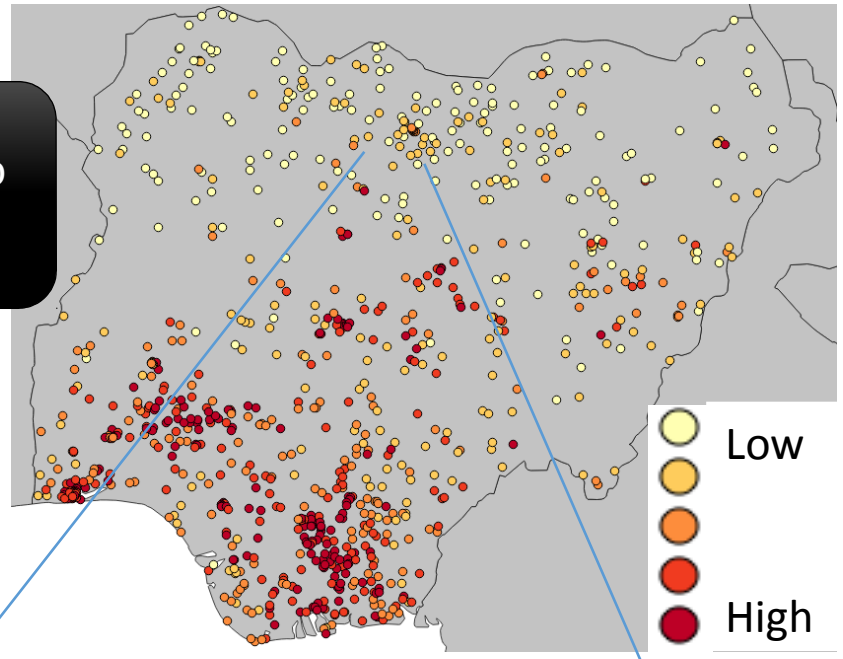


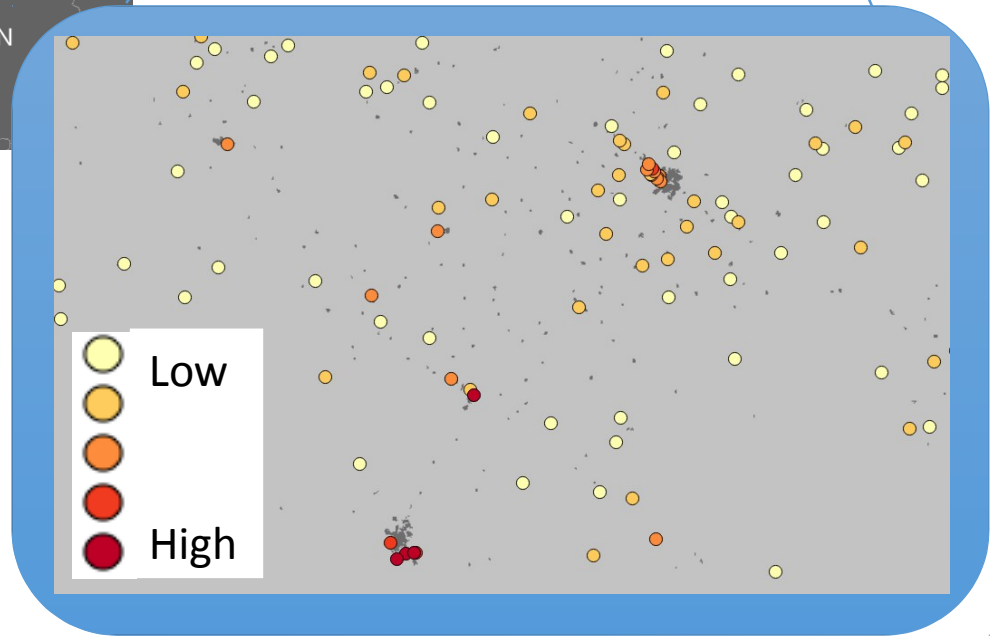
Geolocated Surveys

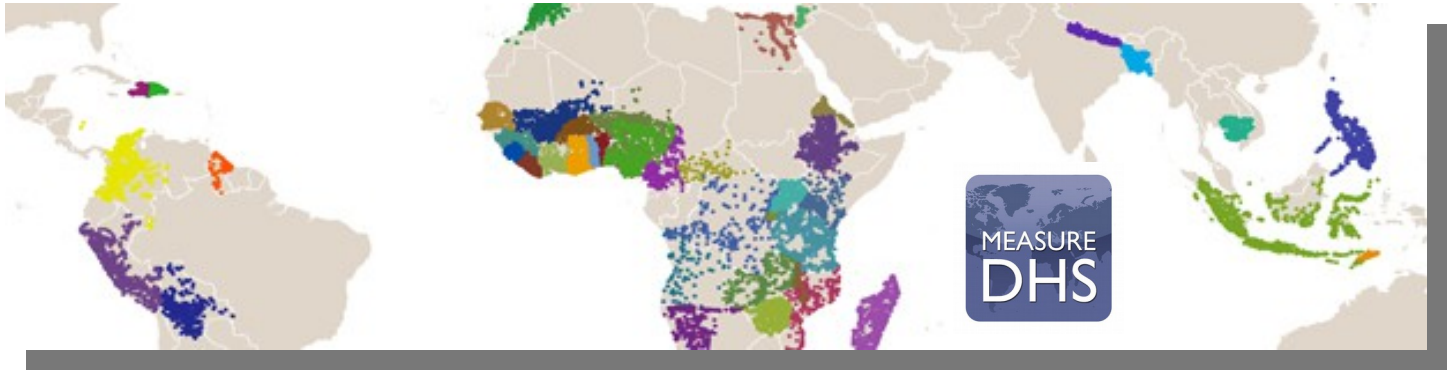




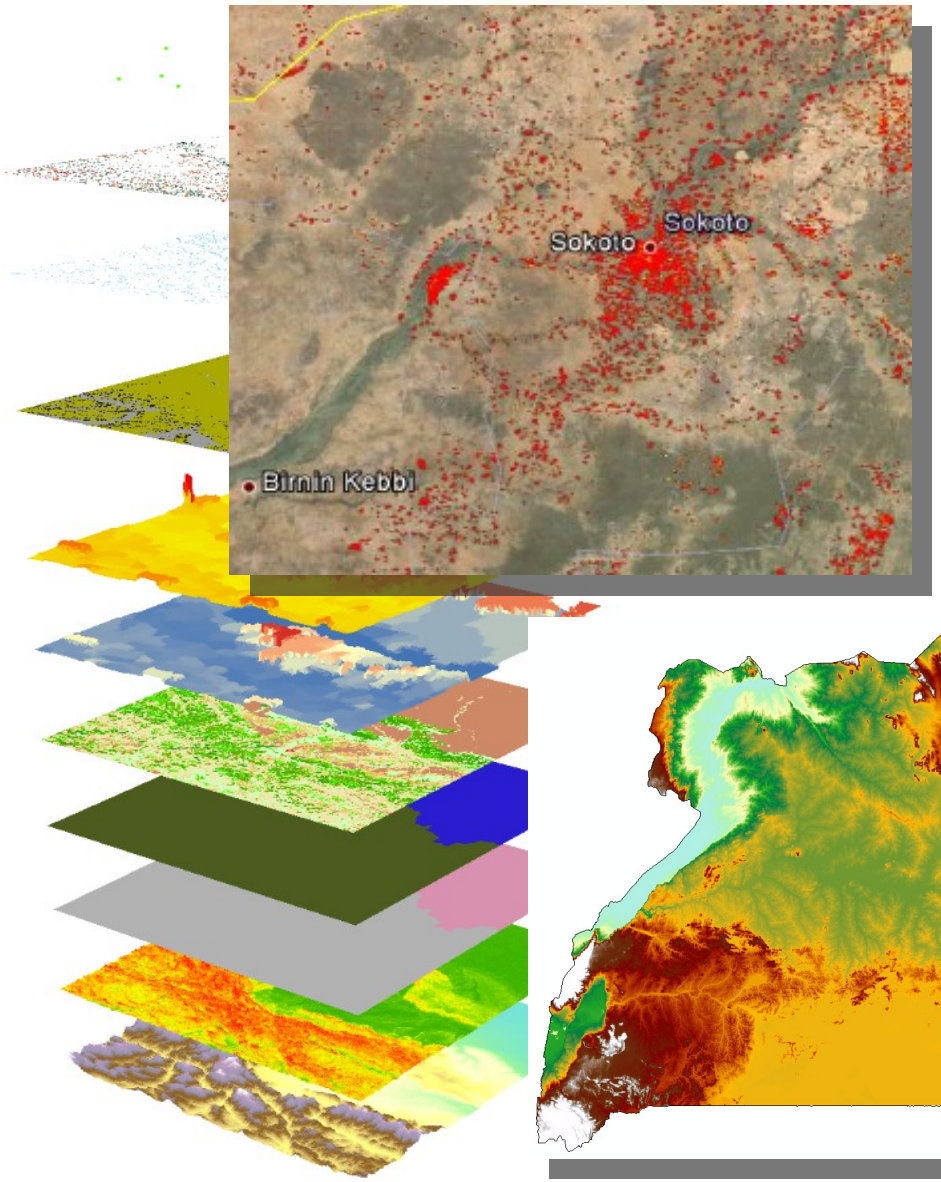
Proportion of women who are literate

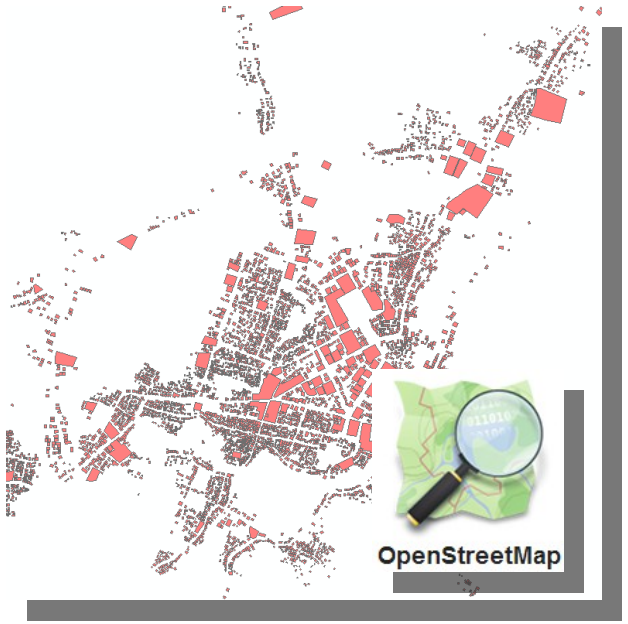


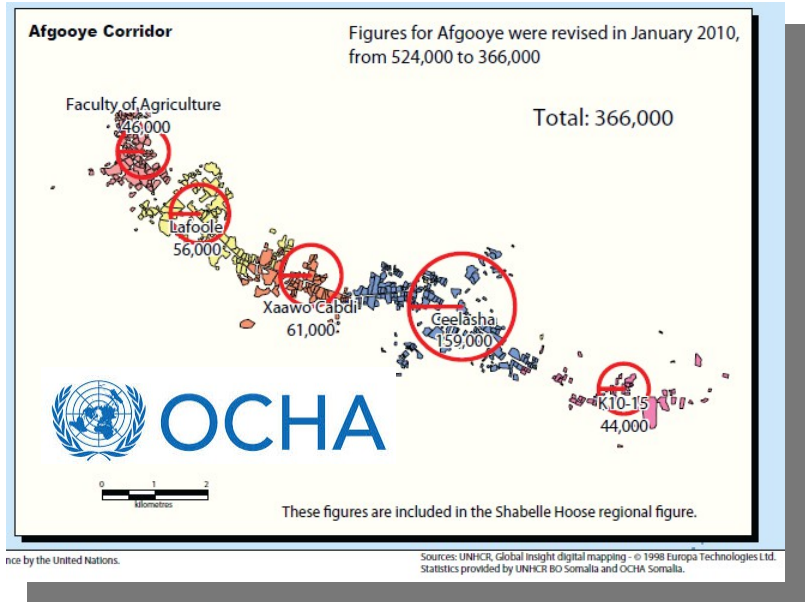


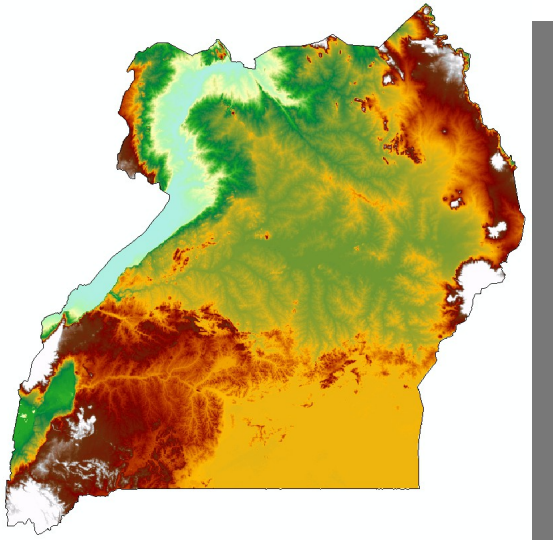


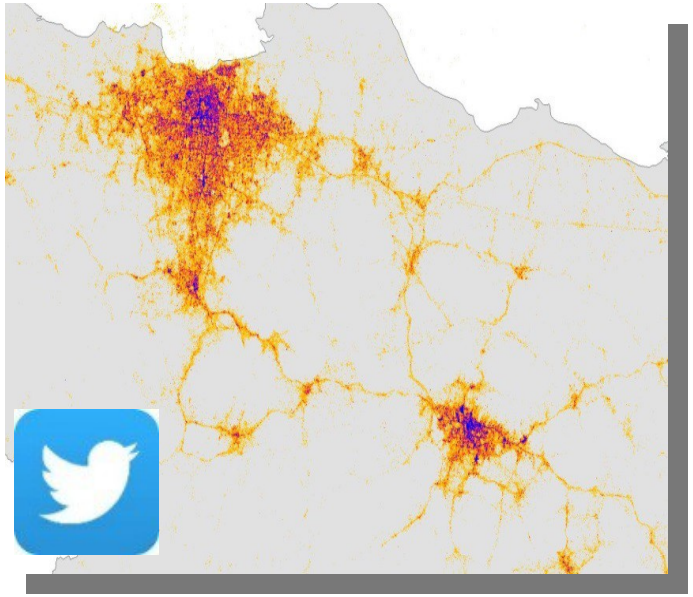
Satellite and GIS data

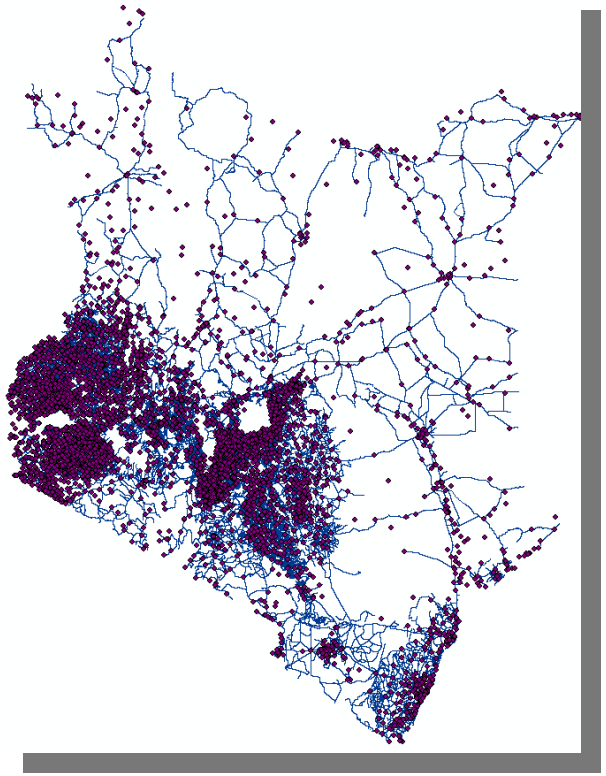




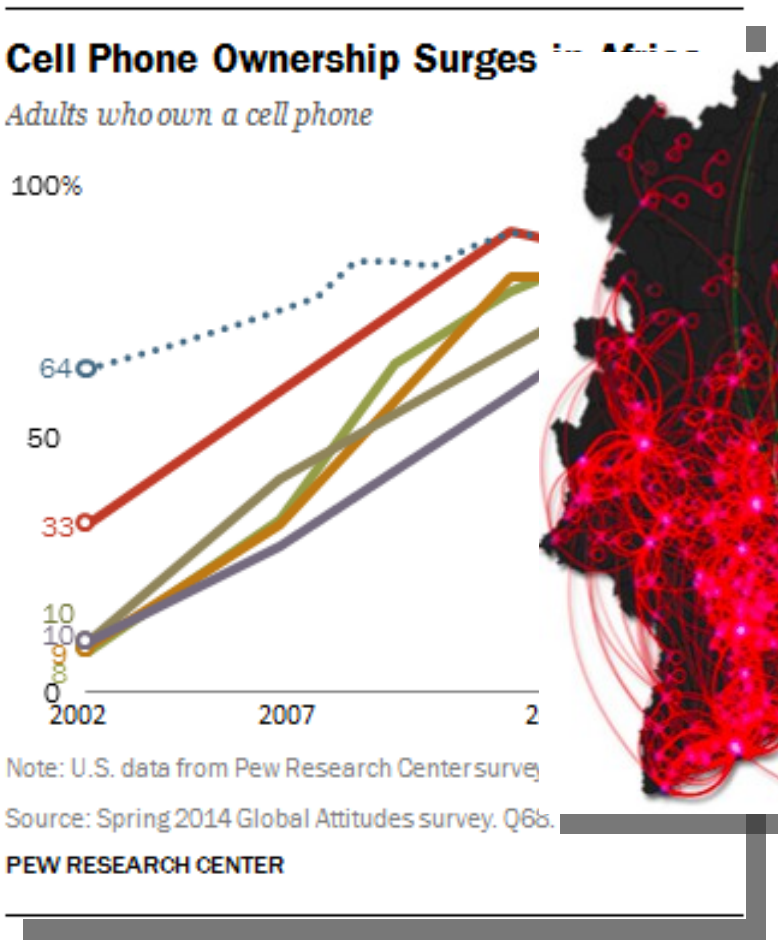




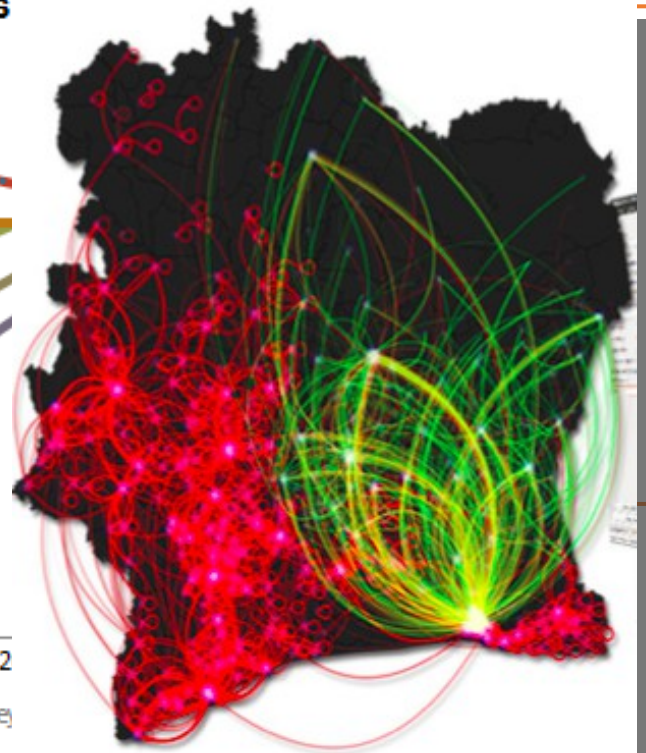




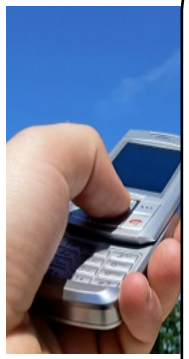
Cellphone Call Detail Records



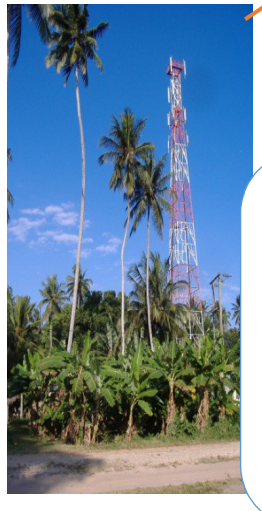
X



User travels to Y and makes a call



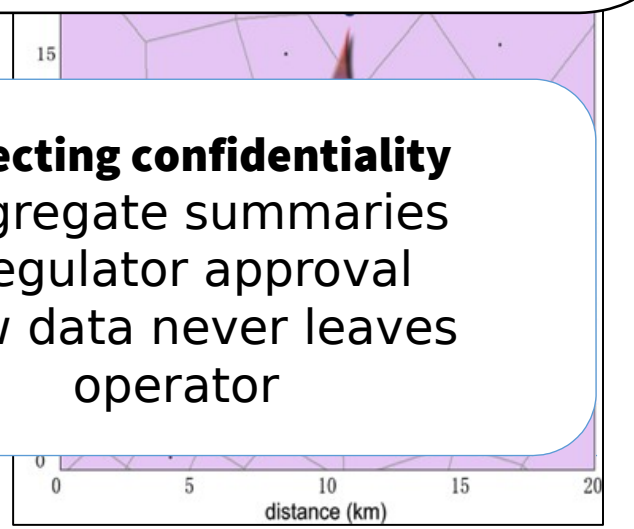
Call routed through near tower



Mobility:
Changing densities, flows, seasonal/permanent migration

Social networks:
Number of contacts, calling patterns

Consumption:
Credit purchase frequencies, top-up amounts



Protecting confidentiality
-Aggregate summaries
-Regulator approval
-Raw data never leaves operator

How can we get there with census data plus these newer datasets?

