Tapping into Facebook's Advertising Audience Estimates

Ingmar Weber

@ingmarweber

Facebook as a Research Tool Webinar Series

April 12, 2022

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Fantastic Collaborators

Joao Palotti
QCRI

Masoomali Fatehkia
QCRI

Ridhi Kashyap
University of Oxford

Also: Reham Al Tamime (QCRI), Ferda Ofli (QCRI), Vedran Sekara (UNICEF), Manuel Garcia Herranz (UNICEF), Kiran Garimella (MIT), Karri Haranko (Aalto University), Francesco Rampazzo (University of Southampton), Marzia Rango (IOM), Emilio Zagheni (MPI for Demographic Research), …

+ more collaborators at QCRI, UN agencies, NGOs, …

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

If you're not paying for the product, you are the product

Facebook et al. collect a lot of information about their users

This information is used to build up user profiles with certain attributes

Advertisers can then use these attributes to target their ads

Example: "Show my advertisement to female Facebook users living in London, aged 25-29, who lived in Poland, and who use an iOS device"

To help with the campaign and budget planning, Facebook provides *audience estimates*

Previous example: "There are 2,600 monthly active users matching these criteria"

This gives a *no-cost, real-time census over the ~ 3 billion Facebook (et al.) users*

How can we use such data for research?

Ads Manager

Search Business

Audience

Define who you want to see your ads. Learn more.

Locations

Qatar

(25.2014, 51.4473) + 2 km

Age 13 - 65+

Gender All | Men | Women

INCLUDE people who match at least ONE of the following

Behaviours > Ex-pats

Lived in India (formerly Expats – India)

Lived in Nepal (formerly Ex-pats – Nepal)

Add demographics, interests or behaviours | Suggestions | Browse

Detailed targeting

and MUST ALSO match at least ONE of the following

Behaviours > Mobile Device User > All Mobile Devices by Operating System

Facebook access (mobile): Android devices

Add demographics, interests or behaviours | Suggestions | Browse

Audience size

Specific — Broad

Your audience selection is broad. This requires a large budget.

Potential reach: 42,000 people

Estimated daily results

Reach

2.7K-17K

Post Engagement

82-510

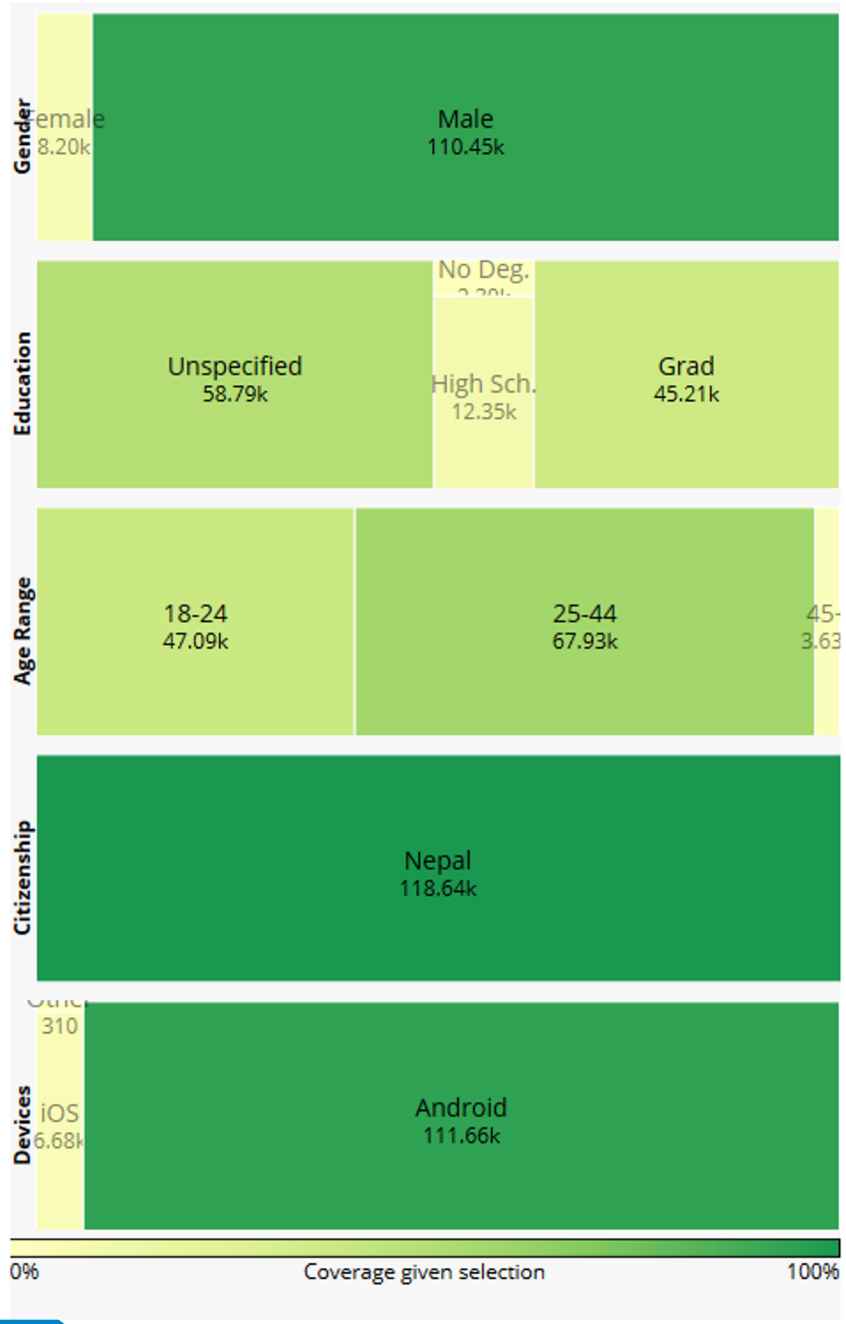
Removing Instagram and Audience Network may result in 45% fewer Post Engagement, based on your past campaign performance. We recommend choosing automatic placements for the best results.

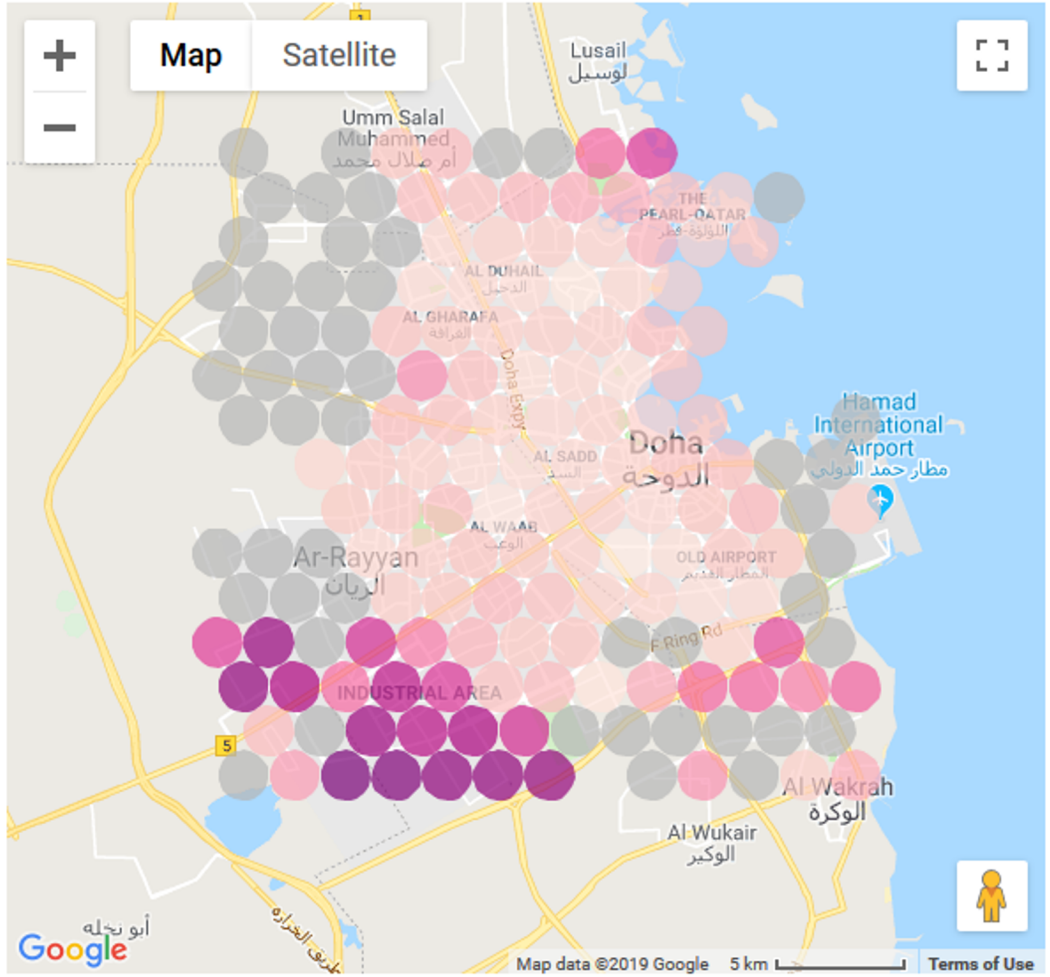
The accuracy of estimates is based on factors such as past campaign data, the budget you've entered and market data. Numbers are provided to give you an idea of performance for your budget, but are only estimates and don't guarantee results.

Were these estimates helpful?

