Harnessing Big Data for National Statistics and Evidence-Based Policymaking

Blantyre, August 2025

Introduction: Big Data for Policymaking

Joshua Blumenstock

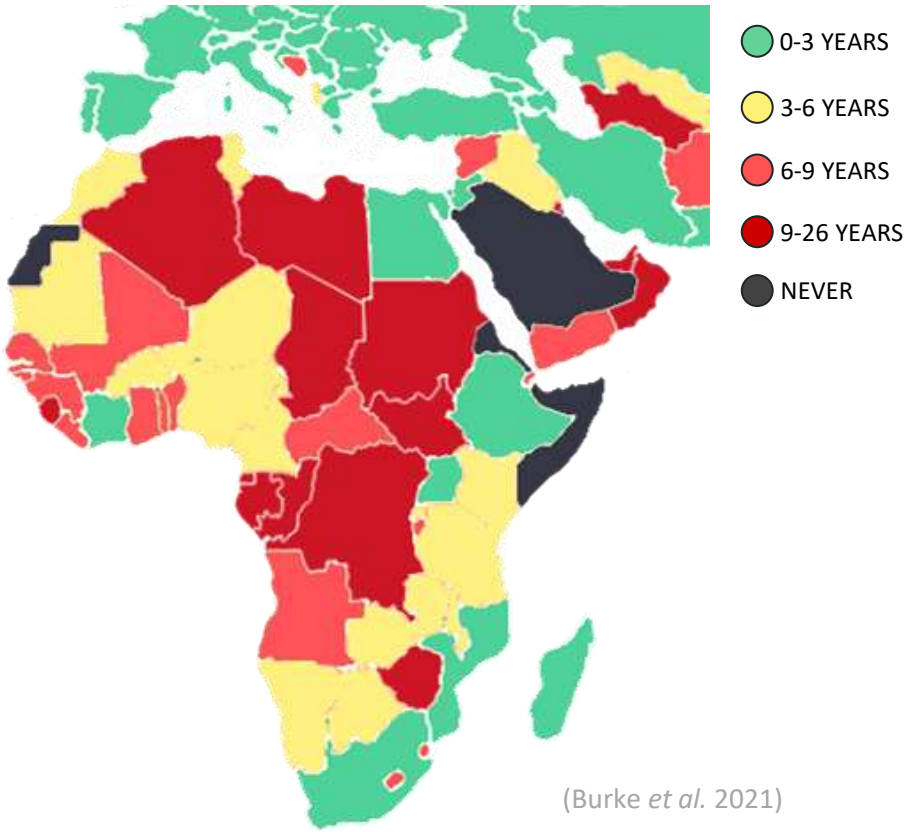
Harnessing Big Data for National Statistics and Evidence-Based Policymaking

Blantyre, August 2025

Motivation: Data Gaps

Many countries lack reliable, up-to-date data

Avg. gap between economic surveys, 1993 - 2021



(Burke *et al.* 2021)

- Typical household survey costs $10+ million; census-scale sweep costs $50+ million
- Less than half of the poorest countries have completed a census in past 10 years

Motivation: Data Gaps

Why are **data gaps** a problem?

nature
International journal of science

WORLD VIEW · 28 FEBRUARY 2018

Kofi Annan

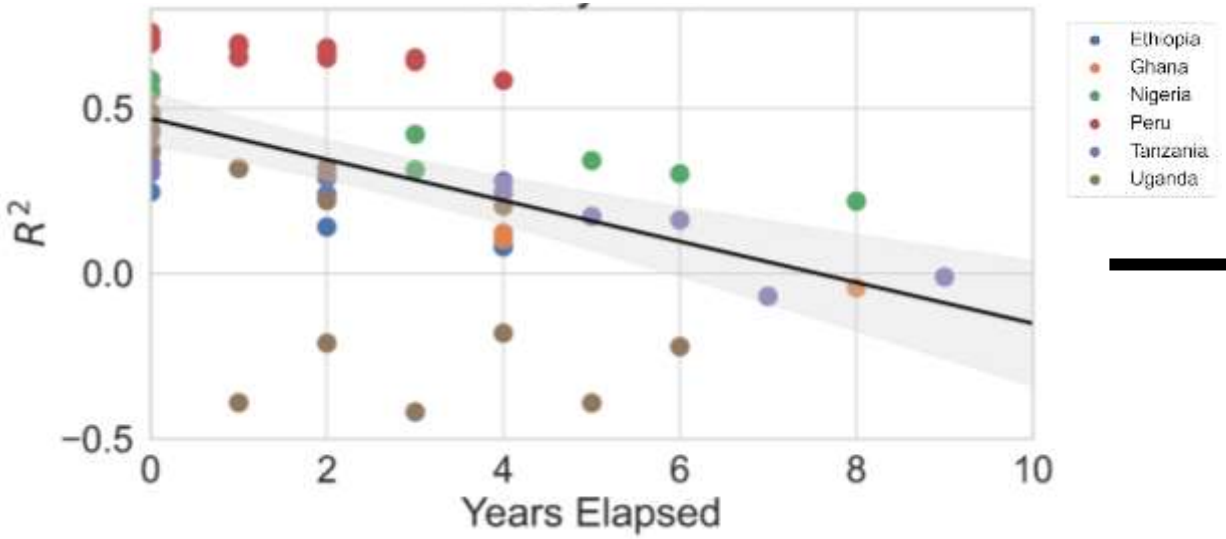
"Data gaps undermine our ability to target resources, develop policies and track accountability. Without good data, we're flying blind. **If you can't see it, you can't solve it**… ."

Example: Social protection programs

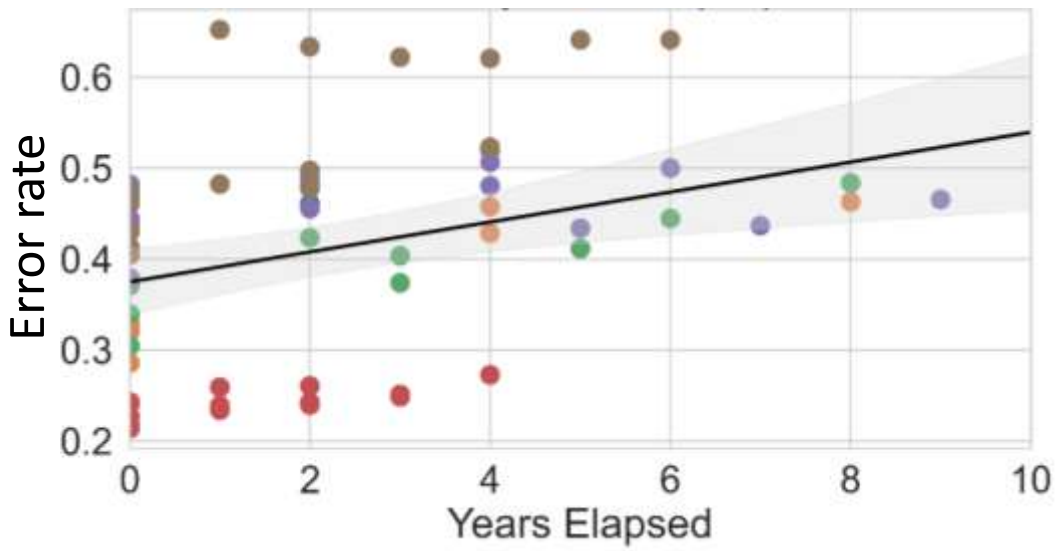
As data go out of date, **policies and interventions are less effective**

- For example, in a social protection program the size of PROGRESA (MEXICO), out of date data would result in **hundreds of thousands of exclusion errors each year**

Accuracy of means testing since date of survey



Error rate of targeted policy



Reference: Aiken, Blumenstock, and Ohlenberg (2025)

This talk: Outline

The potential for **big data** in national statistics and policy

- Motivation
- **What data, and why should we care?**
- Case Study: Nigeria
- Case Study: Togo
- Summary and Reflections

