

# Mapping subnational gaps in internet and mobile adoption using social media data

Digital Demography

Casey F. Breen<sup>1</sup>    Masoomali Fatehkia<sup>2</sup>    Jiani Yan<sup>1</sup>  
Xinyi Zhao<sup>1</sup>    Douglas R. Leasure<sup>1</sup>    Ingmar Weber<sup>3</sup>    Ridhi Kashyap<sup>1</sup>

July 4, 2024

---

<sup>1</sup>University of Oxford

<sup>2</sup>Qatar Computing Research Institute

<sup>3</sup>Saarland University

# Benefits of digital revolution

- ▶ The **digital revolution** has ushered in tremendous societal and economic benefits

# Benefits of digital revolution

- ▶ The **digital revolution** has ushered in tremendous societal and economic benefits
  - ▶ Lower gender inequality, lower maternal/child mortality, higher contraception (Rotondi et al., 2020)

# Benefits of digital revolution

- ▶ The **digital revolution** has ushered in tremendous societal and economic benefits
  - ▶ Lower gender inequality, lower maternal/child mortality, higher contraception ([Rotondi et al., 2020](#))
  - ▶ Boost social connectivity, social learning, access to vital services ([Unwin, 2009](#); [DiMaggio and Hargittai, 2001](#); [Suri and Jack, 2016](#))

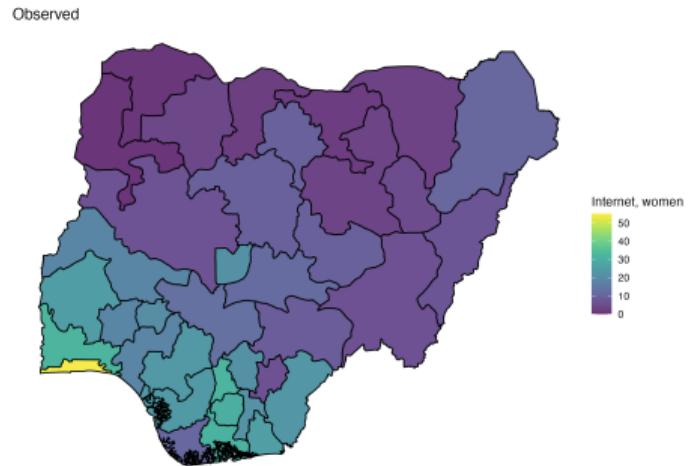
# Benefits of digital revolution

- ▶ The **digital revolution** has ushered in tremendous societal and economic benefits
  - ▶ Lower gender inequality, lower maternal/child mortality, higher contraception ([Rotondi et al., 2020](#))
  - ▶ Boost social connectivity, social learning, access to vital services ([Unwin, 2009](#); [DiMaggio and Hargittai, 2001](#); [Suri and Jack, 2016](#))
  - ▶ Increases levels of education, economic benefits ([Hjort and Poulsen, 2019](#); [Kho, Lakdawala and Nakasone, 2018](#); [Kharisma, 2022](#))

# Benefits of digital revolution

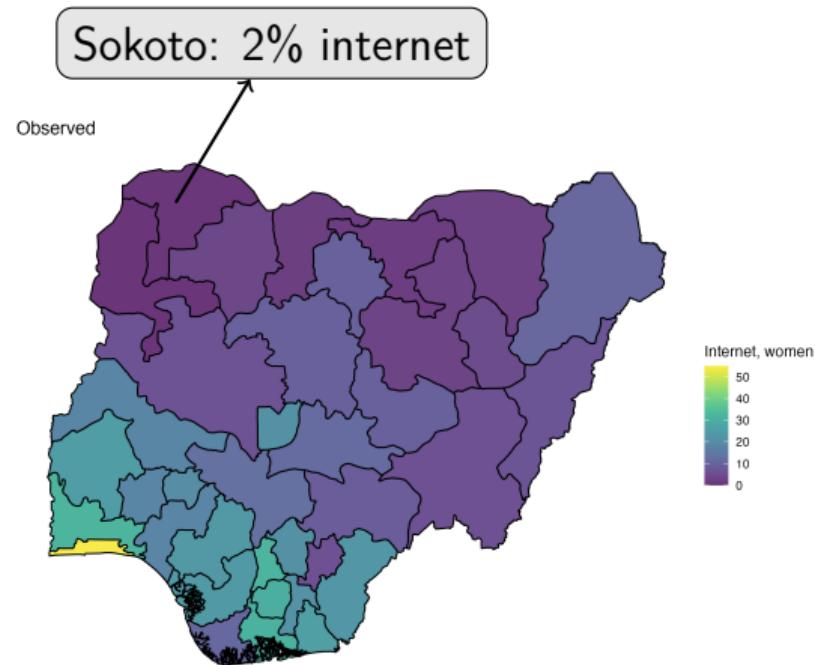
- ▶ The **digital revolution** has ushered in tremendous societal and economic benefits
  - ▶ Lower gender inequality, lower maternal/child mortality, higher contraception ([Rotondi et al., 2020](#))
  - ▶ Boost social connectivity, social learning, access to vital services ([Unwin, 2009](#); [DiMaggio and Hargittai, 2001](#); [Suri and Jack, 2016](#))
  - ▶ Increases levels of education, economic benefits ([Hjort and Poulsen, 2019](#); [Kho, Lakdawala and Nakasone, 2018](#); [Kharisma, 2022](#))
- ▶ Yet large **inequality** in who has access to digital technology...

