# POPULATION ESTIMATION FOR CENSUS SUPPORT

**Doug Leasure**

Leverhulme Centre for Demographic Science
University of Oxford

SICSS-Oxford
6 July 2022

# POPULATION AND HOUSING CENSUS

*Complete enumeration of national population every 10 years*

- **Why it's important:**
  - Allocate representation in government
  - Aportion national budget to administrative units
  - Understand changing demographics

- **Why it's difficult:**
  - Massive logistical challenge even in best circumstances
  - Inaccessibility in conflict areas and remote regions
  - Undercount/omission of vulnerable groups (*e.g. homeless, nomads, immigrants*)

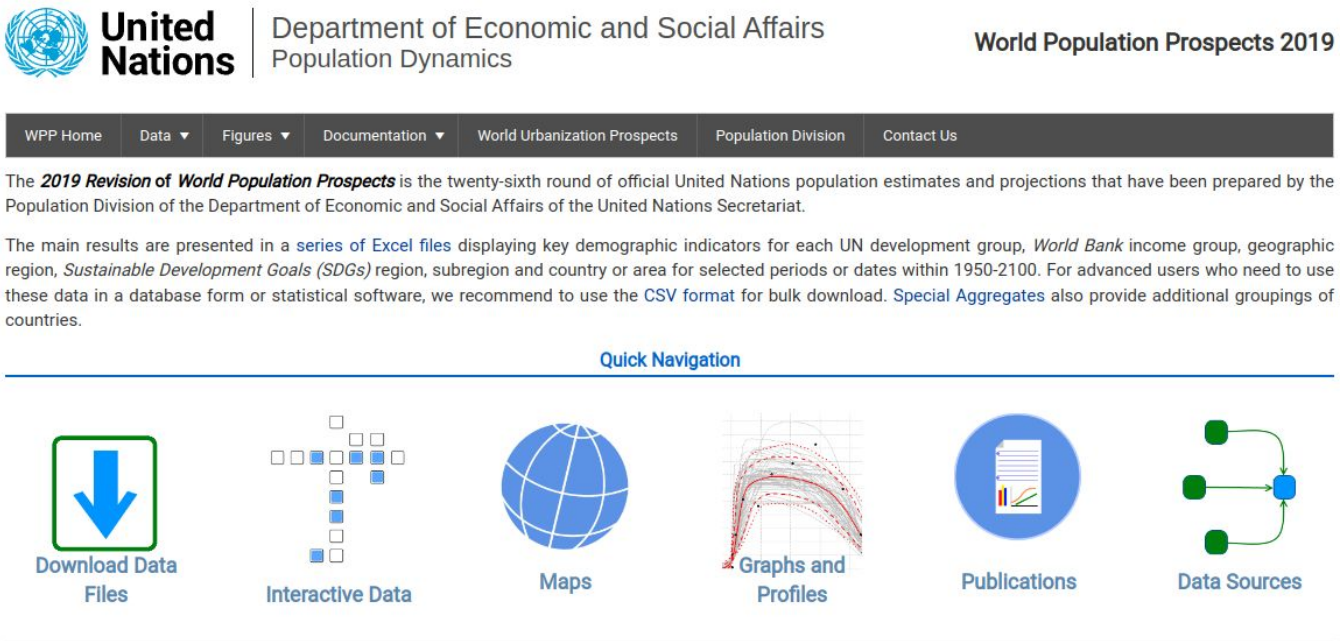
# POPULATION AND HOUSING CENSUS

*Last census…*

- **Lebanon:** 1943
- **Somalia:** 1975
- **Afghanistan:** 1979
- **Democratic Republic of Congo:** 1984
- **Uzbekistan:** 1989
- **Iraq:** 1997
- **Ukraine:** 2001 (severe undercount, particularly for children)
- **Nigeria:** 2006 (widely disputed)
- **Colombia:** 2018 (good quality, but some areas inaccessible)

# UN WORLD POPULATION PROSPECTS

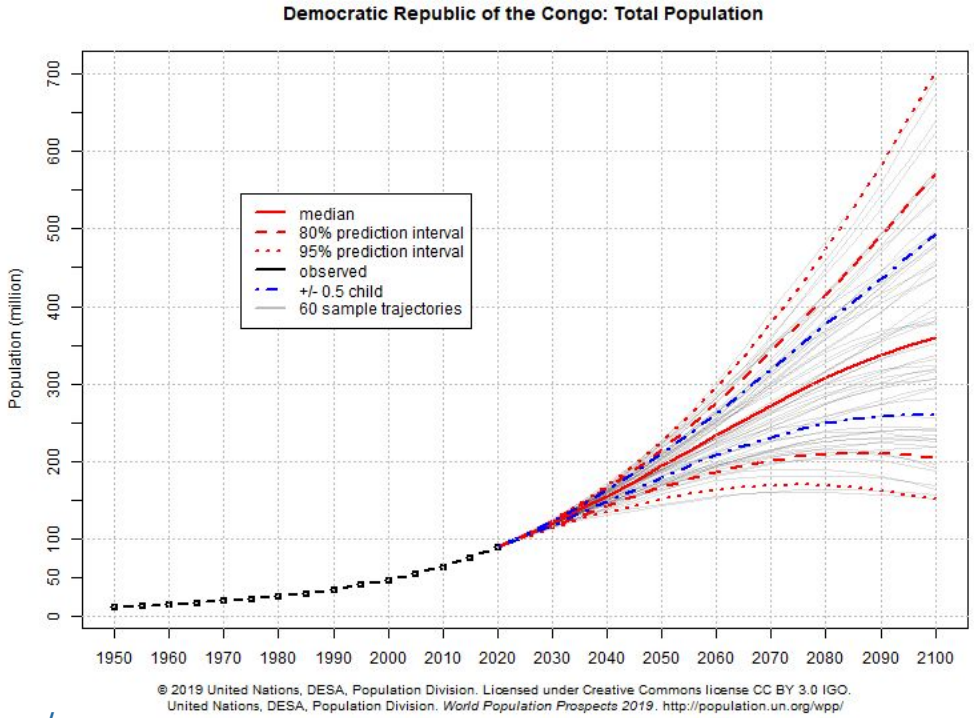
*Population projections for all countries based on estimates of fertility and mortality*




https://population.un.org/wpp/

# UN WORLD POPULATION PROSPECTS

*DRC: Last census in 1984*




https://population.un.org/wpp/

# FILLING DATA GAPS

**Goal:** Up-to-date sub-national population estimates by age and sex

**Question:** Can we use existing data in new ways?

- Up-to-date geolocated primary data collection
  - *Household surveys*
  - *Incomplete census*
  - *Administrative records*

- Geospatial data
  - *Building footprints*
  - *Neighbourhood classification mapping (a.k.a. settlement types)*
  - *Road maps*
  - *Nighttime lights*