QCRI معهد قطر لبحوث الحوسبة Computing Research Institute

جامعة حمد بن خليفة HAMAD BIN KHALIFA UNIVERSITY





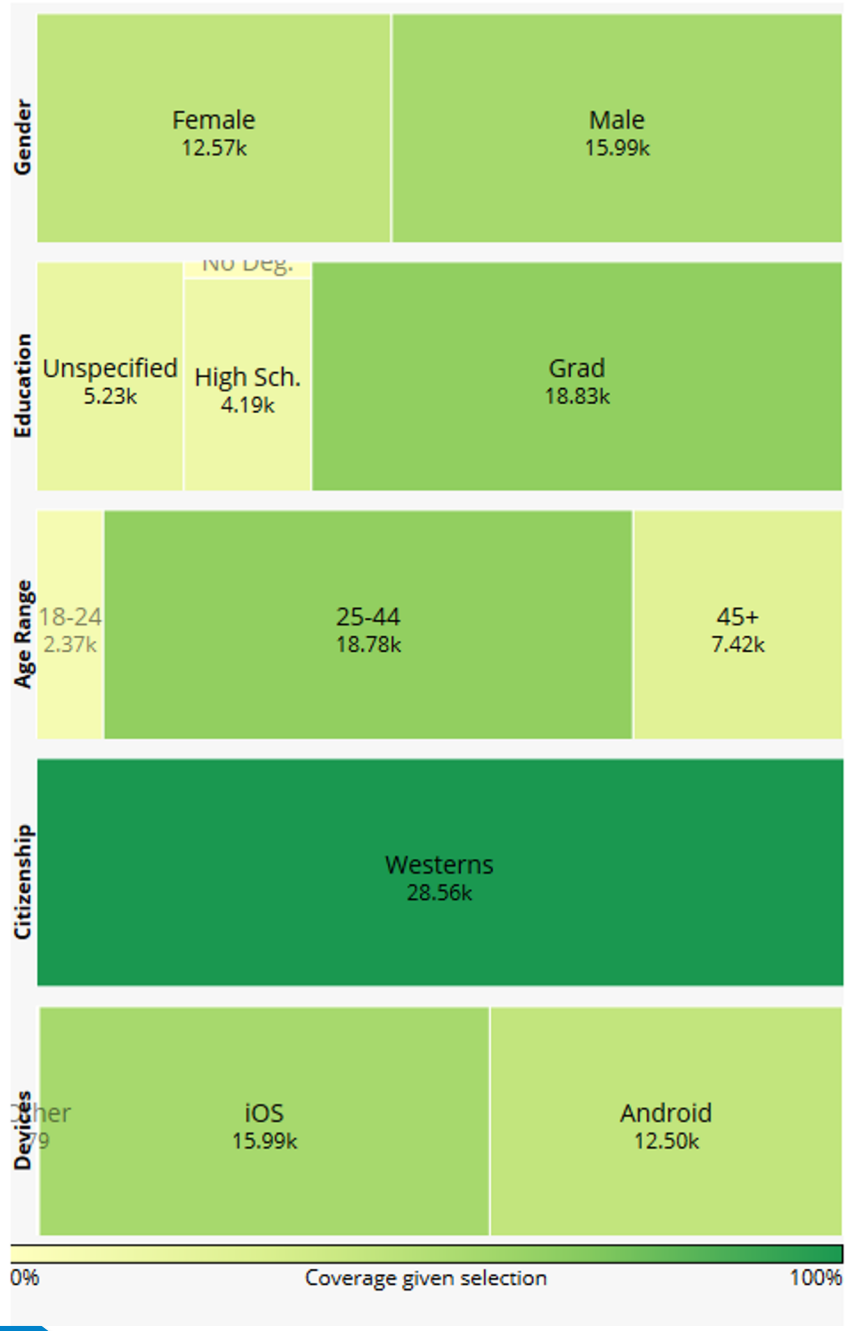
All Locations | No Locations

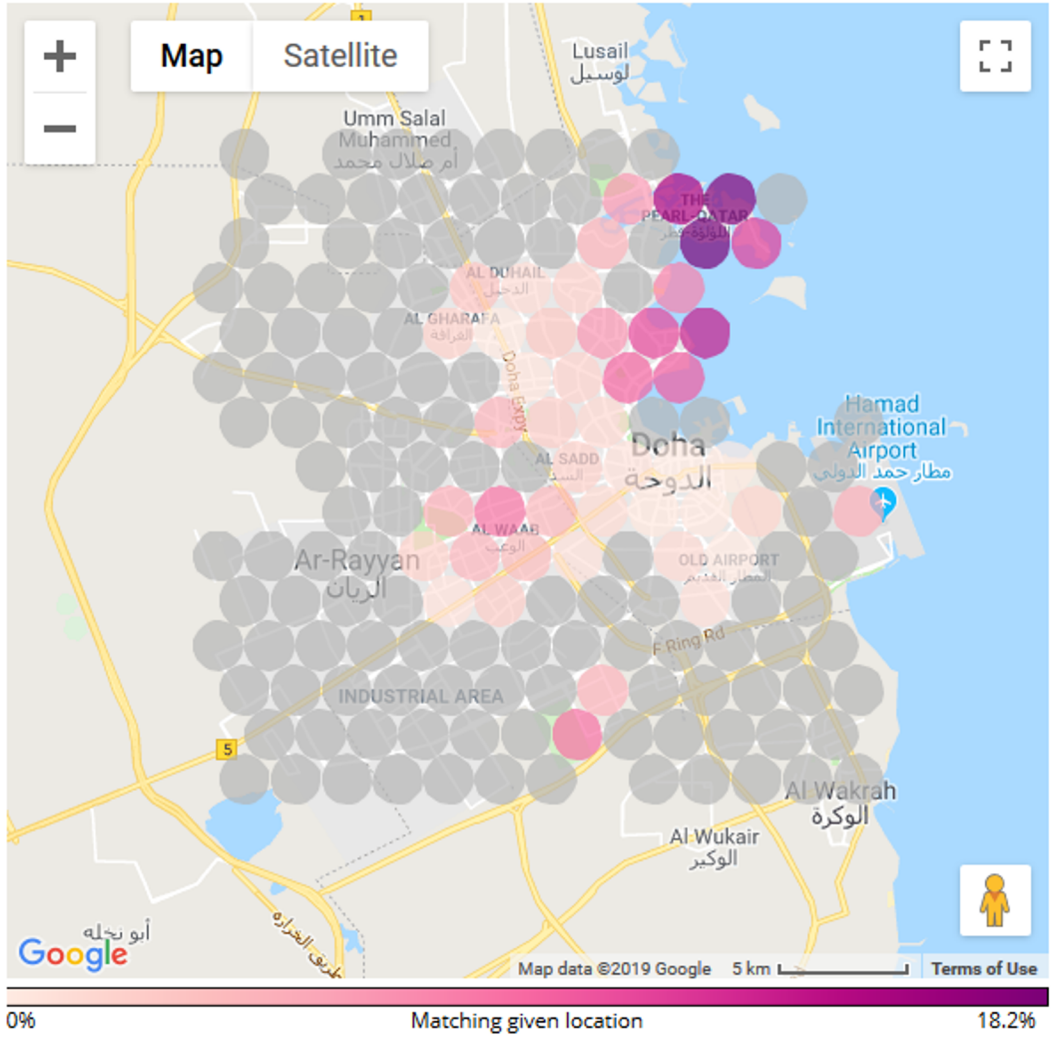
Share what you see:

http://fb-doha.qcri.org

QCRI معهد institute

جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY





0% Matching given location 18.2%

All Locations | No Locations

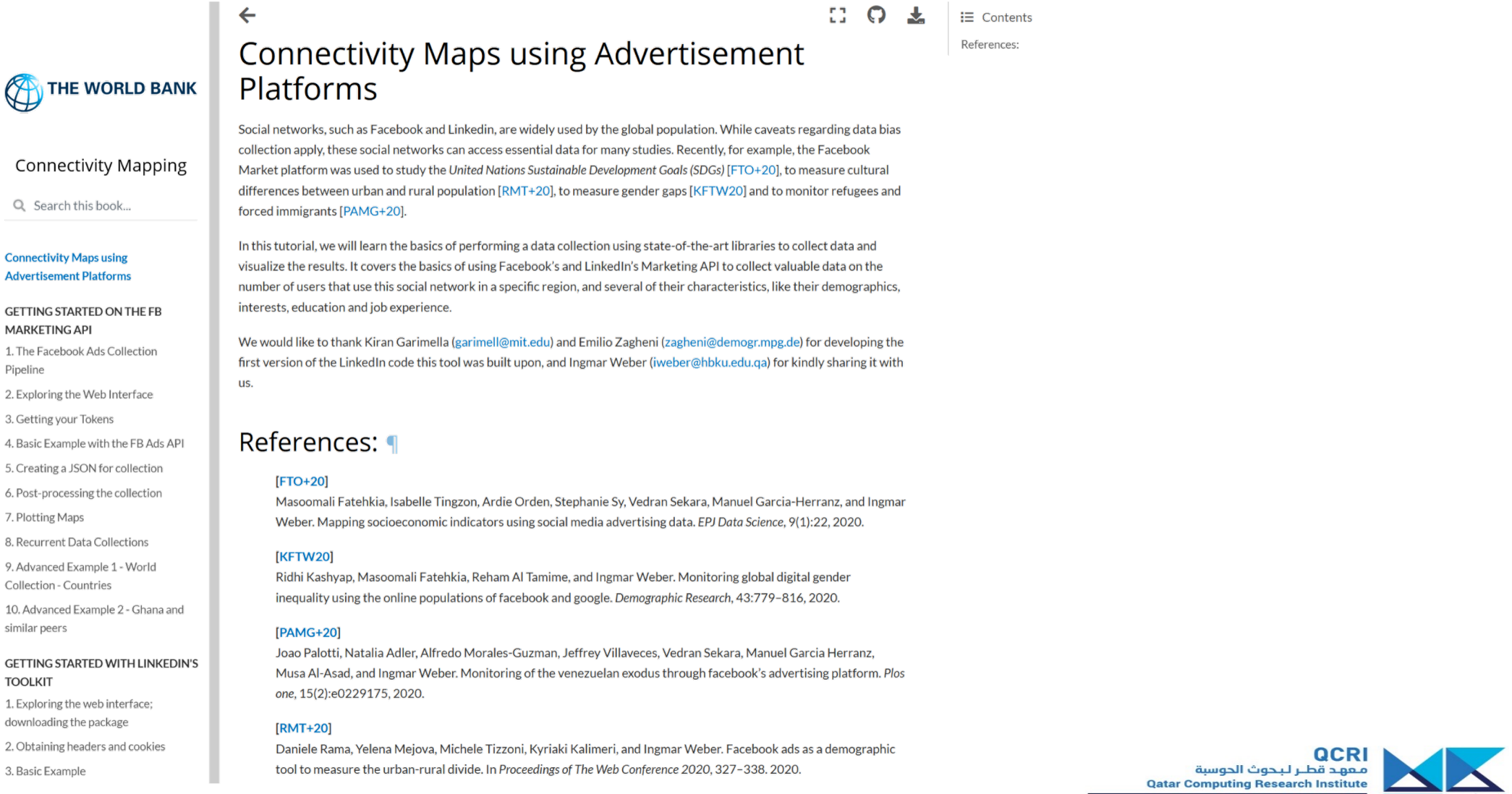
Share what you see:

http://fb-doha.qcri.org

QCRI معهد قطر Institute

جامعة حمد بن خليفة HAMAD BIN KHALIFA UNIVERSITY

Well-Documented API and Wrapper



https://worldbank.github.io/connectivity_mapping/intro.html

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Examples of Targeting Attributes

		Facebook	LinkedIn
Self-declared	Basic	Age, gender	Employment history
	Advanced	Education level, home town, friends	Contacts, skills

This is where most of the inter-app/site tracking comes in

Homework: Access Your Own Data

- Facebook

https://www.facebook.com/your_information

https://www.facebook.com/adpreferences/ad_settings

https://www.facebook.com/off-facebook-activity

- LinkedIn

https://www.linkedin.com/psettings/member-data

Goal: Use Advertising Data to Fill Data Gaps

Data on development indicators is often *slow* (re decennial census), *coarse* (re Uttar Pradesh), and *aggregate* (re women)

Combine different data sources to:

- Improve *recency* (e.g. before/after covid-19)
- Improve *granularity* (e.g. sub-national variation)
- Improve *disaggregation* (e.g. by gender)

Next: Example Applications

Monitor international migration

- FB users who “lived in [country X]” now living elsewhere

Map poverty

- iOS devices more prominent in wealthier areas

Track digital gender gaps

- FB gender gaps mirror internet gender gaps

MONITORING THE VENEZUELAN EXODUS

UNHCR
The UN
Refugee Agency

unicef

iMMAP

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Background on the Venezuelan Exodus