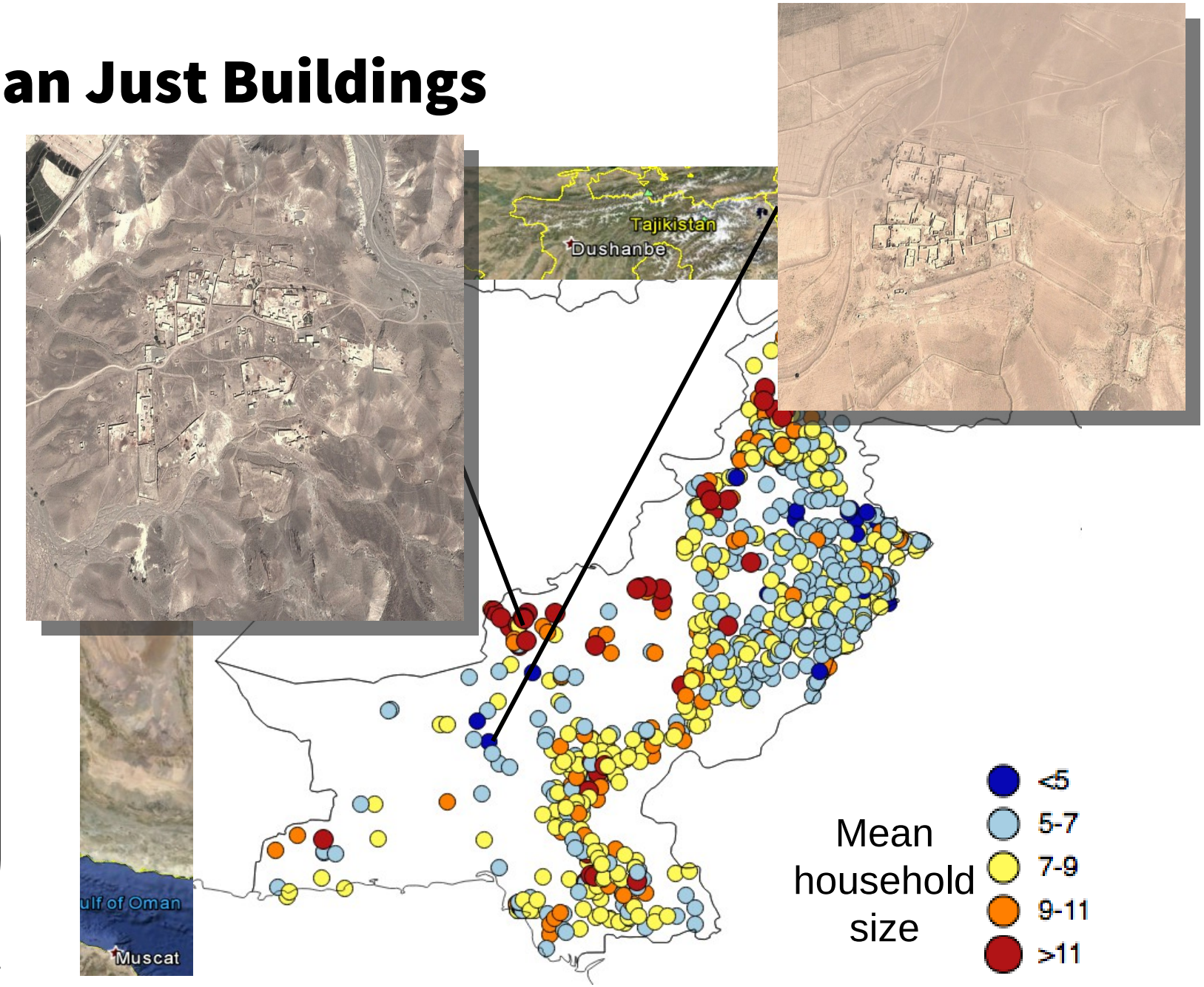
More Information Than Just Buildings

Mapping buildings/settlements is a valuable initial step

But, buildings do not directly translate to people

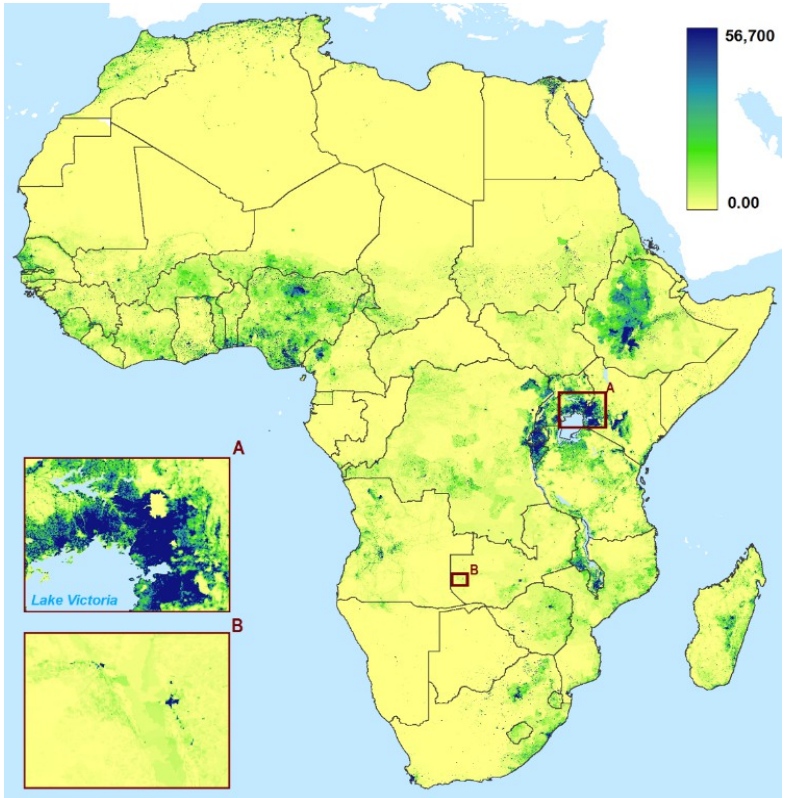
Multiple factors drive differences in distributions and demographics

Importance of *integrating multiple data sources*

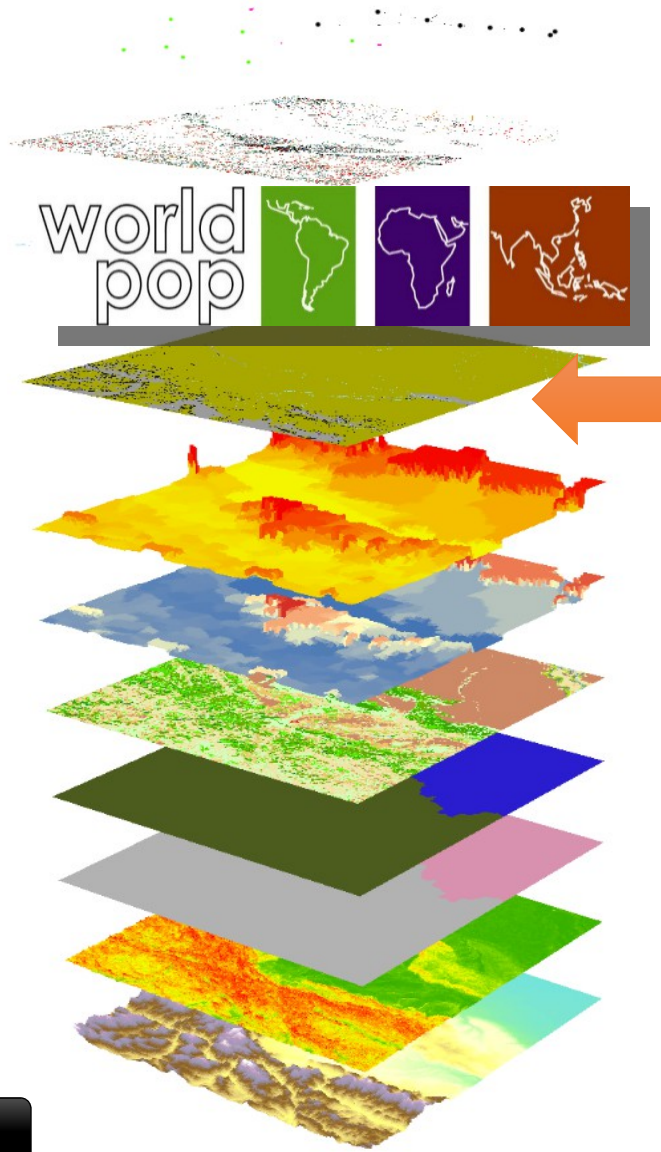


Census Data Disaggregation

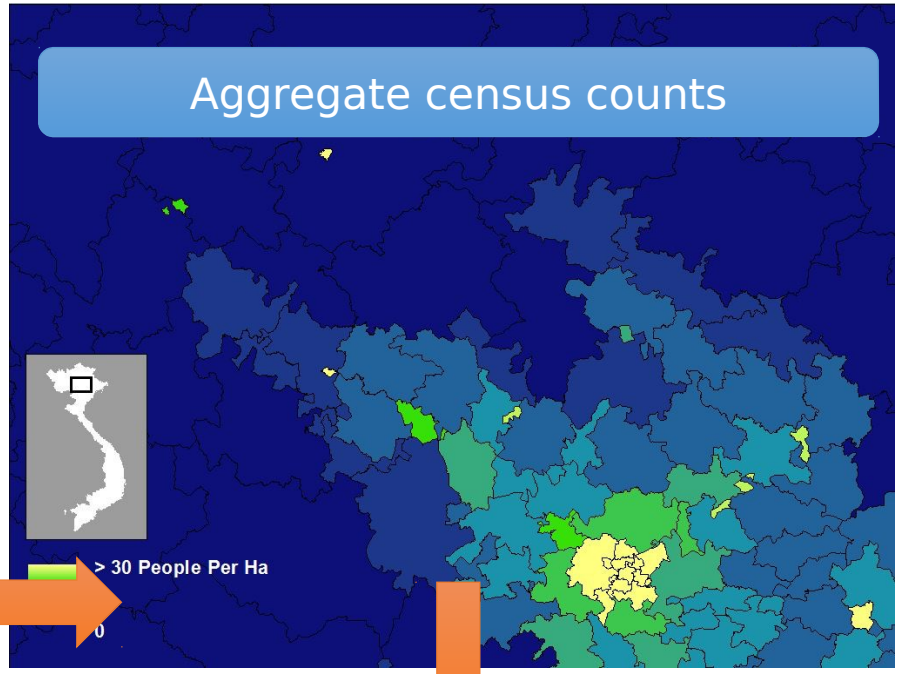
Integration with satellite/GIS data related to human population distribution patterns to disaggregate counts to regular grids



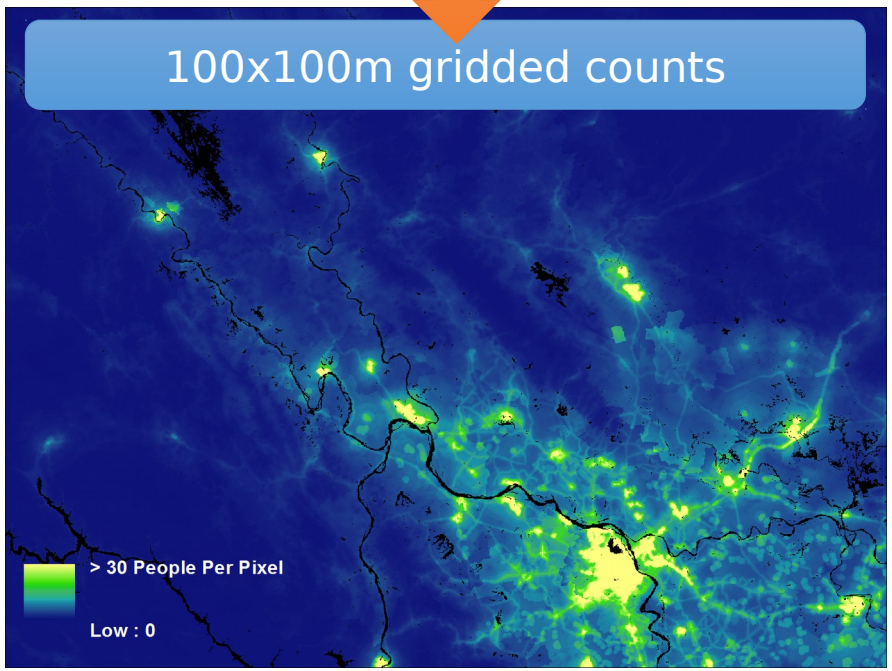
Women of childbearing age per 1x1km 2015



Aggregate census counts

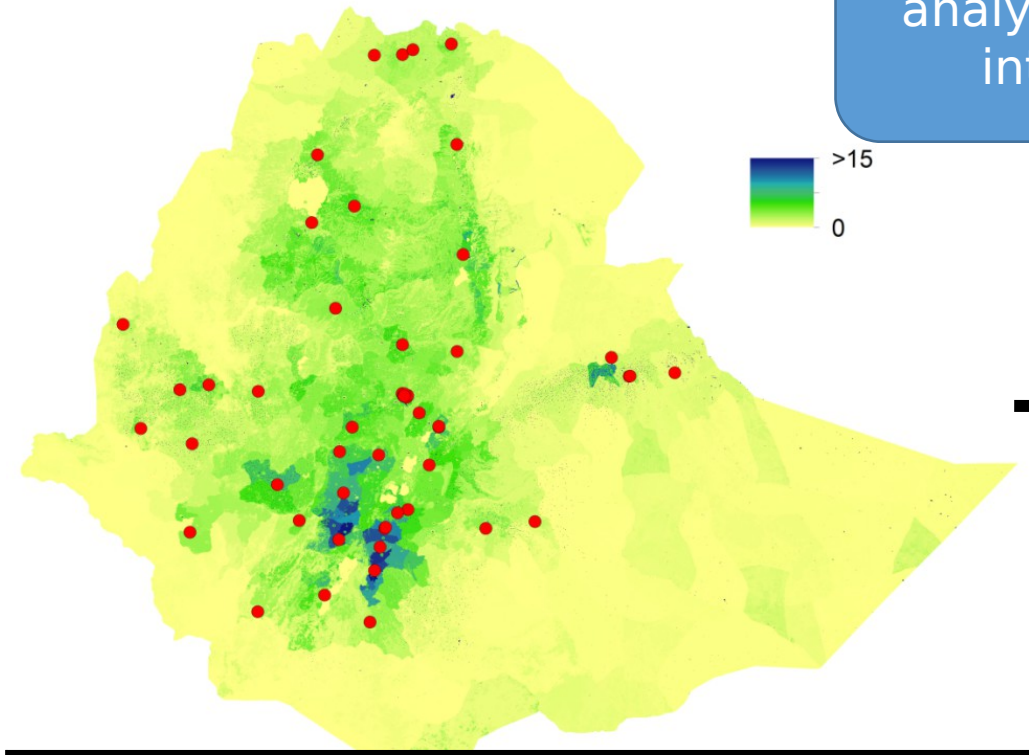


100x100m gridded counts

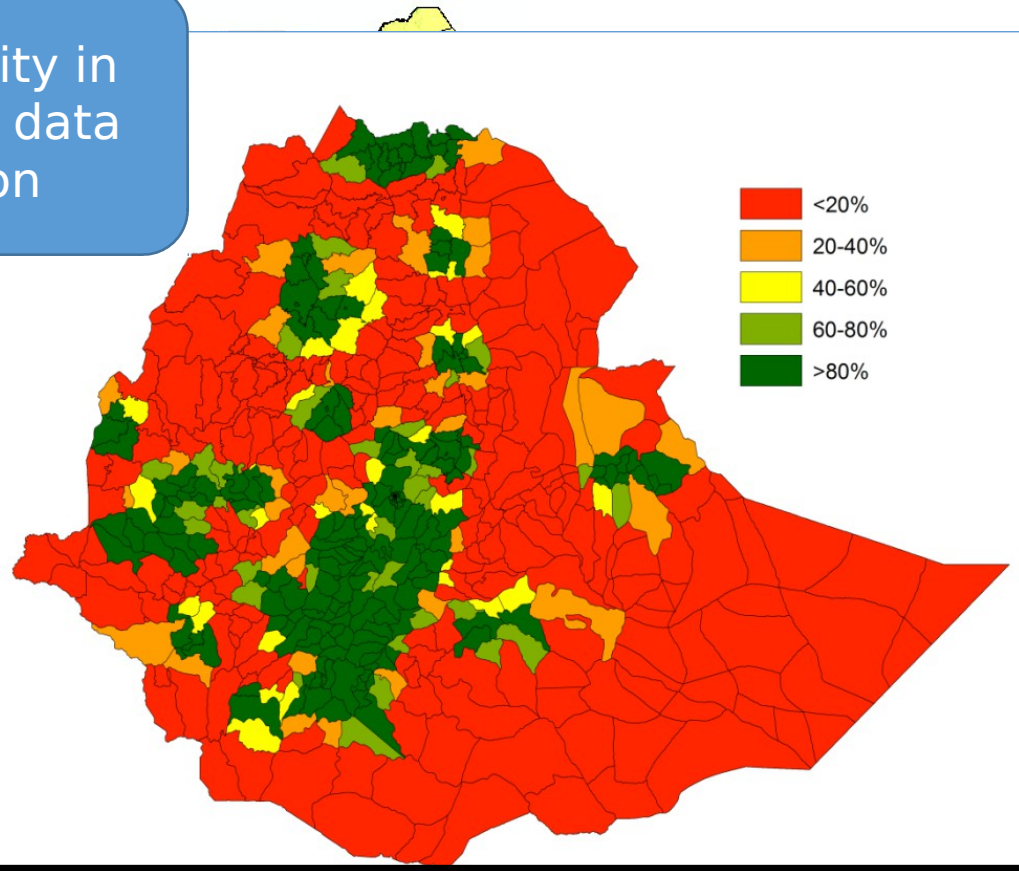


Benefits of ‘Gridded’ Demographic Data

Grids: flexibility in analysis and data integration



Comprehensive Emergency Obstetric and Neonatal Care (CEmONC) Facilities overlaid on grid of women of childbearing age



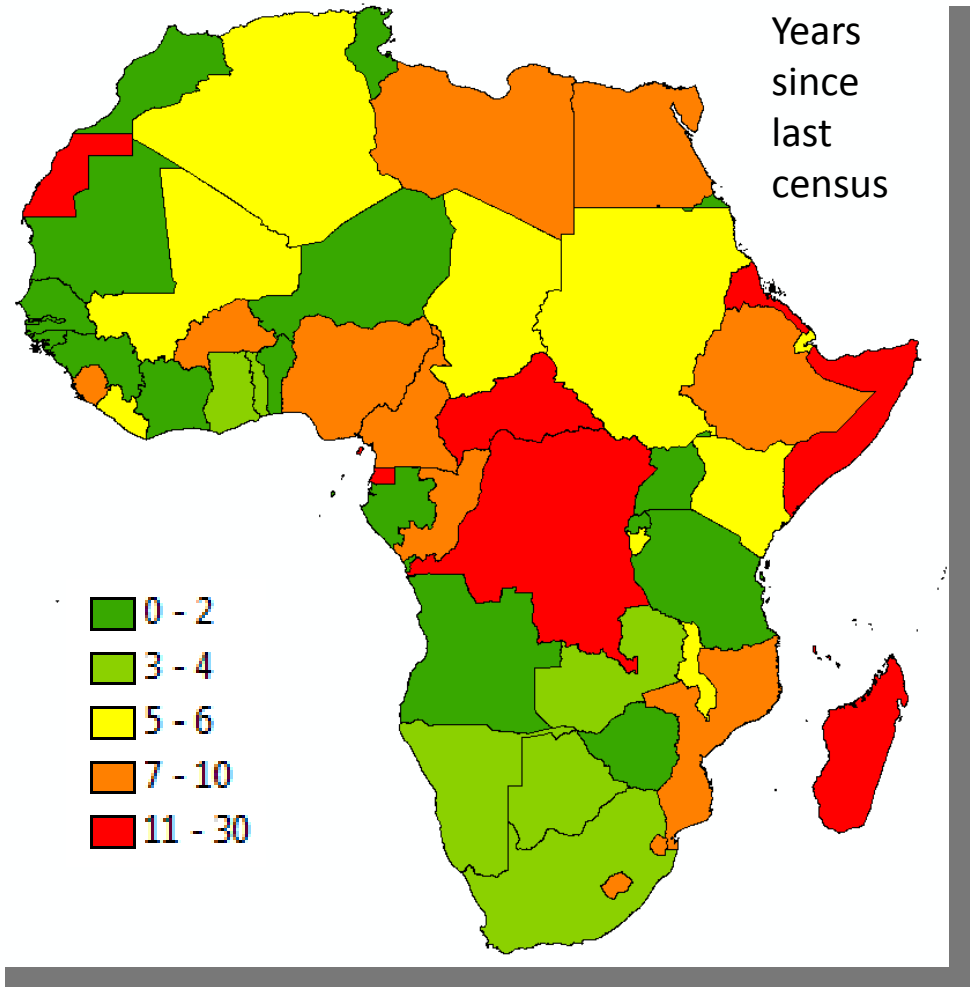
Percentage of women of childbearing age per woreda within 50km of a CEmONC

differing data types

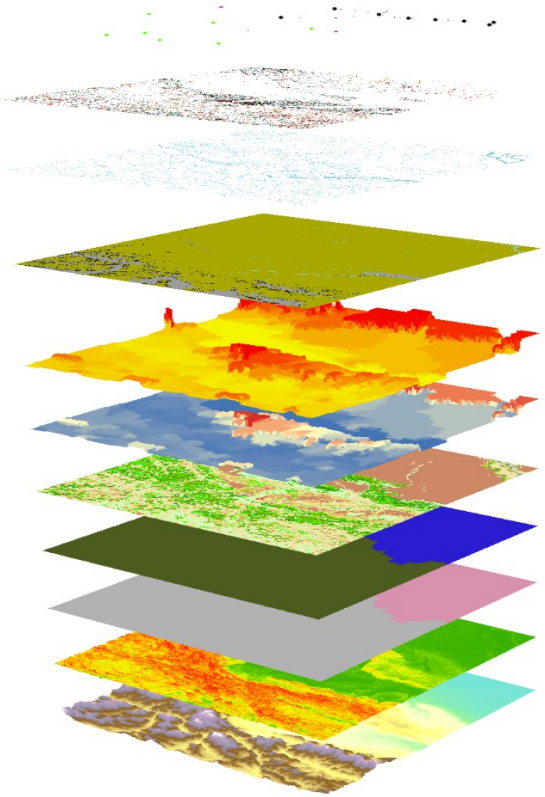
population totals

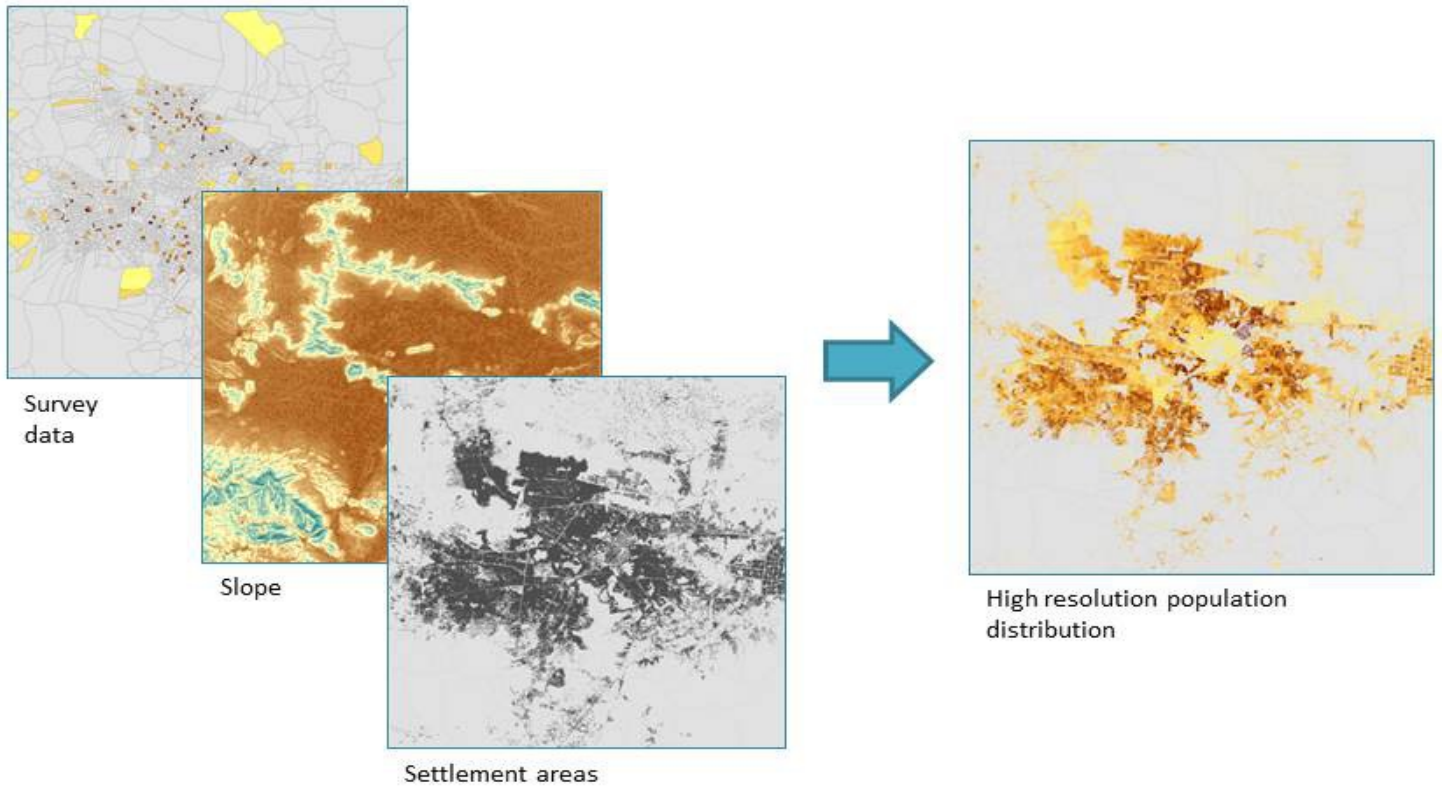
Census Data is not Always Trustworthy

- Every ten years
- Released with delay
- Sometimes much older
- Sometimes manipulated