# Lagos, Nigeria

# GOOGLE BUILDING FOOTPRINTS

## Open Buildings

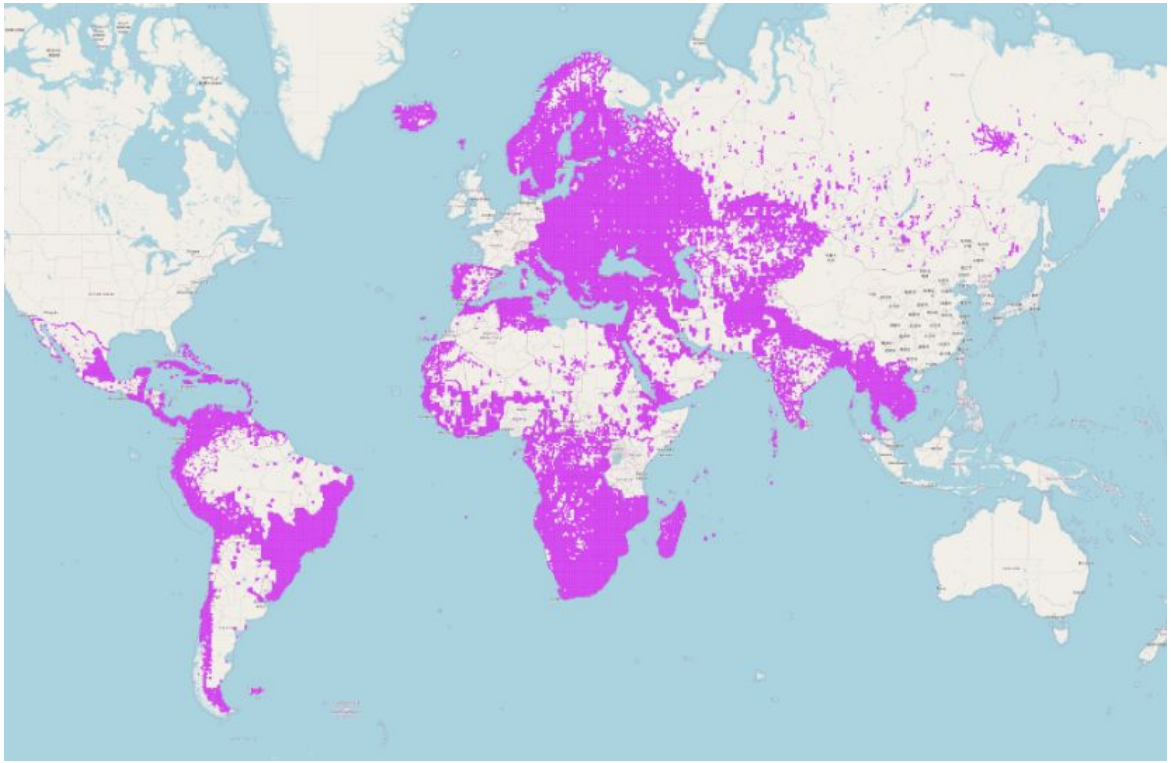
A dataset of building footprints to support social good applications.

Building footprints are useful for a range of important applications, from population estimation, urban planning and humanitarian response, to environmental and climate science. This large-scale open dataset contains the outlines of buildings derived from high-resolution satellite imagery in order to support these types of uses. The project being based in Ghana, the current focus is on the continent of Africa.

Explore | FAQ | Data format | Download

https://sites.research.google.com/open-buildings

# MICROSOFT BUILDING FOOTPRINTS



https://github.com/microsoft/GlobalMLBuildingFootprints

# WORLD SETTLEMENT FOOTPRINT 3D

## (GERMAN AEROSPACE CENTER - DLR)


Remote Sensing of Environment

Volume 270, 1 March 2022, 112877

ELSEVIER

# World Settlement Footprint 3D - A first three-dimensional survey of the global building stock

Thomas Esch ✉, Elisabeth Brzoska, Stefan Dech, Benjamin Leutner, Daniela Palacios-Lopez, Annekatrin Metz-Marconcini, Mattia Marconcini, Achim Roth, Julian Zeidler

https://download.geoservice.dlr.de/WSF2015/

# GLOBAL HUMAN SETTLEMENT LAYER


https://ghsl.jrc.ec.europa.eu/

# NIGHTTIME LIGHTS

EARTHDATA
OPEN ACCESS FOR OPEN SCIENCE


https://www.earthdata.nasa.gov/learn/backgrounders/nighttime-lights

# CASE STUDIES

peanutButter

Aggregate | Disaggregate | About

WorldPop

**Select Country**

Malawi

**Urban Settlements**

Mean people per housing unit

1 — 4.5 — 10

1 1.9 2.8 3.7 4.6 5.5 6.4 7.3 8.2 9.1 10

Mean housing units per residential building

1 — 1.1 — 3

1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3

Proportion building footprints that are residential

0 — 0.49 — 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

**Rural Settlements**

Mean people per housing unit

1 — 4.5 — 10

1 1.9 2.8 3.7 4.6 5.5 6.4 7.3 8.2 9.1 10

Mean housing units per residential building

1 — 3

1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3

Proportion building footprints that are residential

0 — 0.49 — 1

## Do-It-Yourself Gridded Population Estimates

The "aggregate" tool will apply your estimates of people per building to every building and then aggregate buildings to estimate population size for each ~100 m grid cell using a high resolution map of building footprints[1]. See the "About" tab for details.

ℹ Move the sliders on the left panel to adjust settings[2] and then click "Refresh Results" to calculate a summary of the population estimates that will appear in the table below.

ℹ When you are satisfied that the settings and the results are reasonable, click the "Gridded Population Estimates" button to download a 100 meter population grid (geotiff raster, WGS84) created using your settings.

Refresh Results

| | |
|---|---|
| Population Total | 19,326,688 |
| % Urban Population | 22% |
| Urban: Population | 4,262,833 |
| Rural: Population | 15,063,854 |
| Urban: People per building footprint | 2.4 |
| Rural: People per building footprint | 2.2 |
| Urban: Building footprints | 1,757,507 |
| Rural: Building footprints | 6,831,680 |

Gridded Population Estimates

Settings | Source Files

[1] Building footprints for Malawi were based on satellite imagery from 2020 (5%), 2019 (42%), 2018 (27%), 2017 (16%), 2016 (3%), 2015 or earlier (6%) (Ecopia.AI, Maxar 2020; Dooley et al. 2020, 2021).
[2] Default values: "Population Total" represents the year 2020 from UN World Population Prospects (2019); "Mean people per housing unit" is from United Nations (2019) or United Nations (2017).


https://apps.worldpop.org/peanutButter

peanutButton

Aggregate | Disaggregate | About

**Select Country**

Malawi

**Urban Settlements**

Mean people per housing unit

1 | 4.5 | 10

1 1.9 2.8 3.7 4.6 5.5 6.4 7.3 8.2 9.1 10

Mean housing units per residential building

1 | 1.1 | 3

1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3

Proportion building footprints that are residential

0 | 0.49 | 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

**Rural Settlements**

Mean people per housing unit

1 | 4.5 | 10

1 1.9 2.8 3.7 4.6 5.5 6.4 7.3 8.2 9.1 10

Mean housing units per residential building

1 | 3

1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3

Proportion building footprints that are residential

0 | 0.49 | 1

## Do-It-Yourself Gridded Population Estimates

The "aggregate" tool will apply your estimates of people per building to every building and then aggregate buildings to estimate population size for each ~100 m grid cell using a high resolution map of building footprints[1]. See the "About" tab for details.

