



# Measuring Poverty with Mobile Phone Metadata

Joshua Blumenstock, University of Washington

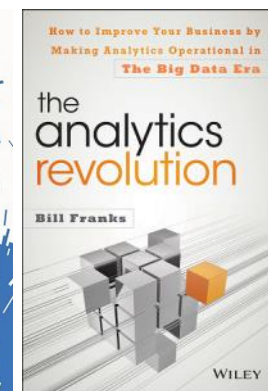
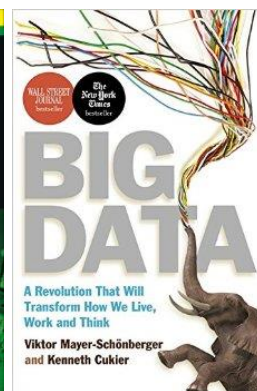
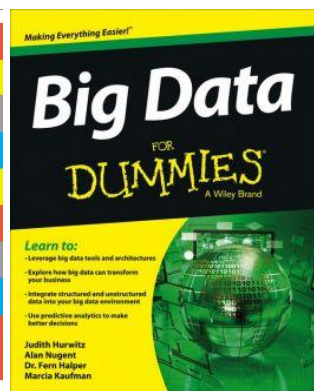
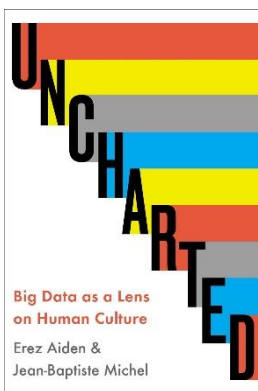
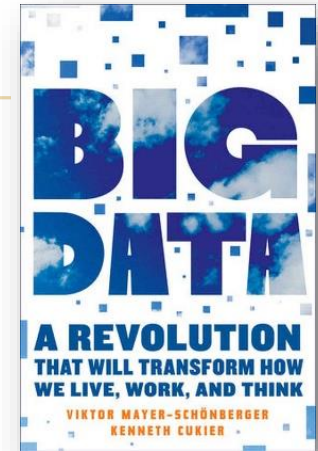
Supported by:



# Motivation: Context

## The “Big Data Revolution”

- 📱 Mobile phones: 96% globally
- 📘 Facebook: 1.5B active users (152M daily in US)
- 🐦 Twitter: 300M active users (500M Tweets/day)
- 🗨️ WhatsApp: 700M monthly active users
- 🛰️ Sensors: 1275 satellites, traffic cams, ...



# Motivation: Context

---



# Motivation: Context

---

## “Poor numbers” from poor countries?

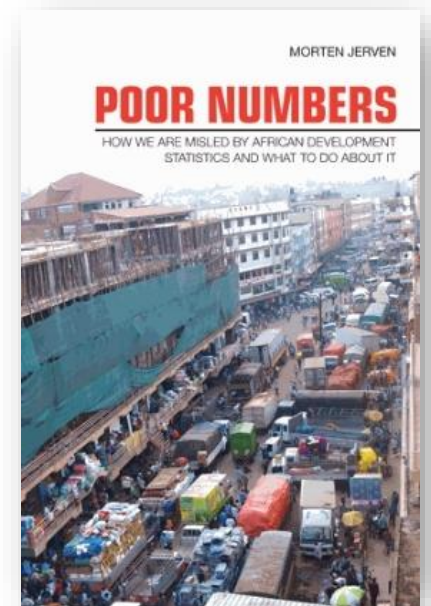
Compare successive rounds of Angolan census:

- 1970 census: 5.6 Million
- 2014 census: 24.4 Million

Official estimate of Ghanaian GDP:

- From US\$6.9 B to US\$11.8 *overnight*

More than a “**statistical tragedy**”



# Motivation: Why this matters

These data are critical to **social science research**

- e.g., understanding growth and inequality (Kuznets 1955, Fields 1989)

And

## The New York Times

Economic View

By SENDHIL MULLAINATHAN APRIL 1, 2016

Consider an unlikely problem: finding the poor. Even in a world riddled with poverty, nearly every government, nonprofit and aid agency struggles with this issue. Where in Kigali should the Rwandan health ministry place a new health clinic? Which rural districts in India should receive rice at subsidized prices? All these decisions require not just knowing poverty exists, but pinpointing areas of greatest need.

001)



# Motivation: The Question

---

Can **new data** and **new methods** improve our ability to **measure and target poverty**?

# Context: The “Mobile Revolution”

“Big” datasets still rare in poor countries



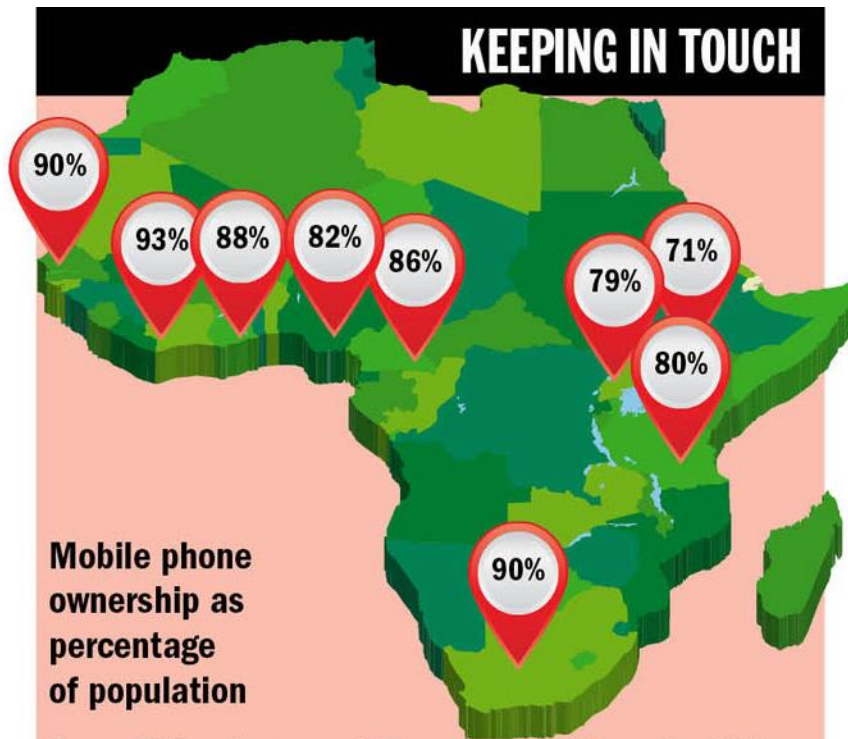
With one prominent exception: 

**3.5 Billion mobile subscribers** in developing countries



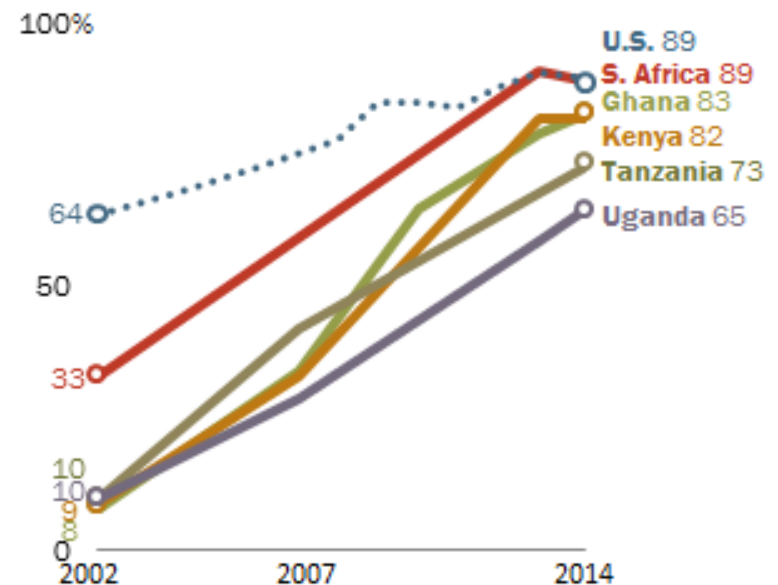
# Background: The “Mobile Revolution”

- From 10% to 90% “SIM penetration” in past 15 years
- 1 new user per second in Nigeria



## Cell Phone Ownership Surges in Africa

Adults who own a cell phone



# Predicting Poverty: Overview

**Question:** Can mobile phone metadata be used to measure poverty and vulnerability?

- Can *useful policy instruments* be constructed from those phone-based measurements?