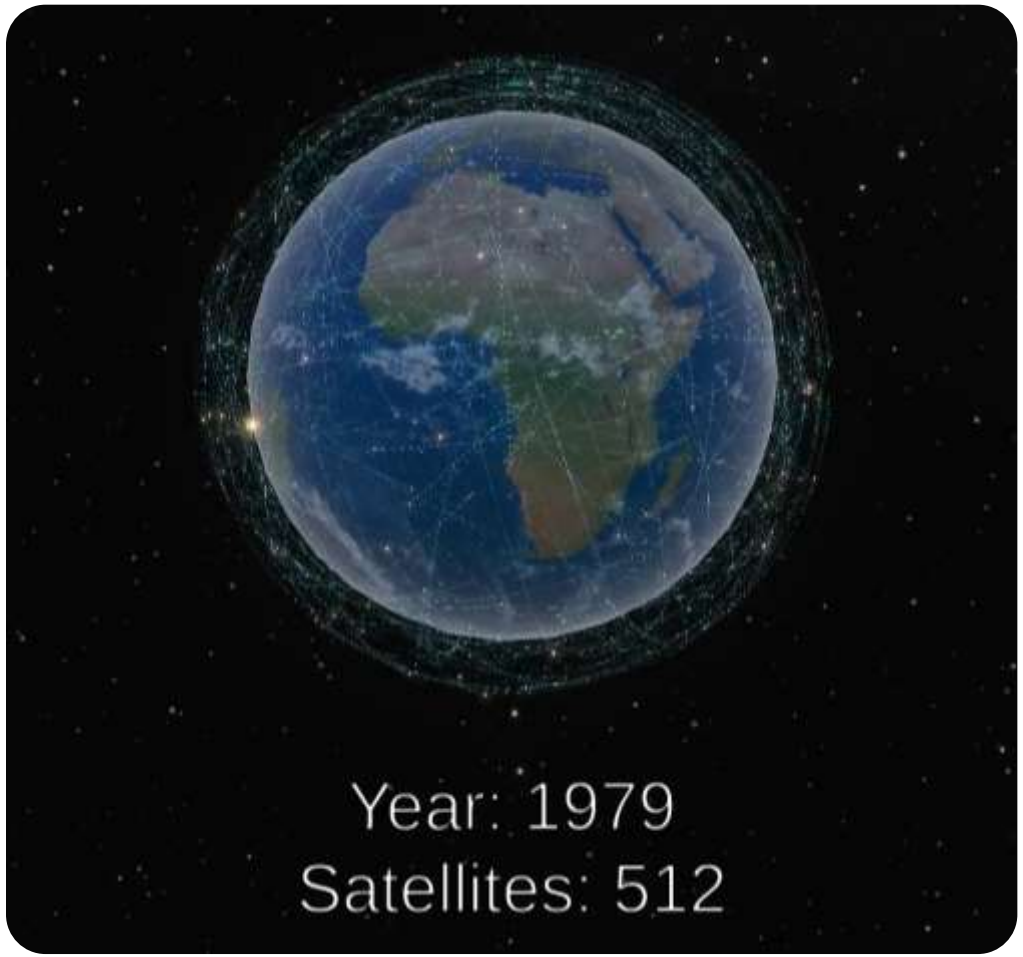
New Data: What are we talking about?

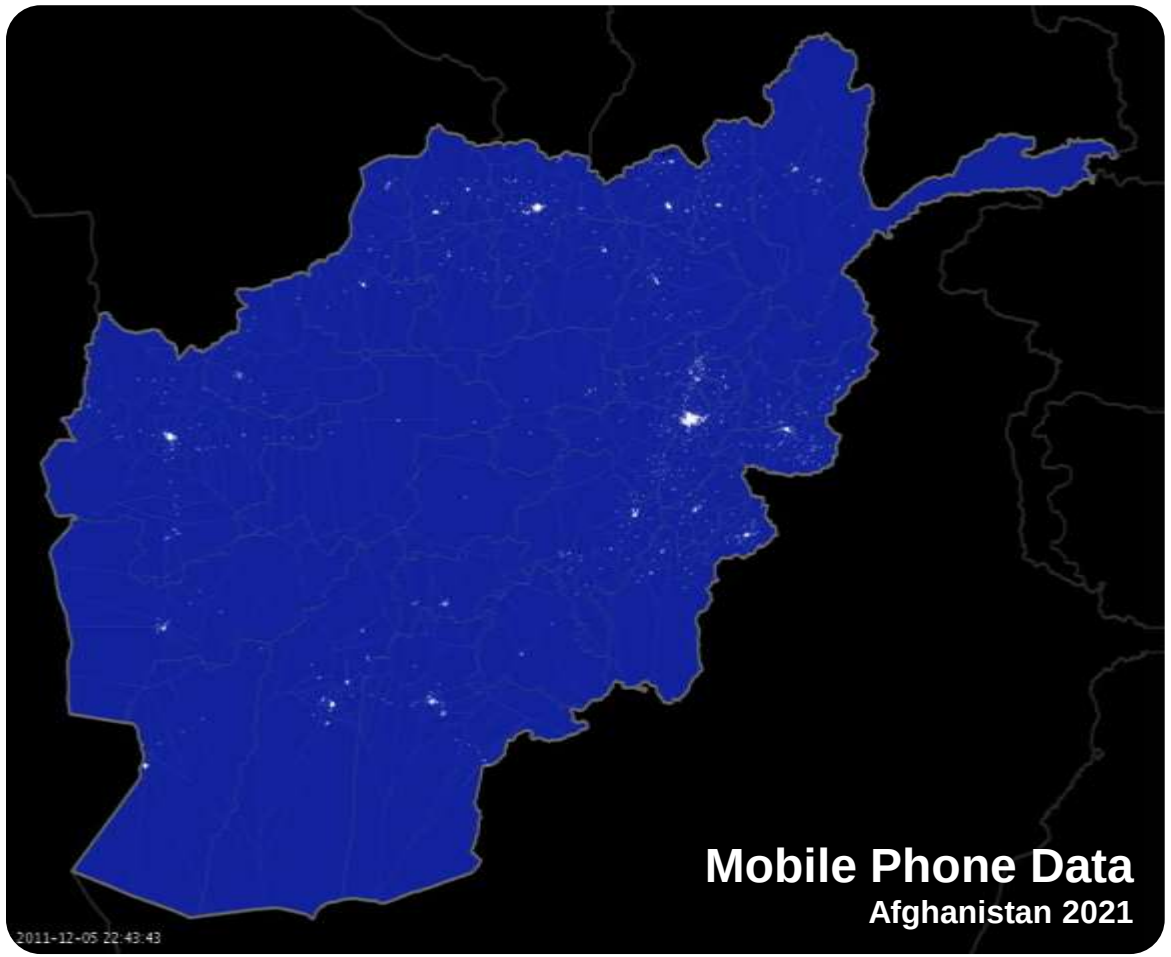
The last decade has seen a proliferation of non-traditional data:

- **Remote sensing:** High-resolution imagery available daily
- **Mobile phones:** 96% mobile phone penetration globally
- **Digital traces:** 3.2 billion active Facebook users
- **Other sources:** …

Satellite and Phone data

New forms of data are available even in poor/remote regions





Big data and policymaking: **Relevance**

Why are these data "suddenly" relevant to policymaking?

