Nigeria Artificial Intelligence Summit Bootcamp

Lagos, 22 Novembre 2019

Wole Ademola, Édith Darin, Sophie Delaporte, Alina Game, Michael Harper

Overview of today's workshop

- Topic: Production and Use of Gridded Population data

- Timetable:
  - First session: 1 hour30 on **Gridded Population Modelling**
  - Second session: 1hour30 on **the toolkit to use Gridded population**
  - Third session: 1 hour30 on a **real world application**
  - Last minutes: explaining our data challenge for you!

Overview of today's workshop

- To follow: all material on Github:

  https://github.com/GRID3/DSNTraining

- If you have any questions throughout the workshop (or after)- tweet us at:

  @GRID3Global

GRID3 NIGERIA

Geospatial Analytics for Nigeria

Session 2—Bayesian Population Modelling

Édith Darin

Bill & Melinda Gates foundation · UKaid from the British people · worldpop · FLOWMINDER.ORG · UNFPA · Center for International Earth Science Information Network, Earth Institute | Columbia University

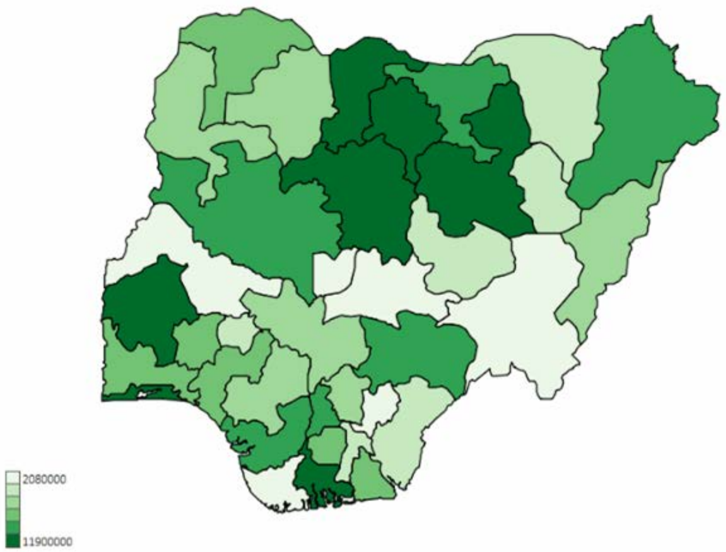
About Me

- Edith Darin
- Bayesian Statistical Modeller
- Focused on developing new methodologies and disseminating knowledge

Structure

- Assessing the issue
- Gridded population estimates
- Data requirements
- Statistical modelling
- Model fit
- Model prediction
- Improvement

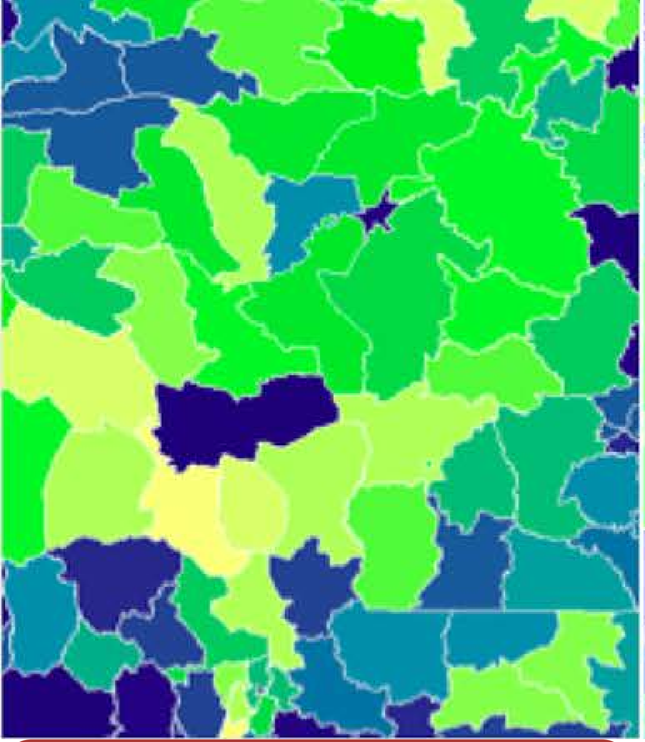
Assessing the issue

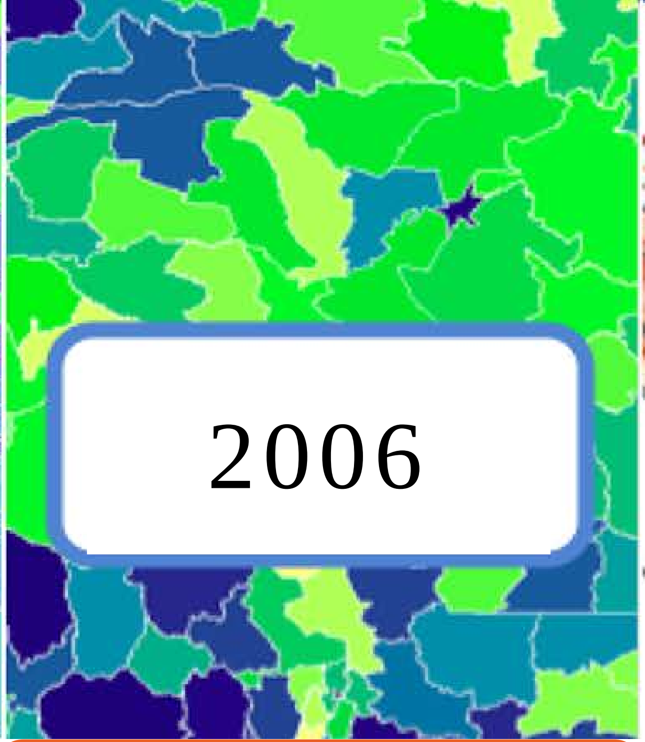


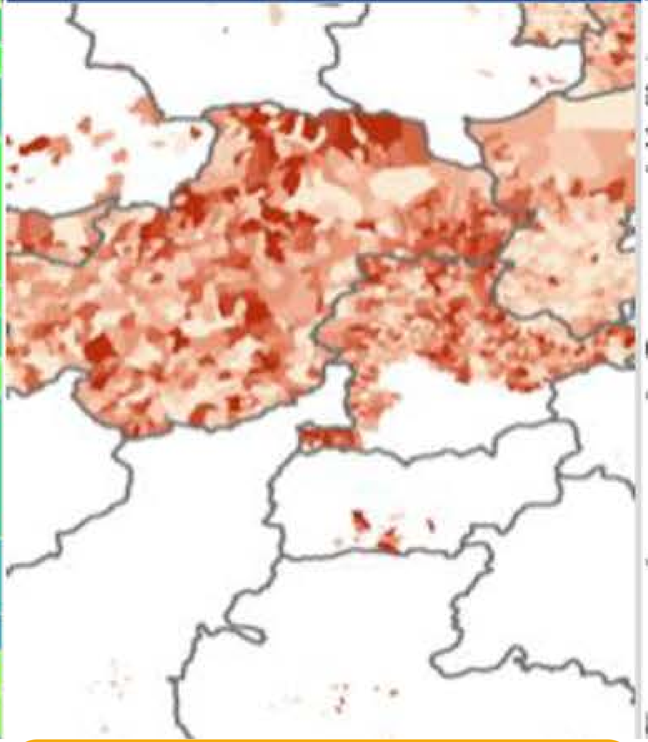
Data from the National Bureau of Statistics

POPULATION FORECASTS 2006 - 2016											
STATE	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
ABIA	2,845,380	2,923,252	3,003,255	3,085,447	3,169,889	3,256,642	3,345,769	3,437,336	3,531,408	3,628,055	3,727,347
ADAMAWA	3,178,950	3,272,489	3,368,781	3,467,906	3,569,948	3,674,992	3,783,127	3,894,444	4,009,037	4,127,001	4,248,436
AKWA/IBOM	3,902,051	4,037,002	4,176,620	4,321,067	4,470,509	4,625,120	4,785,078	4,950,568	5,121,781	5,298,916	5,482,177
ANAMBRA	4,177,828	4,296,460	4,418,461	4,543,926	4,672,954	4,805,646	4,942,106	5,082,440	5,226,760	5,375,177	5,527,809
BAUCHI	4,653,066	4,813,990	4,980,480	5,152,728	5,330,933	5,515,302	5,706,046	5,903,388	6,107,554	6,318,781	6,537,314
BAYELSA	1,704,515	1,754,670	1,806,300	1,859,450	1,914,163	1,970,487	2,028,468	2,088,154	2,149,597	2,212,849	2,277,961
BENUE	4,253,641	4,383,184	4,516,671	4,654,225	4,795,967	4,942,026	5,092,533	5,247,624	5,407,438	5,572,118	5,741,815
BORNO	4,171,104	4,315,360	4,464,605	4,619,012	4,778,758	4,944,030	5,115,017	5,291,918	5,474,937	5,664,285	5,860,183

Challenges with traditional population data







Coarse Resolution

Outdated

Incomplete

Inaccurate, missing populations

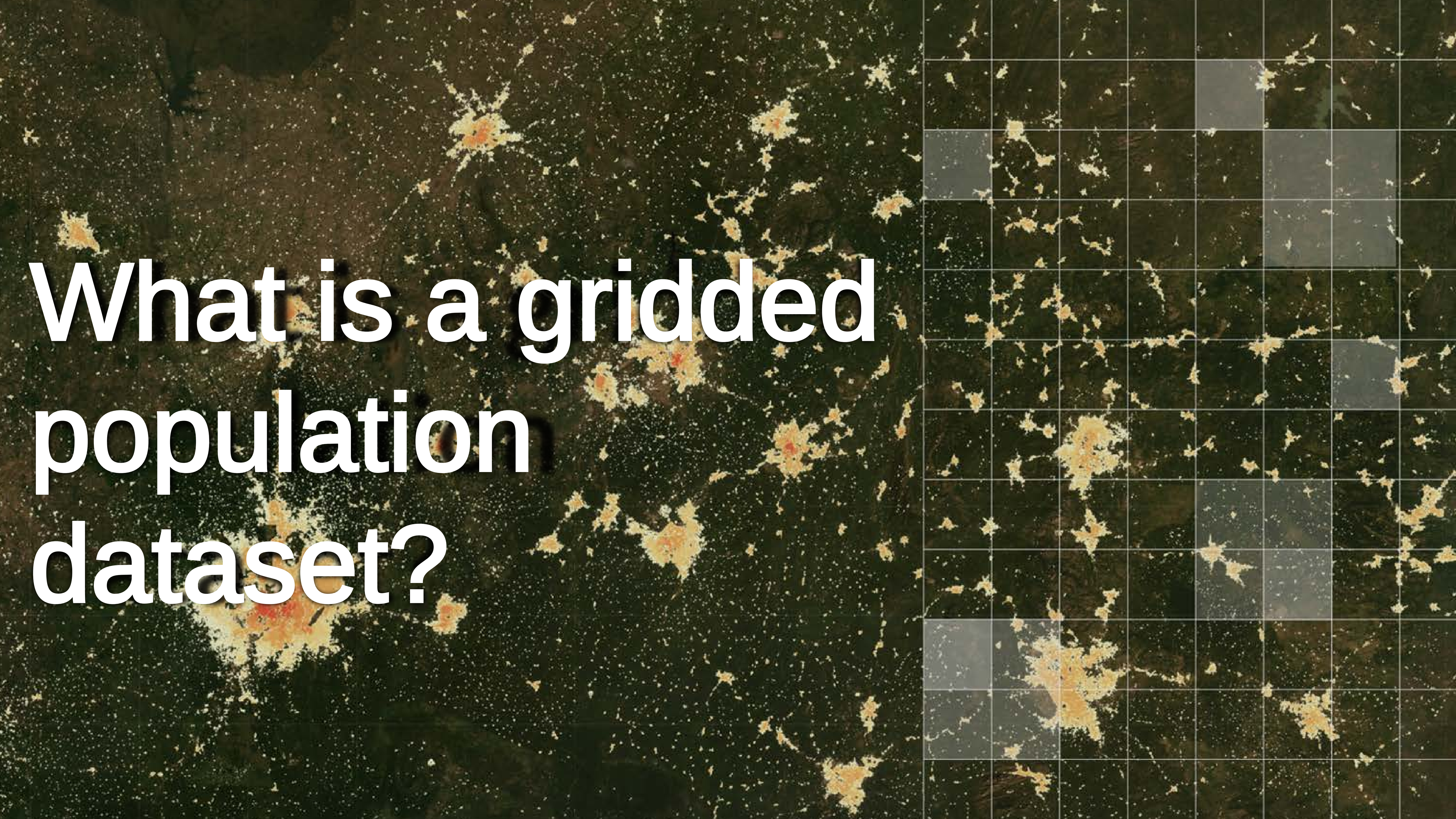
Challenges with traditional population data

Polio Vaccination Campaign

- Bill & Melinda Gates Foundation
- **Aim**: eradication of polio worldwide
- Three remaining countries: Afghanistan, Pakistan, **Nigeria**
- Target: children under five
- Urgent need of accurate and precise population figures

Gridded population estimates

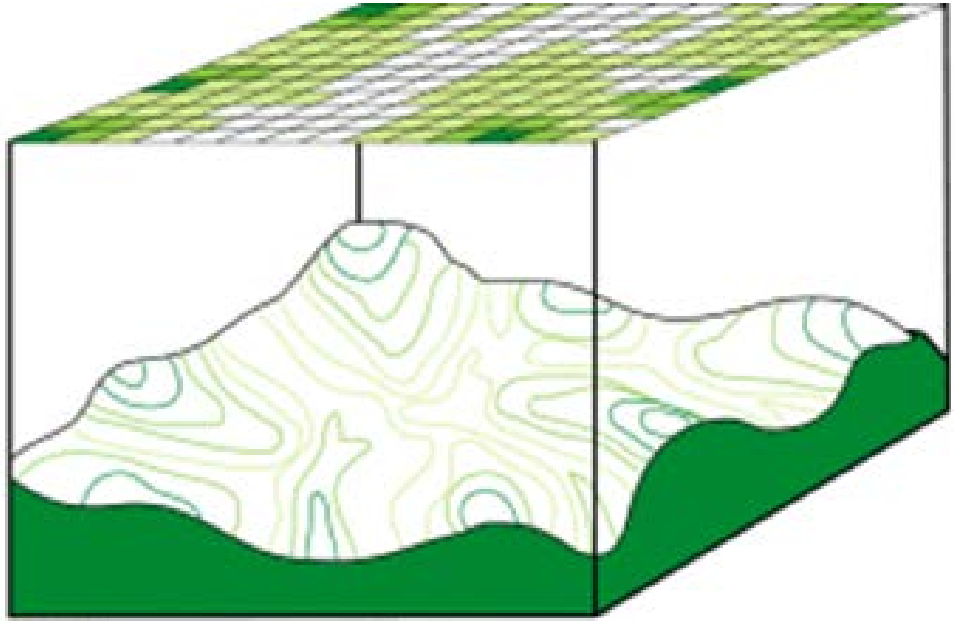
What is a gridded population dataset?



'Raster' GIS format

Spatial continuous data format

The variable as a surface



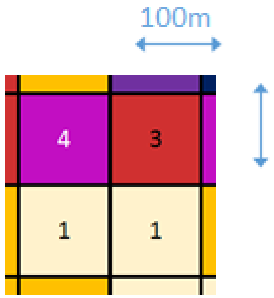
Gridded representation:

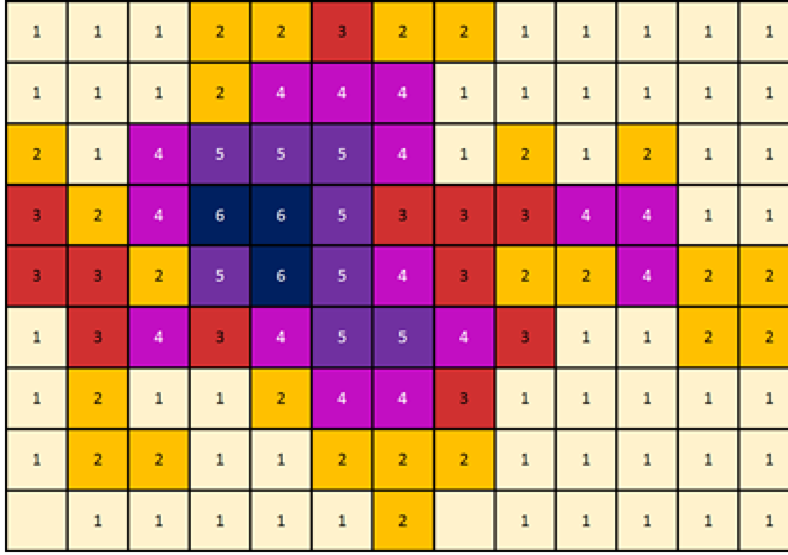
- Set of grid cells
- Uniform cell size and shape
- One value per grid cells

