Using Advertising Data to Model Digital Gender Gaps and Poverty

Ingmar Weber

@ingmarweber

ONS-UNECE Machine Learning Group

March 2, 2022

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

Fantastic Collaborators

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 remote social sensing?

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

Detailed targeting

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**

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

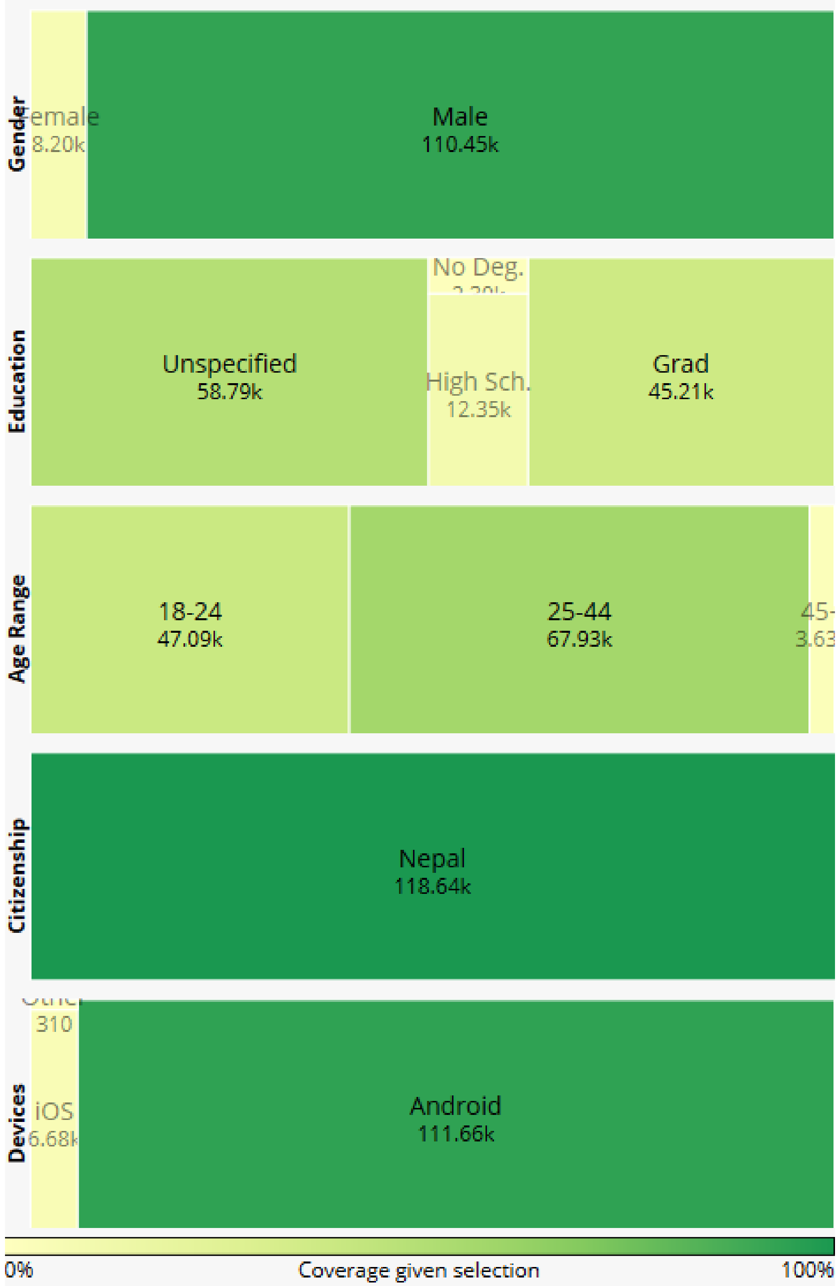
82-510

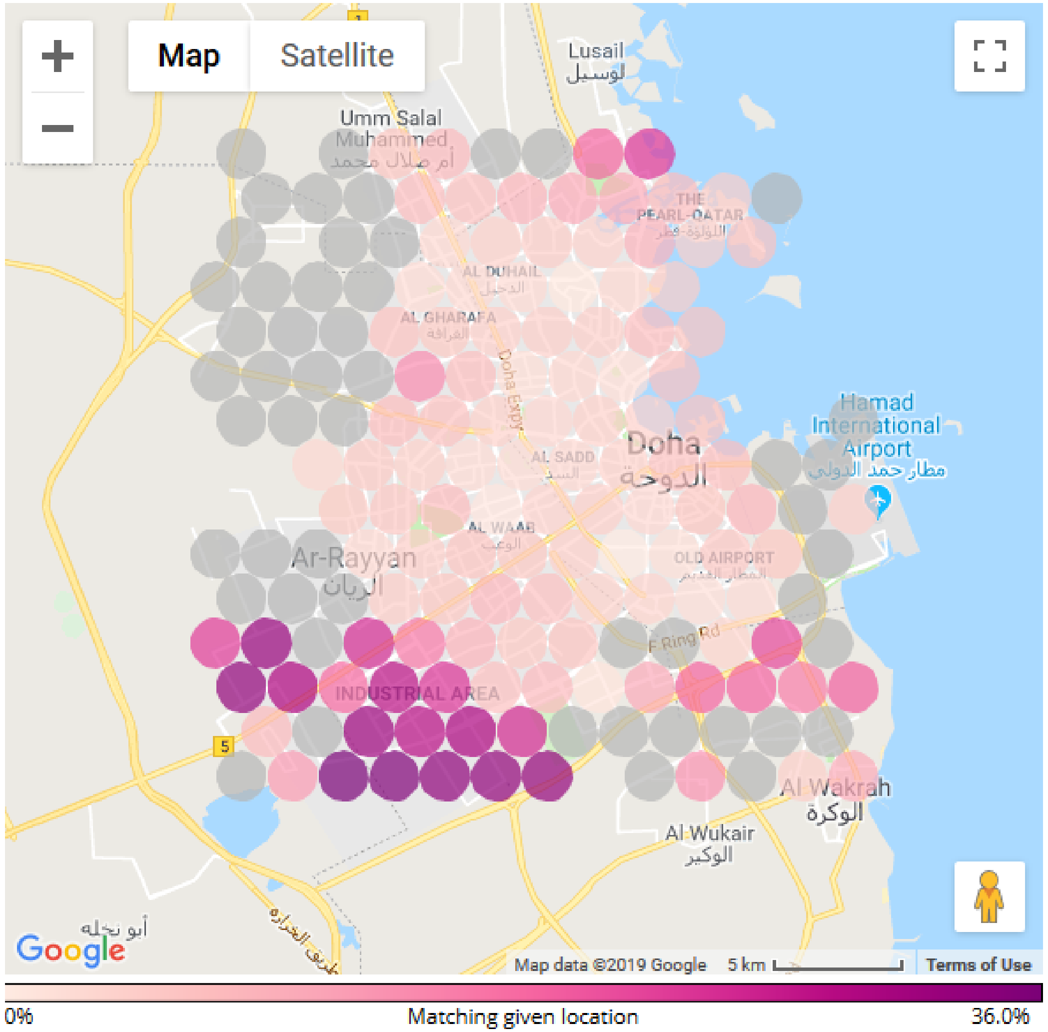
Removing Instagram and Audience Network may result in 45% fewer Post Engagement, based on your past campaign performance. We recommend choosing automatic placements for the best results.

