



What Advertising Data Tells us About the World

SICSS Istanbul

Ingmar Weber

[@ingmarweber](https://twitter.com/ingmarweber)

July 4, 2023

Social Media Advertising Data

Locations

People living in this location

Colombia

Norte de Santander

Include Search locations

Browse



Add locations in bulk

Age

18 65+

Gender

All Men Women

Detailed targeting

Include people who match

Behaviors > Expats

Lived in Venezuela (Formerly Expats - Venezuela)

Audience definition

Your audience selection is fairly broad.

Specific Broad

Estimated audience size: 134,000 - 157,600

Estimates may vary significantly over time based on your targeting selections and available data.

Creates a kind of real-time digital census over 2.3B FB users (+ IG + Messenger)

How many FB users in a location are young or women or from country X or using an expensive device or ... ?

Accessible free of charge!

What Could We Learn About Society?

- Poverty?
- Migration?
- Gender gaps?
- Public health?

Well-Documented API and Wrapper



Connectivity Mapping

Search this book...

Connectivity Maps using Advertisement Platforms

GETTING STARTED ON THE FB MARKETING API

1. The Facebook Ads Collection Pipeline
2. Exploring the Web Interface
3. Getting your Tokens
4. Basic Example with the FB Ads API
5. Creating a JSON for collection
6. Post-processing the collection
7. Plotting Maps
8. Recurrent Data Collections
9. Advanced Example 1 - World Collection - Countries
10. Advanced Example 2 - Ghana and similar peers

GETTING STARTED WITH LINKEDIN'S TOOLKIT

1. Exploring the web interface; downloading the package
2. Obtaining headers and cookies
3. Basic Example

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.

Contents

References:

Important Links (Please Screenshot)

Detailed Tutorial for FB and LinkedIn data

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

Detailed Steps for creating FB Access Token

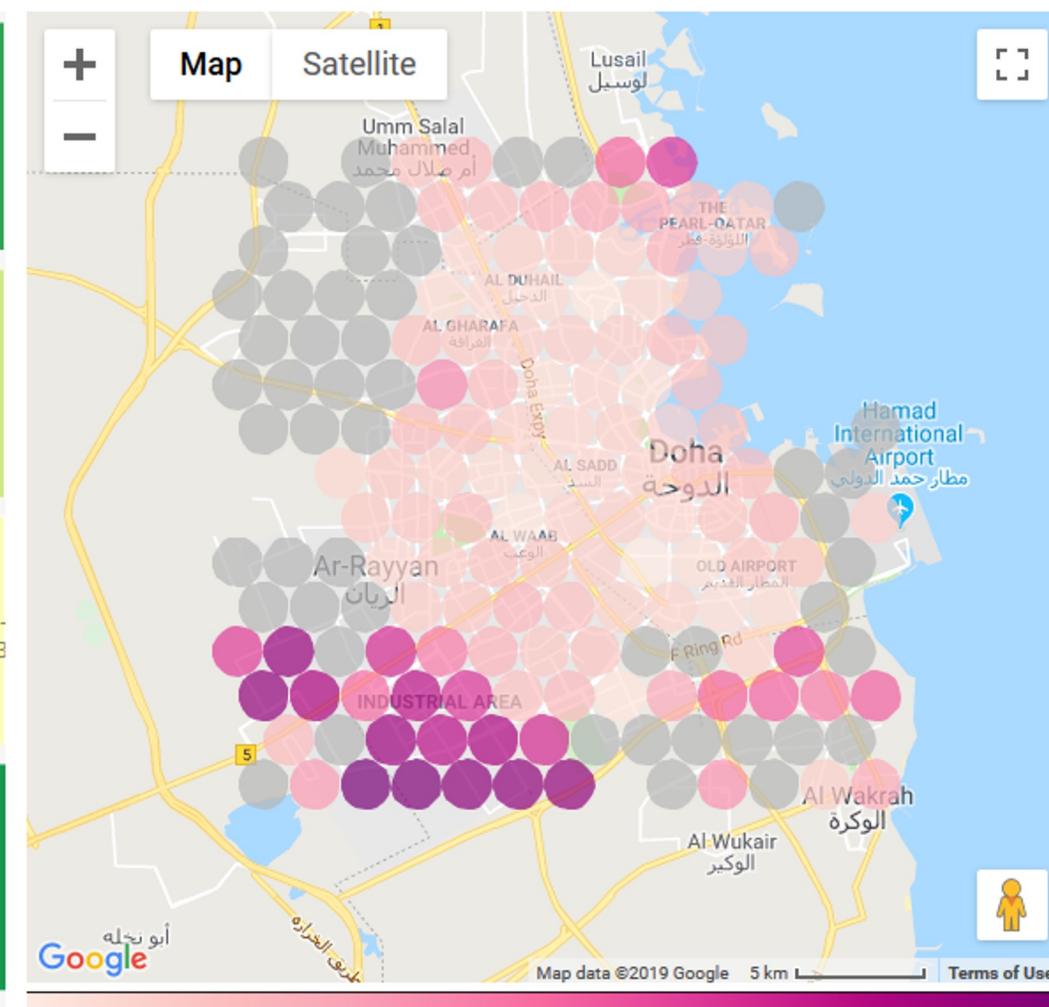
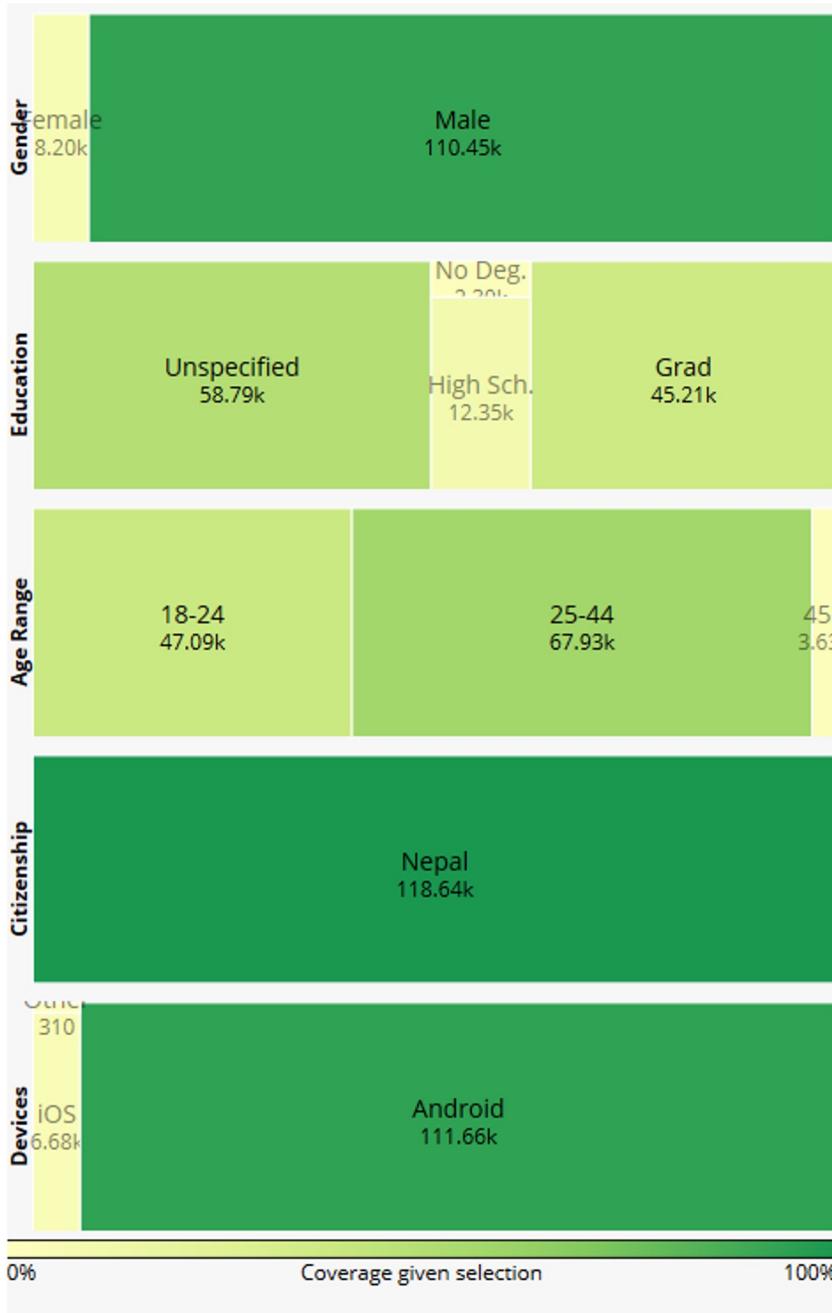
<https://tinyurl.com/krrksu2v>