Spatial resolution

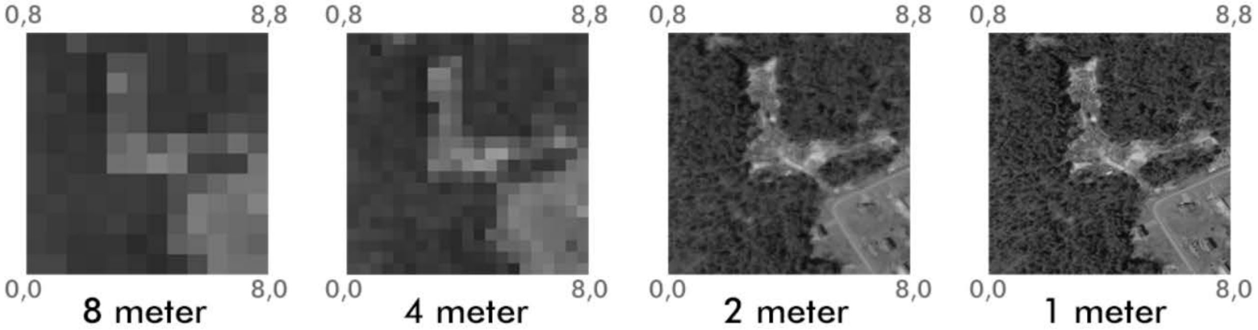
Precision of the area representation

Equivalent to the grid cell size

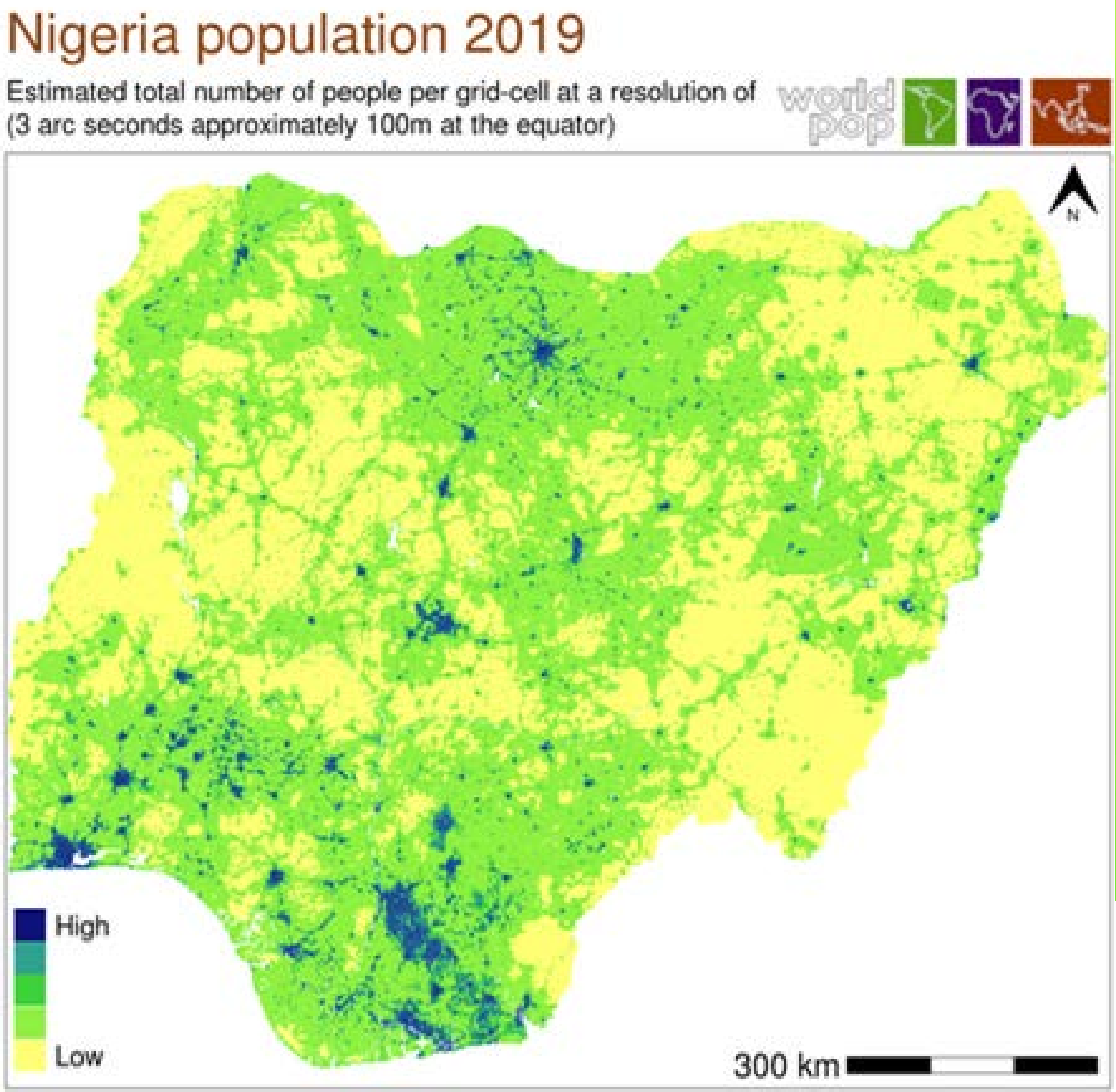


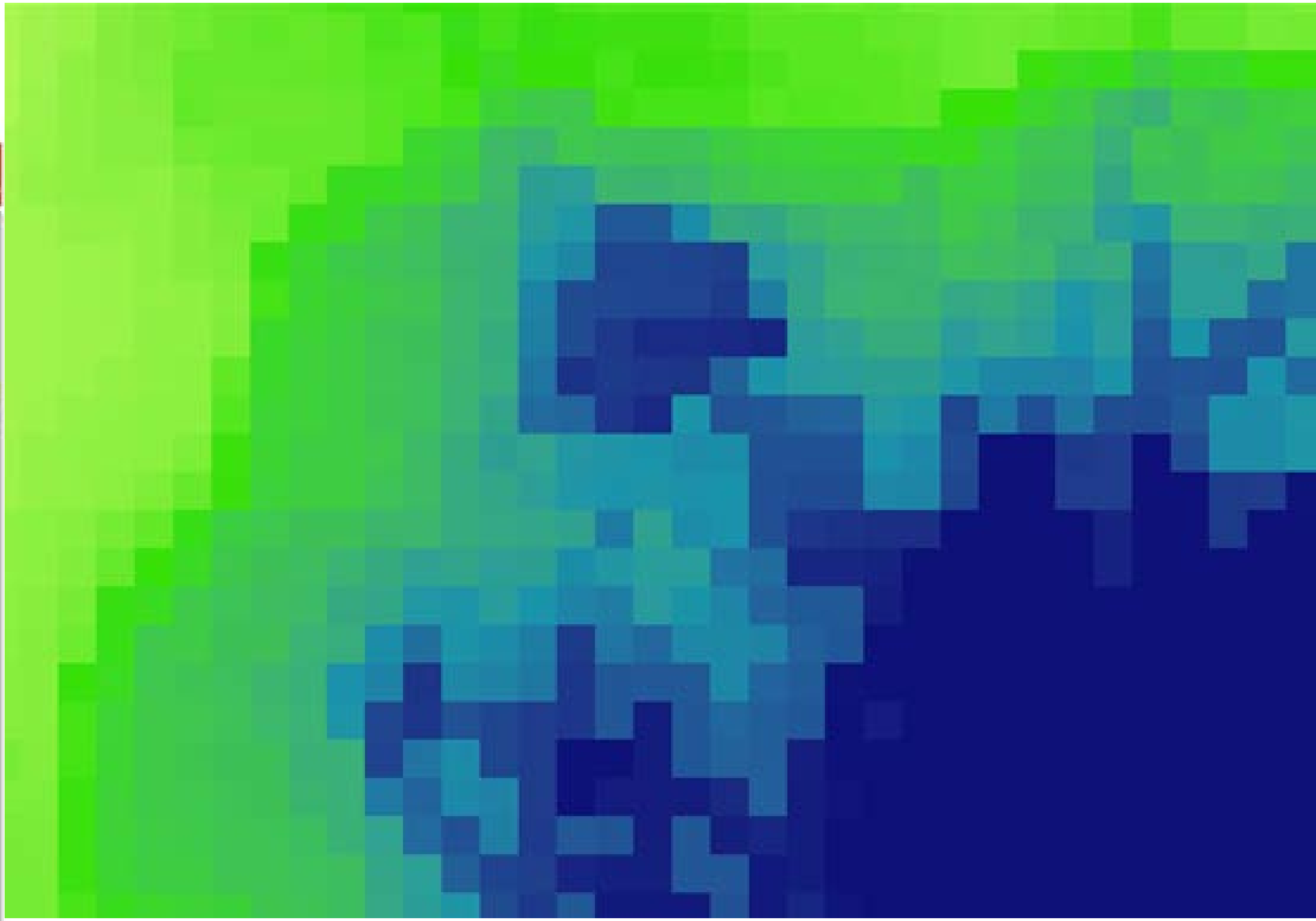


GRID3 output: 100mx100m

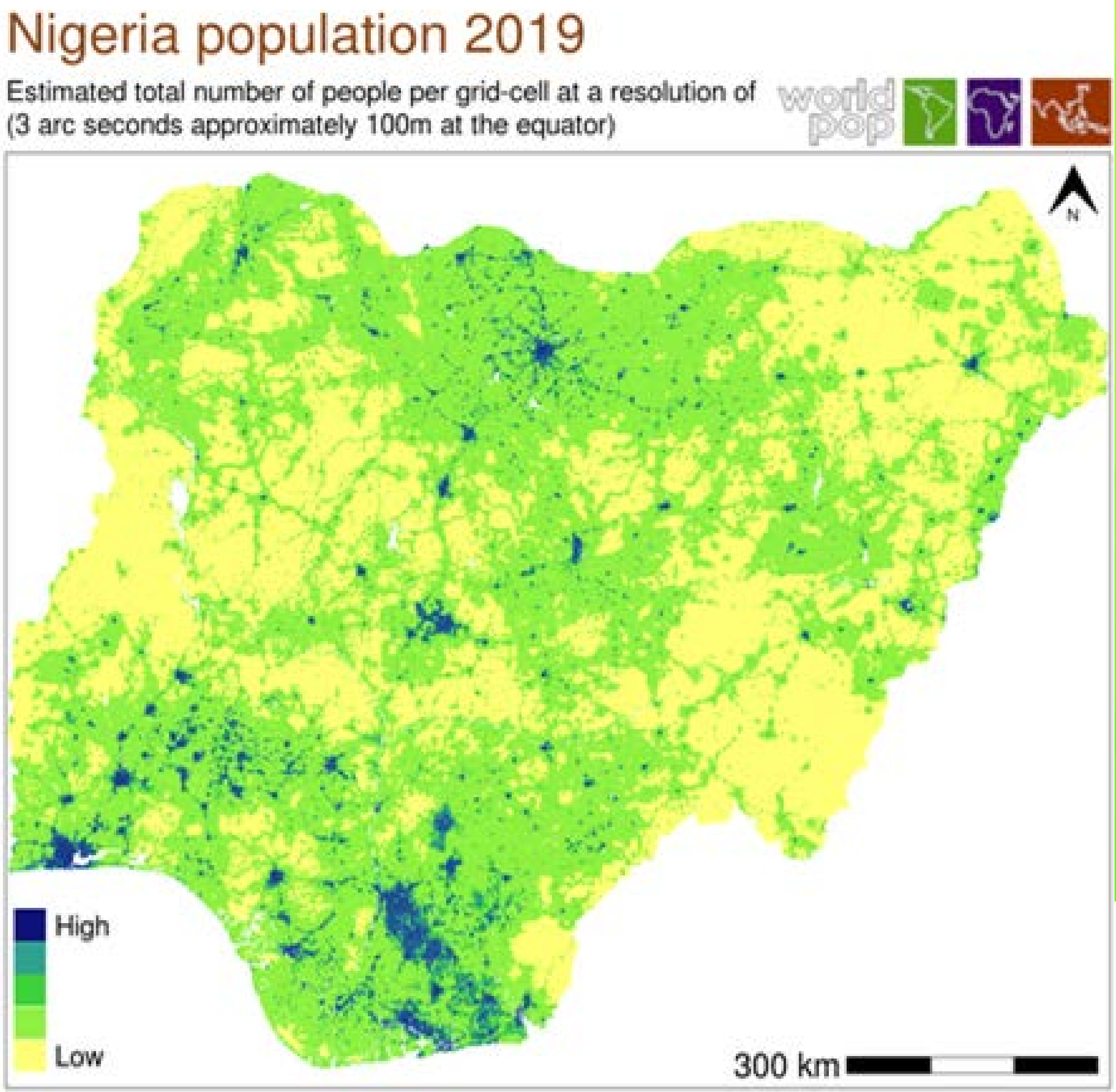


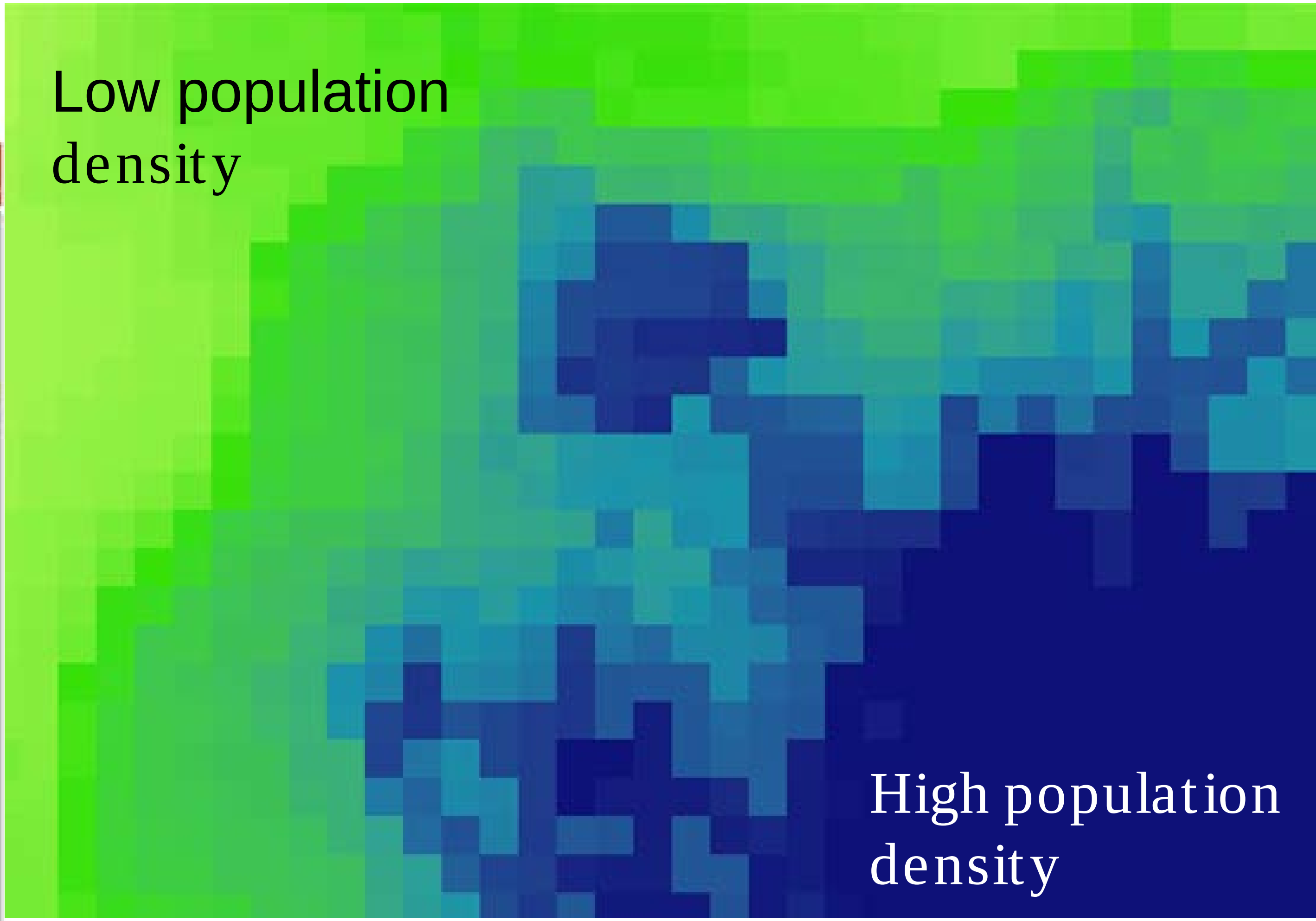
Insight into GIS gridded data format



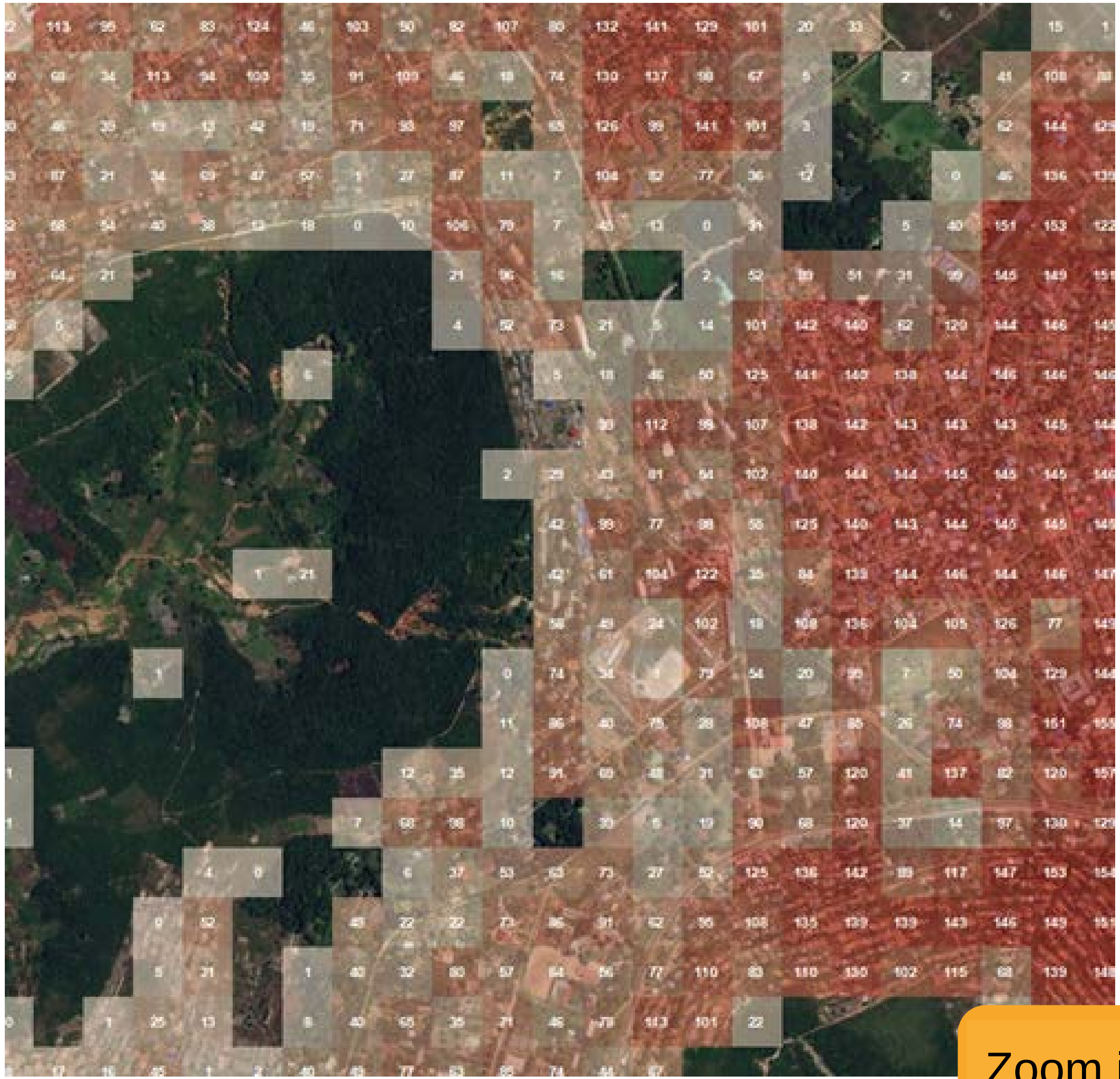


Gridded population - Example of Worldpop dataset





Gridded population - Example of Worldpop dataset



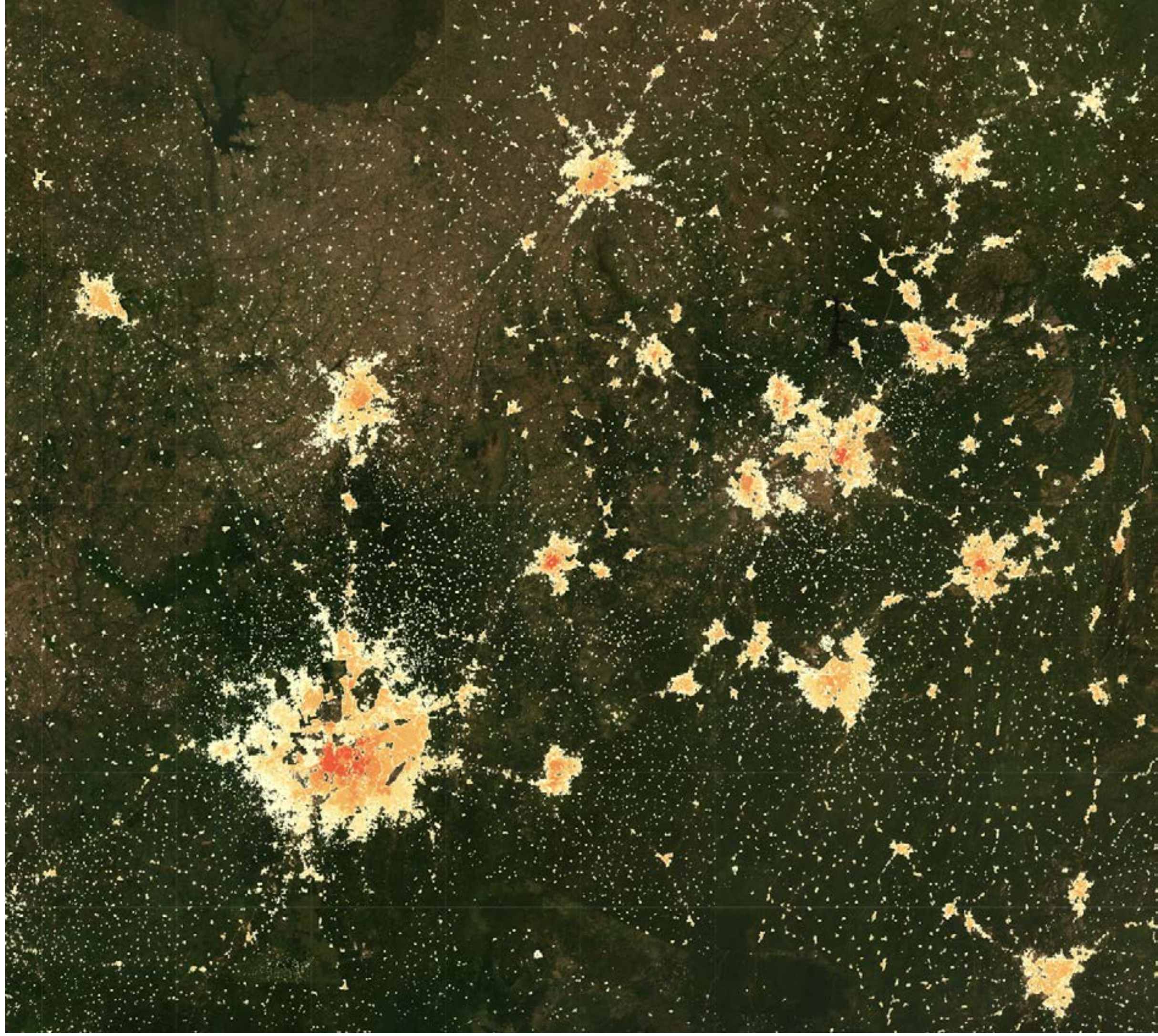
Zoom into a gridded population dataset

Visualisation

A gridded population dataset: what for?

Advantages

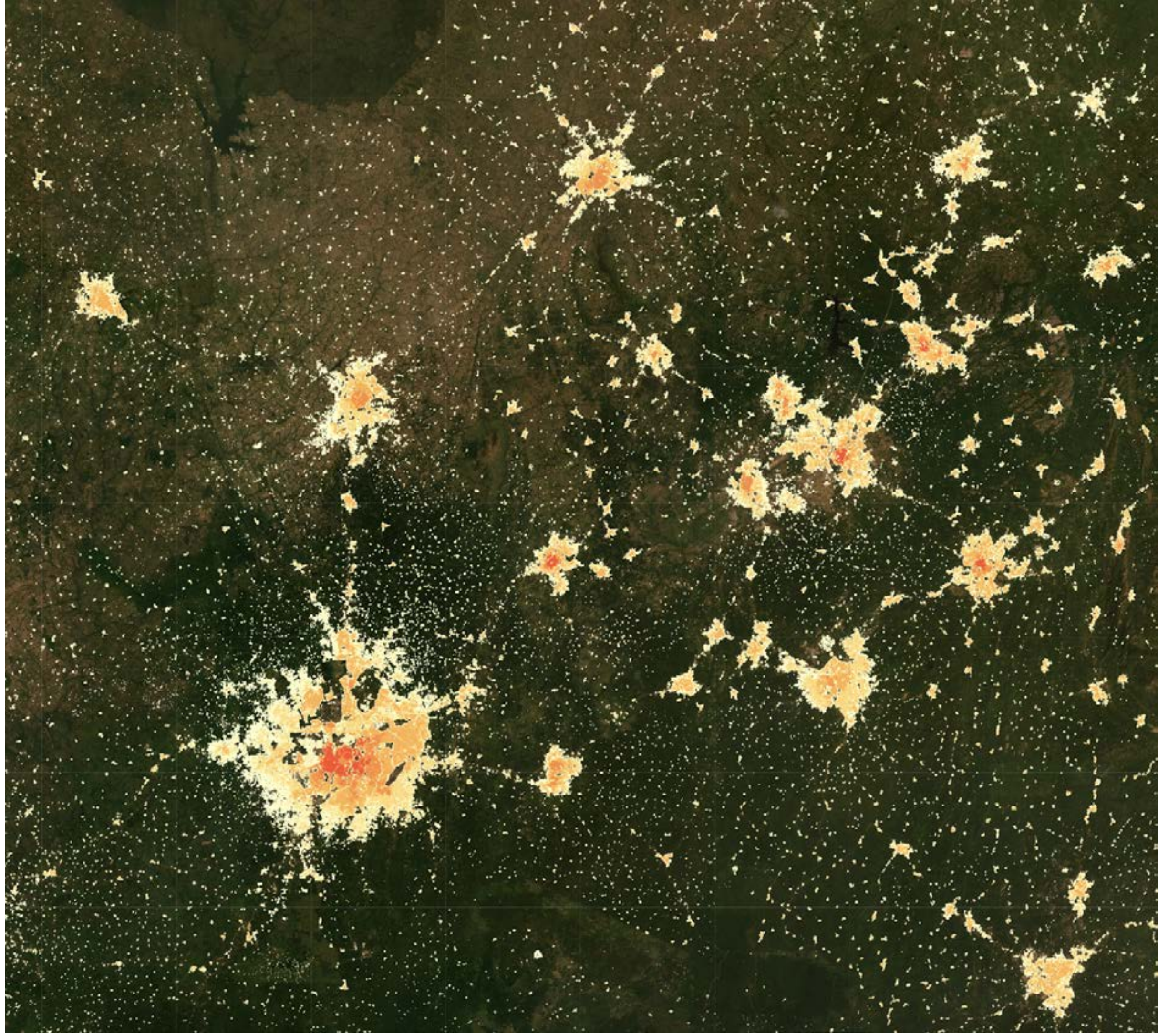
- Fine understanding of **population density variation**
- Good representation of the **spatial distribution**



High-resolution gridded population

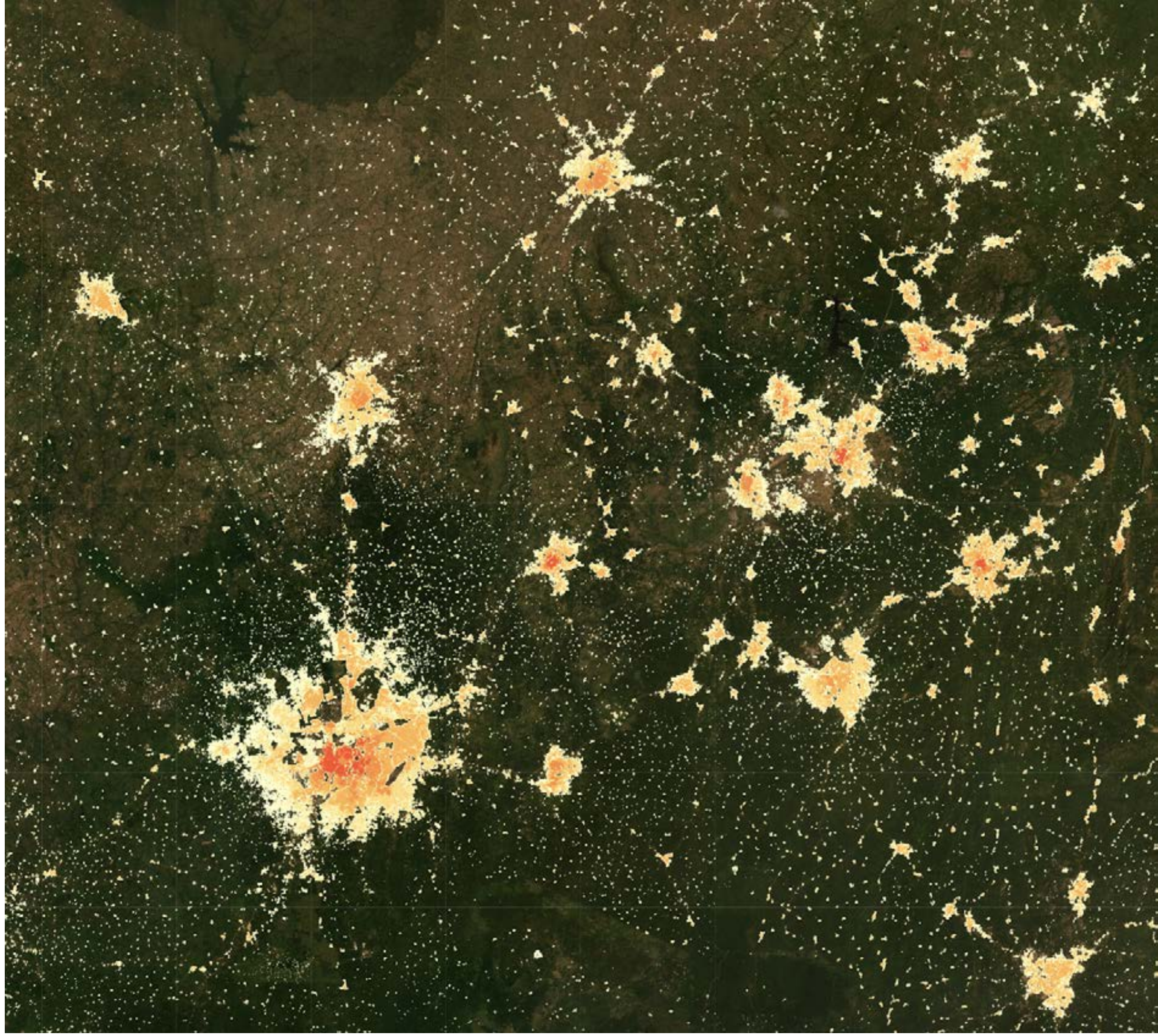
Advantages

- Great **aggregation flexibility**

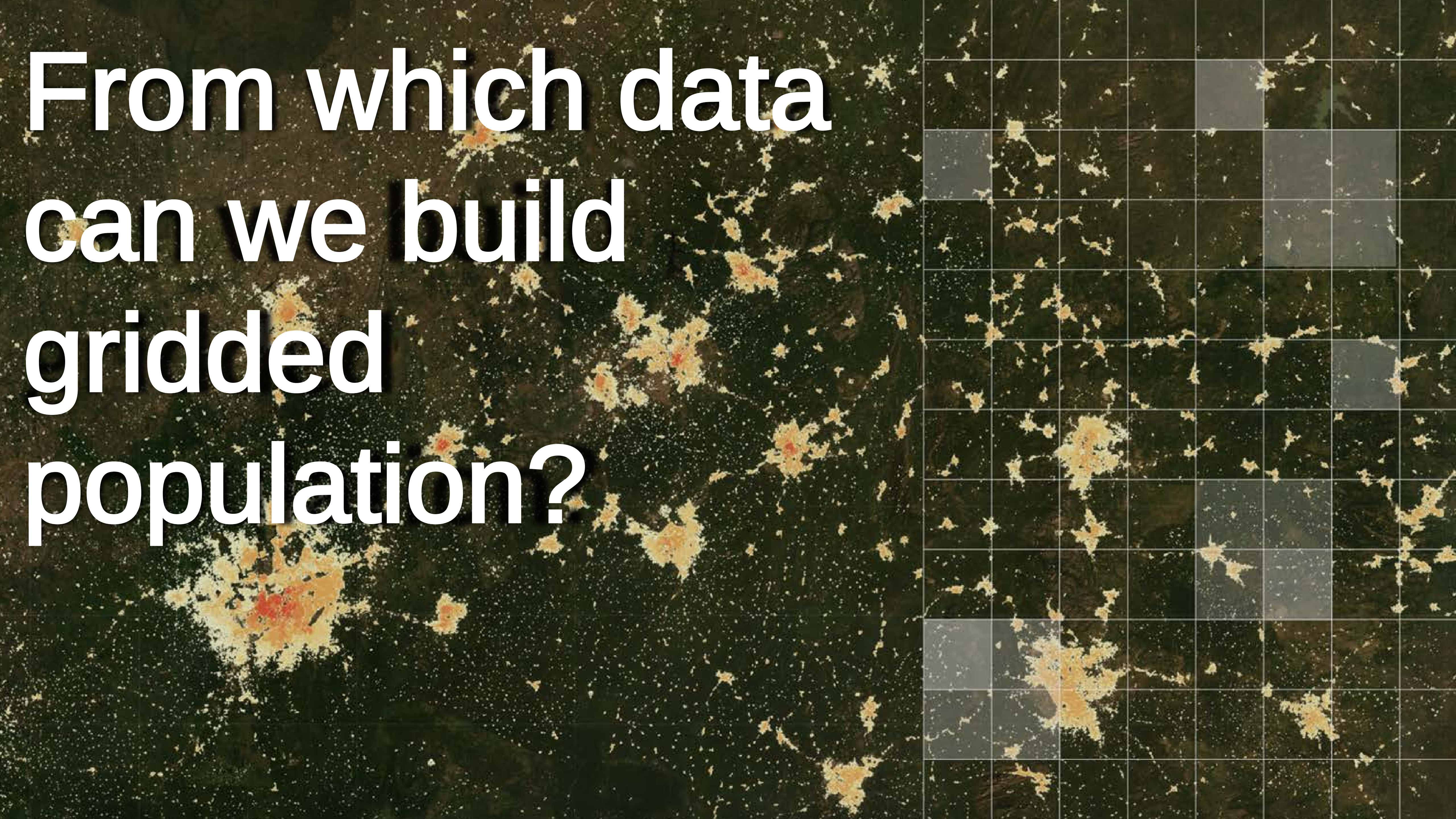


Advantages

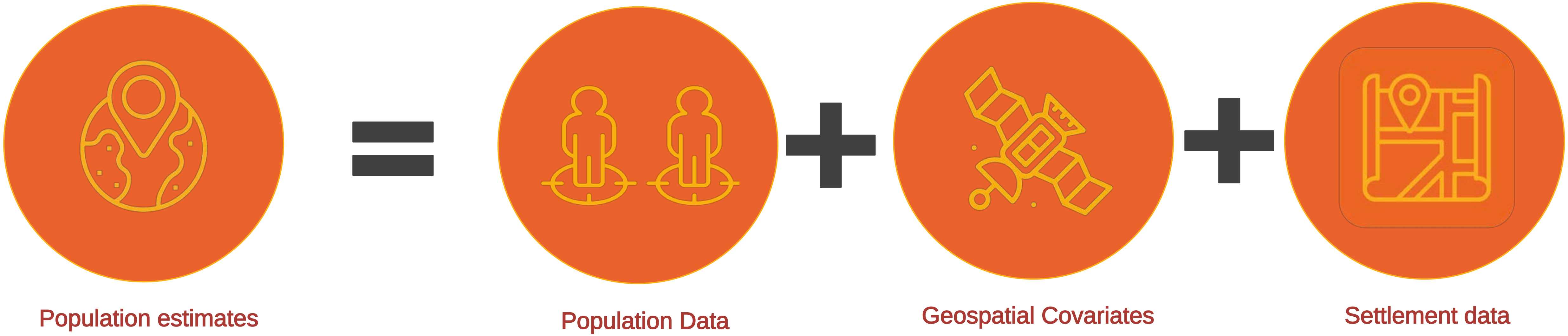
- **Continuous surface** leveraging power for any algorithm
- **Easily integration** with ancillary dataset



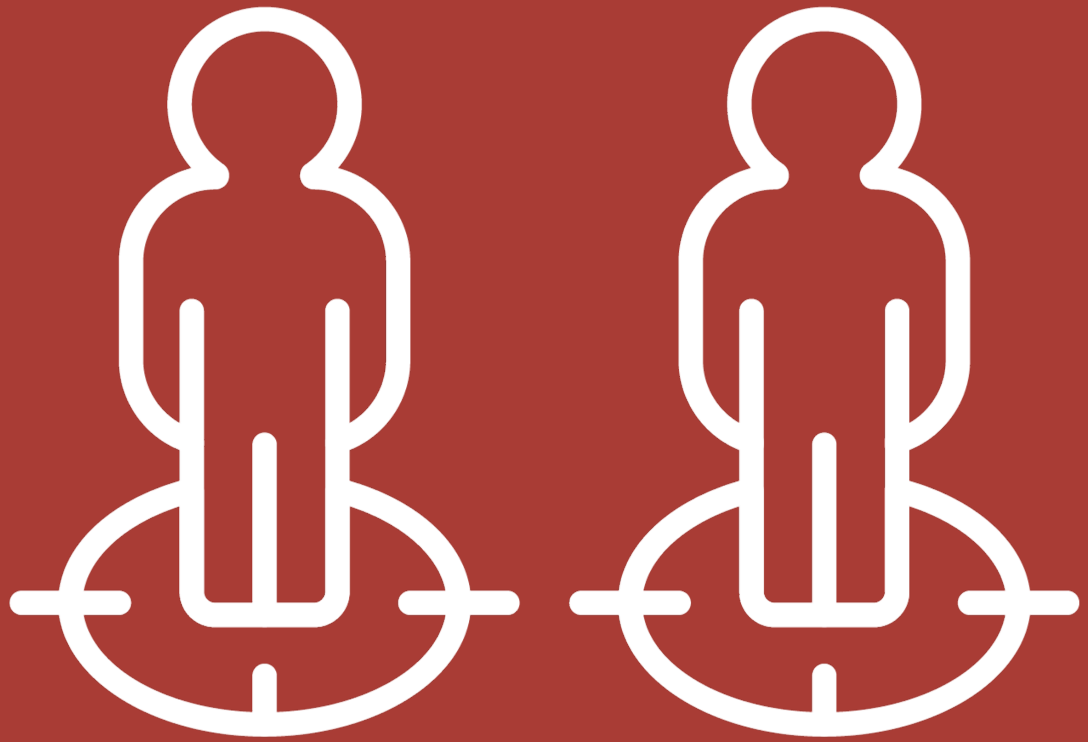
Data requirements


From which data can we build gridded population?

Data input for Population Modelling



Data requirements



Population data

General Key Considerations

Data source

1. Microcensus
2. Partial census
3. Pre-survey household listing

Population data

Microcensus

Years

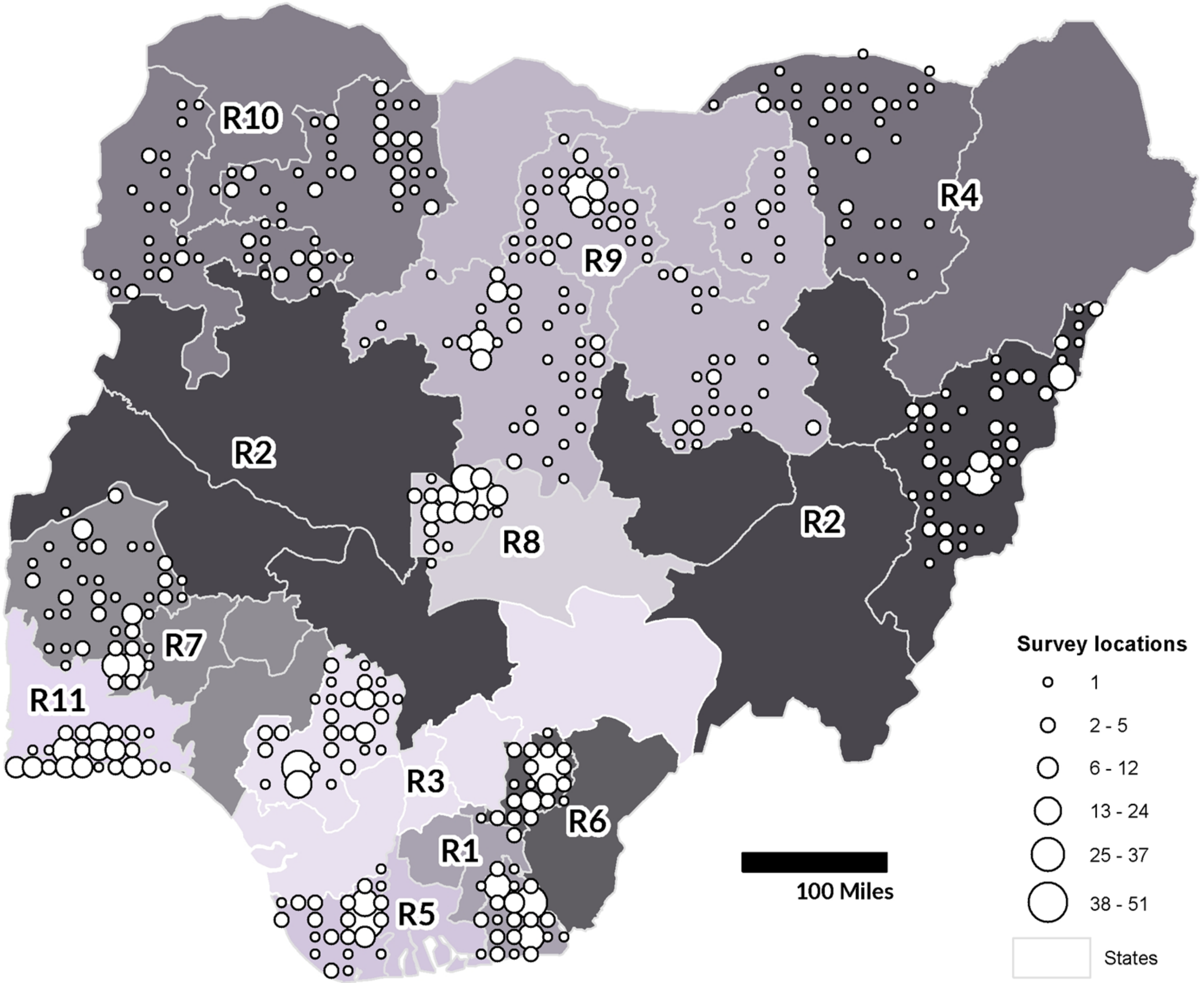
- 2016 - 2017

Collected by

- Oak Ridge National Laboratory
- eHealth Africa

Locations

- 1,141 clusters
- 15 states
- Stratified random sample



Population data for Nigeria

Microcensus

Total enumeration of a well-defined geographical area

One cluster = **3 hectares of settled area**

Three levels:
Housing unit, building, cluster

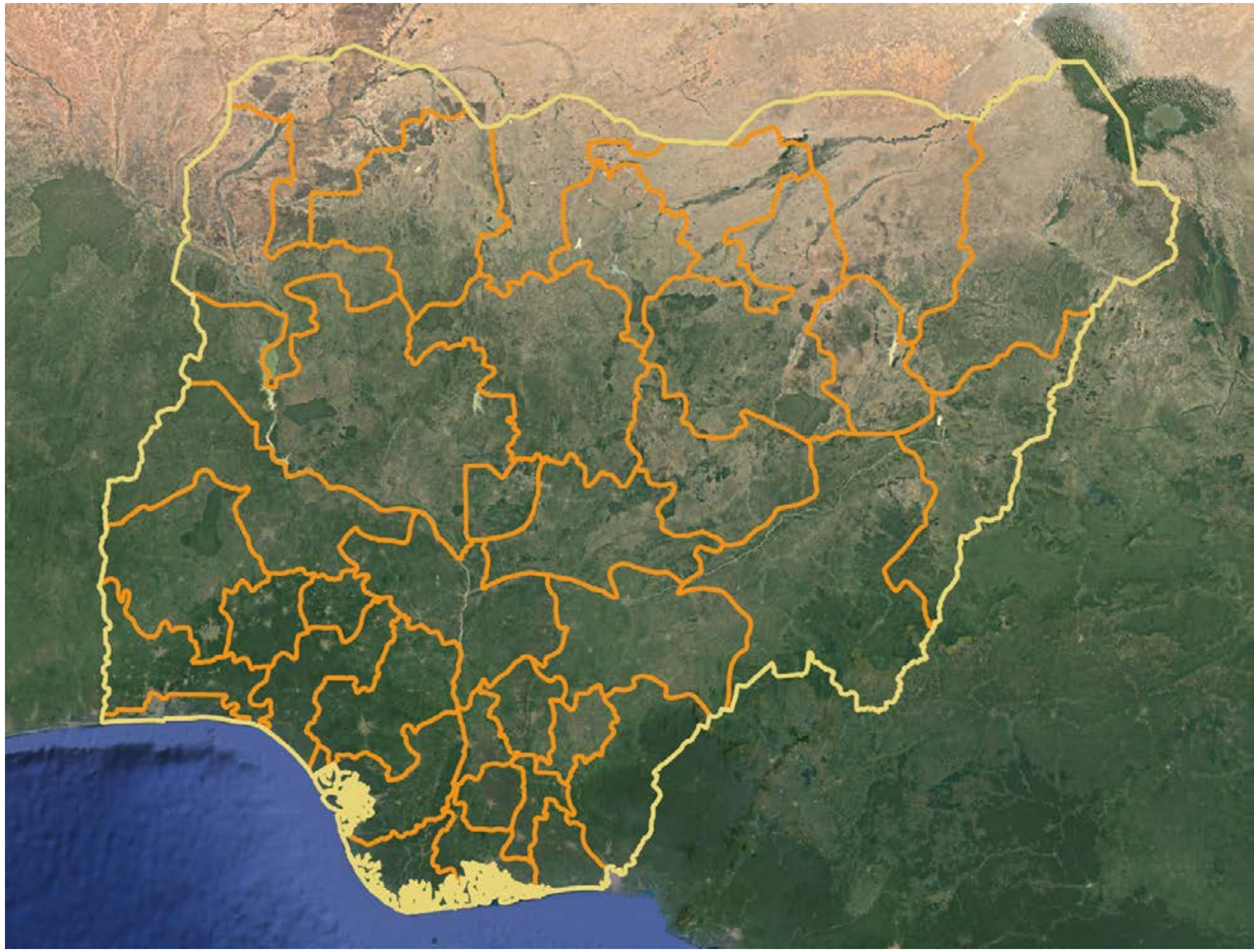
Geospatial covariates

Definition

Data describing the **context** of human settlement

Requirements:

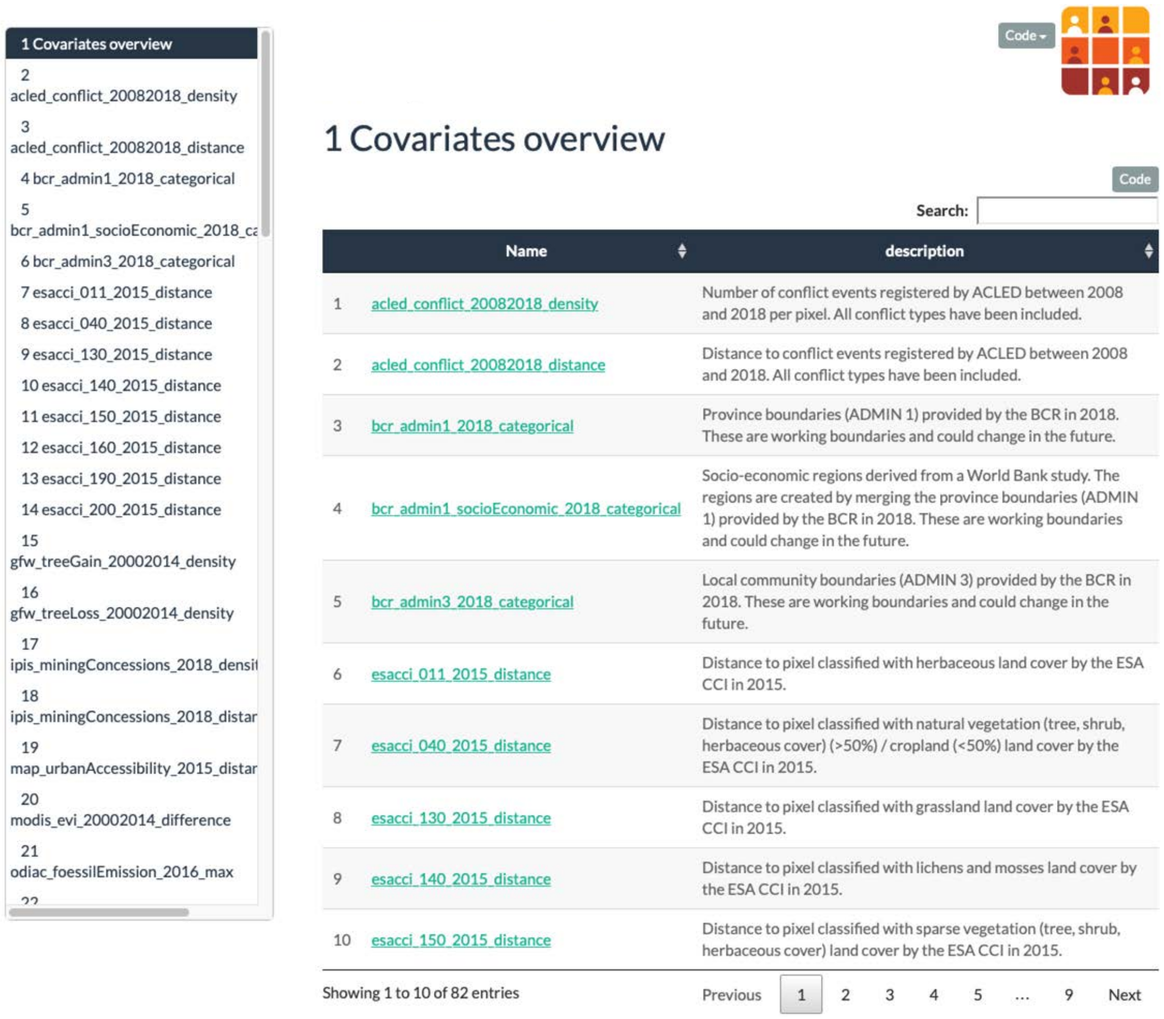
- Influence on the **population spatial distribution**
- Available for the **entire area**
- Include **geographical** information (GPS coordinates …)



Geospatial covariates

Key Considerations

- Physical Context
- Socio-economic Context
- Political Context
- Administrative Context
- Temporal Context


1 Covariates overview
2 acled_conflict_20082018_density
3 acled_conflict_20082018_distance
4 bcr_admin1_2018_categorical
5 bcr_admin1_socioEconomic_2018_ca
6 bcr_admin3_2018_categorical
7 esacci_011_2015_distance
8 esacci_040_2015_distance
9 esacci_130_2015_distance
10 esacci_140_2015_distance
11 esacci_150_2015_distance
12 esacci_160_2015_distance
13 esacci_190_2015_distance
14 esacci_200_2015_distance
15 gfw_treeGain_20002014_density
16 gfw_treeLoss_20002014_density
17 ipis_miningConcessions_2018_densit
18 ipis_miningConcessions_2018_distar
19 map_urbanAccessibility_2015_distar
20 modis_evi_20002014_difference
21 odiac_foessilEmission_2016_max

Code

1 Covariates overview

Code

Search:

	Name	description
1	acled_conflict_20082018_density	Number of conflict events registered by ACLED between 2008 and 2018 per pixel. All conflict types have been included.
2	acled_conflict_20082018_distance	Distance to conflict events registered by ACLED between 2008 and 2018. All conflict types have been included.
3	bcr_admin1_2018_categorical	Province boundaries (ADMIN 1) provided by the BCR in 2018. These are working boundaries and could change in the future.
4	bcr_admin1_socioEconomic_2018_categorical	Socio-economic regions derived from a World Bank study. The regions are created by merging the province boundaries (ADMIN 1) provided by the BCR in 2018. These are working boundaries and could change in the future.
5	bcr_admin3_2018_categorical	Local community boundaries (ADMIN 3) provided by the BCR in 2018. These are working boundaries and could change in the future.
6	esacci_011_2015_distance	Distance to pixel classified with herbaceous land cover by the ESA CCI in 2015.
7	esacci_040_2015_distance	Distance to pixel classified with natural vegetation (tree, shrub, herbaceous cover) (>50%) / cropland (<50%) land cover by the ESA CCI in 2015.
8	esacci_130_2015_distance	Distance to pixel classified with grassland land cover by the ESA CCI in 2015.
9	esacci_140_2015_distance	Distance to pixel classified with lichens and mosses land cover by the ESA CCI in 2015.
10	esacci_150_2015_distance	Distance to pixel classified with sparse vegetation (tree, shrub, herbaceous cover) land cover by the ESA CCI in 2015.