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





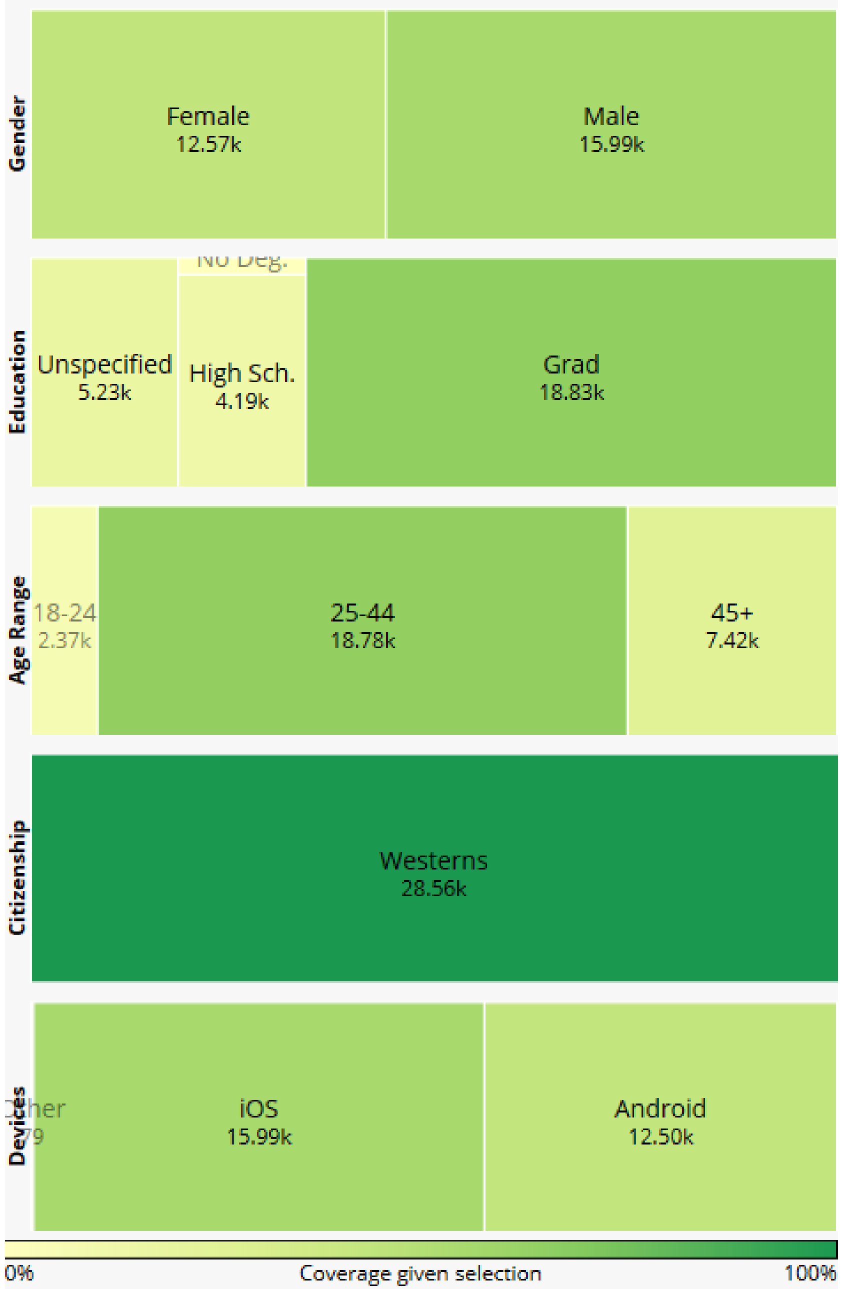
All Locations

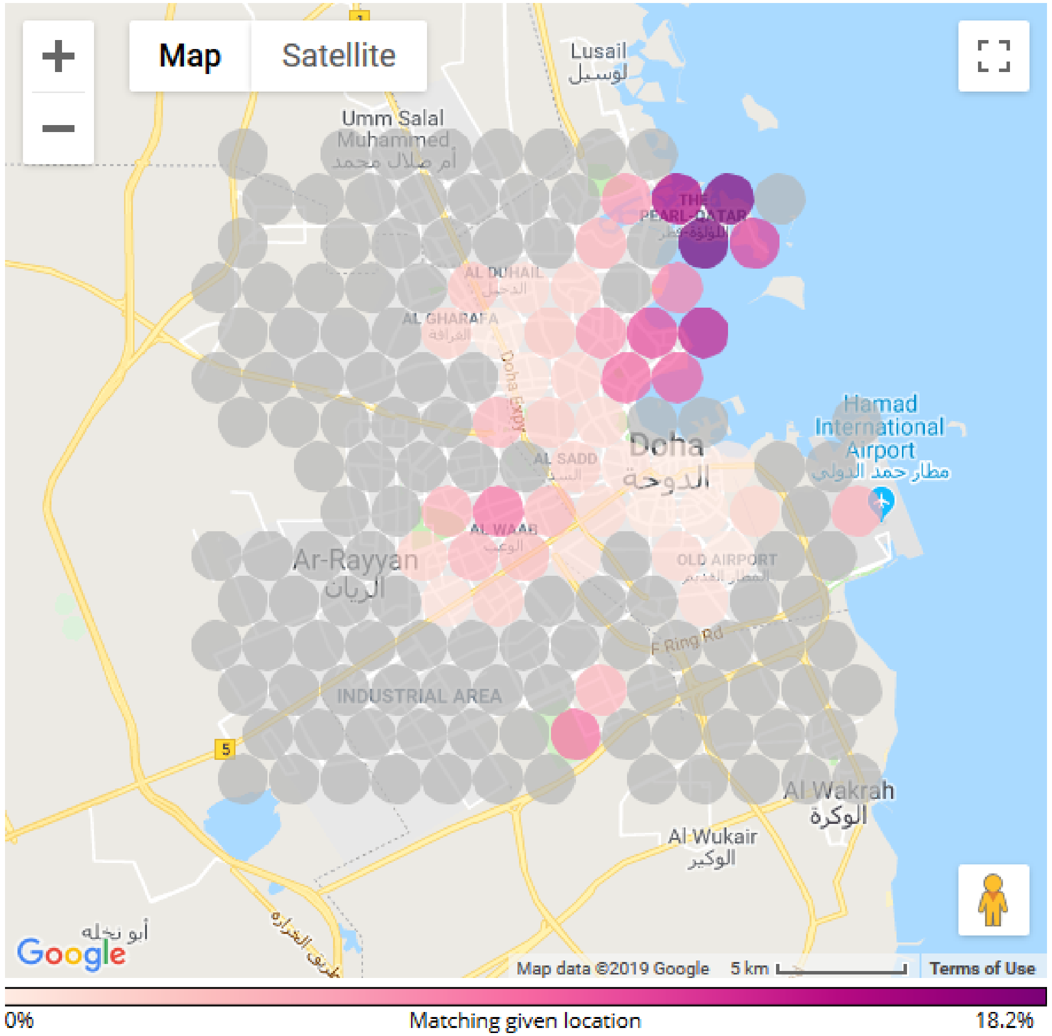
No Locations

Share what you see:

http://fb-doha.qcri.org

QCRI
Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY





All Locations

No Locations

Share what you see:

http://fb-doha.qcri.org

QCRI
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

Well-Documented API and Wrapper

THE WORLD BANK

Connectivity Mapping

Search this book...

Contents

References:

Connectivity Maps using Advertisement Platforms

Social networks, such as Facebook and Linkedin, are widely used by the global population. While caveats regarding data bias collection apply, these social networks can access essential data for many studies. Recently, for example, the Facebook Market platform was used to study the *United Nations Sustainable Development Goals (SDGs)* [FTO+20], to measure cultural differences between urban and rural population [RMT+20], to measure gender gaps [KFTW20] and to monitor refugees and forced immigrants [PAMG+20].

In this tutorial, we will learn the basics of performing a data collection using state-of-the-art libraries to collect data and visualize the results. It covers the basics of using Facebook's and LinkedIn's Marketing API to collect valuable data on the number of users that use this social network in a specific region, and several of their characteristics, like their demographics, interests, education and job experience.