Explosion in data availability

Satellites, remote sensing, mobile phones, digital government records, admin data, networked sensors

Advances in computing and cloud infrastructure

High-speed computing and affordable cloud storage

Policy demand for timely, evidence-based decisions

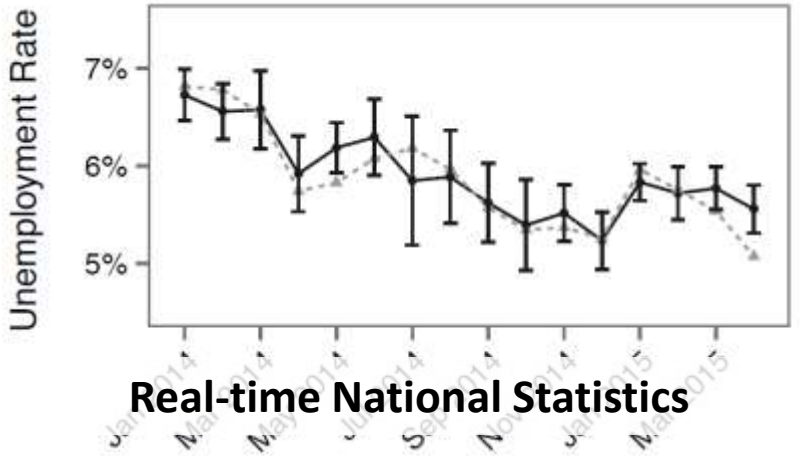
Ability to systematically maximize impact, real-time insights to inform rapid interventions

Improved algorithms and open-source tools

Modern machine learning tools integrated into common frameworks, point-and-click dashboards

Big data and policymaking: **Example Use Cases**

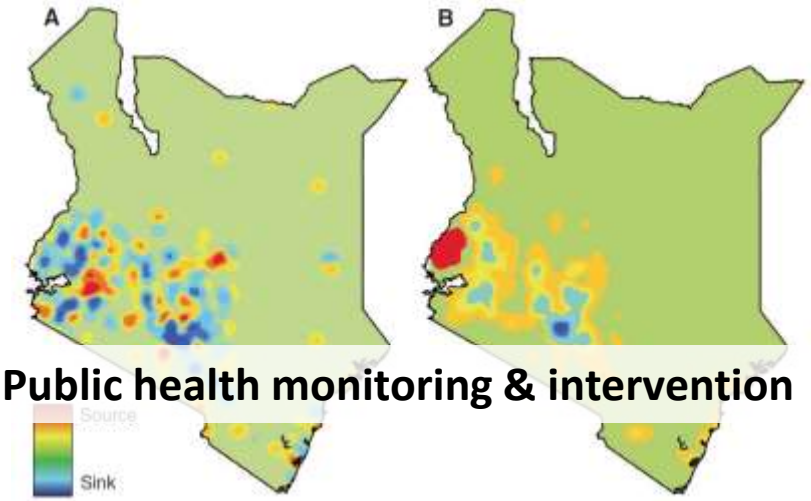
Mapping Poverty and Demographics



Real-time National Statistics

Early warning and crisis response

Urban transport and mobility



Public health monitoring & intervention

Migration and displacement

This talk: Outline

The potential for **big data** in national statistics and policy

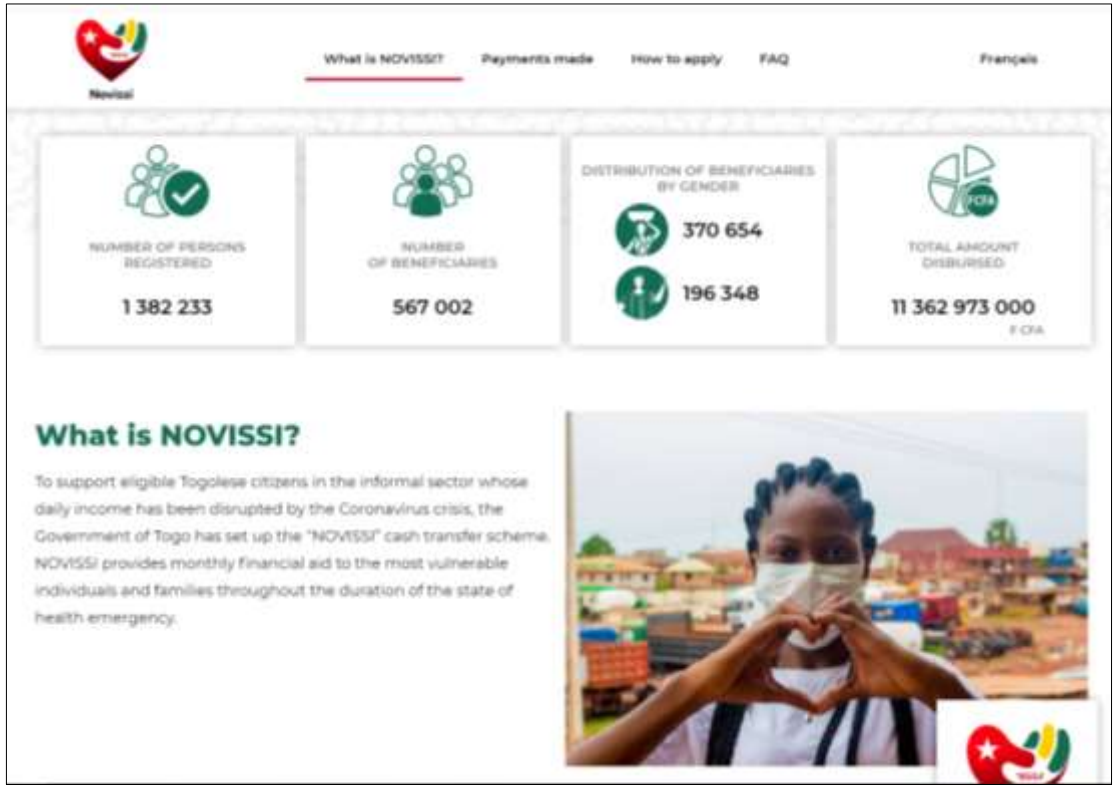
- Motivation
- What data, and why should we care?
- **Case Studies**
- Summary and Reflections

Case Study: Togo's Novissi Program

Case Study: Togo

Novissi: Togo's immediate response to the Covid-19 pandemic

- Beneficiaries received bi-weekly payments of roughly $10 USD
- 100% digital registration (USSD) and payment (mobile money)



The challenge of **targeting**

A key challenge: *Who should be eligible for benefits?*

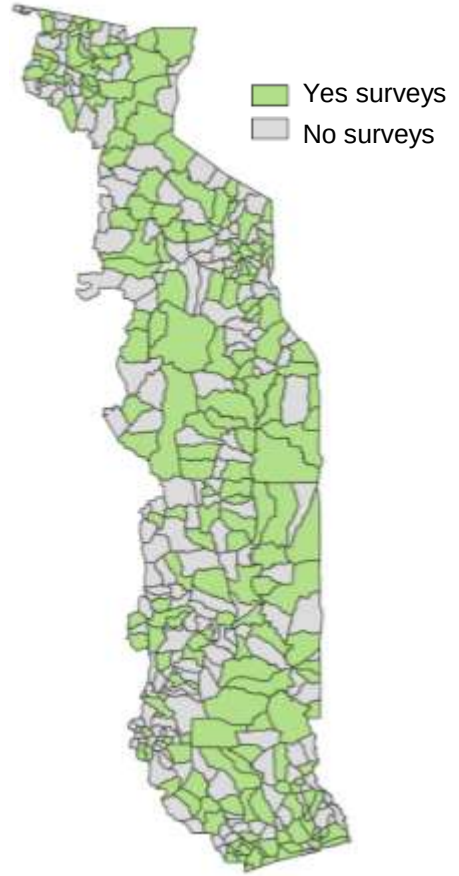
- At the onset of the pandemic, Togo did not have a complete social registry
  - The last census was conducted in 2011
  - Not possible to update in the middle of a public health crisis

- In early 2020, Minister Lawson asked if we could build a poverty map
  - She had seen our work on poverty mapping in Nigeria
  - Her idea was to use those maps to direct resources to the poorest cantons