Showing 1 to 10 of 82 entries

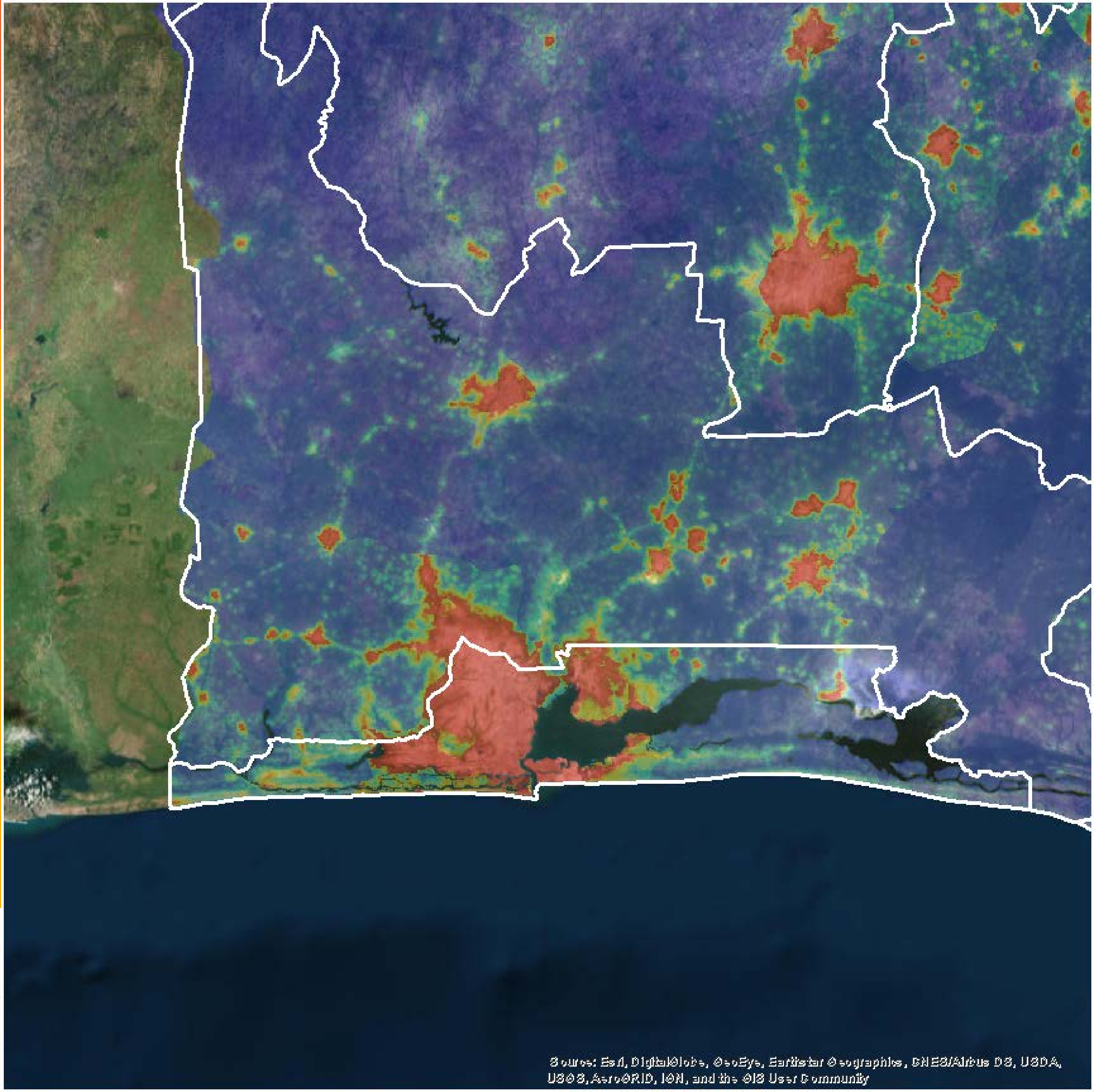
Previous 1 2 3 4 5 … 9 Next

Geospatial covariates

Geospatial Covariates

- WorldPop Global
  worldpop.org

For Nigeria model



Geospatial Covariates

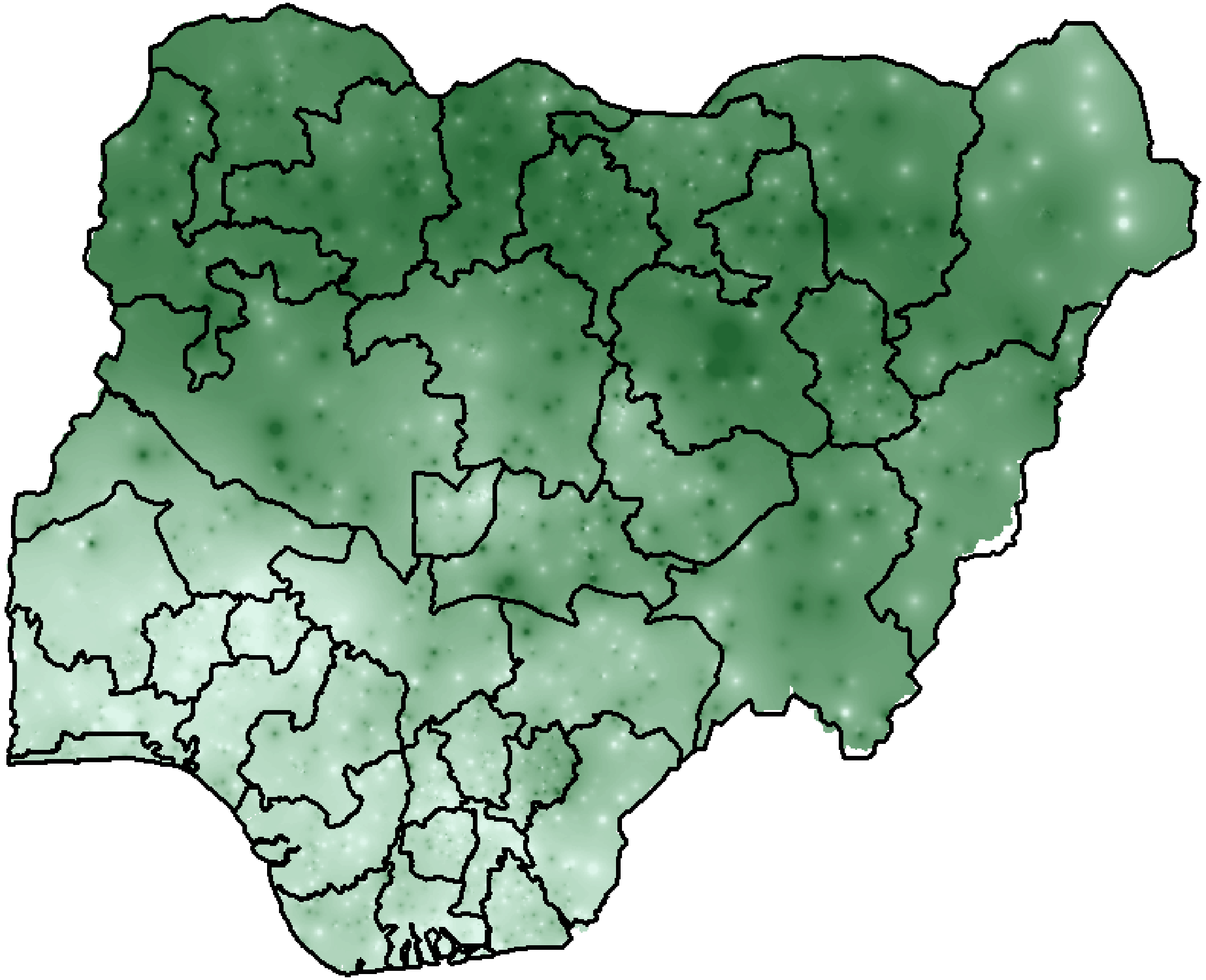
- WorldPop Global
- School density

For Nigeria model

Geospatial Covariates

- WorldPop Global
- School density
- Household size

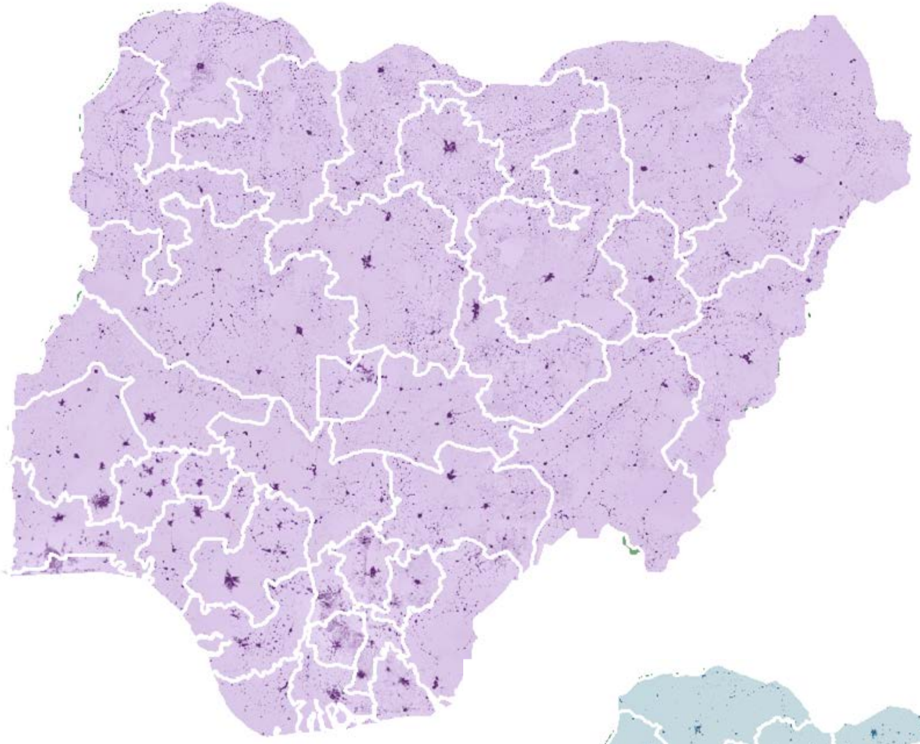
Demographic and health survey

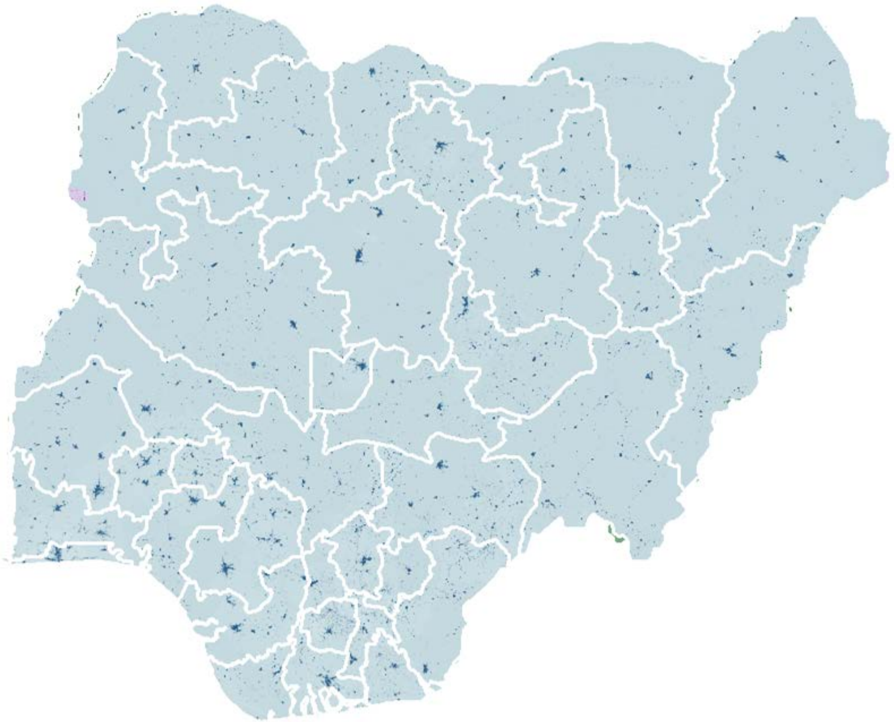


For Nigeria model

Geospatial Covariates

- WorldPop Global
- School density
- Household size
- Nearby (1 km) residential & non-residential settlements





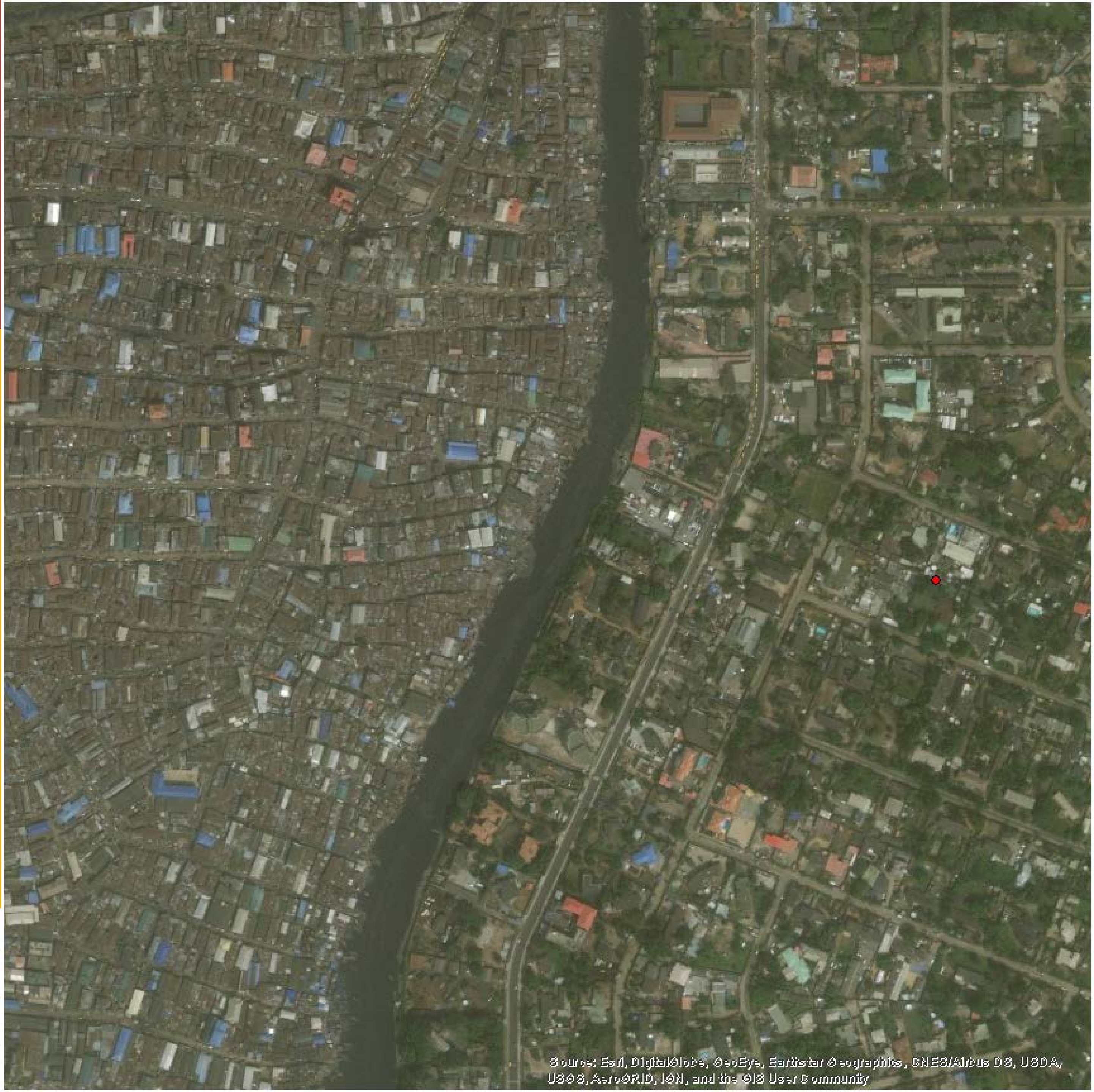
For Nigeria model

Settlement types

Settlement Types

LandScan HD v1.1

Oak Ridge National Lab



For Nigeria model

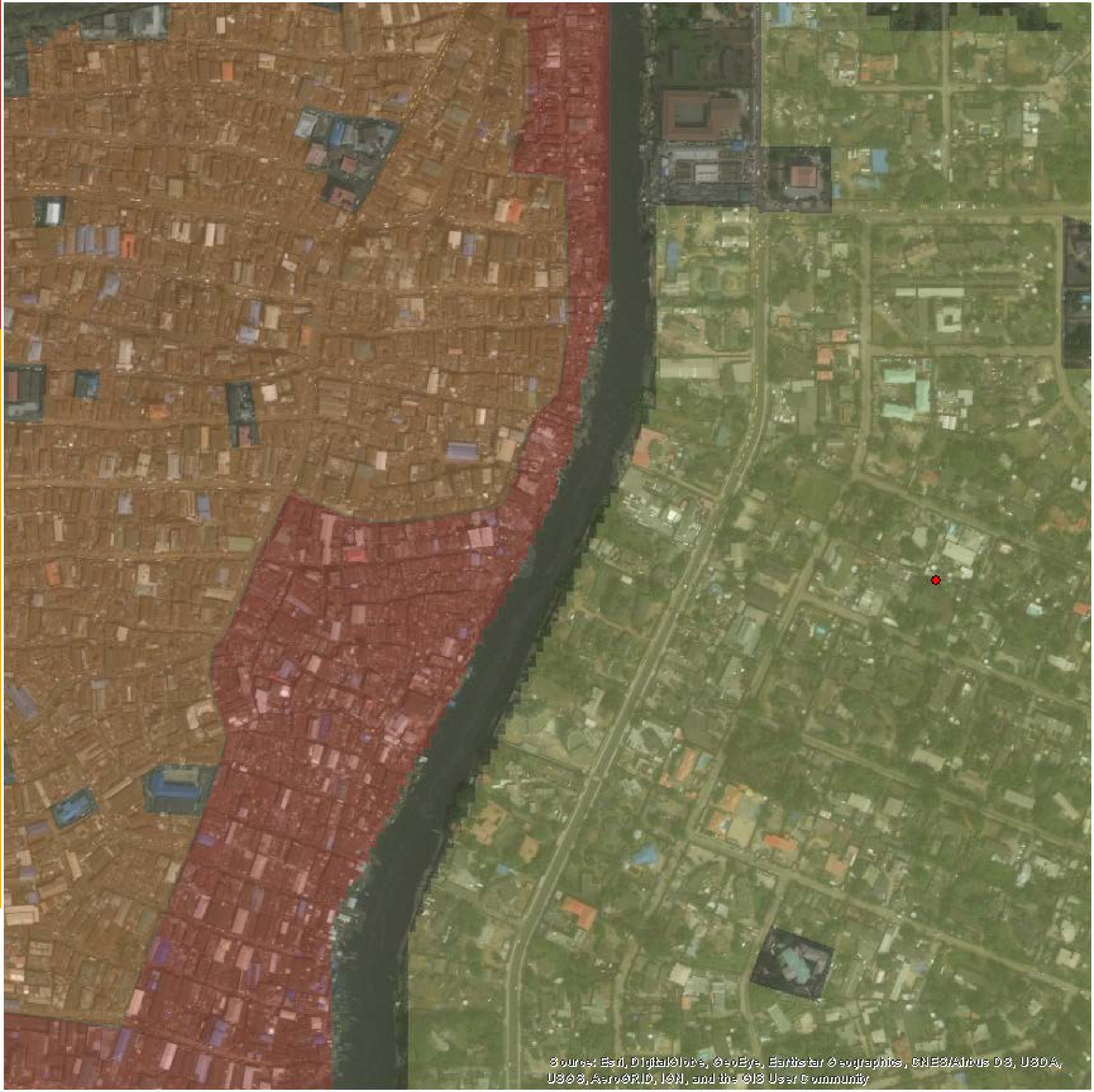
Settlement Types

LandScan HD v1.1

Oak Ridge National Lab

- Urban (A, B, D, F)
- Rural (M)
- Non-residential (Z)

For Nigeria model

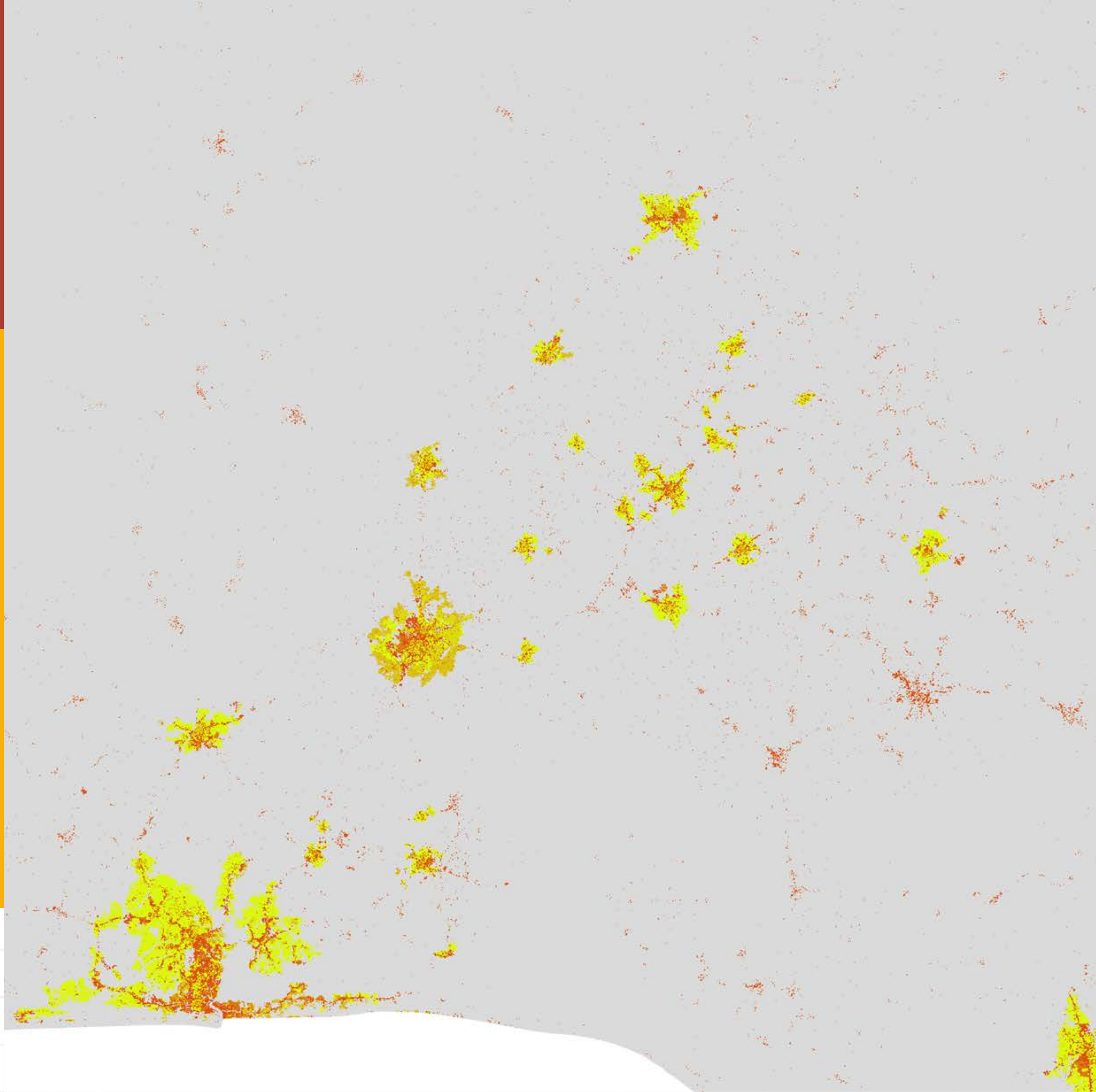


Settlement Types

LandScan HD v1.1

Oak Ridge National Lab

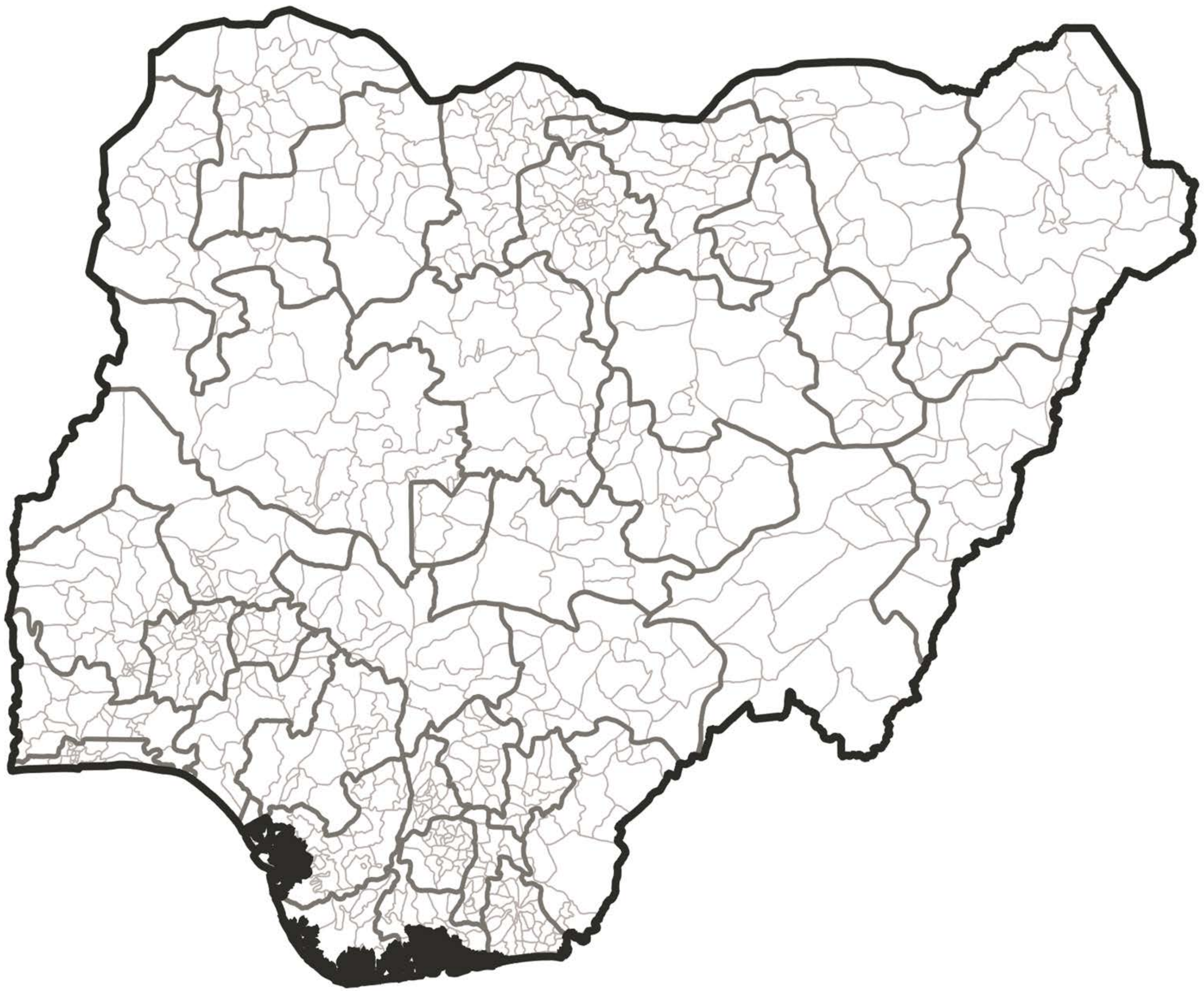
- Urban (A, B, D, F)
- Rural (M)
- Non-residential (Z)