We would like to thank Kiran Garimella (garimell@mit.edu) and Emilio Zagheni (zagheni@demogr.mpg.de) for developing the first version of the LinkedIn code this tool was built upon, and Ingmar Weber (iweber@hbku.edu.qa) for kindly sharing it with us.

References: ¶

[FTO+20]
Masoomali Fatehkia, Isabelle Tingzon, Ardie Orden, Stephanie Sy, Vedran Sekara, Manuel Garcia-Herranz, and Ingmar Weber. Mapping socioeconomic indicators using social media advertising data. *EPJ Data Science*, 9(1):22, 2020.

[KFTW20]
Ridhi Kashyap, Masoomali Fatehkia, Reham Al Tamime, and Ingmar Weber. Monitoring global digital gender inequality using the online populations of facebook and google. *Demographic Research*, 43:779–816, 2020.

[PAMG+20]
Joao Palotti, Natalia Adler, Alfredo Morales-Guzman, Jeffrey Villaveces, Vedran Sekara, Manuel Garcia Herranz, Musa Al-Asad, and Ingmar Weber. Monitoring of the venezuelan exodus through facebook's advertising platform. *Plos one*, 15(2):e0229175, 2020.

[RMT+20]
Daniele Rama, Yelena Mejova, Michele Tizzoni, Kyriaki Kalimeri, and Ingmar Weber. Facebook ads as a demographic tool to measure the urban-rural divide. In *Proceedings of The Web Conference 2020*, 327–338. 2020.

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

Next: Example Applications

Track digital gender gaps

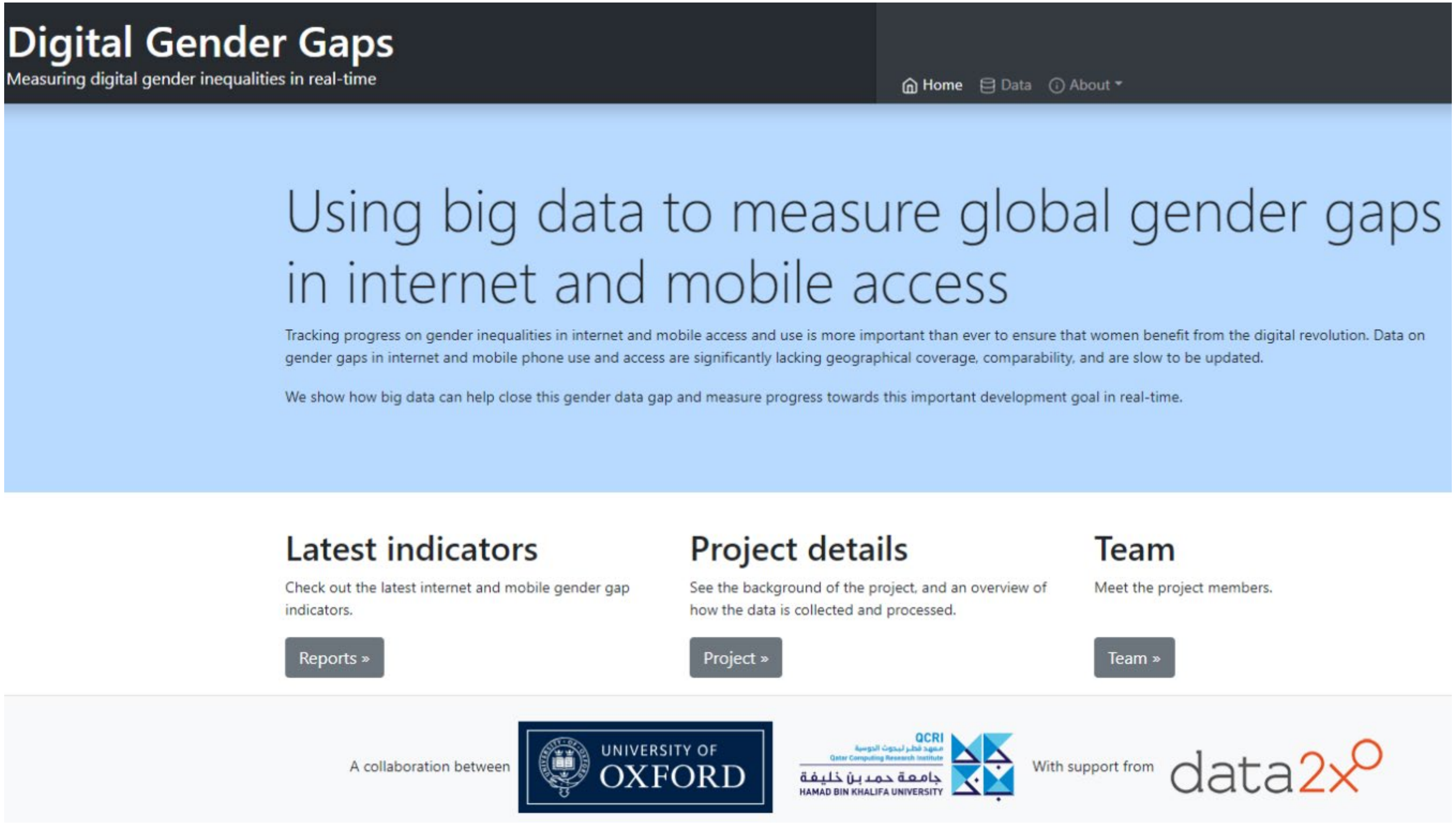
- FB gender gaps mirror internet gender gaps