(Crude) Instagram/Facebook API Tutorial

<https://tinyurl.com/mrycv8t4>

(Crude) Tiktok Advertising API Tutorial

<https://tinyurl.com/yvyxecu8>



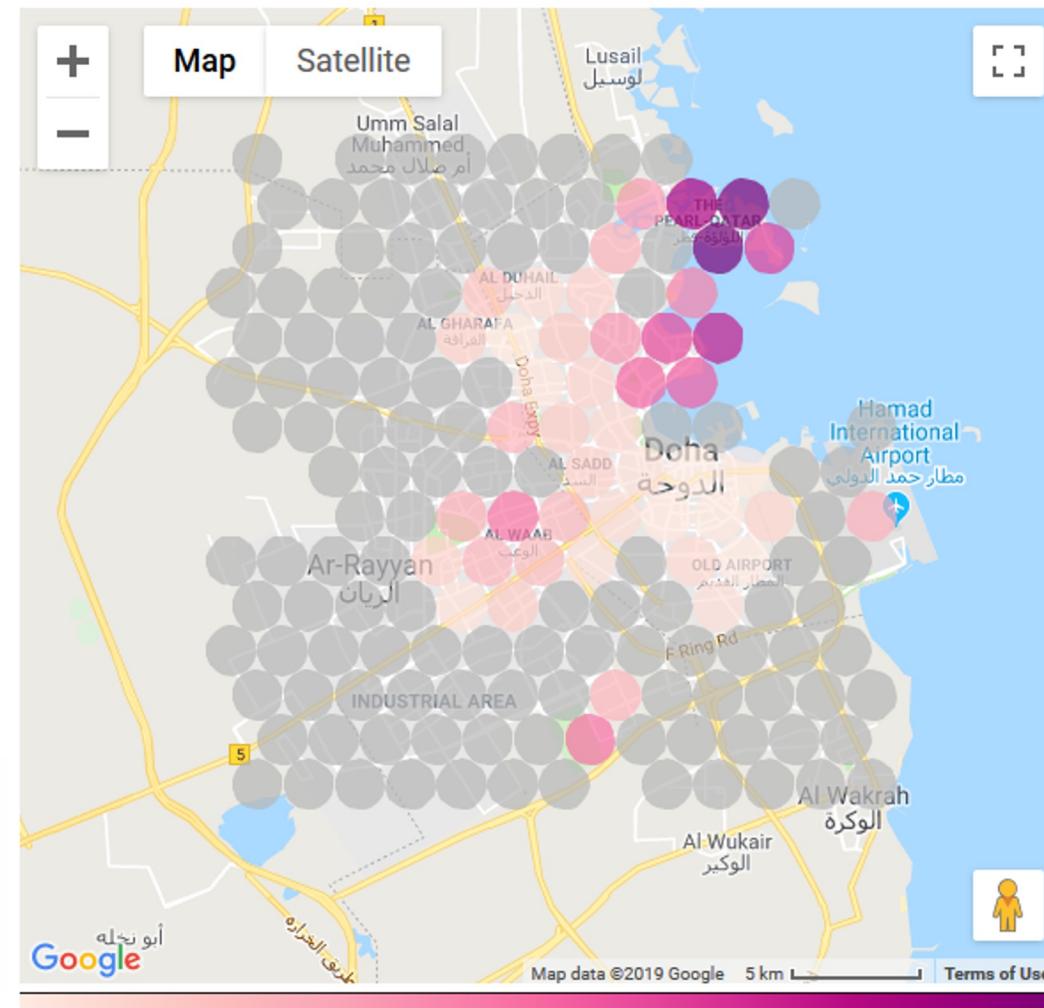
0% Matching given location 36.0%

All Locations

No Locations

Share what you see:





Monitoring the Venezuelan Exodus



Background on the Venezuelan Exodus

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

Unemployment > **40%** (est. 2019, IMF)