For Nigeria model

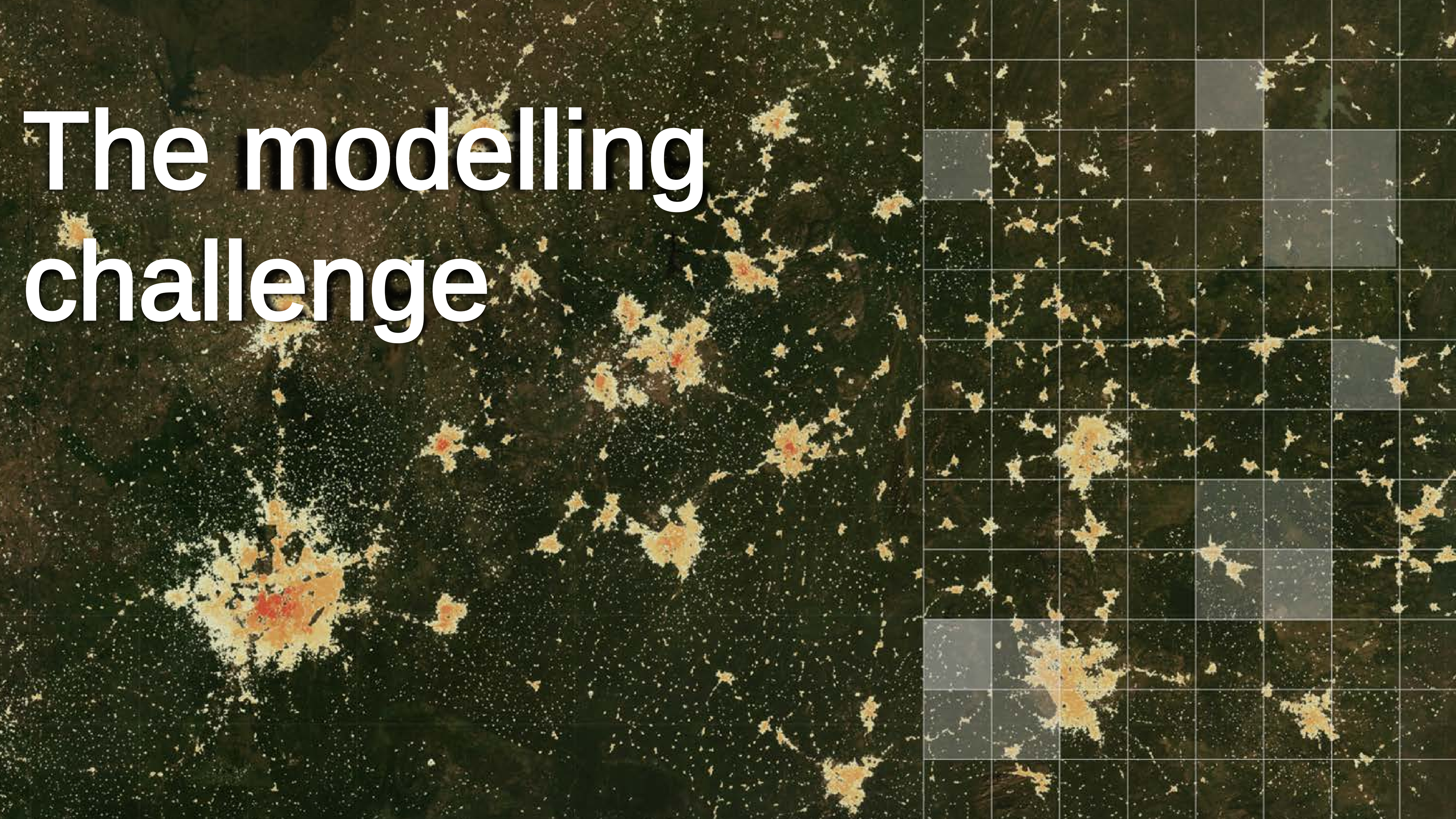
Administrative boundaries

Administrative Boundaries

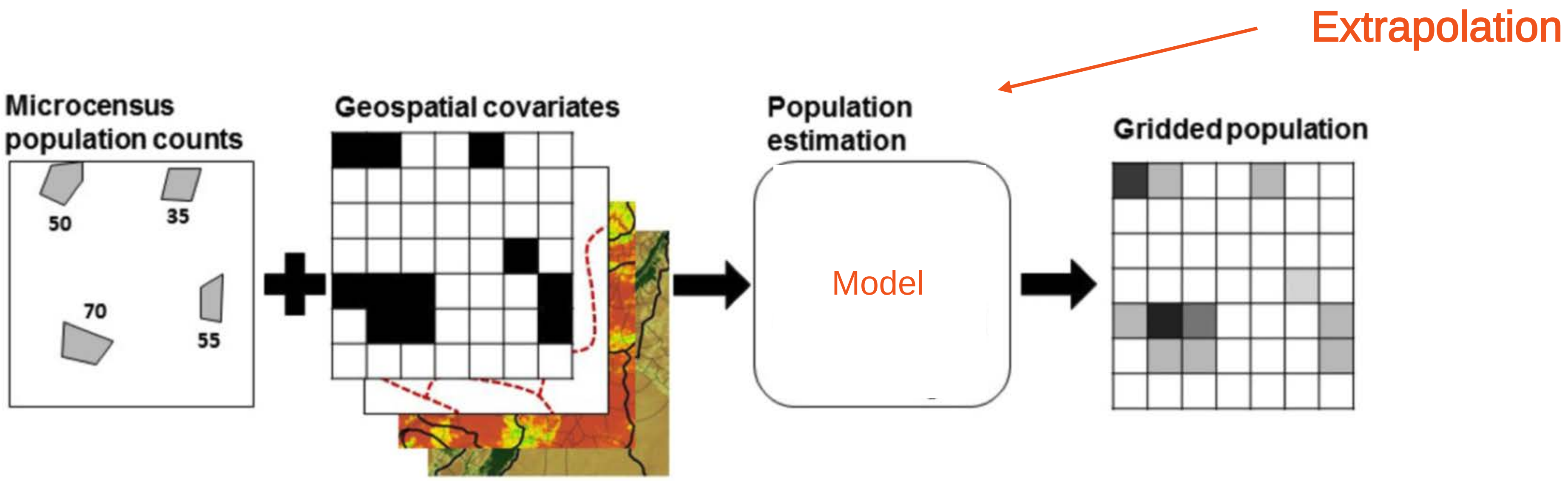


For Nigeria model

The modelling challenge



Bottom -up Modelling



Modelling objectives

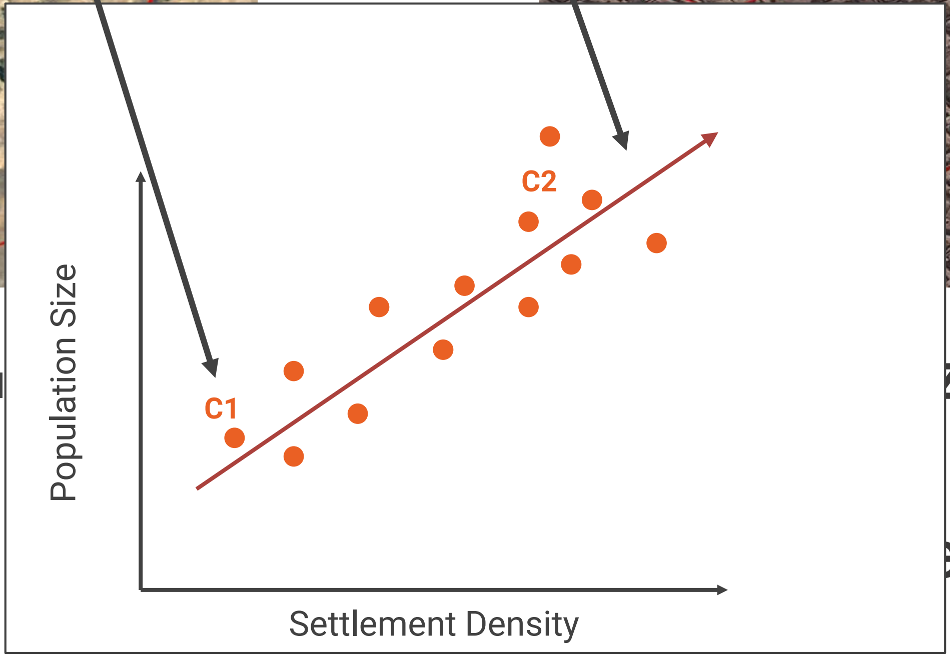
Relationship to population

Cluster 1

Population size | 500 peopl

Settlement density | low

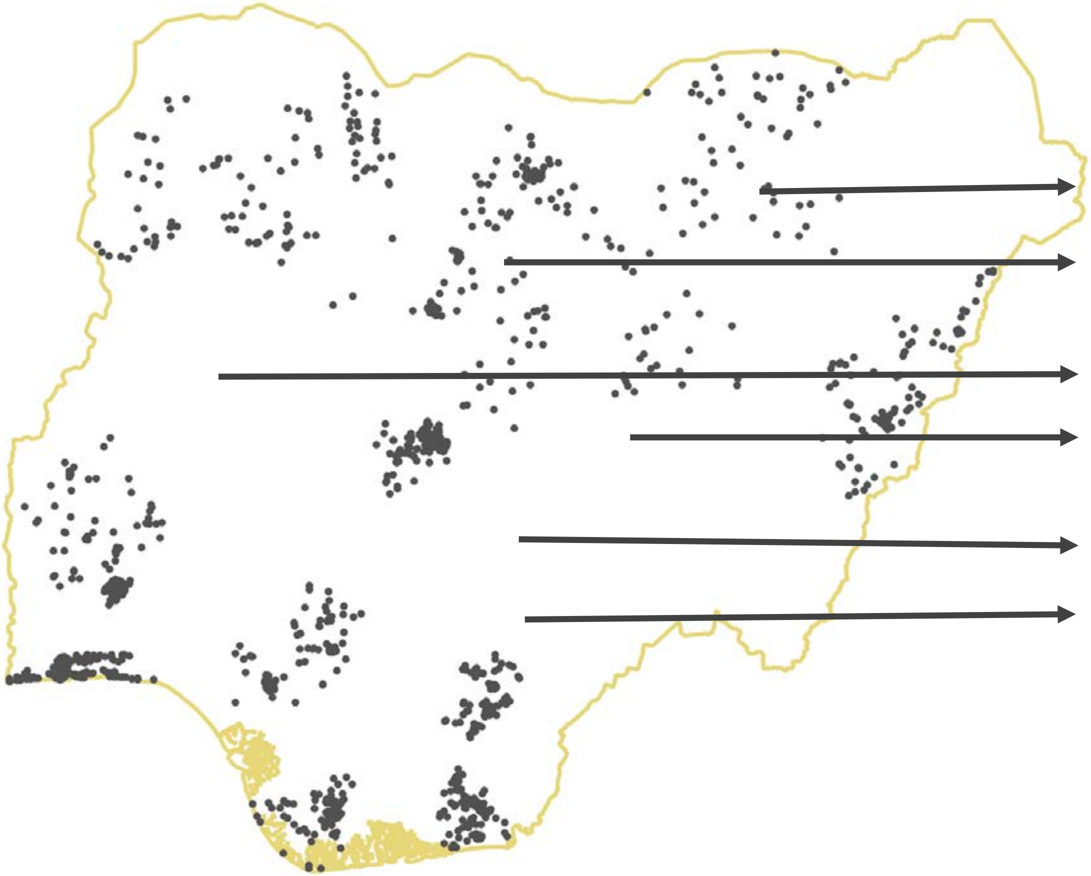
Cluster 2

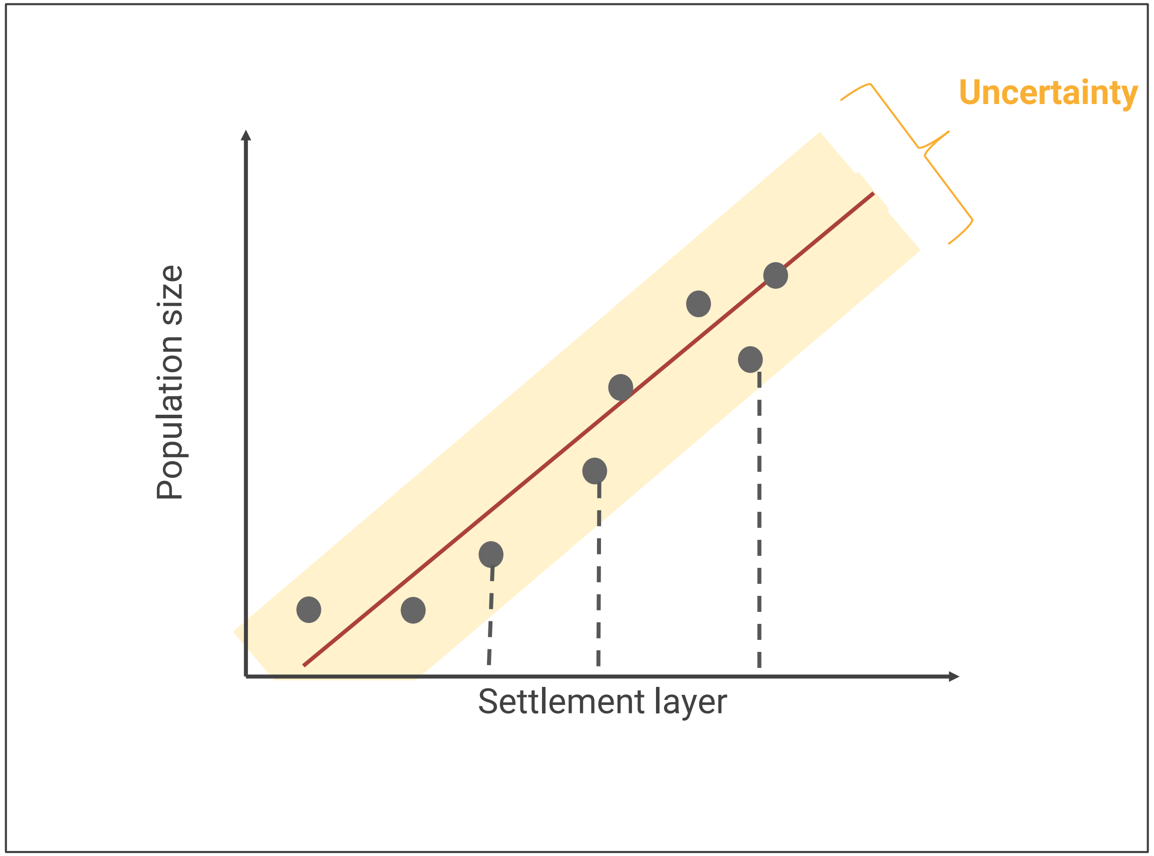


ze| 1,000 people

ensity | high

Prediction in unsurveyed area





Relationship to population

Cluster 1

Settlement density | low

Mean slope | high

Nightlights | low

Distance to urban center | high

Cluster 2

Settlement density | high

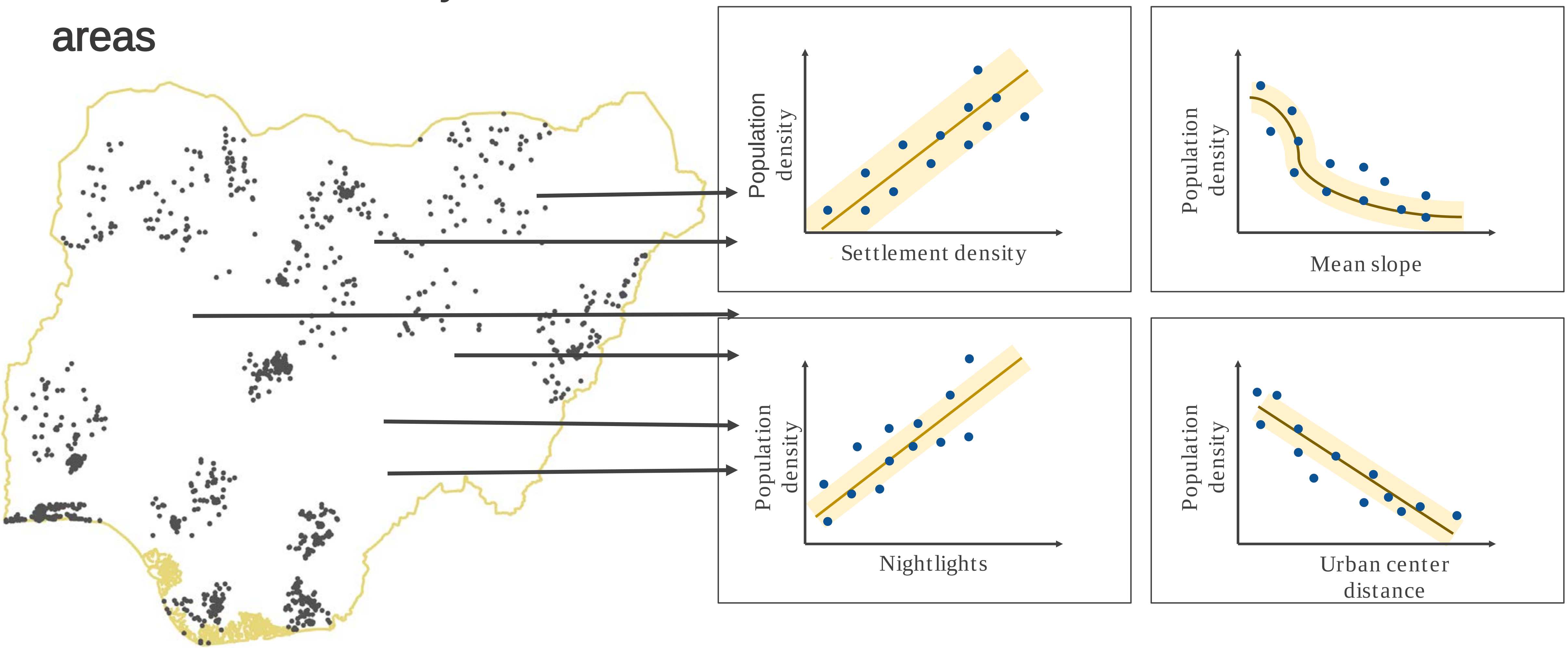
Mean slope | low

Nightlights | high

Distance to urban center | low

Relationship to population

Prediction in unsurveyed areas



Relationship to population

Statistical Modelling

Specificity of Statistical Modelling (SM)

Uncertainty : SMs explicitly take uncertainty into account by specifying a probabilistic model for the data.

Structural : SMs typically start by assuming additivity of predictor effects when building the model.

Empirical : SMs are focused on pre-specified parameters of special interest.

=> **SM is better in low signal/noise environment: world of human outcome**

Stochastic Modelling

- Stochastic = what is random and thus unknown
- Linked to:
  - Impossibility to know all relevant predictors
  - Errors in data collection
  - Imperfect representativity of the sample
  - Individual variation

Deterministic Model

$$N_i = D_i \times A_i$$

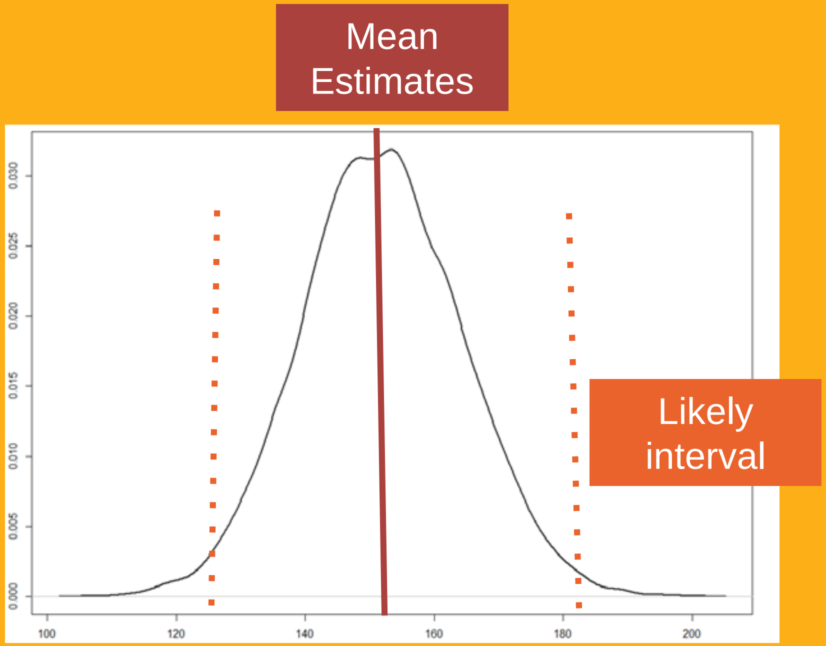
$$\neq$$

Stochastic Model

$$N_i \sim \text{Poisson}(D_i \times A_i)$$

Bayesian Framework

- **Estimates distribution prediction**
  - Most likely values, from which:
    - mean
    - Inferior and superior bounds (95%)
  - Crucial to quantify uncertainty

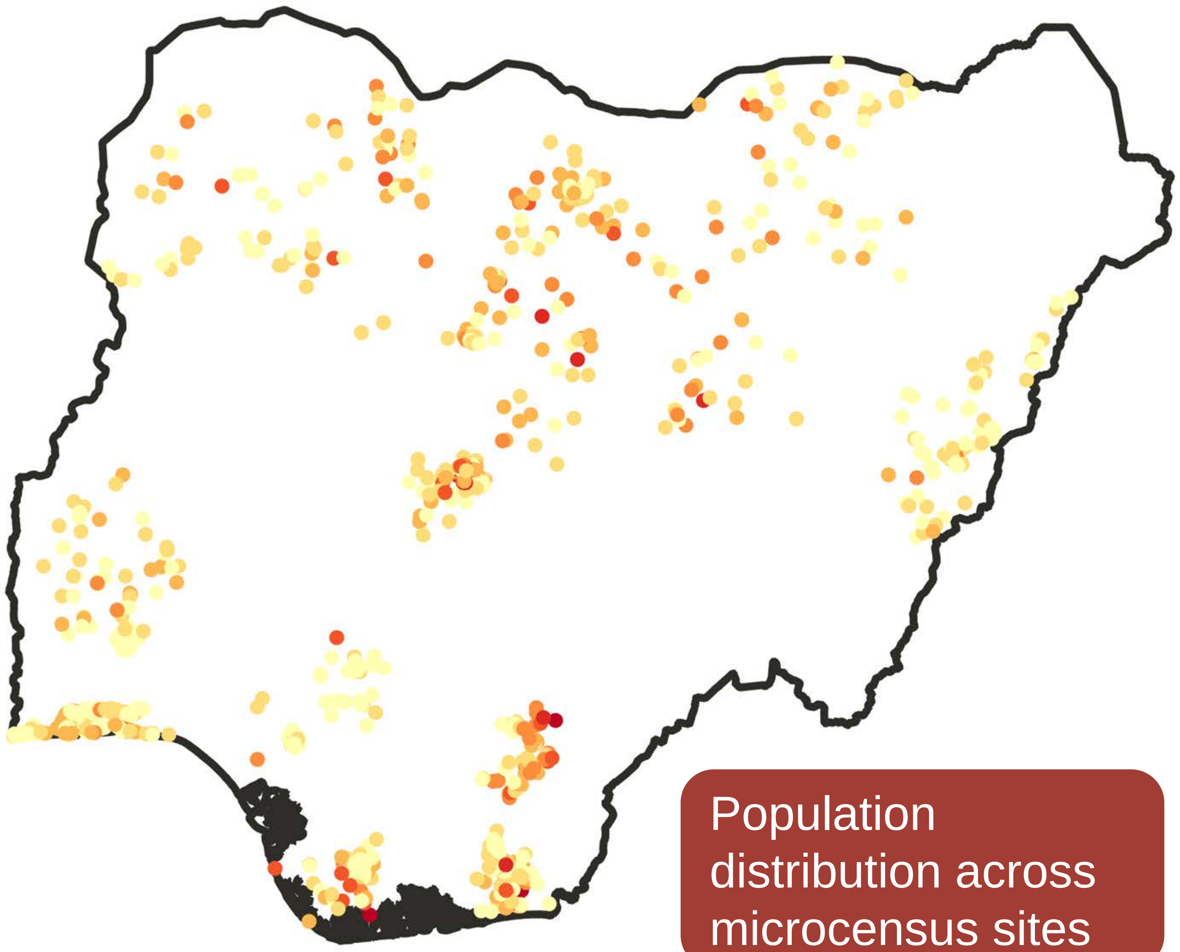


Population estimates distribution of a pixel

Advantages of Statistical Modelling

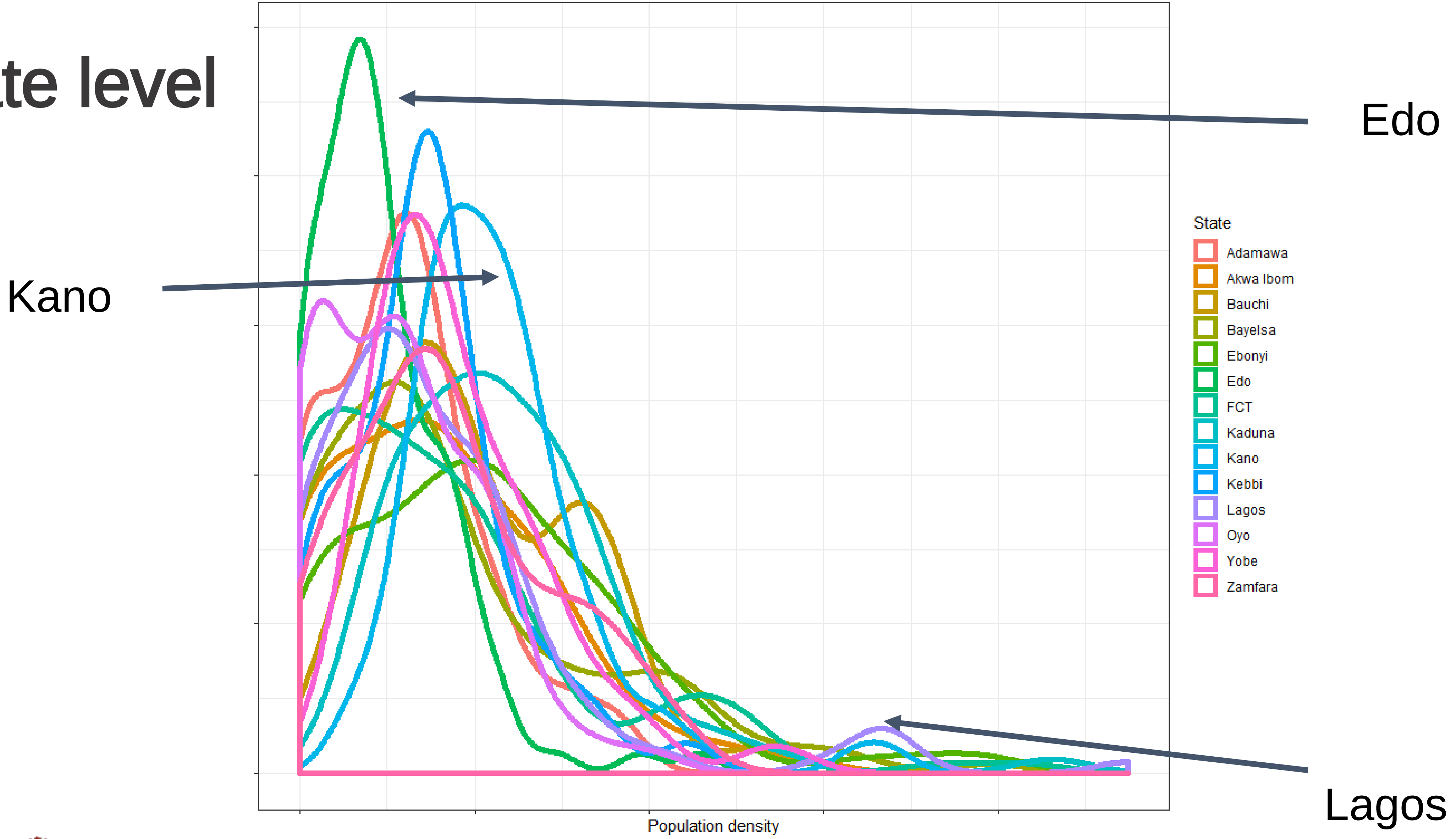
Hierarchical Framework

- Population density variation occurs at different level that can be summarise within a **hierarchical framework**



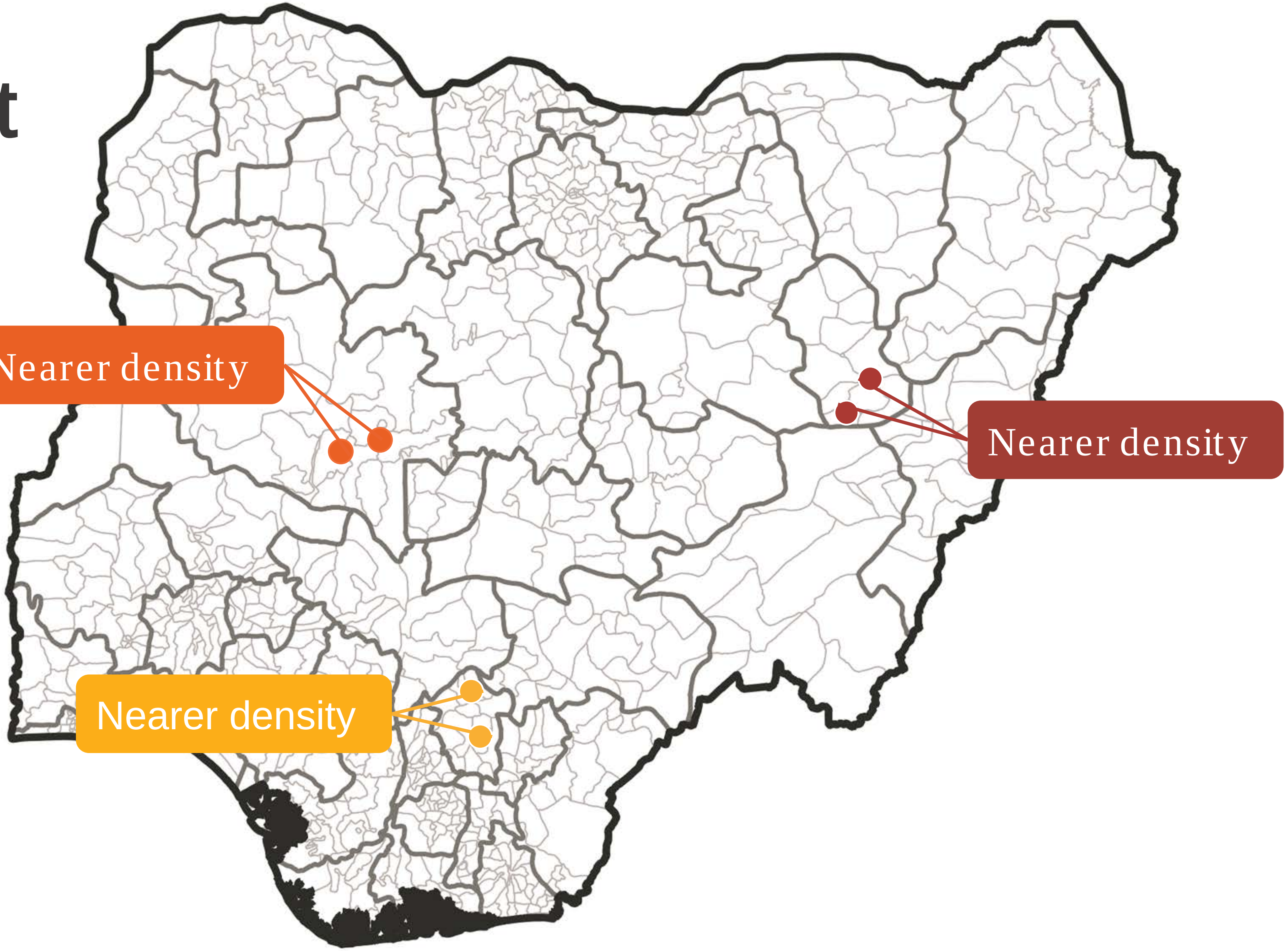
Population distribution across microcensus sites