Map poverty

- iOS devices more prominent in wealthier areas

TRACKING DIGITAL GENDER GAPS

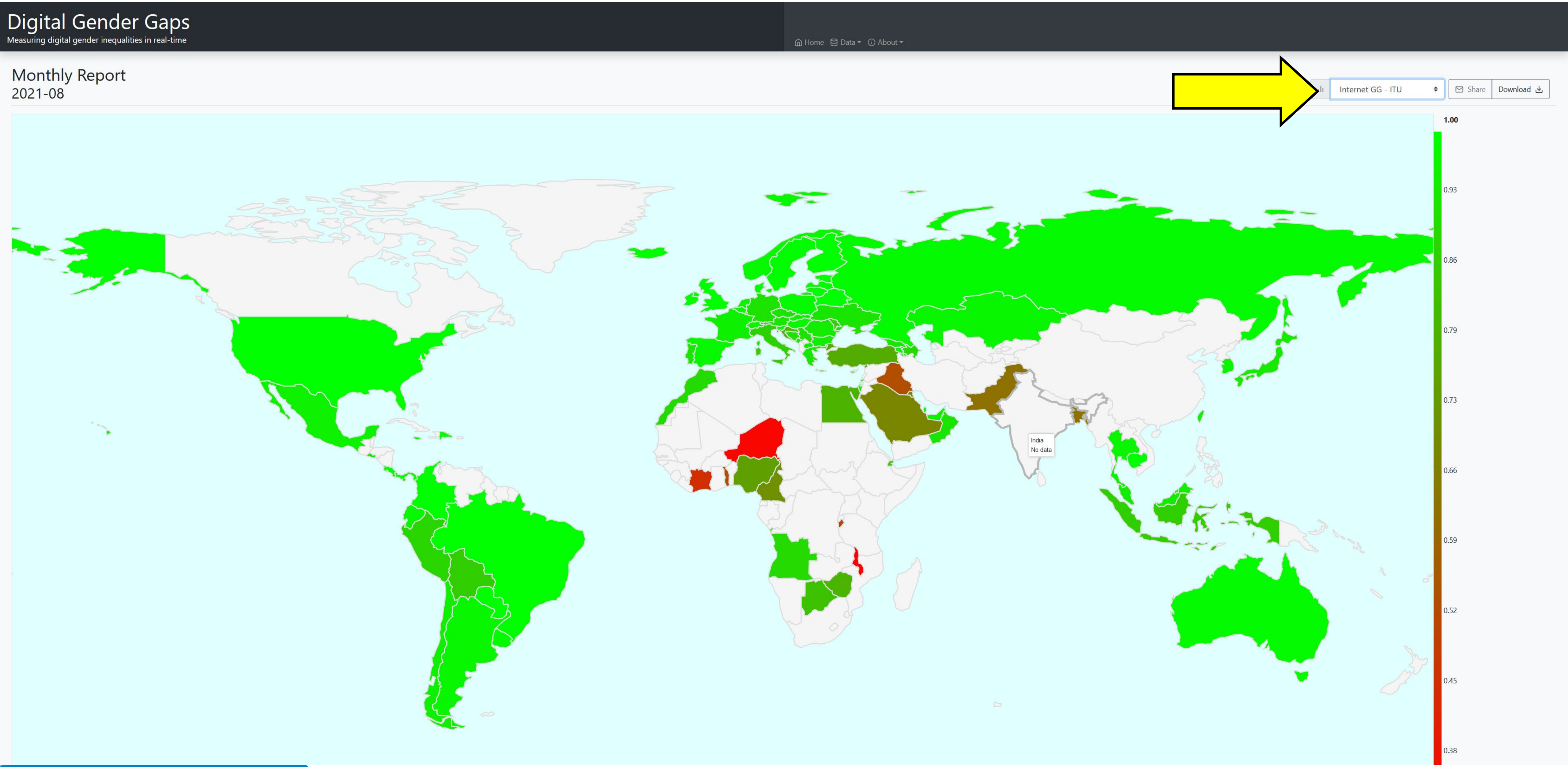
www.digitalgendergaps.org



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