# Adoption of digital technology varies geographically

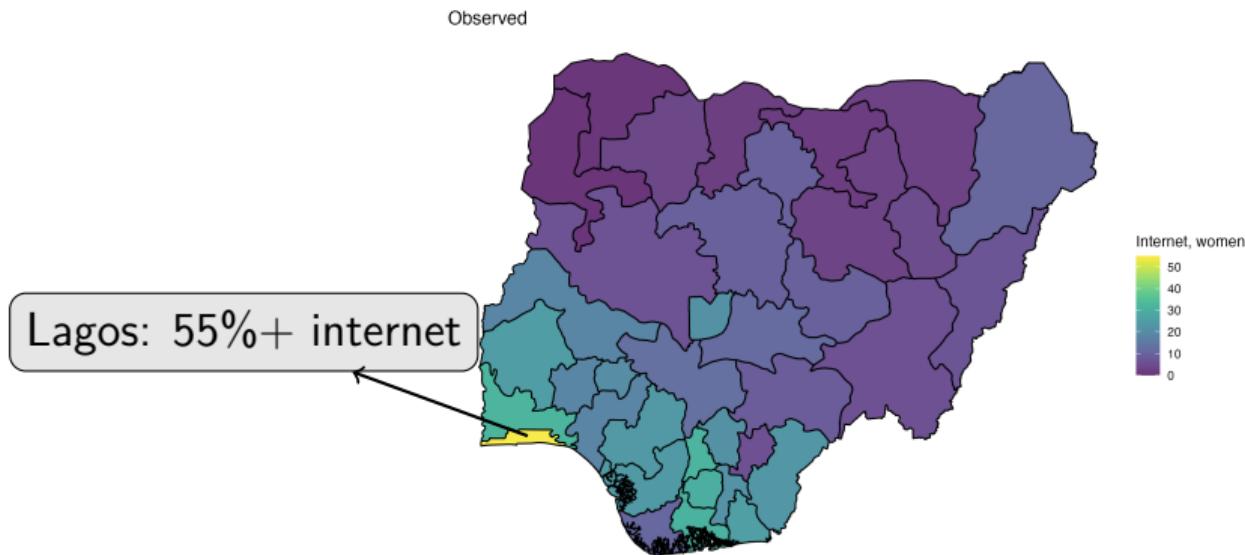


Source: Nigeria, Demographic and Health Survey

# Women using internet, past 12 months



# Women using internet, past 12 months

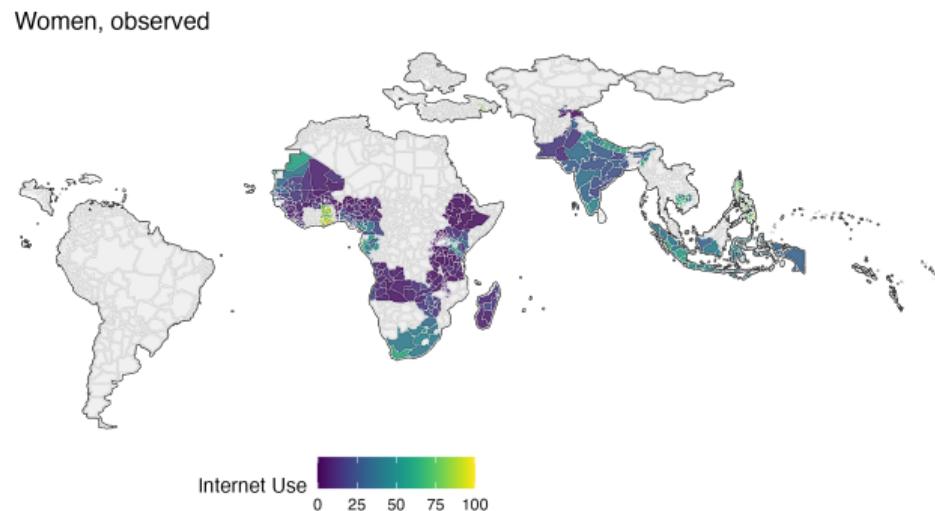


# Develop subnational estimates of adoption

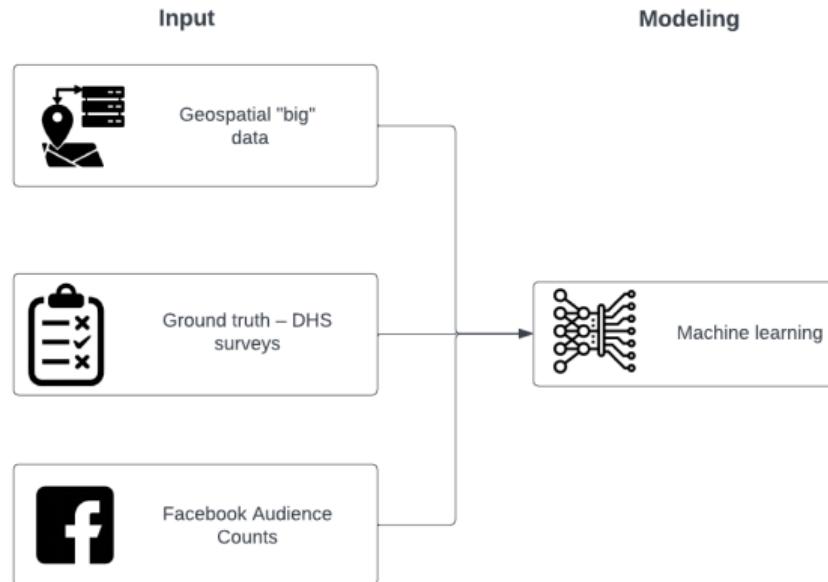
- ▶ **Goal:** Develop estimates of internet and mobile adoption by gender and digital gender gaps