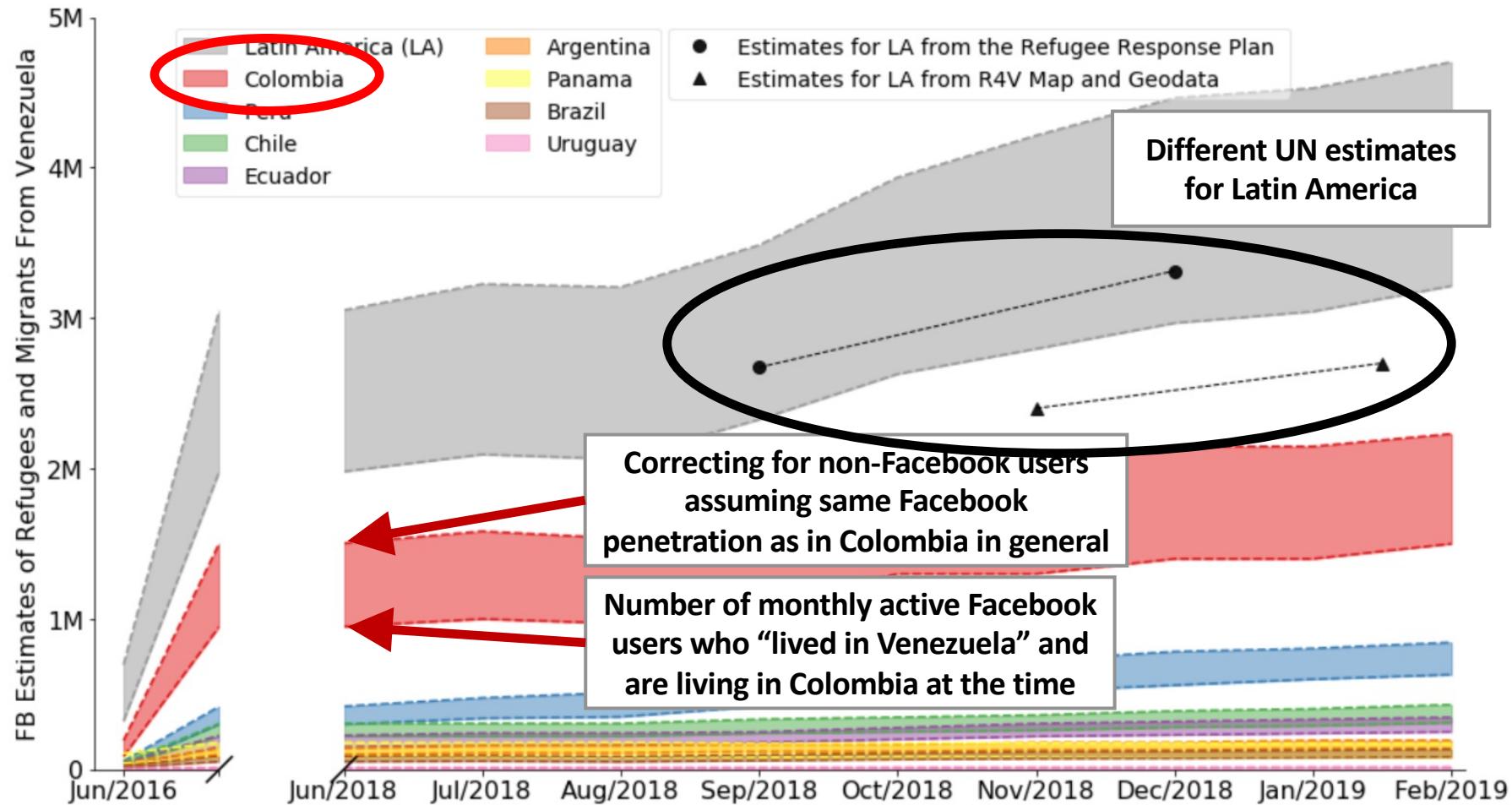
Minimum wage pays < **1000kcal/day**

> **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

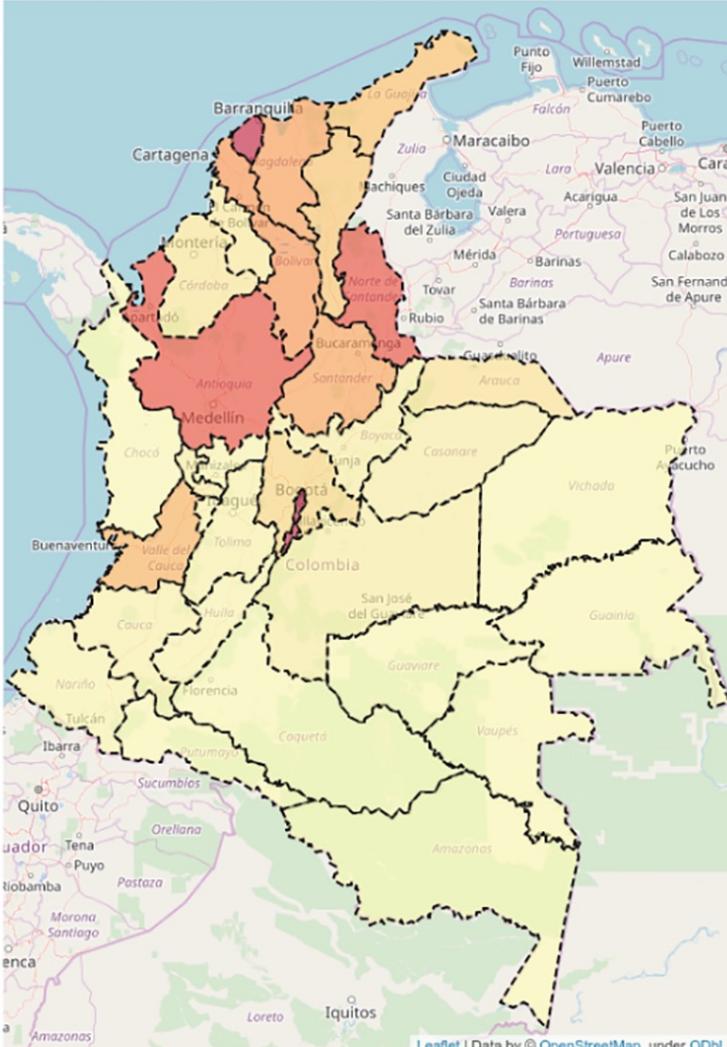
Monitoring Trends in Real-Time



Validation w/ (Few) Available Data



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



(b) Facebook - June 8, 2018

Previously Unavailable Estimates

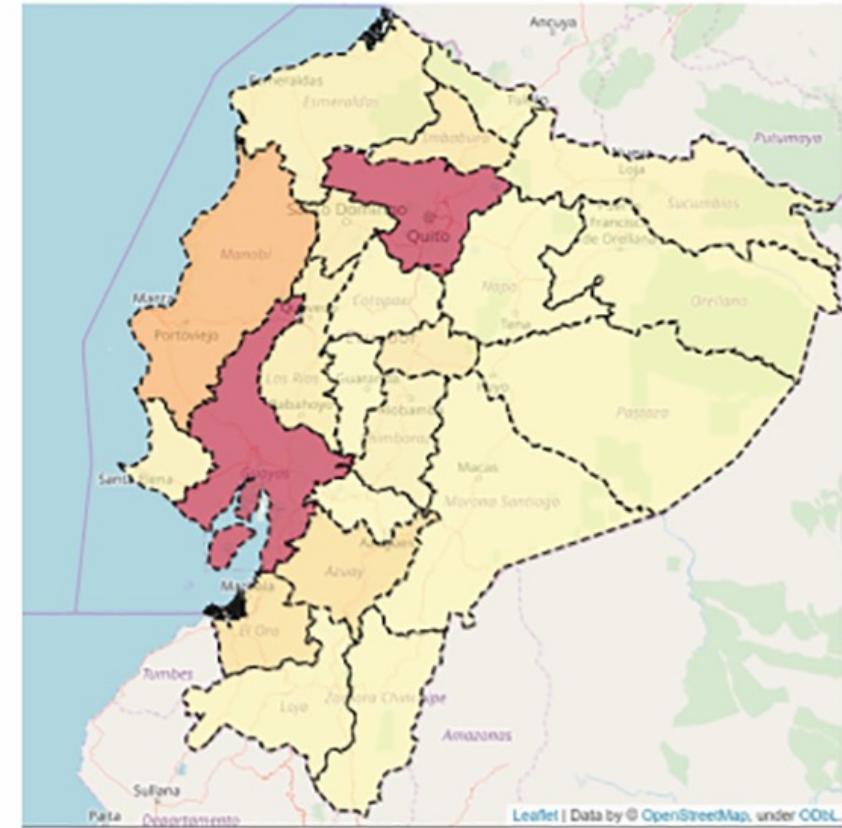


(a) Brazil



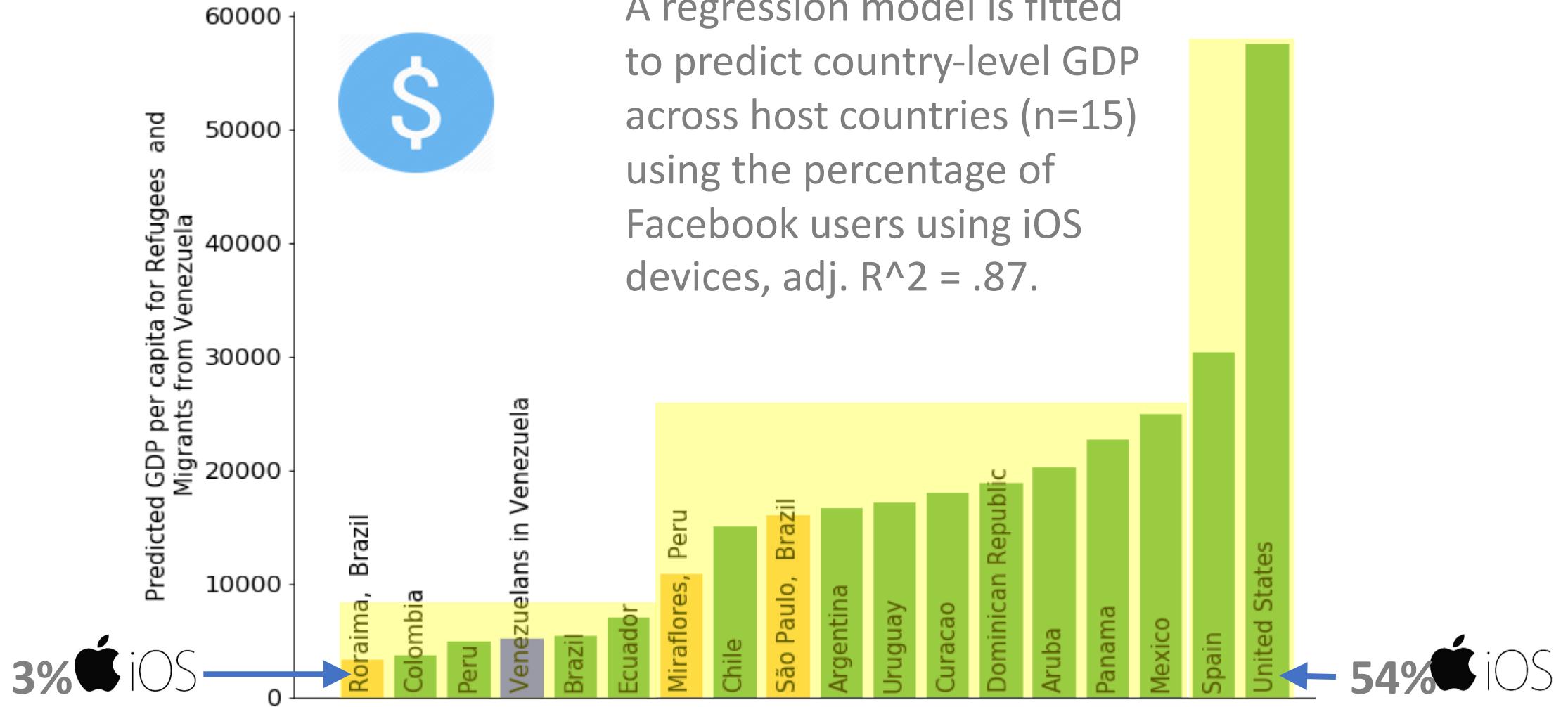
(b) Peru

Facebook. Feb 2019



(c) Ecuador

Predicted Income Based on OS



Operational Impact

“We relied on their [QCRI’s] research analysis for the Venezuela situation. Understanding population movement and density allows humanitarian agencies, like ours, to provide humanitarian assistance and protection in specific areas where population is more concentrated.”

Rebeca Moreno Jimenez
UNHCR Lead Data Scientist
and Innovation Officer

“Knowing this distribution helped redefine and amplify the geographical scope [...]. The recognition of the national scale of the crisis was particularly relevant for UNICEF to develop anti-xenophobia campaigns using Facebook’s chatbots.”

Natalia Adler
Former UNICEF Data, Research,
Policy Manager

Continued Operational Use

[Inicio](#)[Contexto](#)[Sectores Priorizados](#)[Proyectos](#)[Productos de Información](#)[Publicaciones](#)[Contáctenos](#)

LOS RESULTADOS

[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 vivían en Venezuela y ahora viven en el extranjero.