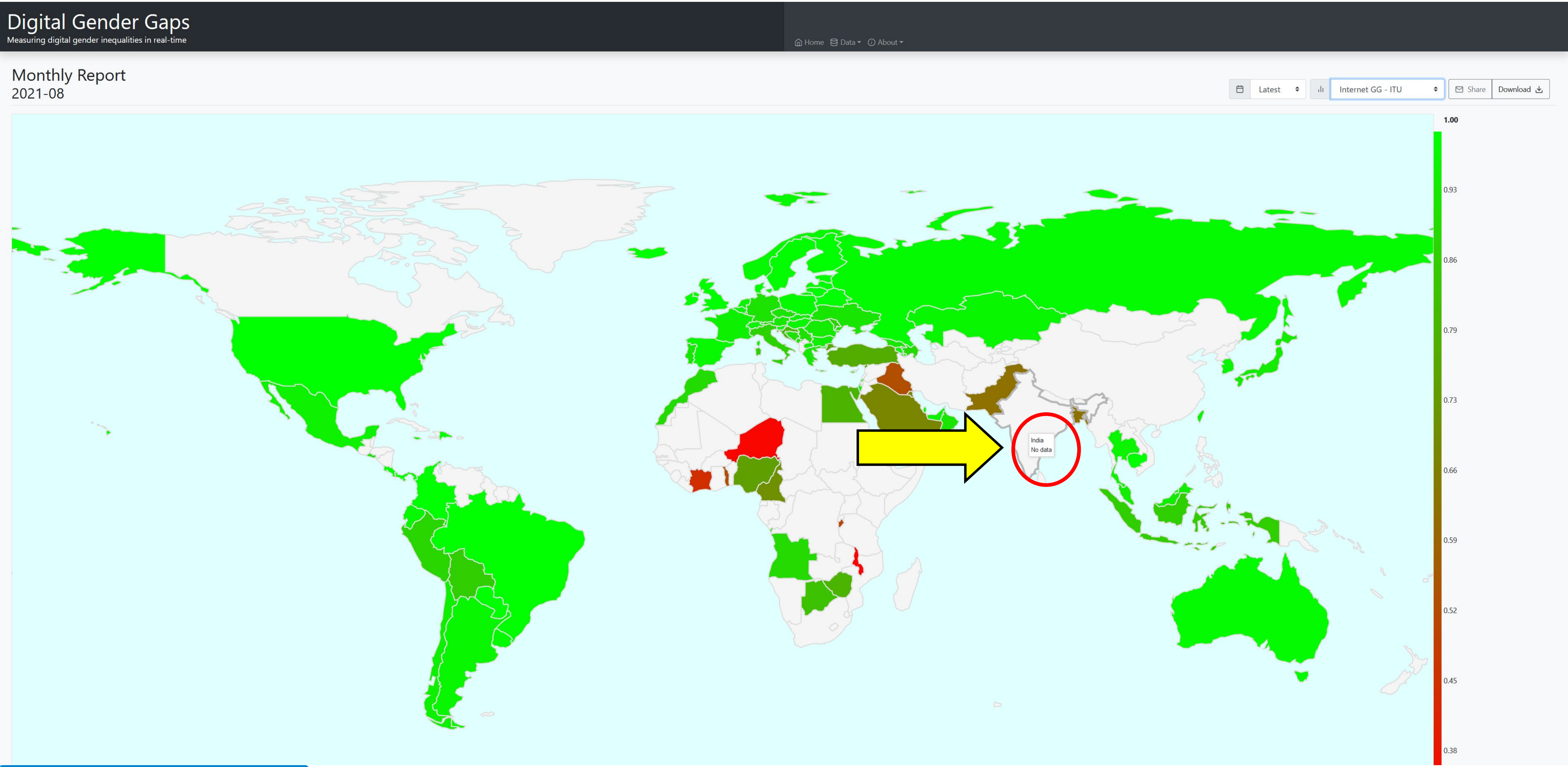
Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org



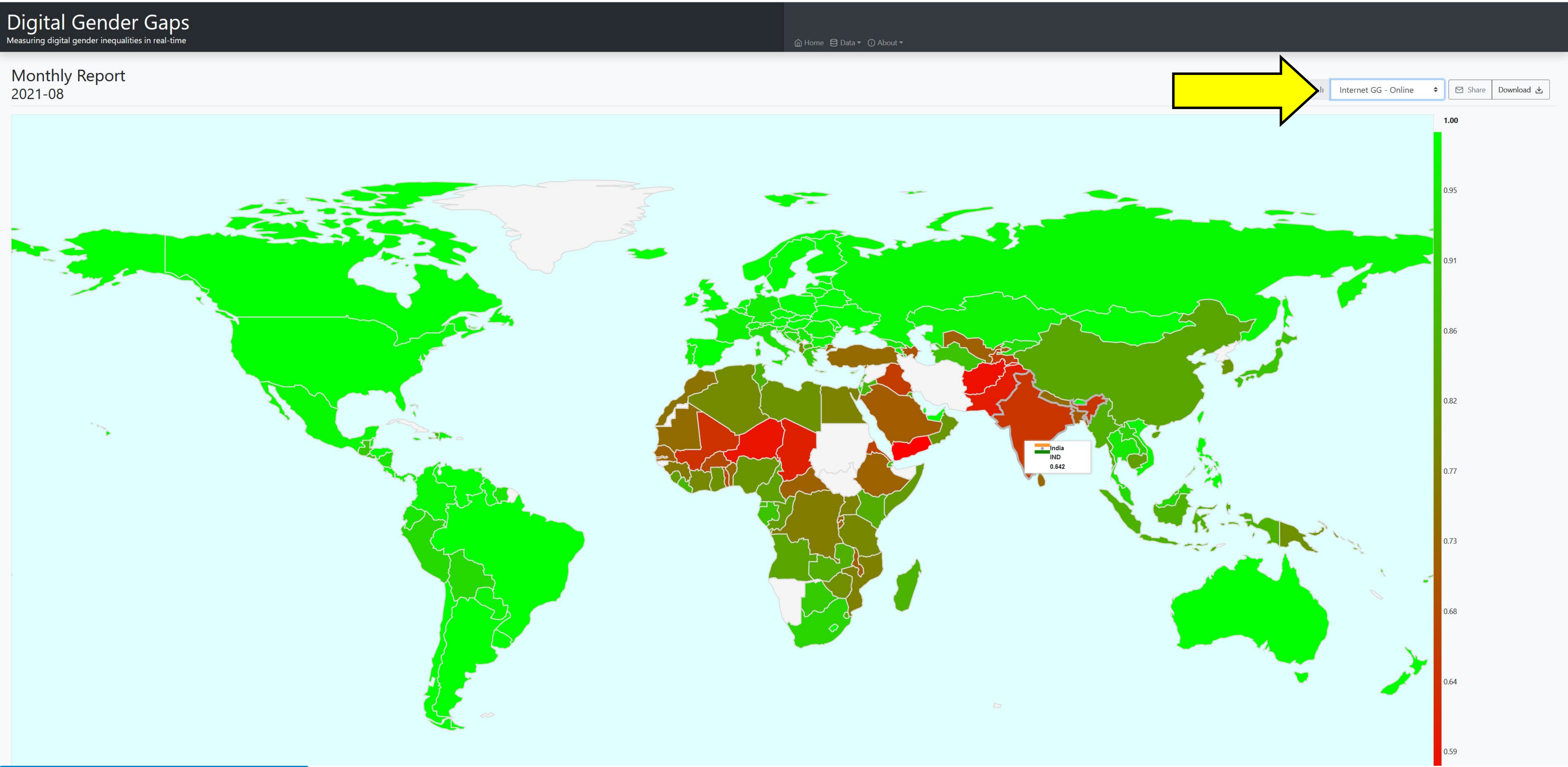
Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org