# Develop subnational estimates of adoption

- ▶ **Goal:** Develop estimates of internet and mobile adoption by gender and digital gender gaps
- ▶ First subnational level
  - ▶ 118 countries, 2,150 subnational units



# Overview of approach



32 countries, with  
ground truth

Introduction  
0

Background  
oooo

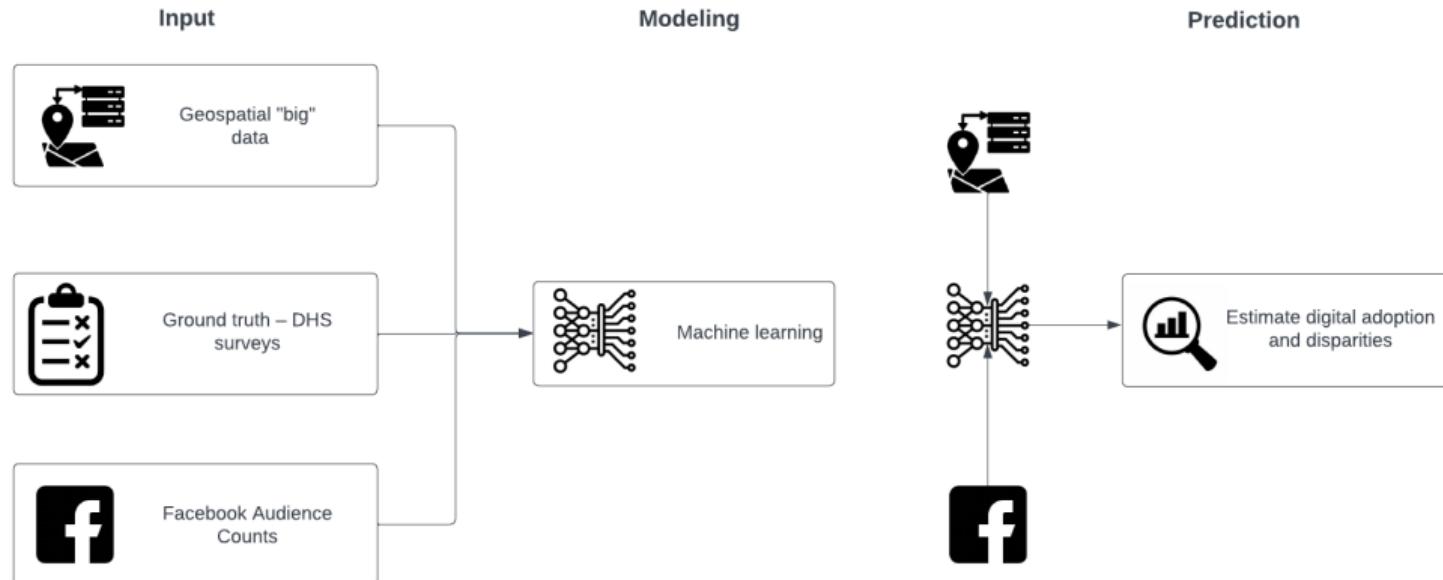
Data + methods  
●ooooooo

Results  
oooooooo

Conclusion  
ooo

References

# Overview of approach



32 countries, with  
ground truth

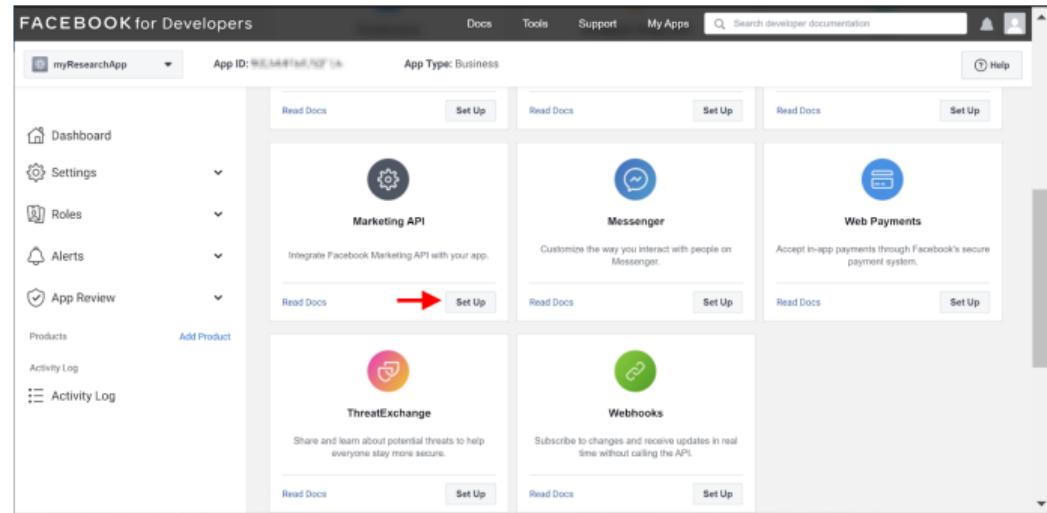
118 countries, with and  
without ground truth