ℹ Move the sliders on the left panel to adjust settings[2] and then click "Refresh Results" to calculate a summary of the population estimates that will appear in the table below.

ℹ When you are satisfied that the settings and the results are reasonable, click the "Gridded Population Estimates" button to download a 100 meter population grid (geotiff raster, WGS84) created using your settings.

Refresh Results

| | |
|---|---|
| Population Total | 19,326,688 |
| % Urban Population | 22% |
| Urban: Population | 4,262,833 |
| Rural: Population | 15,063,854 |
| Urban: People per building footprint | 2.4 |
| Rural: People per building footprint | 2.2 |
| Urban: Building footprints | 1,757,507 |
| Rural: Building footprints | 6,831,680 |

Gridded Population Estimates

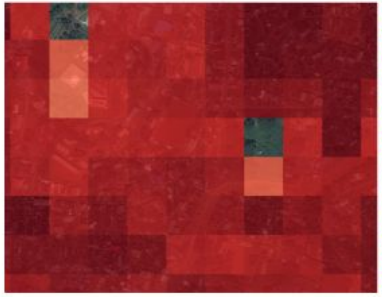
Settings | Source Files

[1] Building footprints for Malawi were based on satellite imagery from 2020 (5%), 2019 (42%), 2018 (27%), 2017 (16%), 2016 (3%), 2015 or earlier (6%) (Ecopia.AI, Maxar 2020; Dooley et al. 2020, 2021).
[2] Default values: "Population Total" represents the year 2020 from UN World Population Prospects (2019); "Mean people per housing unit" is from United Nations (2019) or United Nations (2017).

Satellite Image
(source: ESRI World Imagery)

Building Footprints
(Ecopia.AI & Maxar 2020)


Gridded Population Estimates
(peanutButter)



Gridded Population Estimates
+ Building Footprints


https://apps.worldpop.org/peanutButter

peanutButter

Aggregate | Disaggregate | About

**Select Country**

Malawi

**Urban Settlements**

Mean people per housing unit

1 | 4.5 | 10

1 1.9 2.8 3.7 4.6 5.5 6.4 7.3 8.2 9.1 10

Mean housing units per residential building

1 | 1.1 | 3

1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3

Proportion building footprints that are residential

0 | 0.49 | 1

0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

**Rural Settlements**

Mean people per housing unit

1 | 4.5 | 10

1 1.9 2.8 3.7 4.6 5.5 6.4 7.3 8.2 9.1 10

Mean housing units per residential building

1 | 3

1 1.2 1.4 1.6 1.8 2 2.2 2.4 2.6 2.8 3

Proportion building footprints that are residential

0 | 0.49 | 1

## Do-It-Yourself Gridded Population Estimates

The "aggregate" tool will apply your estimates of people per building to every building and then aggregate buildings to estimate population size for each ~100 m grid cell using a high resolution map of building footprints[1]. See the "About" tab for details.

ℹ Move the sliders on the left panel to adjust settings[2] and then click "Refresh Results" to calculate a summary of the population estimates that will appear in the table below.

ℹ When you are satisfied that the settings and the results are reasonable, click the "Gridded Population Estimates" button to download a 100 meter population grid (geotiff raster, WGS84) created using your settings.

Refresh Results

| | |
|---|---|
| Population Total | 19,326,688 |
| % Urban Population | 22% |
| Urban: Population | 4,262,833 |
| Rural: Population | 15,063,854 |
| Urban: People per building footprint | 2.4 |
| Rural: People per building footprint | 2.2 |
| Urban: Building footprints | 1,757,507 |
| Rural: Building footprints | 6,831,680 |

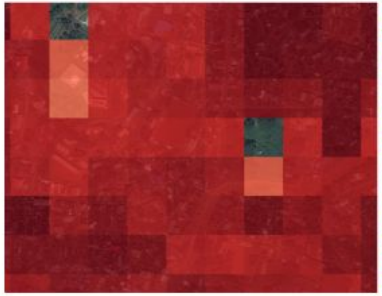
Gridded Population Estimates

Settings | Source Files

Satellite Image
(source: ESRI World Imagery)

Building Footprints
(Ecopia.AI & Maxar 2020)


Gridded Population Estimates
(peanutButter)



Gridded Population Estimates
+ Building Footprints

$$population = buildings \times \frac{households}{building} \times \frac{people}{household}$$


https://apps.worldpop.org/peanutButter

# WHY NOT PEANUT BUTTER?

- It is not fit to ANY observed population data!
- It does not account for spatial variation in people per building
- It does not measure uncertainty (i.e. there are no confidence intervals)