State level



Hierarchical Setting

Local Government Area



Hierarchical Setting

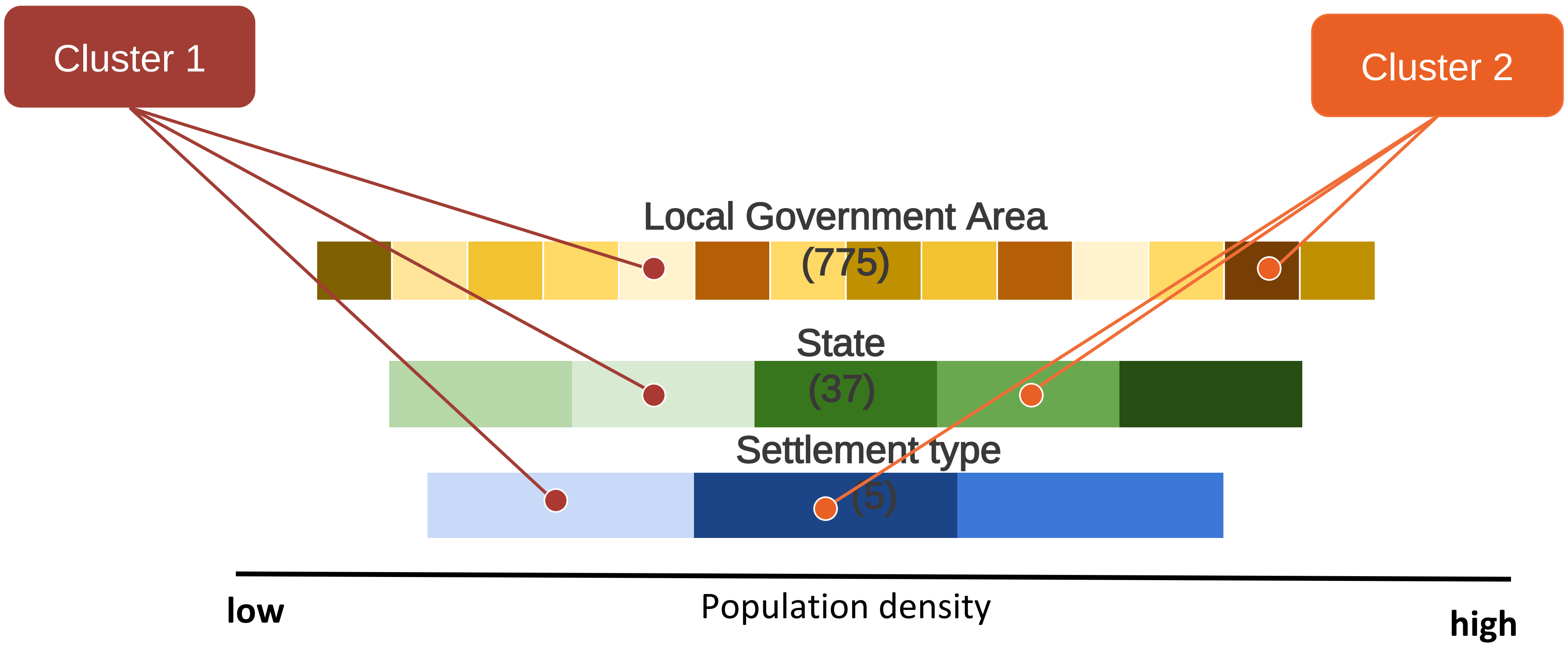
Settlement type

Urban

Rural

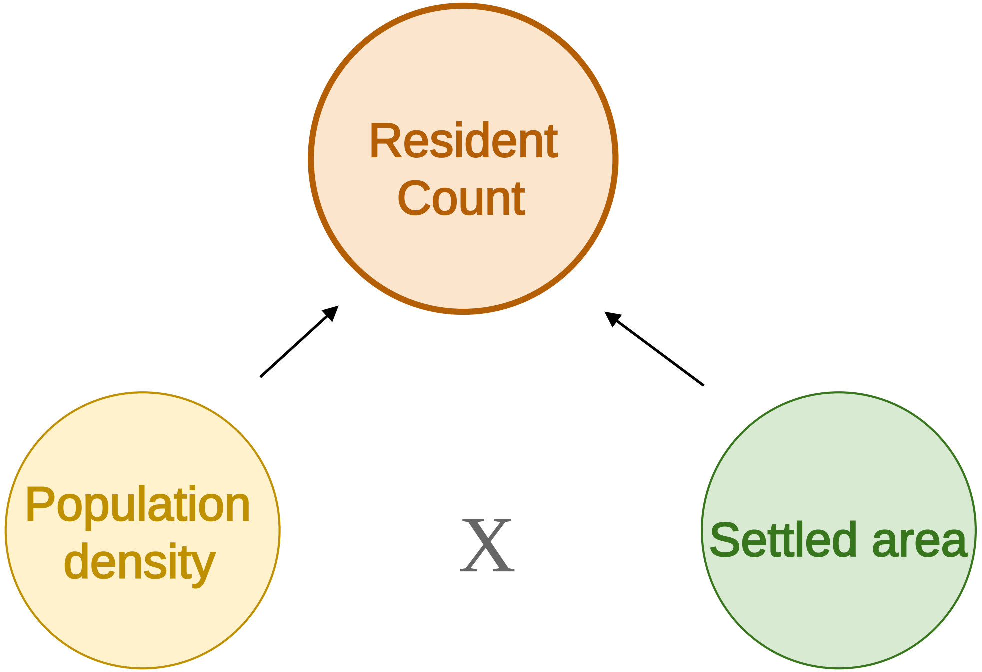
Village

Hierarchical Setting



Hierarchical Setting

End goal

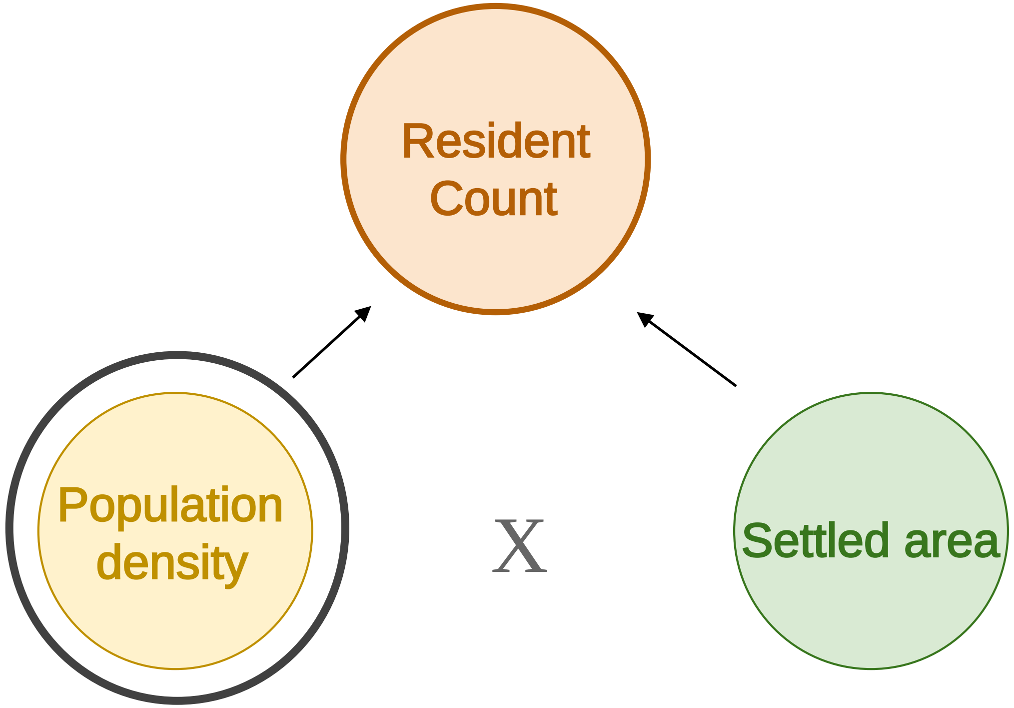


Resident Count | 255 people
Area | 7,0 hectares

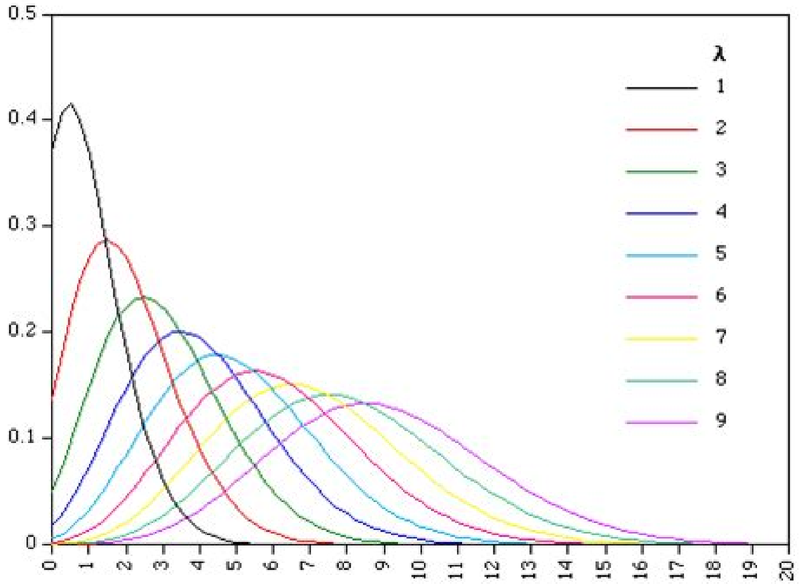
Settled area | 3,5 hectares
Population density | 72 people/hectare

Model structure

Poisson distribution for the Resident Count



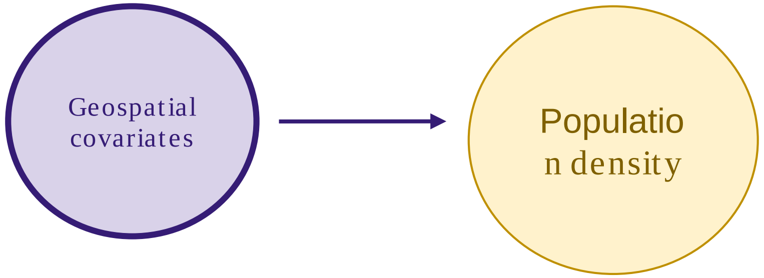
- Resident count:
  - Discrete, positv
    => Poisson distribution
- Parameters:
  - Pop density x Settled area

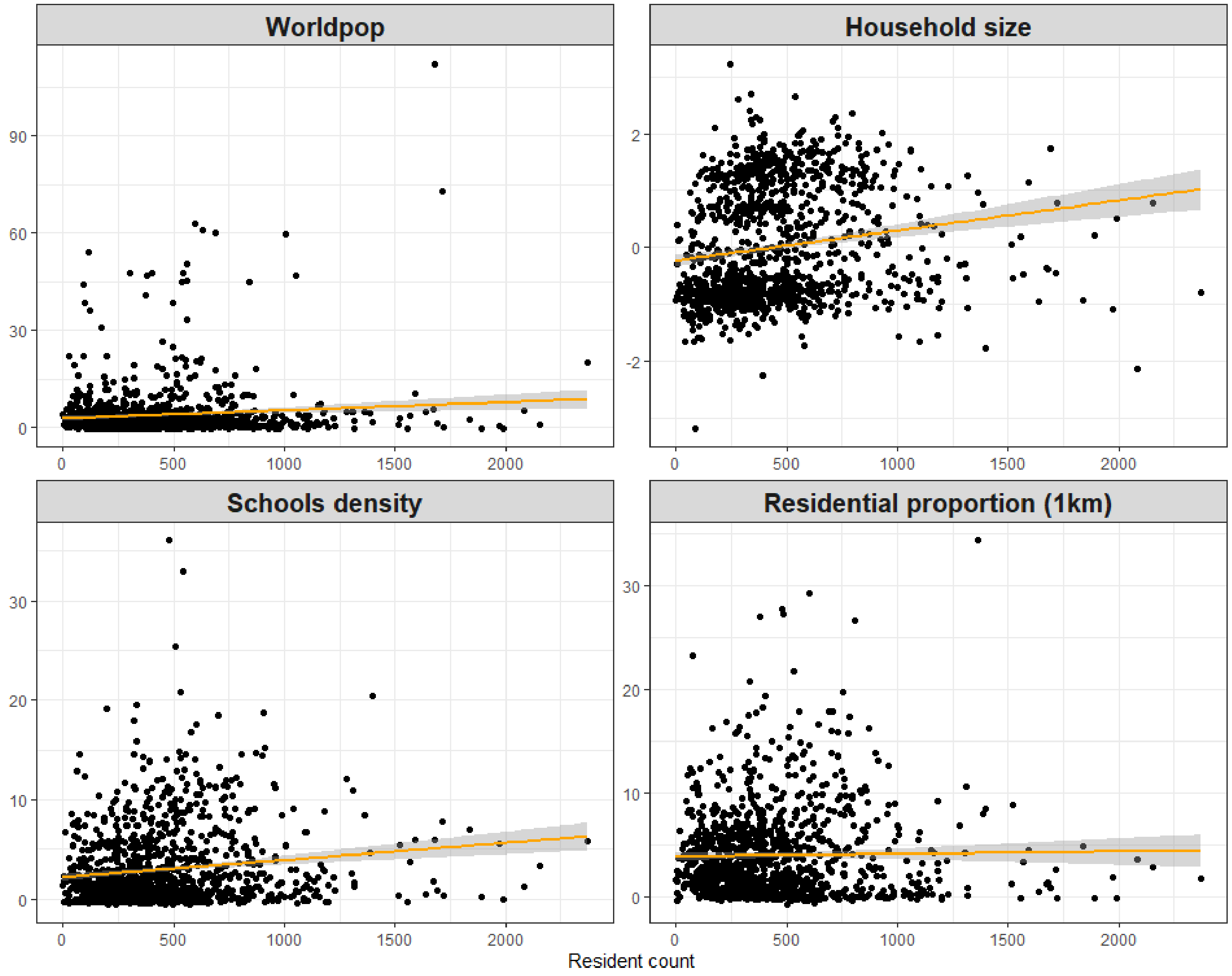


$$N_i \sim \text{Poisson}(D_i \times A_i)$$

Model structure

Predictors

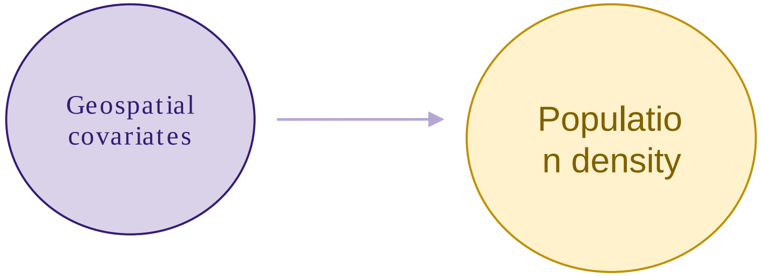


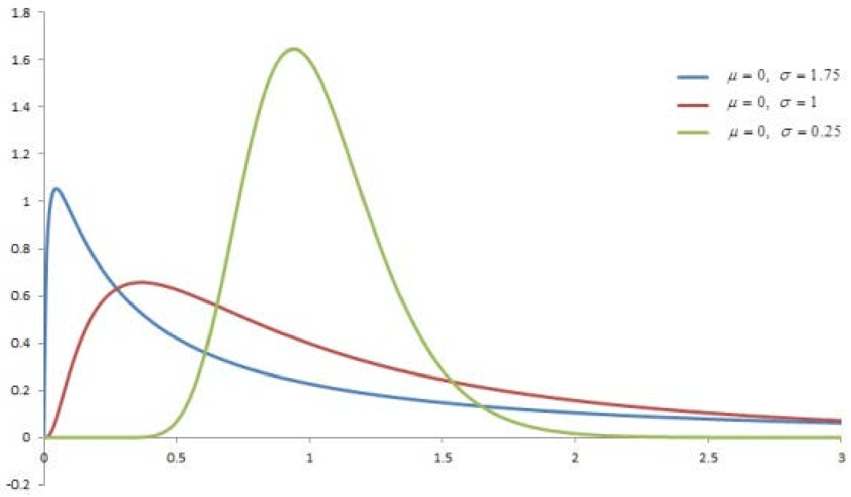


Model structure

Lognormal distribution for population density

- Density population:
  - Continuous, positif
    => lognormal distribution
- Parameters:
  - Mean
  - Variance



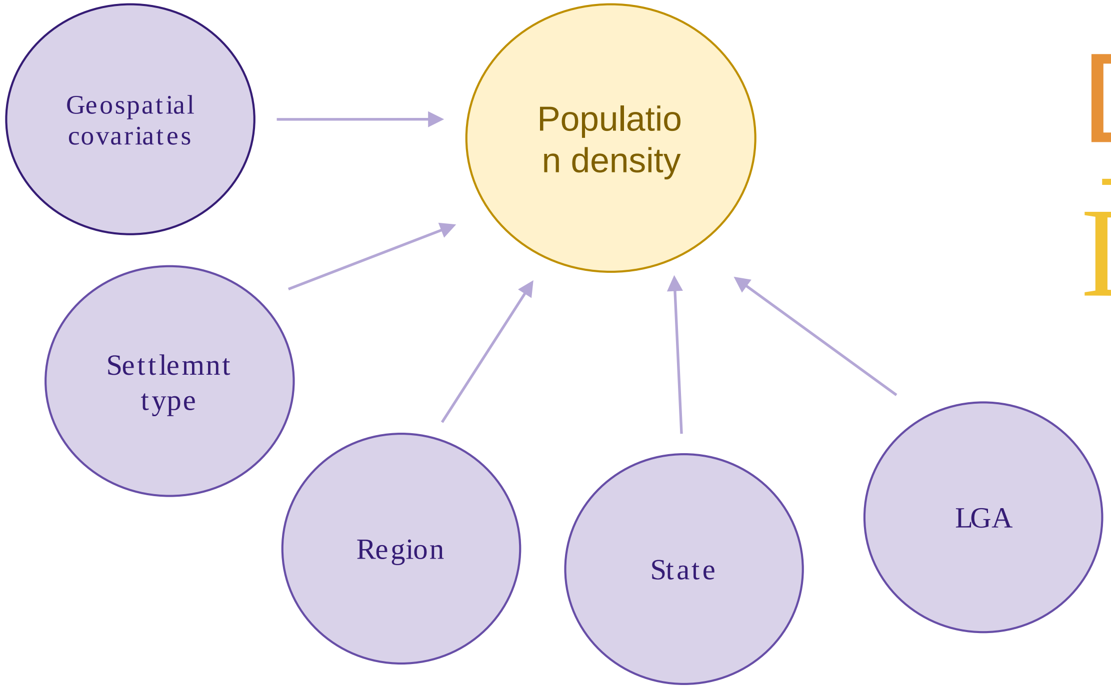


$$D_i \sim logNormal(\bar{D_i}, \sigma)$$

$$\bar{D_i} = \alpha + \beta X_i$$

Model structure

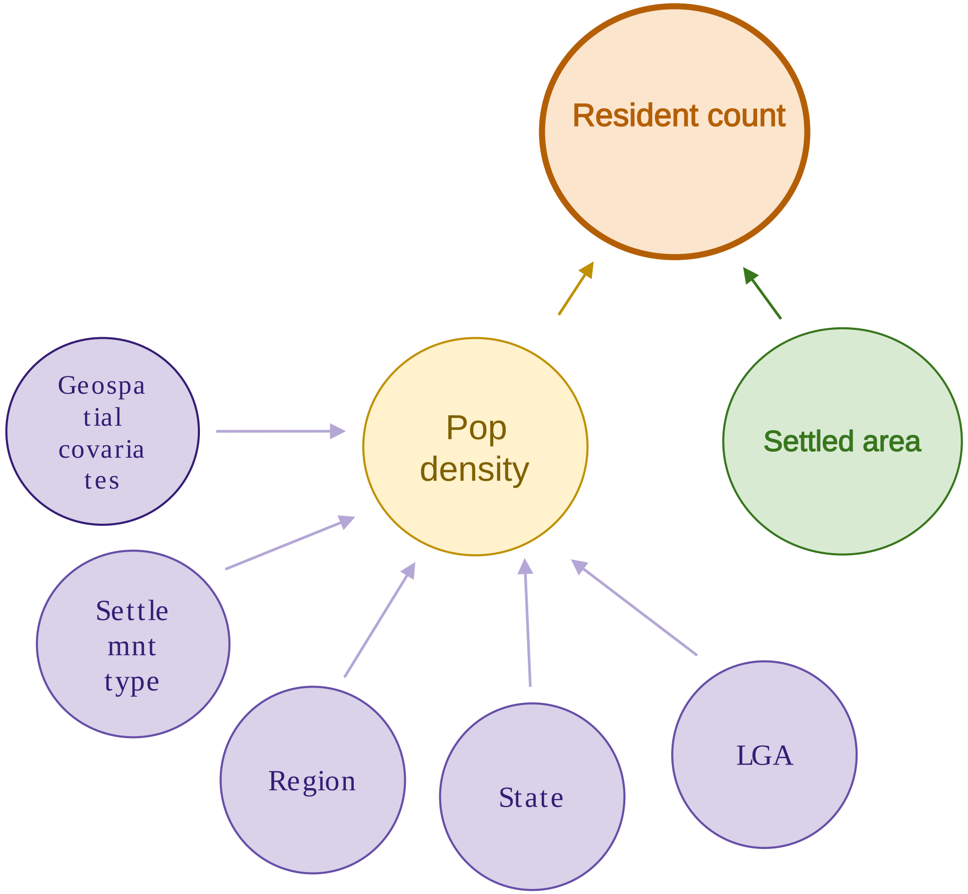
Lognormal distribution for population density



$$Di \sim logNormal(\overline{Di}, \sigma_{t,r,s,l})$$

$$\overline{Di} = \alpha_{t,r,s,l} + \beta\ Xi$$

Model structure



$$N_i \sim \text{Poisson}(D_i \text{ x } A_i)$$

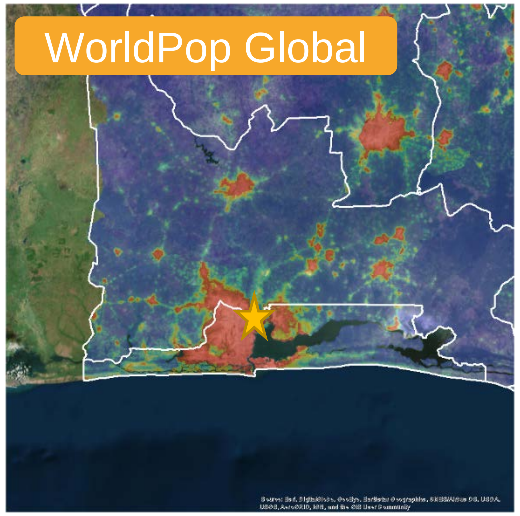
$$D_i \sim \text{logNormal}(\bar{D}_i, \sigma_{t,r,s,l})$$

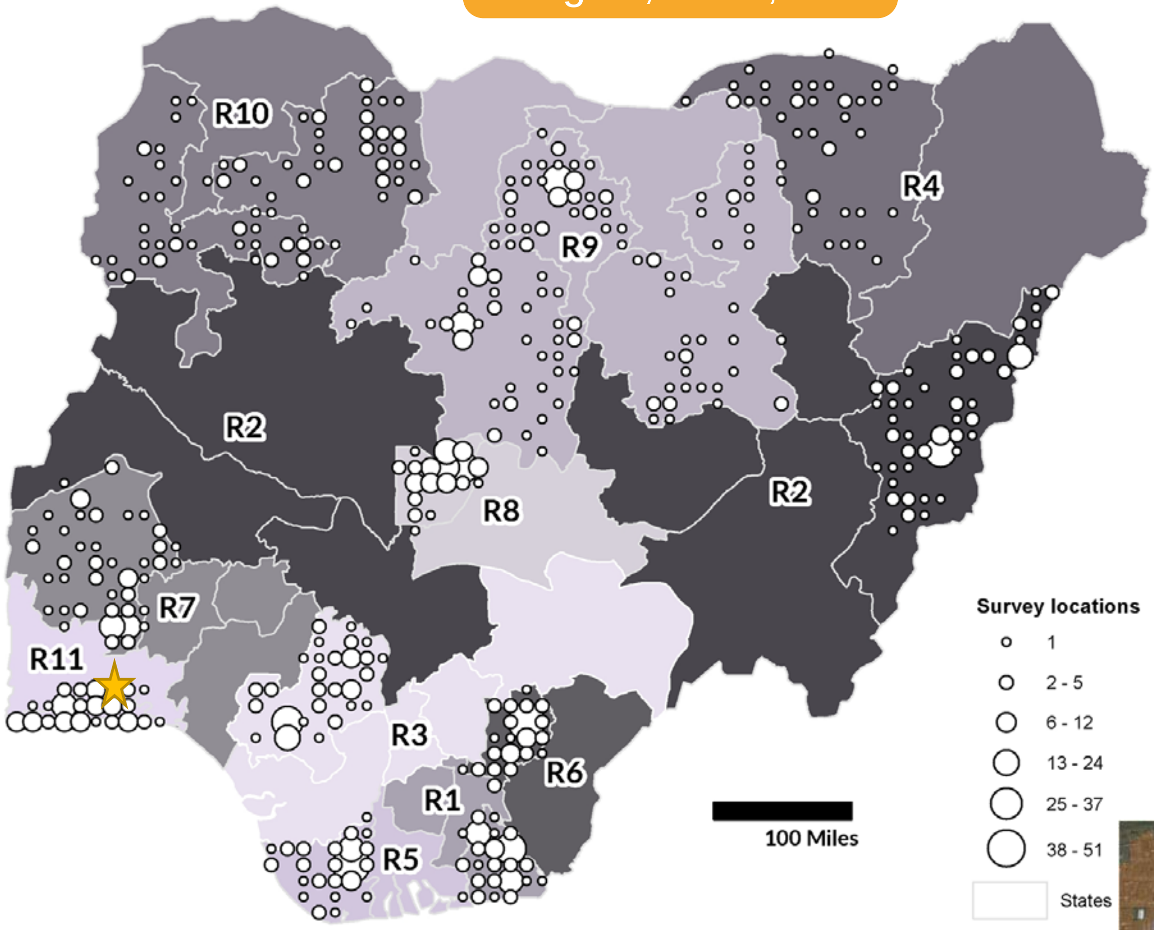
$$\bar{D}_i = \alpha_{t,r,s,l} + \beta X_i$$

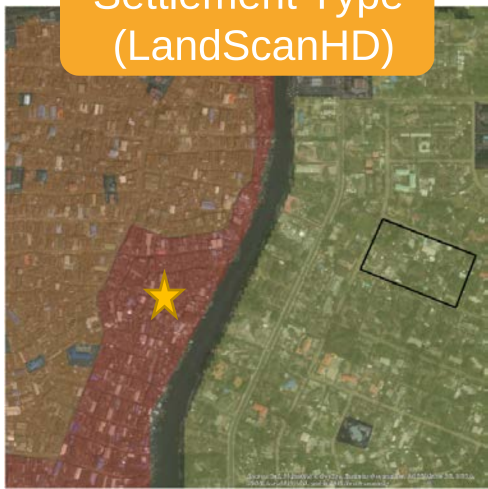
Model structure

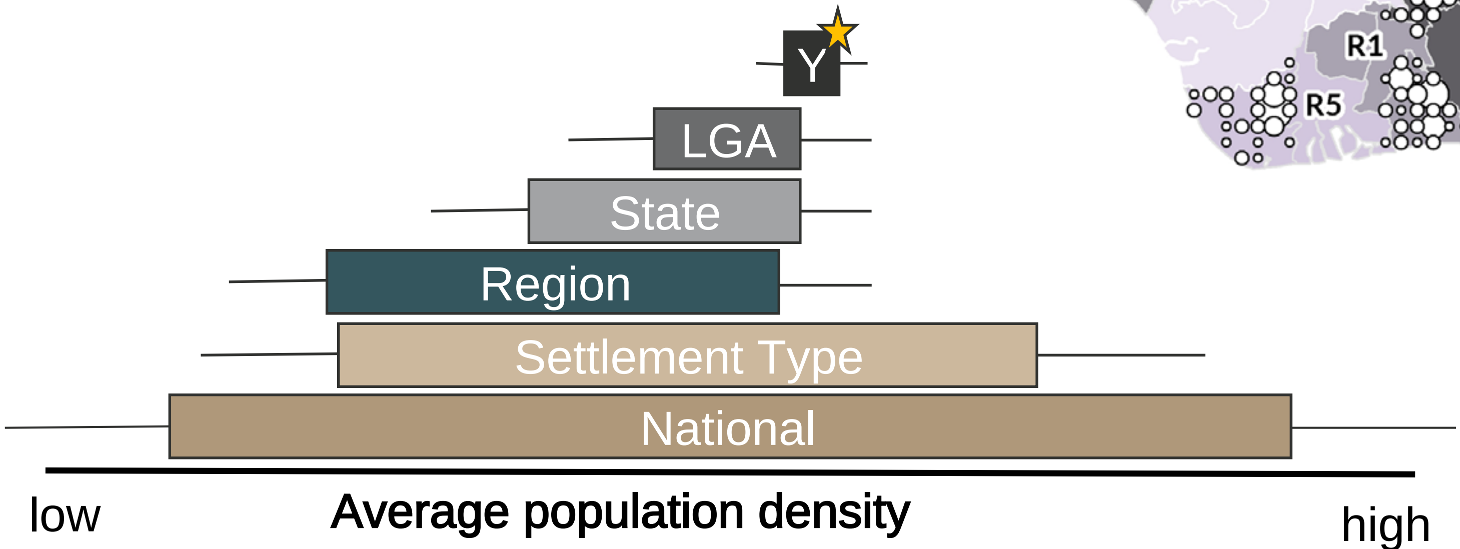
Recap


School

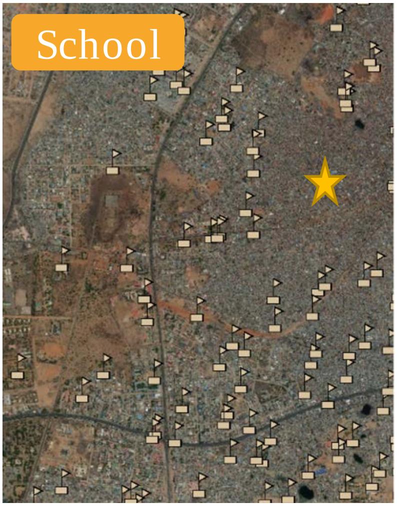

WorldPop Global

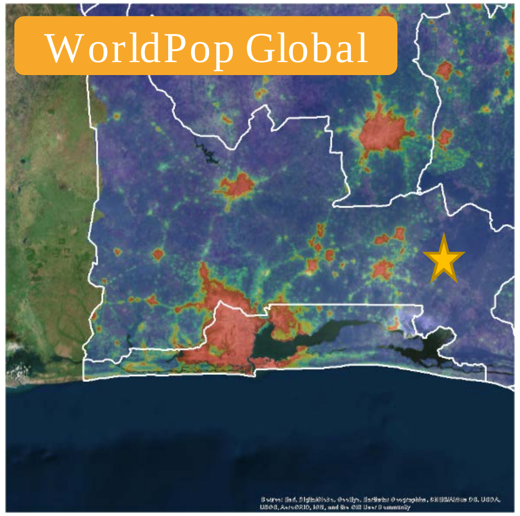

Admin boundaries:
Region, State, LGA
R10
R9
R4
R2
R2
R8
R7
R11
R3
R6
R1
R5
Survey locations
1
2 - 5
6 - 12
13 - 24
25 - 37
38 - 51
States
100 Miles

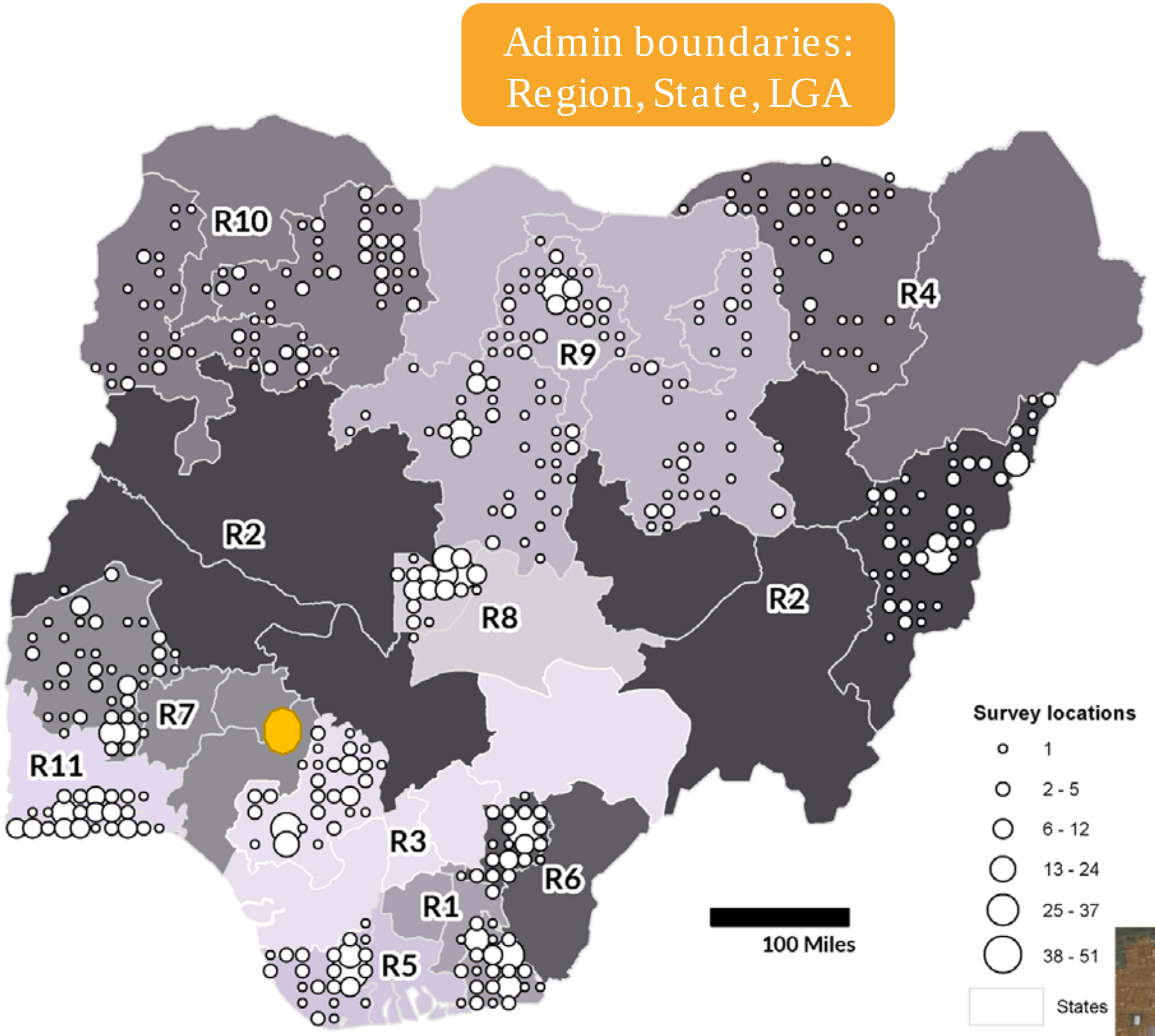

Settlement Type
(LandScanHD)

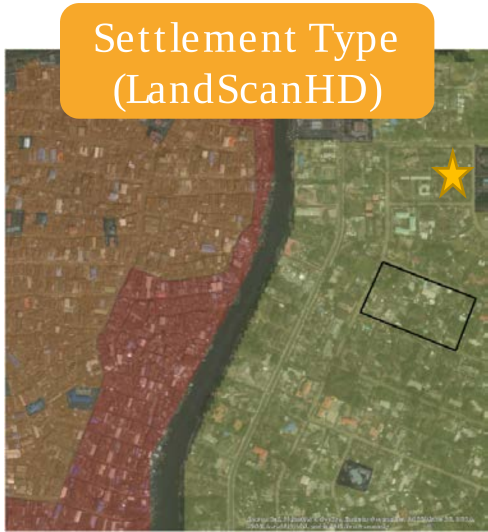

Y
LGA
State
Region
Settlement Type
National
low
Average population density
high

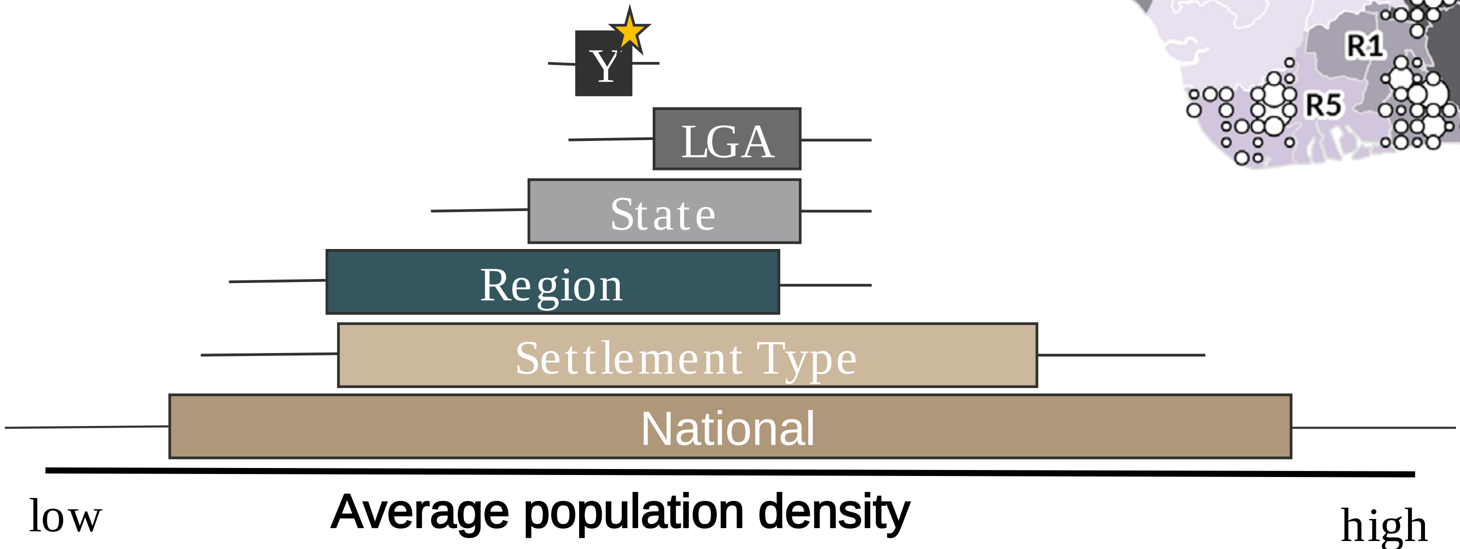
Recap


School


WorldPop Global


Admin boundaries:
Region, State, LGA
R10
R9
R4
R2
R2
R8
R7
R11
R3
R6
R1
R5
Survey locations
1
2 - 5
6 - 12
13 - 24
25 - 37
38 - 51
States
100 Miles