# Ground truth – Demographic and Health Surveys (DHS)

- ▶ Household surveys representative at the first subnational level
  - ▶ Standardized sample design, questionnaire, implementation, etc.
  - ▶ Questions on individual-level internet use and mobile phone use (wave 7 onwards)
- ▶ Focus on 32 different DHS surveys, 2016-2022

# Facebook monthly active users counts

- ▶ Collected through public marketing API
- ▶ Specify geographic region (FB template or custom region)
- ▶ Disaggregated counts by gender, age, device type, etc.



# Big geospatial and population data

- ▶ Include 'offline' predictors that are uniformly available and consistent across subnational units
  - ▶ Satellite-derived nightlights data
  - ▶ Population density (World pop)
  - ▶ Relative wealth index (Meta)
  - ▶ Subnational education index, income index, human development index (HDI), gender development index (GDI)

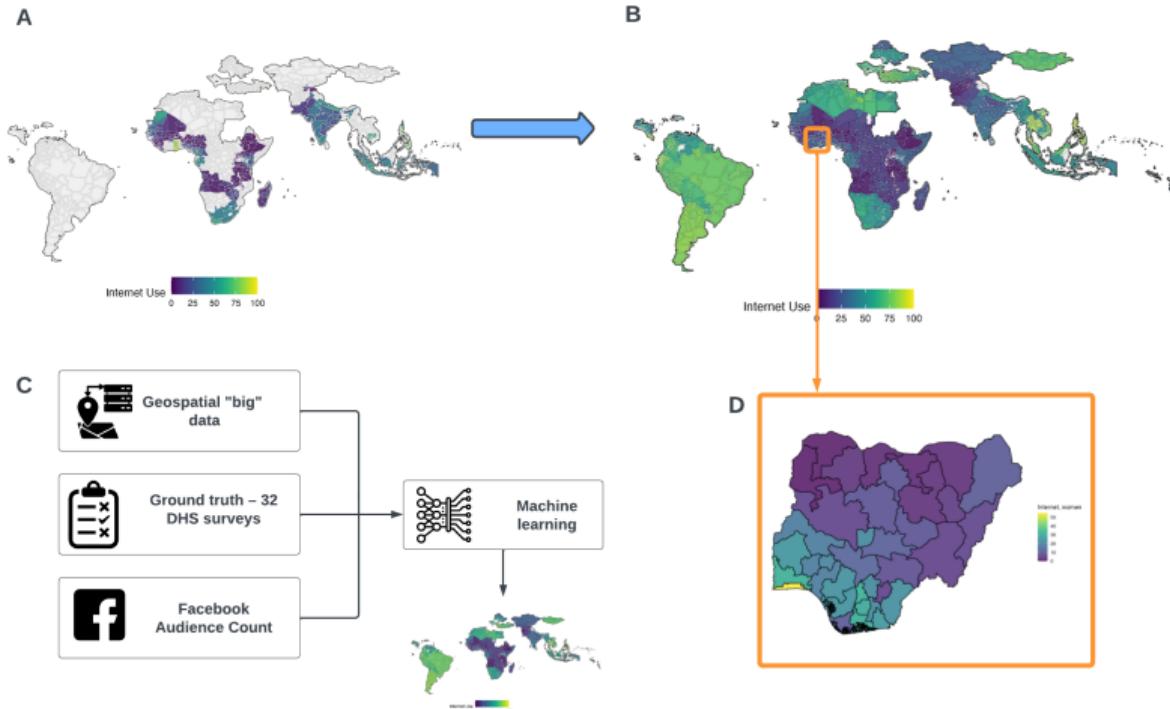
# Outcomes of interest (from DHS)

Indicators	Women	Men	Gender Gap
Mobile Phone Ownership	✓	✓	✓
Internet Use, Past 12 Mo	✓	✓	✓

# Modeling approach - ensemble machine learning

Algorithm	Description
glmnet (Lasso)	Lasso Regression
glmnet (Ridge)	Ridge Regression
glmnet (Elastic Net)	Elastic Net with 50% L1 Ratio
polyspline	Polynomial Spline
ranger	Random Forest with 100 Trees
gbm	Gradient Boosted Machine
glm	Generalized Linear Model
xgboost	Extreme Gradient Boosting
SuperLearner	Ensemble method combining multiple learning algorithms