**While not recommended**, peanut butter works better than expected:

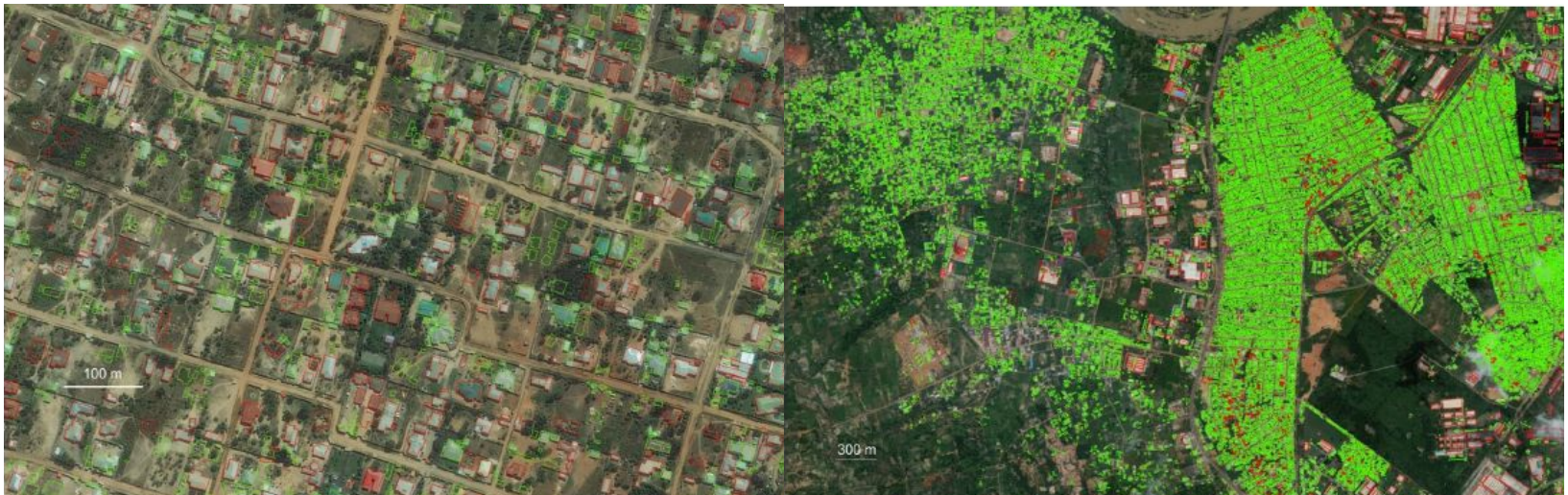
- Calibrated to produce appropriate national population total
- Flexibility for end-user to adjust settings as needed
- Up-to-date and accurate building footprints

# SPATIAL VARIATION IN PEOPLE PER BUILDING

Residential / Non-residential building classification

- *Superlearner* R package
- OpenStreetMaps crowd-sourced training data
- Geospatial covariates (e.g. coverage of impervious surfaces)

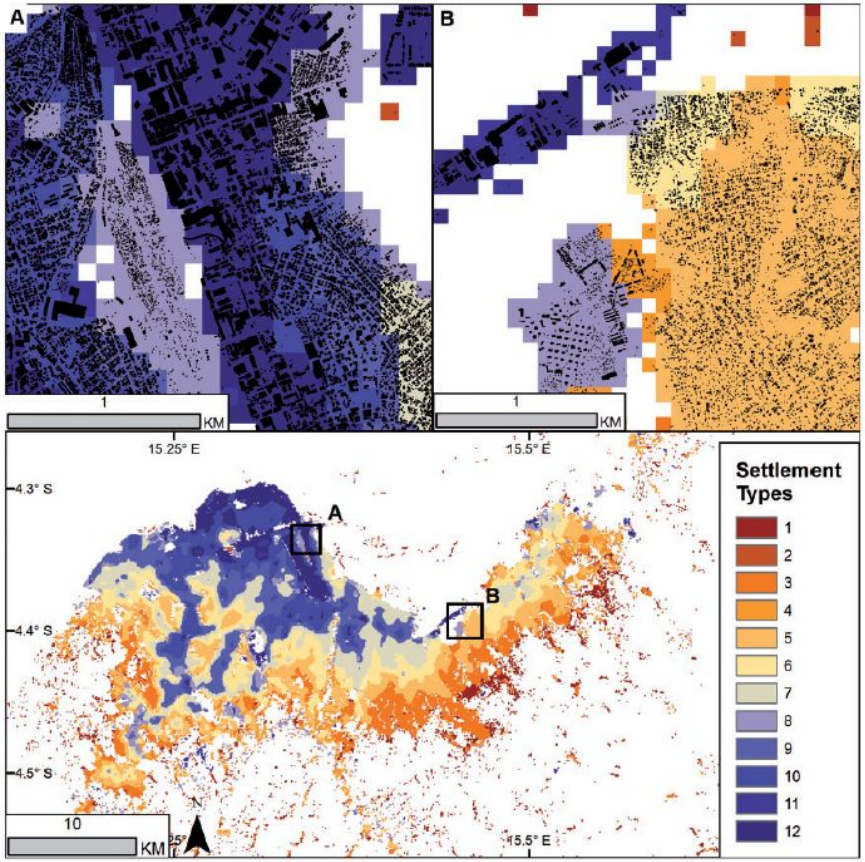
Lloyd et al. 2020. *Remote Sensing* 12



# SPATIAL VARIATION IN PEOPLE PER BUILDING

## Neighbourhood types

- Building footprints
- Metrics of building patterns
  - Size
  - Density
  - Spacing
  - Orientation
  - Regularity
- Mixture model cluster analysis




Jochem et al. 2020. *EPB: Urban Analytics & City Science*

# GHANA AND GUINEA

# CAN WE BUILD ON THE PEANUT BUTTER IDEA?

IPUMS INTERNATIONAL

HARMONIZED INTERNATIONAL CENSUS DATA FOR SOCIAL SCIENCE AND HEALTH RESEARCH

+

# GHANA AND GUINEA

**OBJECTIVE**: 100 m gridded population estimates with age-sex disaggregation and estimation uncertainty

$$population = buildings \times \frac{households}{building} \times \frac{people}{household}$$

**Population Projections**
- National (WPP)
- Sub-national (CIESIN)

* *imprecise estimates*

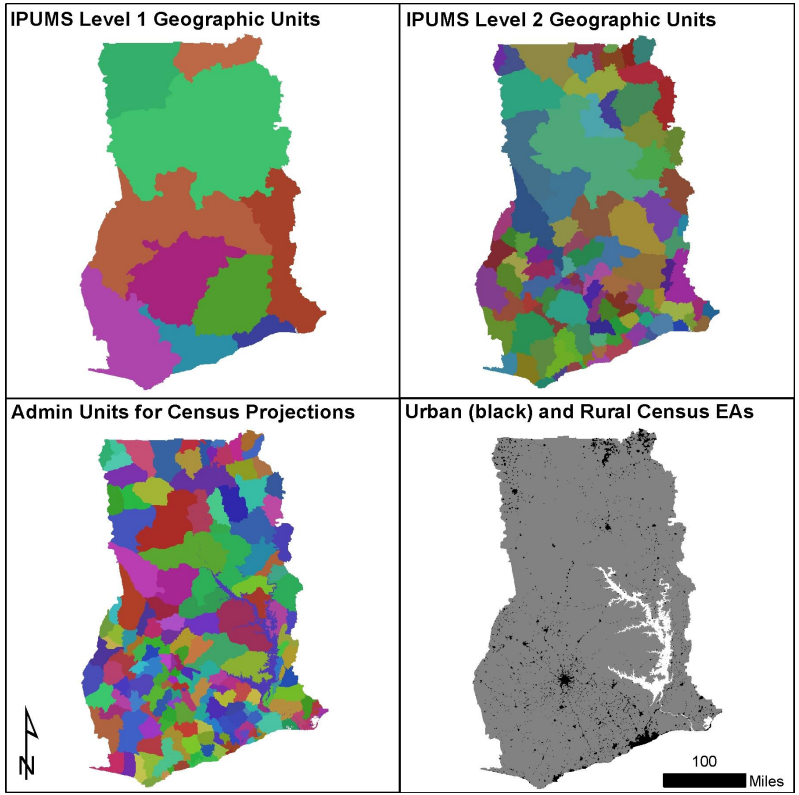
MAXAR

IPUMS INTERNATIONAL

- Representative samples
- Age and sex disaggregated
- 2010 Ghana census
- 2014 Guinea census