Total usuarios conectados

1,845,200

Usuarios conectados Febrero 15

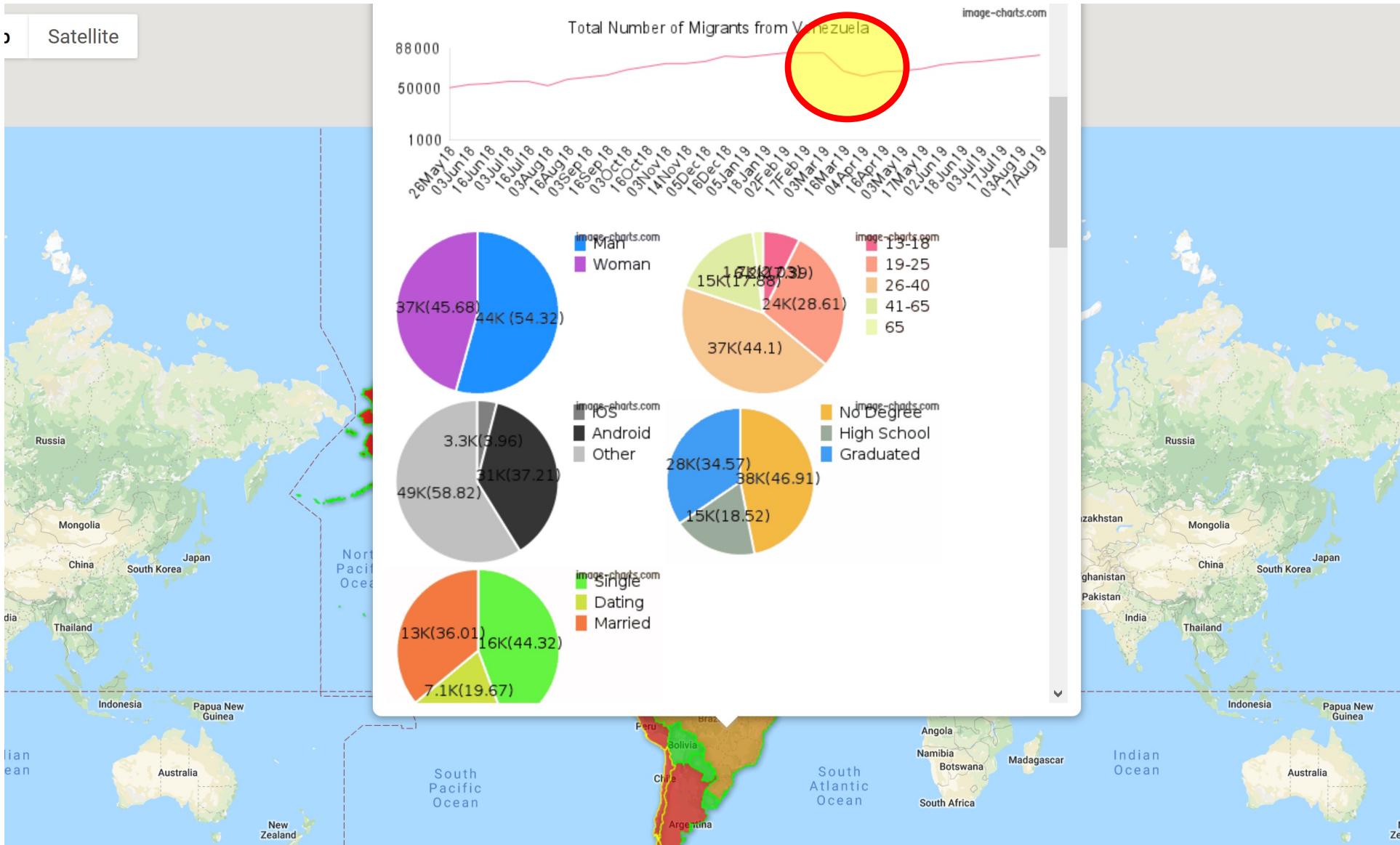
Dato Oficial Migración Colombia

1,729,537

Venezolanos en Colombia

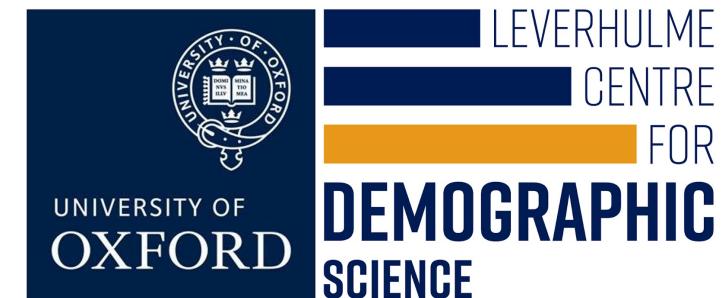
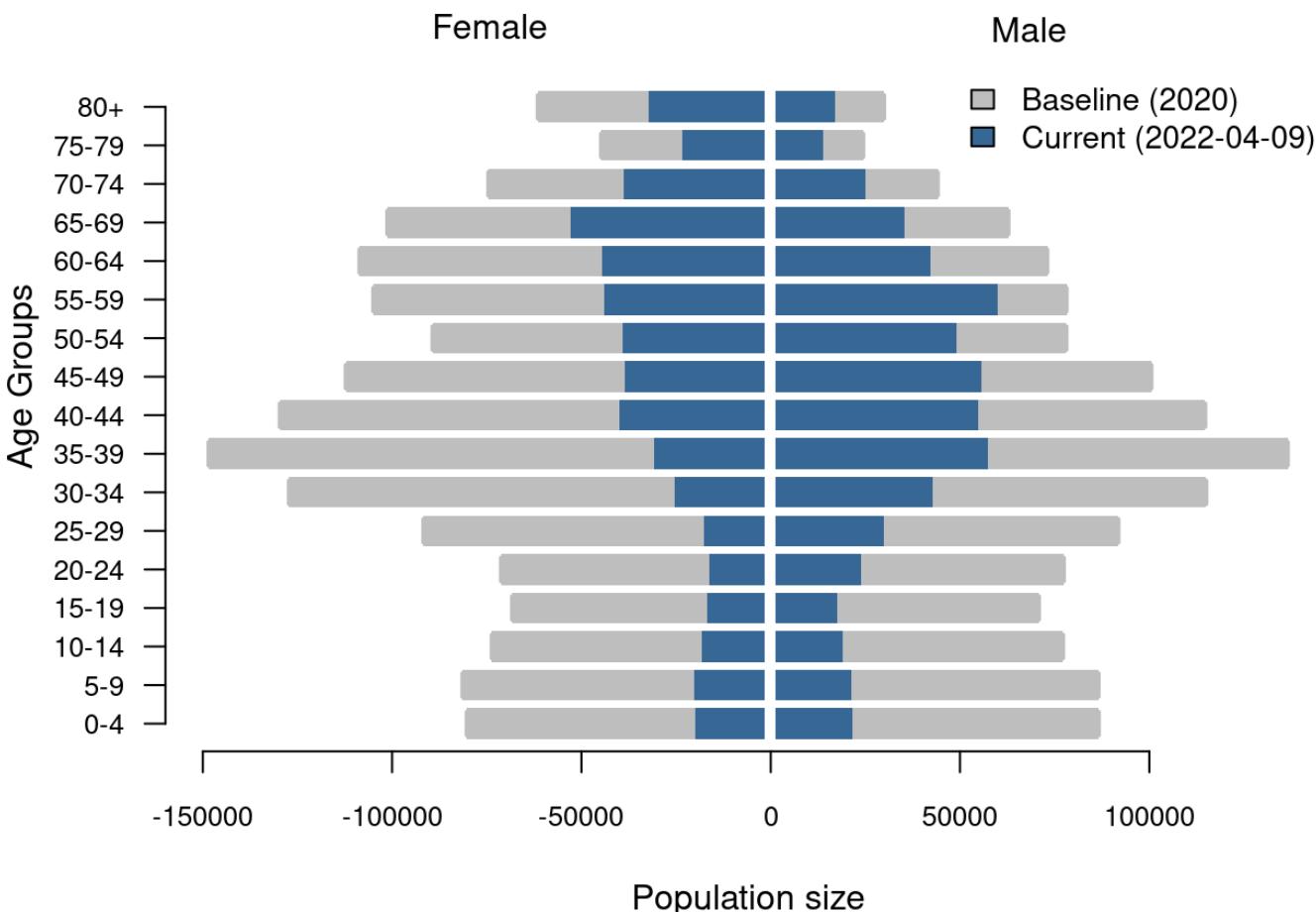
- Los datos son recolectados cada 15 días, mediante el API de mercadeo de Facebook. Los estimados mostrados están basados en los filtros utilizados y en el comportamiento de los usuarios en los últimos 30 días. Obtén mas información sobre el [alcance potencial](#)
- Los estimados presentados no están 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 para 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



What About Ukraine?

Changing demographics in the city of Kyiv



Ukraine Crisis: Monitoring population displacement through social media activity
<https://osf.io/preprints/socarxiv/6j9wq/>

Tracking Digital Gender Gaps

data2x



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

Digital Gender Gaps

Measuring digital gender inequalities in real-time

Home Data About ▾

Using big data to measure global gender gaps in internet and mobile access

Tracking progress on gender inequalities in internet and mobile access and use is more important than ever to ensure that women benefit from the digital revolution. Data on gender gaps in internet and mobile phone use and access are significantly lacking geographical coverage, comparability, and are slow to be updated.

We show how big data can help close this gender data gap and measure progress towards this important development goal in real-time.

Latest indicators

Check out the latest internet and mobile gender gap indicators.

Reports »

Project details

See the background of the project, and an overview of how the data is collected and processed.

Project »

Team

Meet the project members.