# Greatly expanded coverage



Introduction  
0

Background  
0000

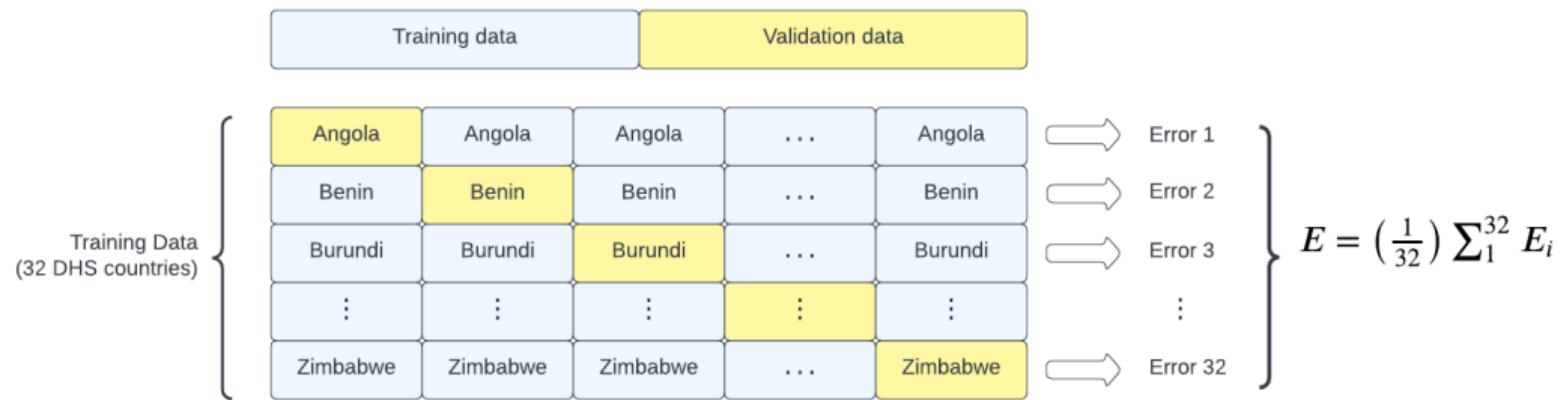
Data + methods  
0000000

Results  
●000000000

Conclusion  
000

References

# Leave-one-country-out cross validation

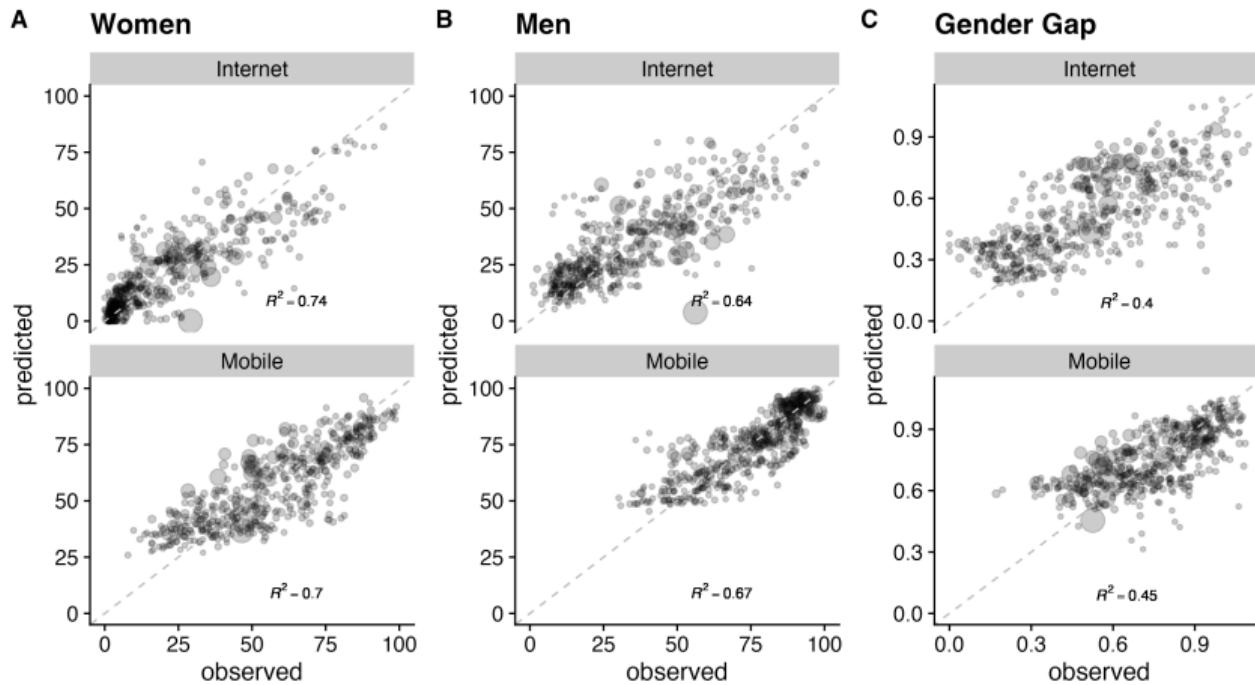


# Validation Metric: $R^2$

$$\begin{aligned} R^2 &= 1 - \frac{SS_{res}}{SS_{tot}} \\ &= 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \end{aligned} \tag{1}$$