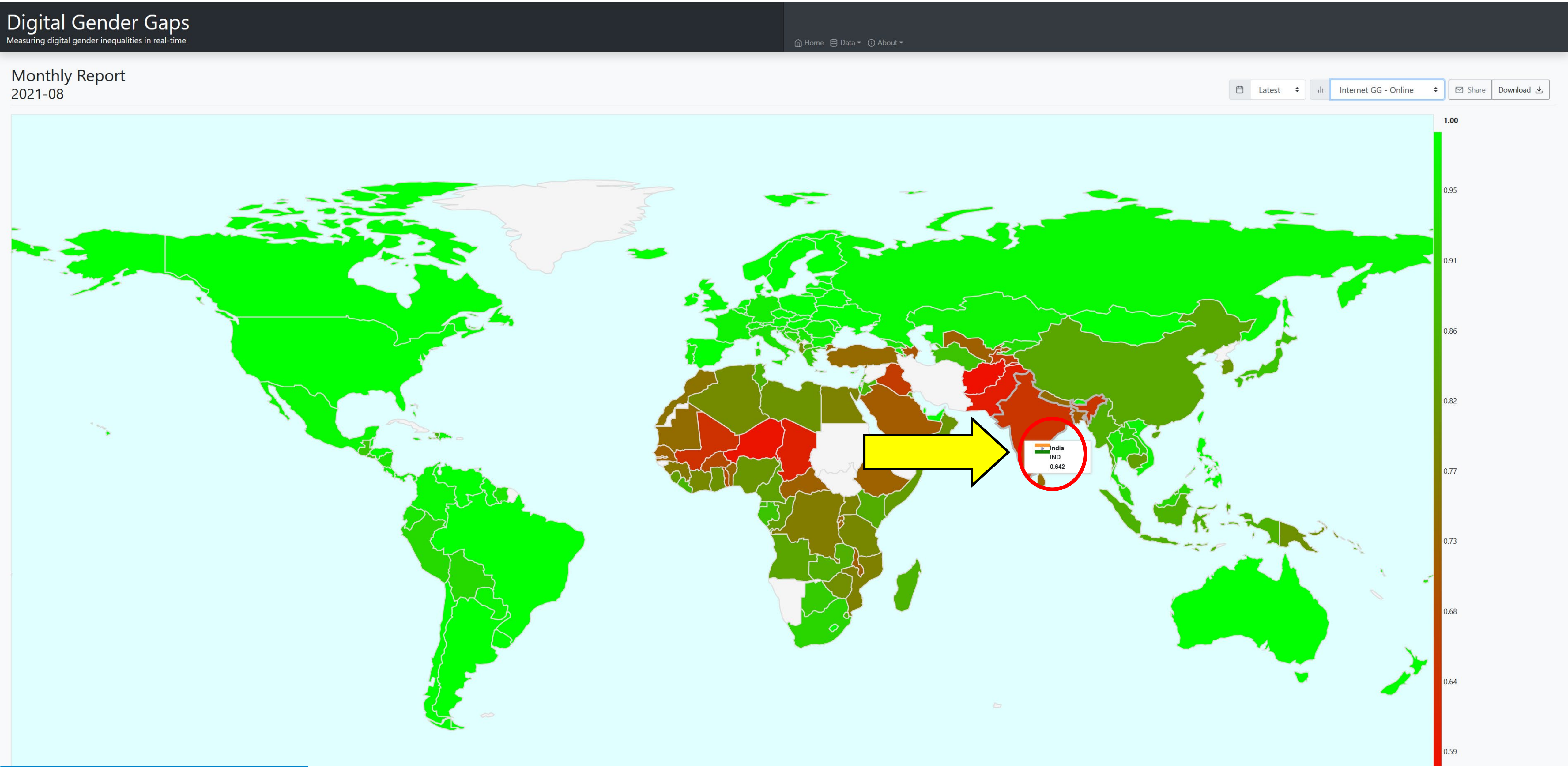
Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org



Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org



Internet Access Gender Gap Predictions

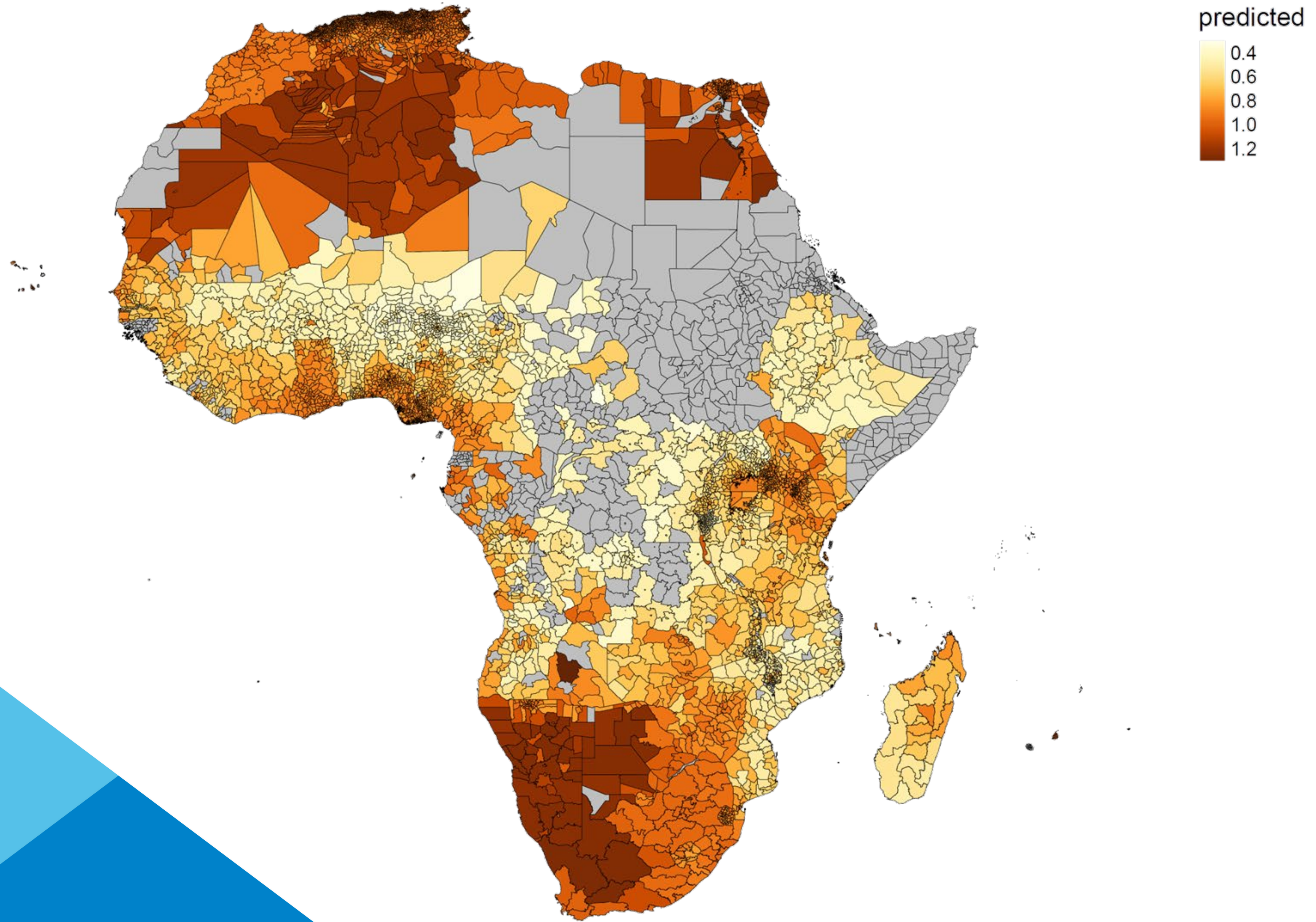
	Online Model	Onl.-Offl. Model	Offline Model
Intercept	0.933***	0.932***	0.933***
	(0.006)	(0.005)	(0.007)
FB GG (age 18+)	0.071***	0.093***	
	(0.011)	(0.017)	
log(GDP per capita)		0.018*	
		(0.008)	
GGGR – Literacy		−0.018	
		(0.016)	
GGGR – Education		−0.019	
		(0.019)	
Internet Penetration			0.040***
			(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
AD 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

MAPPING POVERTY

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.