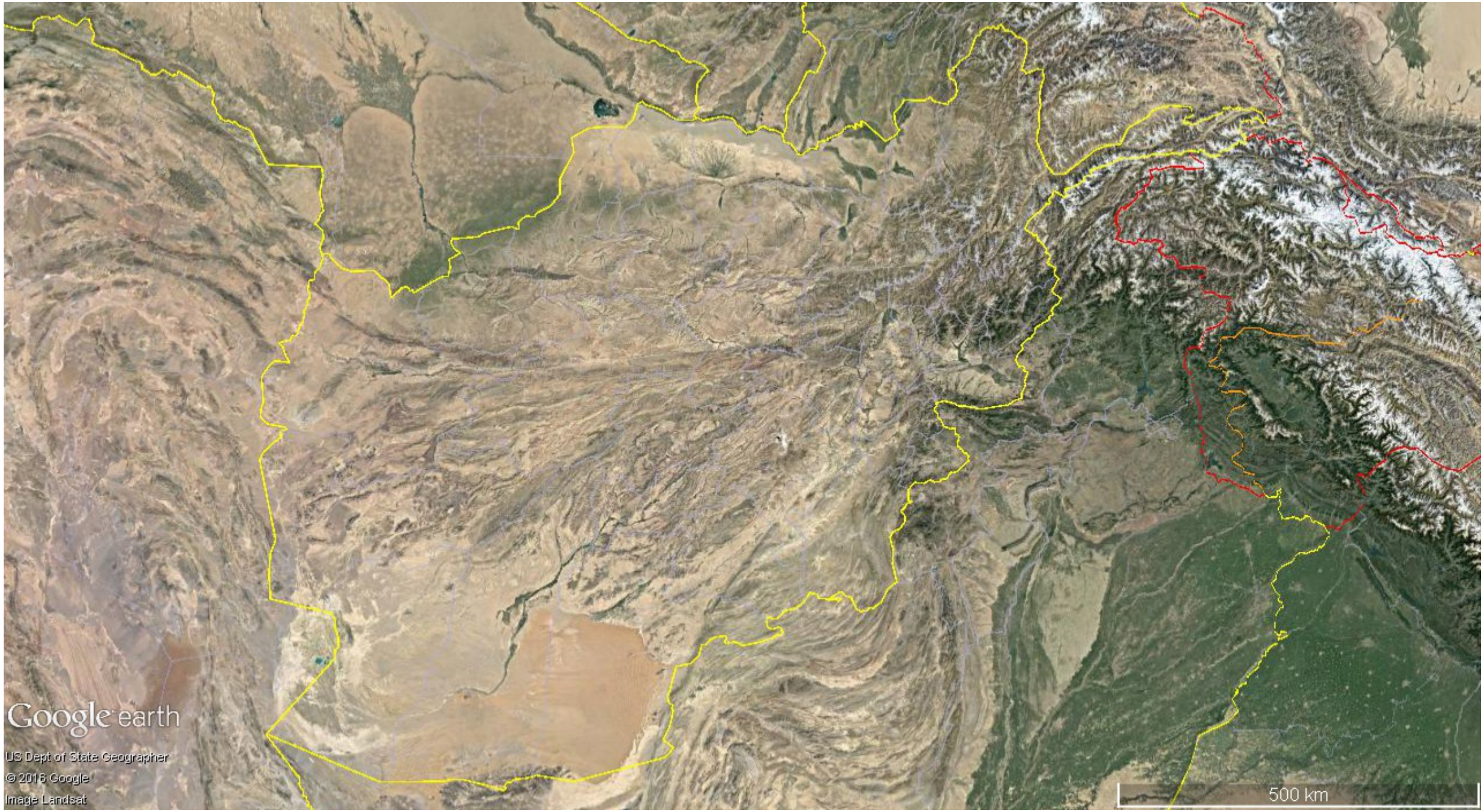


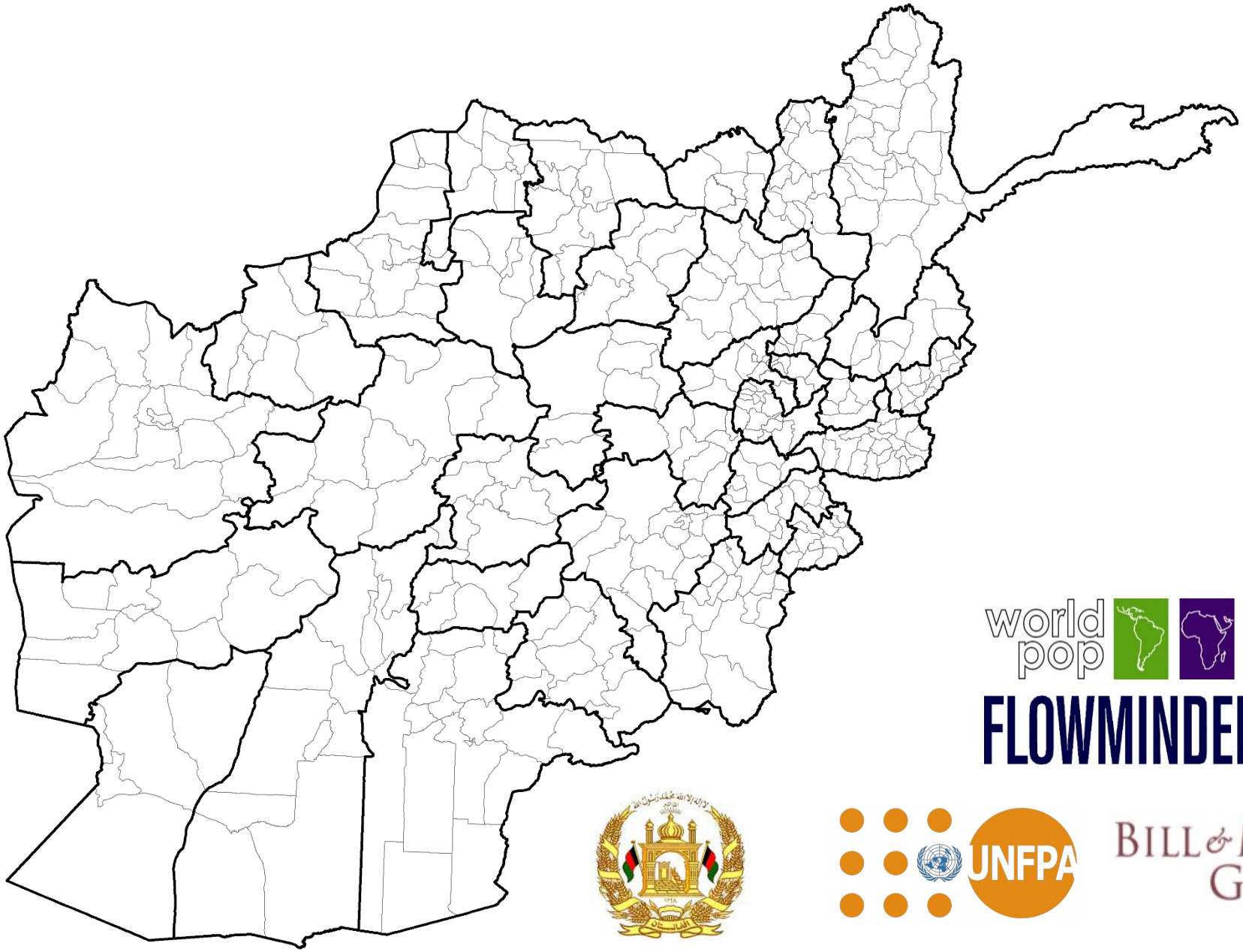
Bottom-up Mapping: Micro-Censuses in Selected Locations