Annual inflation in Venezuela > 10,000,000% (est. 2019, IMF)

Unemployment > 40% (est. 2019, IMF)

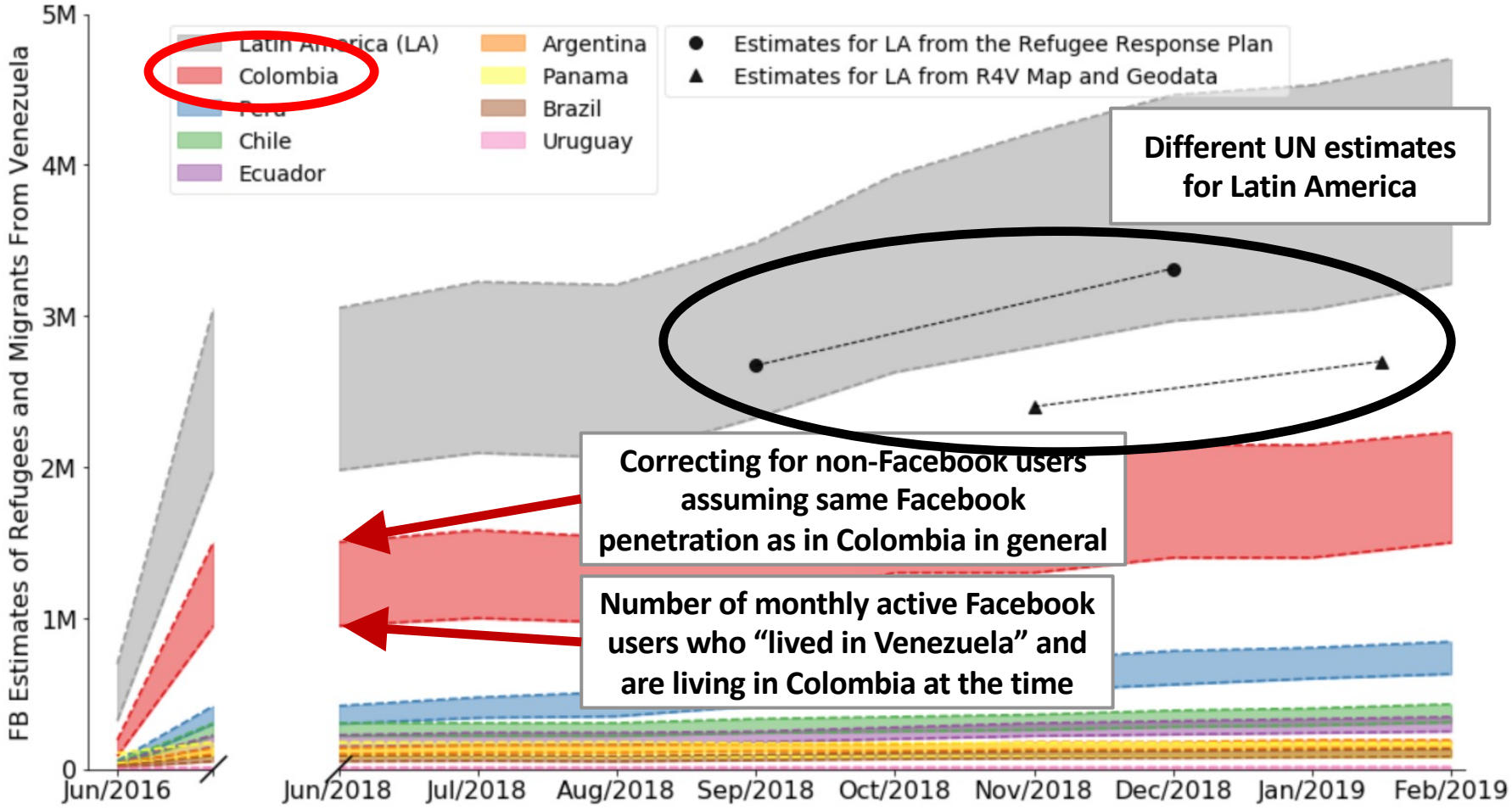
Minimum wage pays < 1000kcal/day

> 4.5 million people have left Venezuela since ~2015

Main destinations: Columbia, Peru, Ecuador, Chile, Brazil

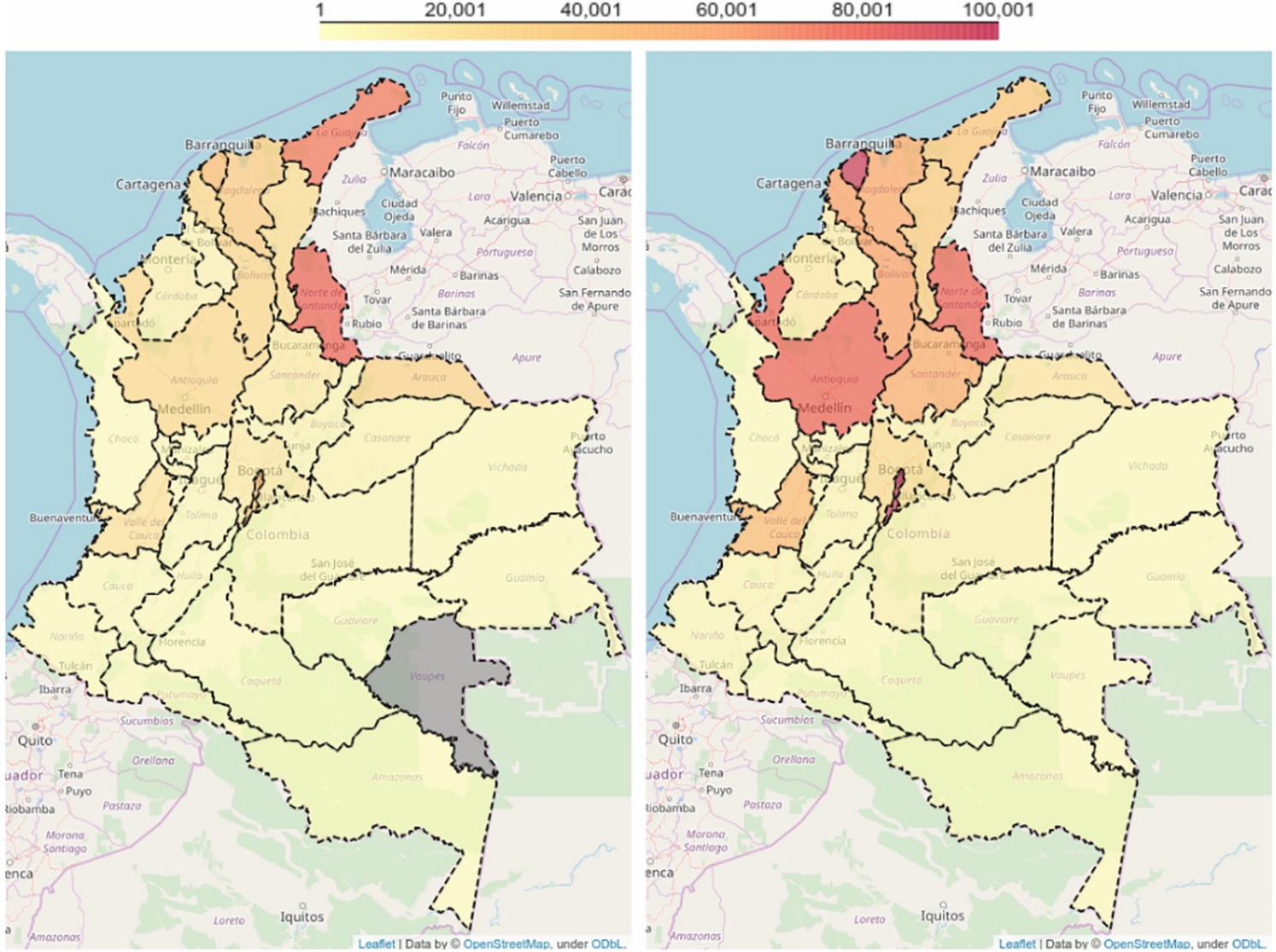
Migration data based on voluntary self-registration is inaccurate, outdated and misses demographic sub-groups

Monitoring Trends in Real-Time



QCRI معهد قطر لبحوث g Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Validation w/ (Few) Available Data

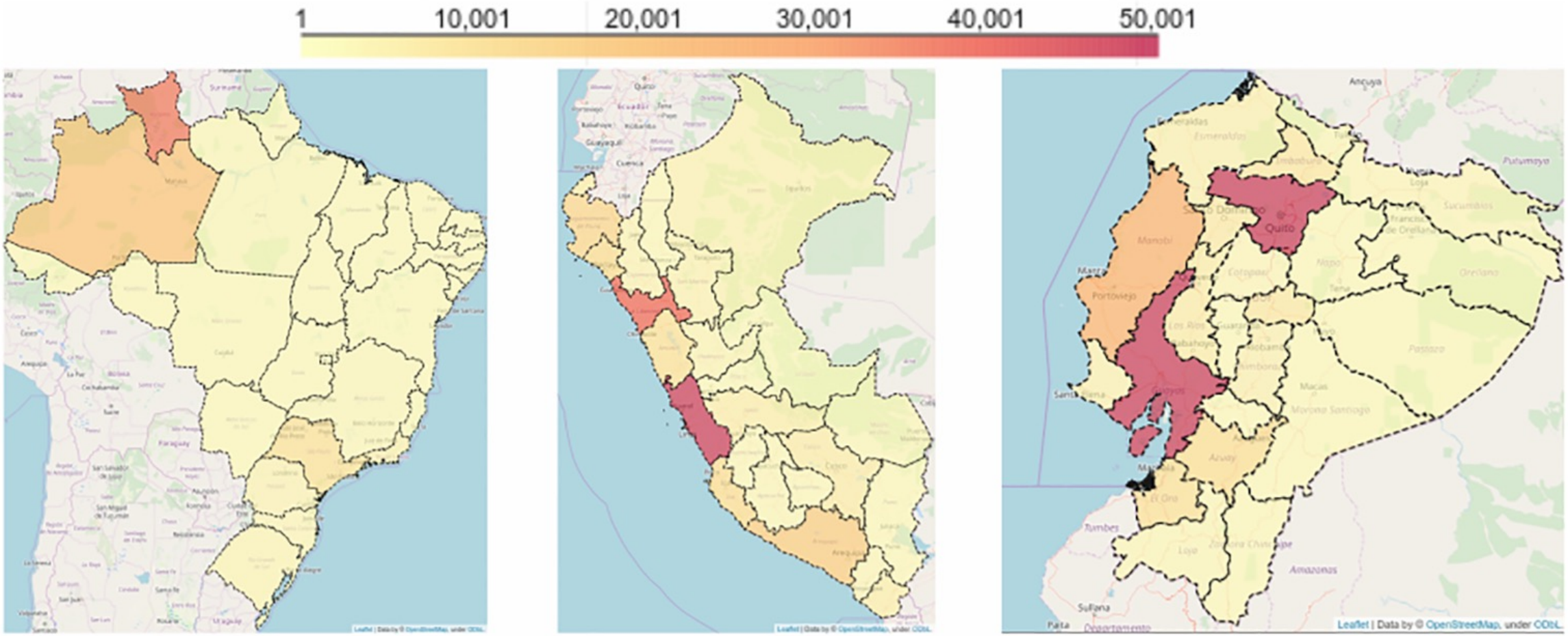


(a) RAMV - April 6 to June 8, 2018

(b) Facebook - June 8, 2018

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Previously Unavailable Estimates