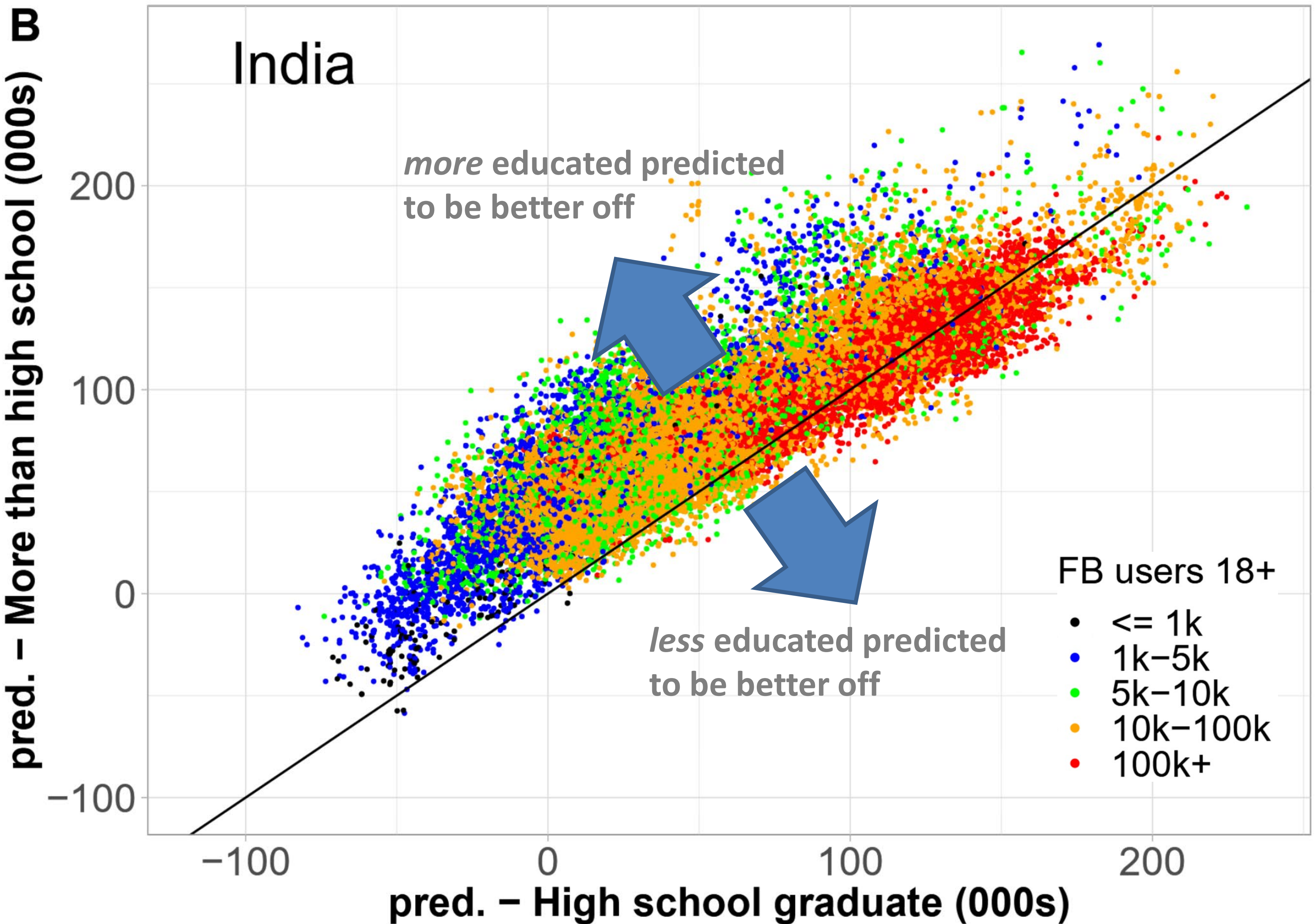
This is a dummy variable “is the location part of [name of region]”.

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)

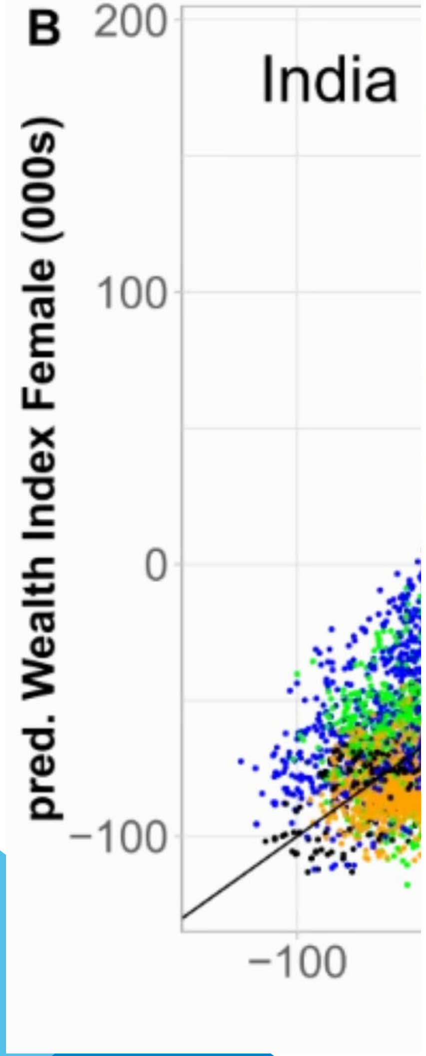
QCRI
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
جامعة حمد بن خليفة
HAMAD BIN KHALIFA UNIVERSITY