Settlement Type
(LandScanHD)


Y
LGA
State
Region
Settlement Type
National
low
Average population density
high

Recap

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

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

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

Data

N = Population size (count)
A = area of settlement (hectares)
x = covariates

Parameters

D = population density (people/hectare)
α = average population density
β = effect of covariate
σ = residual variance in population density

Subscripts

i = location
k = covariate index
t = settlement type
r = region
s = state
l = local government area

Recap

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

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

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

Data

N = Population size (count)
A = area of settlement (hectares)
x = covariates

Parameters

D = population density (people/hectare)
α = average population density
β = effect of covariate
σ = residual variance in population density

Subscripts

i = location
k = covariate index
t = settlement type
r = region
s = state
l = local government area

Model Fit

Technical Note

Programming language

- R, JAGS
- **Packages:** rjags, runjags

Characteristics

- 166’412’498 pixels
- High performance computing cluster
- 24 hearts, 40 CPUs, 64GB RAM
- Executing time: 9 hours

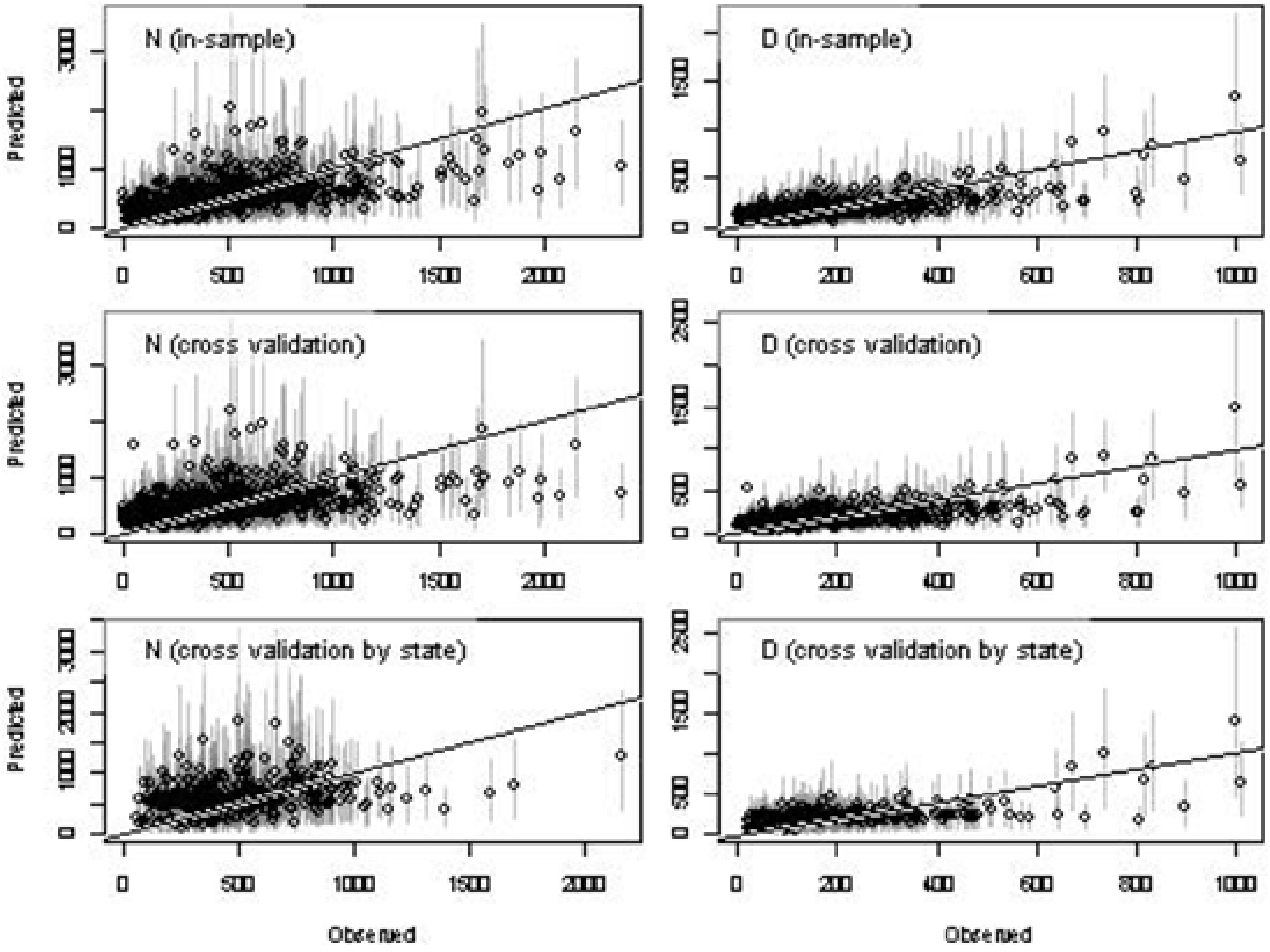
```
model {
  #likelihood function
  for (t in 1:nTrials) { #for each trial
    y[t] ~ dwiener(alpha[WC[t], subject[t]], tau[WC[t], subject[t]],
    ↳ beta[WC[t], subject[t]], delta[WC[t], subject[t]])
  }
  for (s in 1:nSubjects) { #for each subject
    for (w in 1:nWithin) { #for each within-subjects condition
      alpha[w, s] ~ dnorm(muAlpha[w, BC[s]], precAlpha[BC[s]]) T(.1, 5)
      beta[w, s] ~ dnorm(muBeta[w, BC[s]], precBeta[BC[s]]) T(.1, .9)
      tau[w, s] ~ dnorm(muTau[w, BC[s]], precTau[BC[s]]) T(.0001, 1)
      delta[w, s] ~ dnorm(muDelta[w, BC[s]], precDelta[BC[s]]) T(-5, 5)
    }
  }
  #priors
  for (b in 1:nBetween) { #for each between-subjects condition
    precAlpha[b] ~ dgamma(.001, .001)
    precBeta[b] ~ dgamma(.001, .001)
    precTau[b] ~ dgamma(.001, .001)
    precDelta[b] ~ dgamma(.001, .001)
    for (w in 1:nWithin) { for each within-subjects condition
      muAlpha[w, b]  ~ dunif(.1, 5)
      muBeta[w, b] ~ dunif(.1, .9)
      muTau[w, b] ~ dunif(.0001, 1)
      muDelta[w, b] ~ dunif(-5, 5)
    }
  }
}
```

JAGS code

Goodness-of-fit

Two 10-fold Cross-validation:

- Random across samples
- By states



Definition

- **Based on the comparison between predicted and observed** population density for each microcensus:
  - **Bias** = mean of residuals
  - **Imprecision** = standard deviation of residuals
  - **Inaccuracy** = mean of absolute residuals
  - **R2** = correlation between predicted and observed values.

Parameter	Prediction	Bias	Imprecision	Inaccuracy	R2
D_i	In-sample	7	86	61	0.57
D_i	X-val random	8	96	67	0.46
D_i	X-val state	24	121	92	0.40

Goodness-of-fit metrics

Model prediction

Example

1. Data

State | Lagos

LGA | Amuwo Odofin
Settlement | Urban
Settled Area | 0,86 hectares (100% pixel)
Worldpop | 23 p
Schools | 2.84
Household size |1,29
Residential| 1.58

2. Model

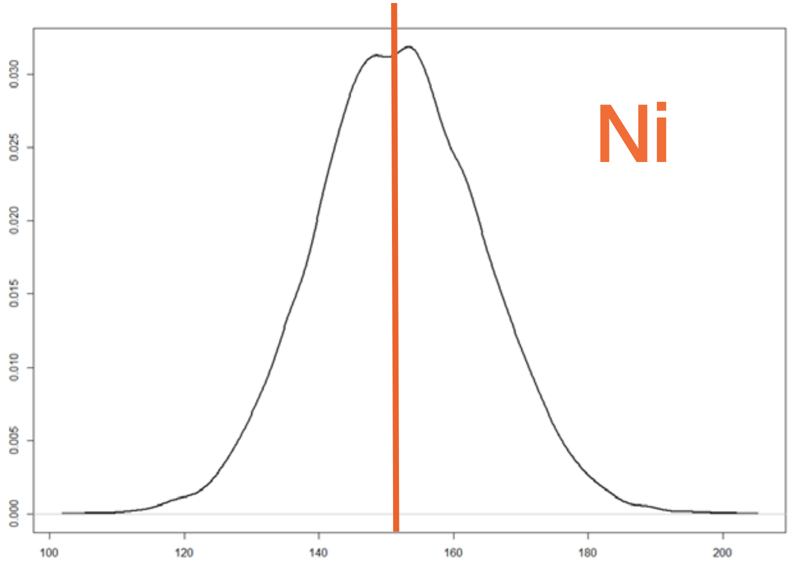
$$N_i \sim Poisson(D_i * 0{,}86)$$

$$D_i \sim logNormal(D_i, 0{,}44)$$

$$D_i = 4{,}3 + 0{,}02 * 23 + 0{,}1 * 2{,}84 + 0.05 * 1{,}29 + 0{,}03 * 1{,}58$$

3. Population estimates

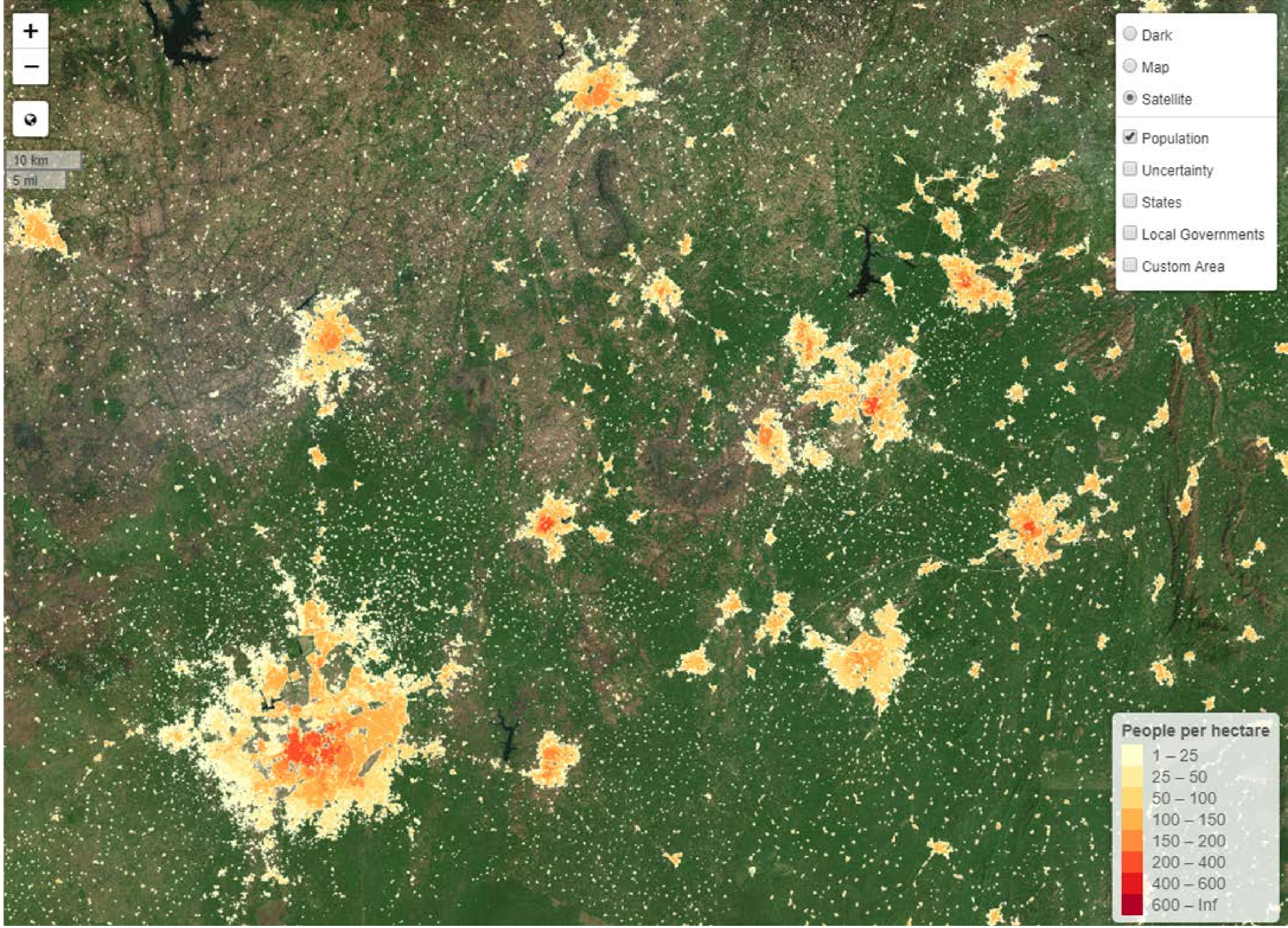
152 people



Mean

Prediction

Gridded population estimates



Output

Population totals

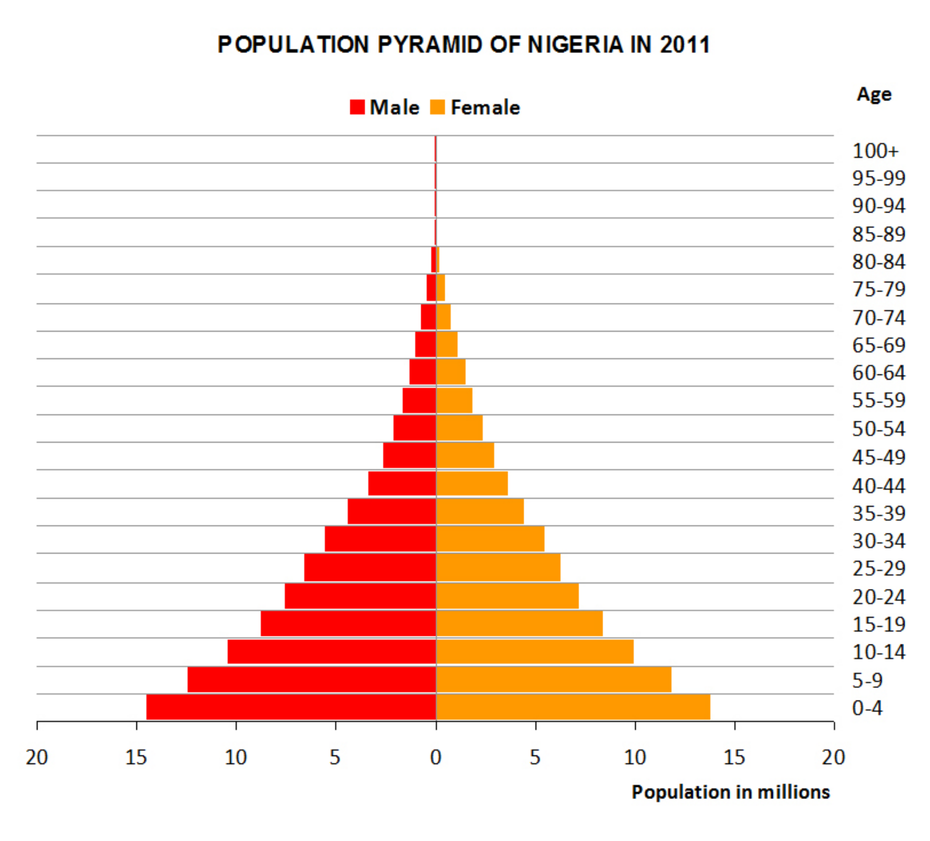
State	Local Government Area	Population	Lower	Upper
Abia	Aba North	248279	153445	400041
Abia	Aba South	382848	251014	608595
Abia	Arochukwu	144510	65579	269464
Abia	Bende	233694	106496	439371
Abia	Ikwuano	151755	71873	284523
Abia	Isiala-Ngwa North	228984	104579	432454
Abia	Isiala-Ngwa South	264367	120401	497555
Abia	Isuikwuato	183752	83106	345038
Abia	Obi Nwga	460609	263579	783035
Abia	Ohafia	198114	114920	334518
Abia	Osisioma Ngwa	558253	381089	841784
Abia	Ugwunagbo	85867	43894	155326
Abia	Ukwa East	66651	30096	127077
Abia	Ukwa West	105540	47965	201129
Abia	Umu-Nneochi	218104	99773	414361
Abia	Umuahia North	307034	210301	463606
Abia	Umuahia South	222744	123620	382898
Adamawa	Demsa	149411	95058	222419
Adamawa	Fufore	239790	185979	314177
Adamawa	Ganye	192641	141492	272963
Adamawa	Girei	181583	142185	228979

Output

What about the age-sex structure?

Interest for policy -making

- Public health policy
  - Vaccination
  - Maternal health
  - Reproductive and sexual health
- Structure investment
  - Ex: school
- M&E assessment
  - Ex: assessment of food security



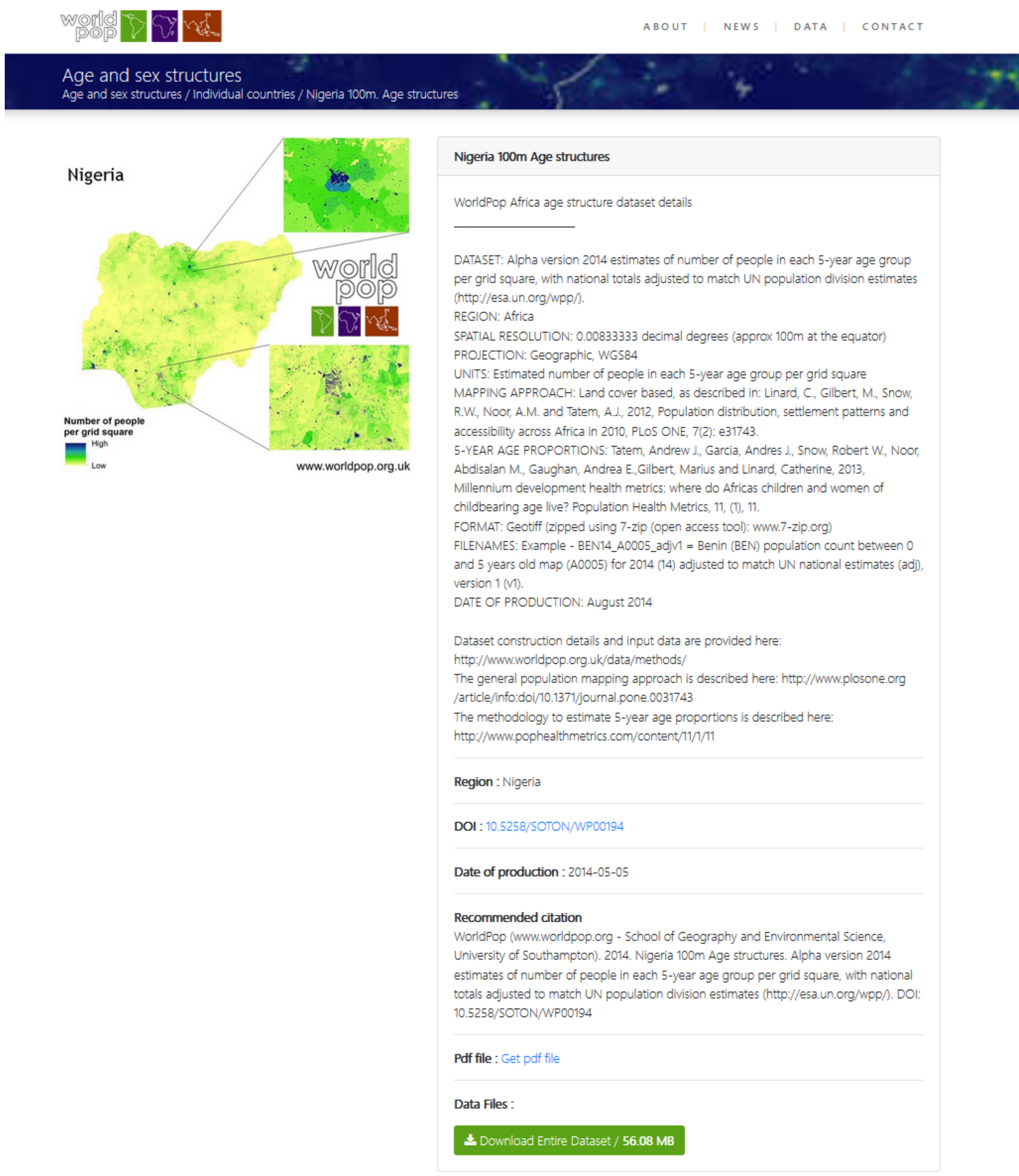
Age and Sex Decomposition

Input data

Age and sex structure, Worldpop 2014

- From DHS 2013 data projected
- Age classes:
  - 0-1 year
  - Then 5 years bucket
  - Above 80 years
- Specific by region
- Open-source:

https://www.worldpop.org/geodata/summary?id=1259



Age and sex structures

Age and sex structures / Individual countries / Nigeria 100m. Age structures

Nigeria 100m Age structures

WorldPop Africa age structure dataset details

DATASET: Alpha version 2014 estimates of number of people in each 5-year age group per grid square, with national totals adjusted to match UN population division estimates (http://esa.un.org/wpp/).
REGION: Africa
SPATIAL RESOLUTION: 0.00833333 decimal degrees (approx 100m at the equator)
PROJECTION: Geographic, WGS84
UNITS: Estimated number of people in each 5-year age group per grid square
MAPPING APPROACH: Land cover based, as described in: Linard, C., Gilbert, M., Snow, R.W., Noor, A.M. and Tatem, A.J., 2012, Population distribution, settlement patterns and accessibility across Africa in 2010, PLoS ONE, 7(2): e31743.
5-YEAR AGE PROPORTIONS: Tatem, Andrew J., Garcia, Andres J., Snow, Robert W., Noor, Abdisalan M., Gaughan, Andrea E.,Gilbert, Marius and Linard, Catherine, 2013, Millennium development health metrics: where do Africas children and women of childbearing age live? Population Health Metrics, 11, (1), 11.
FORMAT: Geotiff (zipped using 7-zip (open access tool): www.7-zip.org)
FILENAMES: Example - BEN14_A0005_adjv1 = Benin (BEN) population count between 0 and 5 years old map (A0005) for 2014 (14) adjusted to match UN national estimates (adj), version 1 (v1).
DATE OF PRODUCTION: August 2014

Dataset construction details and input data are provided here: http://www.worldpop.org.uk/data/methods/
The general population mapping approach is described here: http://www.plosone.org/article/info:doi/10.1371/journal.pone.0031743
The methodology to estimate 5-year age proportions is described here: http://www.pophealthmetrics.com/content/11/1/11

Region : Nigeria

DOI : 10.5258/SOTON/WP00194

Date of production : 2014-05-05

Recommended citation
WorldPop (www.worldpop.org - School of Geography and Environmental Science, University of Southampton). 2014. Nigeria 100m Age structures. Alpha version 2014 estimates of number of people in each 5-year age group per grid square, with national totals adjusted to match UN population division estimates (http://esa.un.org/wpp/). DOI: 10.5258/SOTON/WP00194

Pdf file : Get pdf file

Data Files :

Download Entire Dataset / **56.08 MB**

Age and Sex Decomposition

In the model

1. Data

State | Lagos

LGA | Amuwo Odofin
Settlement | Urban
Settled Area | 0,86 hectares (100% pixel)
Worldpop | 23 p
Schools | 2.84
Household size | 1,29
Residential | 1.58

2. Model

$$N_i \sim Poisson(D_i * 0{,}86)$$

$$D_i \sim logNormal(D_i, 0{,}44)$$

$$D_i = 4{,}3 + 0{,}02 * 23 + 0{,}1 * 2{,}84 + 0.05 * 1{,}29 + 0{,}03 * 1{,}58$$

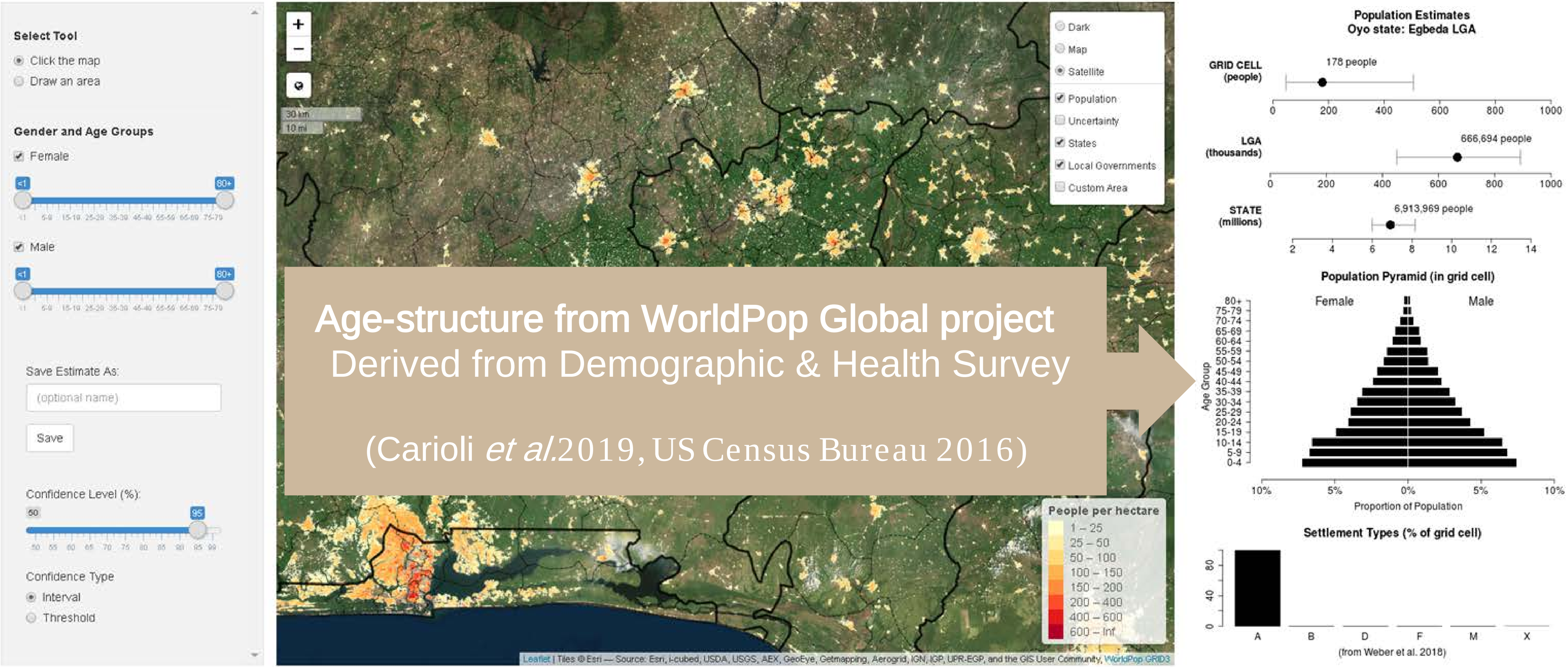
3. Population estimates

152 people

4. Demographic decomposition

female 0-1: 0.01 x 152
female 1-5: 0.06 x 152
...
female >80: 0.002 x 152
male 0-1: 0.01 x 152
male 1-5: 0.05 x 152
...
male >80: 0.002 x 152

Age and Sex Decomposition

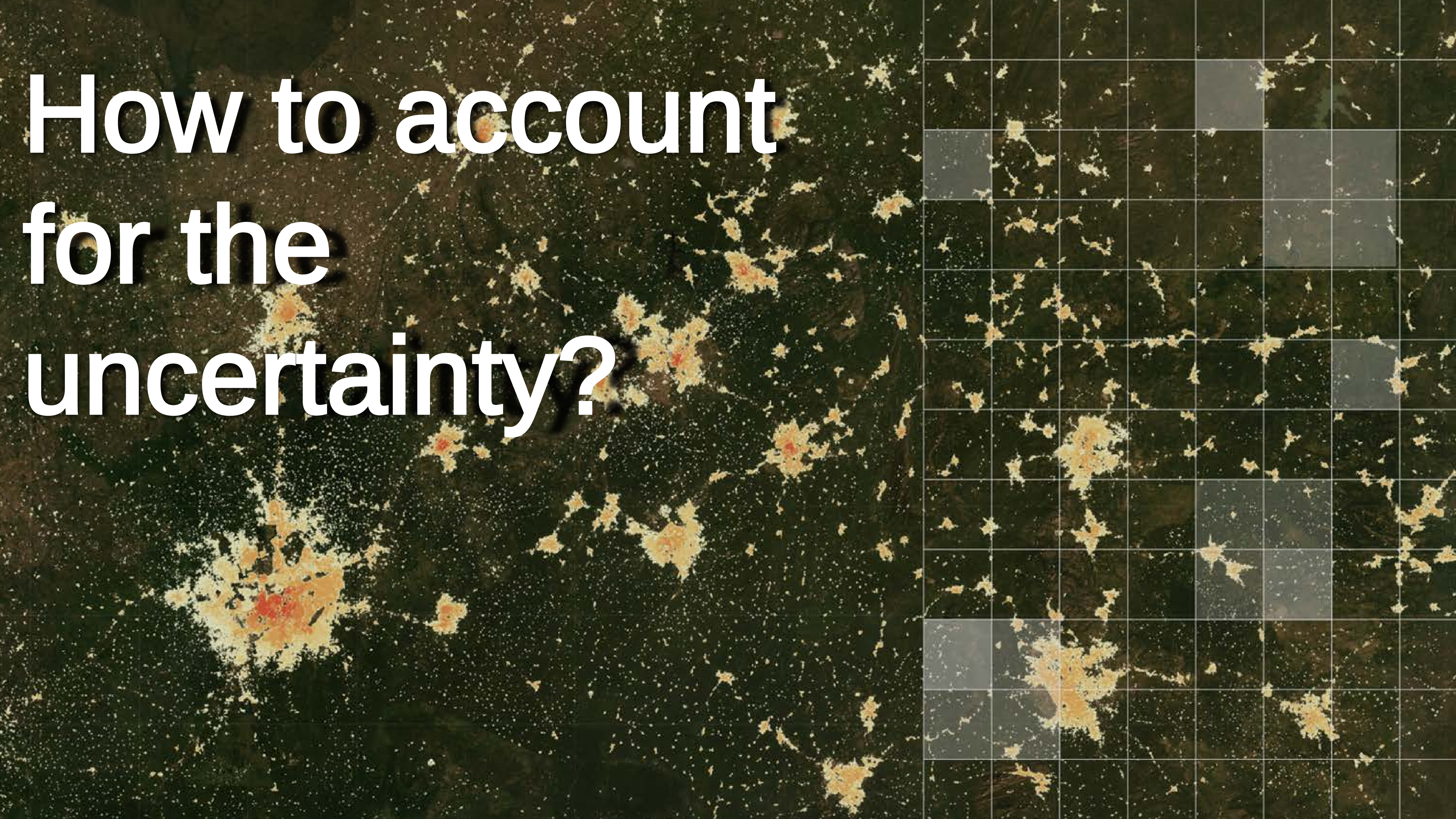


Age-structure from WorldPop Global project

Derived from Demographic & Health Survey

(Carioli *et al.* 2019, US Census Bureau 2016)

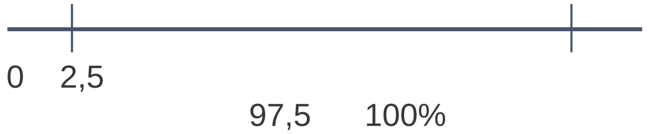
Age and Sex Decomposition


How to account for the uncertainty?