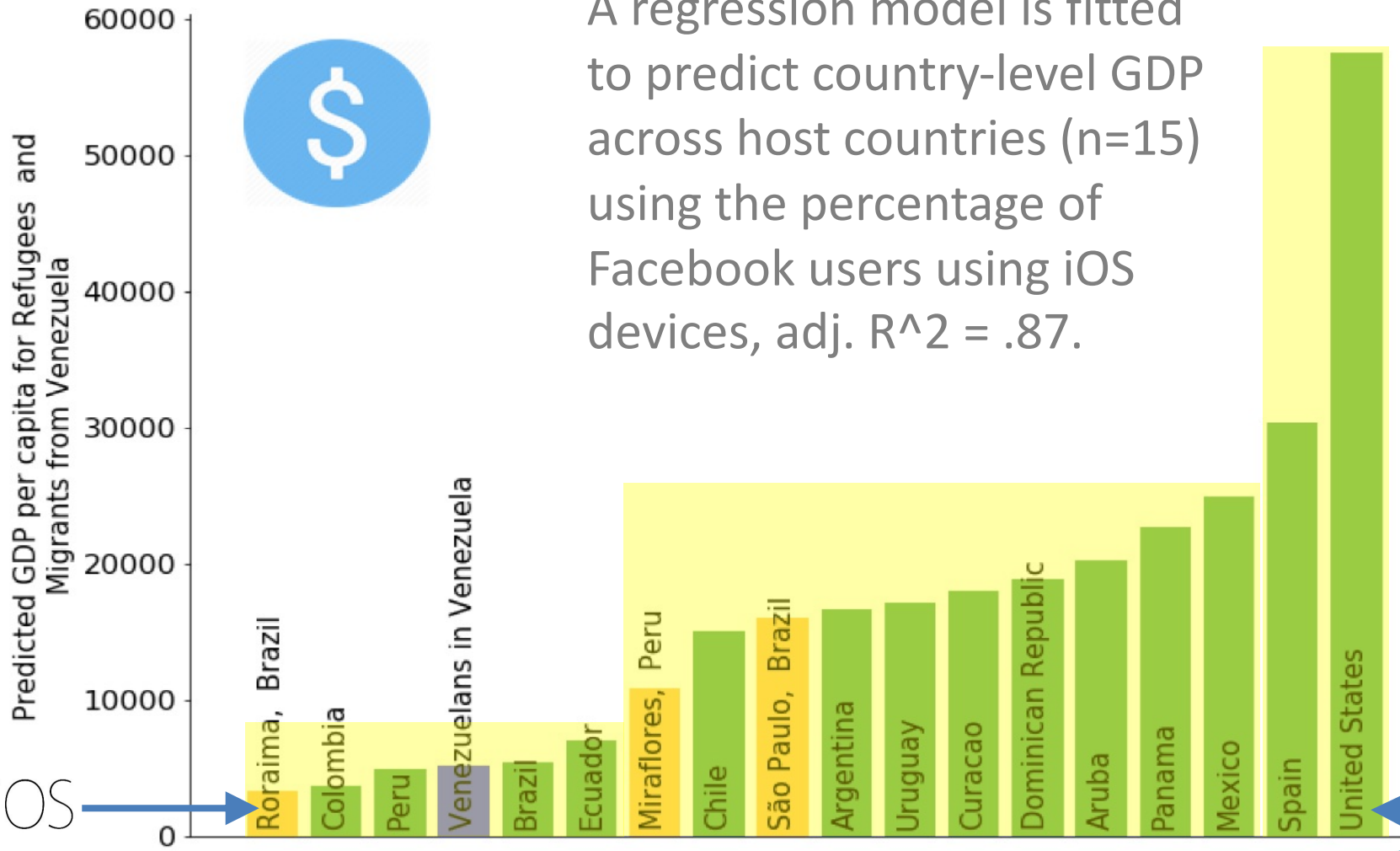


(a) Brazil (b) Peru (c) Ecuador

Facebook. Feb 2019

Predicted Income Based on OS

A regression model is fitted to predict country-level GDP across host countries (n=15) using the percentage of Facebook users using iOS devices, adj. R^2 = .87.



3% iOS

54% iOS

Operational Impact

LOS RESULTADOS

USAID | iMMAP

Detección de Usuarios venezolanos conectados en Facebook

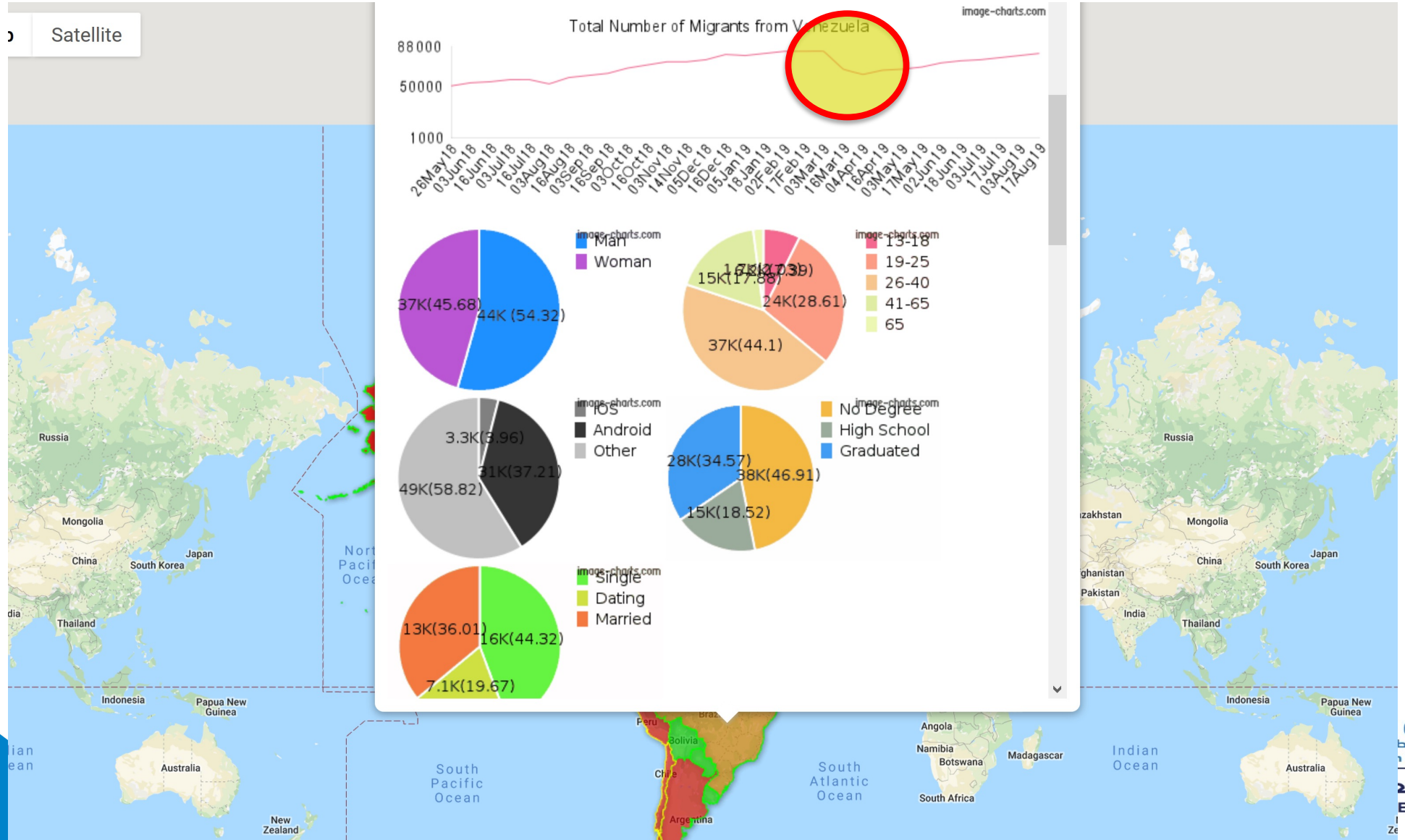
Municipal | Departamental | Regional | Perú | Ecuador | Tendencia

IMMAP localiza a los migrantes venezolanos en América Latina a través del uso de el api de Facebook advertising data mostrando las conexiones de usuarios que antes vivian en Venezuela y ahora viven en el extranjero.

Total usuarios conectados	Dato Oficial Migración Colombia
1,845,200	**1,729,537**
Usuarios conectados Febrero 15	Venezolanos en Colombia

- Los datos son recolectados cada 15 días, mediante el API de mercadeo de Facebook. Los estimados mostrados estan basados en los filtros utilizados y en el comportamiento de los usuarios en los últimos 30 días. Obtén mas información sobre el alcance potencial
- Los estimados presentados no están diseñados para coincidir con censos u otras fuentes oficiales. Facebook no provee censos digitales o conteos de migrantes y/o refugiados. Estos estimados deben ser vistos como una señal par ser utilizada en triangulación.
- Facebook solo provee la definición del comportamiento seleccionado (Expatriados Venezuela). No provee datos estadísticos ni históricos
- El comportamiento depende de la información proporcionada por el usuario en Facebook, su ciudad actual y ciudad de origen y la estructura de la red de amigos (por ejemplo, tener al menos dos amigos de Facebook en el país de origen y dos amigos de Facebook en el país de destino). Leer más: Leveraging Facebook's Advertising Platform to Monitor Stocks of Migrants, ZAGHENI, Emilio. WEBER, Ingmar. GUMMADI, Krishna

Changes to Facebook's Backend



MAPPING POVERTY

unicef

Thinking Machines

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Background on Poverty Mapping

"No Poverty" is the first of the 17 Sustainable Development Goals

- Standard definition of extreme poverty: < $1.25 per day

Available poverty data is often outdated

- Algeria 2011, Jordan 2010, Nigeria 2010, …

Lack of spatially granular data at the sub-city level

- Hard to plan or evaluate targeted poverty interventions

Predicting Poverty and Wealth

Ground truth “Wealth Index” – the dependent variable

- USAID sponsors the Demographic and Health Survey (DHS)
- This survey asks households about asset ownership
- Compiled into a “Wealth Index” for a surveyed location

Features describing a location – the independent variables

- %-age of FB users of population, or with iOS, Wifi, 4G, …

Regression Task – the regression model

- Use Gradient Boosting Machines and other regression models
- Learn to predict the Wealth Index for surveyed locations
- Estimate the Wealth Index for *non*-surveyed locations

Results

This uses the Wealth Index of nearby locations as features.