Give Directly

Global Opportunity Lab
https://gol.berkeley.edu/

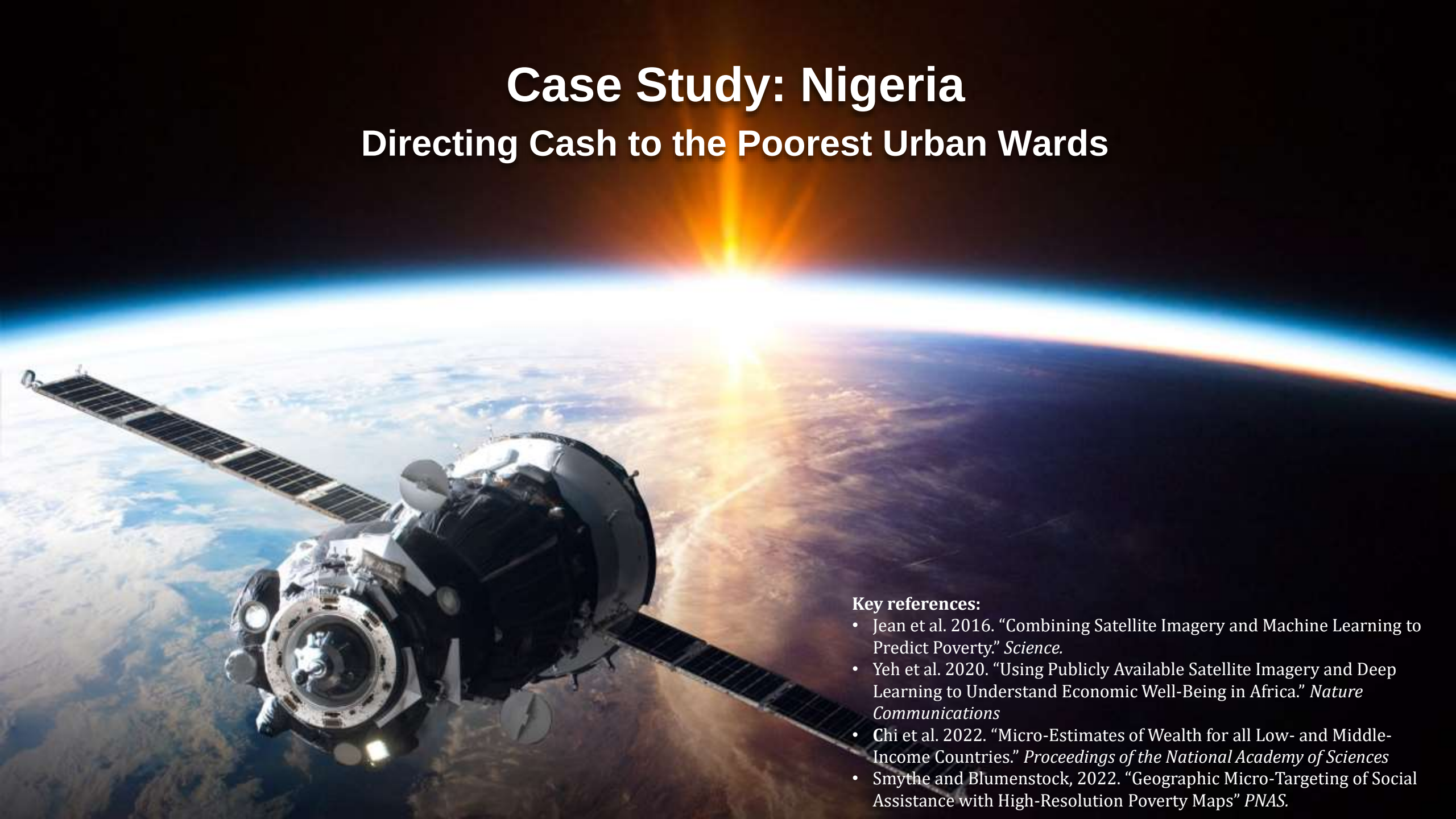
Berkeley
UNIVERSITY OF CALIFORNIA



Map of cantons of Togo

Case Study: Nigeria

Directing Cash to the Poorest Urban Wards



Key references:

- Jean et al. 2016. “Combining Satellite Imagery and Machine Learning to Predict Poverty.” *Science.*
- Yeh et al. 2020. “Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa.” *Nature Communications*
- Chi et al. 2022. “Micro-Estimates of Wealth for all Low- and Middle-Income Countries.” *Proceedings of the National Academy of Sciences*
- Smythe and Blumenstock, 2022. “Geographic Micro-Targeting of Social Assistance with High-Resolution Poverty Maps” *PNAS.*

Nigeria: Policy need

Like Togo, Nigeria had urgent need for an up-to-date poverty map

- World Bank agreed to finance an urban anti-poverty program
- But: existing survey data did not allow for ward-level decisions

← **States (adm-2):** 100% of wards had data

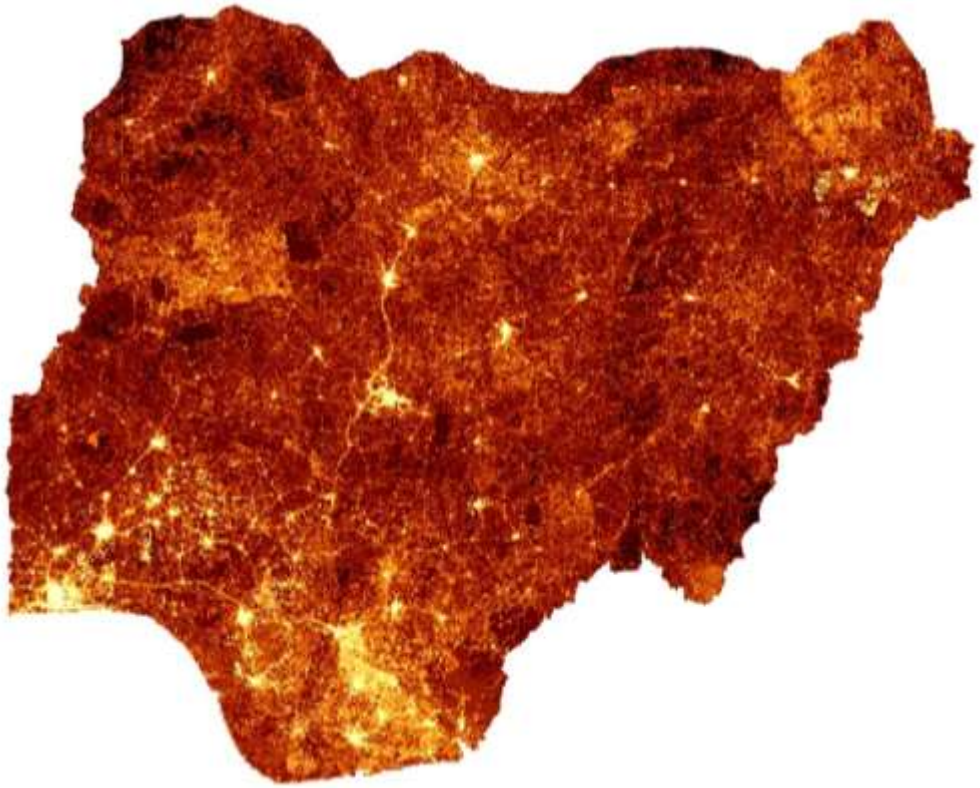
← **LGA (adm-3):** 81.5% of LGAs had data

← **Ward (adm-4):** Only 13.8% of wards had data!

Nigeria: Our objective

Our objective was to use satellite imagery, processed with AI, to produce a high-resolution poverty map for the government

Machine Learning



Nigeria: Intuition

A simple idea 💡 : wealthy regions *look* different from poor regions

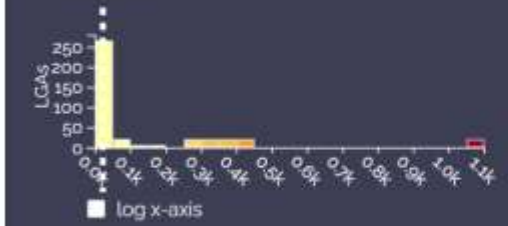
- We use AI to learn these patterns at a national scale

Mexico City, Mexico

Cape Town, South Africa

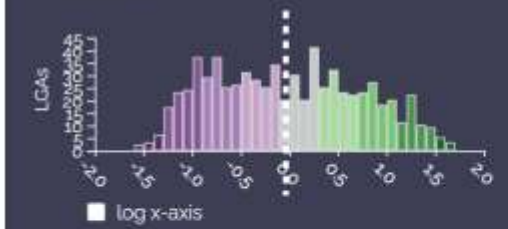
COVID confirmed cases 6,799

as of 2020-05-??