Team »

A collaboration between

UNIVERSITY OF OXFORD

Qatar Computing Research Institute
جامعة حمد بن خالفة
HAMAD BIN KHALIFA UNIVERSITY

With support from

data2x

Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org

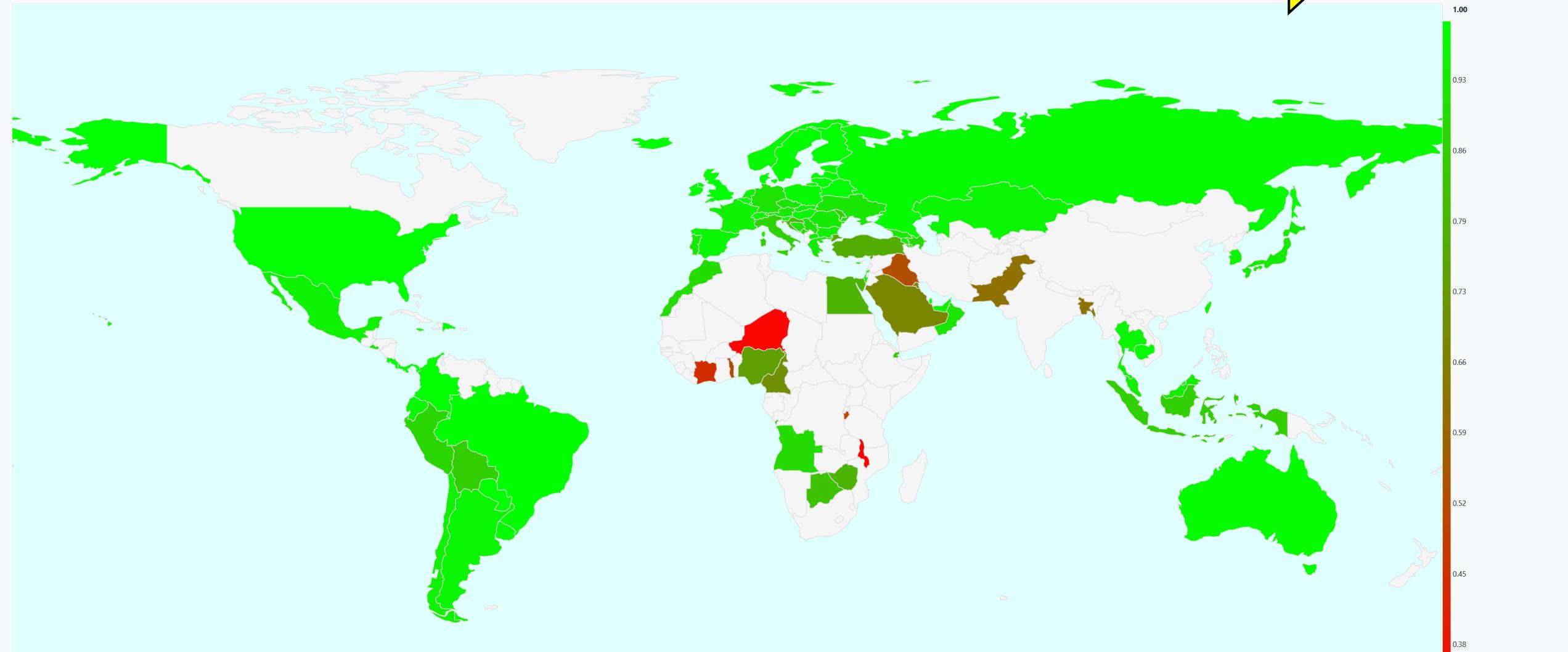
Digital Gender Gaps
Measuring digital gender inequalities in real-time