Ref: Alegana et al. Royal. Soc. Interf. 2015


Google earth
US Dept of State Geographer
© 2016 Google
Image Landsat
500 km

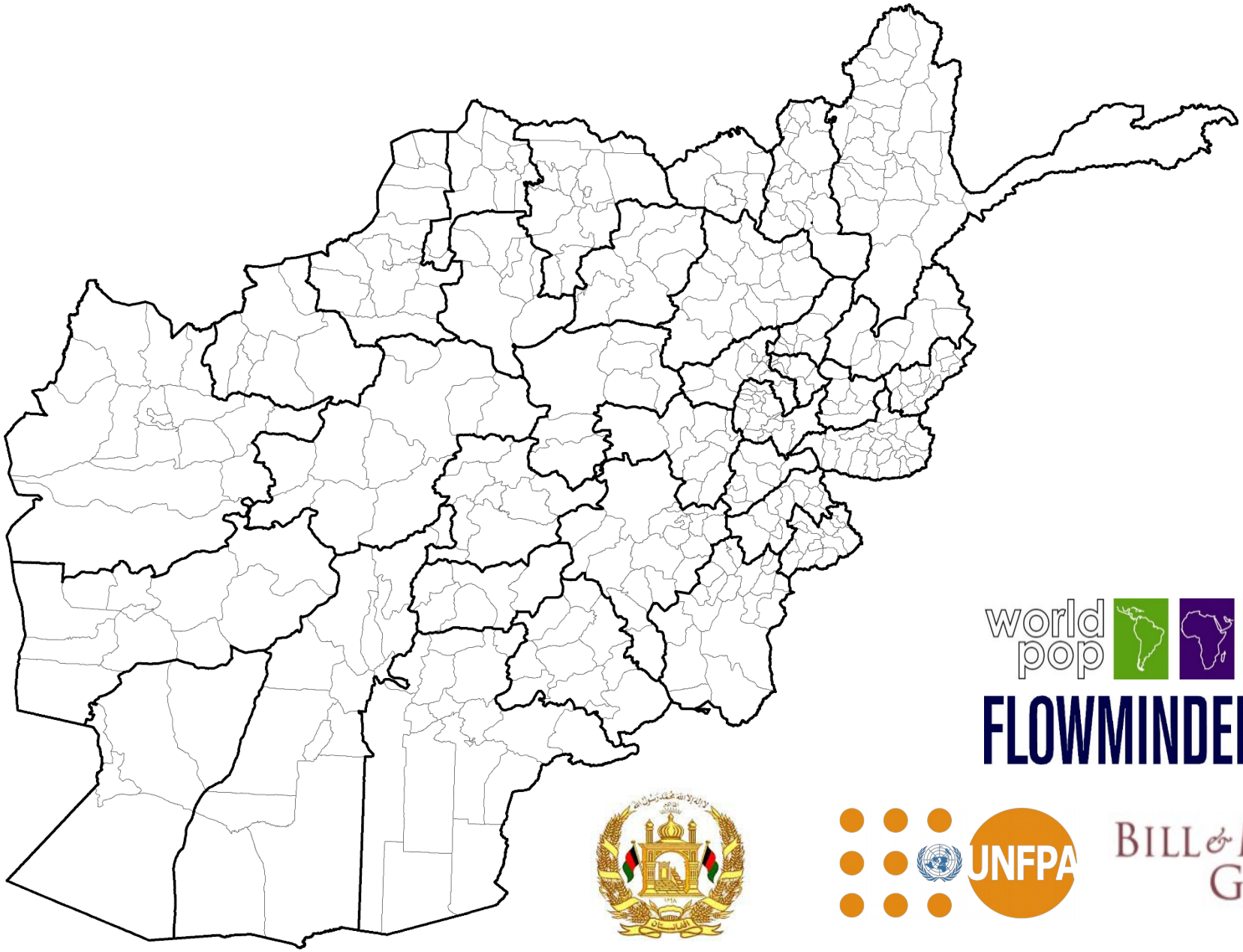


worldpop

FLOWMINDER.ORG

UNFPA

BILL & MELINDA GATES foundation

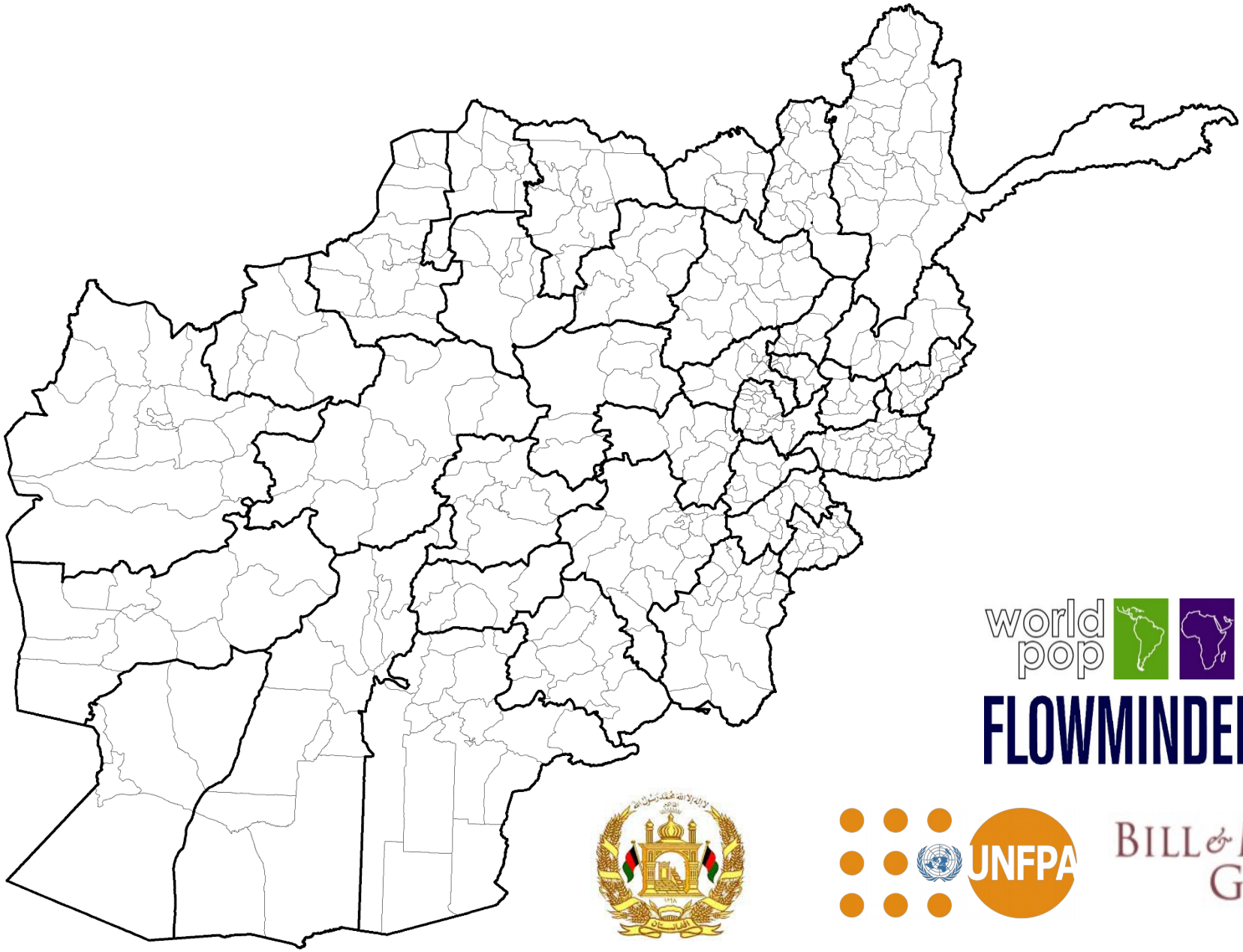


worldpop

FLOWMINDER.ORG

UNFPA

BILL & MELINDA GATES foundation

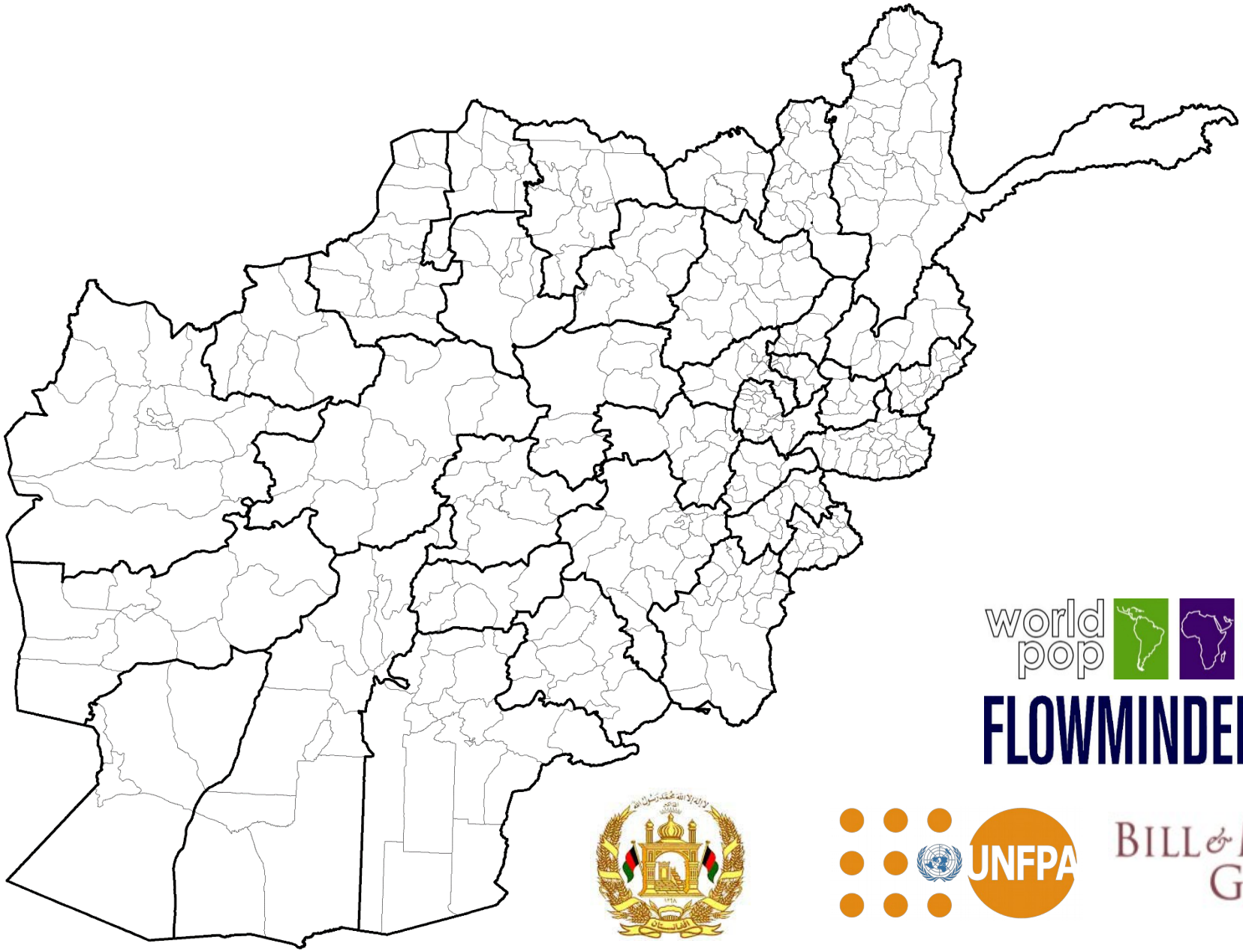


worldpop

FLOWMINDER.ORG

UNFPA

BILL & MELINDA GATES foundation

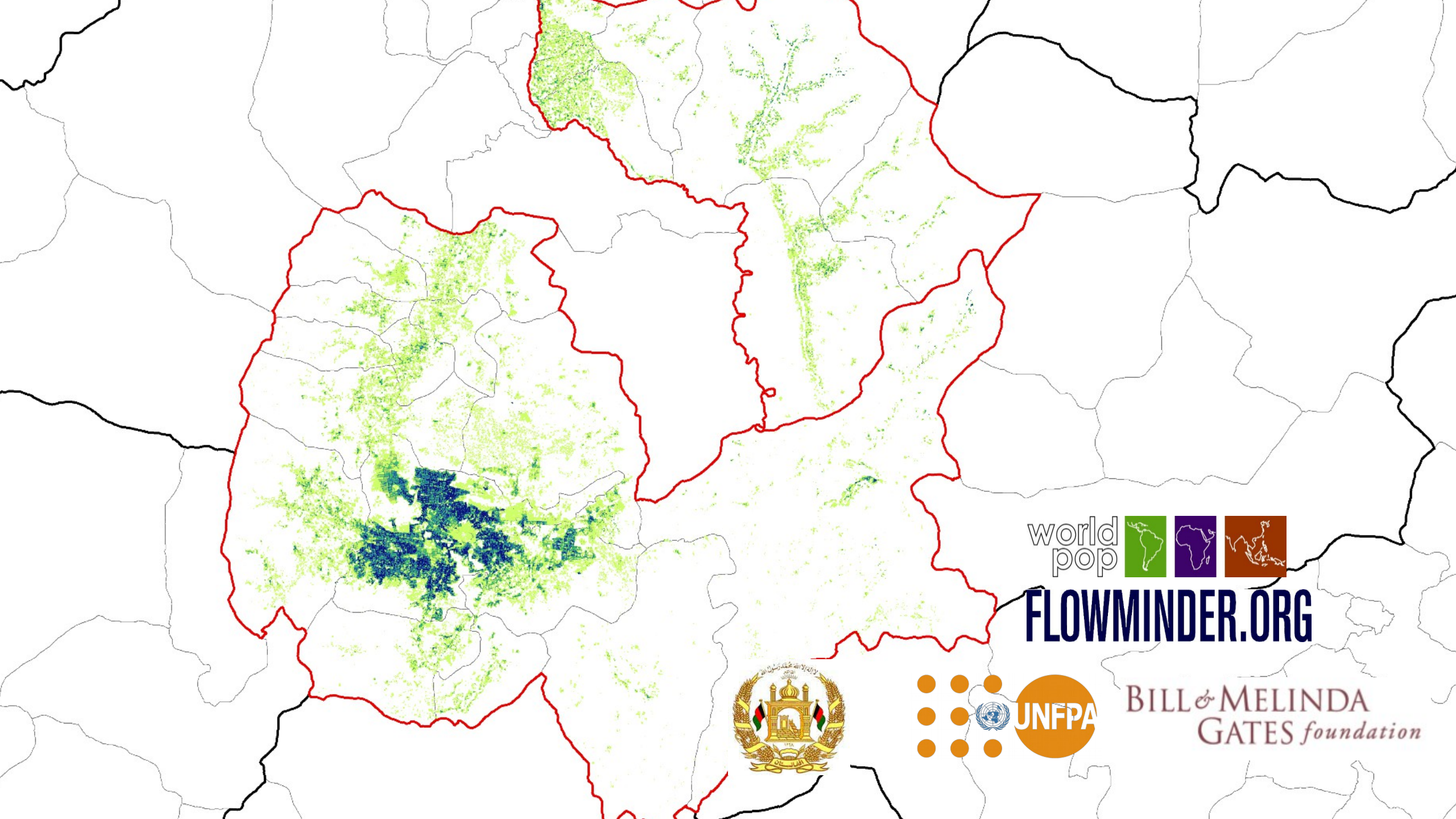


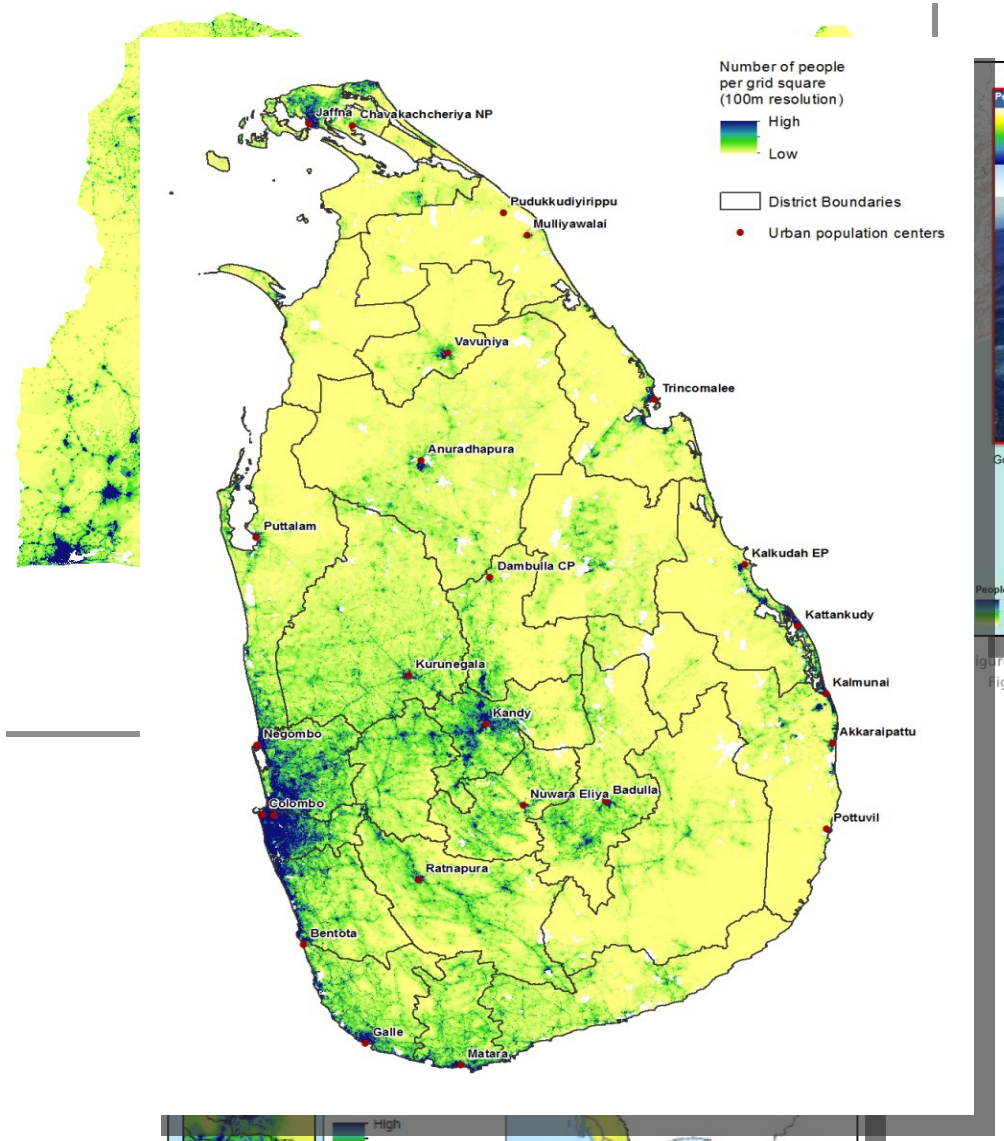
worldpop

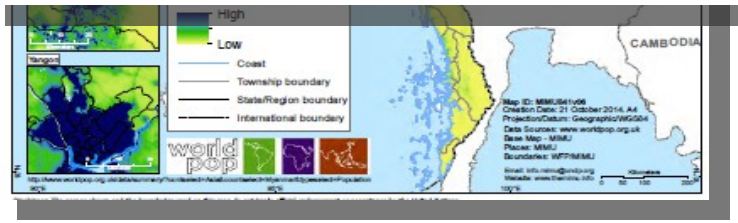
FLOWMINDER.ORG

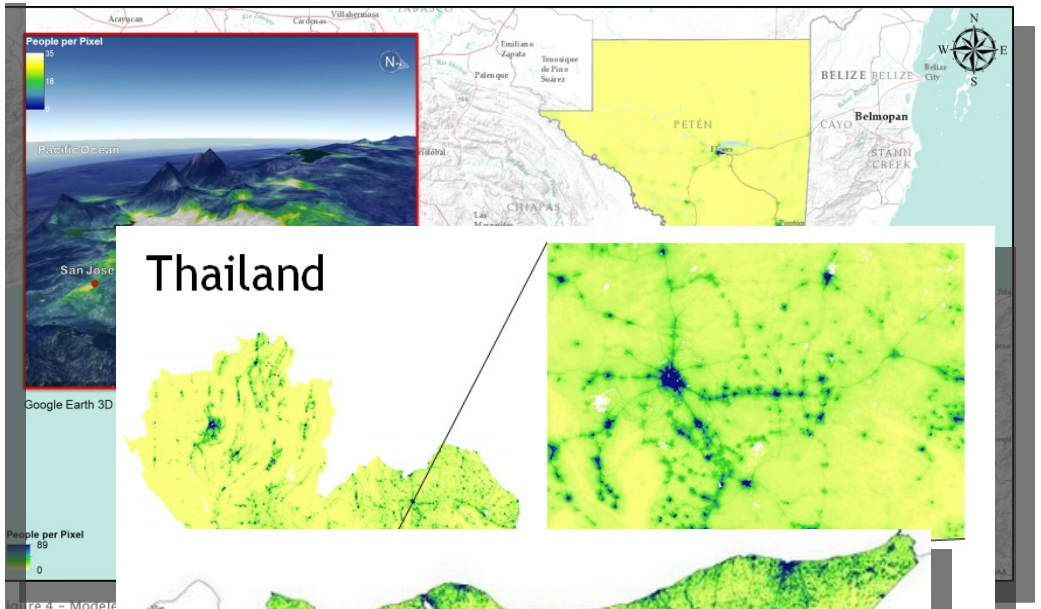
UNFPA

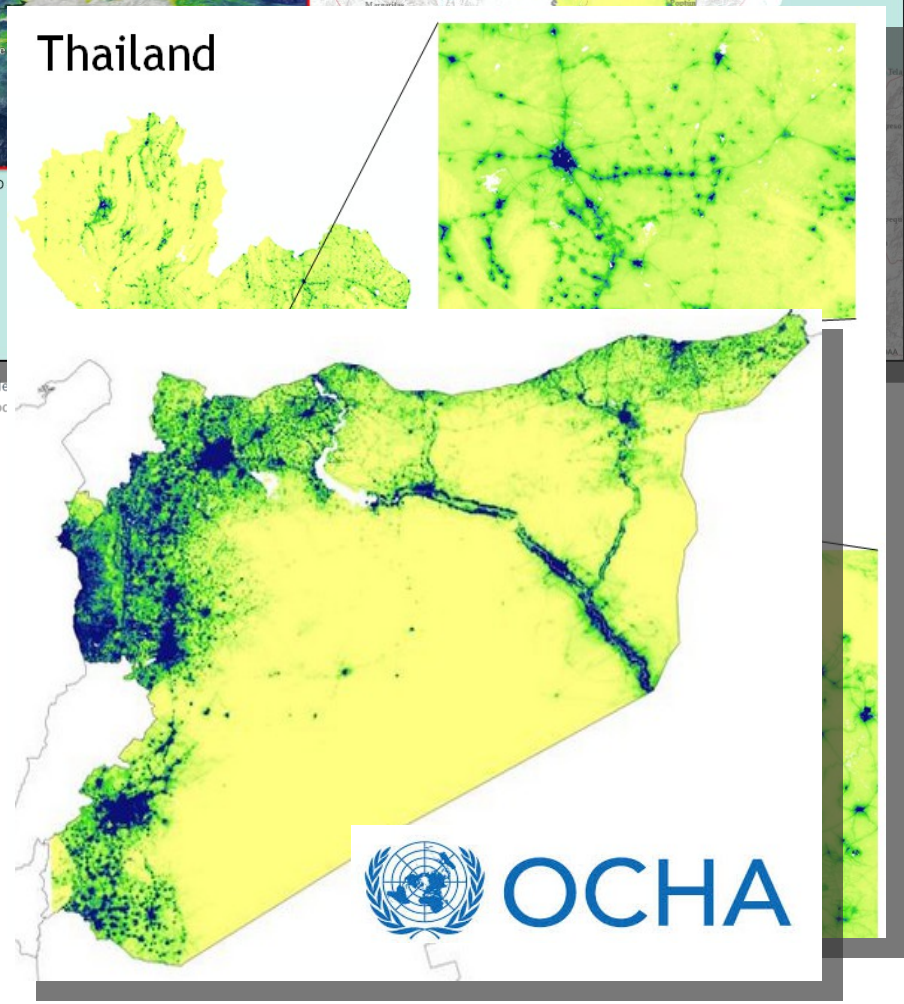
BILL & MELINDA GATES foundation


world pop
FLOWMINDER.ORG
UNFPA
BILL & MELINDA GATES foundation





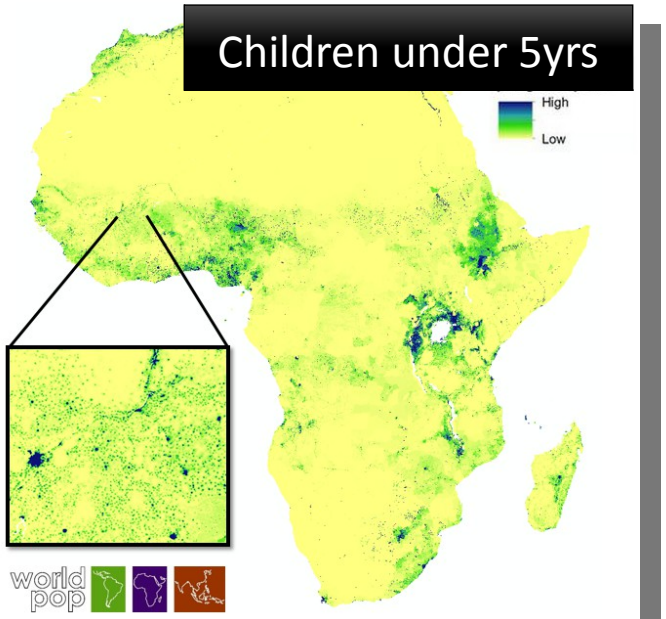


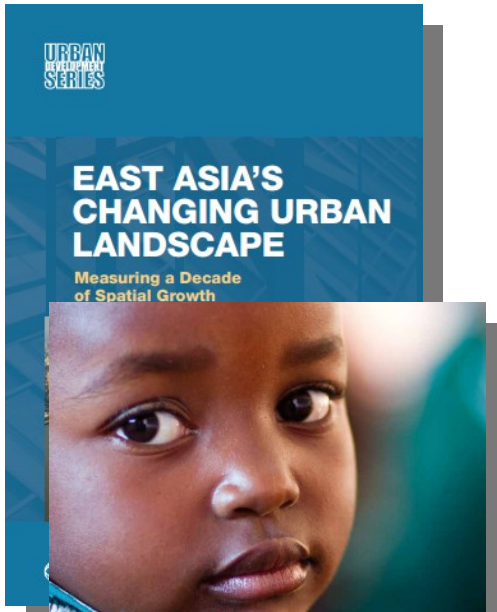


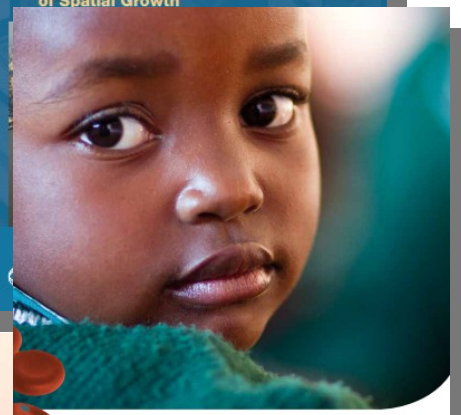
Refs: See www.worldpop.org

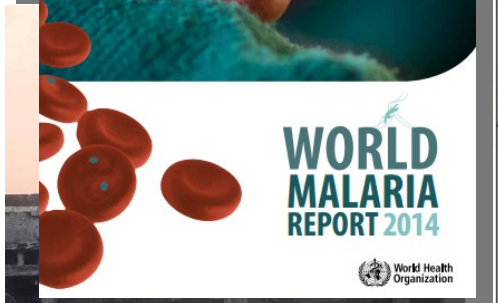
Datasets, Methods in use Operationally

Children under 5yrs

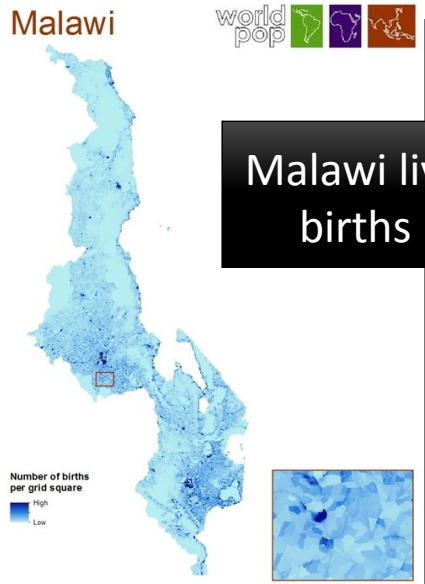


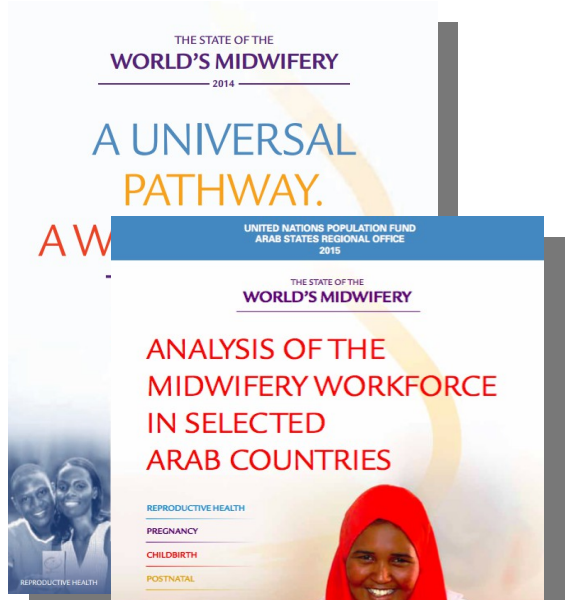




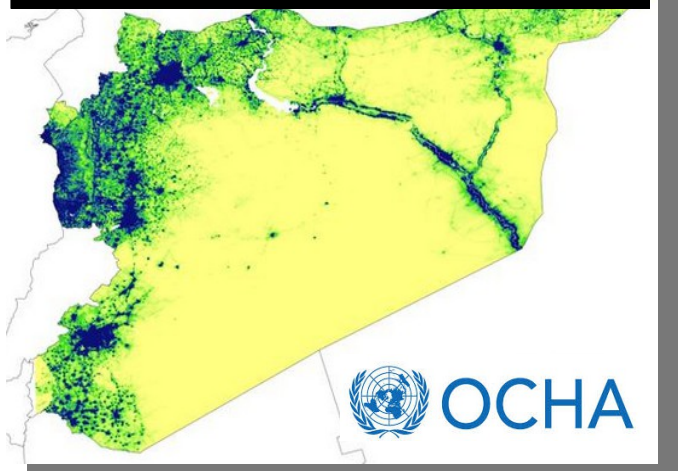


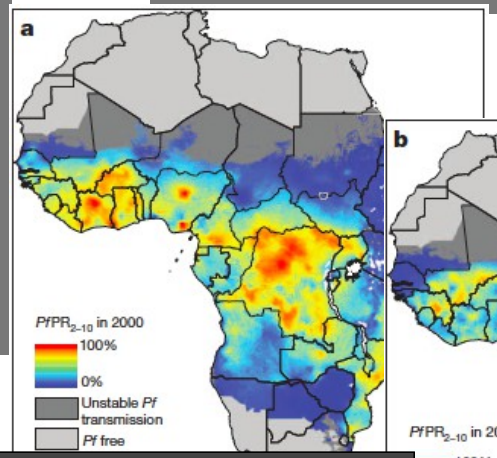
Malawi live births

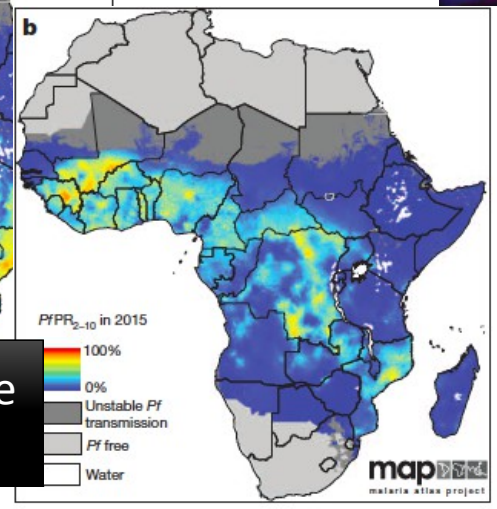




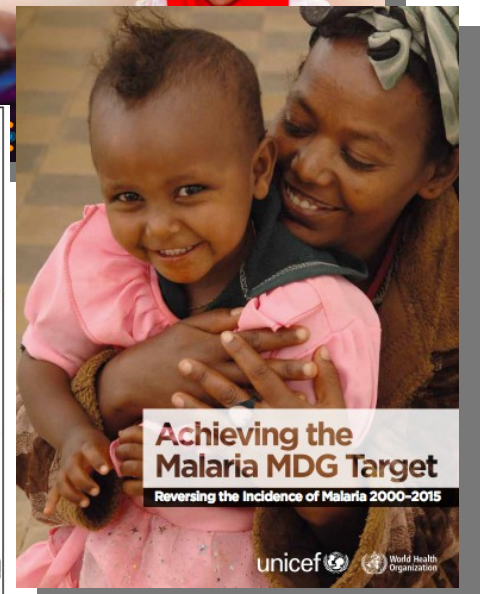
Syria population distribution



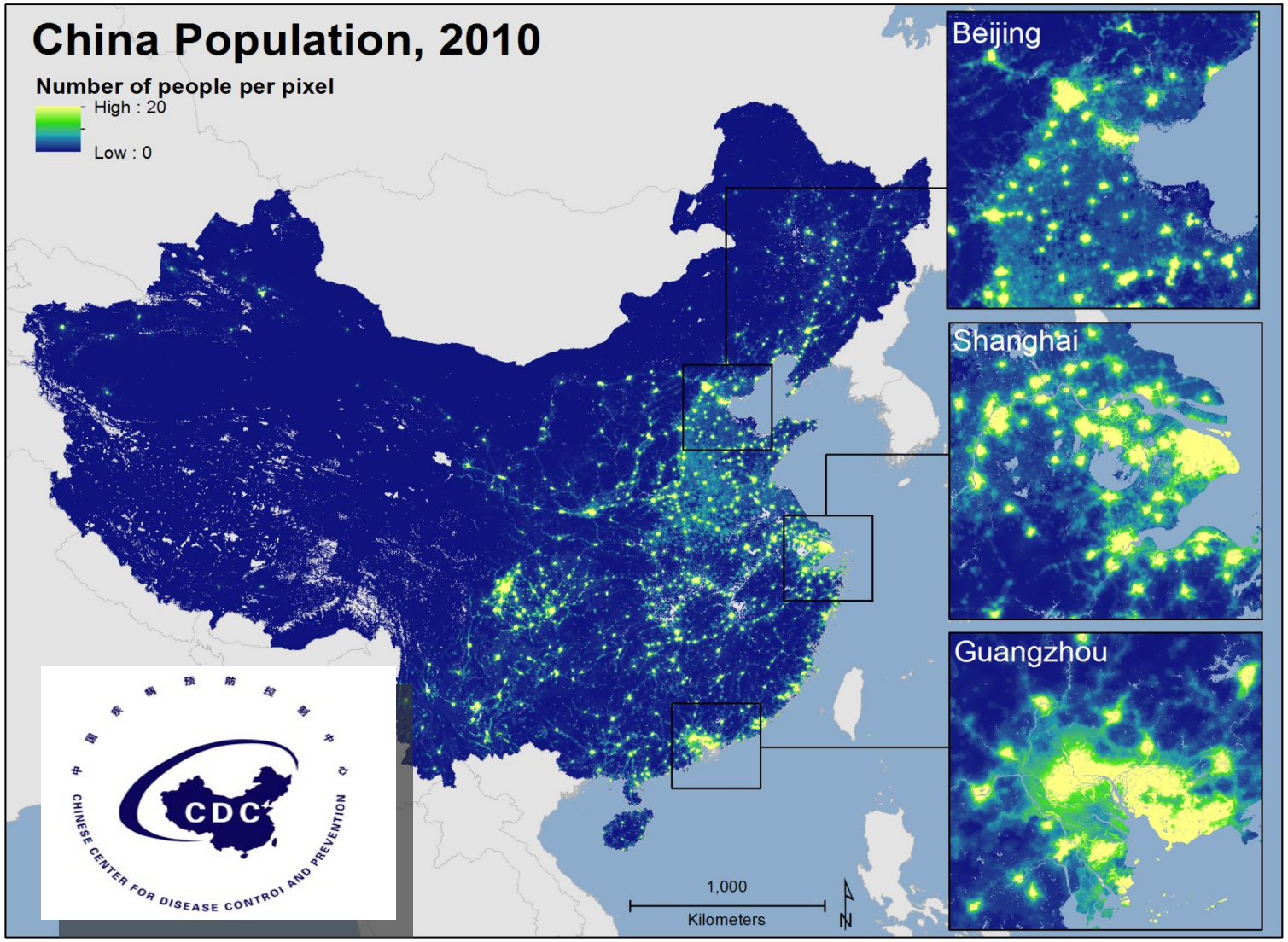




Malaria prevalence 2000-15

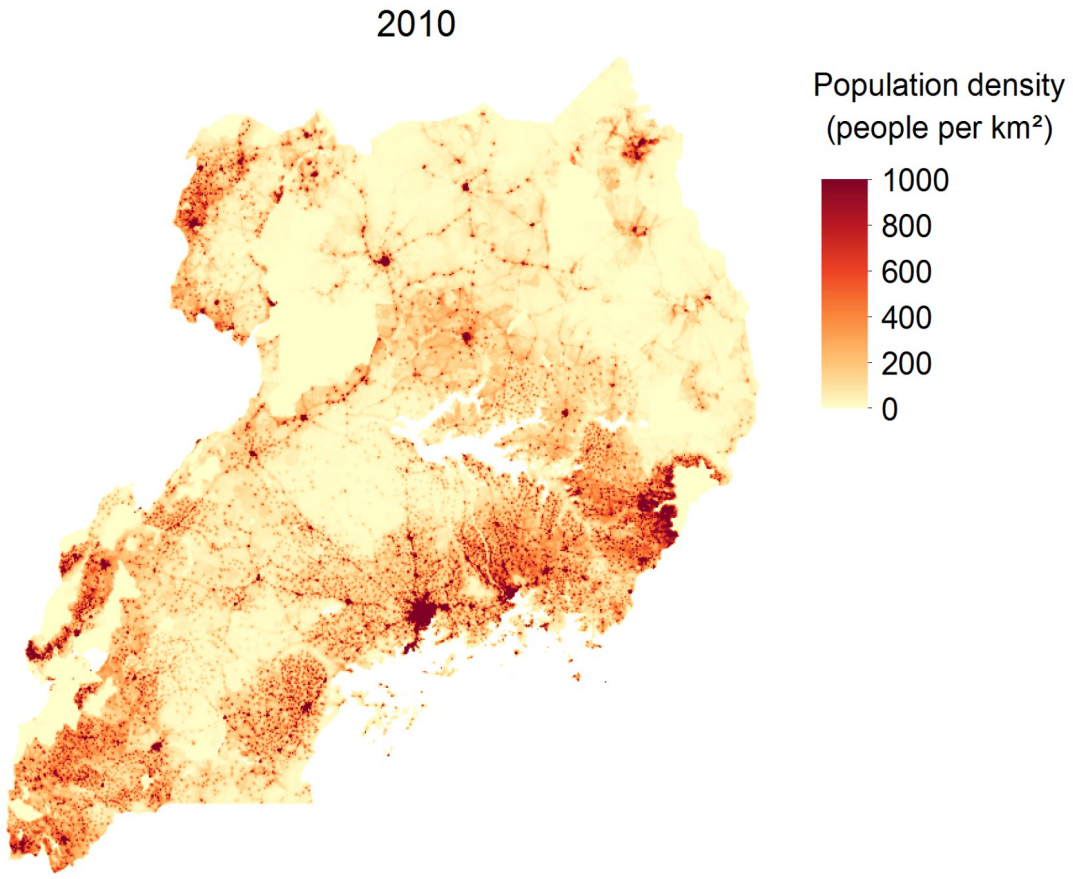


Temporal Change: Looking Back



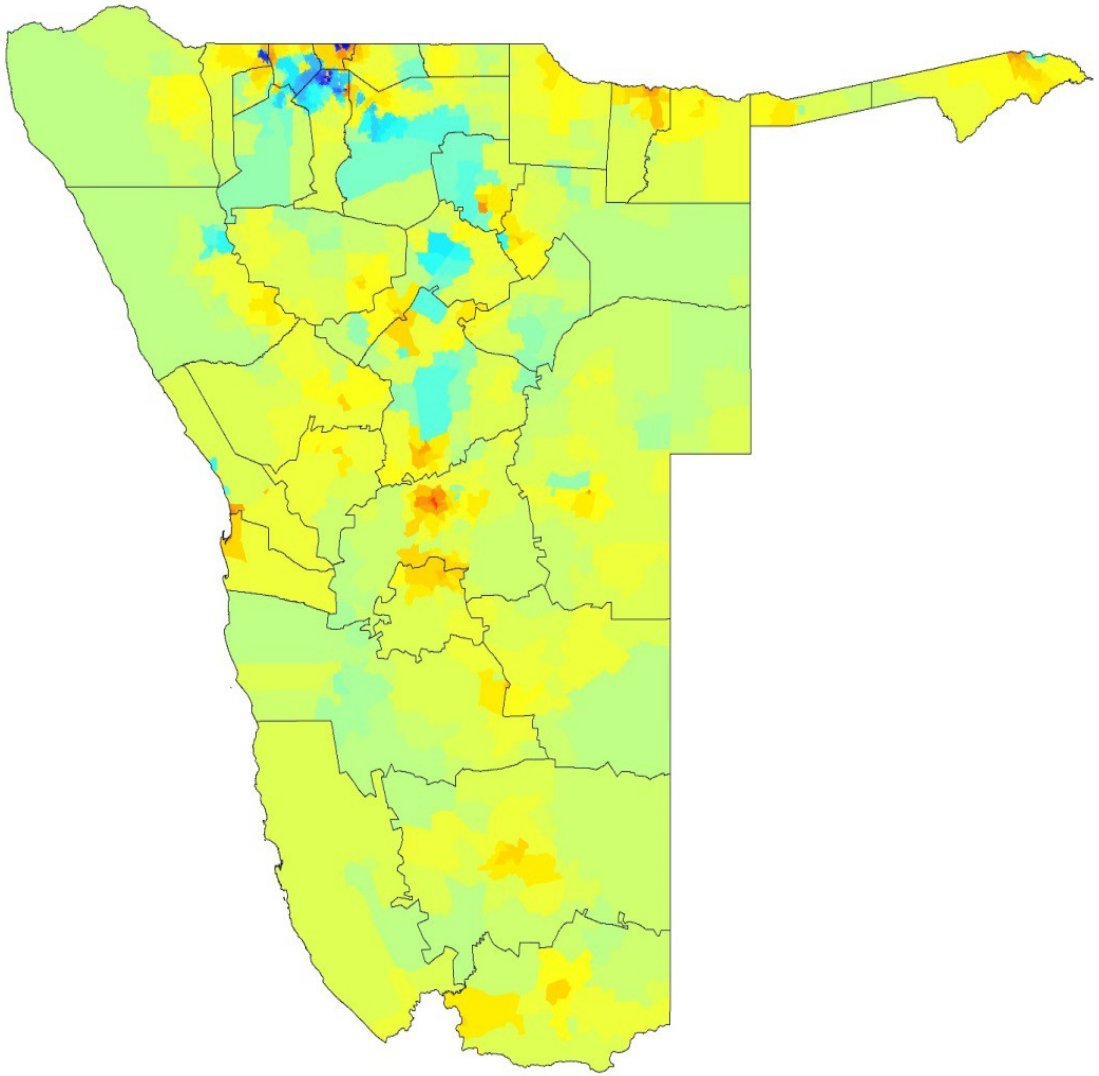
Ref: Gaughan et al. Nature Scientific Data. 2015

Temporal Change: Projecting Forward



Ref: Lindard et al. Applied Geography. 2013

Populations Don’t Stay Put...