**Idea:** Supplement anonymous logs with surveys

```

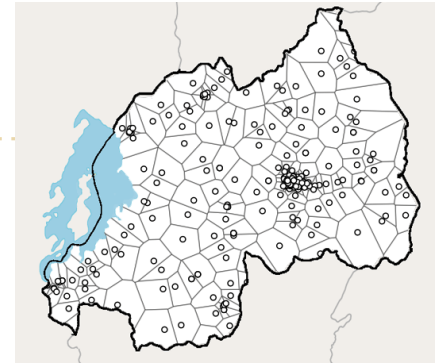
7|1|200750||15436289|21|08110333|08746577||01/01/2008 01:02:04||NBSC10|6351050012584
7|1|200751||15436296|21|00447958879882|03245598||01/01/2008 01:02:04||NBSC20|6351050
5|1|200752||15436303|22|08490767|08110333||01/01/2008 01:02:04||NBSC11|6351050008603
5|1|200753||15436319|22|08539819|08110333||01/01/2008 01:02:04||NBSC11|6351050006511
2|1|200754|2275372|1|15405099|08644907|03185229||01/01/2008 01:01:59|4|0|NBSC21|63510
2|1|200755|2275369|1|15406276|03179473|03162880||01/01/2008 01:01:56|8|0|NBSC21|63510
1|1|200756|2260849|1|15346499|1108743747|08448569||01/01/2008 01:01:25|38|0|NBSC21|S8FD
0|1|223401|2323394|1|15602242|08633883|03219934||01/01/2008 01:04:23|5|2|NBSC21|GRINDM|
5|1|223402||1563592|22|0860723|08110333||01/01/2008 01:04:28||NBSC21|6351050012994
5|1|223403||1563606|22|08431310|08110333||01/01/2008 01:04:28||NBSC21|6351050001313
5|1|223404||15636020|22|03119506|08110333||01/01/2008 01:04:28||NBSC11|6351050015147
5|1|223405||15636025|22|08520255|08110333||01/01/2008 01:04:28||NBSC11|6351050006954
2|1|223406|2310291|1|15547505|08618091|08847090||01/01/2008 01:03:48|40|0|NBSC11|6351
2|1|223407|2290567|1|15469250|08823243|08841803||01/01/2008 01:02:47|10|0|NBSC11|635
5|1|223408||15636091|22|08476273|08110333||01/01/2008 01:04:28||NBSC21|6351050001762
5|1|223409||15636100|22|08754900|08110333||01/01/2008 01:04:28||NBSC21|6351050013637
2|1|223410|2316505|1|15573404|08467584|03140554||01/01/2008 01:03:52|35|0|NBSC11|6351
5|1|223411||15636109|22|003249854307|0032495002530||01/01/2008 01:04:28||NBSC11|206
2|1|223412|2296485|1|15493712|08442497|08427583||01/01/2008 01:03:01|87|0|NBSC11|6351
1|1|223413|2310426|1|15549863|1108640680|08532308||01/01/2008 01:03:50|37|0|NBSC11|NBSC
1|1|223414|2316161|1|15572650|1108641505|08807256||01/01/2008 01:04:00|27|0|NBSC11|S8FD
7|1|223477||15636441|21|08110334|08568931||01/01/2008 01:04:28||NBSC20|6351050008892
4|1|223478|2314273|1|15564440|1108620541|08879103||01/01/2008 01:04:06|21|0|NBSC11|NBSC
0|1|223479|2324828|1|15607541|03217934|456||01/01/2008 01:04:07|21|0|NBSC21|11|131|S8
4|1|223480|2328605|1|15625293|1108774113|08508275||01/01/2008 01:04:27|1|0|NBSC21|NBSC2
2|1|223481|2310704|1|15549576|108222832|08686319||01/01/2008 01:03:38|50|0|NBSC11|6351
1|1|223482|2310931|1|15554565|1108743747|08448569||01/01/2008 01:01:25|38|0|NBSC21|S8FD
4|1|223483|2303553|1|15520311|1108450571|03087322||01/01/2008 01:03:26|62|0|NBSC11|NBSC
2|1|223484|127961|1|15870912|08125108|08890756||01/01/2008 01:04:11|7|0|NBSC11|NBSC20
1|1|223485|127961|1|15870912|08125108|08890756||01/01/2008 01:04:11|7|0|NBSC11|NBSC20
1|1|223486|127961|1|15870912|08125108|08890756||01/01/2008 01:04:11|7|0|NBSC11|NBSC20
2|1|223487|2323697|1|15603148|10312148|08819944||01/01/2008 01:04:17|12|0|NBSC11|6351
4|1|223488|1317246|1|15770241|1103129202|08499326||01/01/2008 09:41:23|13931|NBSC21|NB
5|1|223557||15637118|22|08636277|08110333||01/01/2008 01:04:29||NBSC11|6351050013614
1|1|223558|2312854|1|15556844|1108641415|08809580||01/01/2008 01:03:51|37|0|NBSC11|S8FD
1|1|223559|2312854|1|15556844|1108641415|08809580||01/01/2008 01:03:51|37|0|NBSC11|S8FD
    
```



Call Detail Records

Individual surveys

# Data: Details



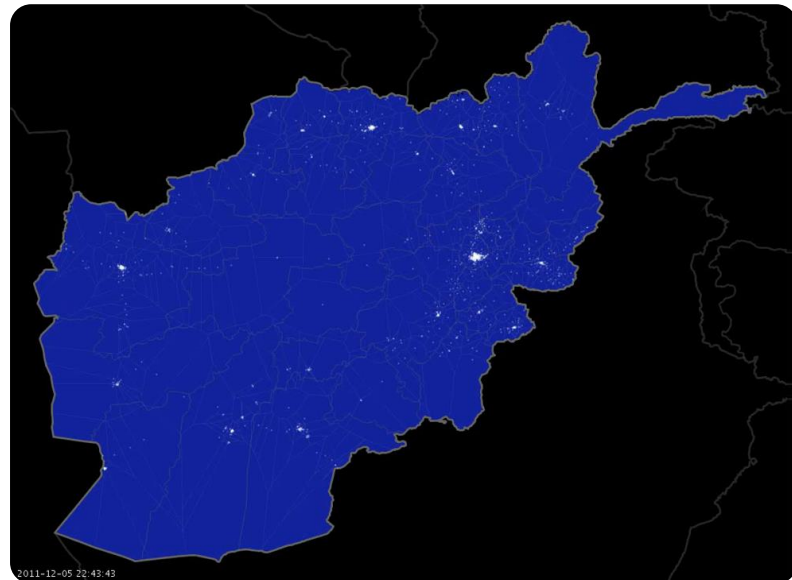
## 1. Call Detail Records (CDR)

- **10 terabytes** of data from monopoly operator
- **Calls, SMS, Recharges:** sender, receiver, time, duration, location
- No demographic or socioeconomic information

A-Party-ID	B-Party-ID	Date	Time	Duration	A-Party-Cell	...
979ae8cd	97939b87	2014-01-04	22:00:11	42	2837	...