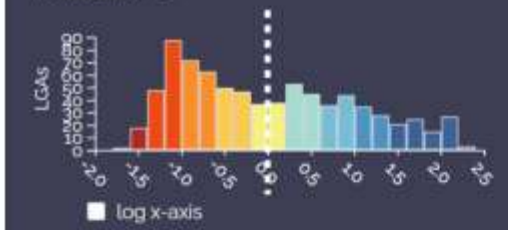
Wealth estimates from survey data

as of 2020-07-14



Wealth estimates from satellite data

as of 2020-07-14



Berkeley
UNIVERSITY OF CALIFORNIA

COVID-19

- Confirmed cases (LGA level) *2020-05-??*
- Counts of confirmed cases (LGA level) *2020-05-??*

Wealth estimates

- None
- Survey (LGA level) *2020-07-14*
- Satellite (LGA level) *2020-07-14*
- Survey (Ward level) *2020-07-14*
- Satellite (Ward level) *2020-07-14*

COVID-19 case counts

0 – 320	0
320 – 530	0.3
530 – 850	0.5
850+	0.8

Wealth estimates - satellite (Wards)

-2.00 – -1.60	0
-1.60 – -1.20	0.1
-1.20 – -0.80	0.2
-0.80 – -0.40	0.3
-0.40 – 0.00	0.4
0.00 – 0.40	0.5
0.40 – 0.80	0.6
0.80 – 1.20	0.7
1.20 – 1.60	0.8
1.60+	0.9

Ward	ilaje_mahin i_3
Wealth estimate (survey)	-0.11
Wealth estimate (satellite)	-0.64

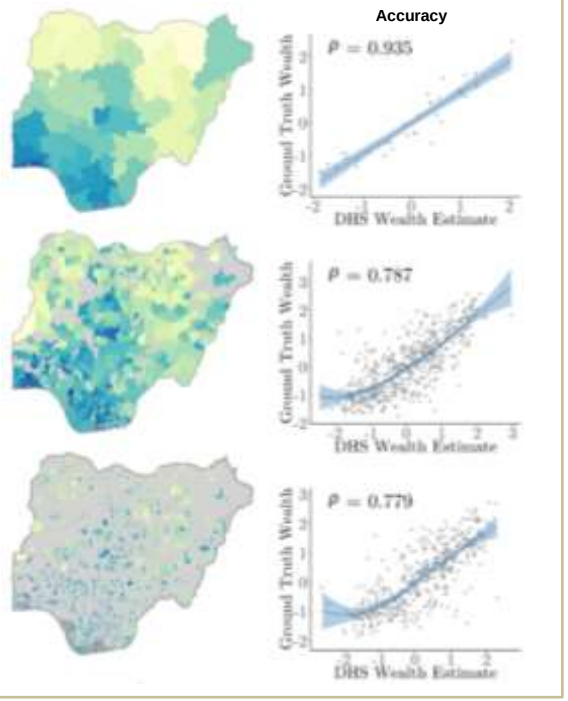
Leaflet | © OpenStreetMap contributors © CARTO

Nigeria: Conclusion

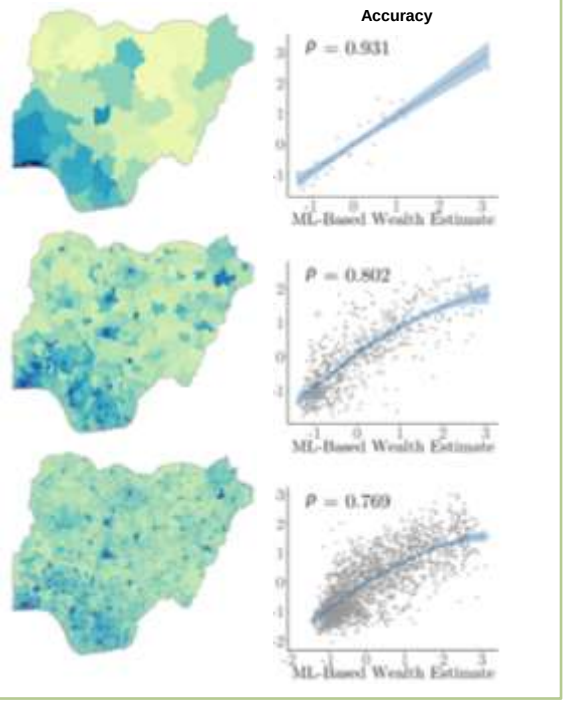
The new poverty maps were as accurate as the old ones

- And: **100% coverage** nationally, at **extremely high resolution**
- They allowed the Nigerian government to prioritize the poorest wards

Old Maps



New Maps



Press coverage

npr

WIRED

FAST COMPANY

These new poverty maps could reshape how we deliver international aid

A Clever Strategy to Distribute Covid Aid—With Satellite Data

Journal article

PNAS

Geographic microtargeting of social assistance with high-resolution poverty maps

Significance

This talk: Outline

The potential for **big data** in national statistics and policy

- Motivation
- What data, and why should we care?
- Case Study: Nigeria
- **Case Study: Togo**
- Summary and Reflections