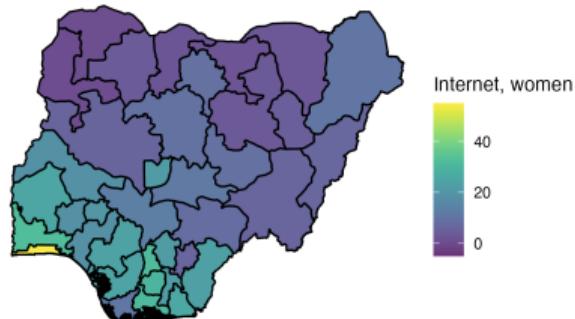
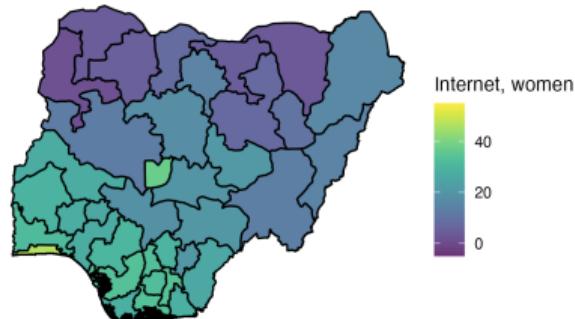
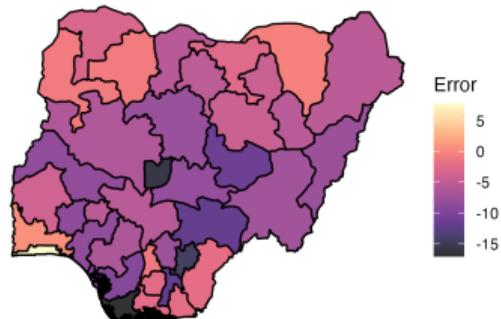
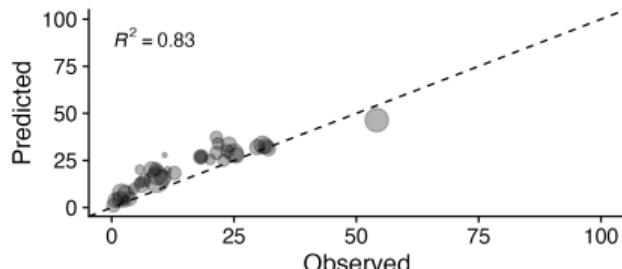
- ▶ 1 = **perfect predictions**
- ▶ 0 = **Mean**