Pop density change per square km

>5000

<-5000

Namibia Pop: 2.3 mill

MTC active subscriptions: 2.1 mill

NOV_12 DEC_12 JAN_13 FEB_13 MAR_13 APR_13 MAY_13 JUN_13 JUL_13 AUG_13 SEP_13

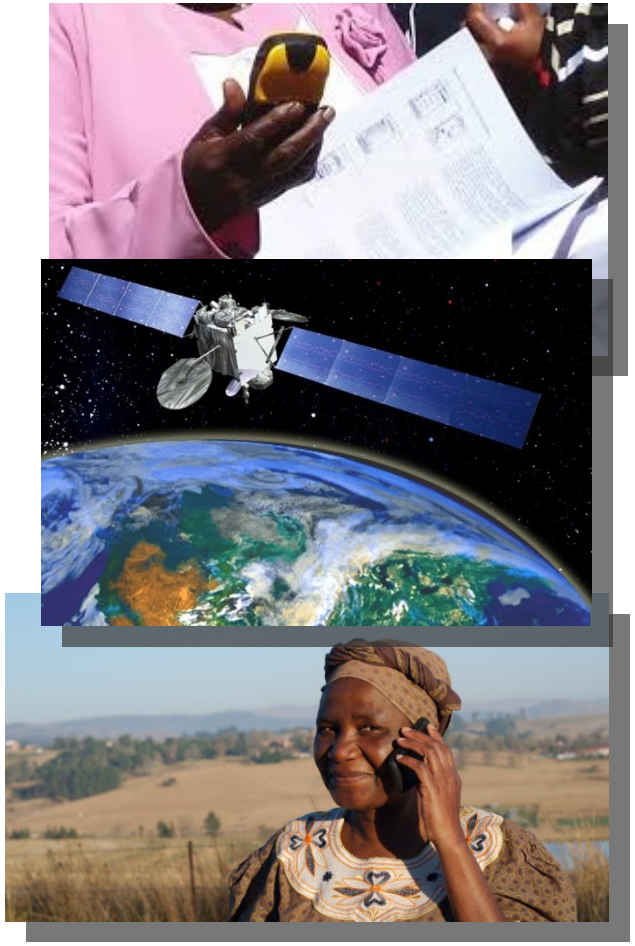
worldpop FLOWMINDER.ORG

Measuring Targets: Population Characteristics and Coverages

SDG targets

- 1.1. Eradicate extreme poverty for all people everywhere
- 2.2. End all forms of malnutrition, including the internationally agreed targets on stunting and wasting in children under 5 years of age
- 3.7. Ensure universal access to sexual and reproductive health-care services
- 4.6. Ensure that all youth and a substantial proportion of adults, both men and women, achieve literacy
- 6.2. Achieve access to adequate and equitable sanitation and hygiene for all

What can we use?



Spatial Data Integration

- Population characteristics measured in household surveys can be strongly related to features we can measure everywhere
- We can use these relationships to predict characteristics into unsampled locations using metrics from census, satellite and cellphone data to create maps of SDG-relevant indicators
- Importance of validation and the measurement and mapping of uncertainty

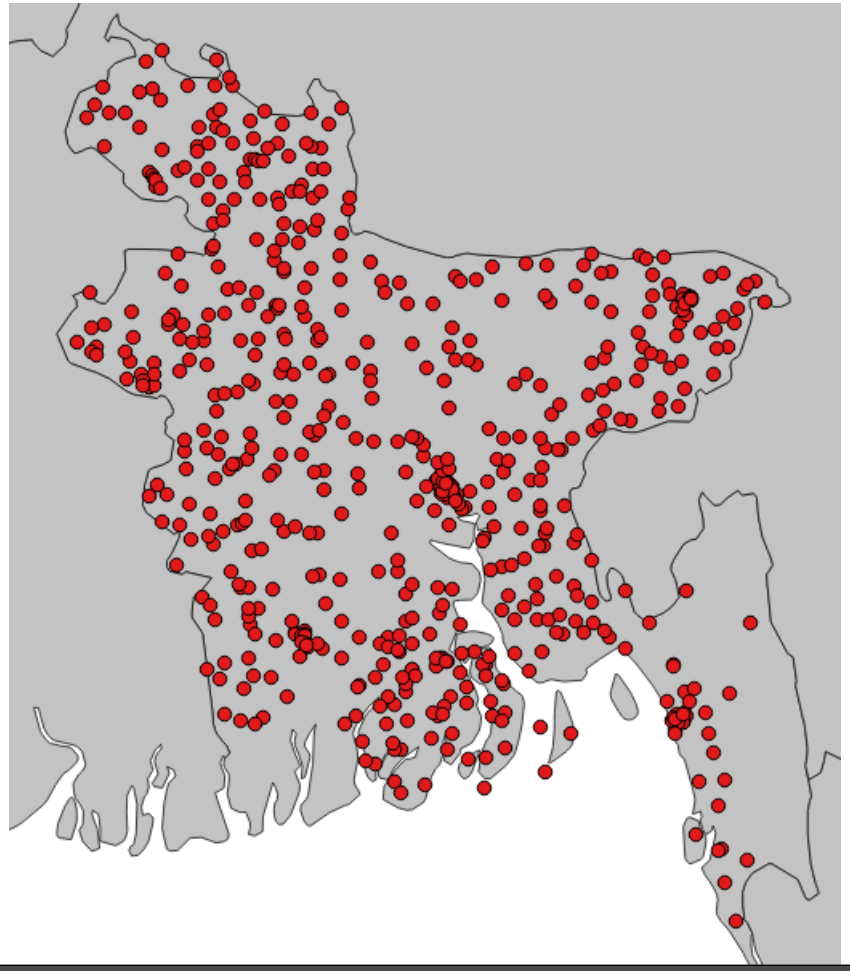
-Increasing distance from major roads = increasing poverty

-Increasing urbanicity = decreasing poverty

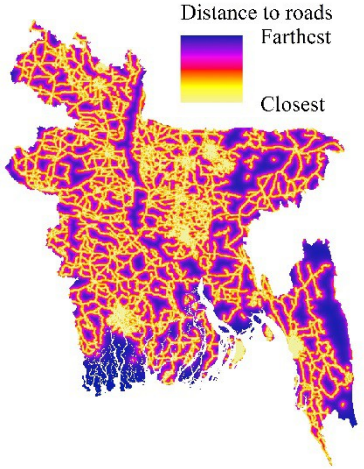
-Wider social network = lower poverty

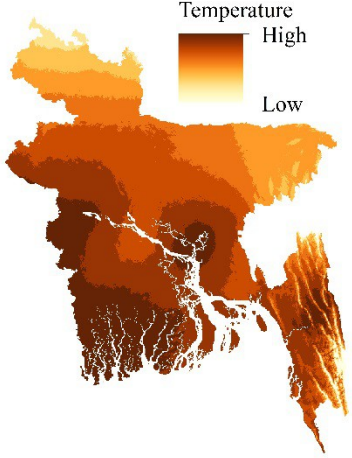
-Large, regular credit top up = lower poverty

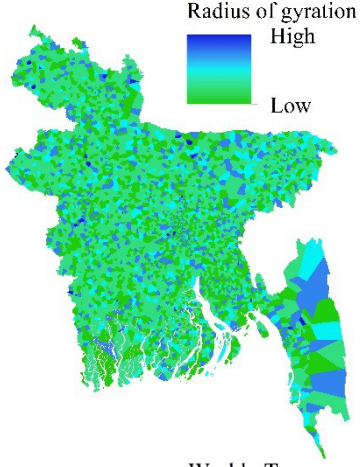
Improving Mapping of Socioeconomic Indicators

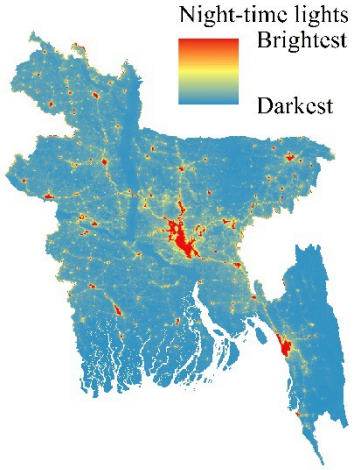


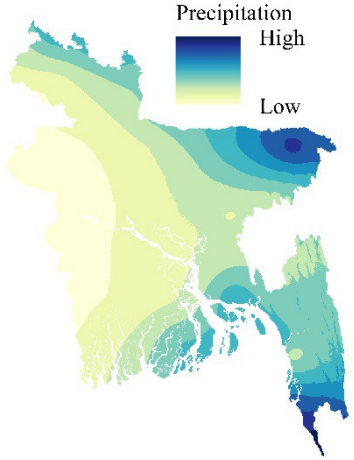
GPS-located survey cluster data

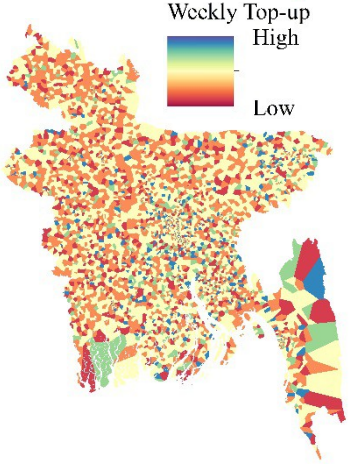


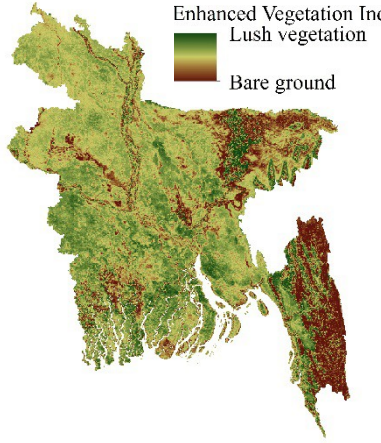


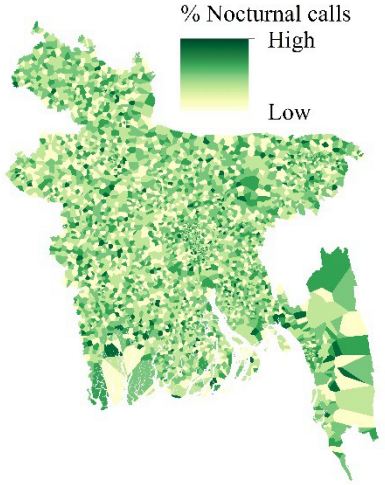


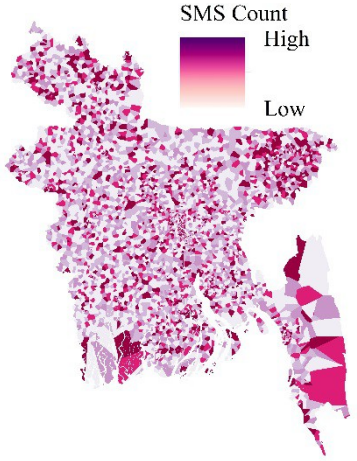




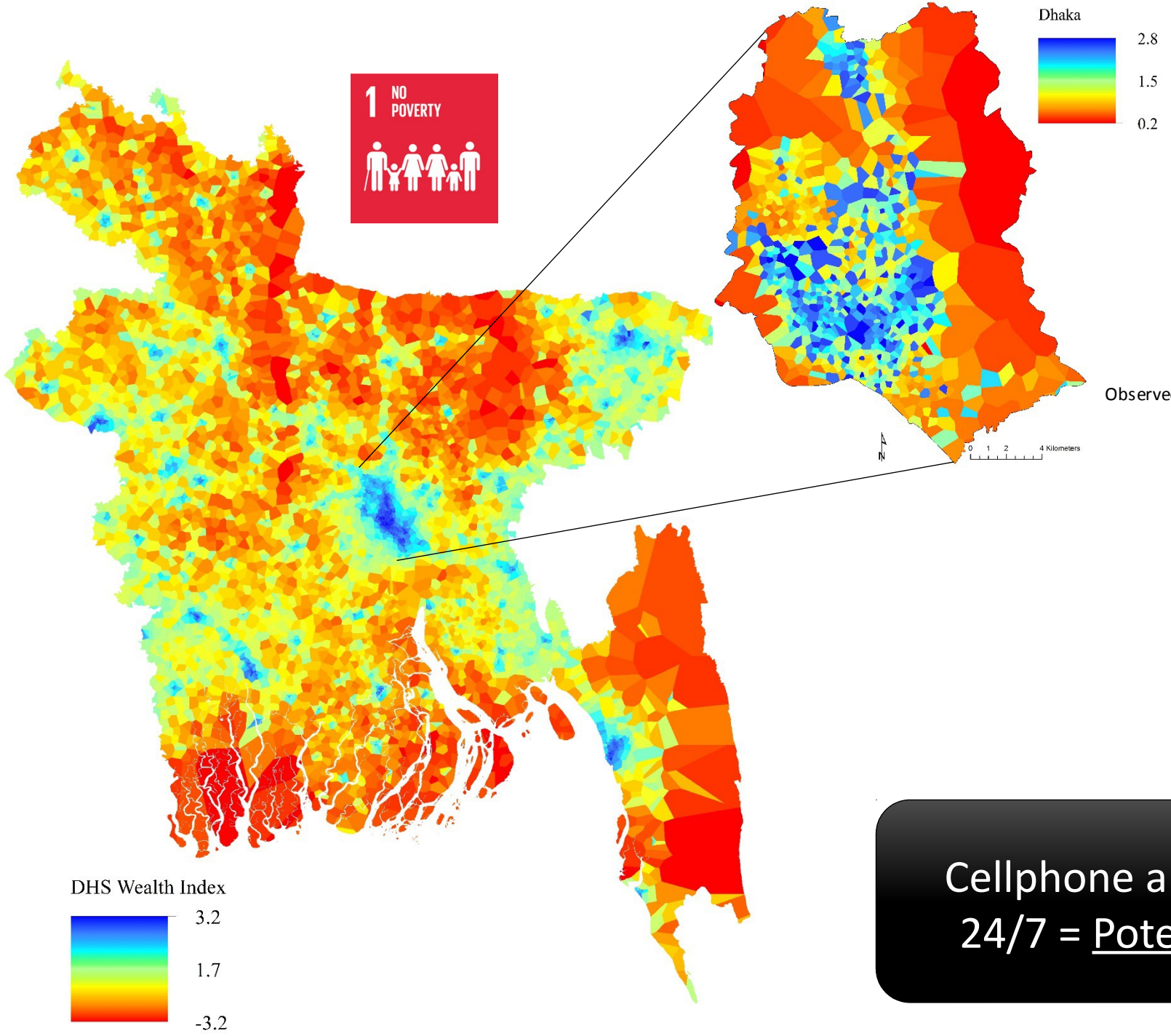


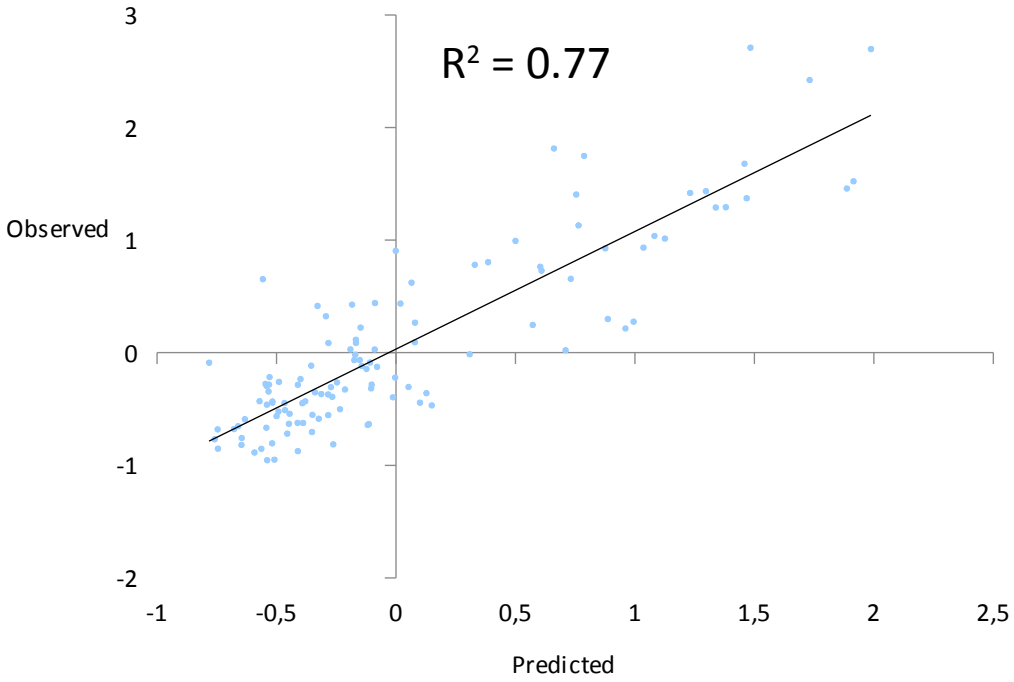






Example Output: Bangladesh wealth index

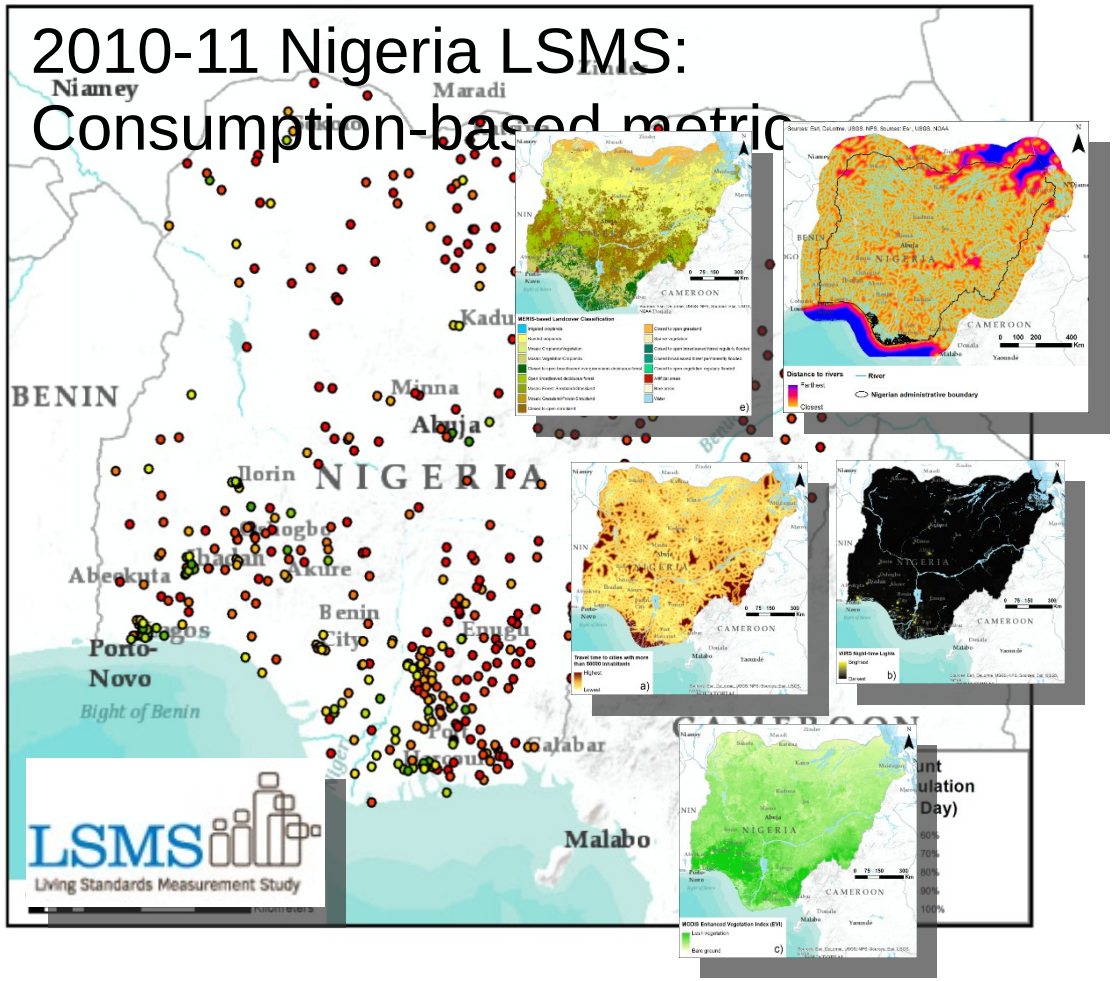


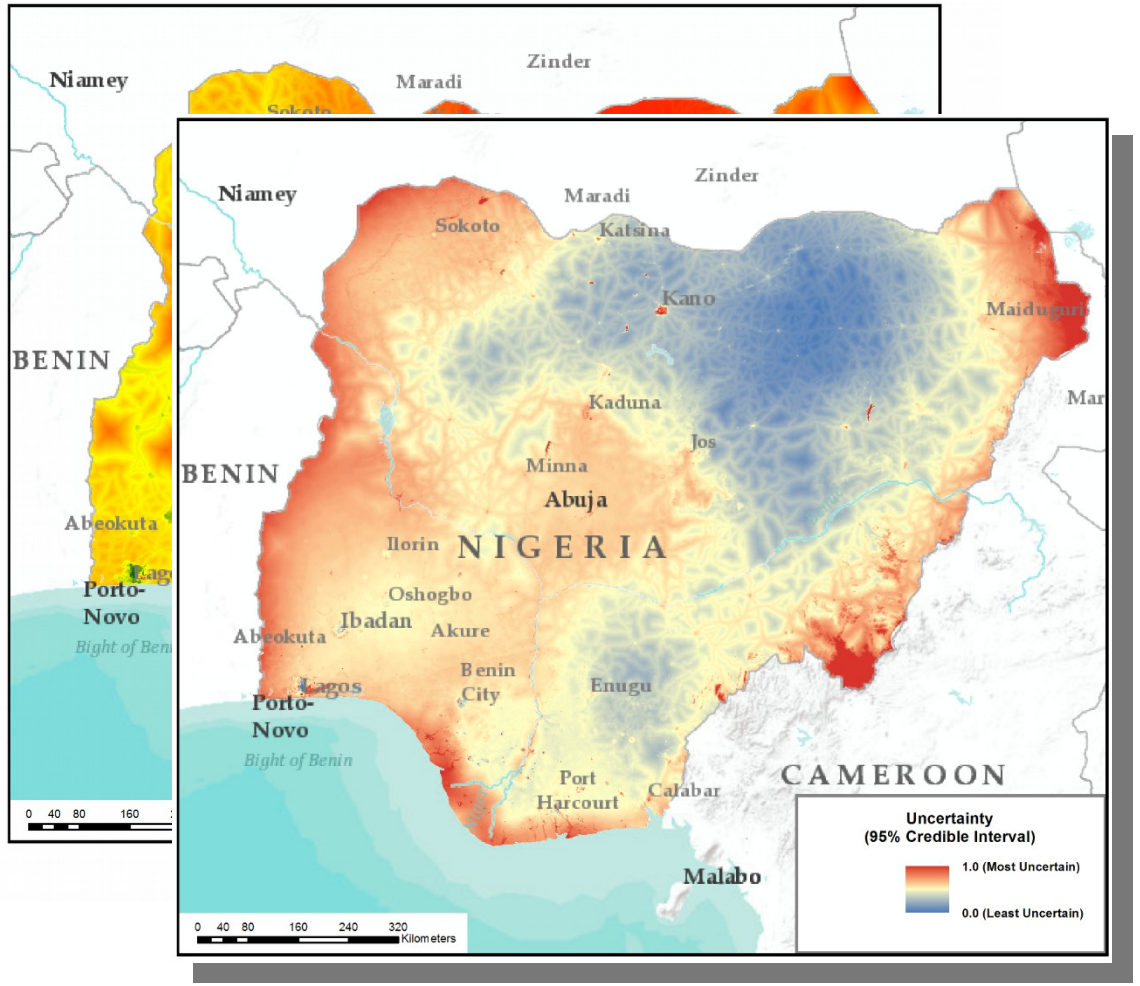


Cellphone and satellite data are collected 24/7 = <u>Potential for ongoing monitoring</u>