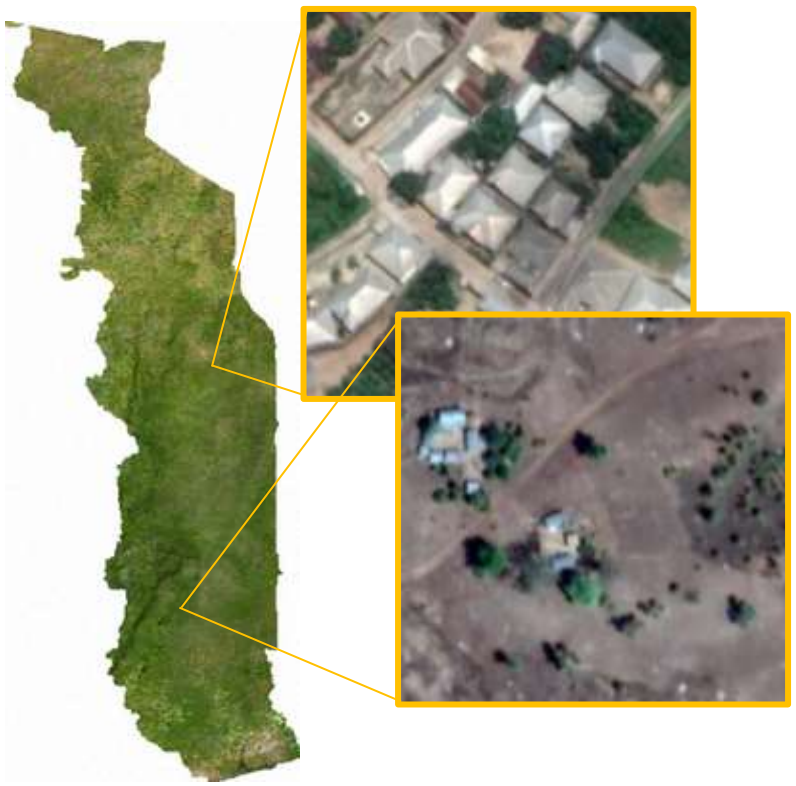
Case Study: Togo

Directing aid to the poorest *individuals*

Togo: Expanding the Novissi program

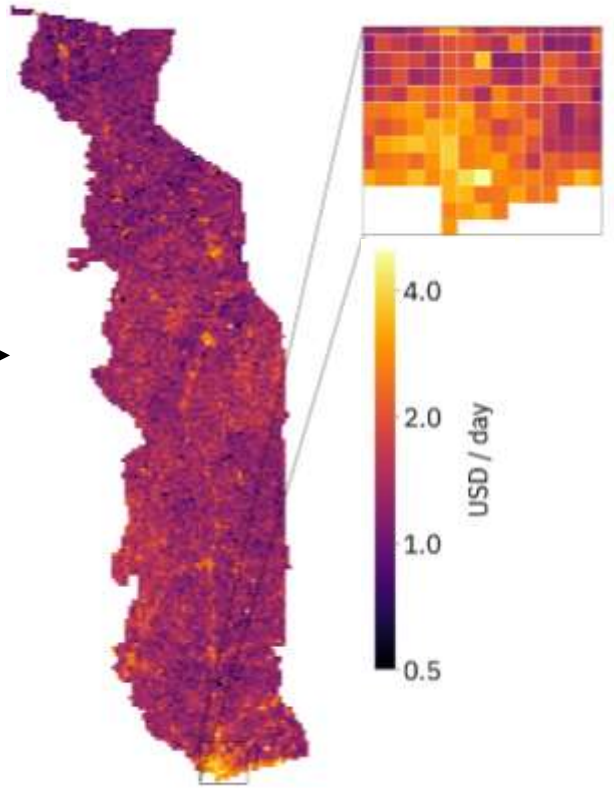
In 2020, we built M. Lawson a poverty map, using similar approach

- Initial plan was to use map to target Novissi to Togo's poorest cantons

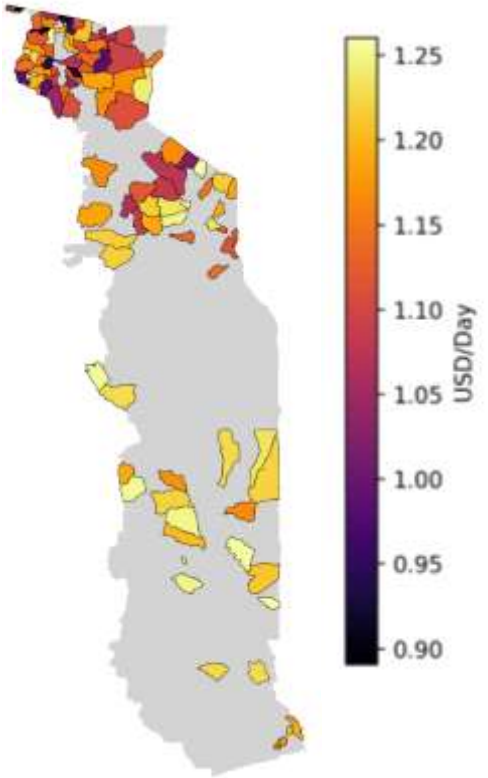


High-resolution satellite imagery

Machine Learning



ML-based poverty maps

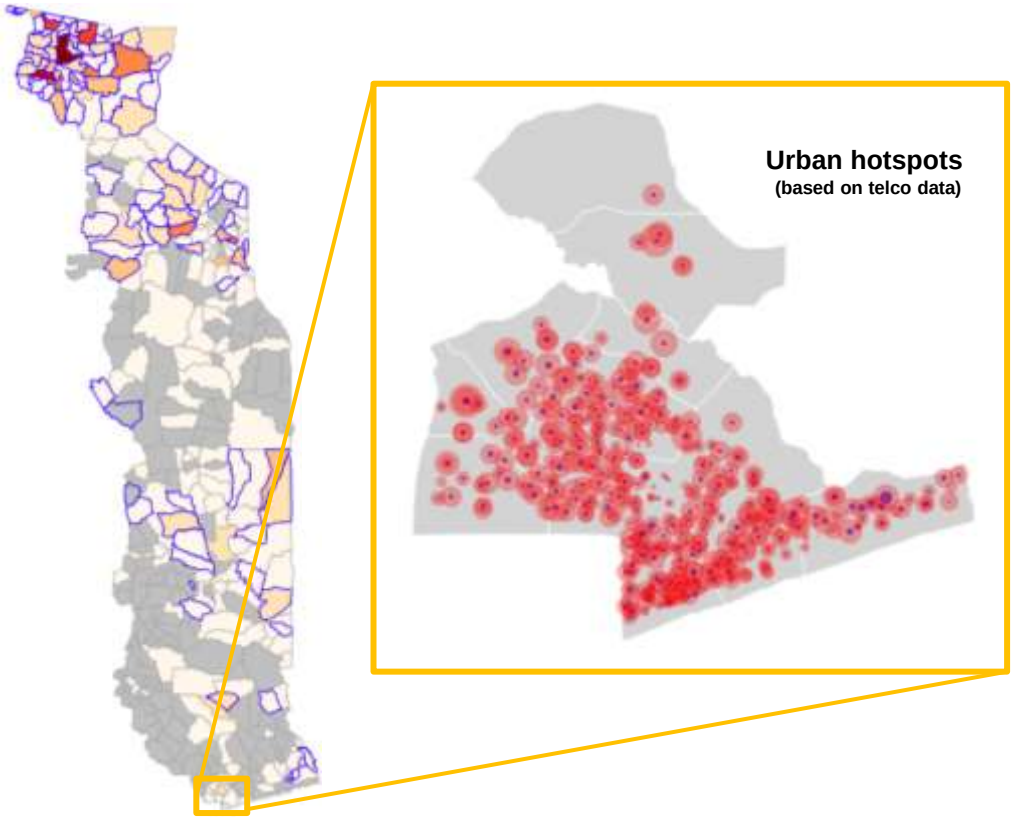


Eligible cantons

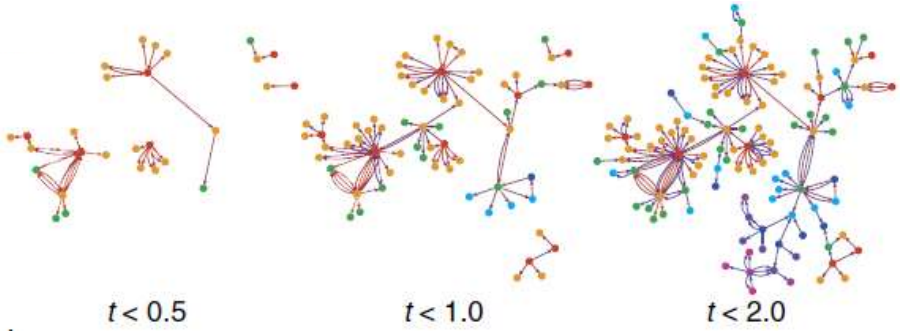
Togo: Mobile phone data integration

Along the way, we helped the gov't track Covid-19 with telco data

- (more on this soon!)