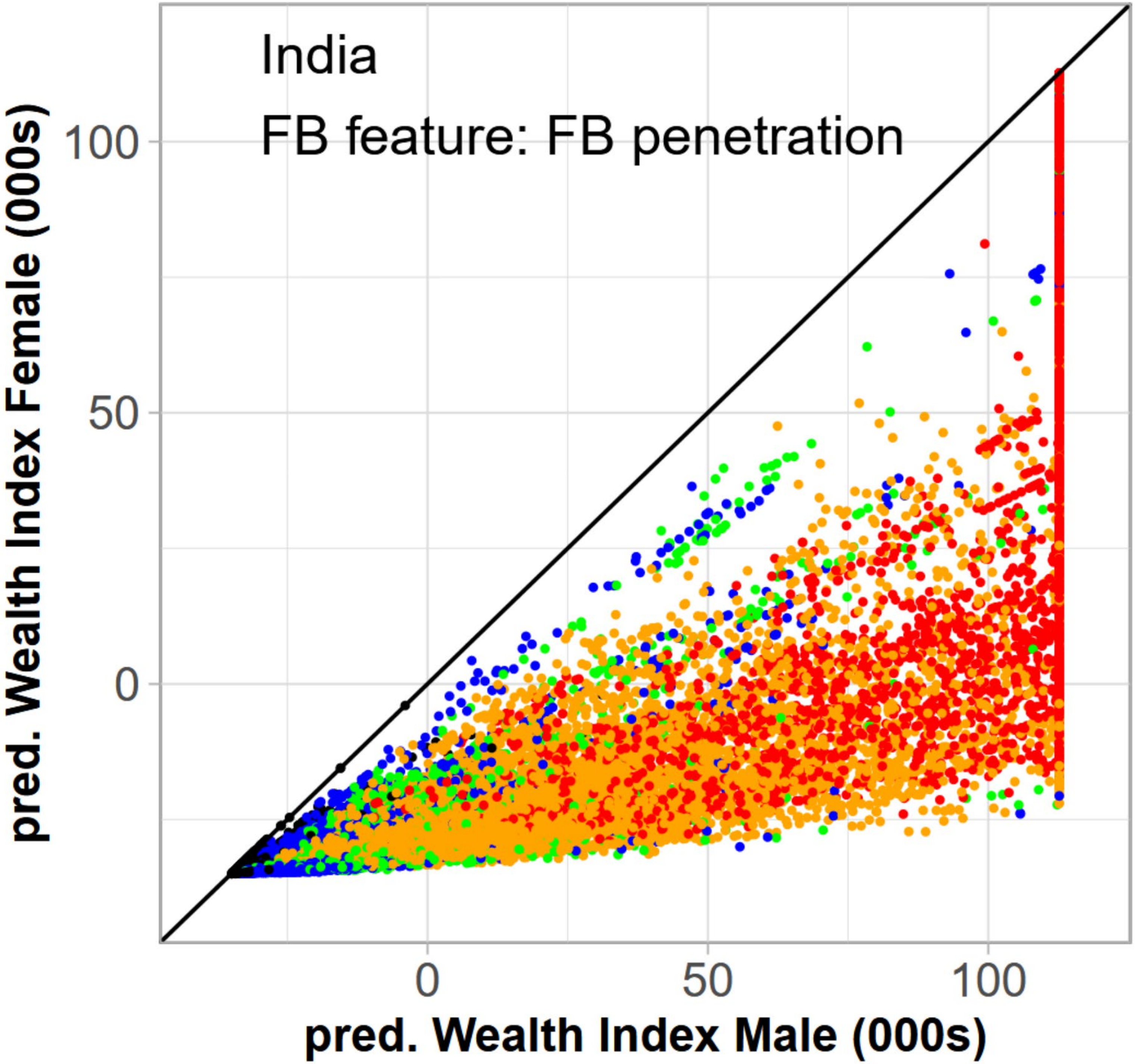
Education-Level-Disaggregated Predictions



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

Gend

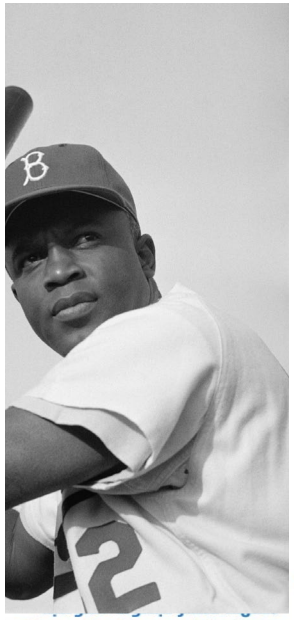




		score
p Index	112	0.668
tion and opportunity	149	0.354
ıent	112	0.962
	150	0.944
ıent	18	0.411

Out of 153

binson Effect



ar Computing Research Institute
جامعة حمد بن .
ıN 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

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

Reflections

- A rare form of “data sharing” with academics by Facebook
  – Has been used for auditing by privacy researchers
- Potential for white-washing of Facebook’s activity
  – Could draw attention away from hate speech and other issues
- Conflicting power dynamics and incentives
  – European Commission wants to both regulate and work with FB
- Great “bang for the buck” for capacity building
  – Less labor and resource-intensive than using satellite imagery

CfP Social Media Analysis for Humanitarian and Development Challenges

Special Track at ACM GoodIT

Three paper types:

1. Analysis and modeling papers
2. Experience reports and systems papers
3. Discussion and vision papers

Submission deadline	23 May 2022
Notification	20 June 2022
Conference	7-9 Sep 2022

Program Committee

Girmaw Abebe Tadesse, IBM
Joseph Aylett-Bullock, UN Global Pulse & Durham University
Aleksandra Berditchevskaia, Nesta
Ciro Cattuto, Data Science for Social Impact, ISI Foundation, Italy
Meeyoung Cha, KAIST & IBS
Samuel Fraiberger, World Bank
Vanessa Frias-Martinez, University of Maryland
Manuel García-Herranz, UNICEF
Rayid Ghani, CMU
Geoff Gilbert, University of Essex
Bahia Halawi, Data Aurora
Katherine Hoffmann-Pham, UN Global Pulse
Yennie Jun, UN Global Pulse
Bruno Lepri, Bruno Kessler Foundation
Yee Man Margaret Ng, University of Illinois
Leonardo Milano, UN OCHA
Trevor Monroe, World Bank
Ayako Kawano, UN Global Pulse
Wuraola Oyewusi, Data Science Nigeria
Daniela Paolotti, Data Science for Social Impact, ISI Foundation, Italy
Marzia Rango, IOM
Julie Ricard, Data-Pop Alliance
Kit Rodolfa, CMU
Bahman Rostami-Tabar, Forecasting for Social Good/Cardiff University
Albert Ali Salah, Utrecht University
Vedran Sekara, IT University of Copenhagen
Javier Teran, UN OCHA
Michele Vespe, European Commission, Joint Research Centre
Jeffrey Villaveces, iMMAP
Martin Waehlisch, UN DPPA
Emilio Zagheni, Max Planck Institute for Demographic Research
Eduardo Zambrano, IOM

https://www.socialmedia4good.org/

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