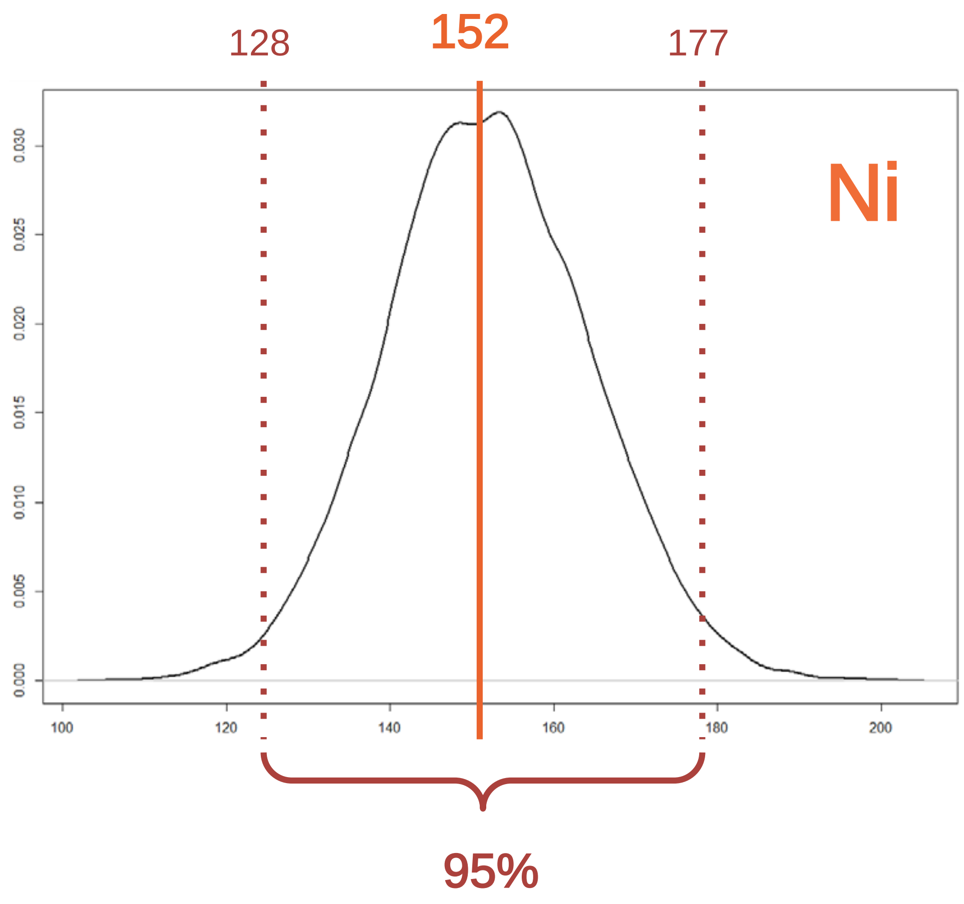
Computation

- Choose a **confidence level**: **95%**

- Compute corresponding quantiles

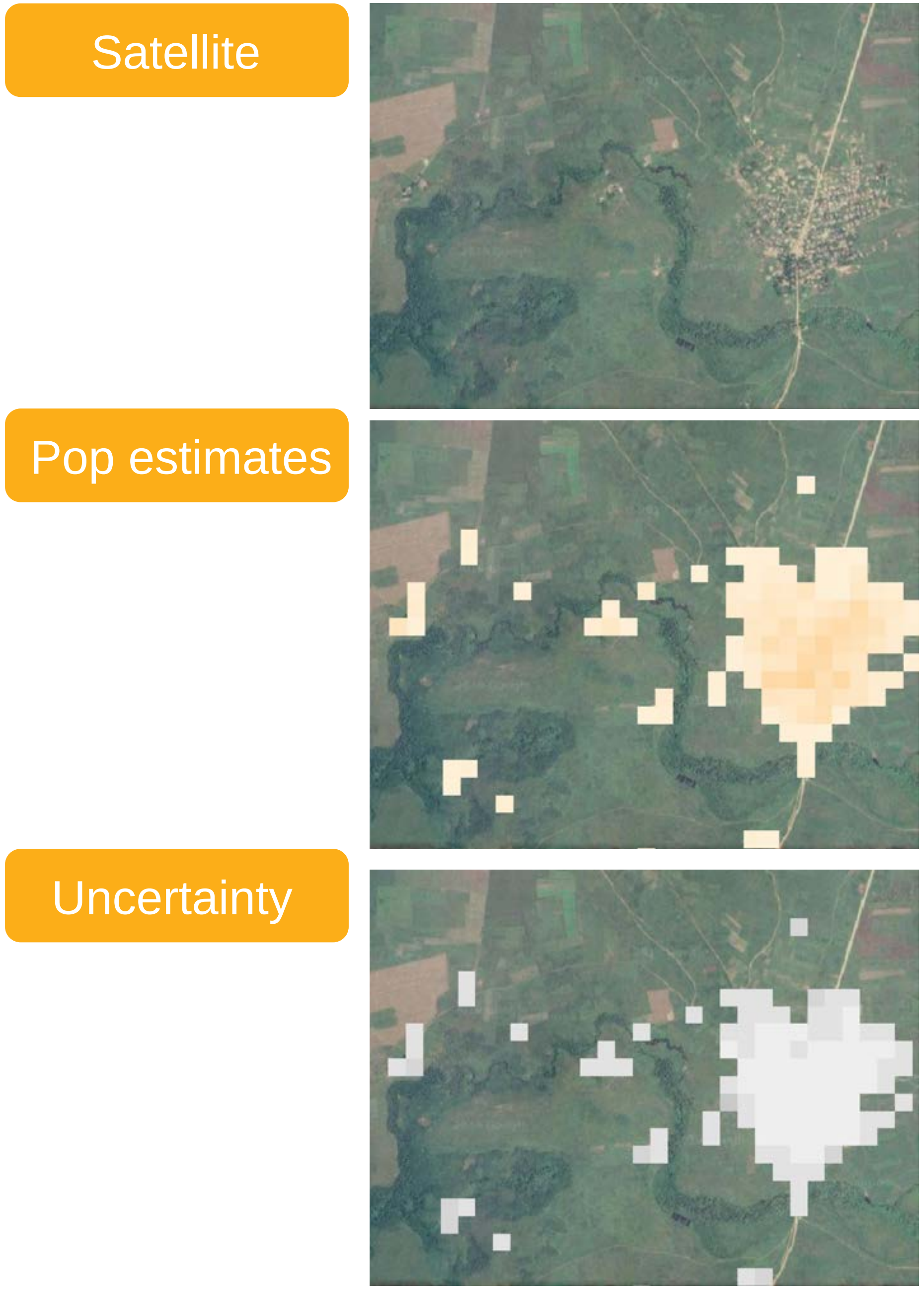


- **Uncertainty** = (sup- inf)/mean
  - (177-128) / 152 = 0.32



Output

- Estimates interval in table format
- Uncertainty raster



Areas for improvement

Known limitations

Assumes that no people live in areas mapped as unsettled or non-residential

Assumes that microcensus perfectly counts the number of people in each surveyed area

Year of population estimates:

2016-2017 microcensus determines population estimates in each grid cell

2013-2014 settlement map determines spatial distribution of population across Nigeria

Population estimates are missing from some areas near the Nigeria border where covariates were not available

Schools are not mapped consistently across the country and information about school capacity (# of students) was not available

Model updates

New building footprint layer from Maxar and Ecopia

New definition of hierarchy integrated feedbacks from the alpha review

New microcensus round:

- 2019 population data
- Samples in non-residential area
- Finer resolution
- Better monitoring of missing household

Covariates updates for region near the border

Keep tuned for model updates!

GRID3 NIGERIA

Break

5 min

GRID3 NIGERIA

Geospatial Analytics for Nigeria

Session 2- Part 2: Use and application of gridded population outputs

Michael Harper

BILL & MELINDA GATES foundation
UKaid from the British people
worldpop
FLOWMINDER.ORG
UNFPA
Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

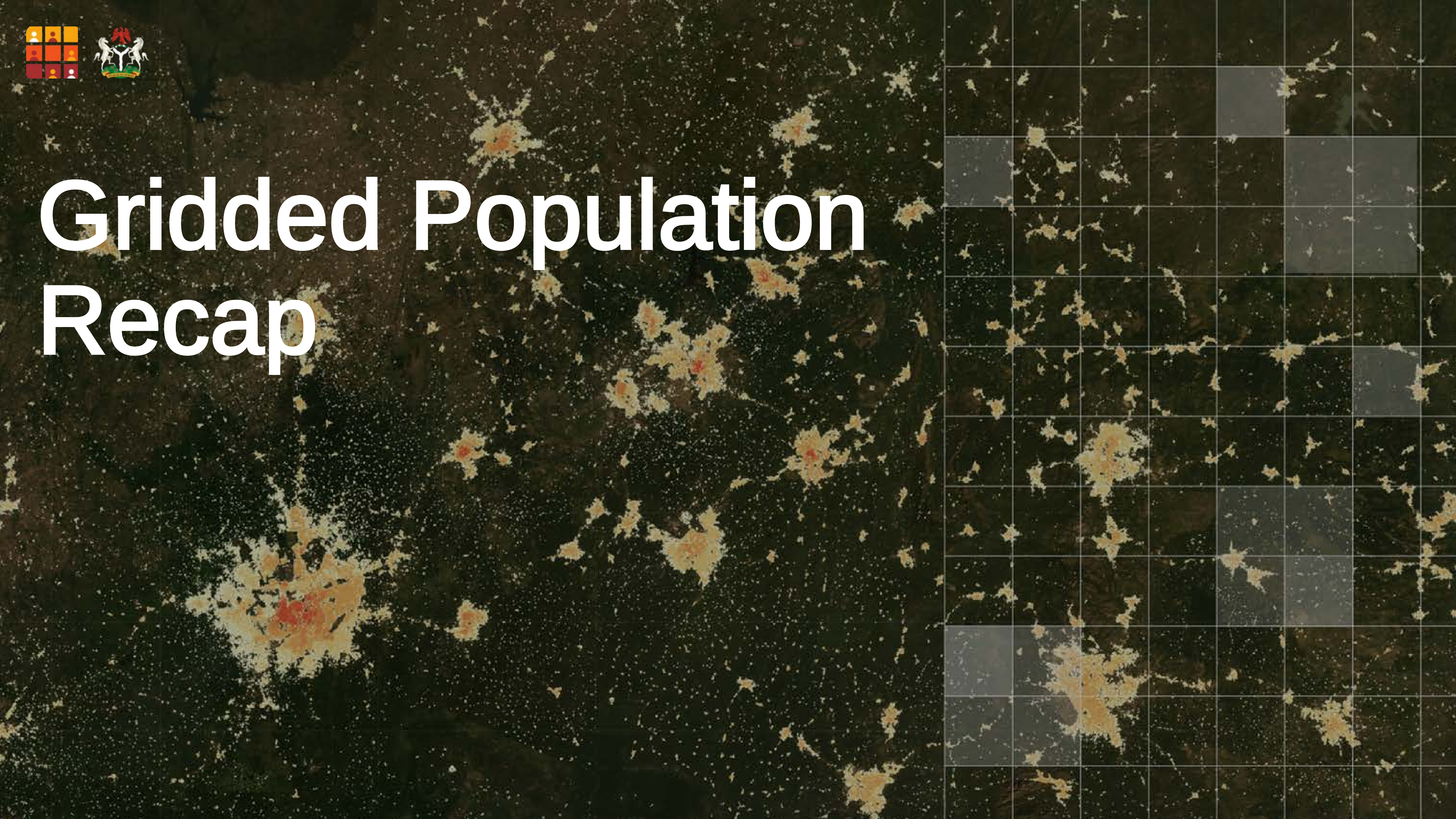
About Me

- Implementation Analyst
- PhD in spatial statistics with focus on renewable energy
- Interested in how we can **apply spatial data**

Learning Objectives

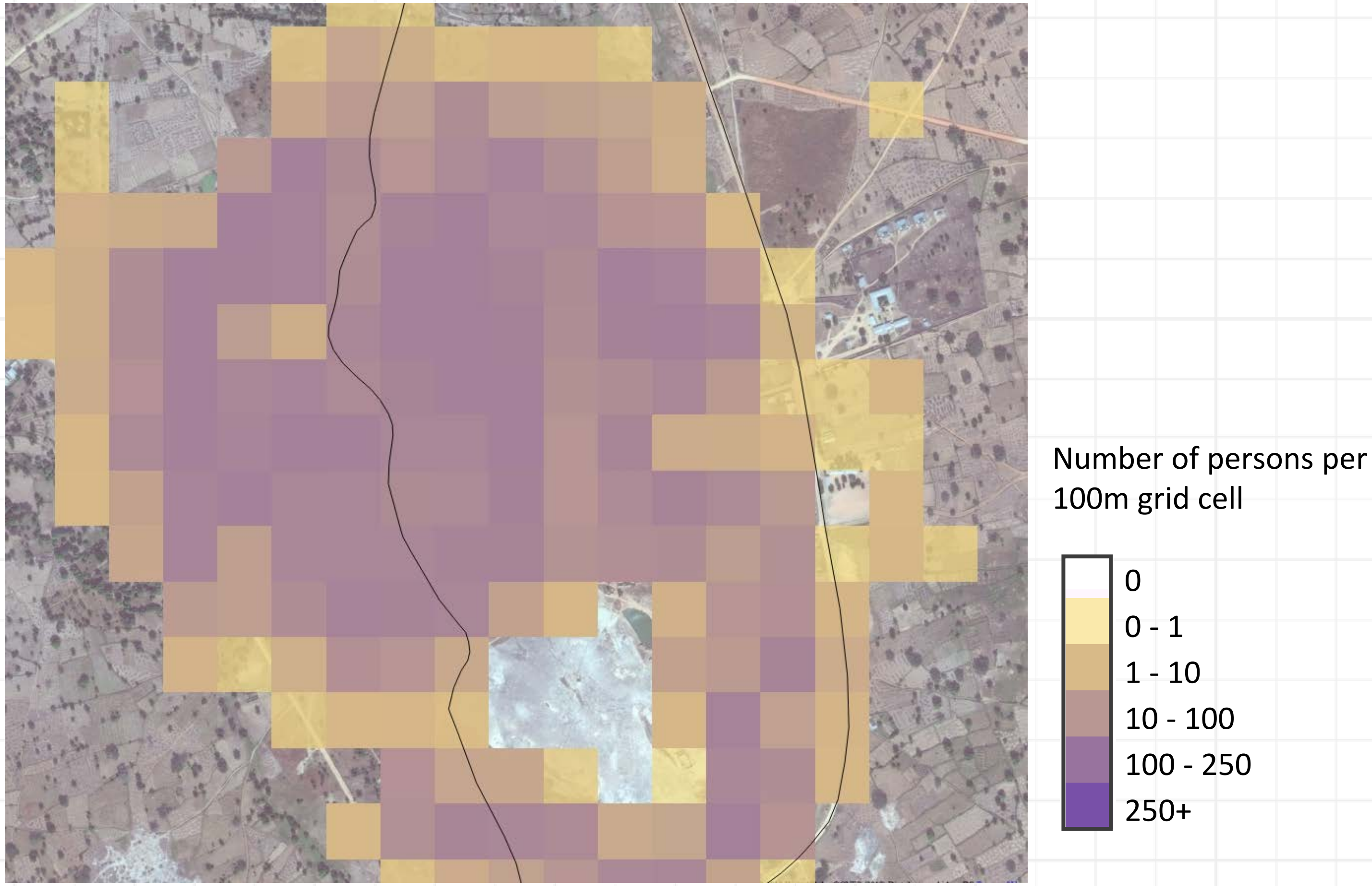
- **Recap** the value of gridded population datasets
- **Highlight** how population data can be used to support decision-making
- Explain essential **geospatial processing** methods for gridded population datasets

Gridded Population Recap



Gridded populations

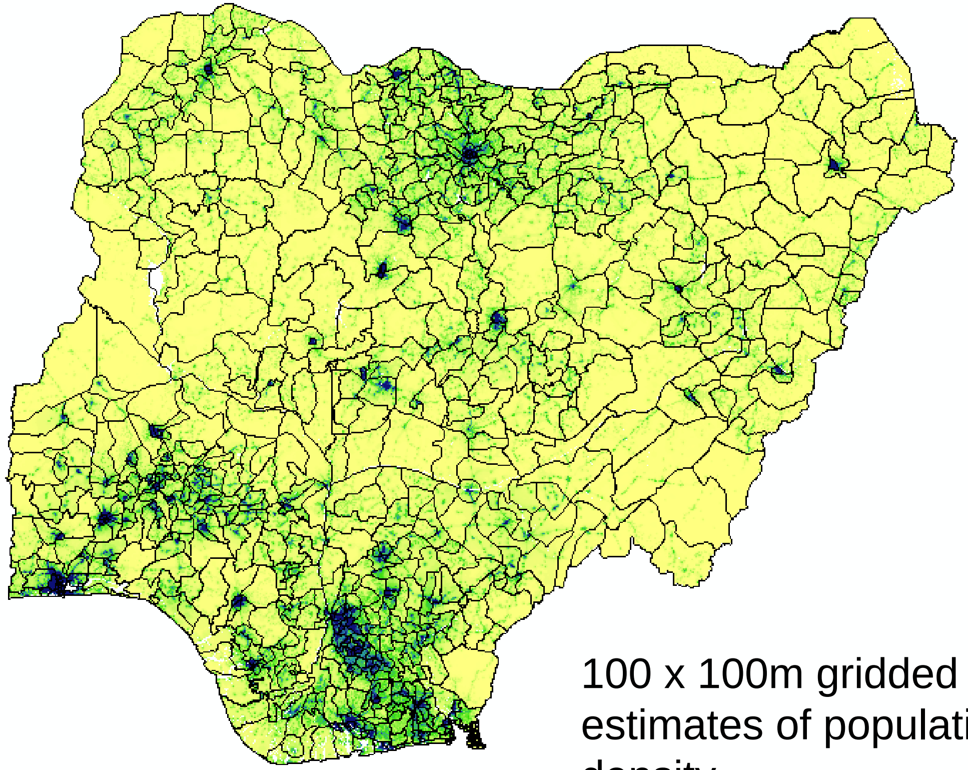
- **Gridded resolution**
  - Typically **100m or 1Km cells**
  - **Consistent** and **comparable**
  - Enable integrating different **ancillary data types**
- **Advantages**
  - **Flexible aggregation** at different administrative levels
  - Finer distributions **within enumeration boundaries**



GRID3 Nigeria Population Estimates v1.1, WorldPop and Flowminder, 22-02-2019.

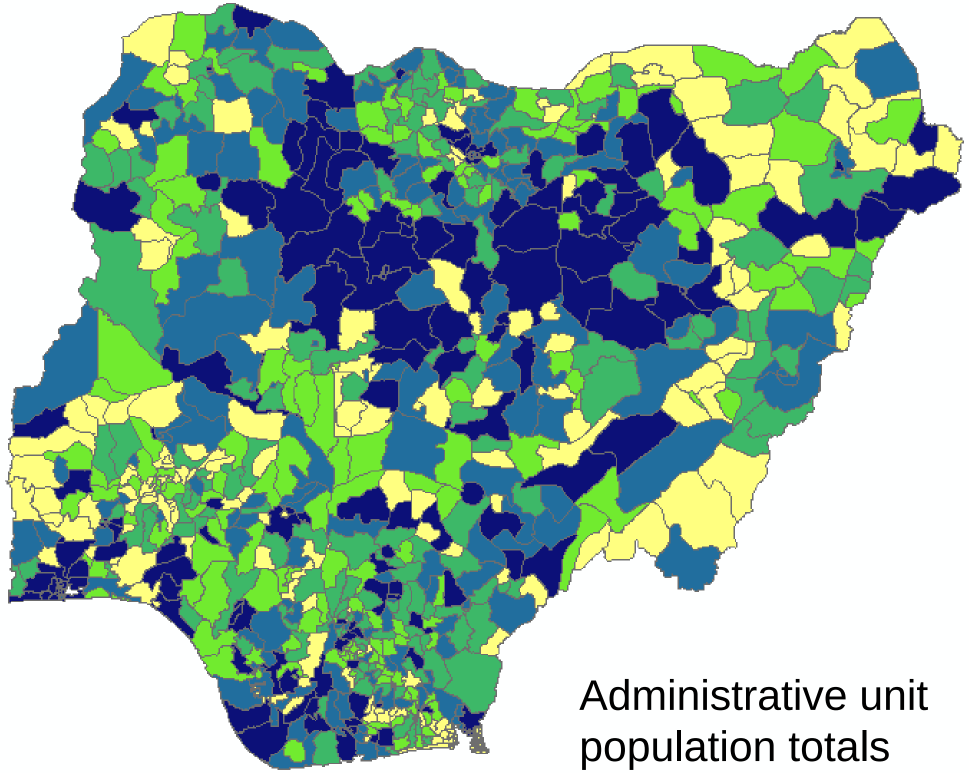
Thinking in grids

Grids: flexibility in analysis and data integration



100 x 100m gridded estimates of population density

Grids: flexibility in summarisation to any administrative unit level



Administrative unit population totals

GRID3 Population Estimates

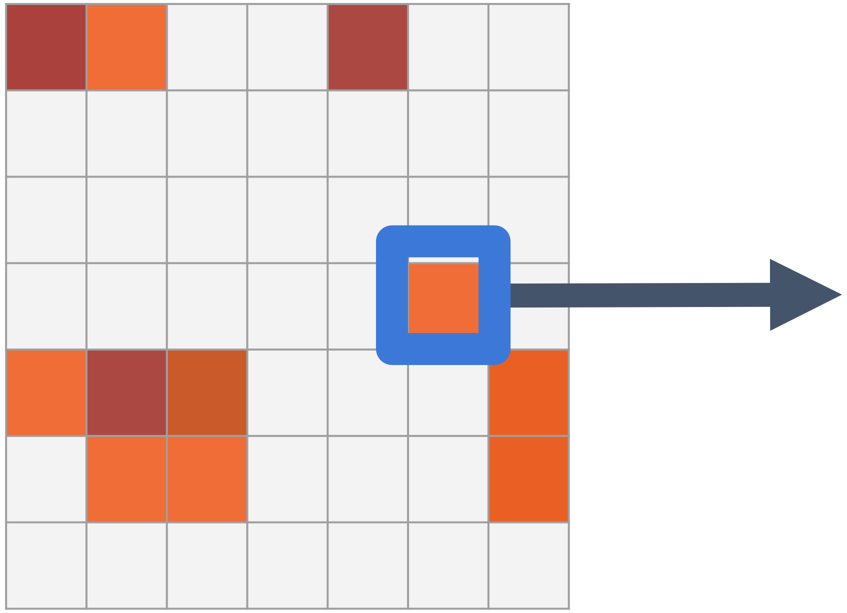
High Resolution

Measure of Uncertainty

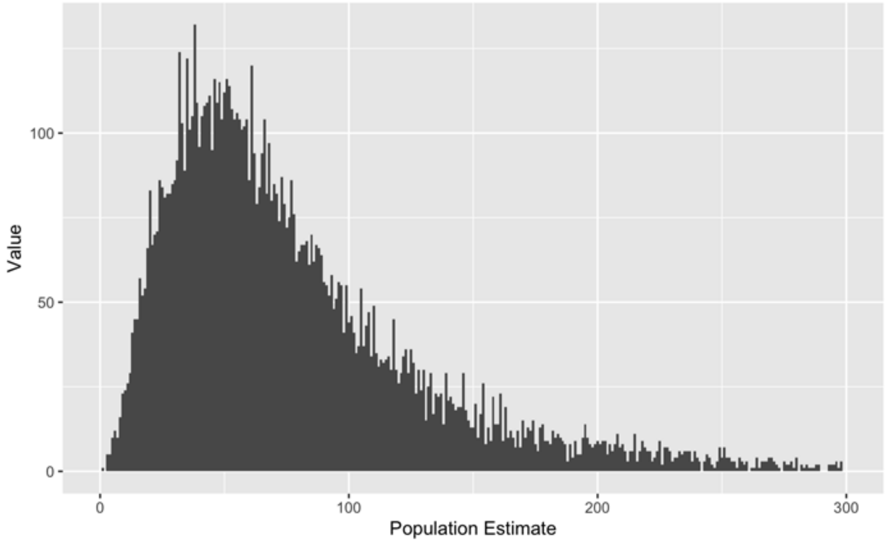
Age Sex Structures

Uncertainty at grid cell level

Population Raster



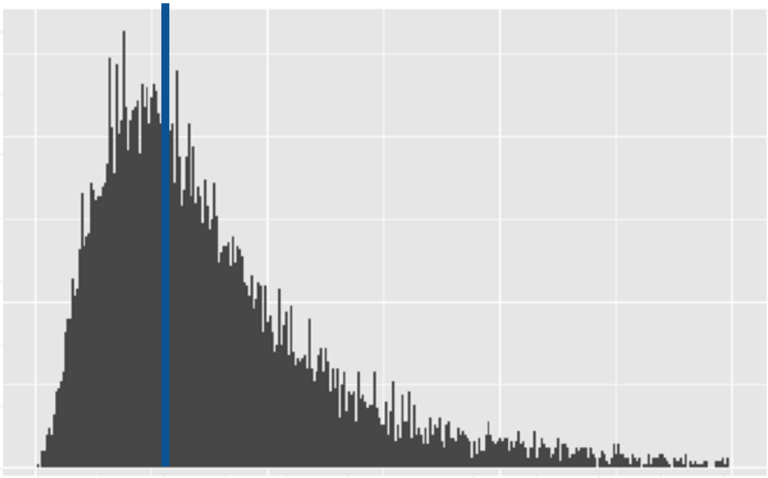
Model Distribution



Each cell has 1000s of modelled estimates

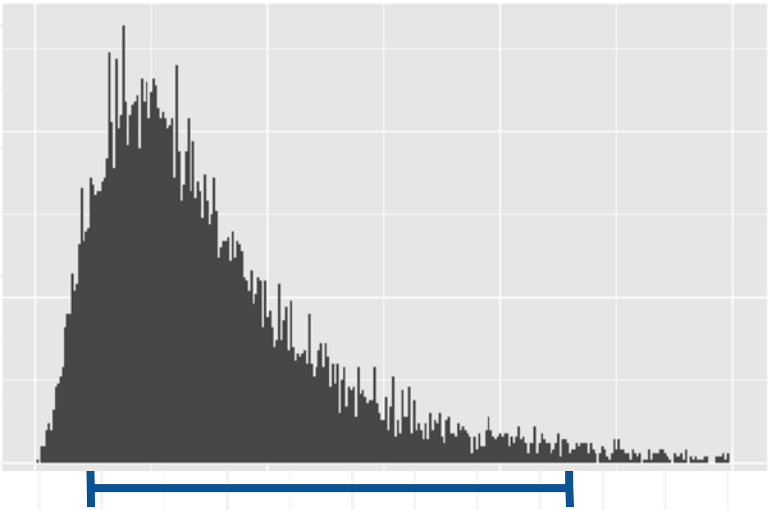
Ways of interpreting uncertainty

What is the mean value?



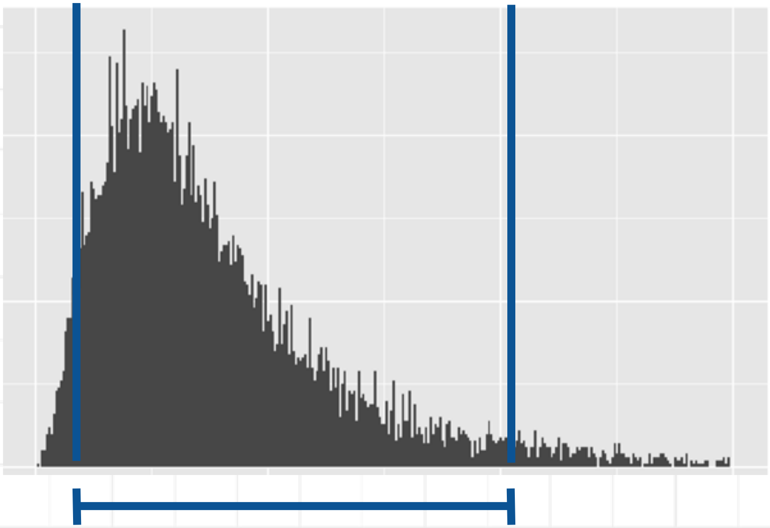
?

What is the confidence interval?



? ?

What is the chance of the population being between these values?



?
%

Getting value from population data

Insight from population data

Question

Other Datasets

Population Data

- Geoprocessing
- Spatial analysis
- Machine learning
- AI

Insights

Types of Questions

- *How many people are there in an area?*
- *Do we have enough ____ in an area?*
- *How many people in need to travel more than ___km to access ________?*
- *Where should we build ______?*
- What would be the cheapest way to provide ______ to the population?
- Where should we send out copies of the AI textbook?

Other datasets include

Spatial Data

- Infrastructure locations
- administrative boundaries
- Facility catchment areas
- Landscape features
- Elevation data

Economic Data

- Cost of materials/supplies
- Transport costs
- Land value

Tabular

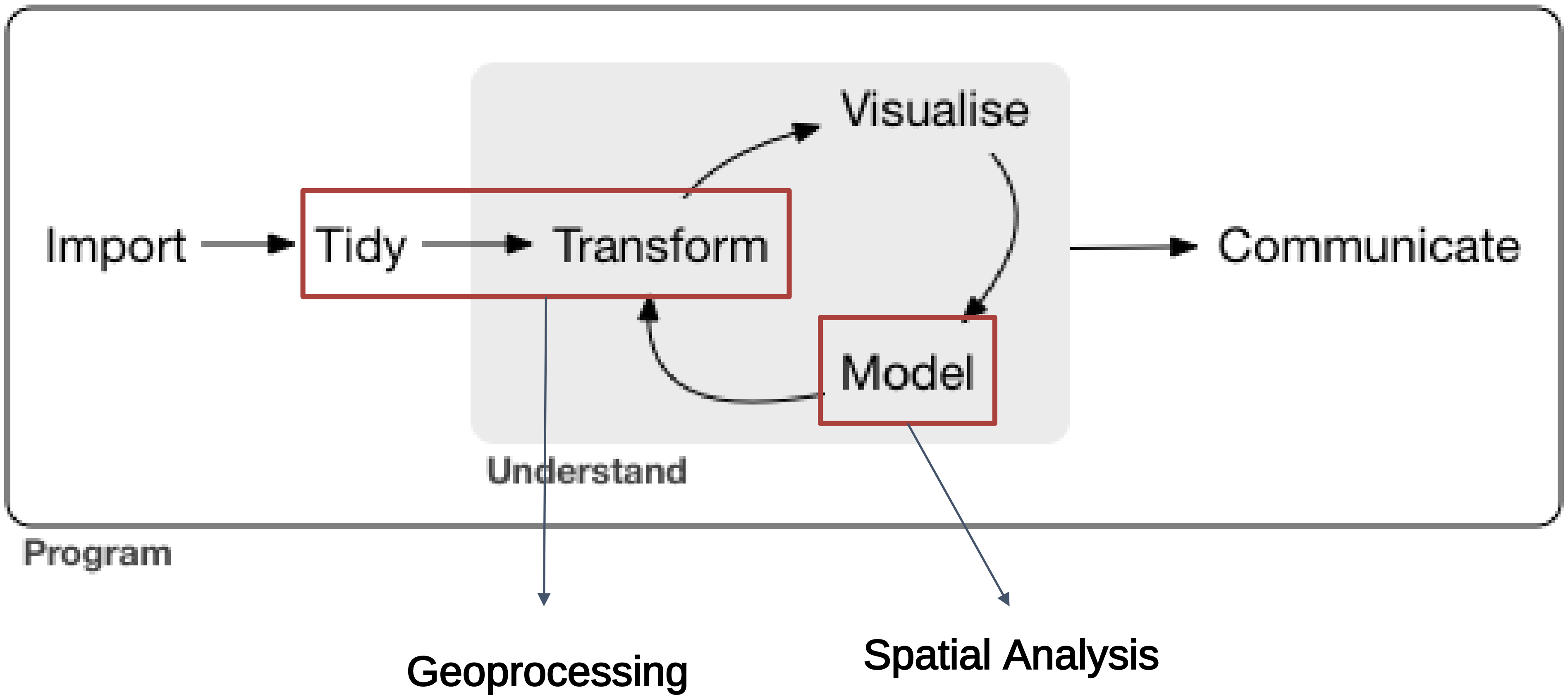
- Government statistics
- Health records
- School attendance figures

Qualitative Information

- Local knowledge
- Newspaper reports

Geoprocessing and spatial analysis

Typical data analysis workflow



R for Data Science, https://r4ds.had.co.nz/

Types of Methods

Geoprocessing

A GIS operation used to **manipulate GIS data**.

Input Data → Geoprocessing → Output Data

Geoprocessing allows for definition, management, and analysis of information used to form decisions.

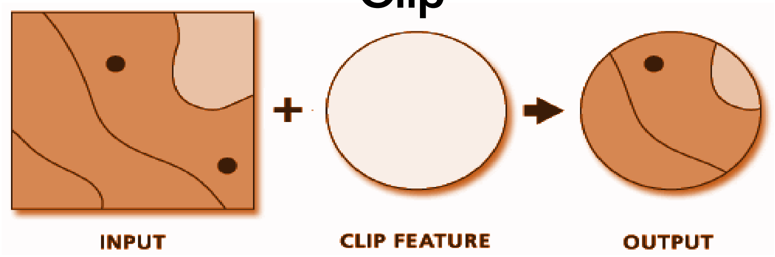
Spatial analytics

The process of **examining the locations**, attributes, and **relationships** of features in spatial data.

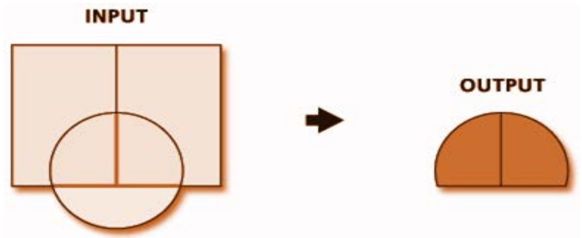
Spatial analysis extracts or creates new information from spatial data.

There are many geospatial algorithms ...