2. Phone surveys

3. Validation data



# Data: Details

---

## 1. Call Detail Records (CDR)

## 2. Phone surveys

856 respondents

- Geographically stratified
- Representative of “active subscribers”
- Informed consent, no PII collected

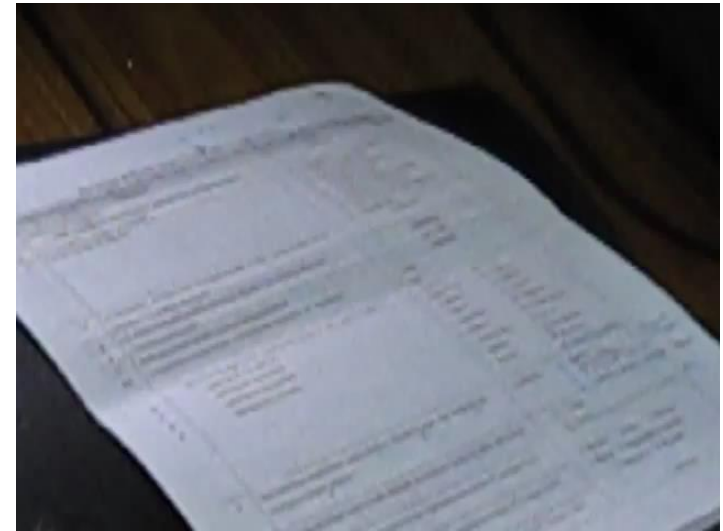
25-minute survey

- Today’s focus: self-reported assets and housing characteristics

Ongoing work

- Panel surveys in Afghanistan ( $N=1,200$ ) and Ghana ( $N=5,000$ )

## 3. Validation data



# Data: Details

---

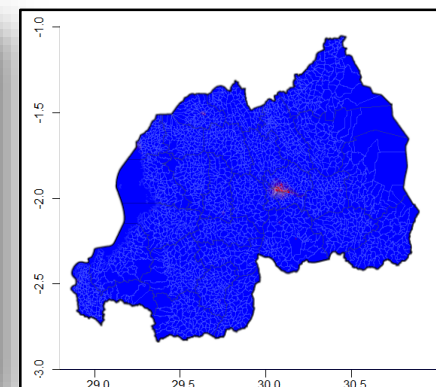
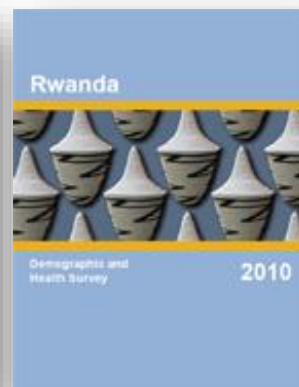
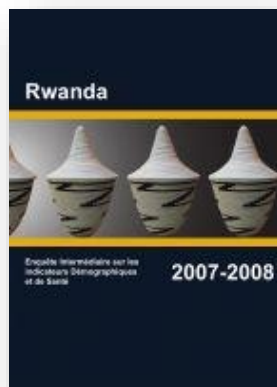
1. Call Detail Records (CDR)
2. Phone surveys

## 3. Validation data

### Rwanda Demographic and Health Surveys

- DHSV: 7,377 households, 2007 - 2008
- DHSVI: 12,792 households, 2010 – 2011

### Satellite luminosity data



# Methods: Overview

---

## 1. Feature engineering

[\[Details\]](#)

Convert call data into quantitative metrics

## 2. Supervised learning

[\[Details\]](#)

Trained on sample of survey respondents (N=856)

## 3. Out of Sample Prediction

[\[Details\]](#)

Generate predicted values for all 1.5M subscribers

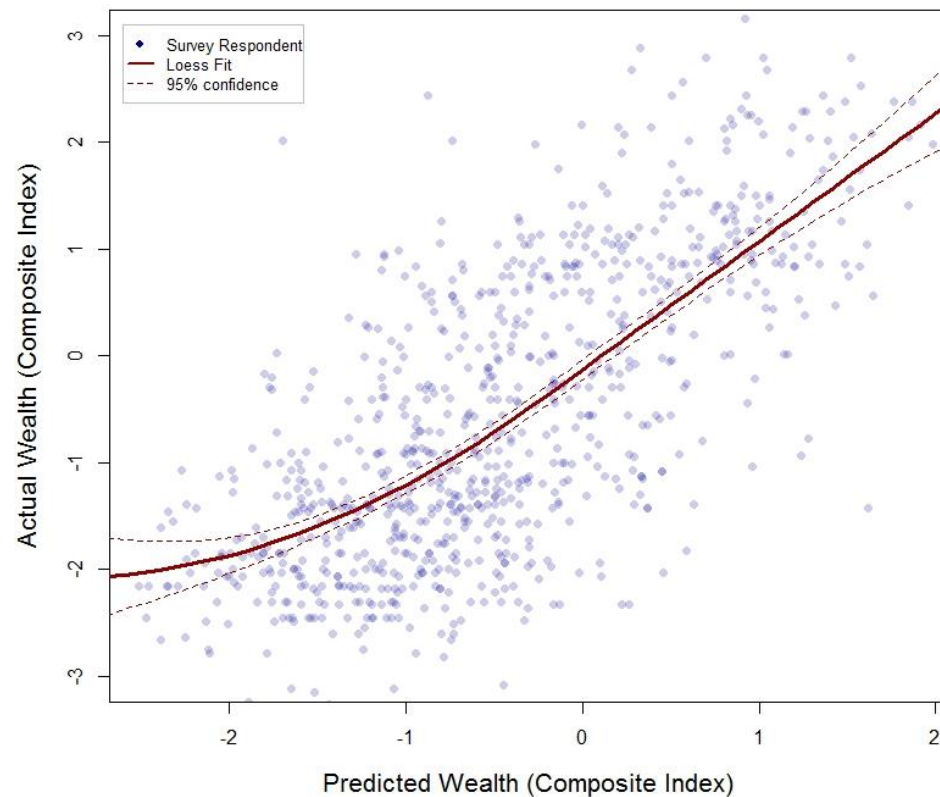
## 4. Validation

[\[Details\]](#)