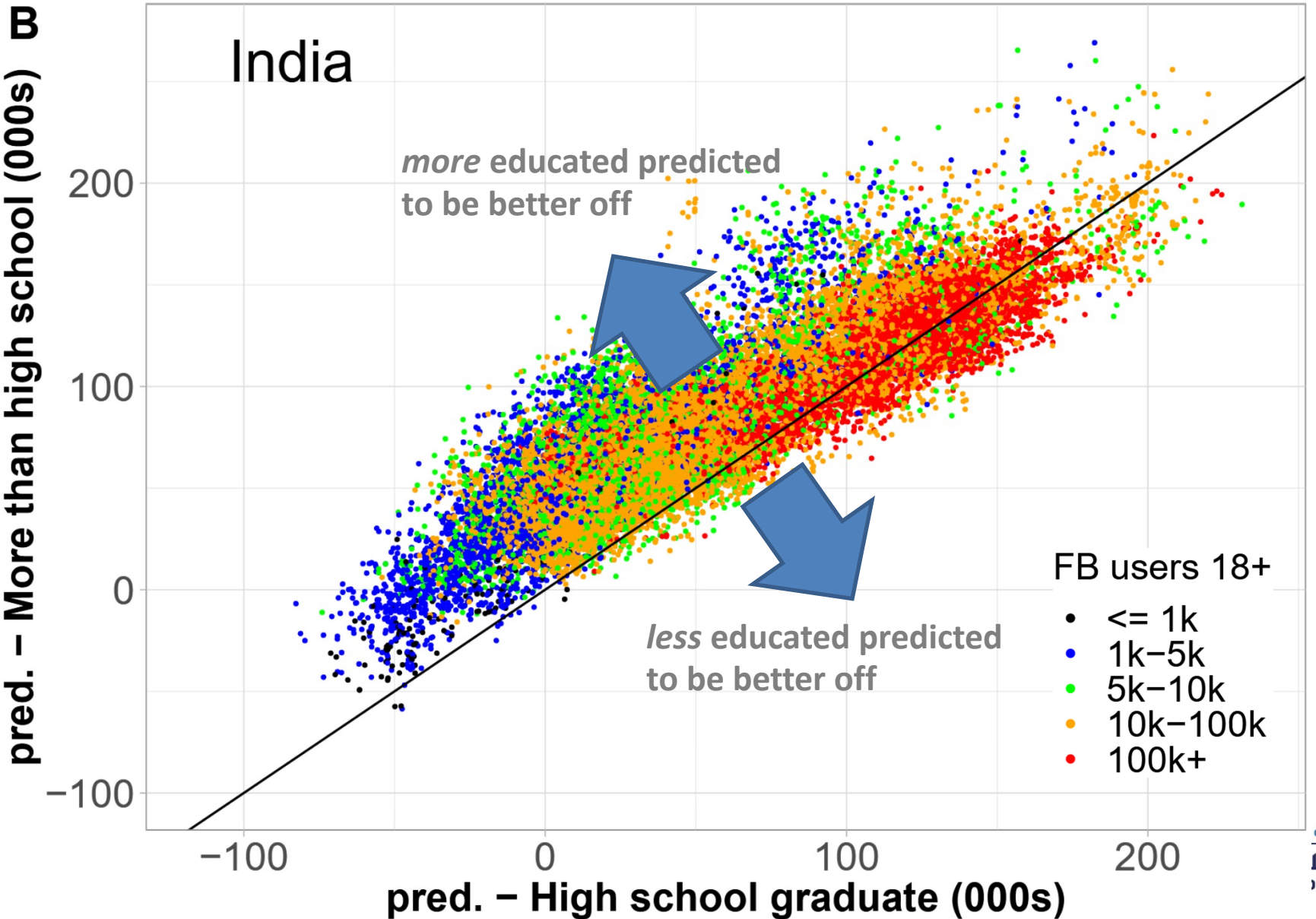
This includes the FB penetration, computed using high resolution settlement layer information.

This is a dummy variable “is the location part of [name of region]”.

Model features					
Interpolated DHS Wealth Index		X			X
Facebook features			X	X	X
Log population density				X	X
Regional indicators				X	X
Philippines (*N* = 1205)	R^2	0.480	0.608	0.627	0.630
	RMSE	50,983	44,218	43,099	42,965
India (*N* = 28,043)	R^2	0.652	0.563	0.691	0.728
	RMSE	46,810	52,502	44,149	41,394

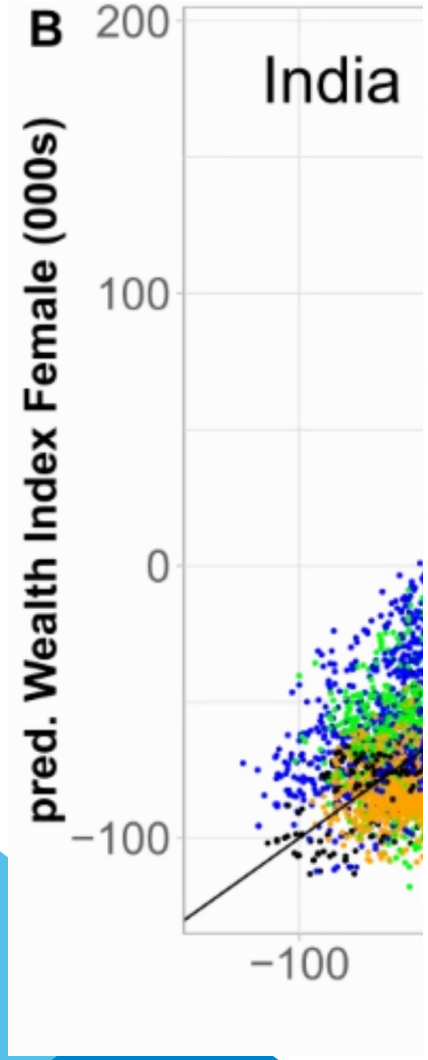
upper bound R^2 = .845/.838 (PH/IN) (due to noise)

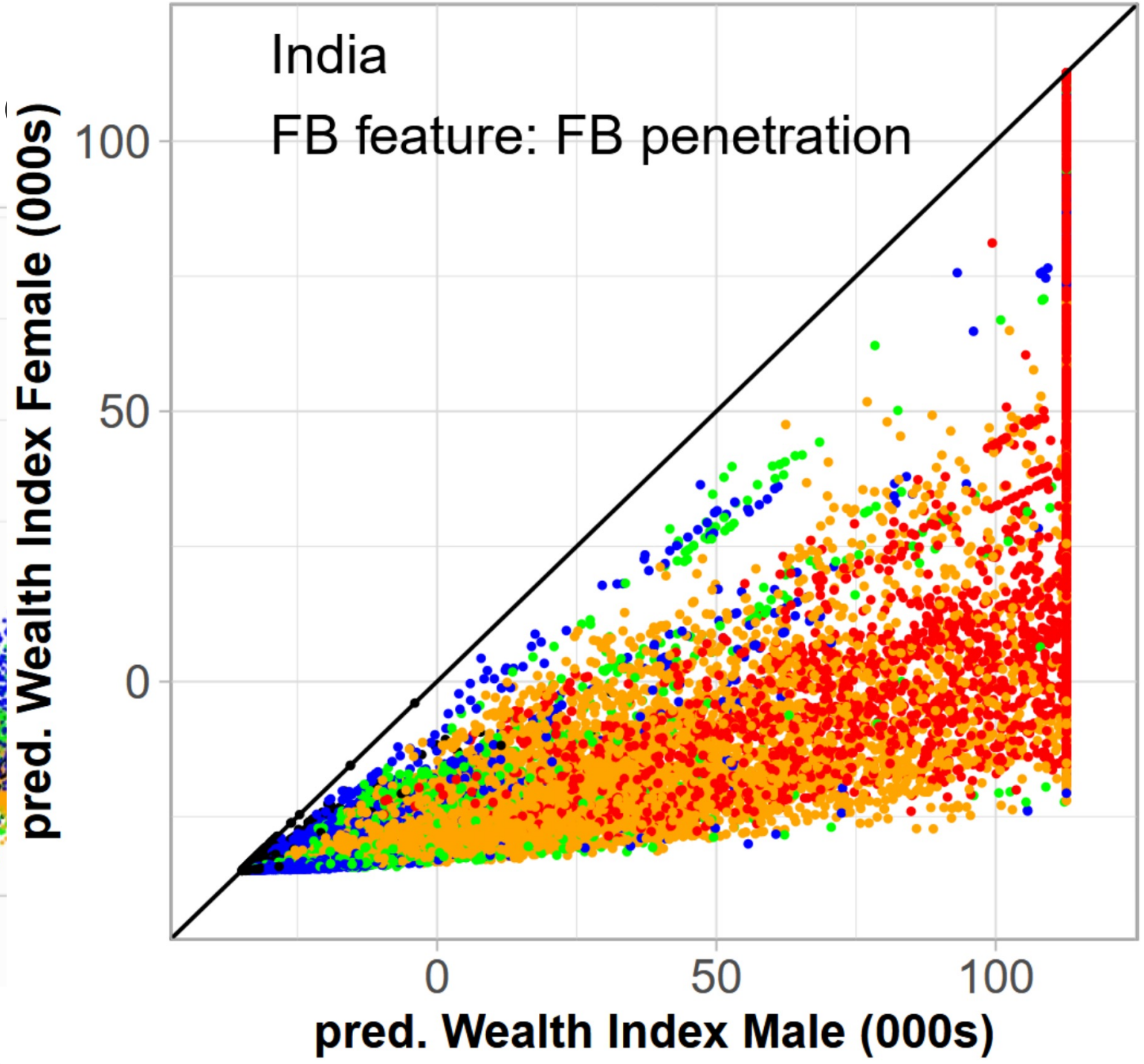
Education-Level-Disaggregated Predictions



QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Gend

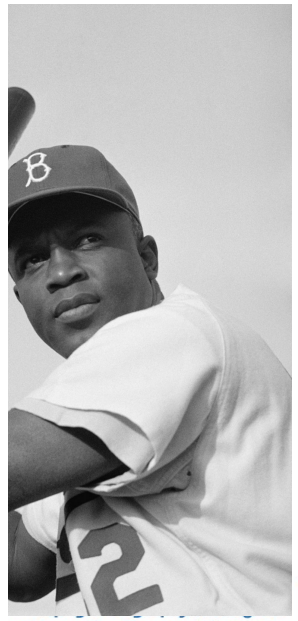




		score
p Index	112	0.668
tion and opportunity	149	0.354
ient	112	0.962	
	150	0.944
ent	18	0.411

Out of 153

binson Effect



ar Computing Research Institute
جامعة حمد بن
N KHALIFA UNIVERSITY

Features Used in Deployment

Understanding poverty in the Philippines with artificial intelligence

December 22, 2020 case-study computer-vision geospatial machine-learning mapping open-source openstreetmap poverty remote-sensing satellite-imagery sdg sustainable-development-goals business development government healthcare ngo telecom

SUMMARY

Thinking Machines helped the United Nations Development Programme (UNDP) generate nationwide wealth estimates in the Philippines with machine learning:

- AI model estimated the wealth index across the Philippines, based on the 2017 Demographic and Health Survey
- AI model was trained on open geospatial data sourced from OpenStreetMap, Facebook Marketing API, VIIRS Nighttime Lights, Land Surface Temperature, NDVI, etc.
- Rolled out the model to 18 square kilometer grids nationwide with better performance and at the fraction of the cost of our previous model
- Generated granular and nationwide map of wealth estimates

IMPACT

Developed a model that generates **low-cost, reliable, and granular poverty estimates** at scale in under a minute.

https://stories.thinkingmachin.es/poverty-mapping-artificial-intelligence/

TRACKING DIGITAL GENDER GAPS

data2x

UNIVERSITY OF OXFORD

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Background on Digital Gender Gaps

There are ~4.7 billion internet users worldwide (~60% of population)

But access across genders is not equal

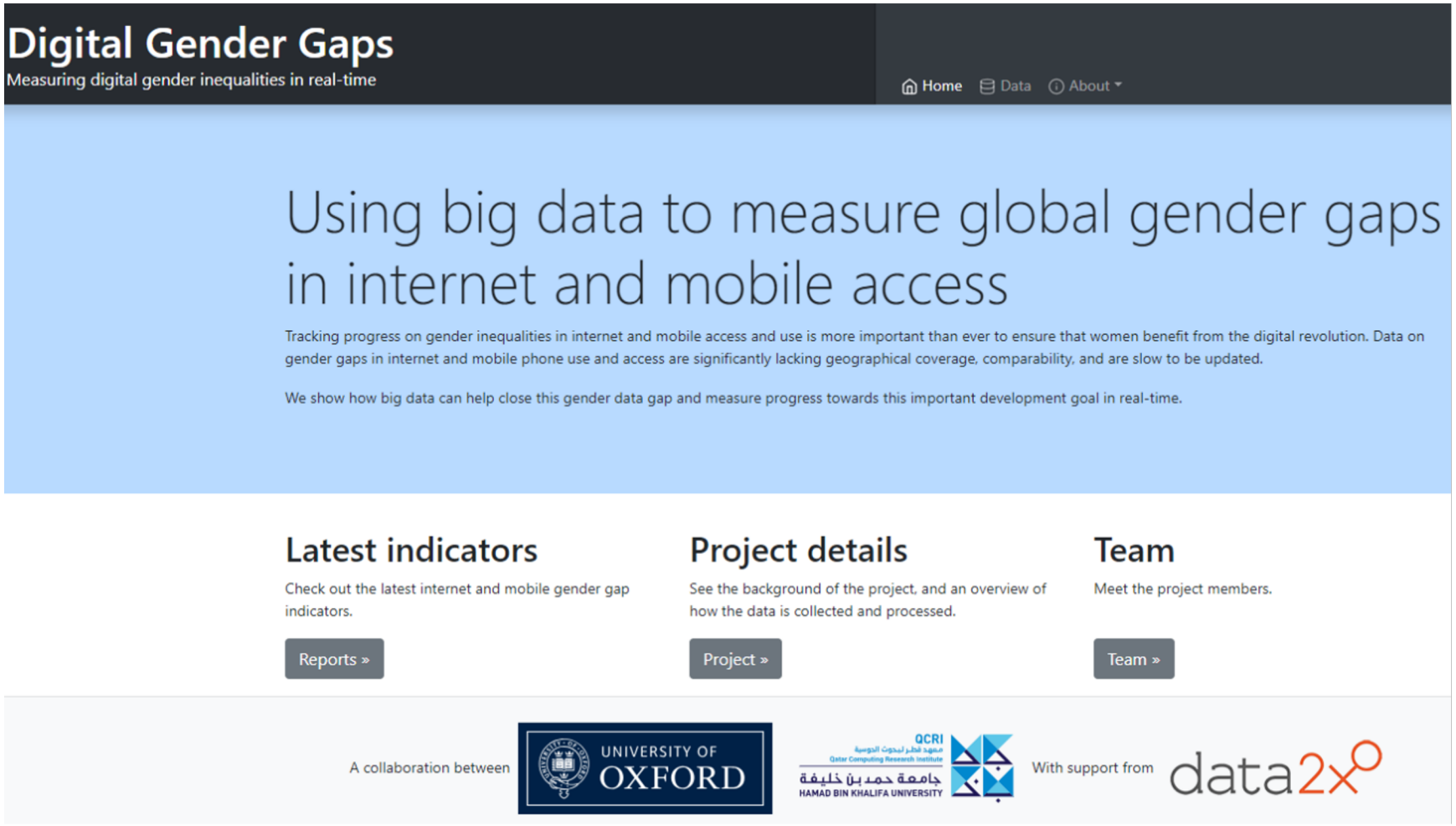
Niger: 1 woman for every 3 men with internet access

Iraq: 1 woman for every 2 men with internet access

Official data only exists for half the world's countries

Gender equality is one of the Sustainable Development Goals

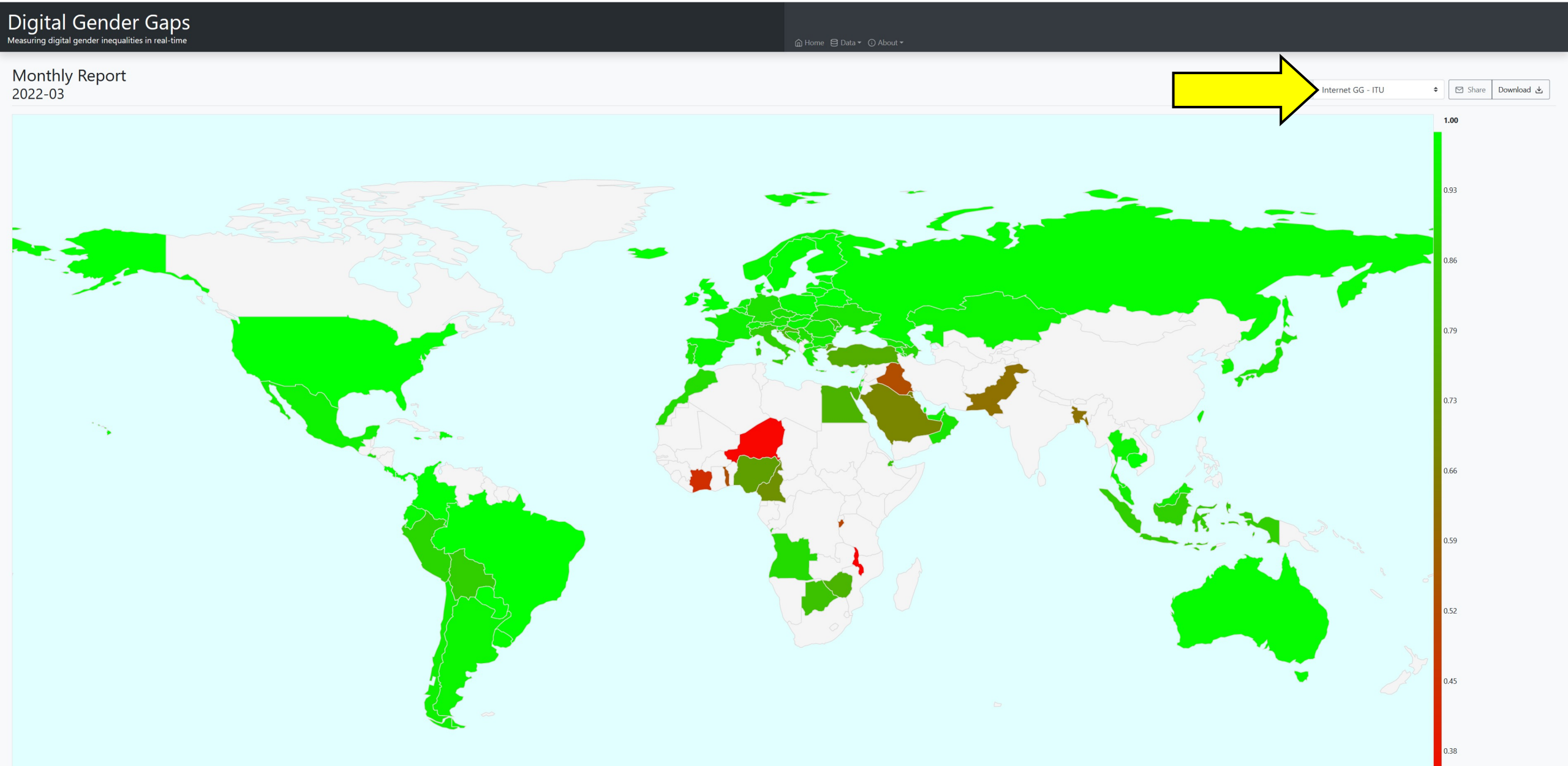
www.digitalgendergaps.org



QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

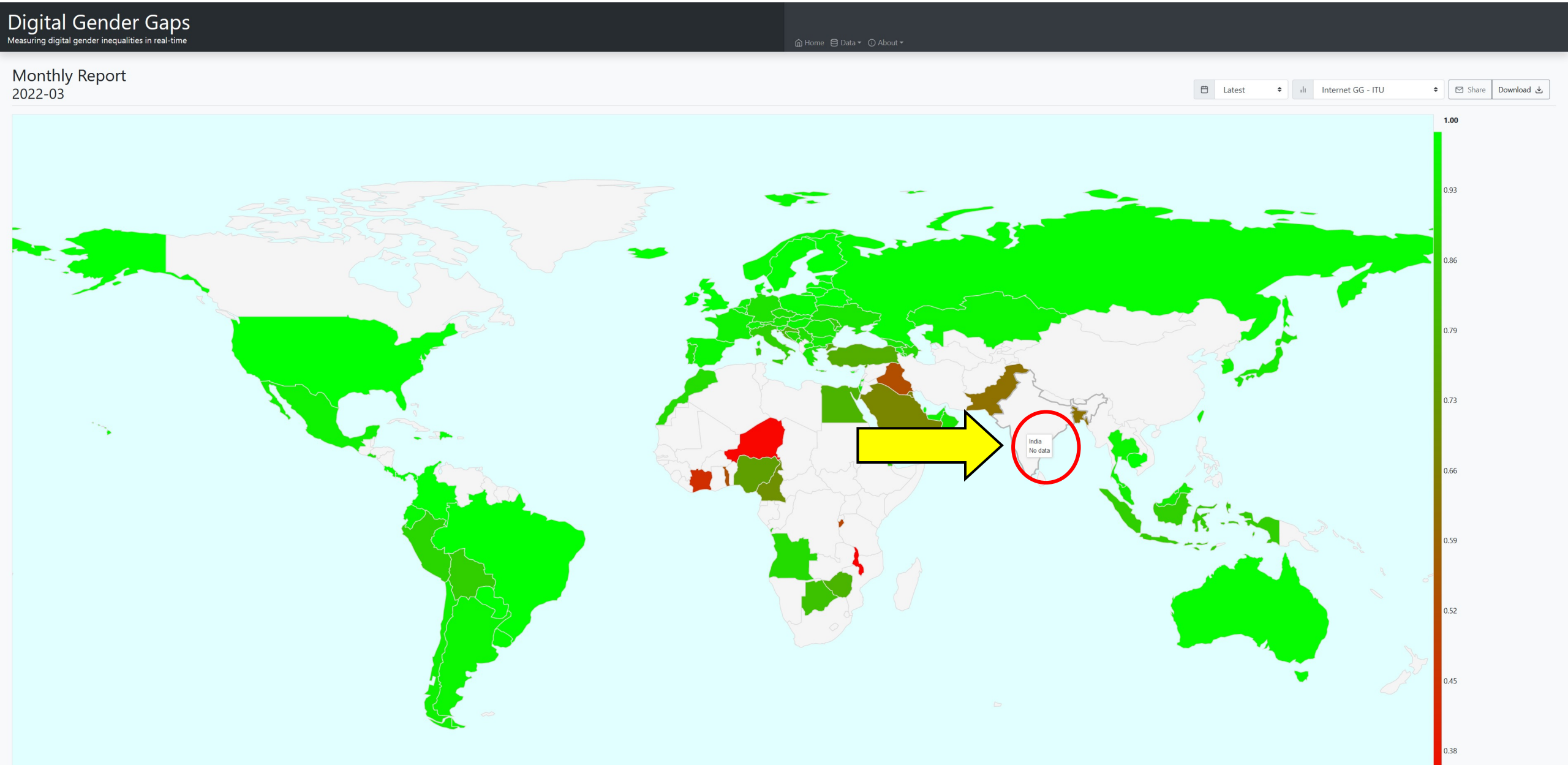
Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org