Monthly Report
2022-03

Home Data About

Internet GG - ITU

Share Download



Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org

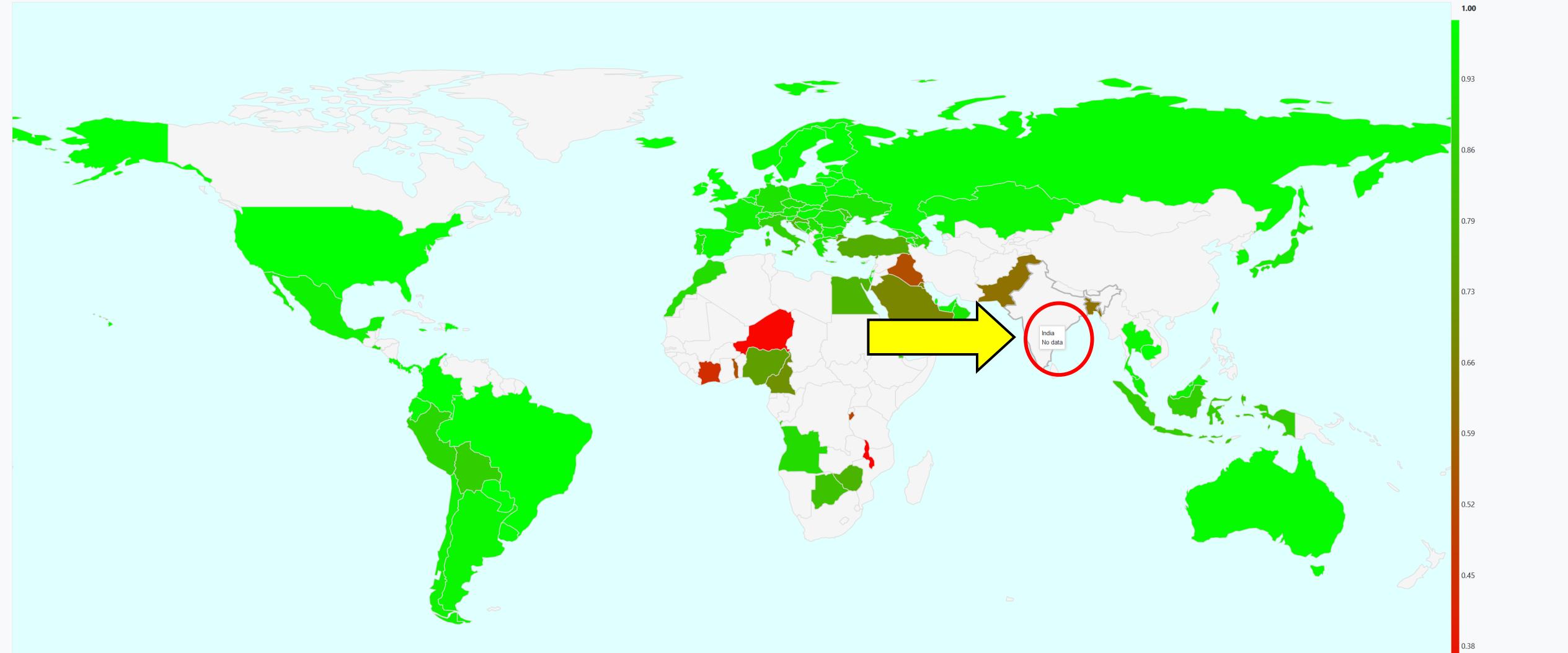
Digital Gender Gaps

Measuring digital gender inequalities in real-time

Home Data About

Monthly Report
2022-03

Latest Internet GG - ITU Share Download



Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org

Digital Gender Gaps
Measuring digital gender inequalities in real-time

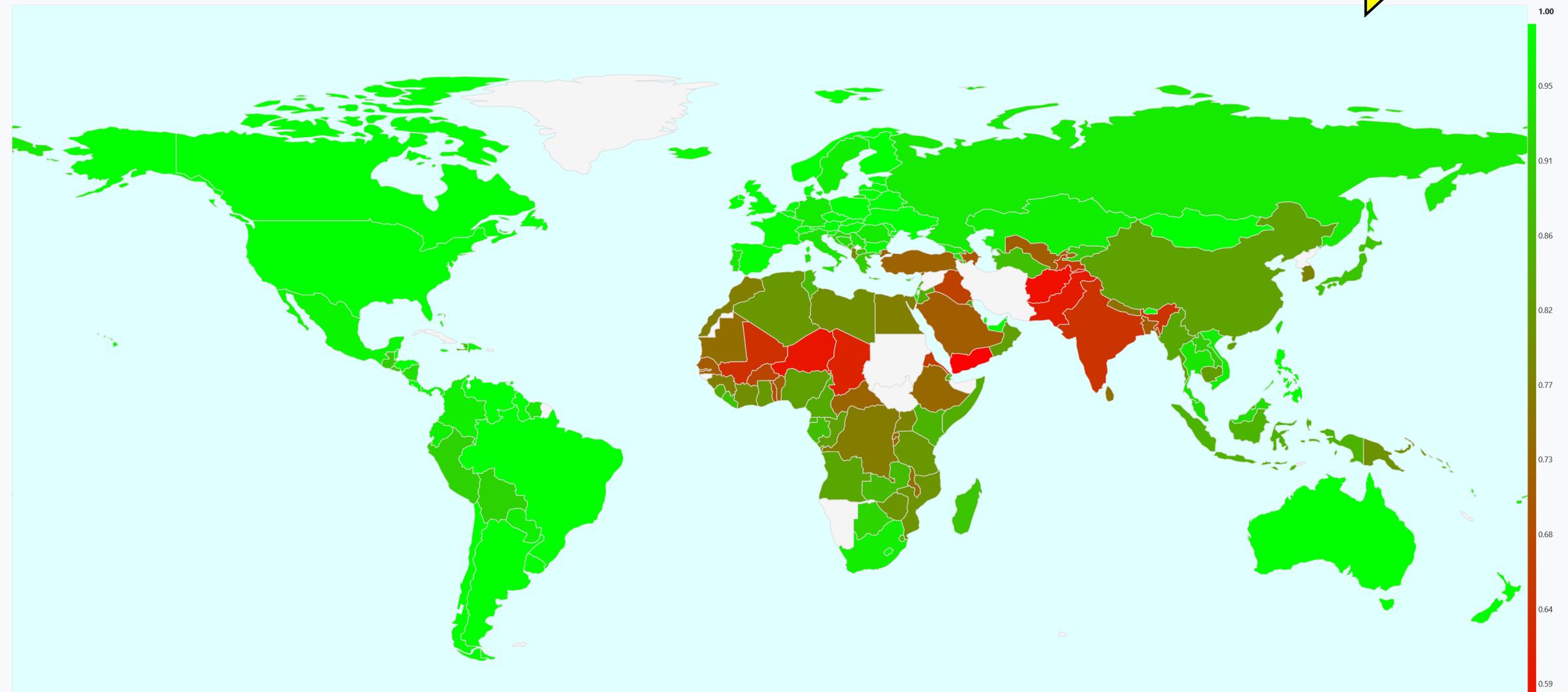
Home Data About

Monthly Report
2022-02

Internet GG - Online

Share

Download



Joint work with Masoomali Fatehkia and Ridhi Kashyap

www.digitalgendergaps.org

Digital Gender Gaps

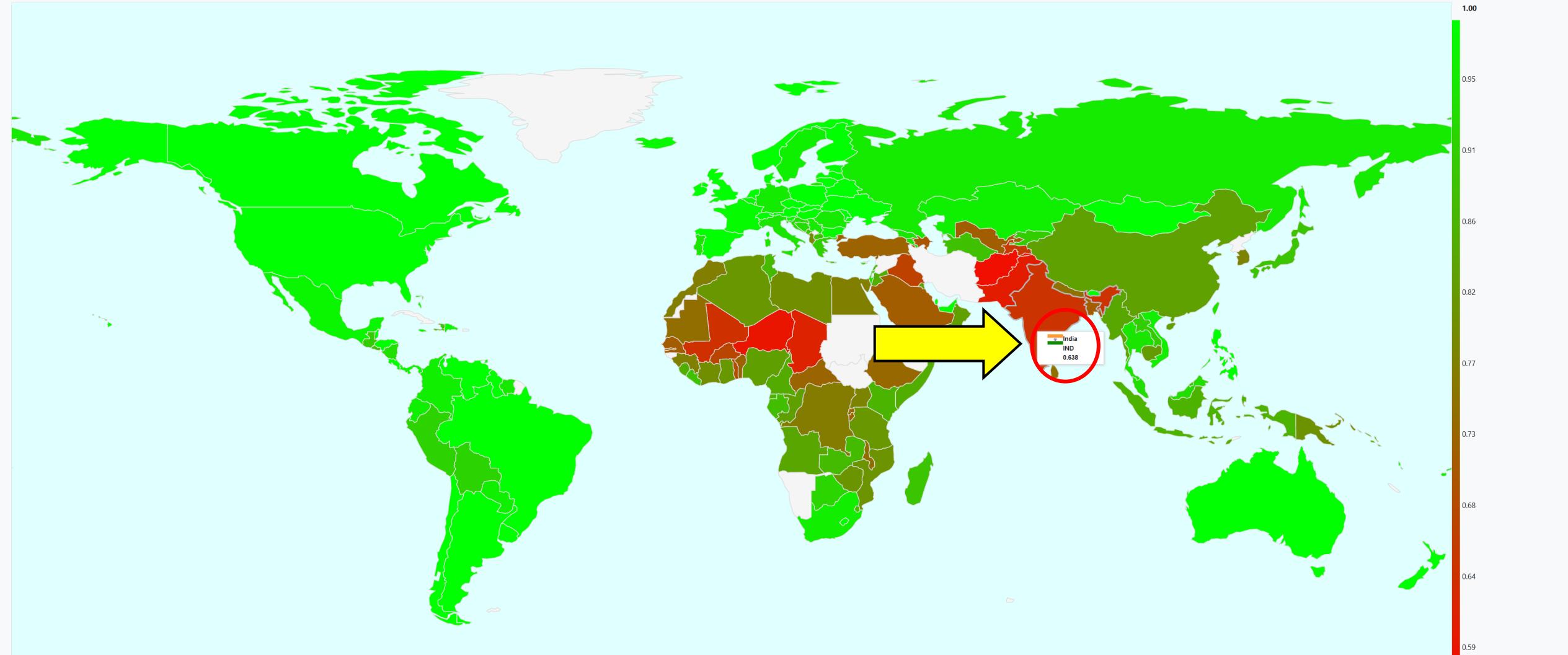
Measuring digital gender inequalities in real-time

Home Data About

Monthly Report

2022-02

2022-02 Internet GG - Online Share Download



Internet Access Gender Gap Predictions

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

Part of SDGs Today Portal

[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**

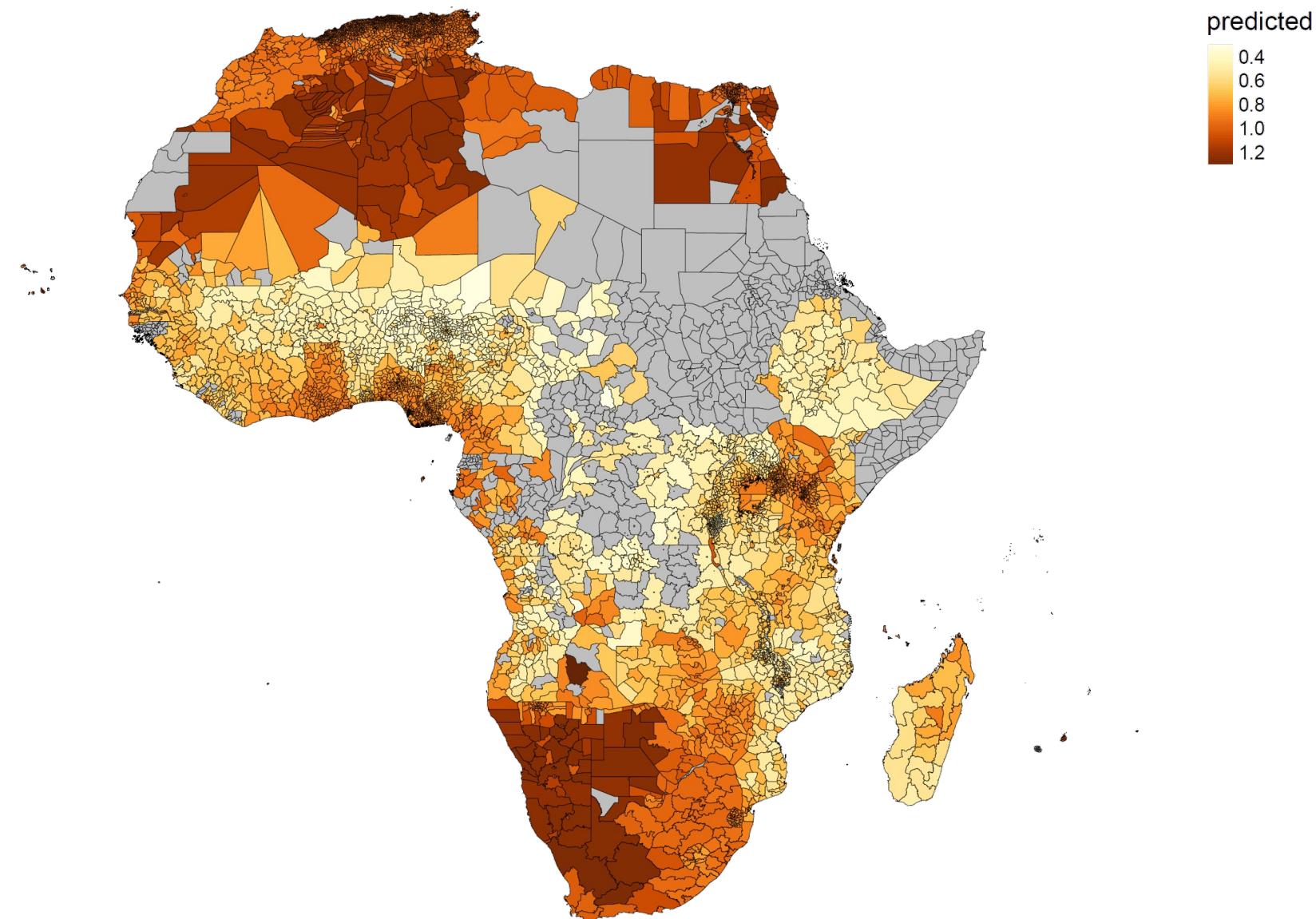
Inter-Parliamentary Union

[SDG 5 - Gender Equality](#)**Digital Gender Gap**

University of Oxford, QCRI, Data2X

[SDG 5 - Gender Equality](#)

Ongoing: Subnational Digital Gender Gaps in Africa



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

91% population coverage

Mapping of Wealth and Poverty



Background on Poverty Mapping



Data

Survey mean consumptio...