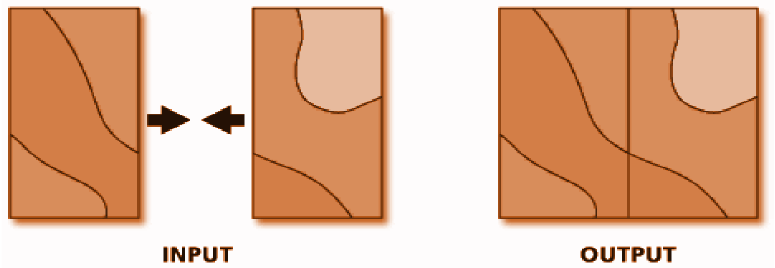
Clip



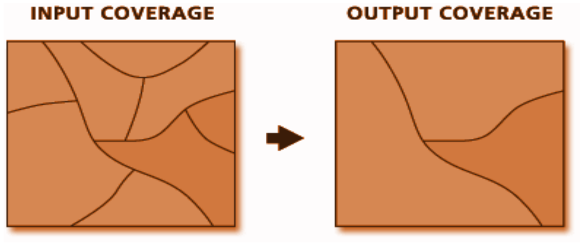
Intersect



Merge



Dissolve



\+ 100s more

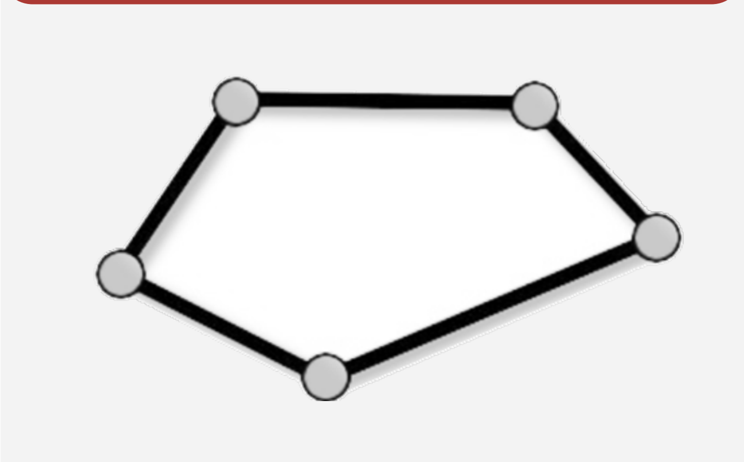
… but we will focus on Geospatial algorithms for gridded population data

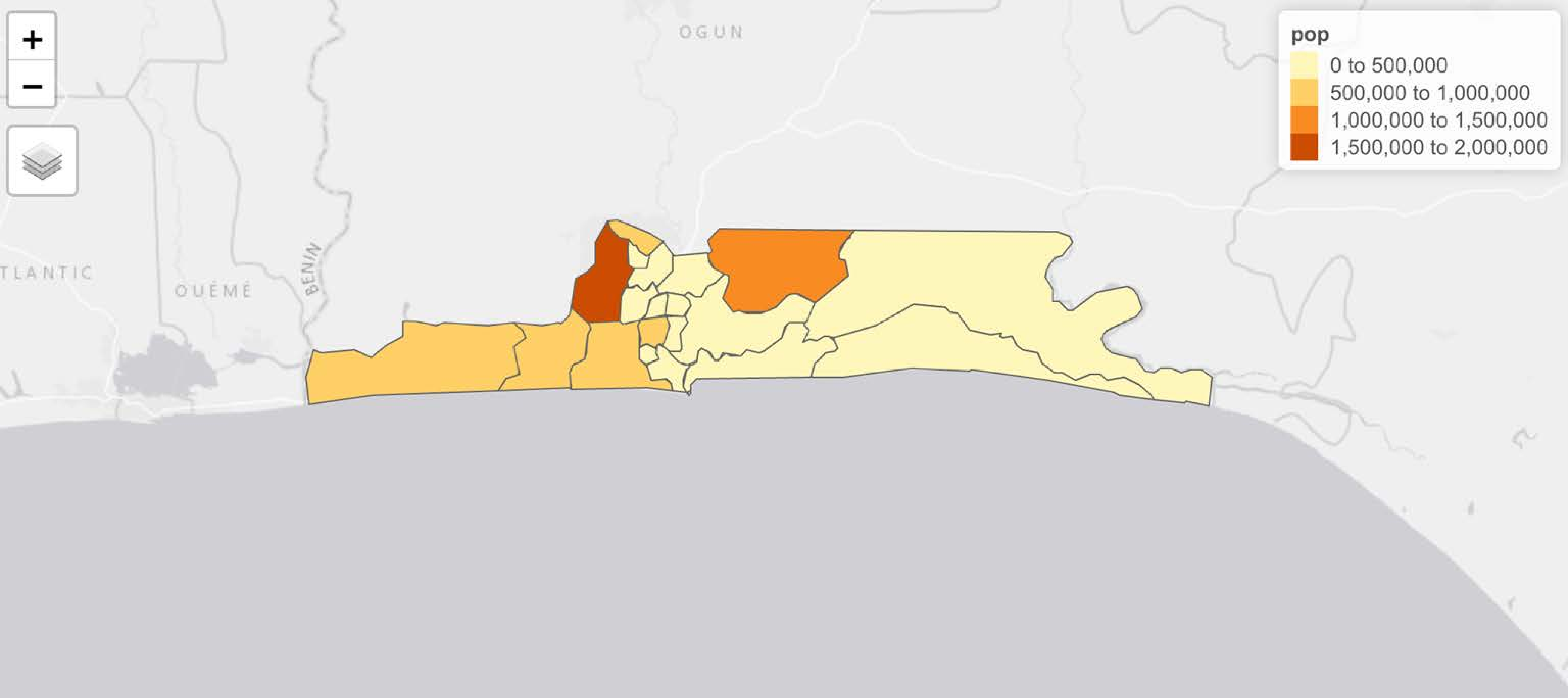
Image credits: Environmental Systems Research Institute 2019

Two Main Scenarios

Points and Lines

Polygons

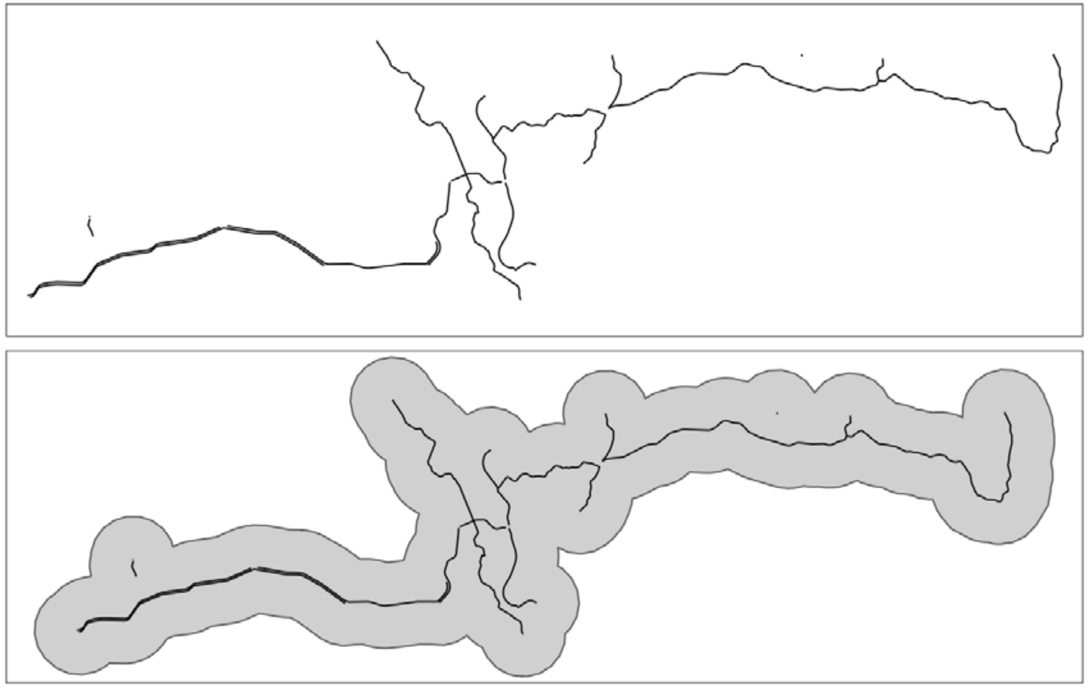


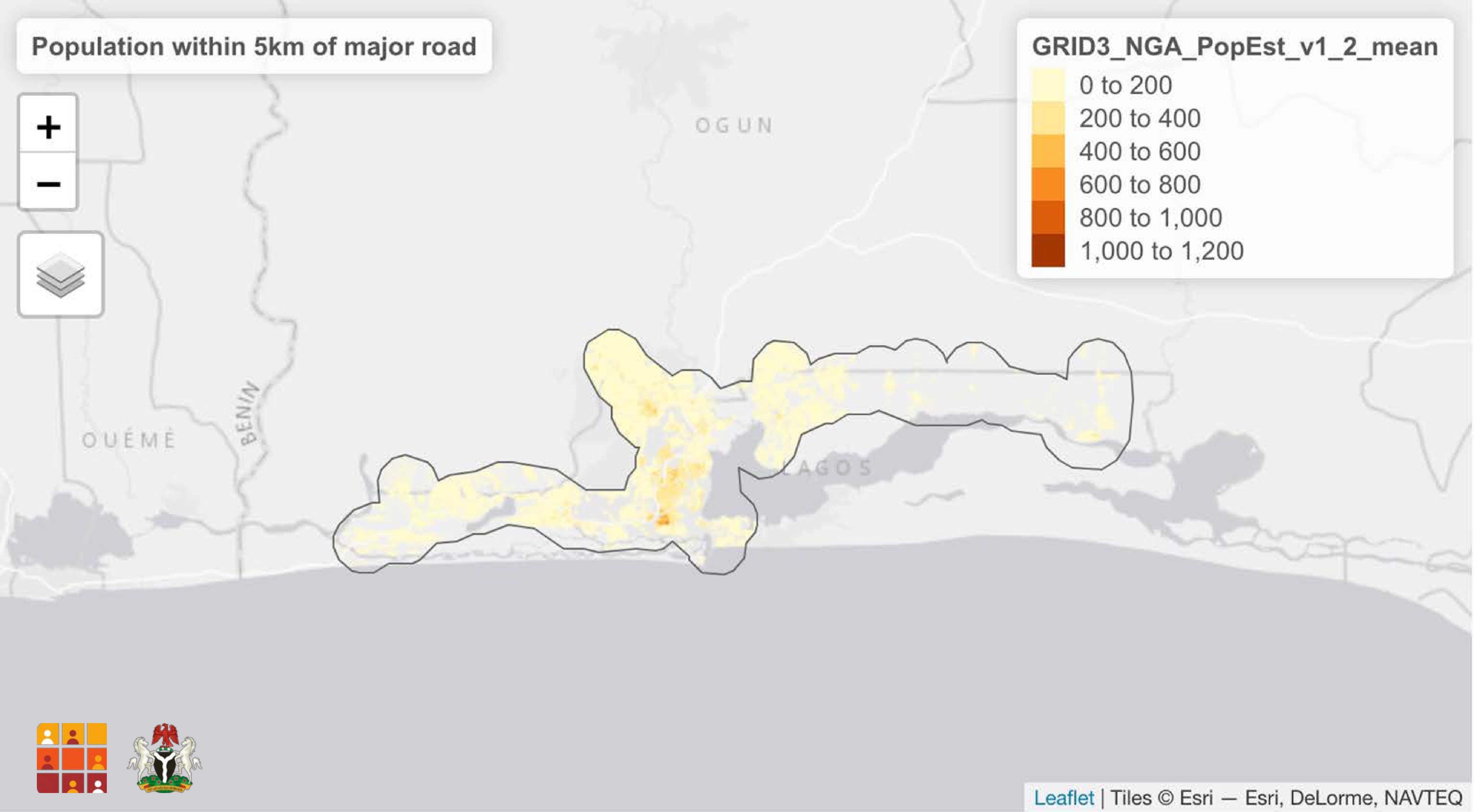


Population within polygons

Population Near Points or Lines

- We must determine the area that we are interested around the object
- Most likely to calculate a buffer




Population within 5km of major road
GRID3_NGA_PopEst_v1_2_mean
0 to 200
200 to 400
400 to 600
600 to 800
800 to 1,000
1,000 to 1,200
OGUN
OUÉMÉ
BENIN
LAGOS
Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ

Instructions

Go to this link to RStudio Cloud to run the code in realtime:

https://rstudio.cloud/project/741928

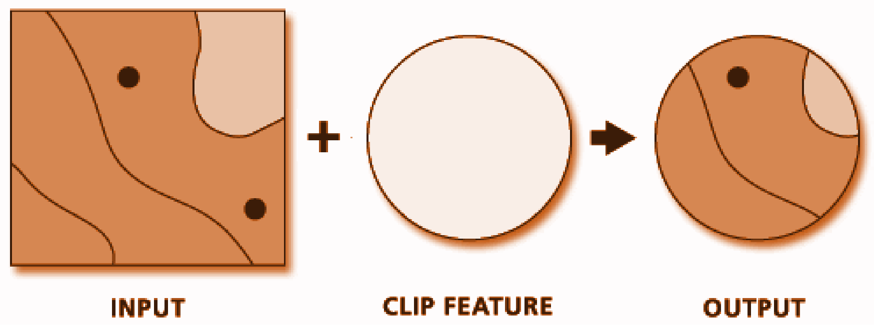
If you can't use RStudio cloud, follow with the html file:

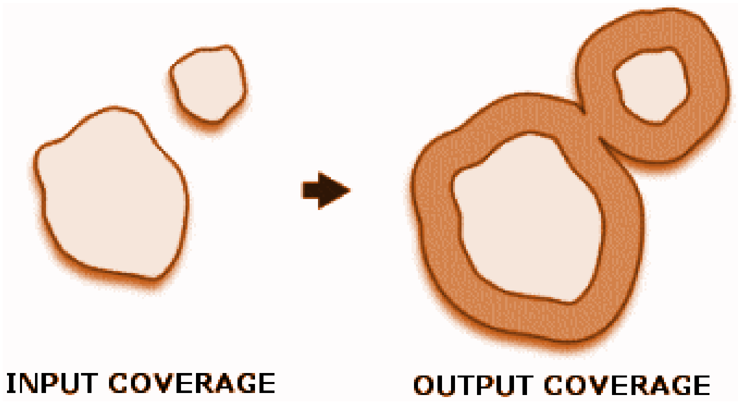
bit.ly/GRID3_Practical1

Building Geospatial Toolkits

We will often need to chain multiple tools together

How many people live within locations are within 2km of a road?

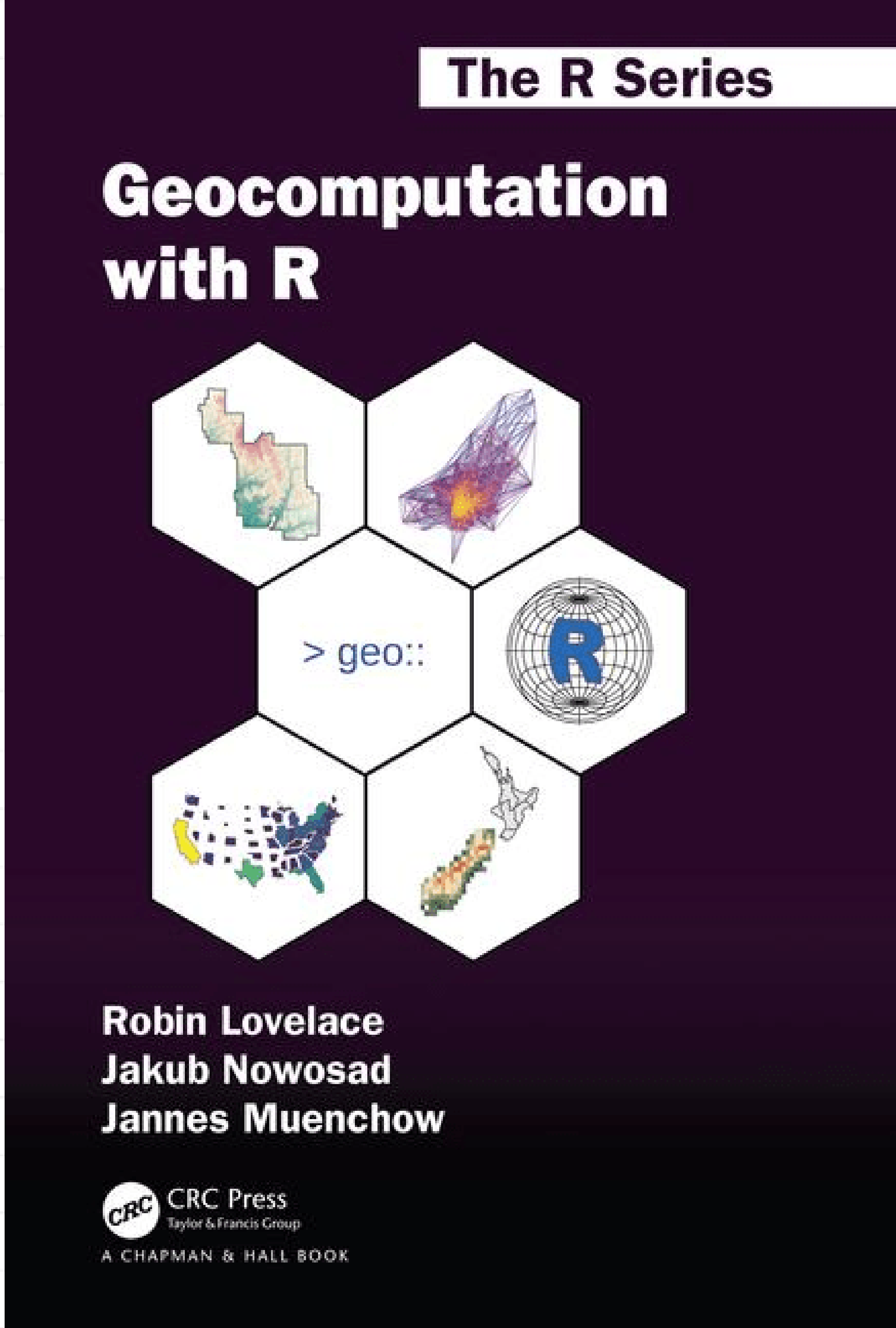




Next Steps

- Interested in R spatial analysis?
- Lots of free online material:
- Good source for spatial packages on R:https://www.r -spatial.org/

https://geocompr.robinlovelace.net/



GRID3 NIGERIA

UNITY AND FAITH, PEACE AND PROGRESS

Geospatial Analytics for Nigeria

Session 2 Part 3: Application of GRID3 data to solve real-world problems

Alina Game

BILL & MELINDA GATES foundation
UKaid from the British people
worldpop
FLOWMINDER.ORG
UNFPA
Center for International Earth Science Information Network
EARTH INSTITUTE | COLUMBIA UNIVERSITY

About Me

- Alina Game
- GIS Analyst
- Interest in spatial data applications for decision-making

Learning Objectives

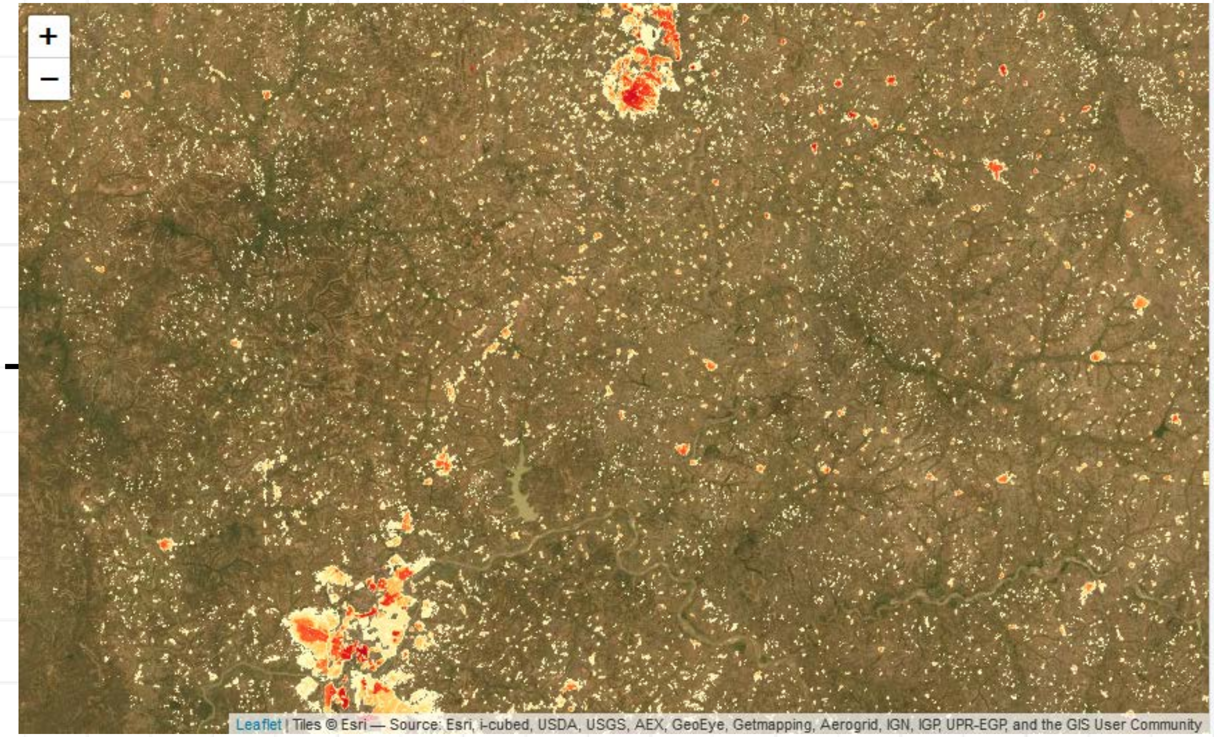
- Gain an understanding of the GRID3 datasets and their applications.
- Understand how GRID3 data can be used for decision-making.
- Get inspired on ways you can use GRID3 data in your work.

Intro to R

- Statistical Computing
- Main data science toolkit
- Scientific Research

R for Geospatial Analysis

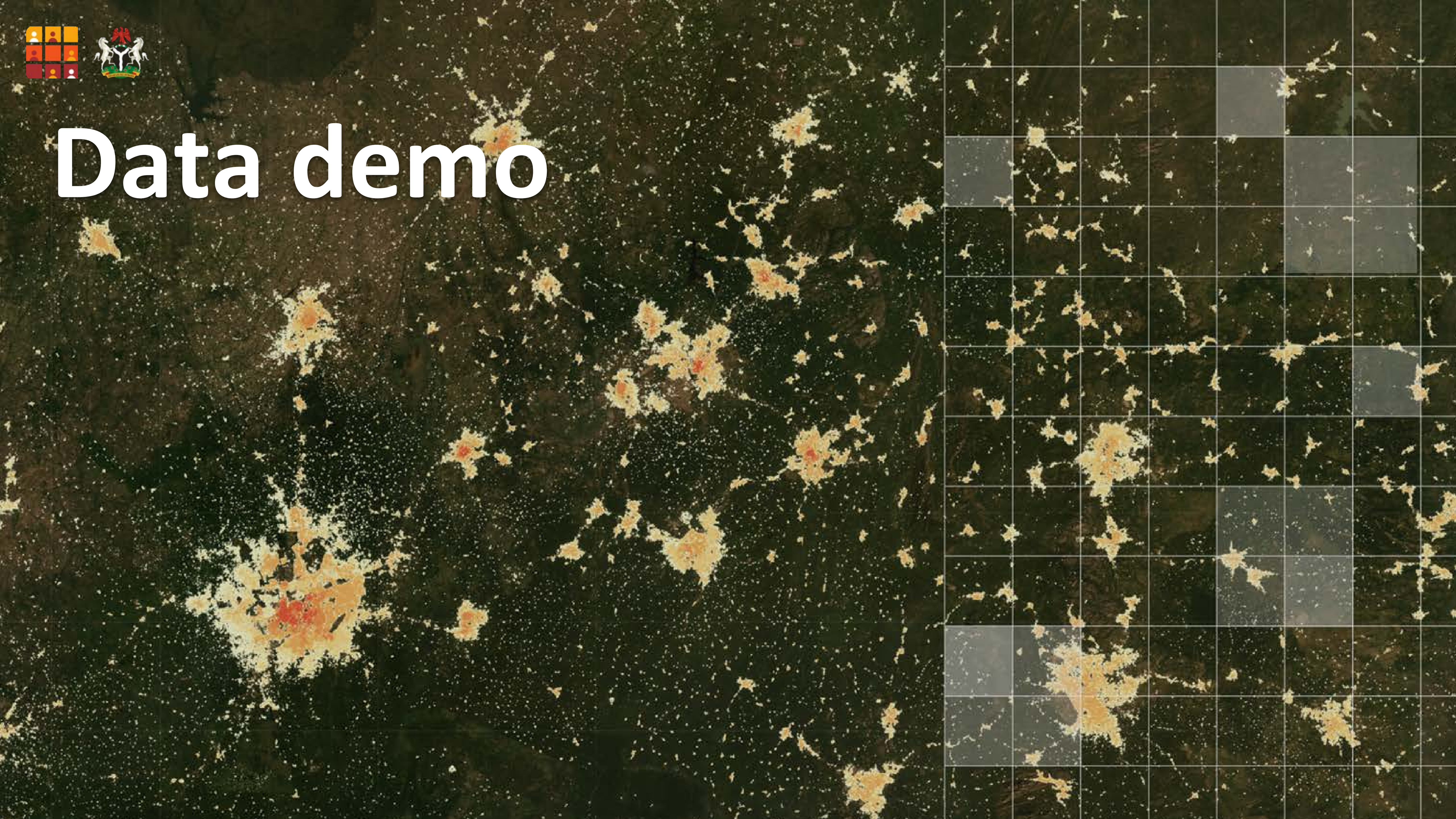
- Reproducibility
- Advanced capabilities for managing - analysing spatial data
- Visualisations



R markdown

- A combination of R and Markdown (a simple markup language)
- Save and execute code within the report
- Generate high quality reports directly from the analysis

Data demo



Data Demo

Assessing coverage of health facilities using GRID3 population estimates for maternal health in Kaduna State.

Using GRID3 Data:

- Population data for women age 14-49
- Ward and state boundaries
- Health facility locations

Instructions

Go to this link to RStudio Cloud to run the code in realtime:

https://rstudio.cloud/project/741928

If you can't use RStudio cloud, follow with the html file:

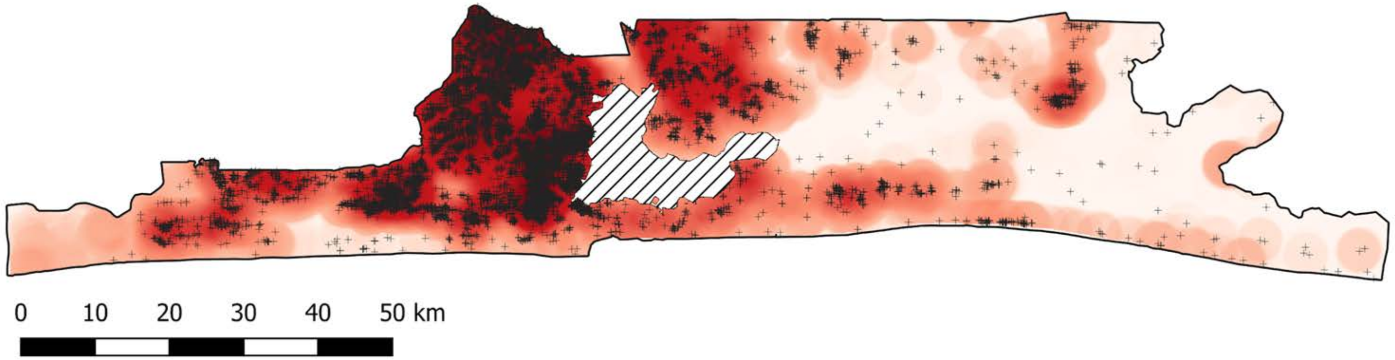
bit.ly/ GRID3_Practical2

Further Work

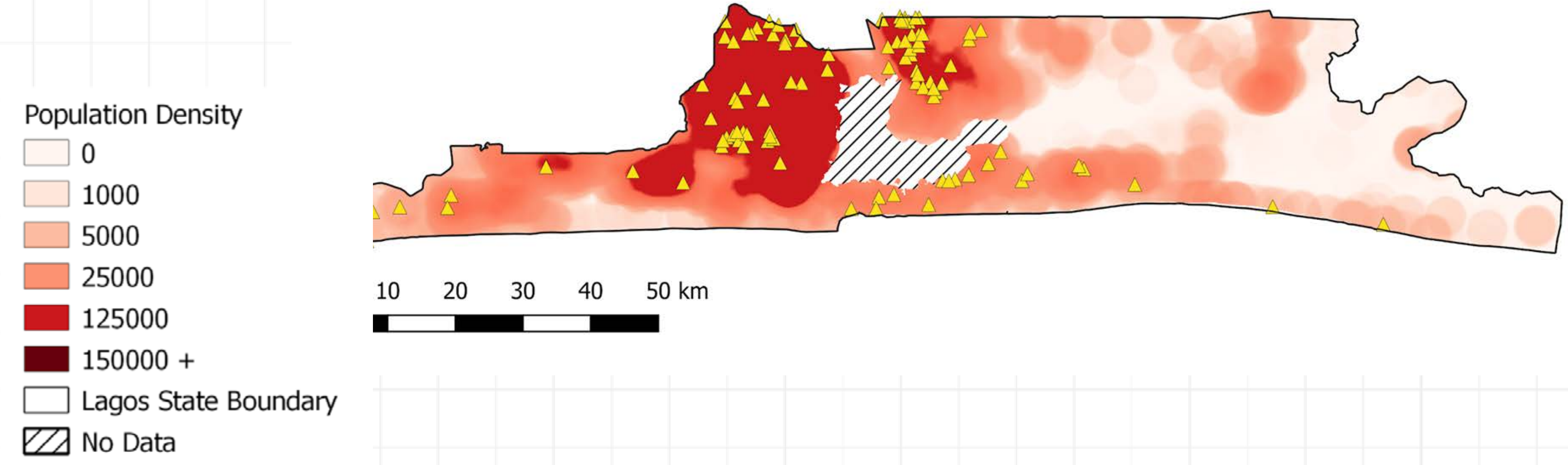
Optimisation of the placement of new schools and expansion of existing schools

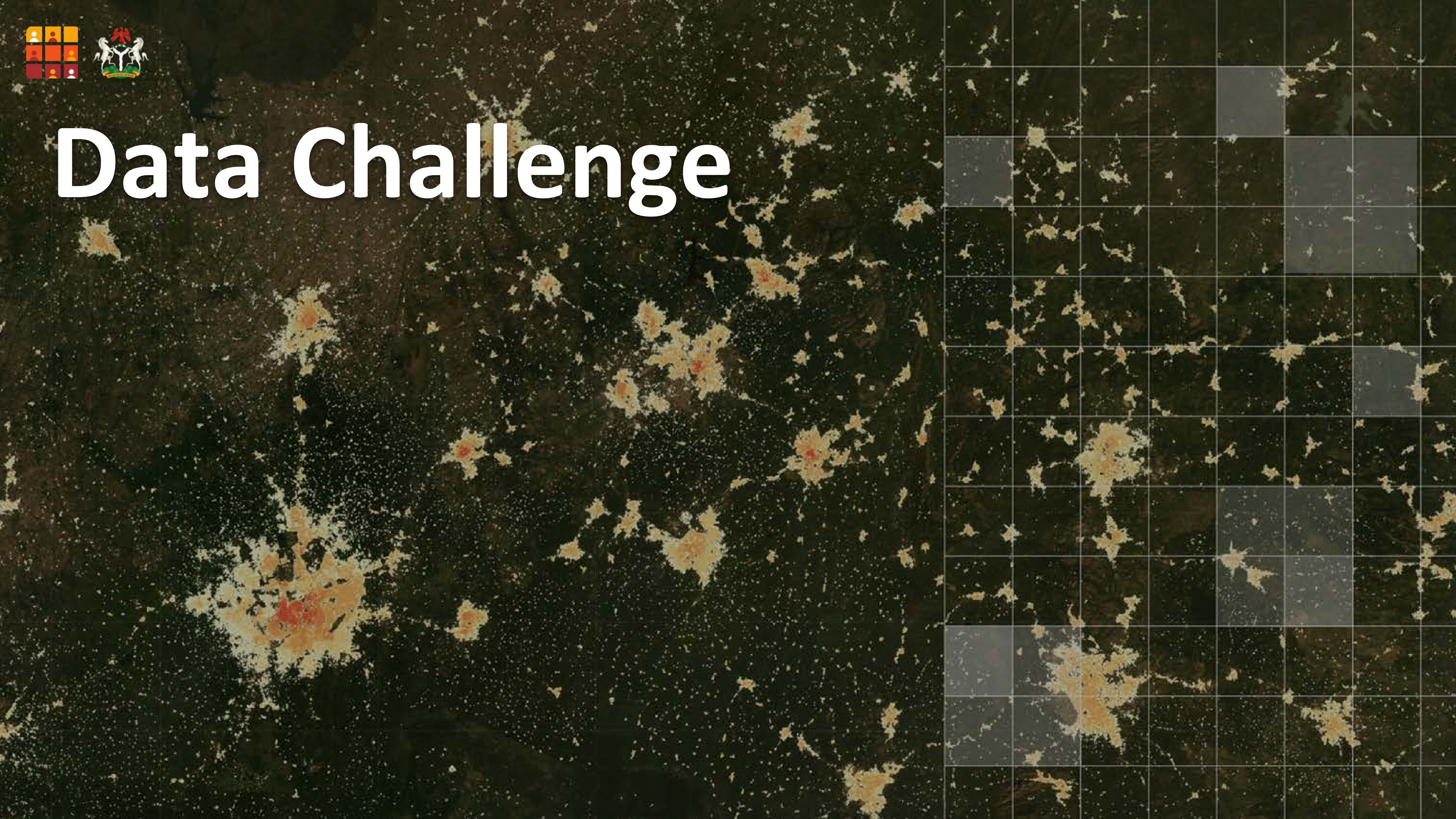
- Optimising availability
- Optimising availability + accessibility

Existing Schools



Simulated School Placement (100 schools)




Data Challenge

The Challenge

Which problem can be solved with gridded population data in Nigeria?

What are we looking for?

- **Code is not required**, we are interested in the idea, but you should have an idea of how it could be implemented
- **Could the idea have high impact?**
- **Creativity is encouraged**
- **Is the idea achievable?**
- **Think about more than just the data analysis.** How might your data analysis support decision-makers? Who are the stakeholders?

Prize

- Mentorship with a GRID3 analyst to develop your idea
- Feature on the GRID3 website as a showcase data application

Submission

Deadline Friday 6th December

Submit your proposals at

http://bit.ly/333IjnO

Visit the GRID3 Nigeria Portal to download our data: grid3.gov.ng

For more information, contact us: **info@grid3.org**

For project updates and announcements, visit us online at:

@GRID3Global www.grid3.org

Or follow our partners on Twitter at @Flowminder, @WorldPopProject, @UNFPA, @PopDevUNFPA, and @CIESIN