Compare aggregated predictions to DHS survey data

# Results: Predicting composite wealth

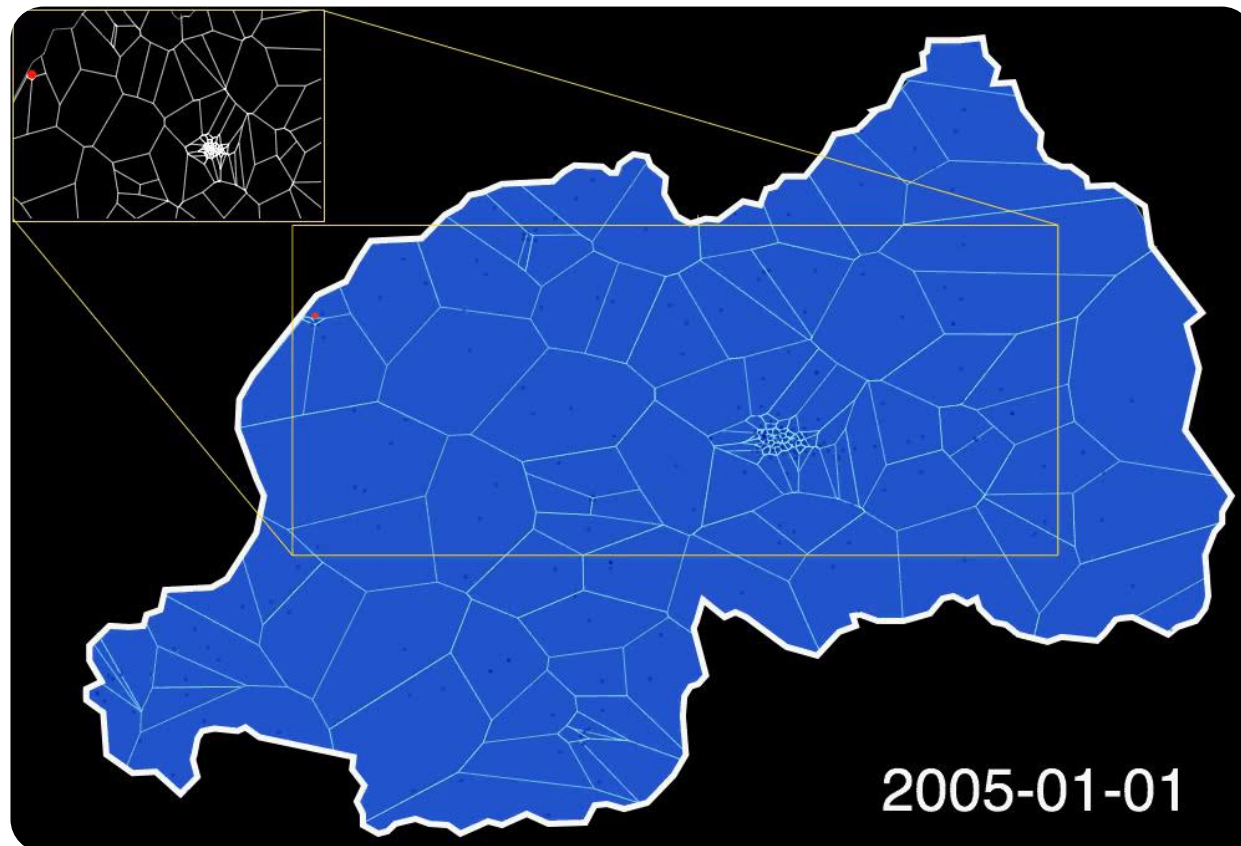
	Elastic Net		Random Forest	
	$r$	$R^2$	$r$	$R^2$
Optimal DFA-based model	0.68	0.46	0.63	0.40
“Intuitive” 5-feature model	0.44	0.20	0.37	0.14
Single-feature model	0.61	0.38	0.46	0.22