Search data e.g. GDP, population, Indonesia

1

“No Poverty” is the first of the 17 SDGs

Available data is often outdated

Lack of spatially granular data at the sub-city level

Lack of gender disaggregated data

All Countries and Economies

| Country | Most Recent Year | Most Recent Value |
|---------|------------------|-------------------|
| Mali | 2009 | 2.8 |
| Nepal | 2010 | 5.2 |
| Jordan | 2010 | 16.9 |

Ground Truth and Training Data



- Representative sample of ~20 households for each “cluster”
 - Philippines (PH): n=1.2k, India (IN): n=28k
 - Sampled to be nationally representative
- Asset ownership based “Wealth Index”
 - Ownership of fridge, motorcycle, mobile phone, sanitation, ...
 - Combined using Principal Component Analysis
- Most recent data PH: 2017; IN: 2015-2016
 - “now-casting” (FB Data is from 2019)

Less-Than-Golden Ground Truth

- Sampling noise
 - Wealth index depends on particular households
 - Expected $R^2 = .955/.973$ (PH/IN, bootstrap estimate)
- Spatial perturbation
 - True location is (x,y) , but reported at (x',y') to protect privacy
 - Expected $R^2 = .885/.860$ (PH/IN, simulations)
- Combined
 - **Expected $R^2 = .845/.838$ (PH/IN)**
 - “Expected upper bound”, “explainable variance”

Standard Regression Setup

The DHS Wealth Index is the target (y)

Use the FB data as features (x)

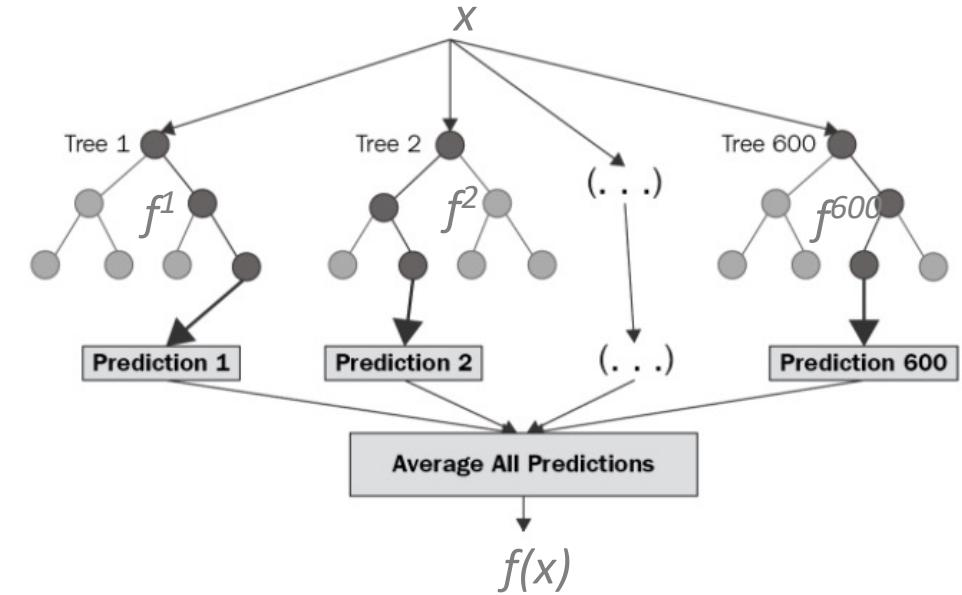
Try to find a function f such as $y \approx f(x)$

Use a gradient boosting regression trees as f

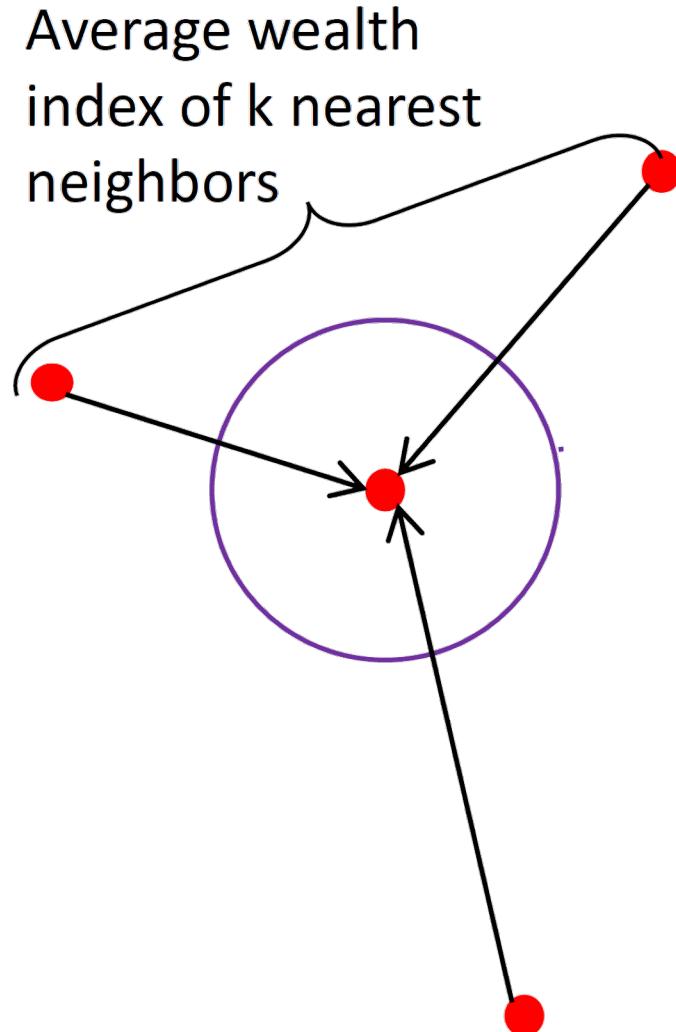
Worked better than linear methods

Fitted and evaluated using 10-fold cross validation.

No innovation on this part



Baseline: Interpolation with Nearest Ground Truth



| | $k = 1$ | $k = 3$ | $k = 5$ | $k = 10$ |
|-------------|---------|---------|---------|----------|
| Philippines | 0.597 | 0.686 | 0.687 | 0.681 |
| India | 0.739 | 0.793 | 0.796 | 0.788 |

r correlation between ground truth WI and average of k NN

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 |

Recall: upper bound R² = .845/.838 (PH/IN) (due to noise)

Beyond Accuracy: Who's Left Behind

“Leave no one behind (LNOB) is the central, transformative promise of the 2030 Agenda for Sustainable Development and its Sustainable Development Goals (SDGs).”

Common in Machine Learning to drop sparse cases

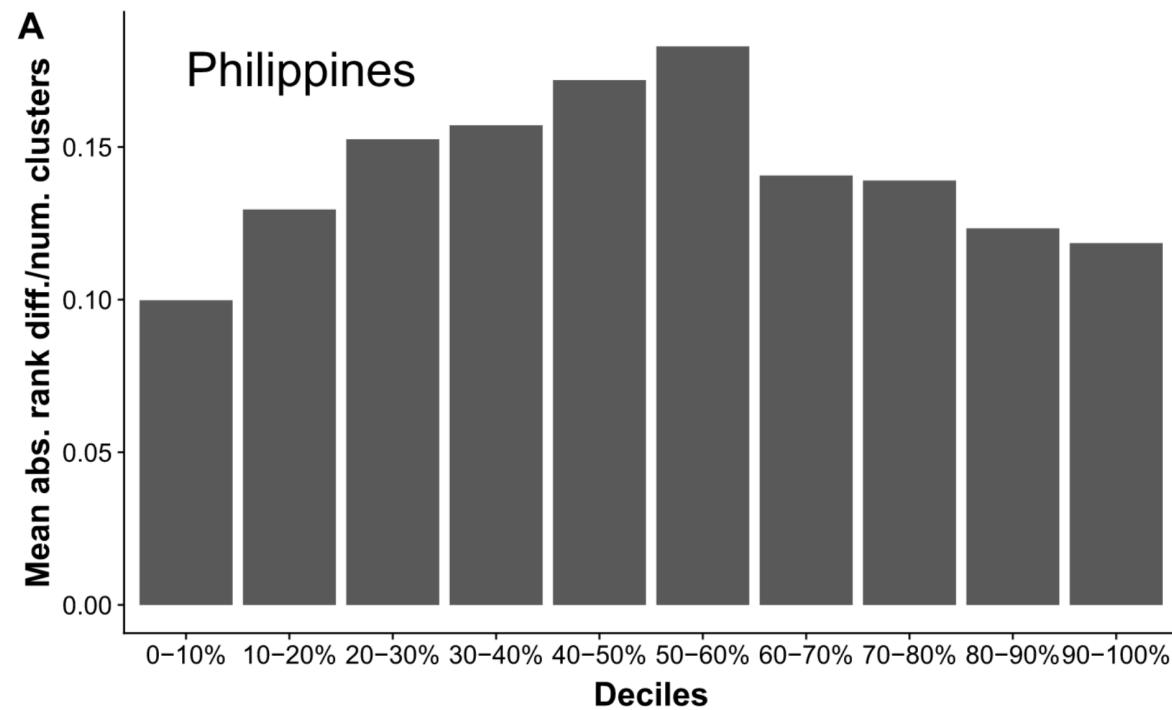
Table 1 Breakdown of the data for each country for clusters with at least one surveyed household

| | Philippines | India |
|--------------------------------------------------------|-------------|---------------|
| Number of DHS clusters | 1249 | 28,524 |
| Clusters missing geo-location | 36 | 131 |
| Geo-located DHS clusters | 1213 | 28,393 |
| Clusters with <100 FB users 18+ | 8 | 350 |
| Clusters with \geq100 FB users 18+ | 1205 | 28,043 |
| Clusters with >1000 FB users 18+ | 1043 | 25,316 |
| Median number of households surveyed (DHS) | 23 | 21 |

Had to drop 8 (of 1213 for PH, 0.7%) and 350 (of 28,393 for IN, 1%) clusters as no FB data could be collected. Dropped locations about $\frac{1}{2}$ standard deviation below average Wealth Index: some poor/rural areas removed.