# GHANA AND GUINEA

Input data used different spatial units and geolocation was very coarse



# GHANA AND GUINEA

Implemented as a Bayesian statistical model (using Stan software):

$$pop_i \sim \text{Poisson}(bldg_i \times hpb_{t_i,u_i} \times pph_{t_i,g_i})$$
$$pop_{i,k} \sim \text{Multinomial}(pop_i, \theta_{t_i,r_i,k})$$

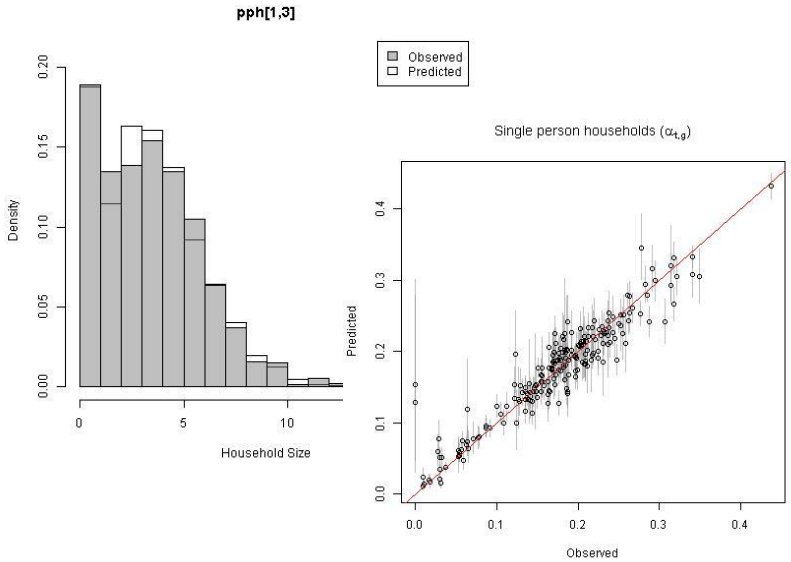
**Note:** Full model not shown

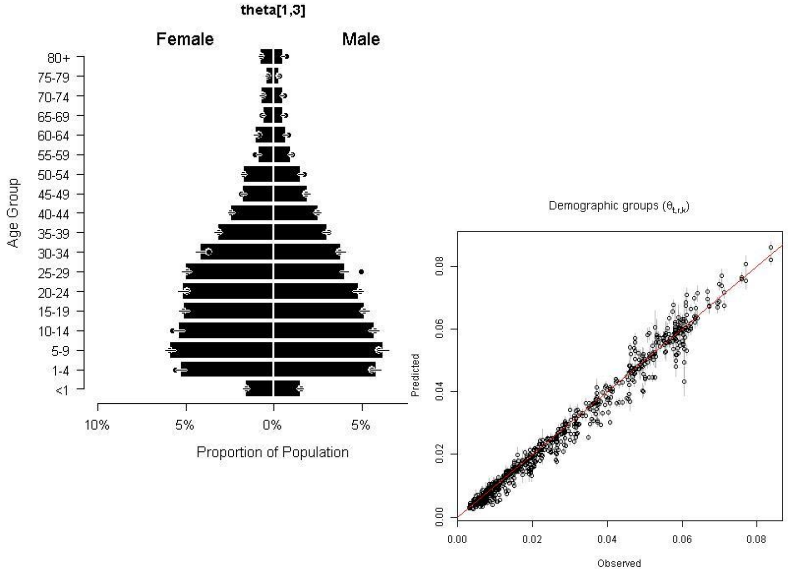

https://data.worldpop.org/repo/docs/leasure2020bayesian/leasure2020bayesian.pdf

https://mc-stan.org/

# GHANA AND GUINEA

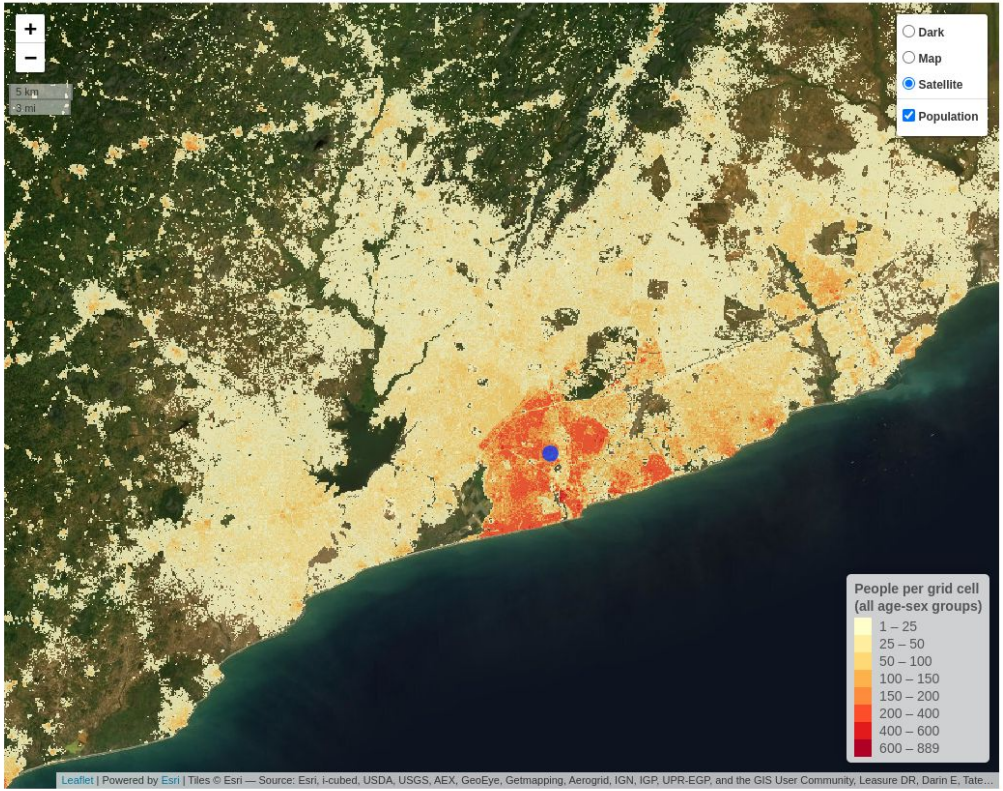
Probabilistic estimates of **people per housing unit**, **housing units per building**, and **age-sex proportions**:

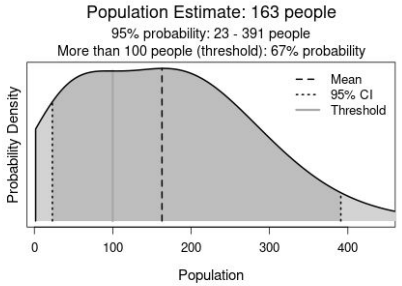


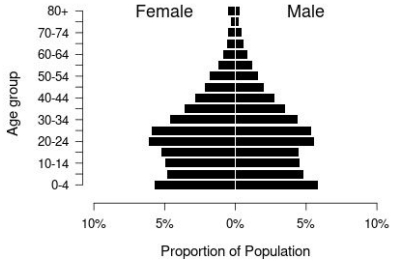




https://data.worldpop.org/repo/docs/leasure2020bayesian/leasure2020bayesian.pdf

# GHANA AND GUINEA




https://apps.worldpop.org/woprVision *(select data "GHA v2.0" or "GIN v1.0")*

# NIGERIA

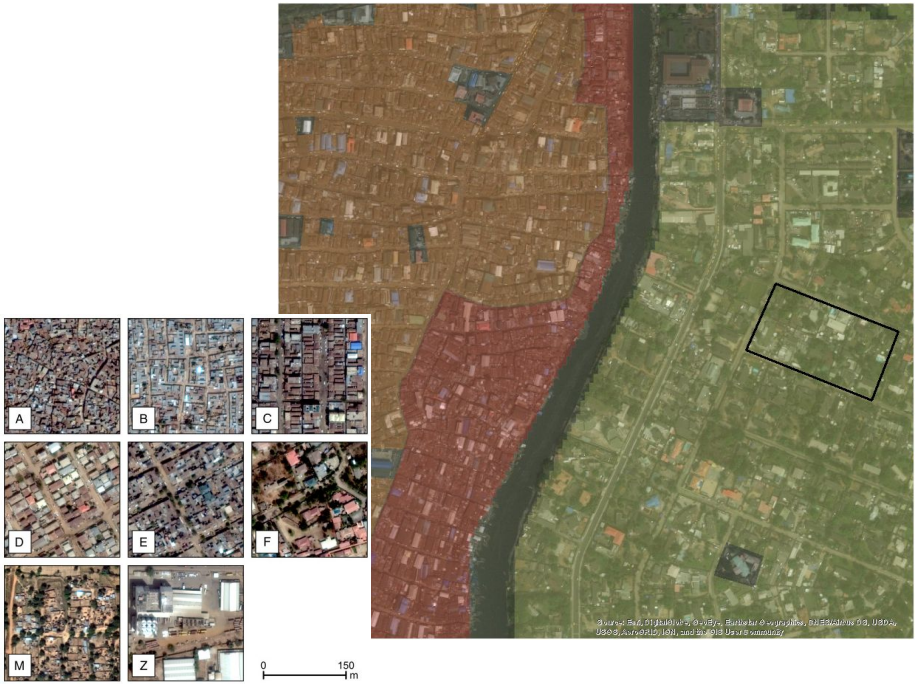
# NIGERIA

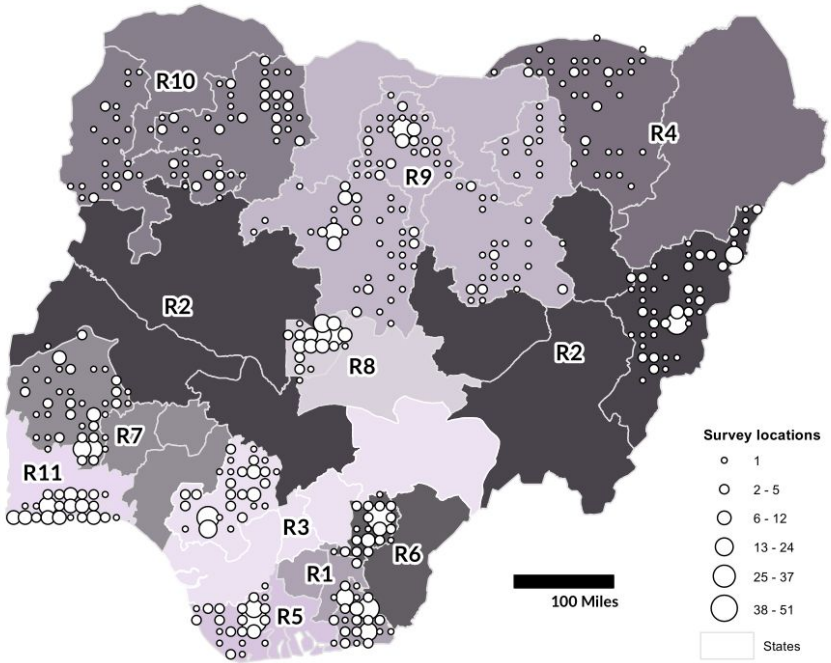
Estimating population sizes to support Polio eradication

- Last census was 2006 and it was widely criticized
- Bill and Melinda Gates Foundation were focused on Polio eradication, particularly in northern Nigeria
- Some areas are inaccessible due to security concerns from Boko Haram
- Need accurate high-resolution estimates of children under 5 for targeted Polio vaccination campaigns

# NIGERIA

## Oak Ridge National Laboratory *settlement map* and *microcensus survey*





Weber et al. 2018. *Remote Sensing of Environment* 204: 786-798

Leasure et al. 2020. *Proceedings of the National Academy of Sciences* 117(39): 2413-4179

# NIGERIA

Bayesian statistical model with geospatial covariates

$$N_i \sim Poison(D_i A_i)$$

$$D_i \sim LogNormal(\overline{D}_i, \sigma_{t,r,s,l})$$

$$\overline{D}_i = \alpha_{t,r,s,l} + \sum_{k=1}^{K} \beta_k x_{k,i}$$

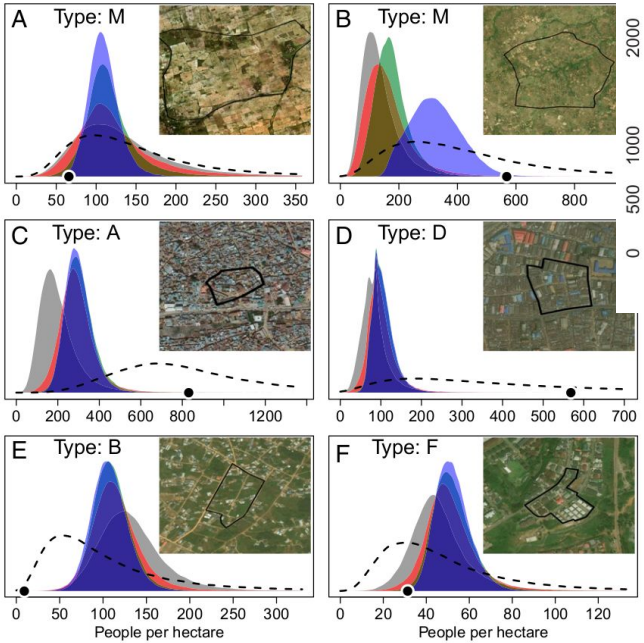
Geospatial covariates (x):