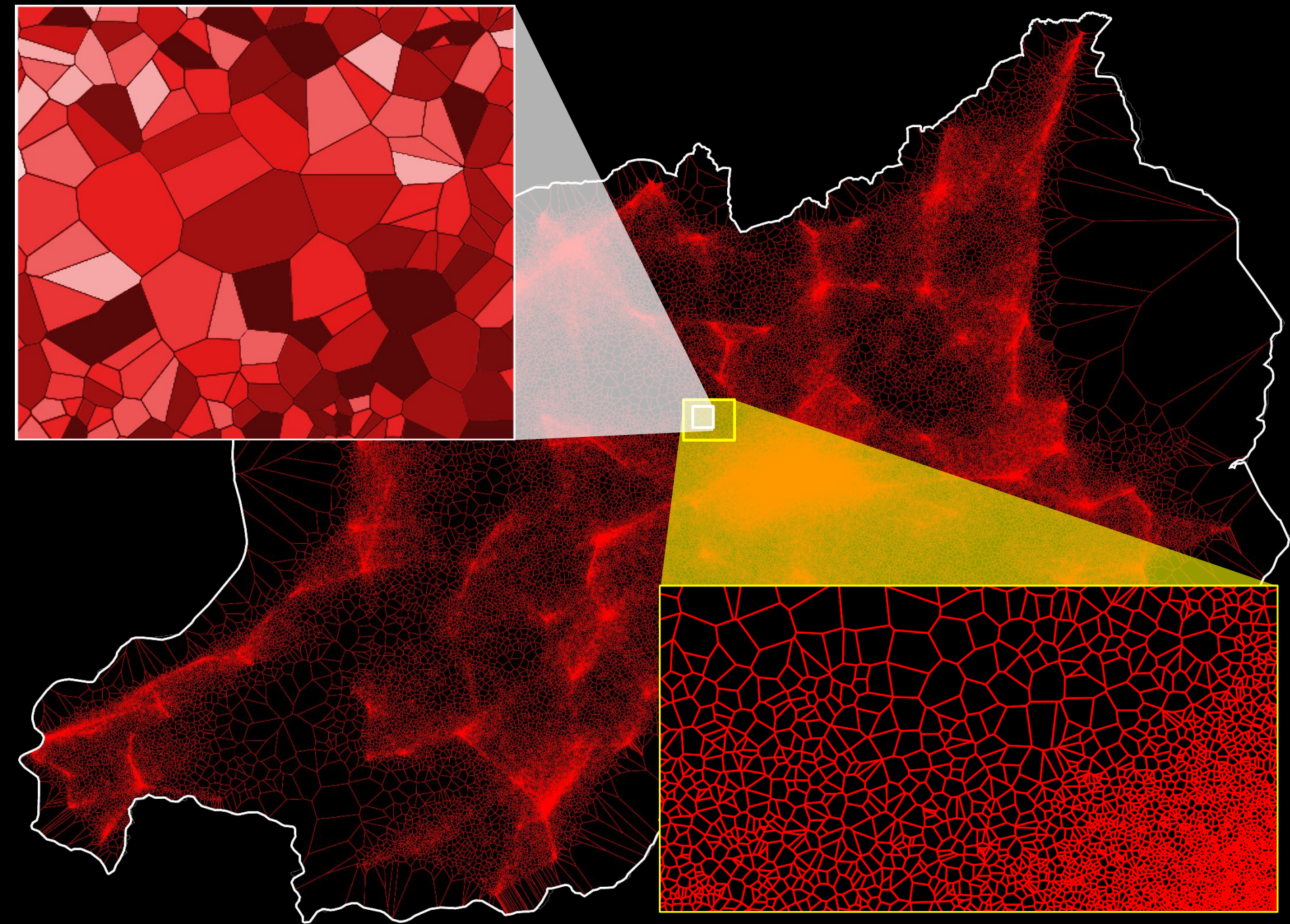


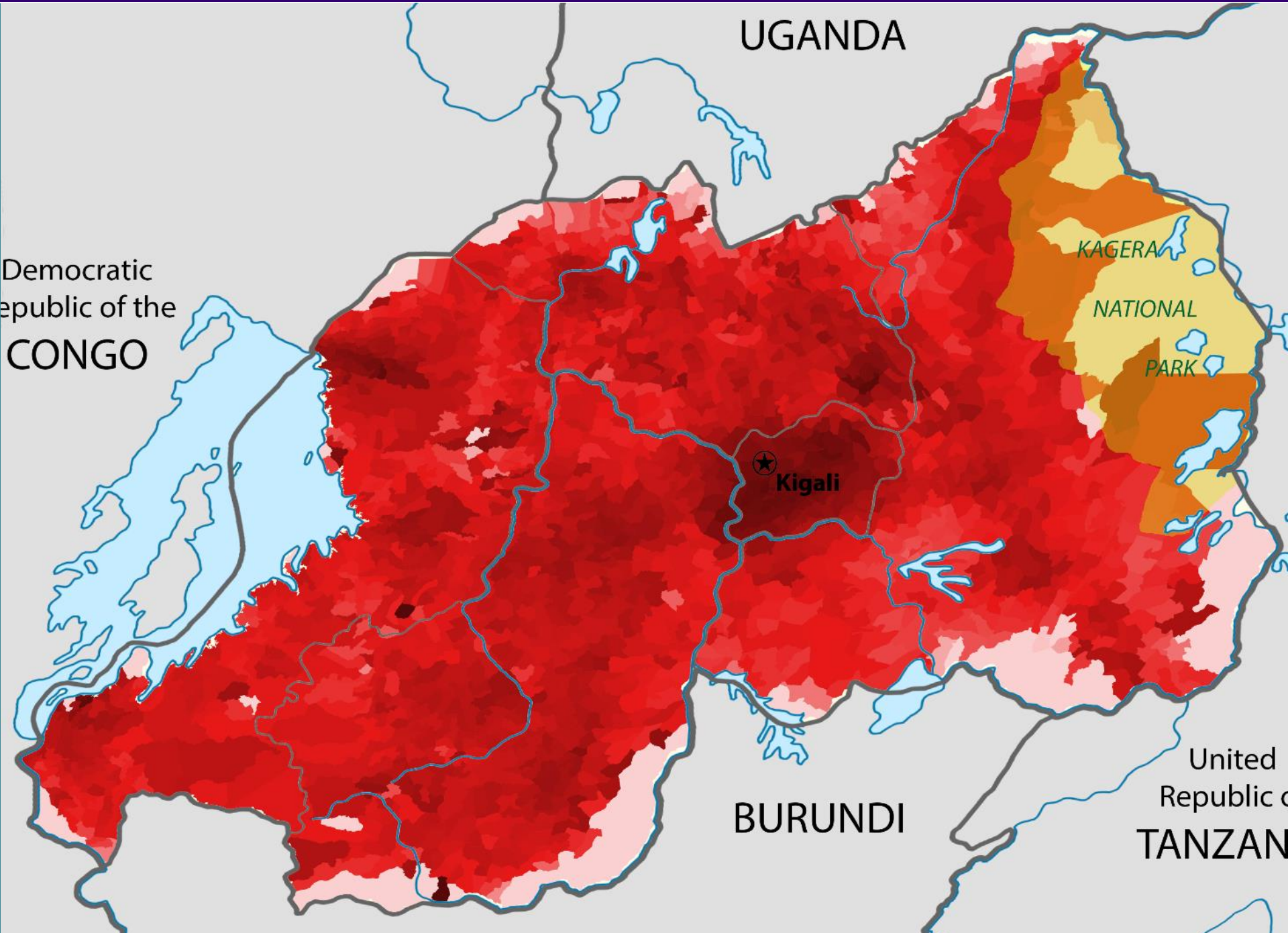
[\[Details\]](#)

# Results: Out-of-sample projections

Use model trained on small sample ( $n=856$ ) to predict wealth of an entire nation ( $n=\text{millions}$ )