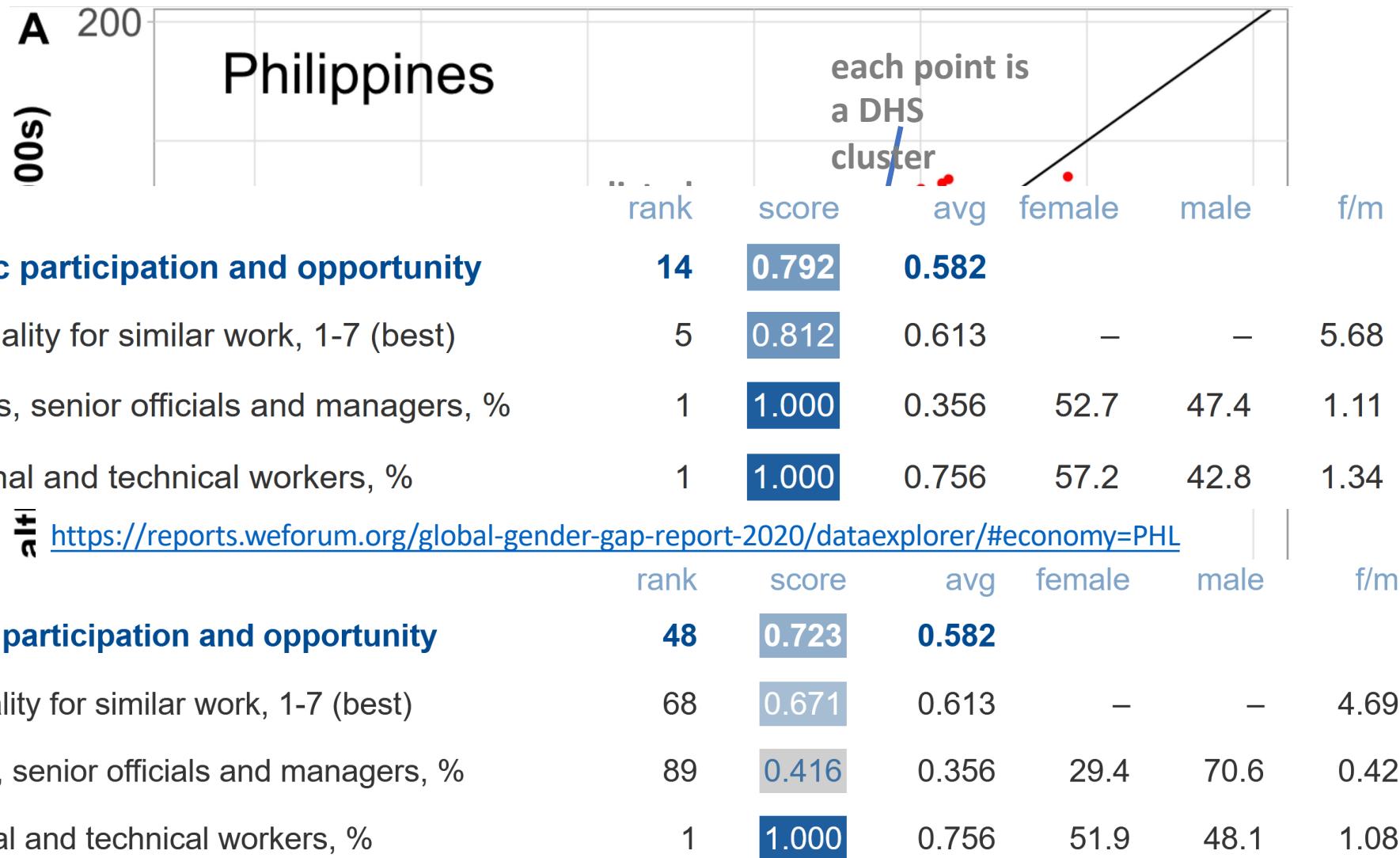
Beyond Accuracy: Performance Across Spectrum

Does the model only work on the poor or the rich?



Model most wrong where ranking is hardest (the “middle”)
But no performance collapse for lowest Wealth Index

Beyond Accuracy: Gender-Disaggregated Predictions

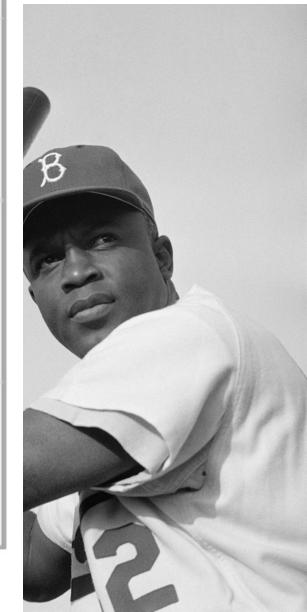
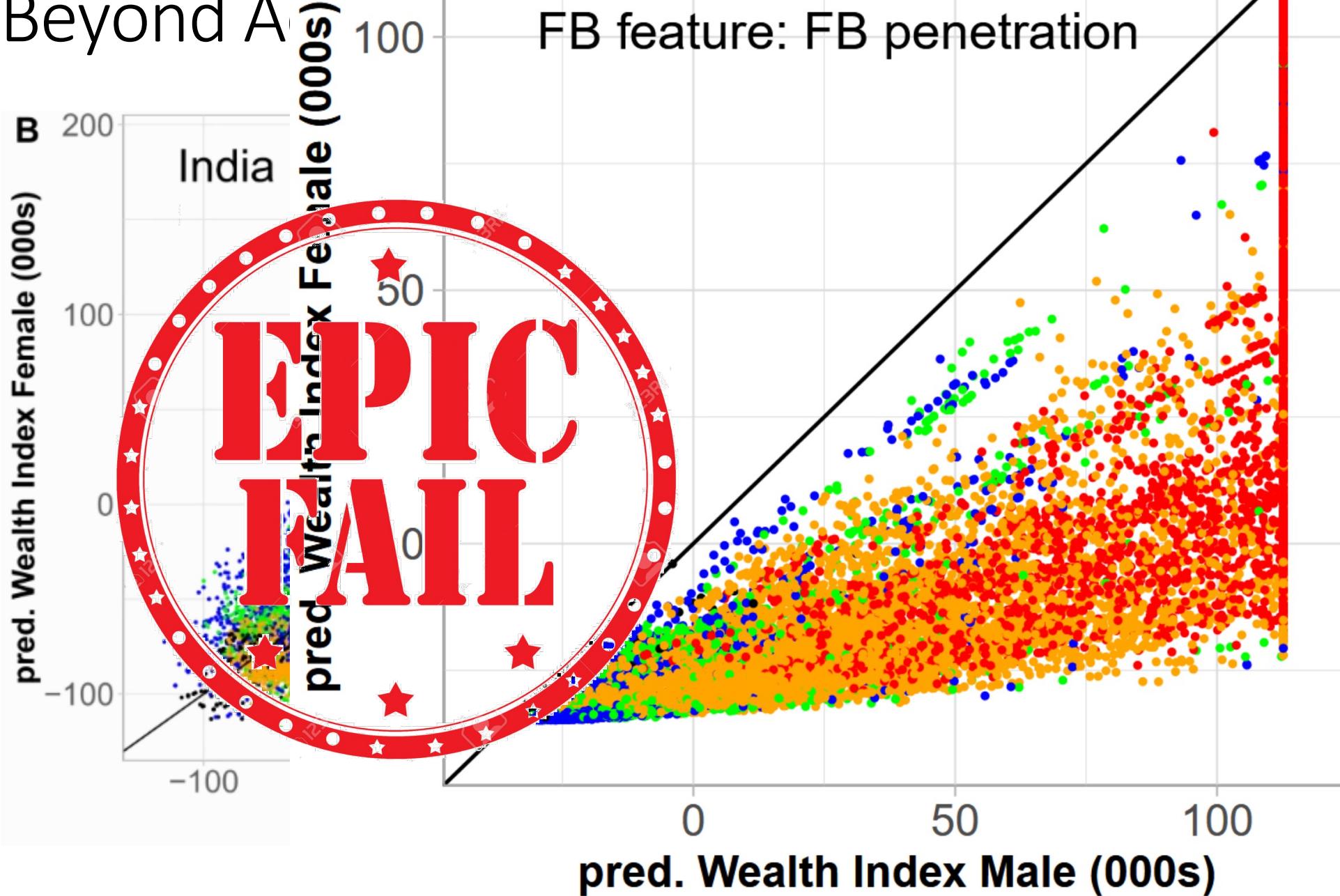


Beyond A

India

FB feature: FB penetration

dictions



Used as Part of UNDP's Operational Planning

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/>

Questions? Comments?

Important Links Again (Please Screenshot)

Detailed Tutorial for FB and LinkedIn data

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

Detailed Steps for creating FB Access Token

<https://tinyurl.com/krrksu2v>

(Crude) Instagram/Facebook API Tutorial

<https://tinyurl.com/mrycv8t4>

(Crude) Tiktok Advertising API Tutorial

<https://tinyurl.com/yvyxecu8>

More Links

- Data visualizations
 - <https://fb-nyc.qcri.org/>, <https://fb-doha.qcri.org/>, <https://fb-lagos.qcri.org/>
- Advertising platforms to sign up for
 - <https://de-de.facebook.com/business/tools/ads-manager>
 - <https://www.tiktok.com/business/de>
 - <https://forbusiness.snapchat.com/>
 - <https://www.linkedin.com/campaignmanager/>
 - <https://map.snapchat.com/>

Thanks!

iweber@cs.uni-saarland.de