# Overall predictive accuracy



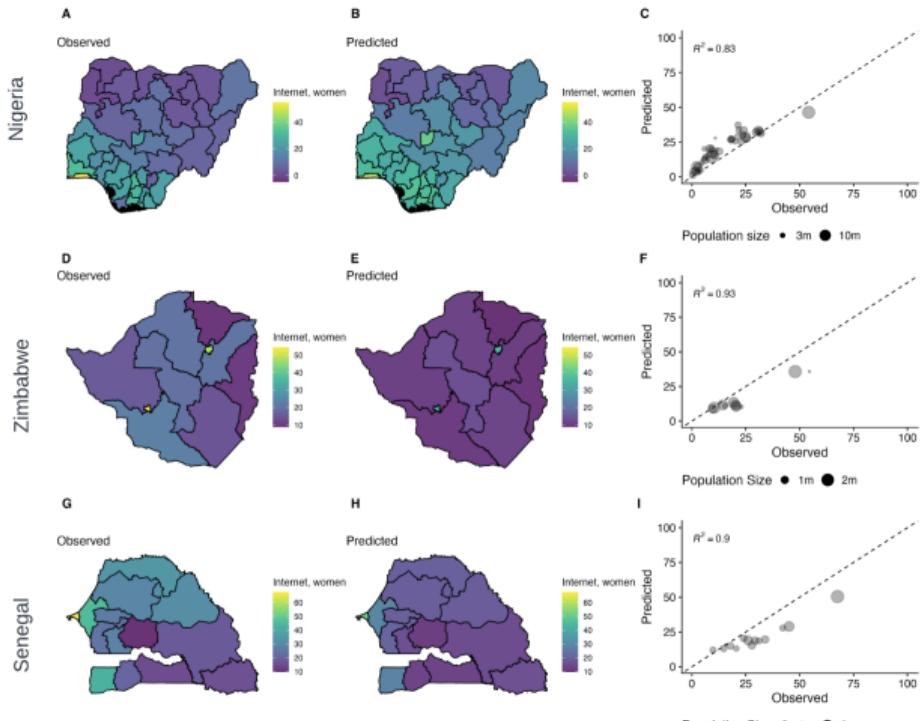
Population Size • 300k • 1m • 3m • 10m

# Results for Nigeria (Leave-one-country-out)

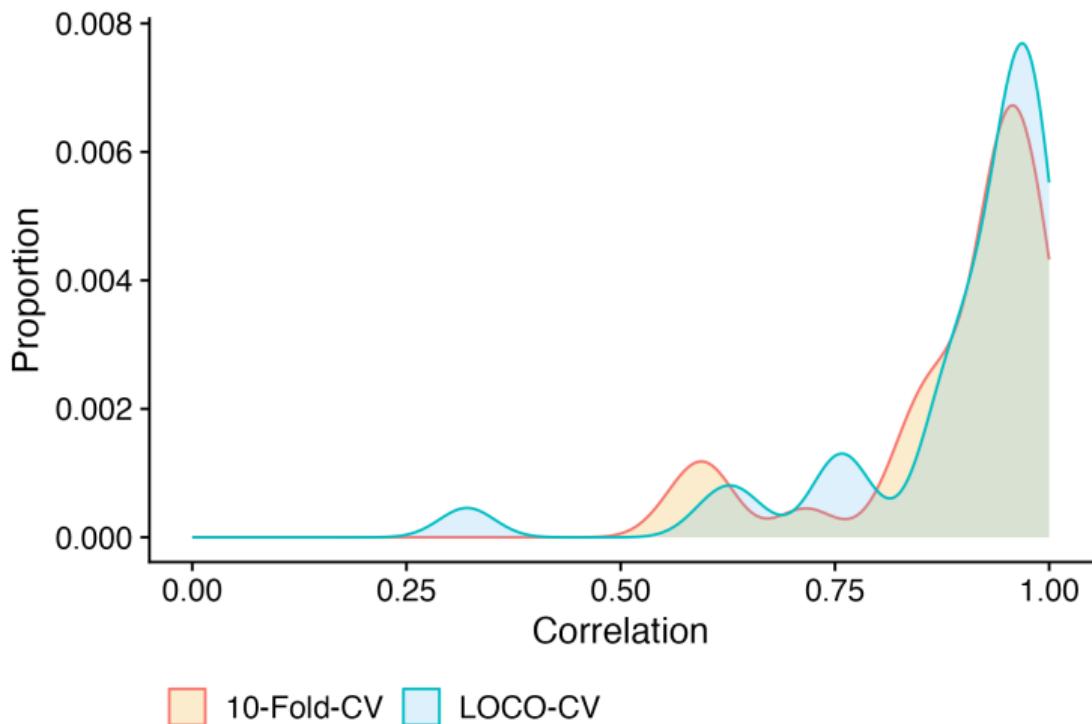
**A** Observed**B** Predicted**C** Error**D**

Population size • 3m ● 10m

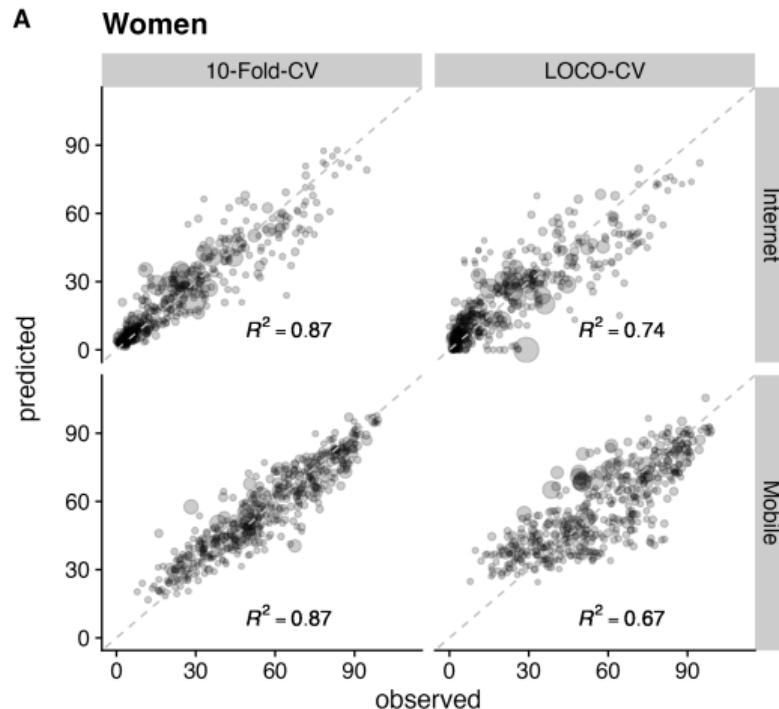
# Error by country



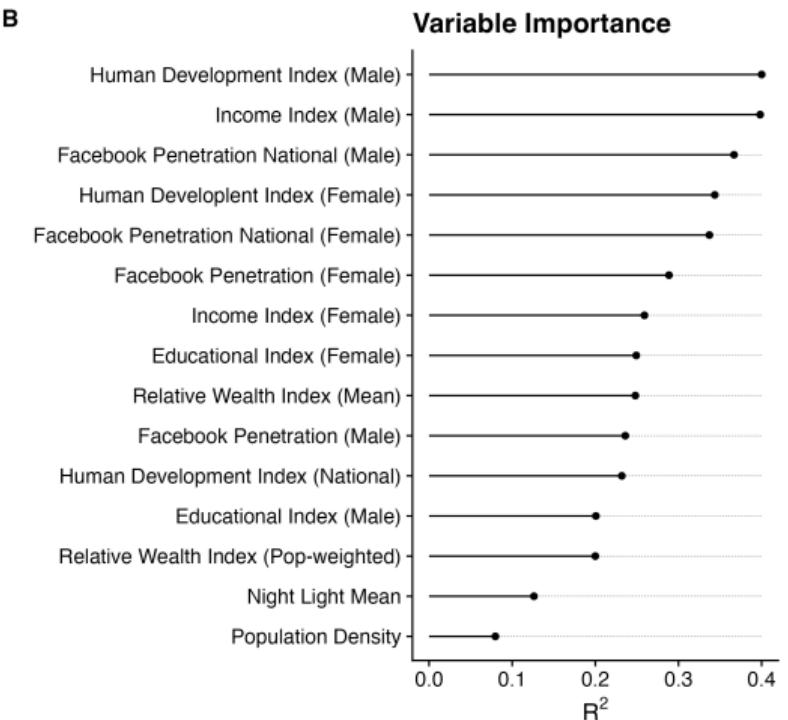
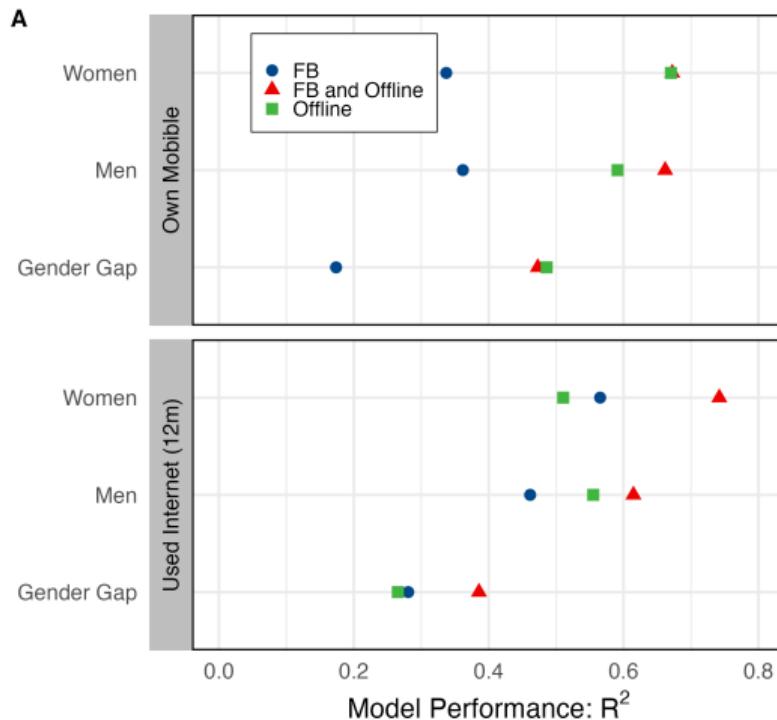
# Within-country performance