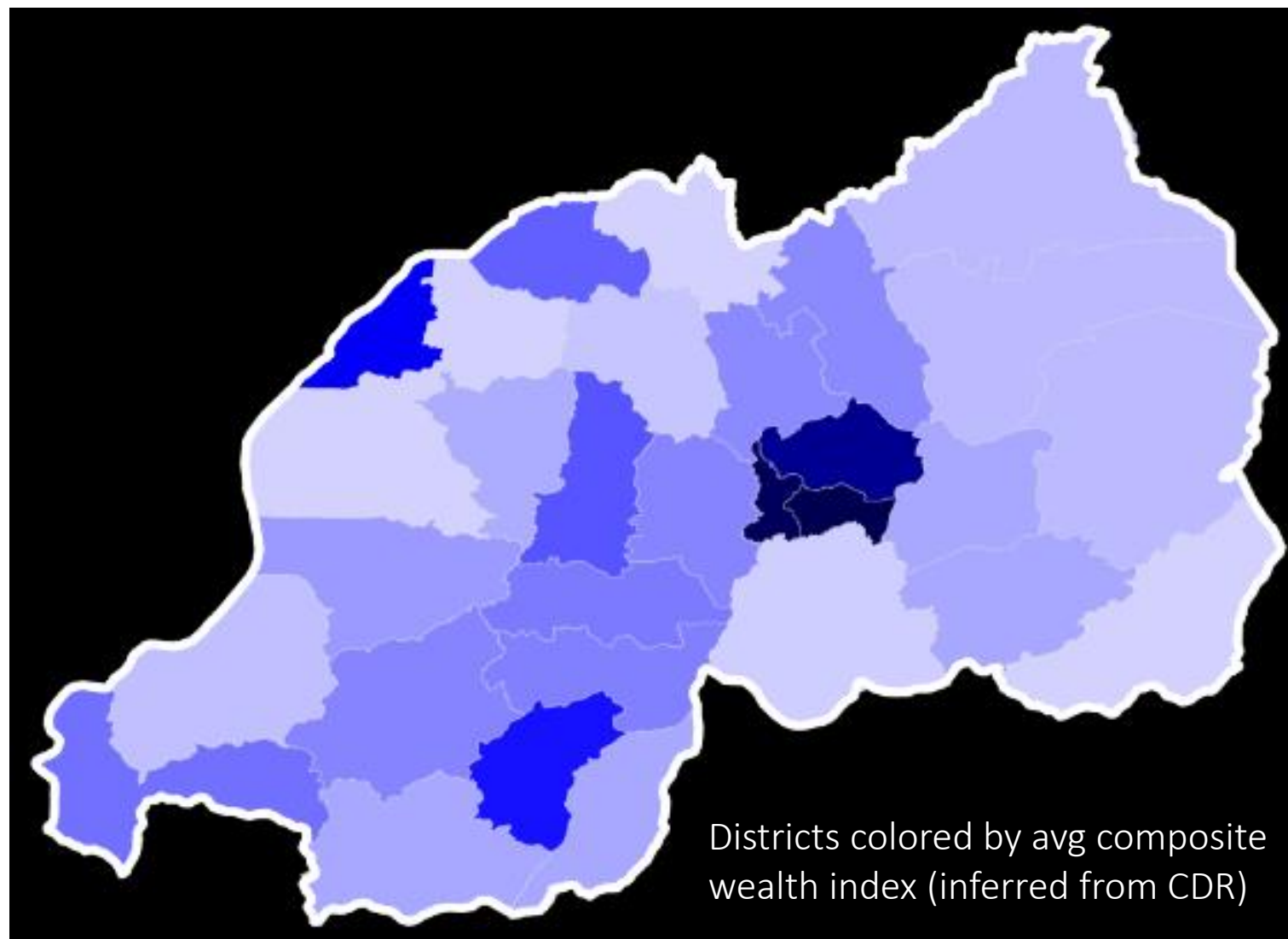






# Results: Out-of-sample aggregates

---

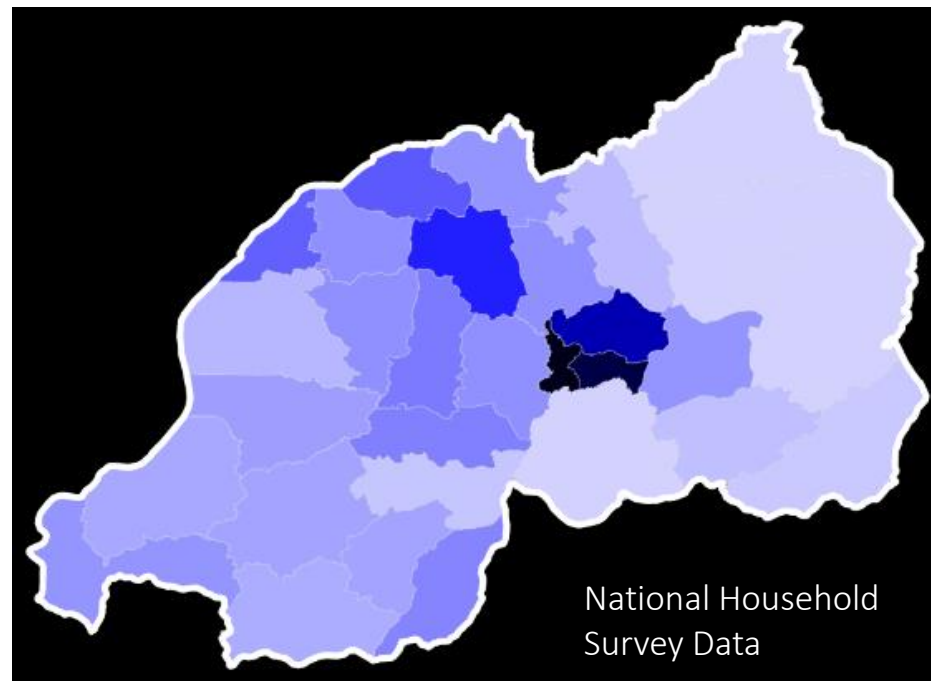
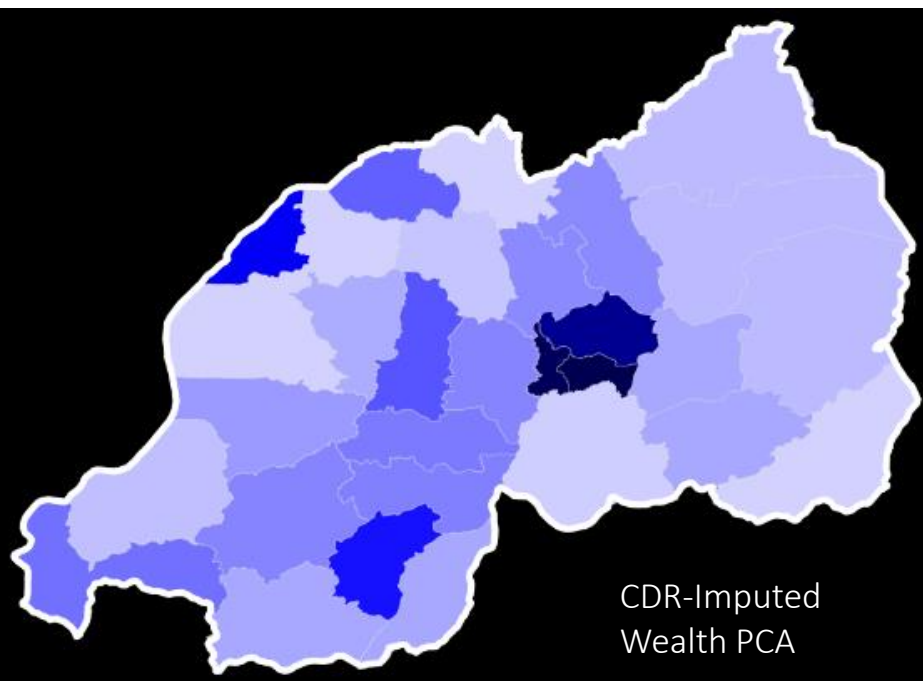


# Validation: Visual comparison

## District comparison ( $N=30$ )

Calculate average wealth in 2010 DHS

- Uses same principal component projection from CDR

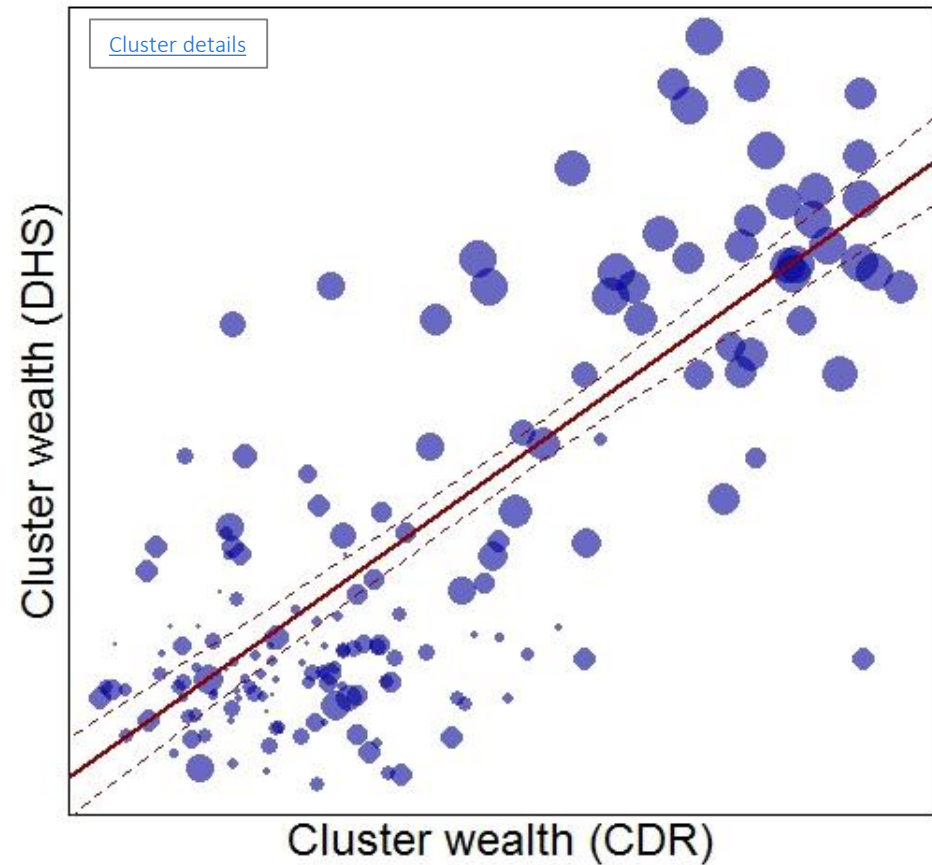
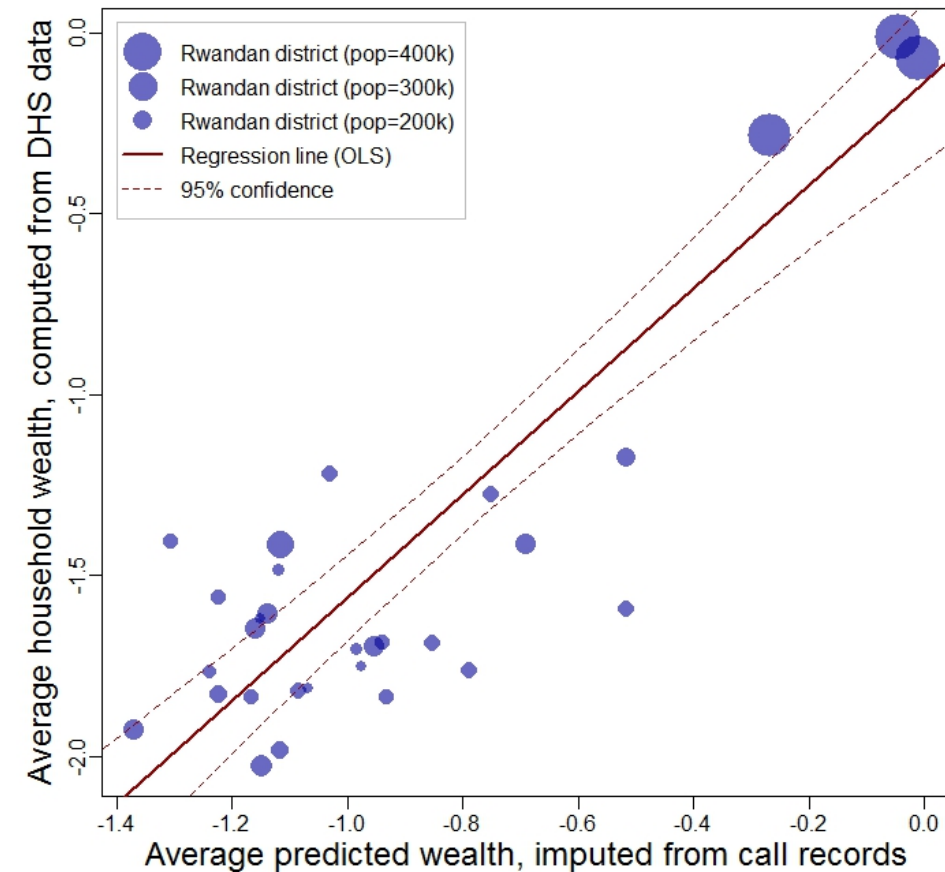