Predicted cases
(based on telco data)



Diffusion network

Journal article

Article | Open access | Published: 29 June 2021

Public mobility data enables COVID-19 forecasting and management at local and global scales

Cornelia Ilin, Sébastien Annan-Phan, Xiao Hui Tai, Shikhar Mehra, Solomon Hsiang & Joshua E. Blumenstock

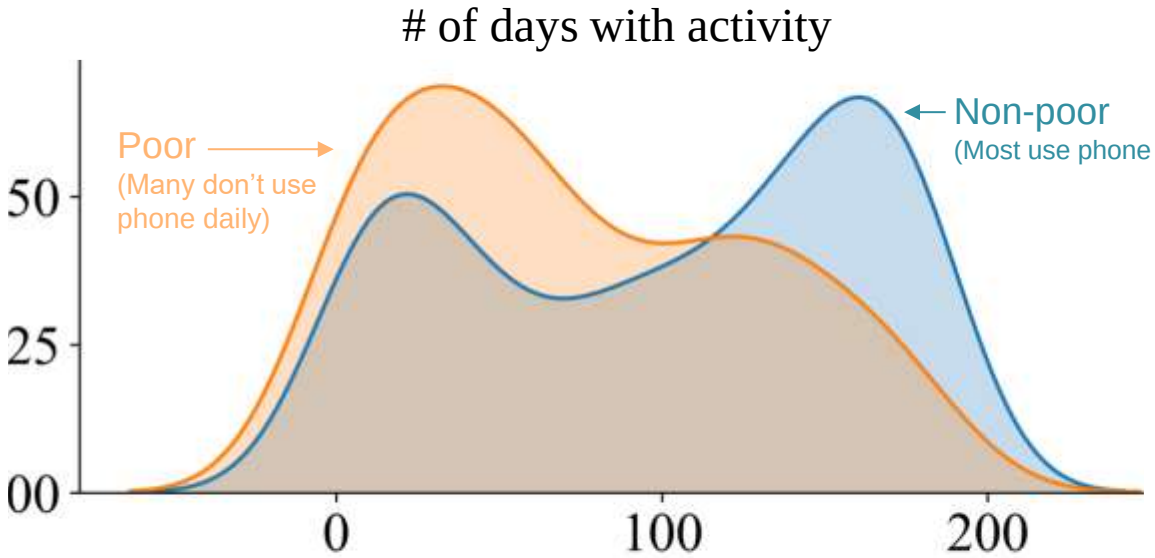
Scientific Reports 11, Article number: 13531 (2021) | Cite this article

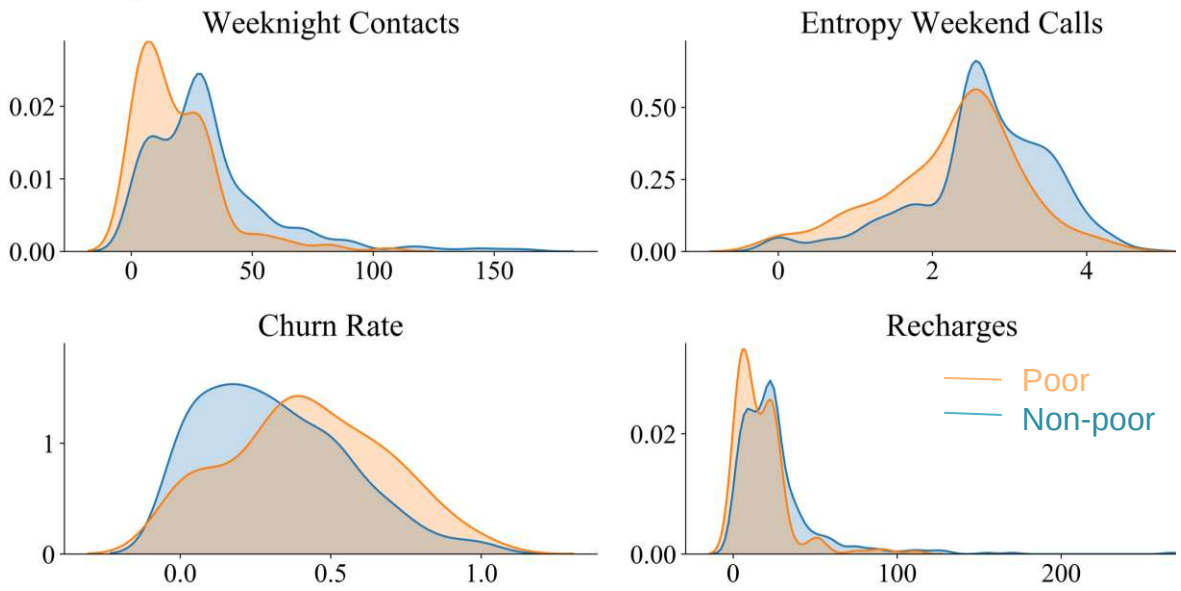
9585 Accesses | 125 Citations | 17 Altmetric | Metrics

Togo: Targeting with phone data

This led us to ask: **Can phone data help us prioritize poorest?**

- The idea 💡 : Wealthy people use phones differently than poor people
- In Rwanda, we'd shown that predictive algorithms can accurately estimate wealth





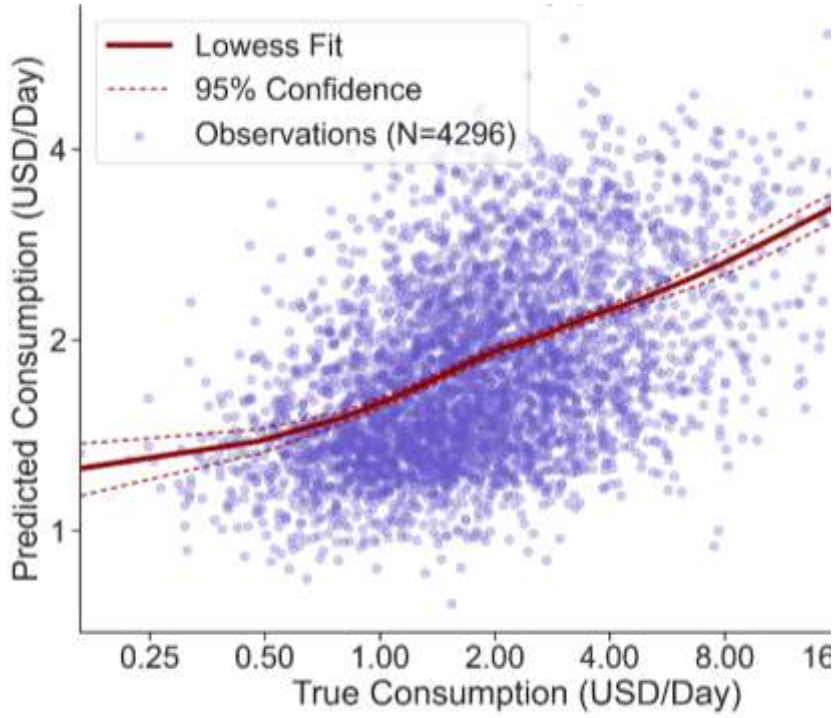
Togo: Targeting with phone data

In Togo, we found that using phone data (and machine learning), we could **predict the total consumption** of each mobile subscriber

Asset	Consumption	PMT	Asset Index
Prediction Accuracy (N = 4,171)			
Machine Learning	0.46	0.62	0.74
"Intuition" model	0.13	0.16	0.11