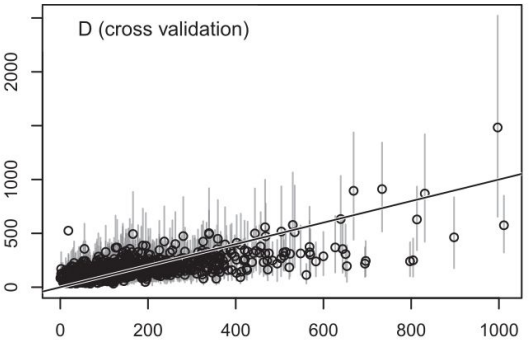
- **Region**, **state**, and **local govt.**
- **Settlement type**
- **School density** within 1km
- **Average household size** spatially interpolated from DHS
- **Census projections** of total pop.
- **Residential area** within 1 km
- **Non-residential area** within 1 km

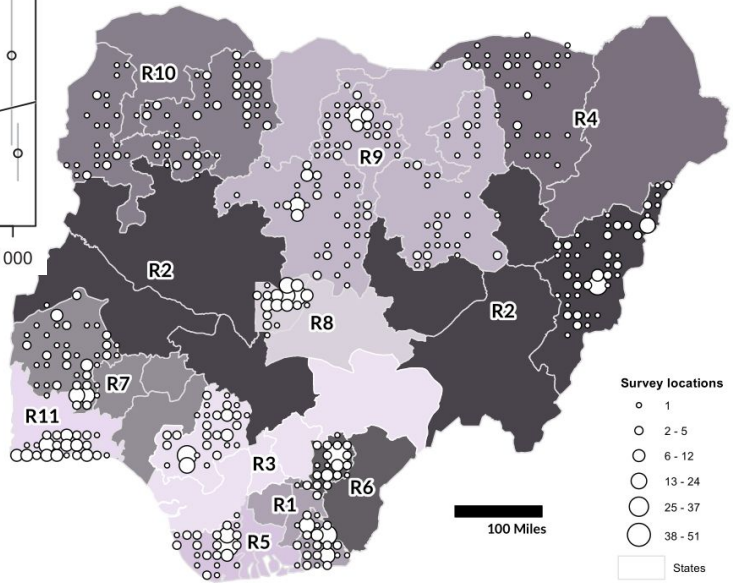
Leasure et al. 2020. *Proceedings of the National Academy of Sciences* 117(39): 2413-4179

# NIGERIA

## Predicting population density and accounting for uncertainty



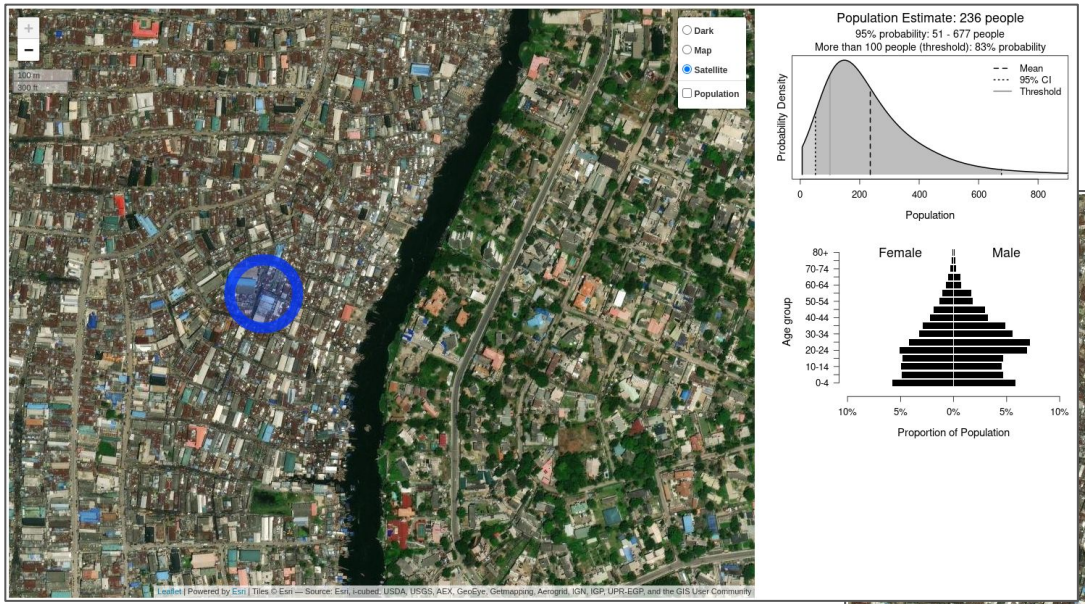



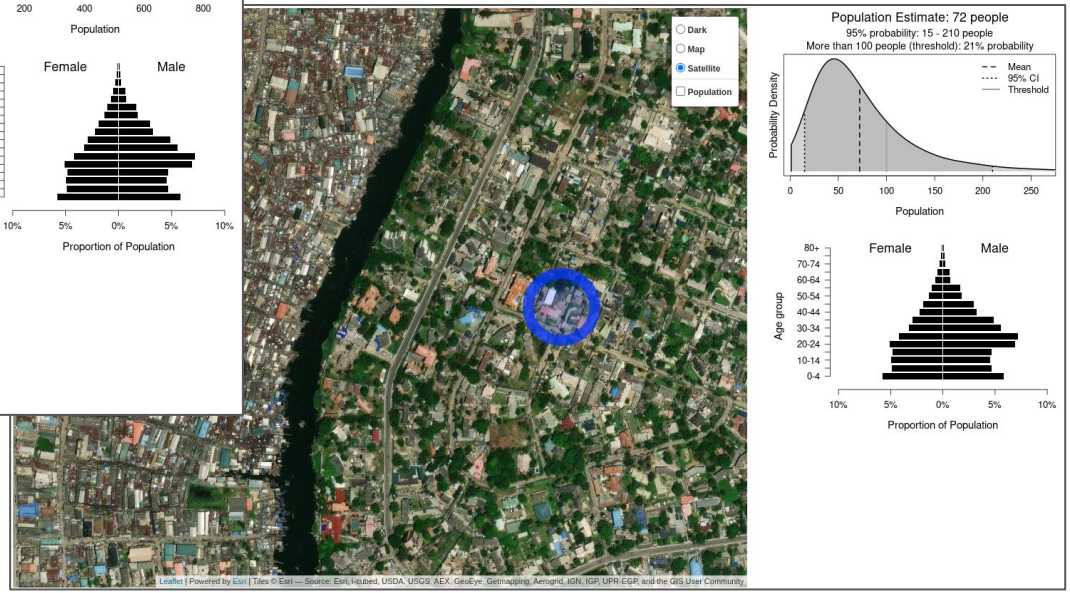




Leasure et al. 2020. *Proceedings of the National Academy of Sciences* 117(39): 2413-4179