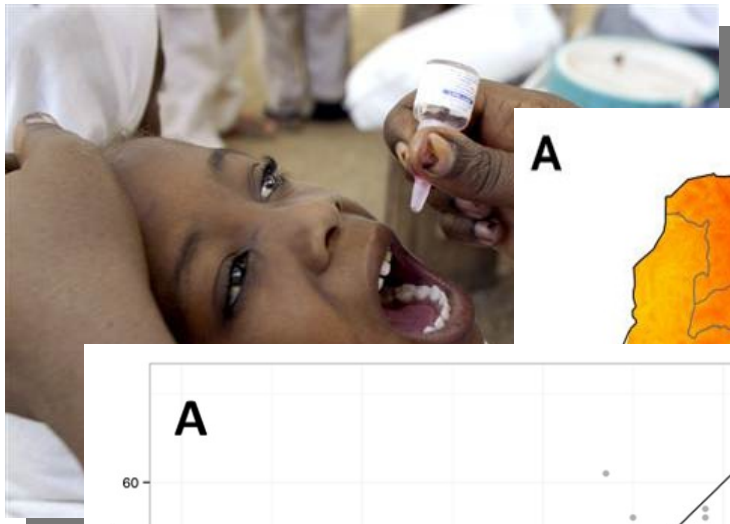
High-Resolution Poverty Maps

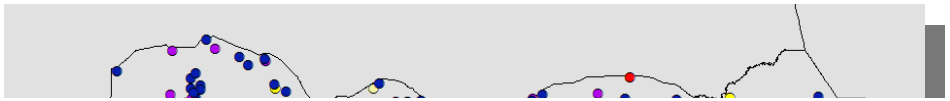
2010-11 Nigeria LSMS:
Consumption-based metric

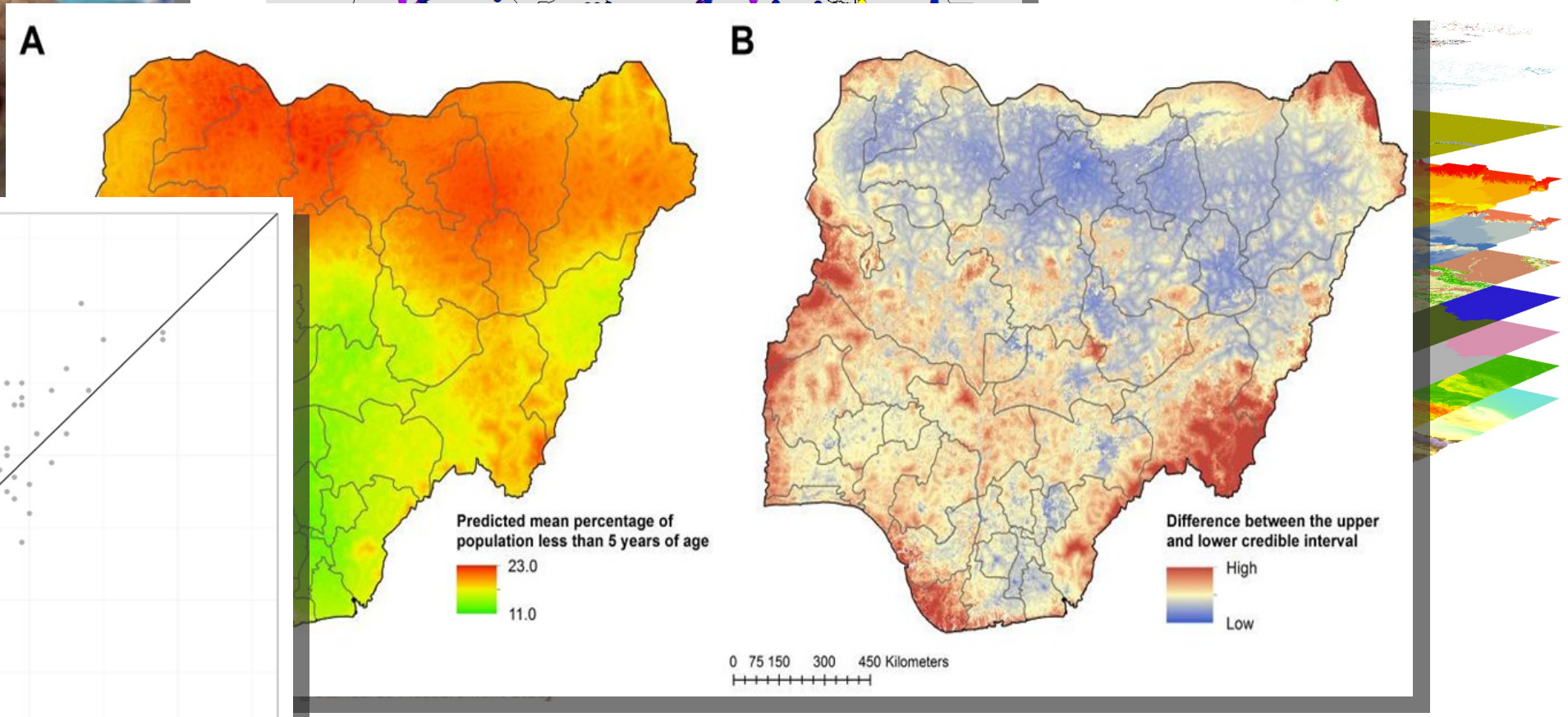


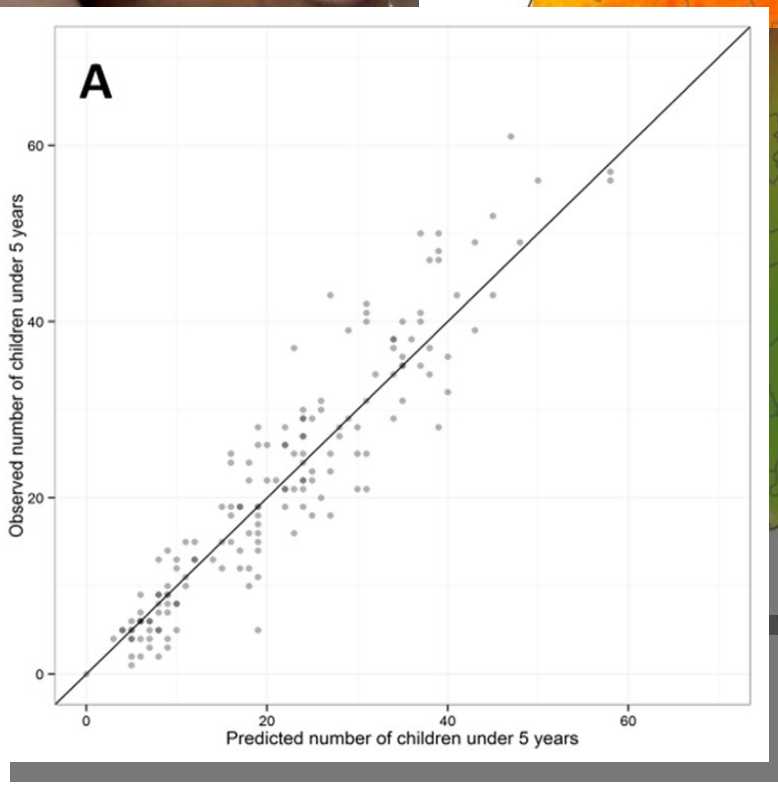


Polio Vaccination in Nigeria: Where are the <5s?

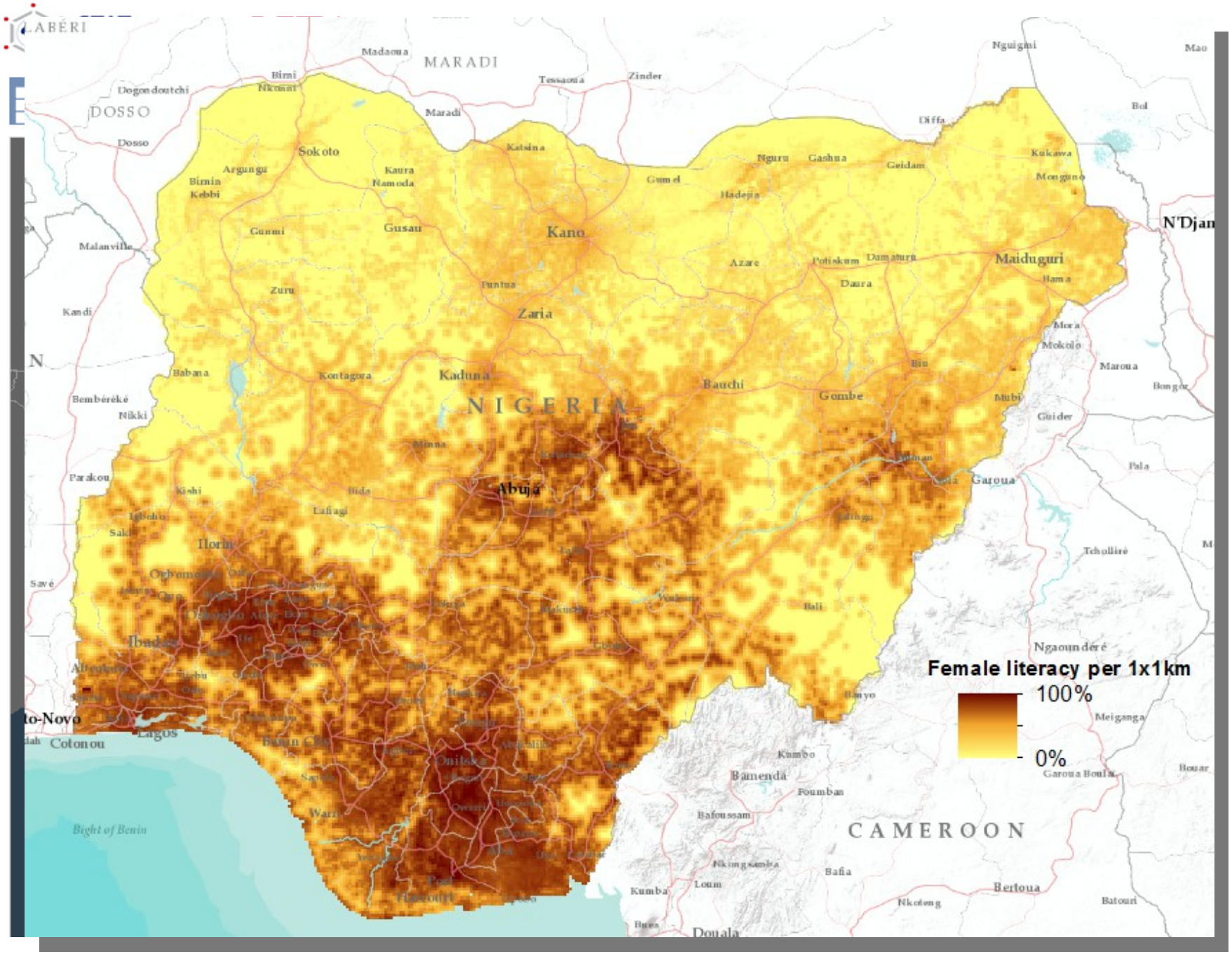




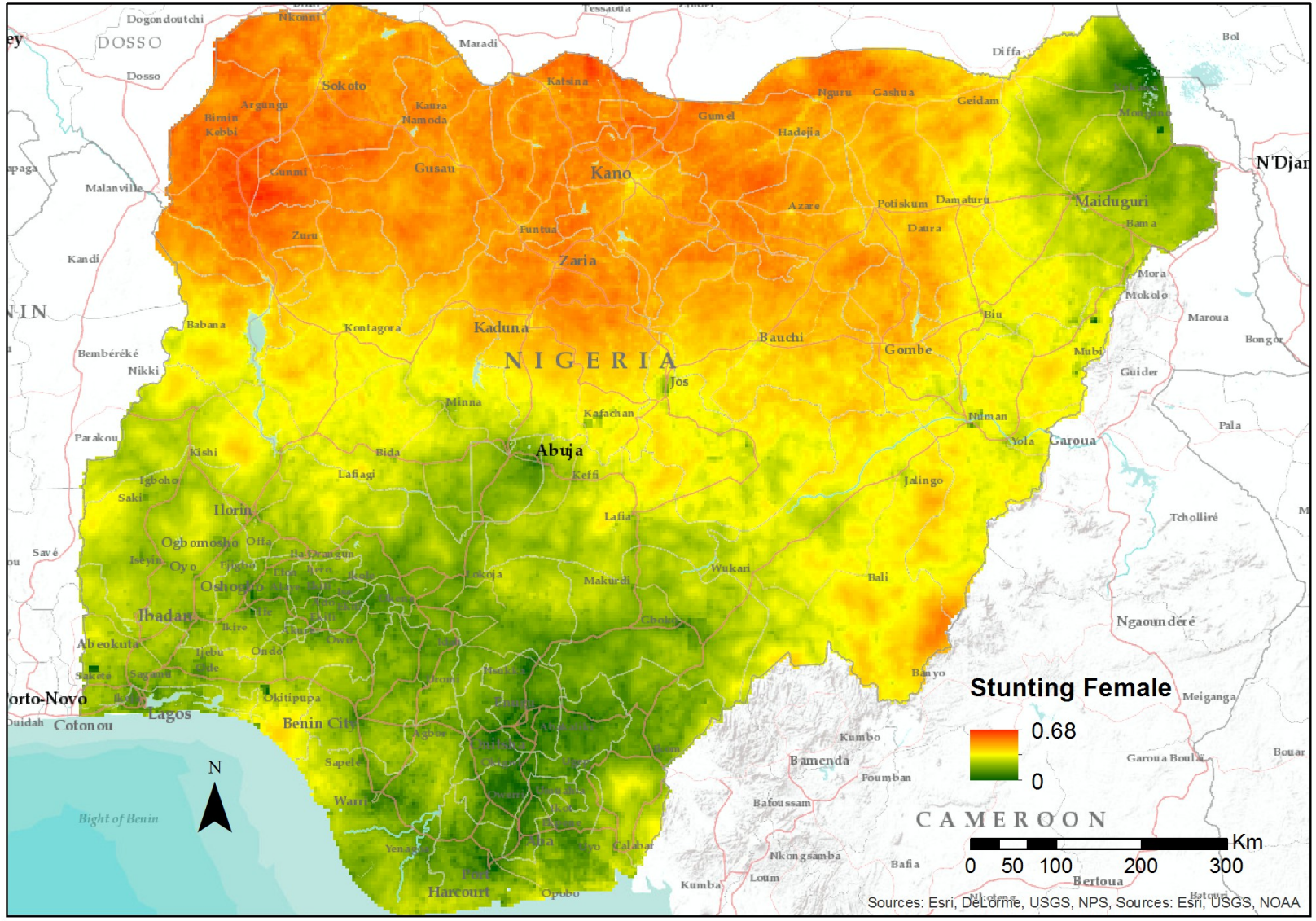




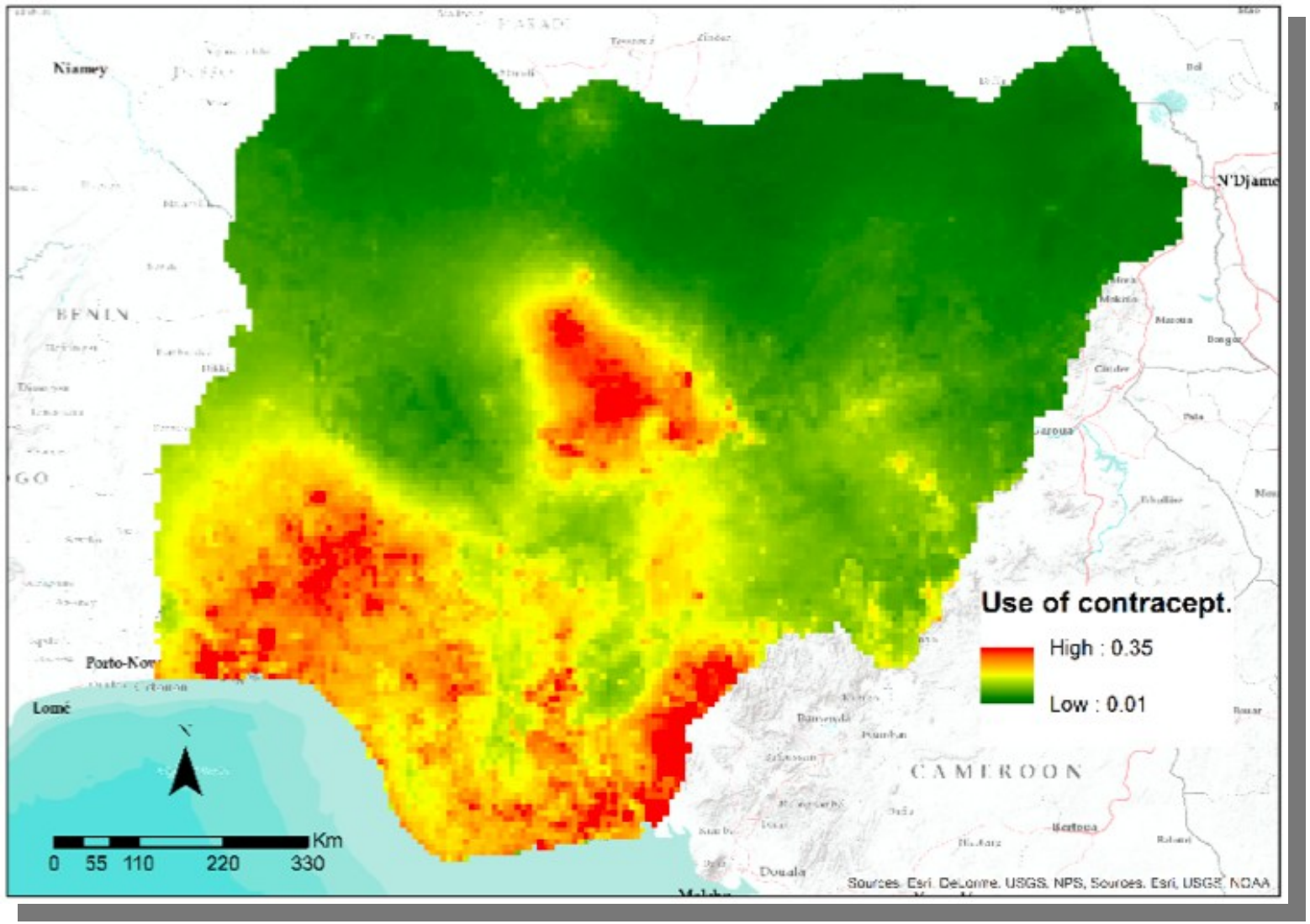
Nigeria: Proportion females who are literate