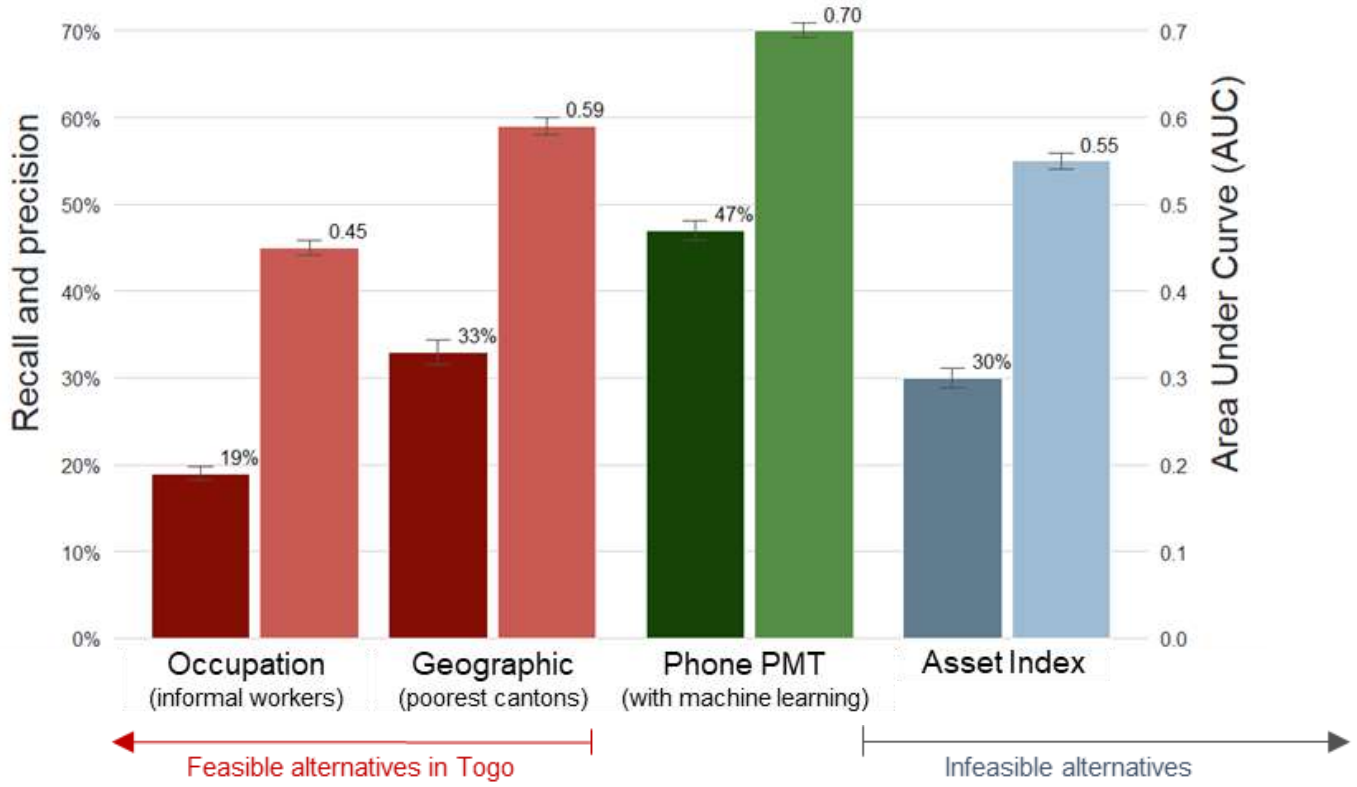
Correlation between predicted value and "ground truth" measured in survey

Predicted vs. actual consumption



Togo: Key finding

Using a combination of phone data and machine learning, the government **reduced exclusion errors** by roughly 50%



Article | Open access | Published: 16 March 2022

Machine learning and phone data can improve targeting of humanitarian aid

Emily Aiken, Suzanne Bellue, Dean Karlan, Chris Udry & Joshua E. Blumenstock

Nature 603, 864–870 (2022) | Cite this article

60k Accesses | 99 Citations | 495 Altmetric | Metrics

This talk: Outline

The potential for **big data** in national statistics and policy

- Motivation
- What data, and why should we care?
- Case Studies
- **Summary and Reflections**

Togo and Nigeria: Summary

In Togo and Nigeria, big data allowed policymakers to **design and implement more effective policies**

Nigeria: Satellite data

Output: High-resolution poverty maps

Result: Cash targeted to poorest urban wards

Togo: Mobile phone data

Output: Predicted poverty of each subscriber

Result: Cash targeted to poorest subscribers

Big Data for Policy: Summary

To unlock such applications and innovation requires careful planning (more on this later today!)

Regulation and Governance

Frameworks required to enable ethical, inclusive, and responsible use of data

Access and Incentives

Institutional and financial support for innovation; incentives for companies to share data

Capacity

Invest in training, tools, and partnerships; especially local data science capacity

Collaboration

Leverage universities, international agencies, local innovators, and local industry

Big Data for Policy: Summary

The **policy need**

- Traditional data sources are costly to maintain, and often incomplete
- This hinders policymakers' ability to make effective, evidence-based decisions

The **opportunity**

- New data sources are now widespread, even in remote areas
- Advances in computing & AI make it possible to analyze big data in real-time

The **challenges**

- Big data and AI are not a silver bullet!
- They should complement, not replace traditional approaches
- Innovation must be responsible and equitable

Thanks!

Contact Info

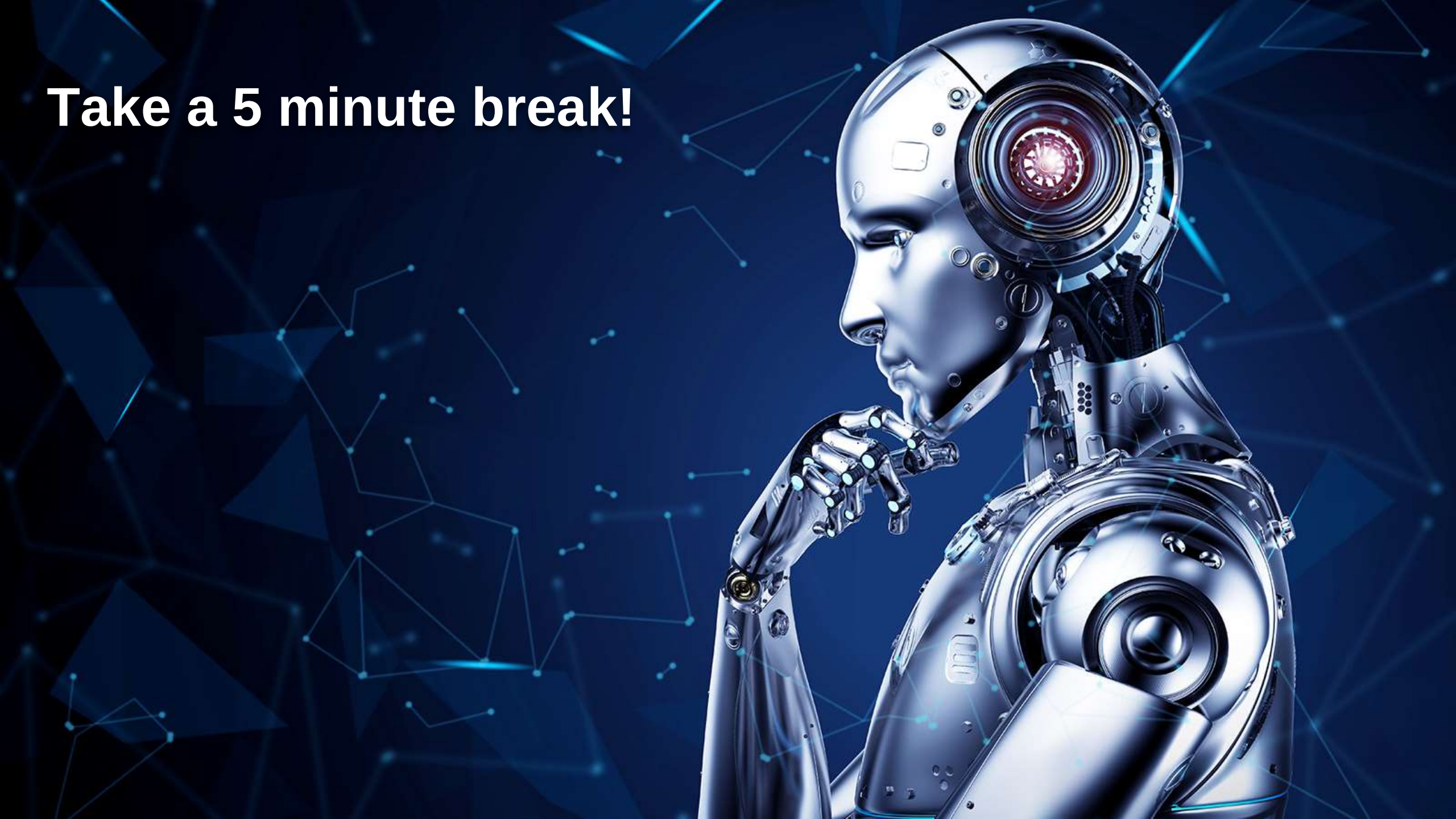
www.jblumenstock.com

jblumenstock@berkeley.edu

@jblumenstock

Global Opportunity Lab

https://gol.berkeley.edu/

The Team (part of it, in Rwanda last year)

Take a 5 minute break!