# NIGERIA

## Mapping Bayesian population predictions




https://apps.worldpop.org/woprVision/ *(select data "NGA v1.2" or "NGA v2.0")*

# COLOMBIA

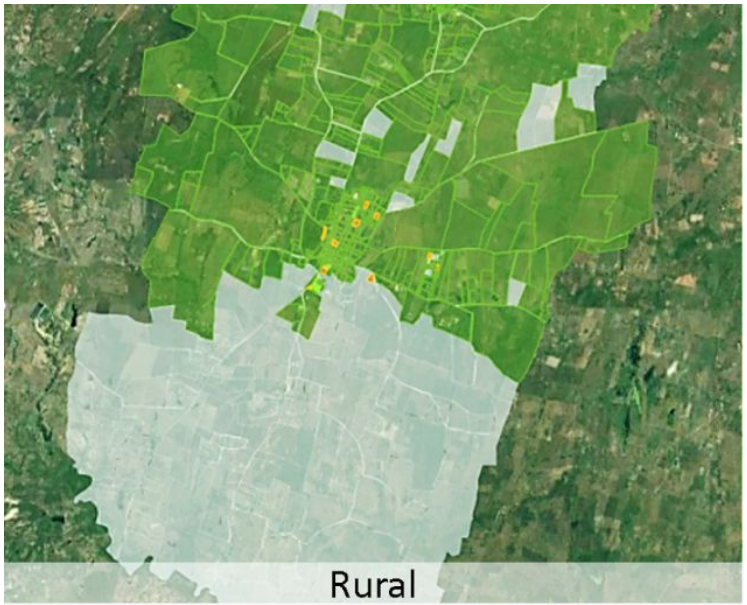
# COLOMBIA

Local knowledge and satellites to fill census gaps

- Last census was 2018
- Some areas were inaccessible due to vast jungle terrain and insecurity
  - Tends to result in undercounts of ethnic minority groups and indigenous populations
- Social cartography workshops asked citizens to map their communities to help census preparations
- German Aerospace Center provided satellite-derived estimates of building area for every 90 m grid square in the country

# COLOMBIA

## Addressing 2018 census omissions in remote regions





**Green**
>90% enumerated

**Orange**
10 - 90% enumerated

**Grey**
<10% enumerated

# COLOMBIA

## Geospatial predictors with full coverage

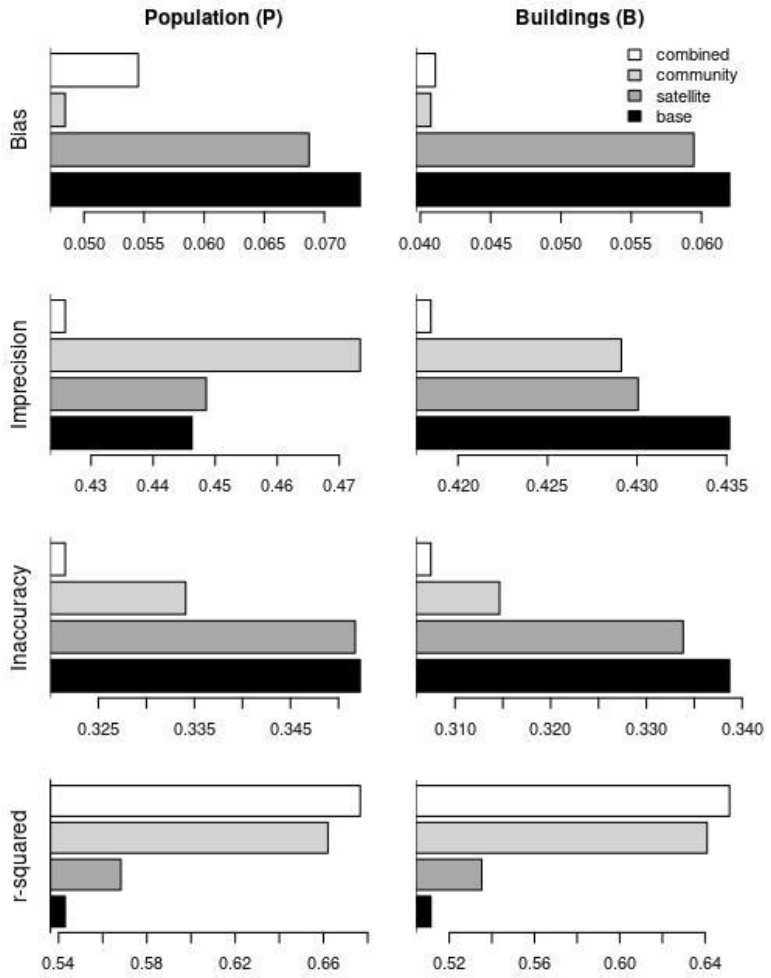
- Social Cartography Workshops:
  - Estimate dwellings
  - Estimate families
  - Estimate population
- Satellite-derived:
  - Building area per 90 m grid square
- Other Geospatial covariates:
  - Distance to city center
  - Intensity of nighttime lights
  - Elevation
  - Poverty index
  - Schools


# COLOMBIA

Bayesian statistical models (not shown)

Four competing models:

1. **base**: standard geospatial predictors (included in all models)
2. **satellite**: building area per 90 m grid square
3. **community**: social cartography estimates
4. **combined**: satellite + community



# COLOMBIA

- Combination of local knowledge and remotely sensed buildings was more effective than either alone
- Remotely-sensed *counts* of buildings may be better
- Social cartography workshops
  - Engaged underrepresented populations in the census
  - Improved population estimates for inaccessible areas

# CONCLUSIONS

# CONCLUSIONS

Geospatial data for population estimation

- Building footprints are a valuable source of information
  - Becoming more openly available globally
- Existing household surveys or other population data can be repurposed for population estimation to fill critical data gaps
- Bespoke methods (statistical or machine learning) often needed to combine data sources for population estimation
- Don’t forget:  “geospatial data” can include low-tech data like social cartography.

# THANK YOU

Questions?  Suggestions?  Want to get involved?

Contact me:  **douglas.leasure@sociology.ox.ac.uk**