# Leave-one-country-out cross validation is more conservative



# Most important predictors



# Next steps and future opportunities

- ▶ Modeling of trends over time
- ▶ Modeling residuals / better uncertainty quantification

# Summary

- ▶ Using Facebook audience counts **greatly expands** our ability to accurately predict internet adoption in countries with no ground truth

# Summary

- ▶ Using Facebook audience counts **greatly expands** our ability to accurately predict internet adoption in countries with no ground truth
- ▶ Huge **disparities** in access to mobile and internet technologies between and within countries

# Summary

- ▶ Using Facebook audience counts **greatly expands** our ability to accurately predict internet adoption in countries with no ground truth
- ▶ Huge **disparities** in access to mobile and internet technologies between and within countries
- ▶ New opportunities to study **population-level impacts** of digital technology using these subnational estimates

# Thank You

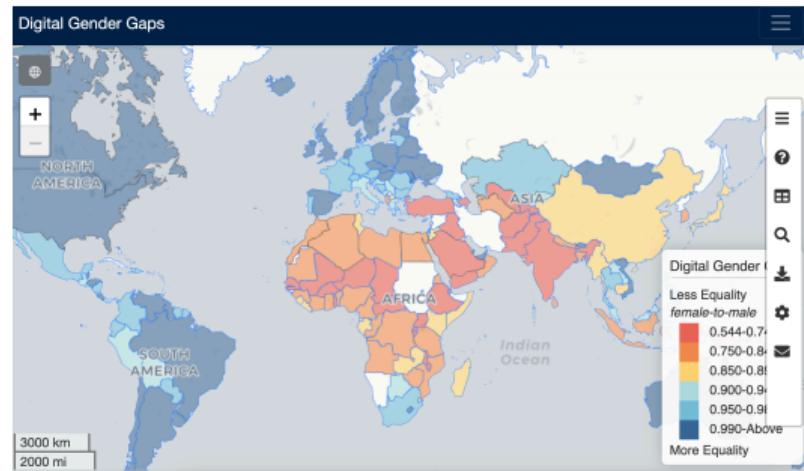
## Funders:

- ▶ Bill and Melinda Gates Foundation (INV-045370)
  - ▶ PI: Ridhi Kashyap
- ▶ Leverhulme Trust (Grant RC-2018-003) for the Leverhulme Centre for Demographic Science

## Contact:

 caseybreen

 casey.breen@demography.ox.ac.uk



[digitalgendergaps.org](http://digitalgendergaps.org)

# References

- DiMaggio, Paul and Eszter Hargittai. 2001. "From the 'Digital Divide' to 'Digital Inequality': Studying Internet Use as Penetration Increases." p. 25.
- Hjort, Jonas and Jonas Poulsen. 2019. "The Arrival of Fast Internet and Employment in Africa." *American Economic Review* 109(3):1032–1079.
- Kharisma, Bayu. 2022. "Surfing Alone? The Internet and Social Capital: Evidence from Indonesia." *Journal of Economic Structures* 11(1):8.
- Kho, Kevin, Leah K Lakdawala and Eduardo Nakasone. 2018. "Impact of Internet Access on Student Learning in Peruvian Schools."
- Rotondi, Valentina, Ridhi Kashyap, Luca Maria Pesando, Simone Spinelli and Francesco C. Billari. 2020. "Leveraging Mobile Phones to Attain Sustainable Development." *Proceedings of the National Academy of Sciences* 117(24):13413–13420.
- Suri, Tavneet and William Jack. 2016. "The Long-Run Poverty and Gender Impacts of Mobile Money." *Science* 354(6317):1288–1292.
- Unwin, P. T. H. 2009. *ICT4D: Information and Communication Technology for Development*. Cambridge University Press.