# Validation: Quantitative comparison [\[urban/rural\]](#)

**CDR Vs. DHS** (30 districts, MP-owning HH's):  $\rho = 0.916$

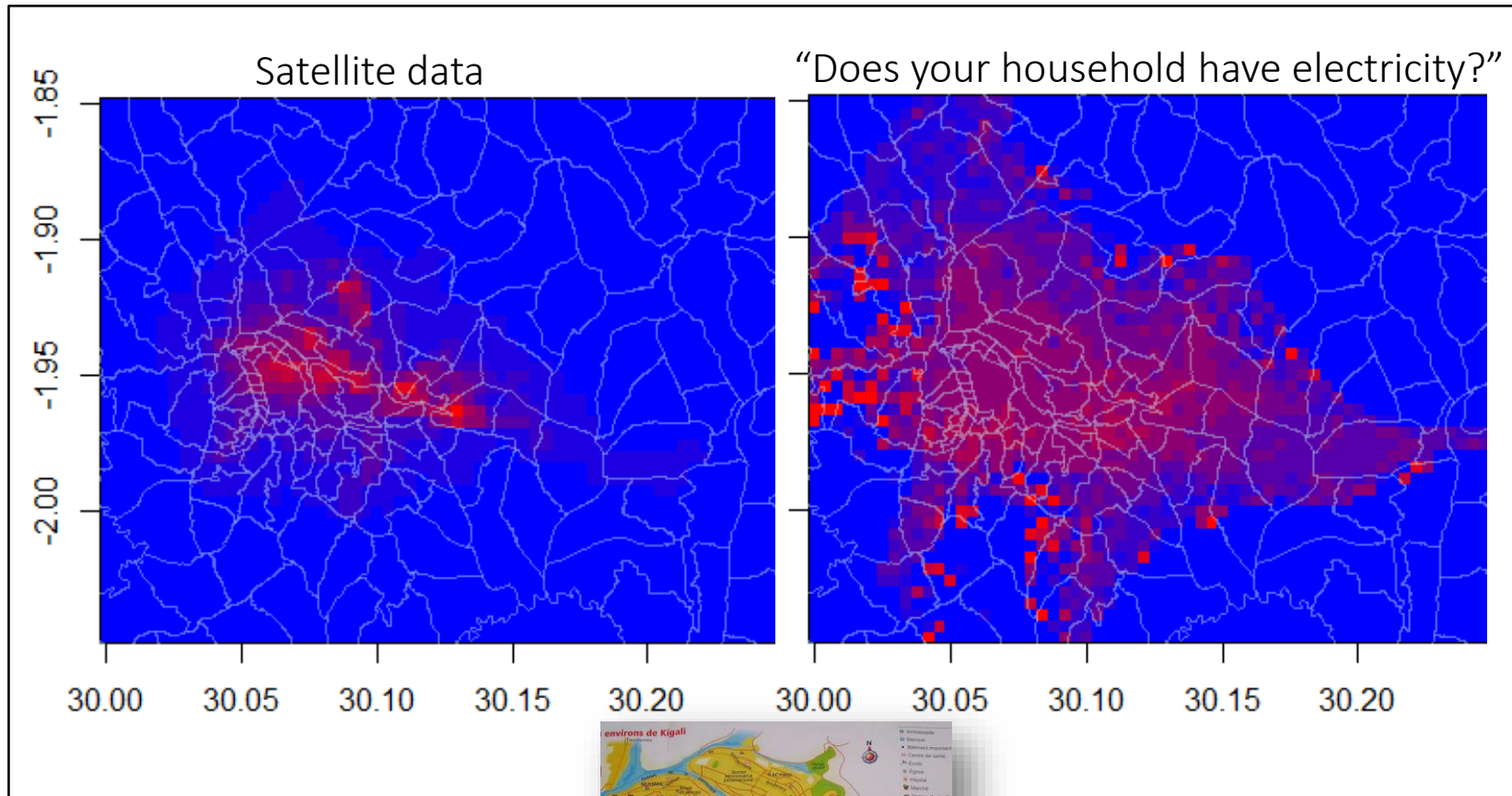
CDR Vs. DHS (30 districts, all HH's):  $\rho = 0.917$

CDR Vs. DHS (492 clusters, all HH's):  $\rho = 0.799$



# Validation: Micro-level validation?

[\[Details\]](#)



# Application 1: Interim national statistics

“Poor numbers”

Dean Jolliffe and Peter Lanjouw (2014)

**Table 6.1** Countries with outdated censuses

Country	Year of last census	Years since last census
Somalia	1986	28
Congo, Dem. Rep. (planned for 2015)	1984	30
Eritrea	1984	30
Afghanistan (2011 ongoing Socio-Demographic and Economic Survey by province)	1979	35
Lebanon	1943	71

*Source:* United Nations Department of Economic and Social Affairs, Population Division, based on analysis in July 2014 of the implementation of the World Population and Housing Census program since 1948.

These data are **critical inputs to public policy**

- Infrastructure investment, aid distribution, political redistricting

CDR-based method **creates new options**

[\[details\]](#)

- Rwanda: **CDR estimates outperform** 5-year old national survey
- Advantages: 500 times cheaper, 20 times faster

## Application 2: Targeting of humanitarian aid

Cash transfer programs are one of the world's **most common social protection programs**

How to identify eligible recipients?

- Nearly impossible to verify informal income
- Proxy means tests (Skoufias et al. 2001, Tabor 2002)
- Community-based targeting (Alderman 2002)
- Satellite imagery (Haushofer & Shapiro 2013, Jean et al. 2016)
- Serious limitations to all of these (Alatas et al. 2010, Niehaus 2013)

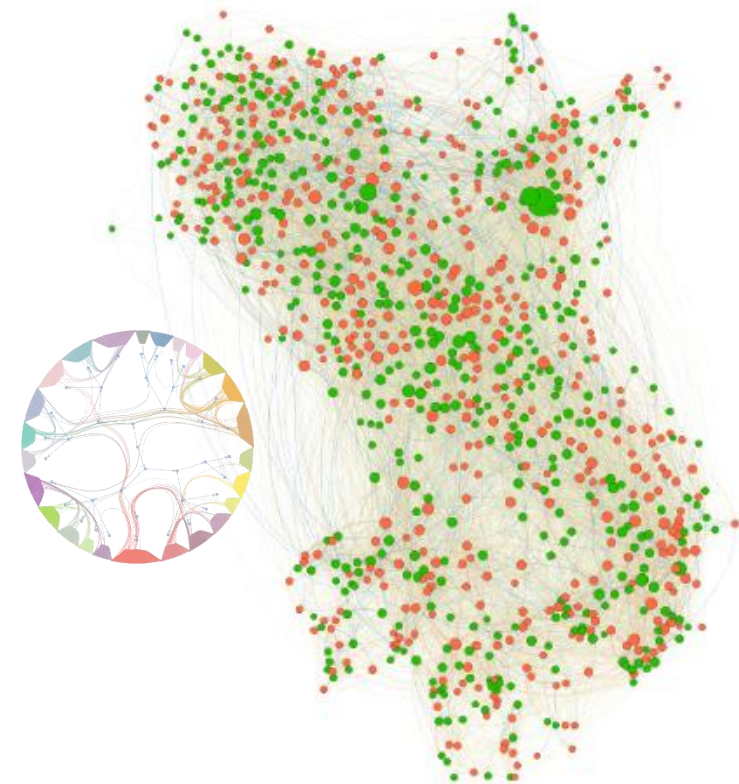
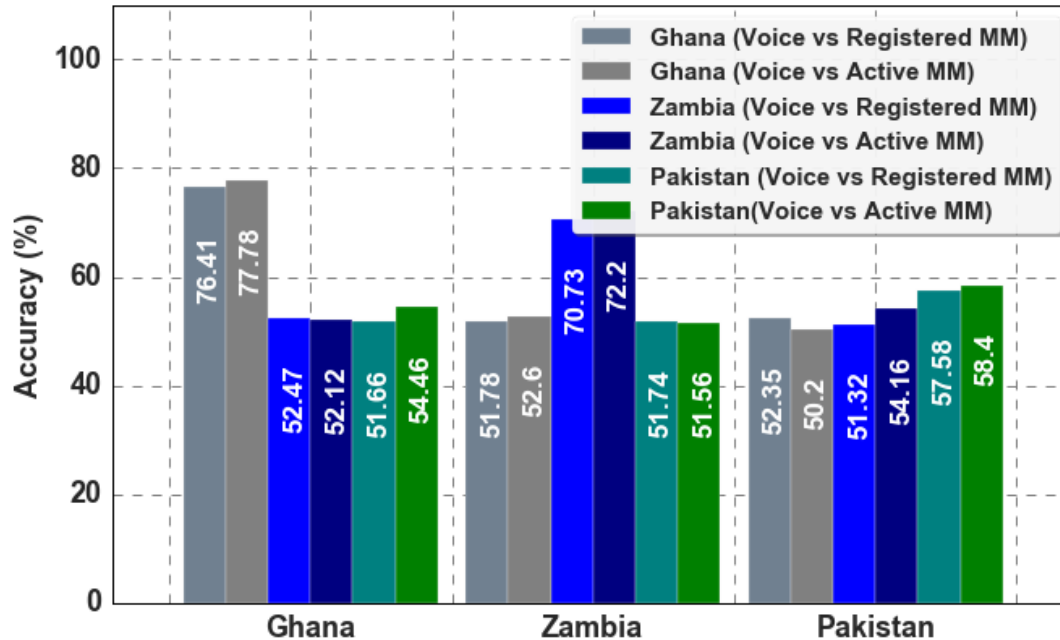


**Can CDR improve targeting in data-poor contexts?**

# Application 3: Financial services for the poor

Who is most likely to **adopt and use** digital financial services?

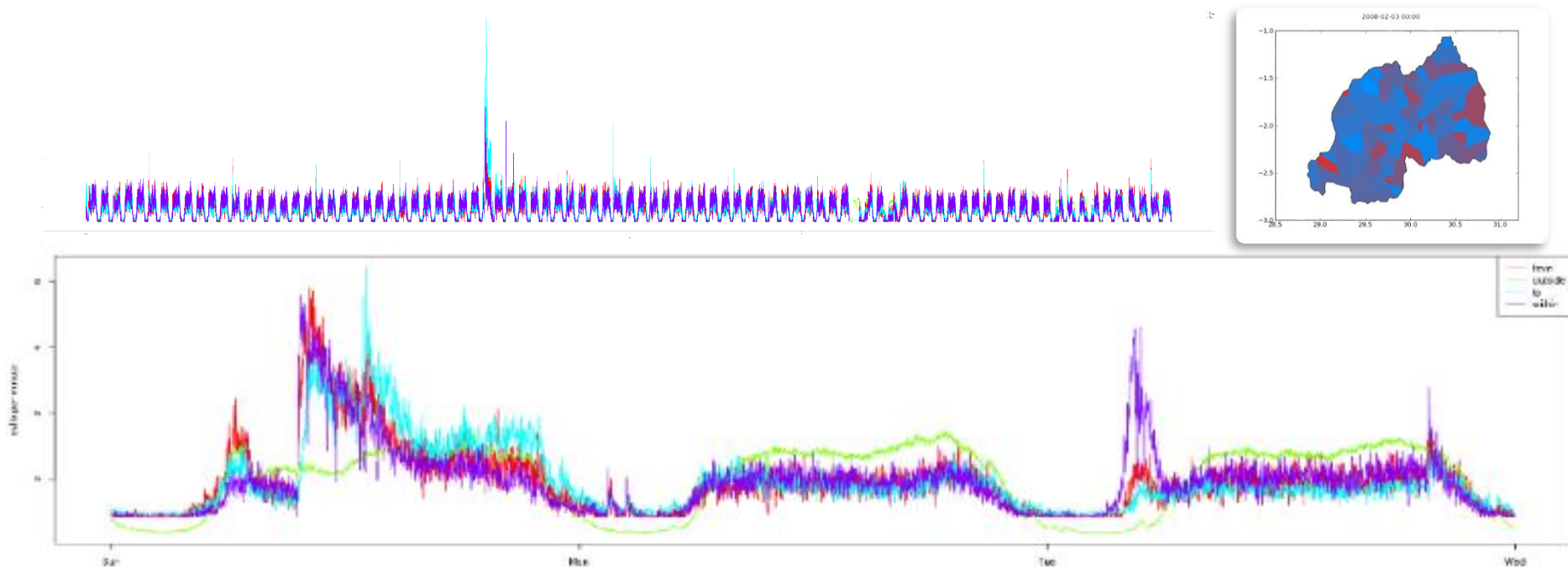
Predicting Mobile Money Adoption



# Application 4: Real-time impact evaluation

Is it possible to **monitor** the state of an entire nation in **real time**?

- To detect shocks and monitor vulnerability
- To track economic trends
- To **evaluate policy impact**



# Limitations and challenges

---

## Non-representative mobile adoption

- Not critical in Rwanda, but may matter in other contexts

## “Algorithmic Dynamics”

- Requires re-basing of model, additional research

## Strategic incentives

## Data access and privacy

# Summary

---

## Primary results

1. **New method** for feature engineering
2. Can predict **individual poverty** from CDR
3. **Micro-predictions validated** by macro data
4. Near-term **policy applications**
  - Measurement
  - Targeting
  - Financial services for the poor
  - Evaluation

A network graph visualization overlaid on a map of a region. The nodes are represented by white and light blue circles of varying sizes, connected by a dense web of thin blue lines. The background is a dark blue map with white outlines of landmasses. A small orange square highlights a specific node in the central part of the network.

Thanks for your attention!

Questions, comments, suggestions?

[joshblum@uw.edu](mailto:joshblum@uw.edu) / [www.jblumenstock.com](http://www.jblumenstock.com)