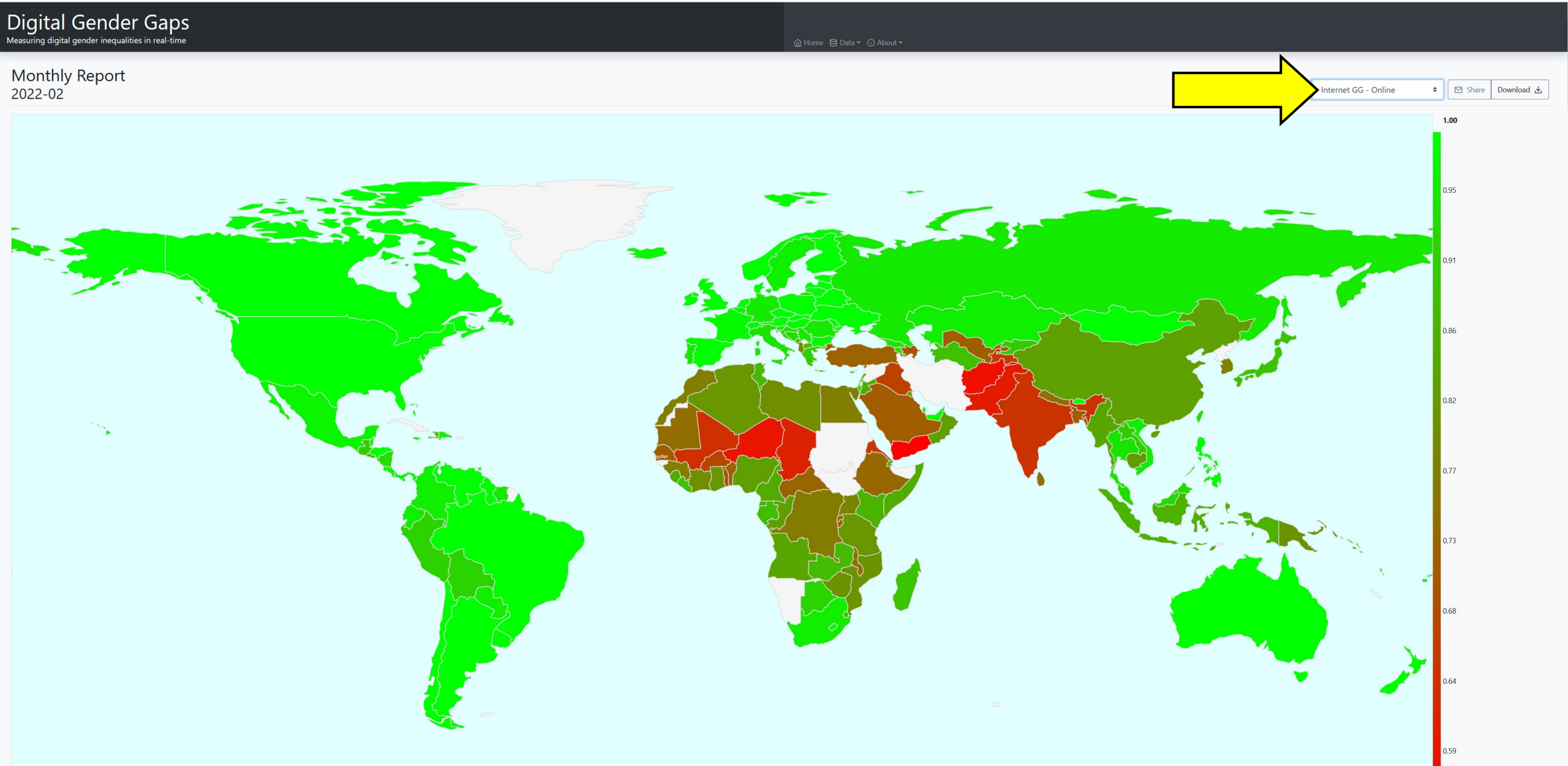
Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org



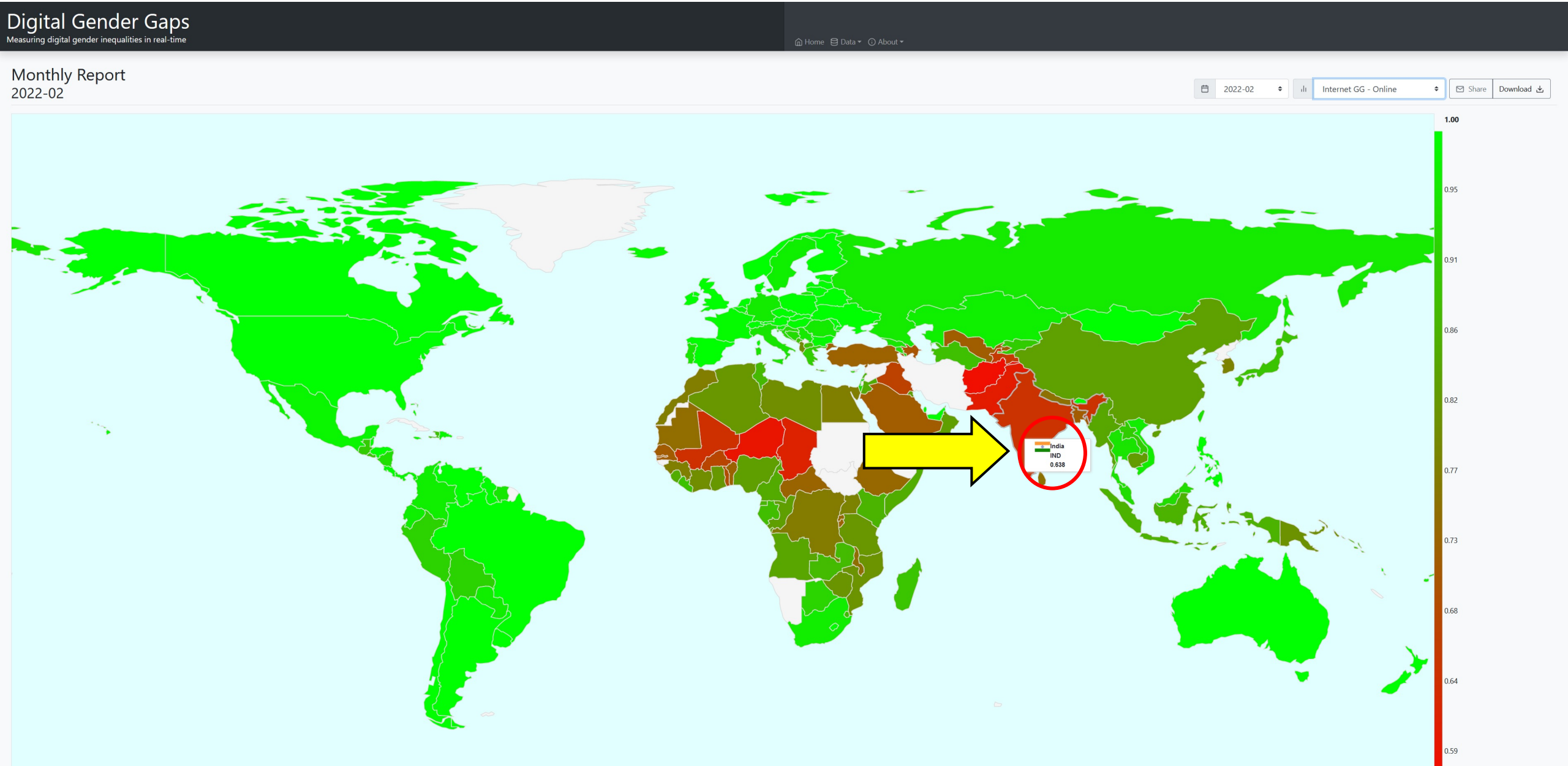
Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org



Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org



Internet Access Gender Gap Predictions

	Online Model	Onl.-Offl. Model	Offline Model
Intercept	0.933***	0.932***	0.933***
	(0.006)	(0.005)	(0.007)
FB GG (age 18+)	0.071***	0.093***	
	(0.011)	(0.017)	
log(GDP per capita)		0.018*	
		(0.008)	
GGGR – Literacy		−0.018	
		(0.016)	
GGGR – Education		−0.019	
		(0.019)	
Internet Penetration			0.040***
			(0.009)
GGGR – Tertiary Educ.			0.032
			(0.021)
GGGR – Economy			0.043**
			(0.014)
GGGR Score			−0.024
			(0.012)
Adjusted R-squared	0.691	0.791	0.615
# predicted countries[a]	152	127	132

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$.

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Part of SDGs Today Portal

SDGs Today

Data Hub | Storytelling | Education | About | Submit your data

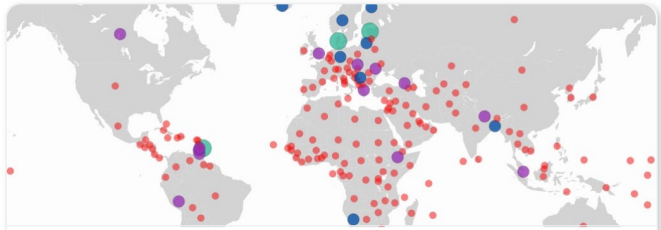
Data Hub

5 GENDER EQUALITY

SDG 5: Achieve gender equality and empower all women and girls

SDG 5 calls for an end to gender discrimination in all forms by 2030 and promotes equal opportunity for all women and girls in terms of education, career, and sexual/reproductive rights. Review the latest reference metadata information provided by the UN System and other international organizations on data and statistics for SDG 5 here. SDGS Today works with various data communities to curate, produce, and feature new data sources and methods that can complement official SDG data. Interact with our ArcGIS maps and dashboards, explore our StoryMap collections, review the metadata, access the underlying data provided by our partners and discover other relevant resources on SDG 5 in our Data Hub.