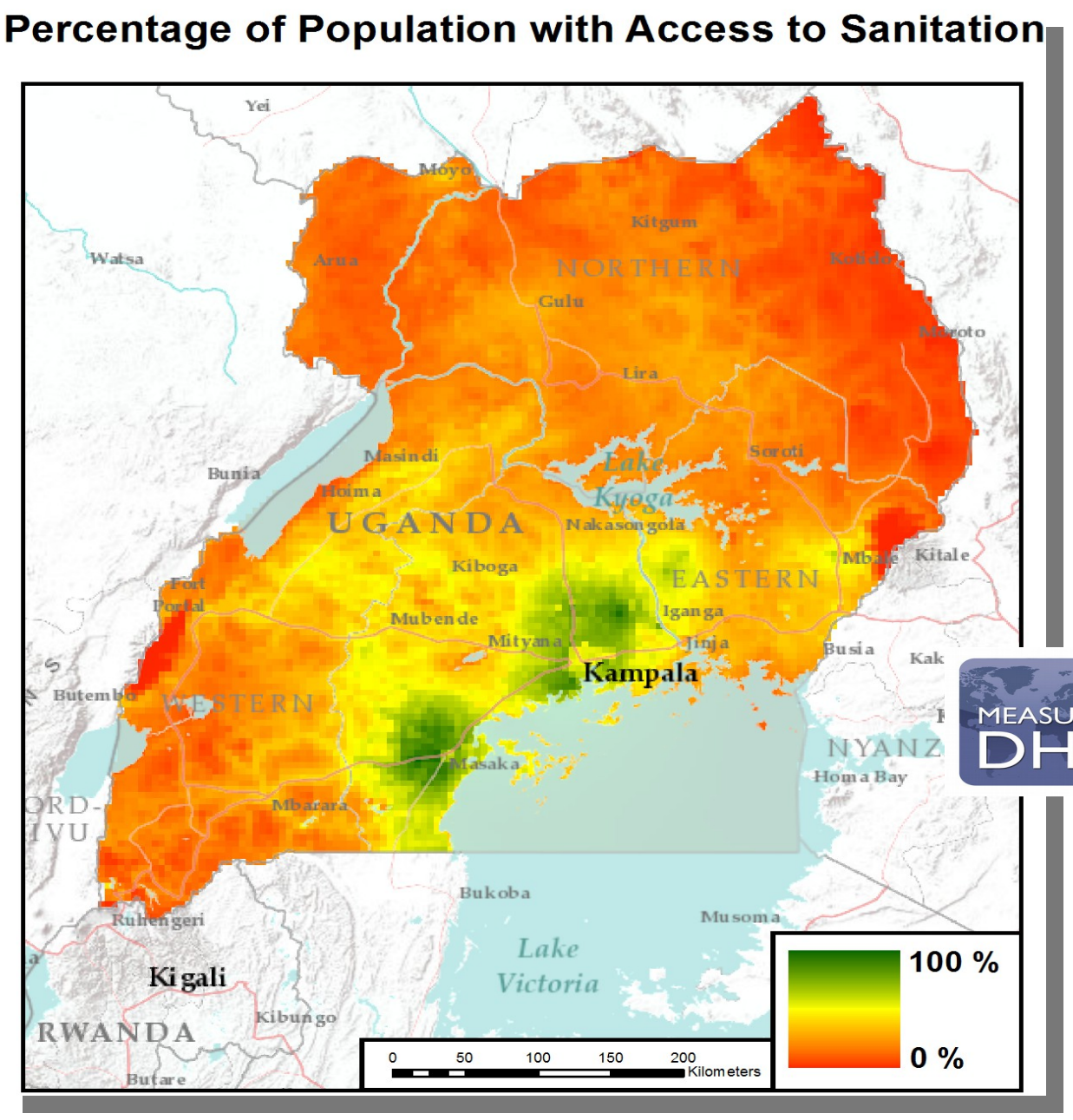
Stunting among girls Nigera



% women of reproductive age using modern contraceptive methods

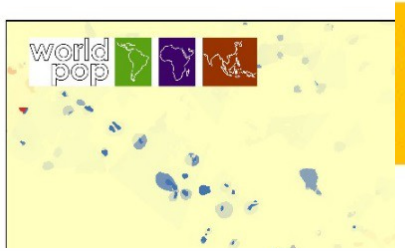


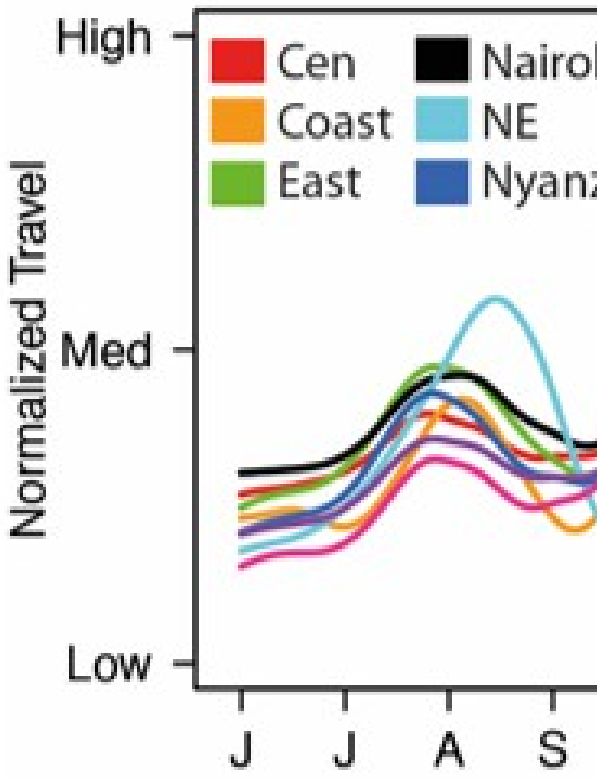
Uganda: access to sanitation



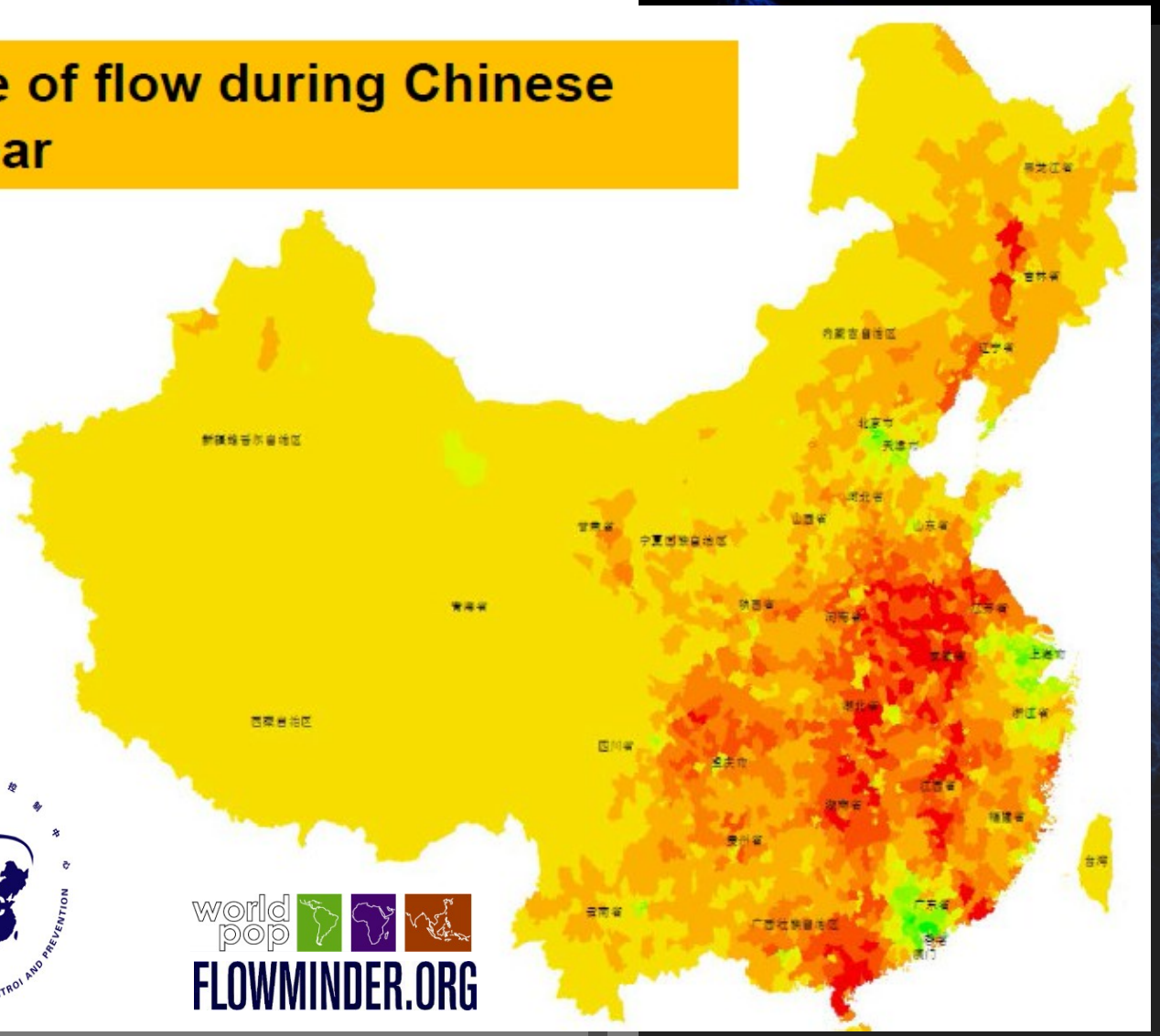
Mapping Displacement and Movements due to Climate Change and Disasters

Mapping Population Movements





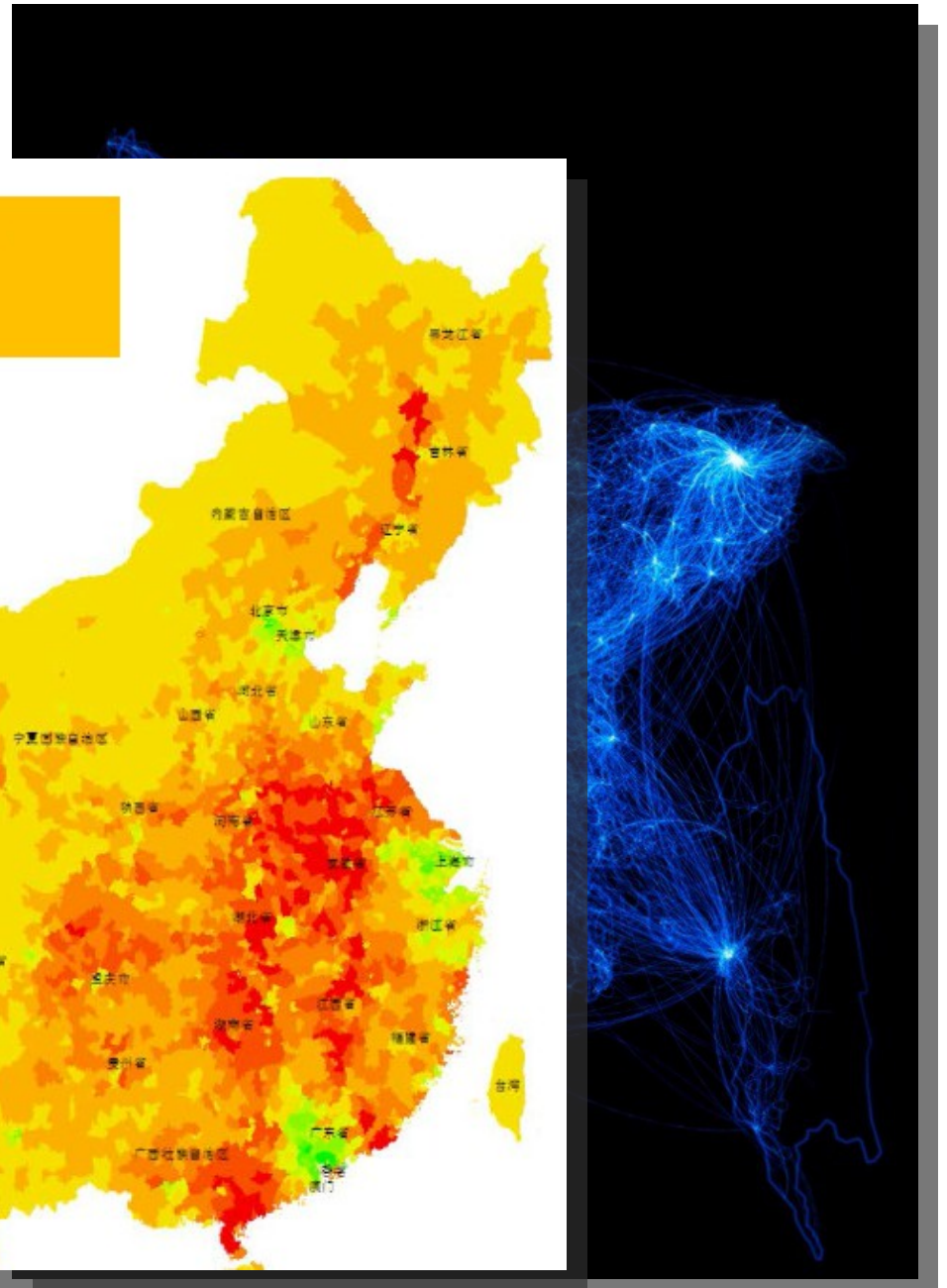
Change of flow during Chinese New Year



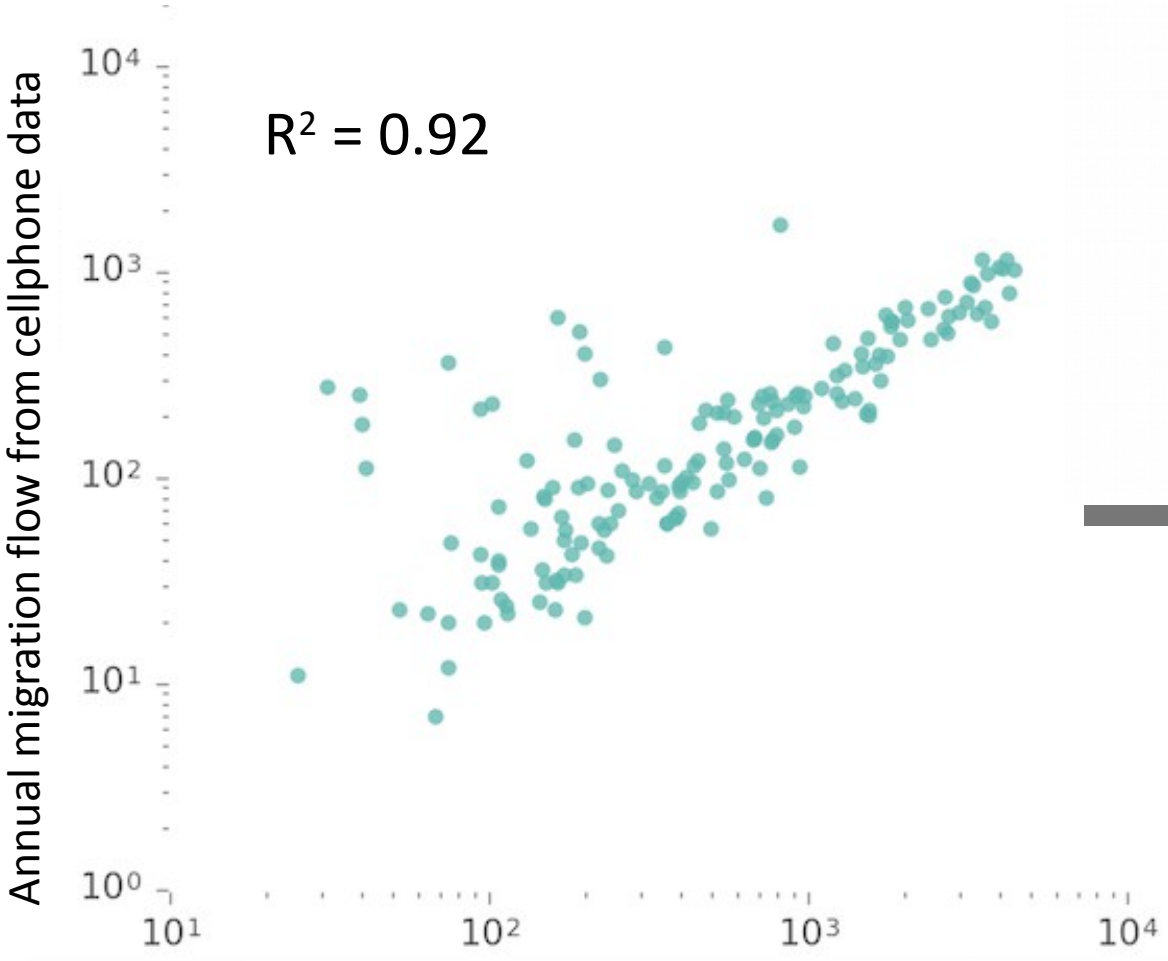
CHINESE CENTER FOR DISEASE CONTROL AND PREVENTION

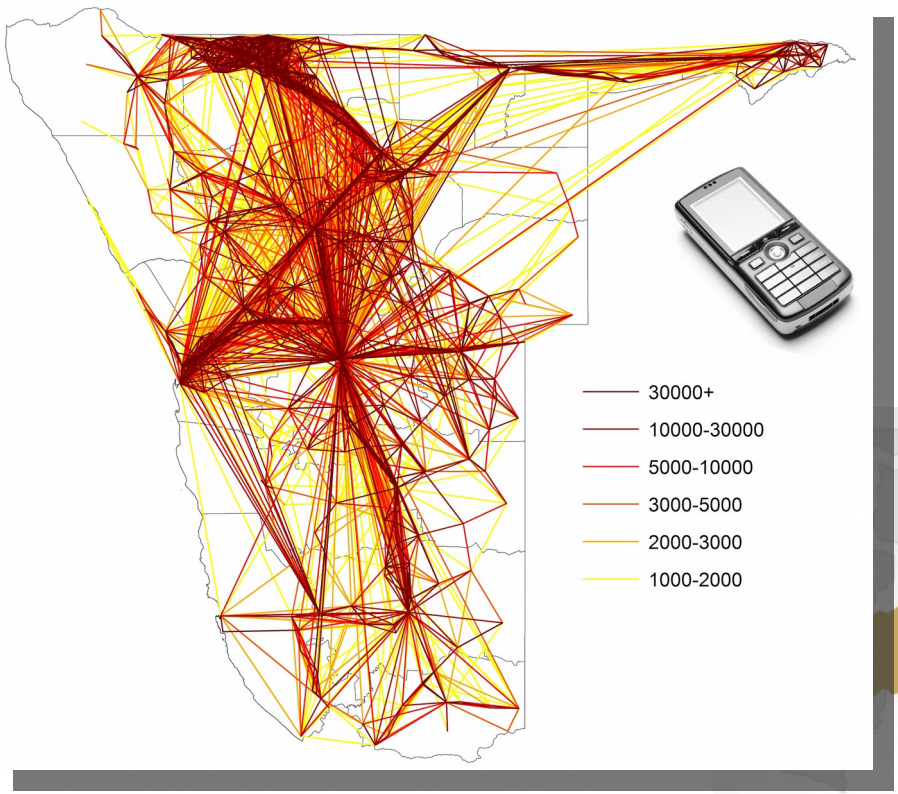
world pop

FLOWMINDER.ORG

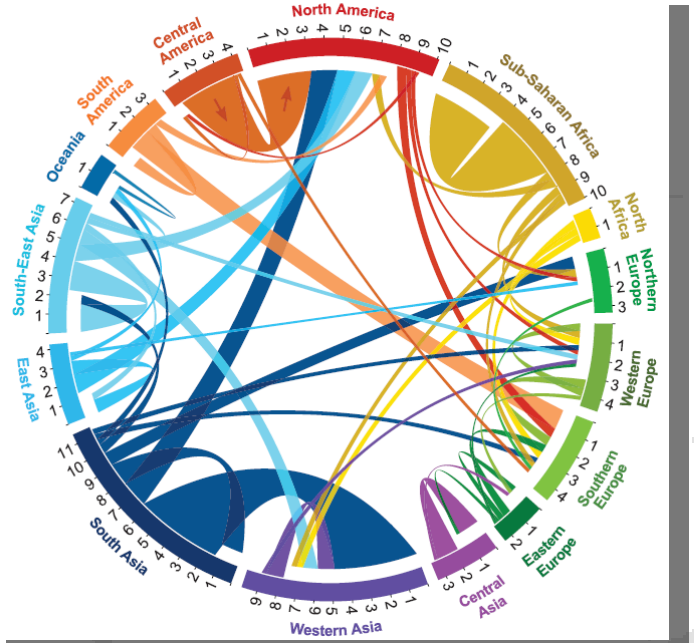


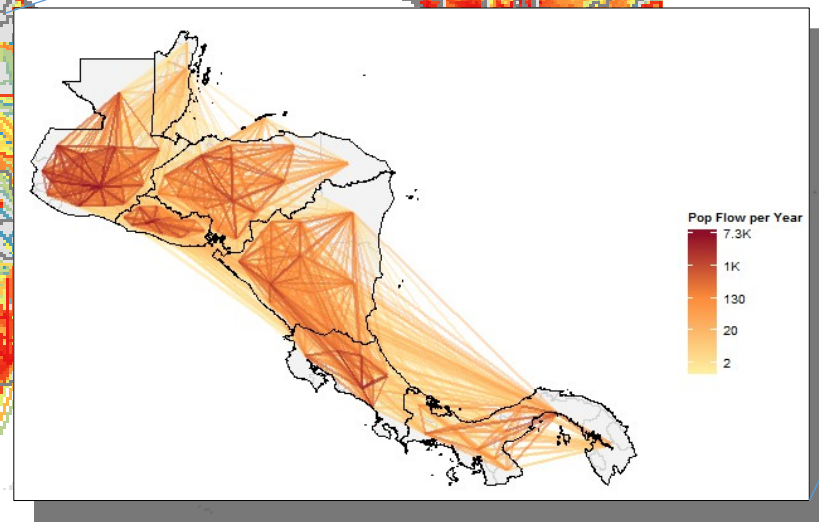
Measuring Migration

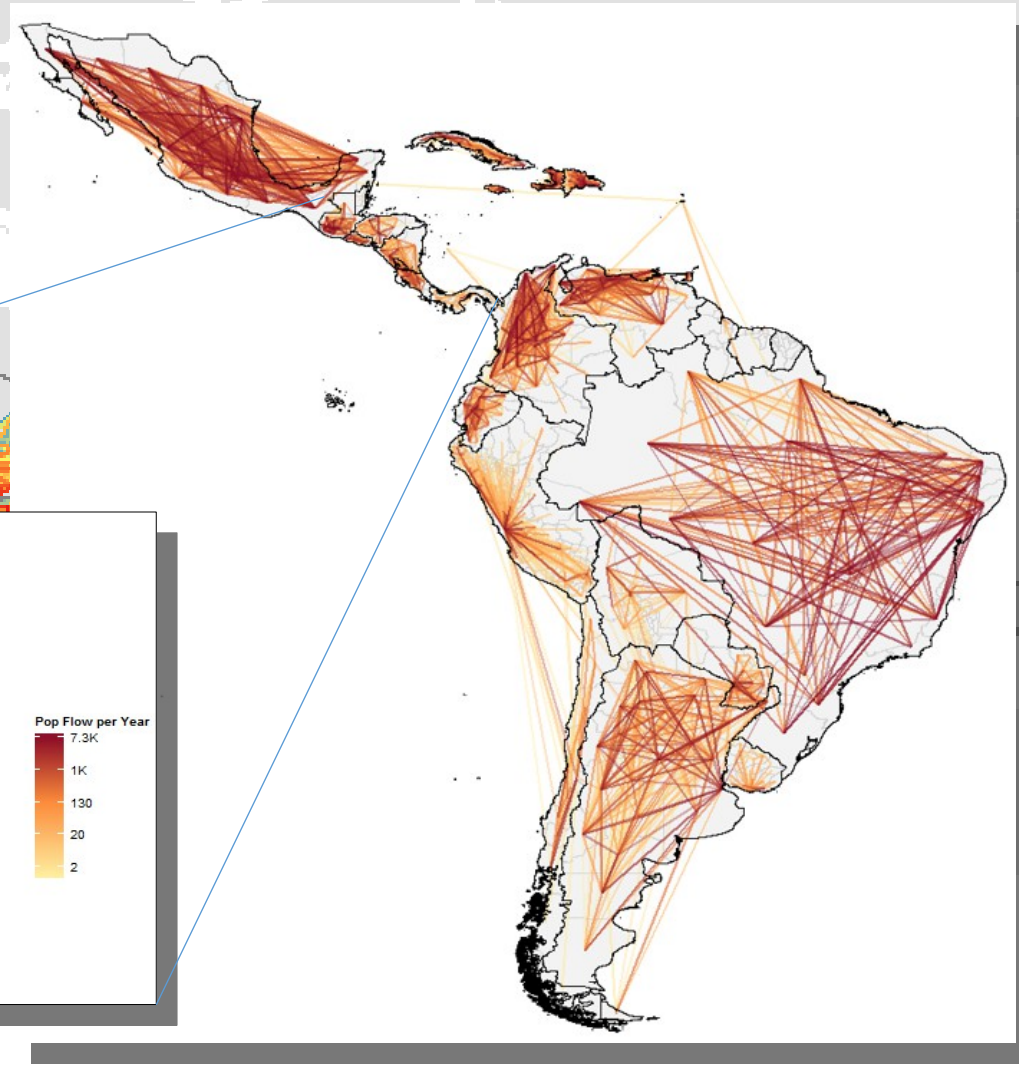




Mapping Global Connectivity







Preparedness and Response to Disasters: Haiti Earthquake 2010

Reference: Bengtsson et al. PLoS Medicine 2011

Hurricane Matthew: Estimated population movement, 24 October 2016

- The map shows the estimated distribution of people for whom their home Section Communale in the pre-hurricane period was in either Grande Anse, Sud or Nippes département, and as of 24 October had moved to another Section Communale.
- Estimates are based on movements of de-identified SIM cards which made or received at least one call pre-hurricane and on 24 October 2016.
- The SIM card movements combined with available population data derived from estimates for the year 2015[1].
- The table lists the locations with the largest number of arrivals

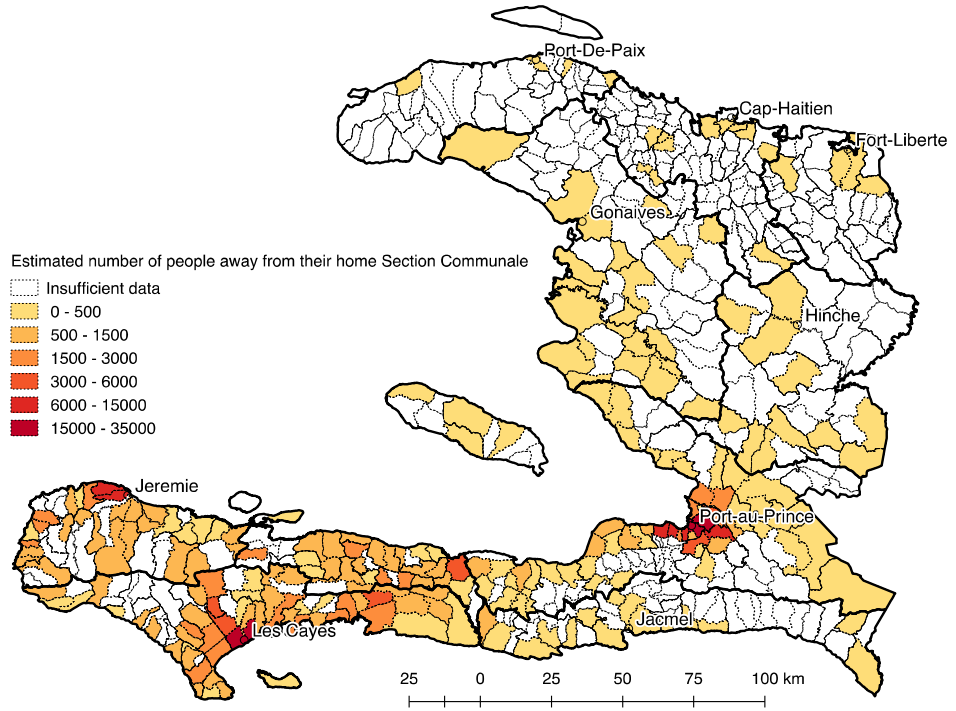
Location	Population pre-hurricane	Persons arrived	Ratio (%)
Port-au-Prince (Metropolitan area)	2870000	85700	3
Bourdet (Les Cayes)	71600	30100	42
Fond Rouge Daiyer (Jeremie)	27100	6950	26
Fond Rouge De Torbec (Jeremie)	26800	6790	25

[1] https://data.humdata.org/dataset/estimated-population-of-haiti-2015

Estimated population away from their home Section Communale[2]:

HOME DEPARTMENT:	GRANDE ANSE	SUD	NIPPES
POPULATION AWAY FROM HOME:	77500	132000	51000
% AWAY FROM HOME:	18%	17%	15%

24 October 2016, location of people away from their home Section Communale (out of those living pre-hurricane in Grande Anse, Sud and Nippes only)[3]



[2] Of the people normally resident within the given Département, we estimate the total number away from their home Section Communale on the given day.
[3] Section Communales are left blank where insufficient data is available.

Contacts:
linus.bengtsson@flowminder.org +41 78 964 88 28
chris.brooks@flowminder.org +44 7815 944 012

Flowminder.org is a non-profit organization registered in Stockholm, Sweden. Digicel is a mobile operator in Haiti

Produced 26 Oct 2016 v1.0

Predictive Modeling

First Project (MDEEP) on Mobile data and Climate Displacement: 2013 Bangladesh Cyclone Mahasen

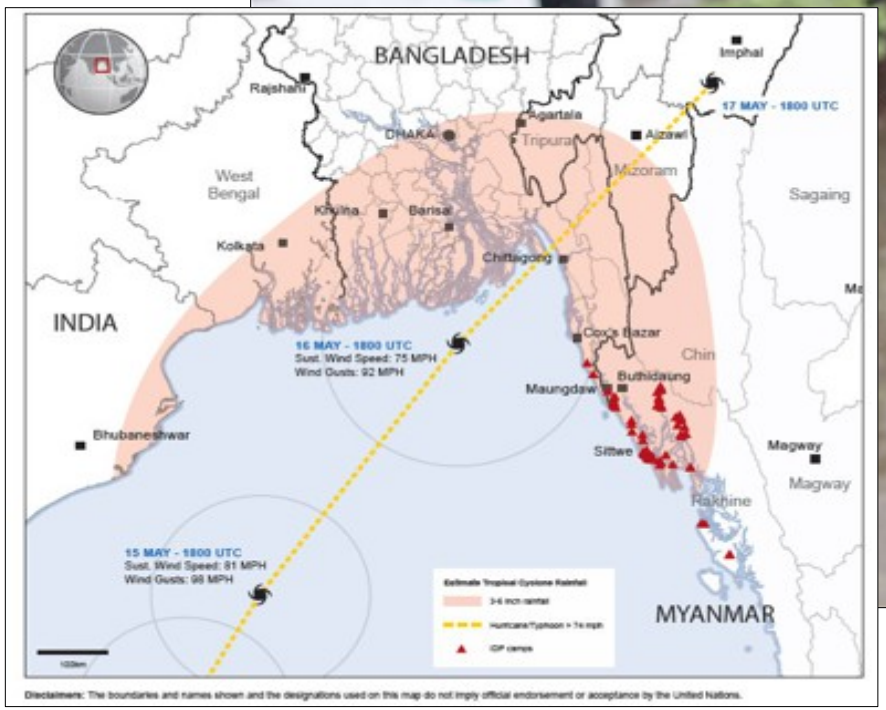
ICCCAD
International Centre for Climate Change and Development
at
Independent University, Bangladesh
in partnership with BCAS and IIED

grameenphone

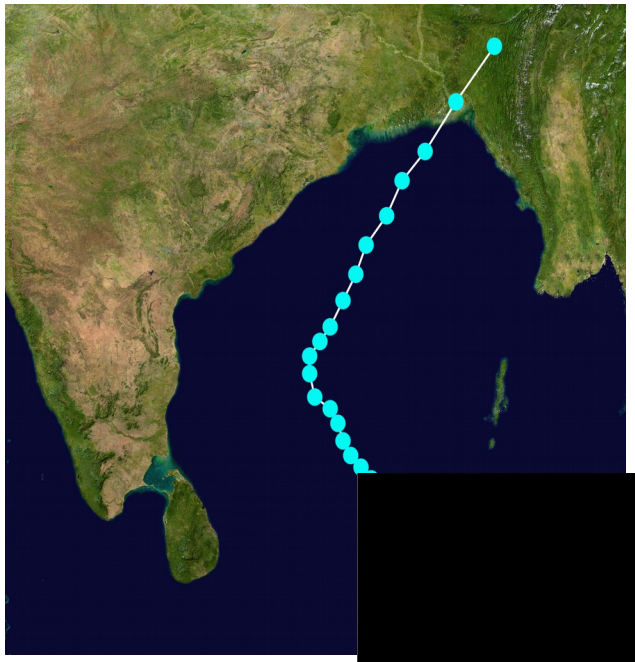
FLOWMINDER.ORG

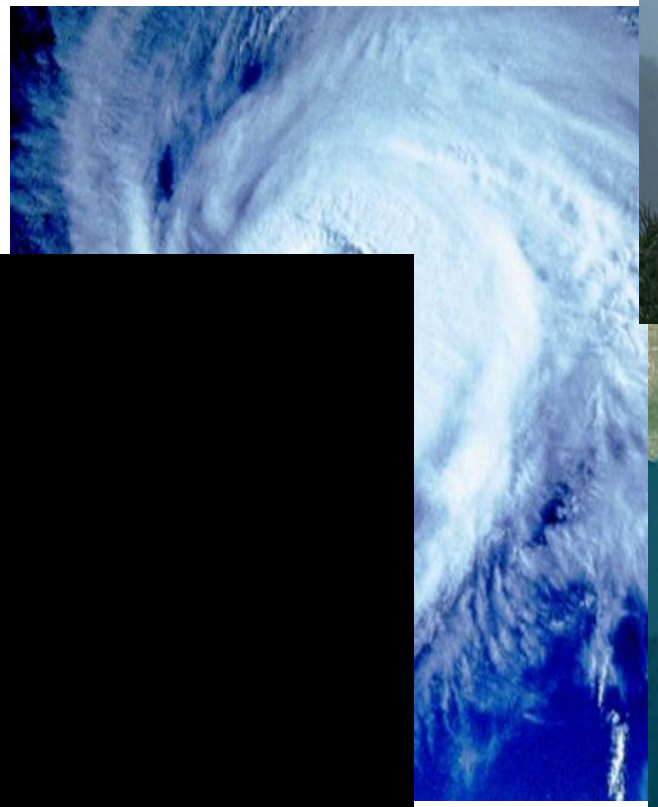
telenor group

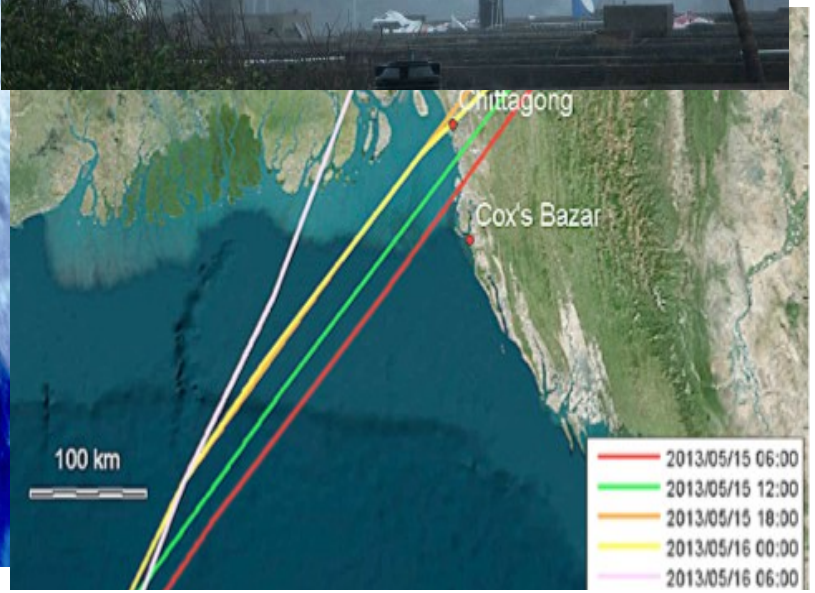
UNITED NATIONS UNIVERSITY
UNU-EHS
Institute for Environment and Human Security



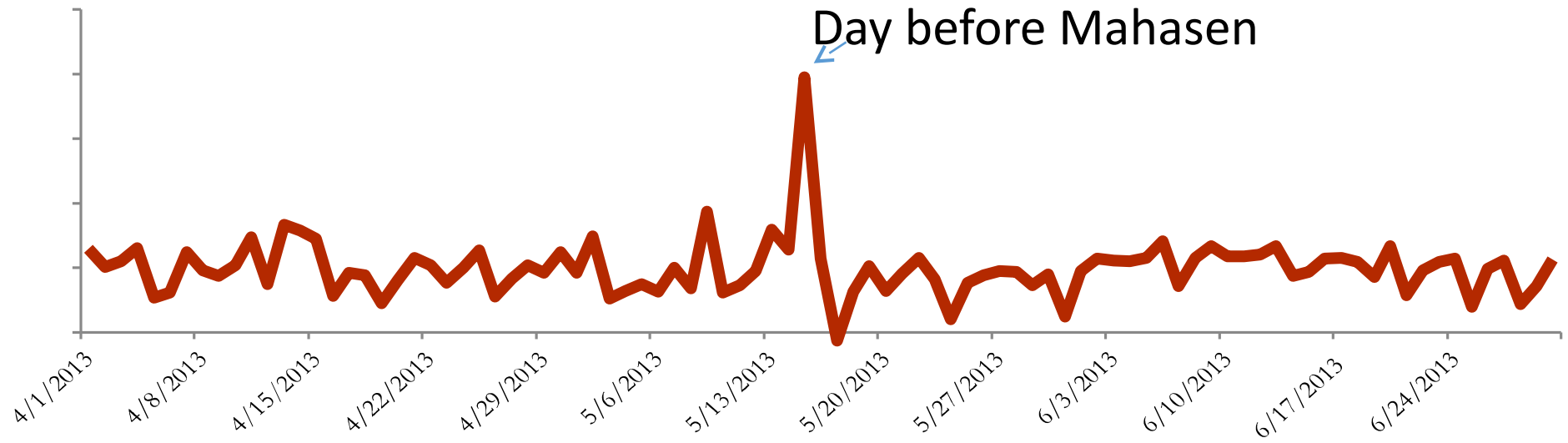
Mahasen: 16 May







Large Spending on Airtime Before Mahasen: Ability to Communicate is Paramount in a Disaster



Ref: Lu et al. Climatic Change. 2016; Global Environmental Change. 2016

ICCCAD
International Centre for Climate Change and Development
at
Independent University, Bangladesh
in partnership with BCAS and IIED

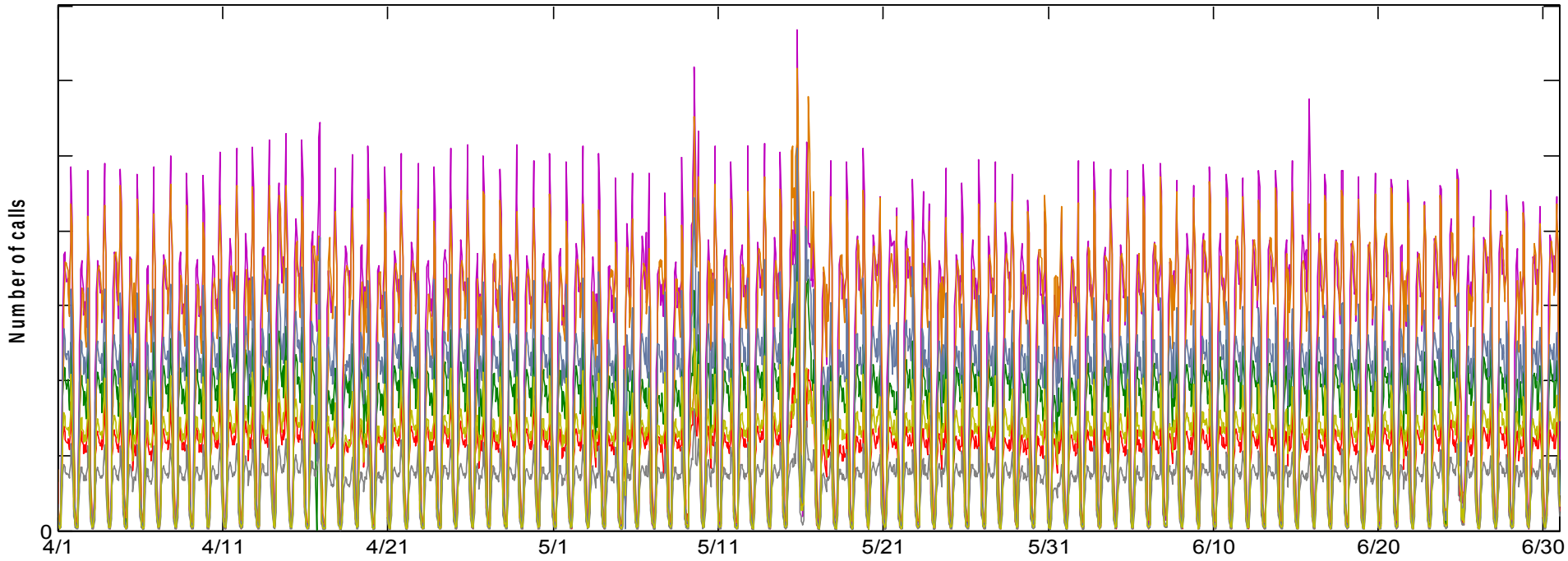
grameenphone

telenor group

FLOWMINDER.ORG

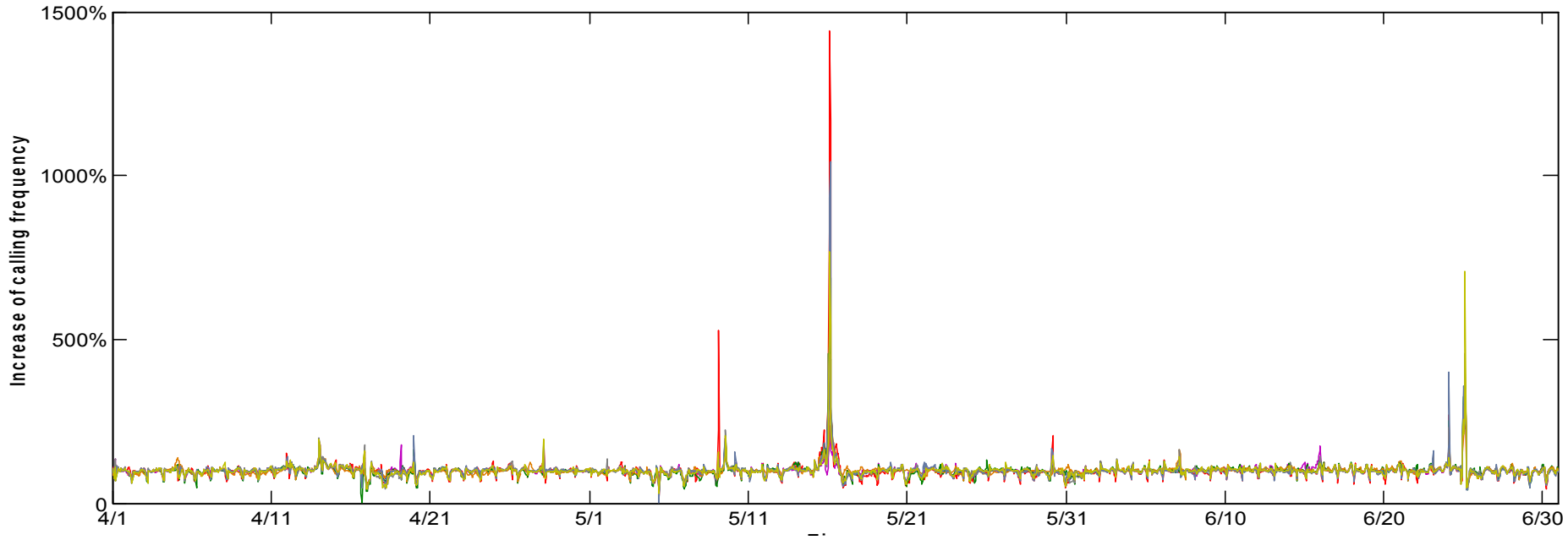
UNITED NATIONS UNIVERSITY
UNU-EHS
Institute for Environment and Human Security

Call Frequencies Follow a Daily Rythm



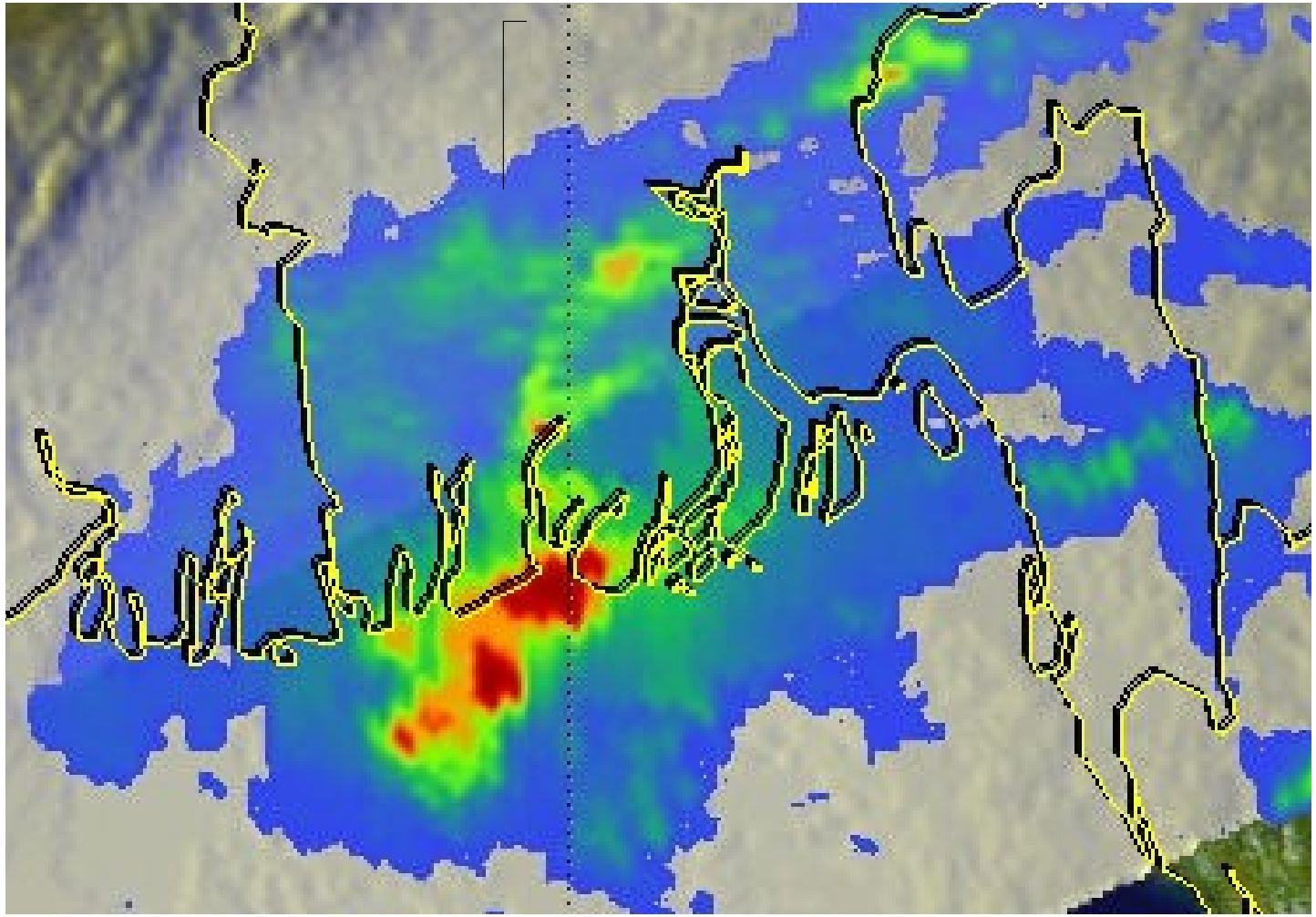
Ref: Lu et al. Climatic Change. 2016; Global Environmental Change. 2016

Call Frequency Increased 1500% During Mahasen



Ref: Lu et al. Climatic Change. 2016; Global Environmental Change. 2016

Increase in Calling Frequency Likely Indicate Impact



3.32 am on 16 May

Ref: Lu et al. Climatic Change. 2016; Global Environmental Change. 2016

Nepal Earthquake 2015: Setup & 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)

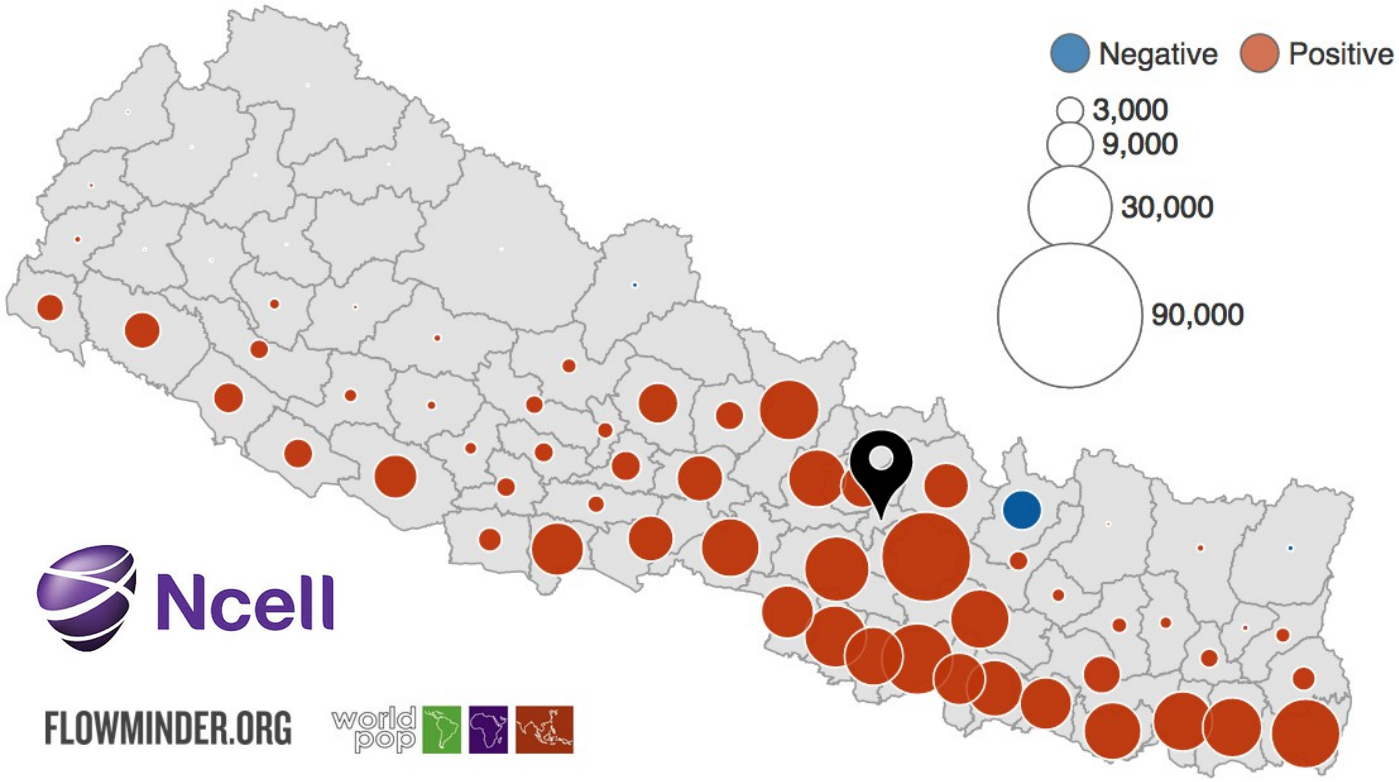
2. Kathmandu Valley

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



Ref: Bengtsson et al. Sci Rep. 2015; Wesolowski et al. PLoS Currents 2014; Wilson et al. PLoS Current 201

Global Mobile Industry Recognition (Feb 2016)

"A brilliant example of how the application of big data analysis to mobile technologies can be used to accelerate emergency aid, and provide intelligence to help prepare for future disasters."


STAR TREK

The Star Trek Fallacy

1. Data is part of **the tool, not the solution**
   – Issue-driven vs. data-driven problem solving.
2. Remote sensing data and analytics can **augment but not replace census and surveys**
3. Very few studies of bias in mobile operator data
4. Mobile network data does not have stable properties:
   - Fundamental characteristics (subscribers) **constantly changing**.
   - Mobile data is **heterogenous** – market/operators.
   - **Representativeness** – what does a SIM card represent?
   - **Realtime mobile data without validation = realtime mistakes**

Summary

- In producing estimates for different geographical scales and time periods, the integration of multiple types of data to compliment traditional sources is often required
- Novel datasets (e.g. phones, satellite) are prone to biases, but each has advantages over census data in terms of the frequency of measurement and coverage
- Methods to account for biases, reporting uncertainties and providing clear metadata/documentation to inform users are all important
- ***Great potential in complimenting traditional sources to build strong demographic databases for measuring progress towards the SDGs***

www.worldpop.org

@WorldPopProject

FLOWMINDER.ORG

www.flowminder.org

@Flowminder