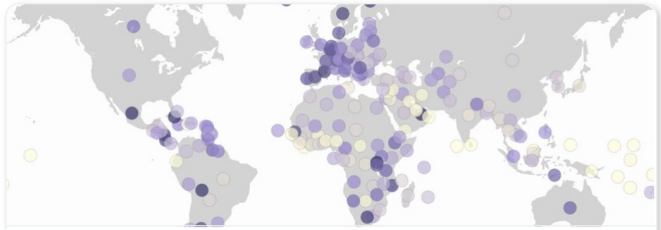
Datasets for SDG 5 | More About SDG 5



Female World Leaders

United Nations

SDG 5 - Gender Equality



Percentage of Women in Parliament

Inter-Parliamentary Union

SDG 5 - Gender Equality

Digital Gender Gap

University of Oxford, QCRI, Data2X

SDG 5 - Gender Equality

https://sdgstoday.org/sdg/gender-equality

جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Joint work with Ridhi Kashyap, Masoomali Fatehkia and Ian Knowles

Ongoing: Trends in Afghanistan

DW Made for minds.

ASIA

Will the Taliban restrict internet access in Afghanistan?

MORE ON THE TALIBAN'S RETURN TO POWER IN AFGHANISTAN

The Taliban say that they want to ensure internet access in Afghanistan, but they could face substantial technical and financial challenges to keep it running. Afghans say they fear more surveillance and censorship.

ALJAZEERA News Afghanistan Features Economy Opinion Video

News | Social Media

As Taliban returns, Afghan influencers go dark on social media

Prominent social media influencers go dark or flee, while residents and activists scramble to scrub their digital lives.

Search jobs My account Search The Guardian For 200 years International edition

Global development

#DoNotTouchMyClothes: Afghan women's social media protest against Taliban

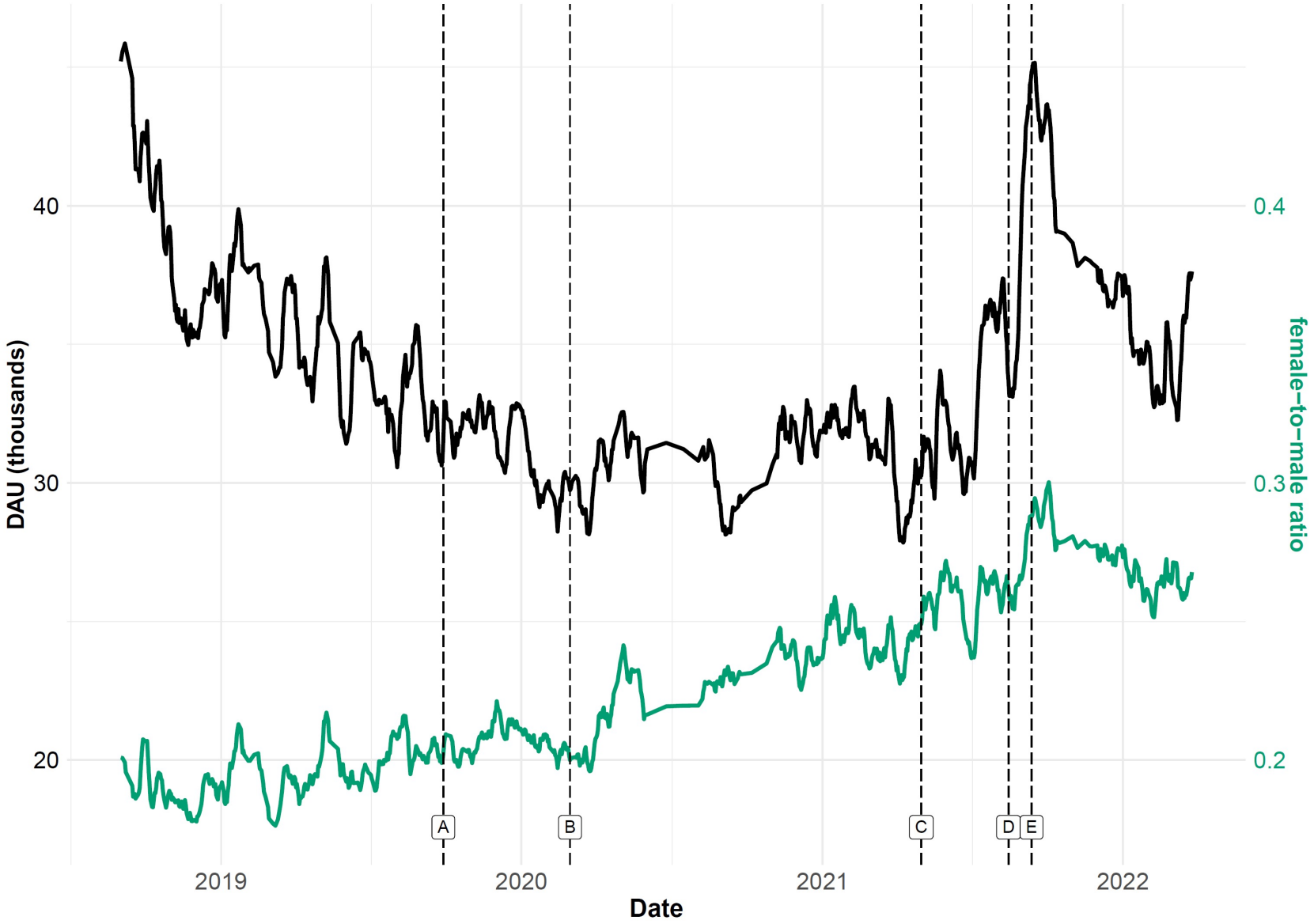
Women around the world are sharing pictures of themselves in traditional colourful clothes in a campaign against the new strict dress code for female students

QCRI معهد قطر لبحوث الحوسبة Qatar Computing Research Institute

HAMAD BIN KHALIFA UNIVERSITY جامعة حمد بن خليفة

Joint work with Ridhi Kashyap, Masoomali Fatehkia and Ian Knowles

Ongoing: Changes in Facebook Daily Active Users

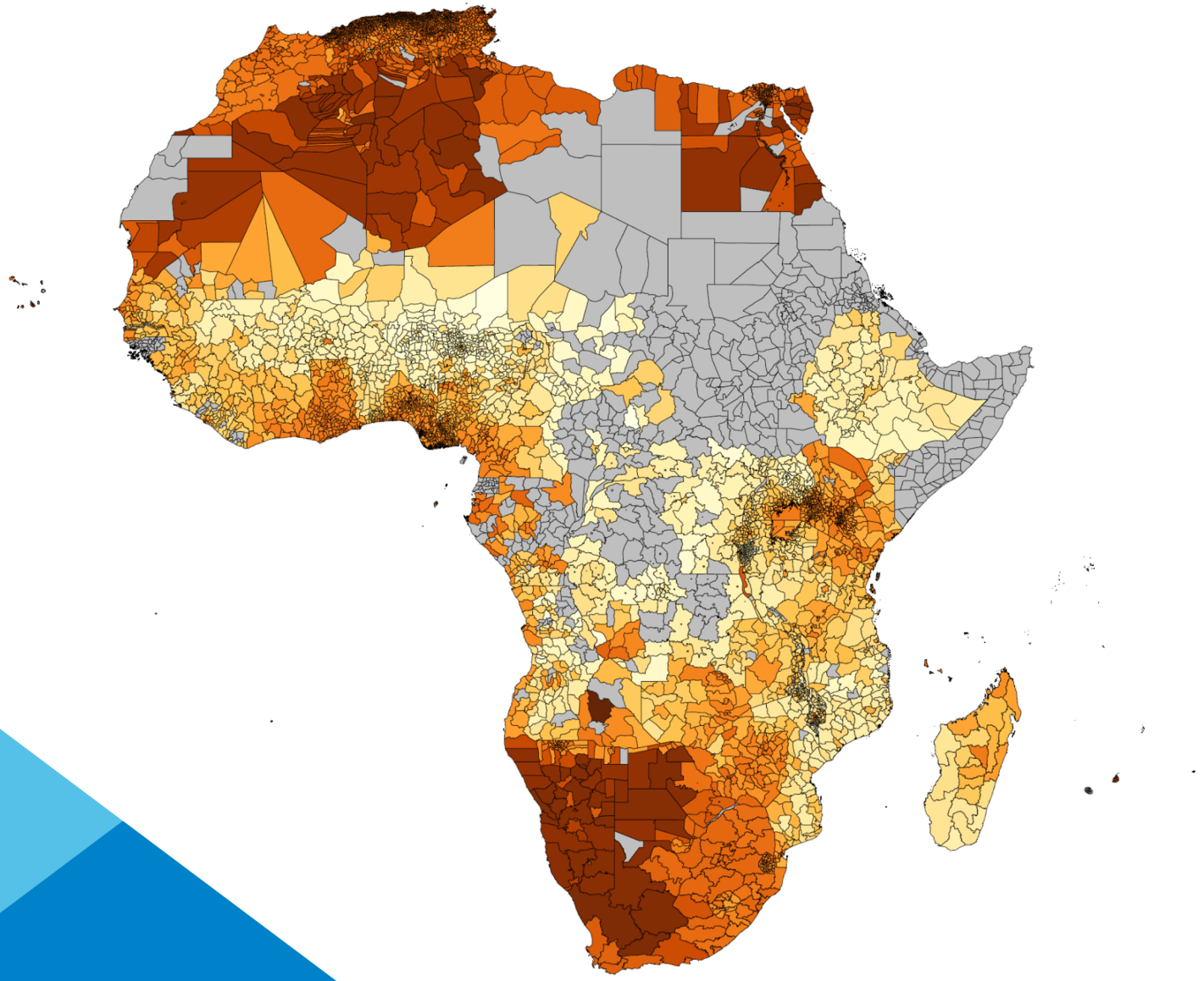


Changes in Facebook usage for 15-19 year old women

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Joint work with Masoomali Fatehkia, Ridhi Kashyap, and Douglas Leasure

Ongoing: Subnational Digital Gender Gaps in Africa



Spatial variation in the predicted f-to-m ratio of mobile phone ownership

91% population coverage

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

CLOSING THOUGHTS

Recap: Advertising Audience Estimates

+ Facebook, LinkedIn, Weibo, Snapchat, Google, ...
+ (Relatively) real-time estimates
+ Uses anonymous and aggregate data
+ Gender, age, location, device type, ….
+ Free of charge

- Blackbox inference for many attributes
- Non-representative, biased sample
- Gender dichotomy used in advertising
- Usage patterns change over time
- No historic data available
- Risk of misuse

“Better data will lead to better decisions!”

Maybe. But remember global warming?

Better data can help to highlight systematic issues, but more and better data will not fix those issues.

Thanks!

Happy to collaborate.

iweber@hbku.edu.qa

https://ingmarweber.de/publications/

QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY