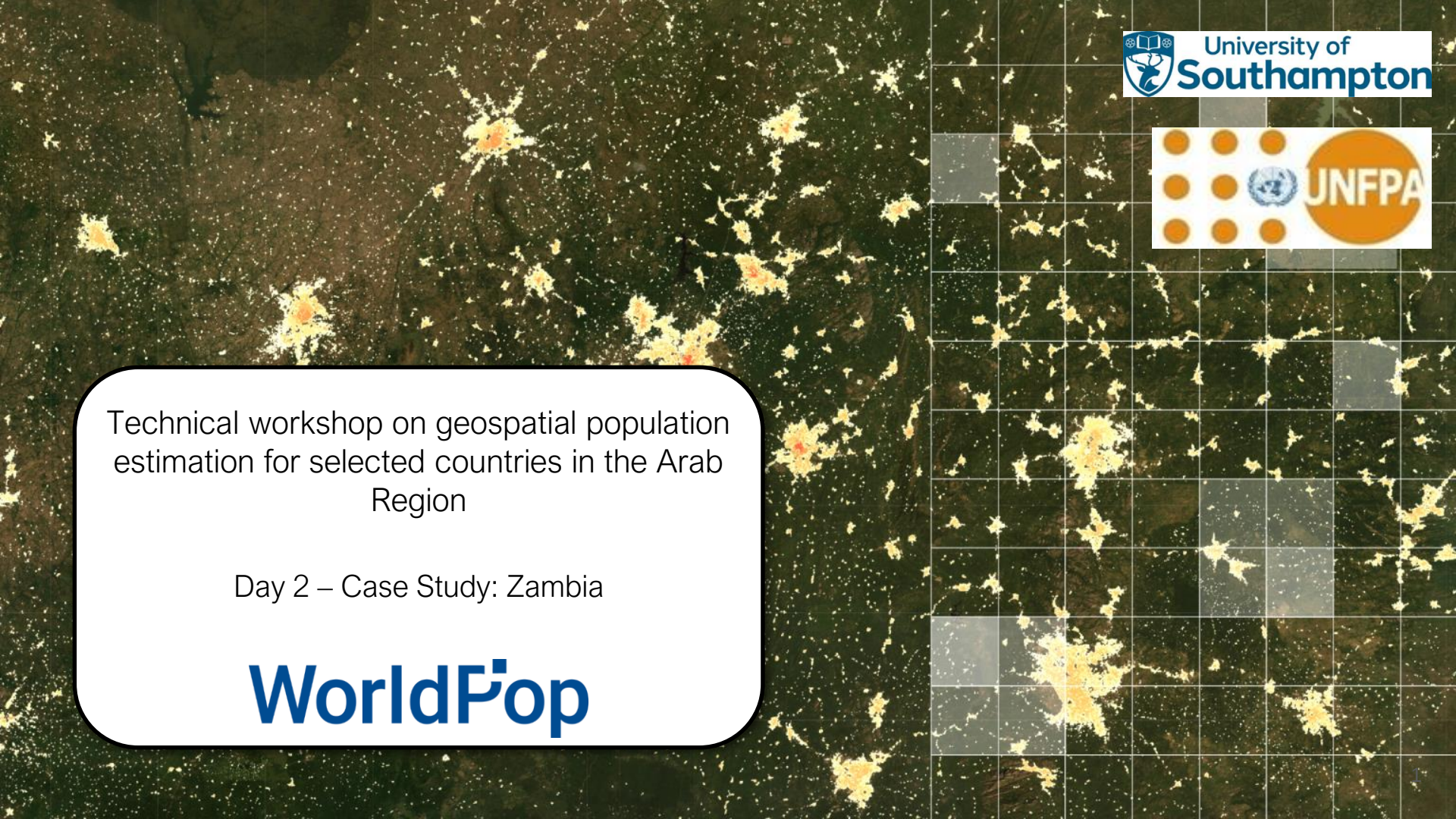


University of Southampton

UNFPA

Technical workshop on geospatial population estimation for selected countries in the Arab Region

Day 2 – Case Study: Zambia

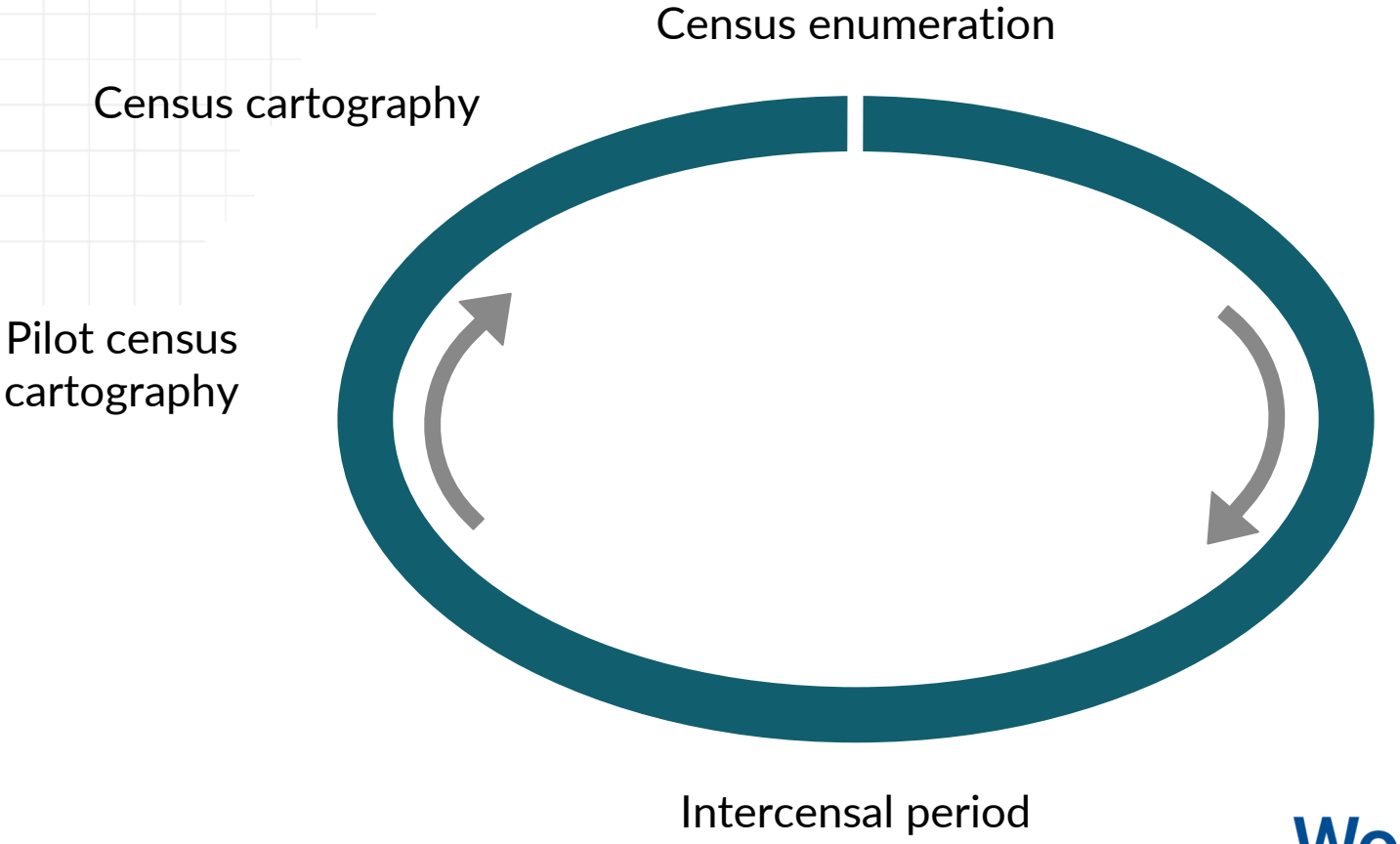
WorldPop

Background: Zambia

- Last census in 2010
- 2020 census postponed due to funding challenges and Covid-19
- Subnational variations in fertility, mortality and migration have made projections uncertain and inaccurate
- Strong need for new estimates for operational purposes and to support planning for next census
- Recent partial enumeration/survey sample data have been collected

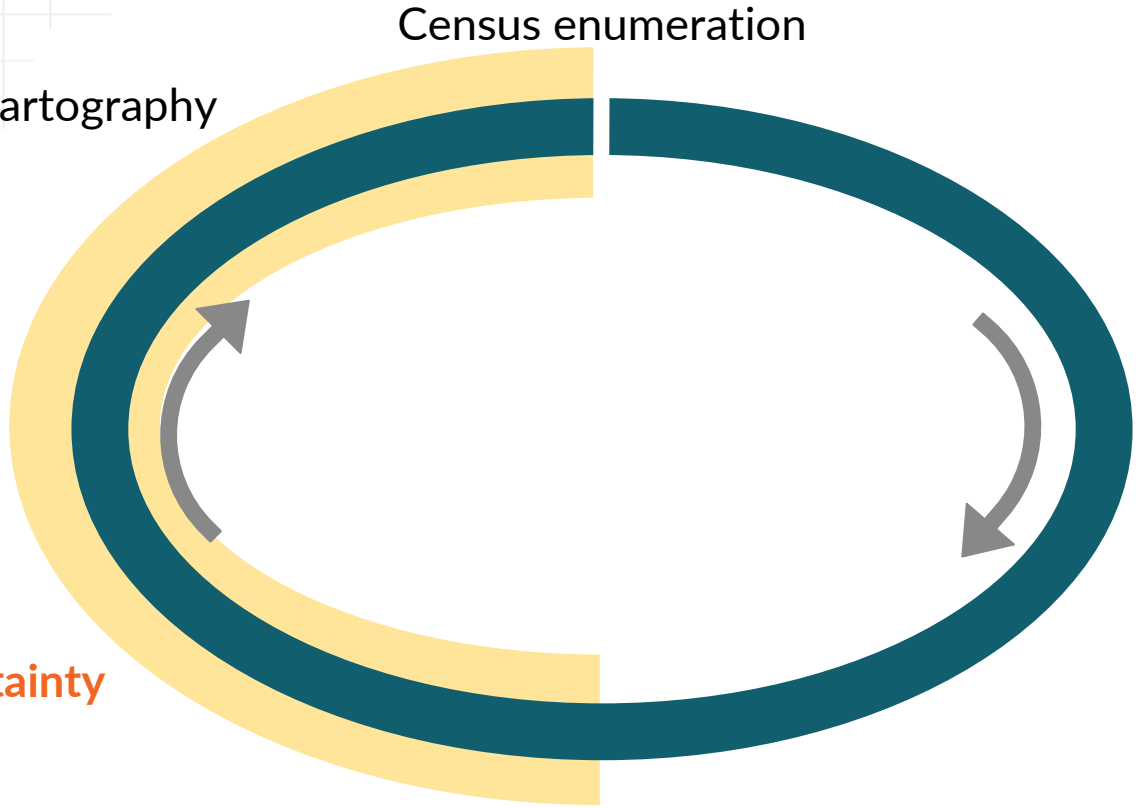
WorldPop

Stages of the census cycle



WorldPop

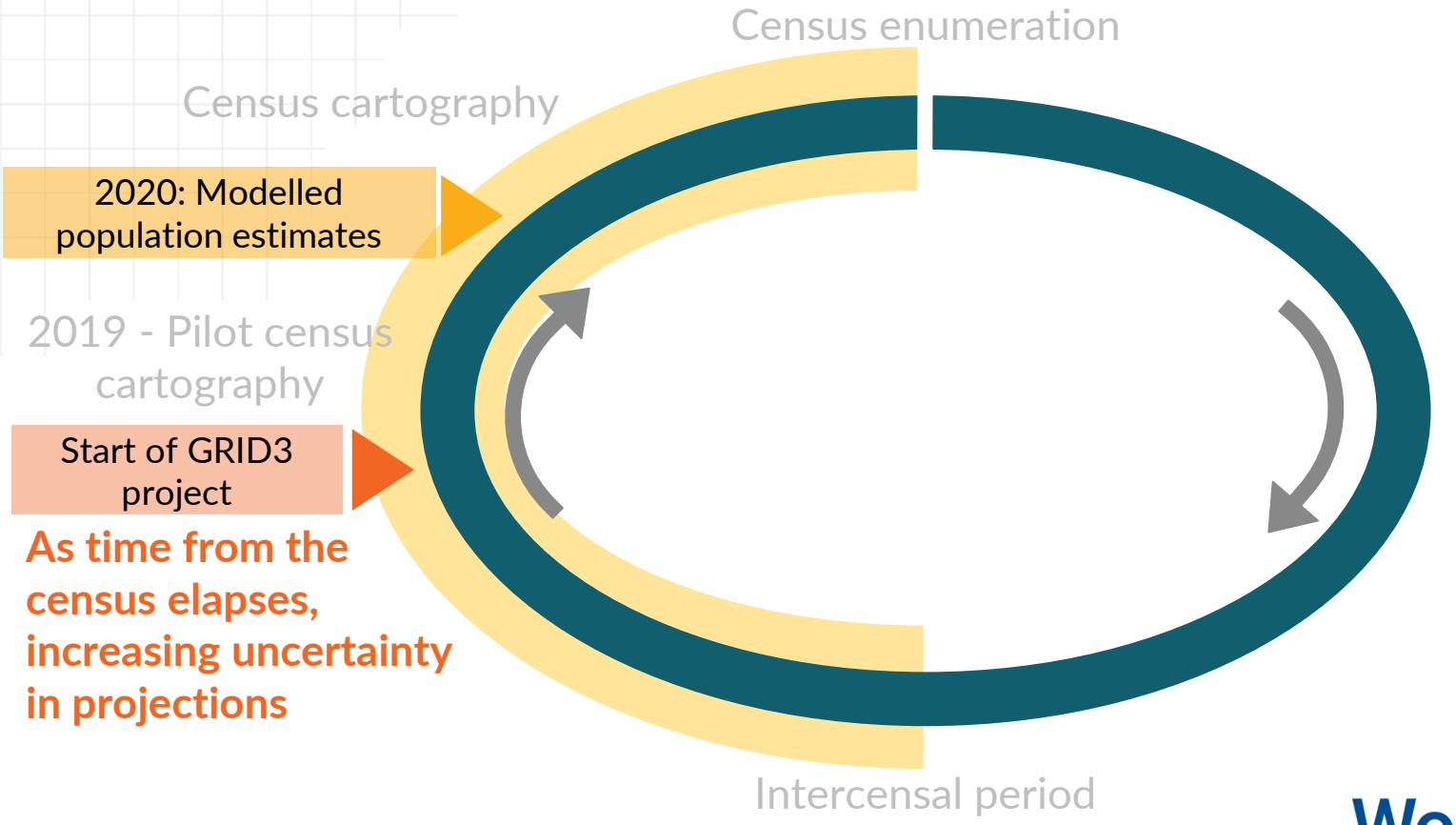
Stages of the census cycle



As time from the census elapses, increasing uncertainty in projections

WorldPop

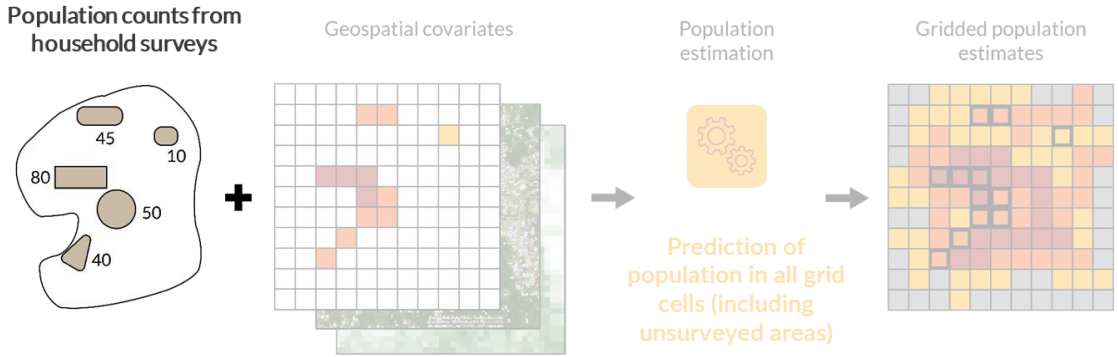
Stages of the census cycle: Zambia



WorldPop

Input datasets

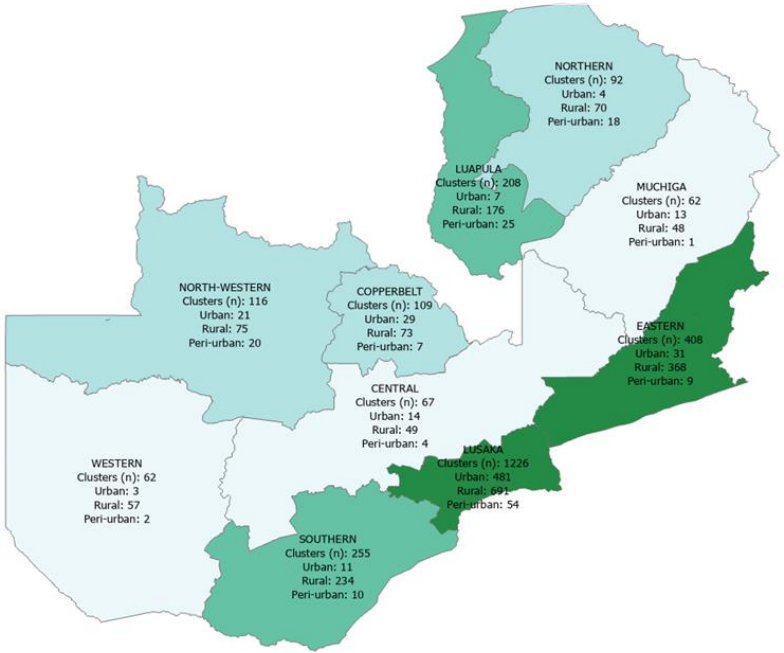
Population counts



3 sources of population counts:

- Pilot mapping/cartography
- Livestock census survey
- Saving Mothers, Giving Life survey

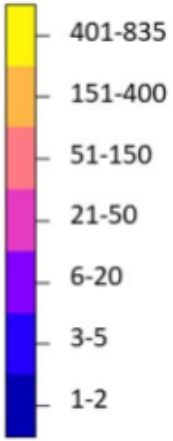
All survey datasets included counts of population enumerated for each household, with household or building locations (latitude and longitude) recorded using GPS-enabled devices.

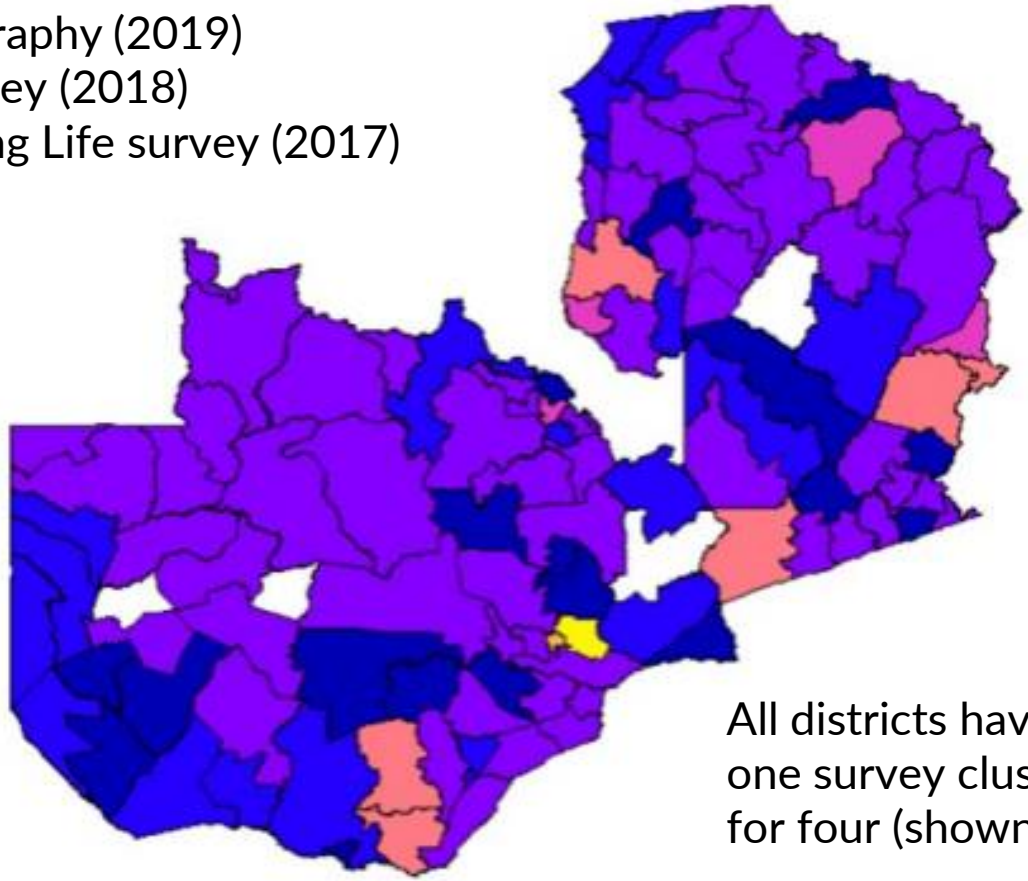


3 sources of population counts:

- Pilot mapping/cartography (2019)
- Livestock census survey (2018)
- Saving Mothers, Giving Life survey (2017)

Counts of survey clusters per district

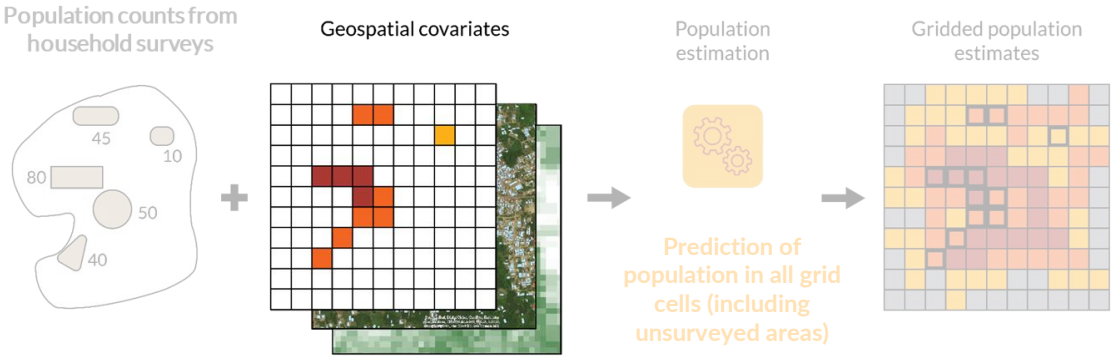




All districts have at least one survey cluster, except for four (shown in white):

Input datasets

Geospatial covariates



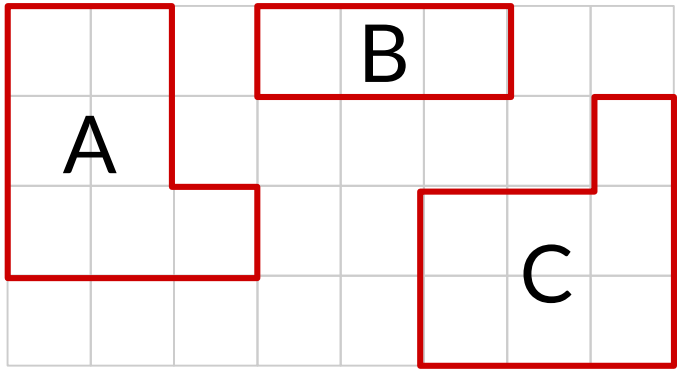
- Predictive variables related to population density, and associated with the **built** and **natural environment** are considered, such as building density and vegetation cover.
- Geospatial covariates are **spatially harmonised** so that all variables have the same spatial resolution and grid cell alignment → geospatial covariate stack
- Summarised for each population survey cluster location

Input datasets

Geospatial covariates

- Summary statistics calculated for each population survey cluster location

Clusters



Raster

3	3	4	2	3	3	4	4
0	3	4	3	0	3	0	3
3	1	4	2	2	2	0	0
0	1	0	3	0	0	0	3

Cluster	Mean value
A	2.43
B	2.67
C	1.14

Input datasets

Building footprint derived covariates

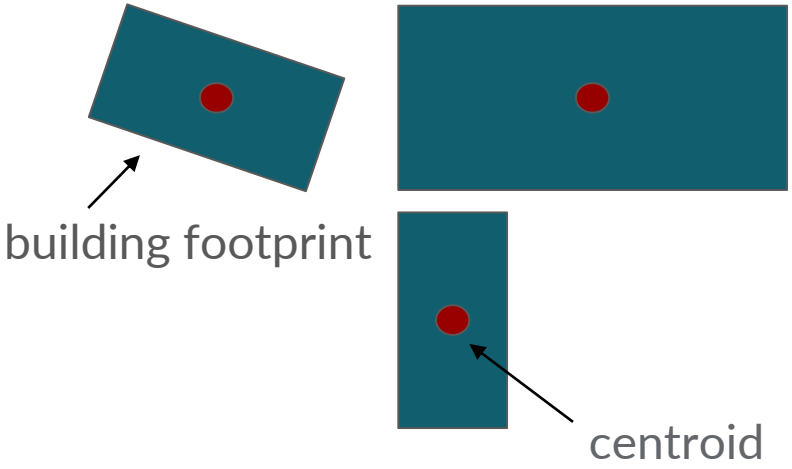
- Building footprints from Ecopia-Maxar (DigitizeAfrica), created using feature extraction techniques with satellite imagery.
- 93% of the building footprints were extracted from imagery dating from 2017 to 2019.
- No residential/non-residential distinction.
- Very large buildings (>750m^2) filtered out.

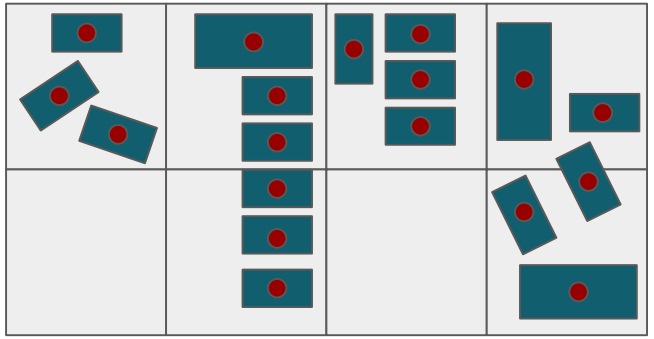
Input datasets

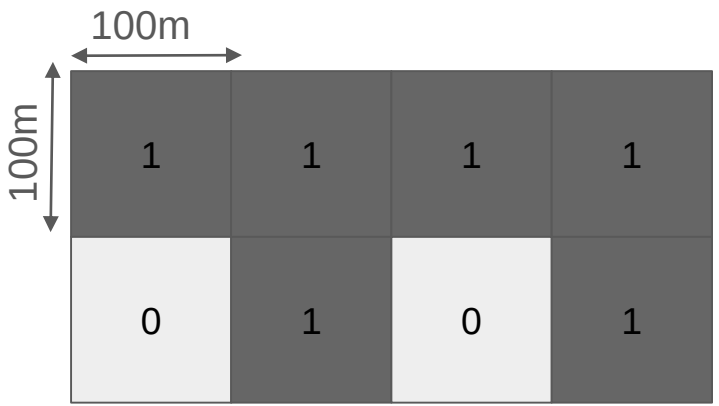
Geospatial covariates - settlement

All grid cells with 1 or more buildings are considered to potentially have population

- Based on building footprint centroid



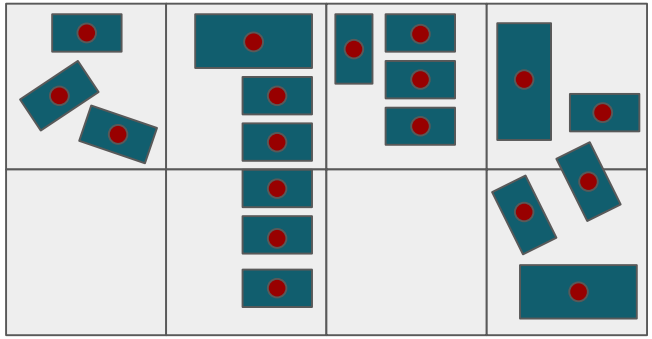




Input datasets

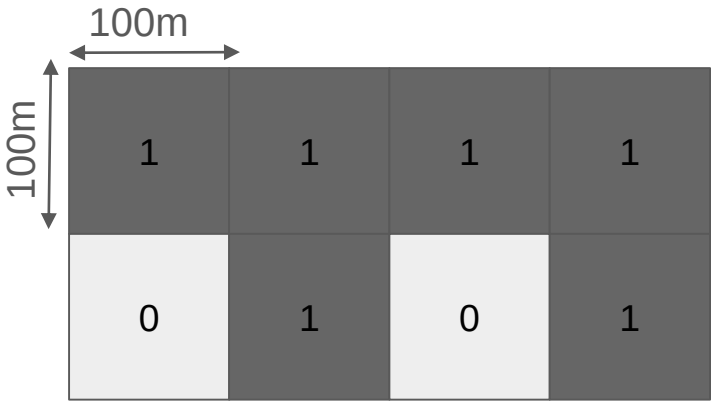
Geospatial covariates - settlement

Population density calculated as population per total building area (as determined from building footprints).



Settled grid cells classified based on size of settlement, considering contiguous grid cells:

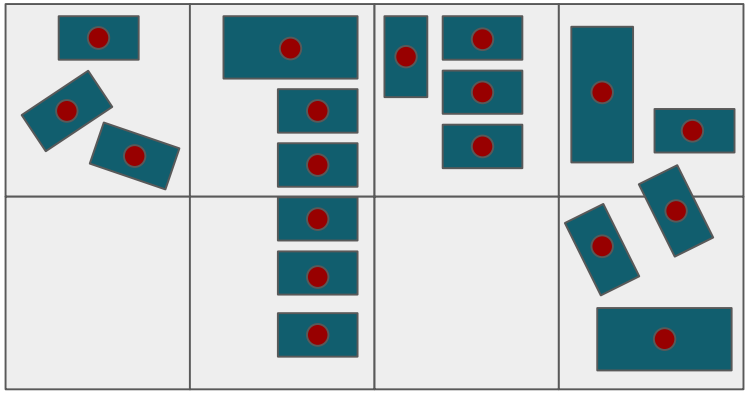
- Rural: clumps of <500 grid cells
- Small urban: 500-1500 contiguous grid cells
- Large urban: - >1500 contiguous grid cells



Input datasets

Geospatial covariates - settlement

- Deriving metrics from building footprints:
  - Building count
  - Building area (min, max, mean, total, sd)
  - Building perimeter (min, max, mean, total, sd)
  - Count of building vertices (min, max, mean, total, sd)



3	3	4	2
0	3	0	3

Building count

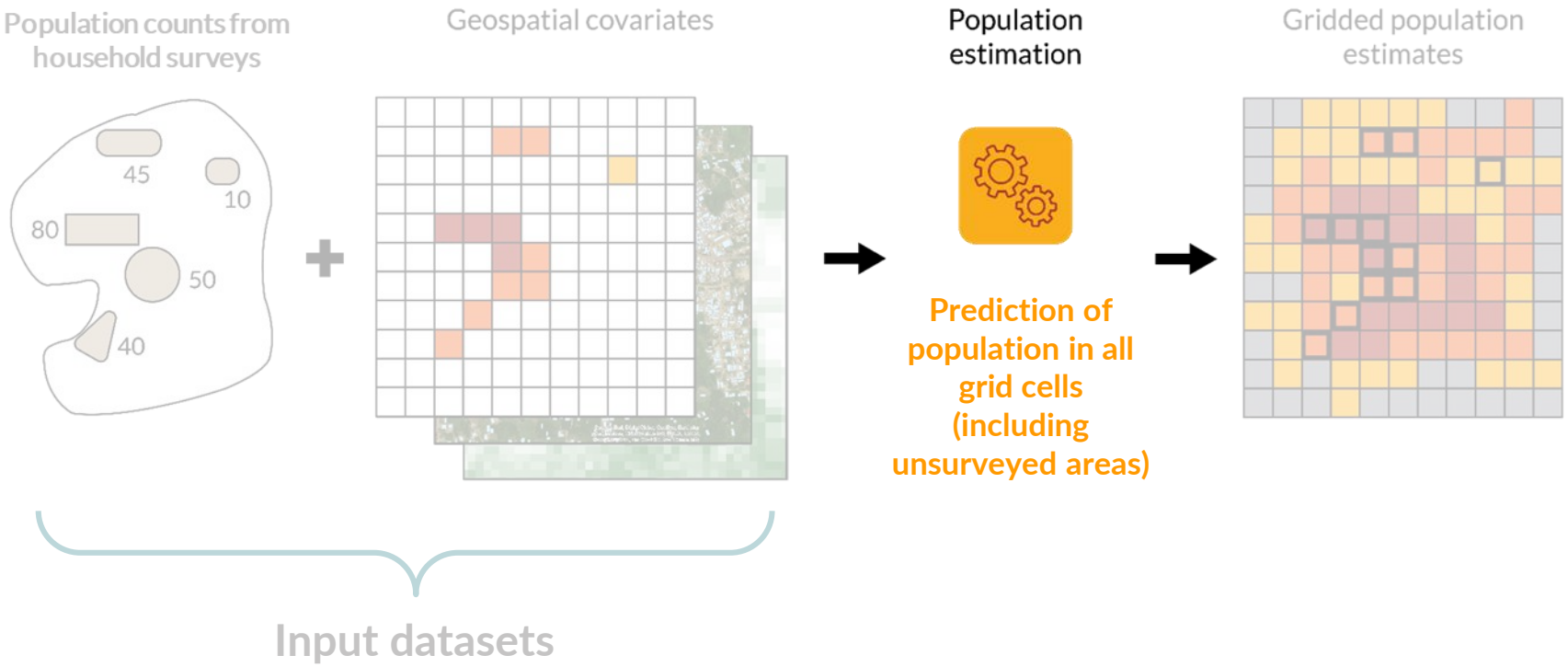
2	2.8	2	2.6
0	2	0	2.4

Mean building area

4	8	4	8
0	4	0	8

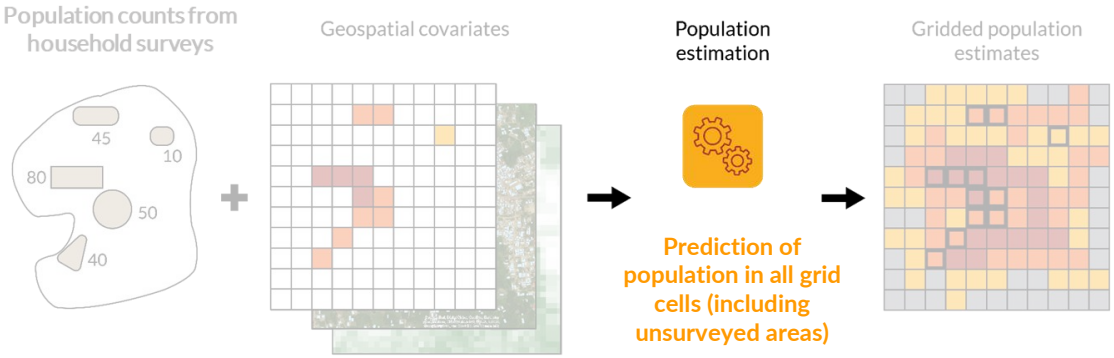
Maximum building perimeter

The “bottom-up” approach



Population estimation

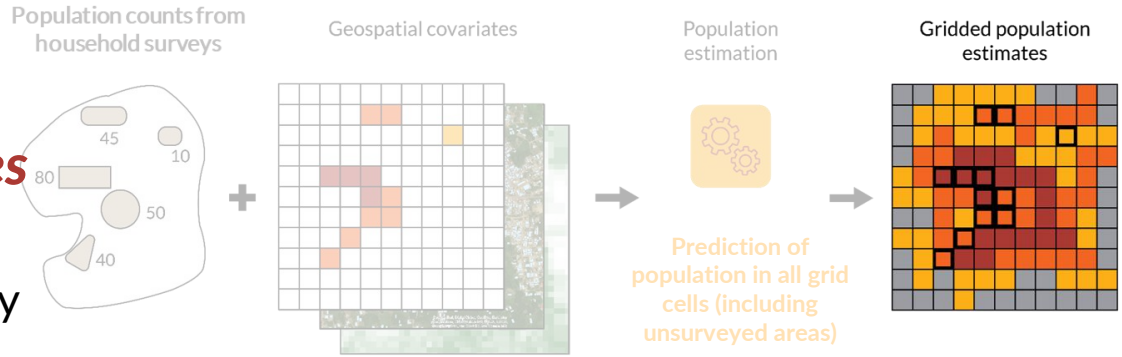
Bayesian statistical model



- **Core model** (used to predict) describes the relationship between people per building area and three key variables:
  - Mean building area (of buildings in a cluster or pixel)
  - Building density
  - Variation in building area
- **Sub-model** to accurately parameterise the core model
  - Account for recorded residential households that could not be enumerated in the survey data
  - Sample weights (Livestock census)

The output

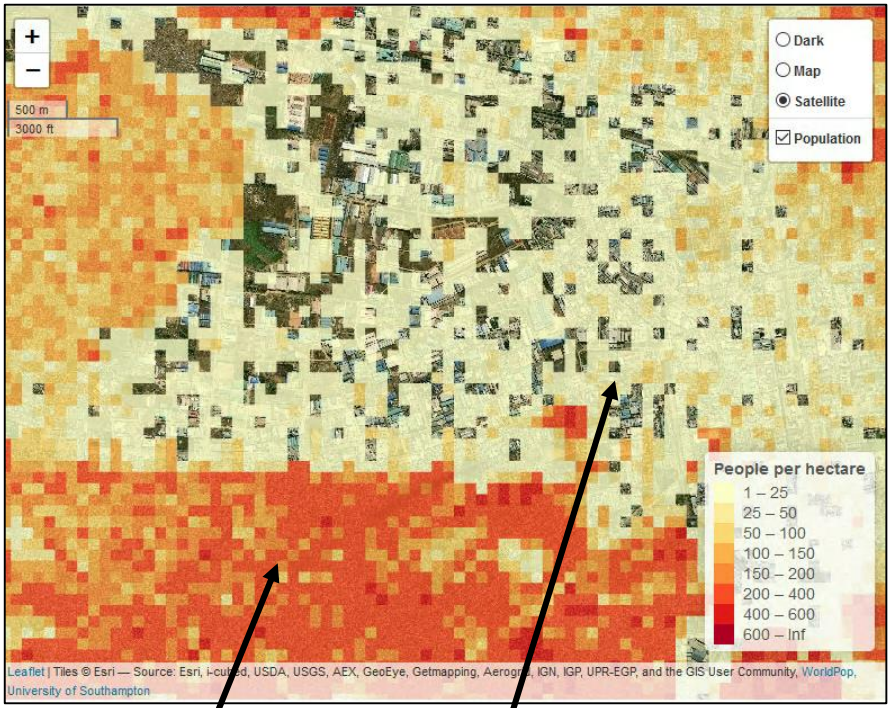
Gridded population estimates



- Output from model for every **~100m x 100m grid cell**
- For every grid cell, we calculated the **probability distribution of population counts** (i.e. posterior predictions), using a grid cell's covariate values and the parameterised model
- The posterior predictions provide, for every grid cell, a **mean population estimate** and a statistical measure of **uncertainty**
- Grid cell level posterior predictions are available in SQL database

“Bottom-up” modelled population estimates

Population estimates (predictions) for each 100m grid cell across Zambia



Good contrast in predicted population counts between areas of dense residential buildings & industrial areas

High pop. count

Low pop. count

Data accessibility

Data is available to download from:

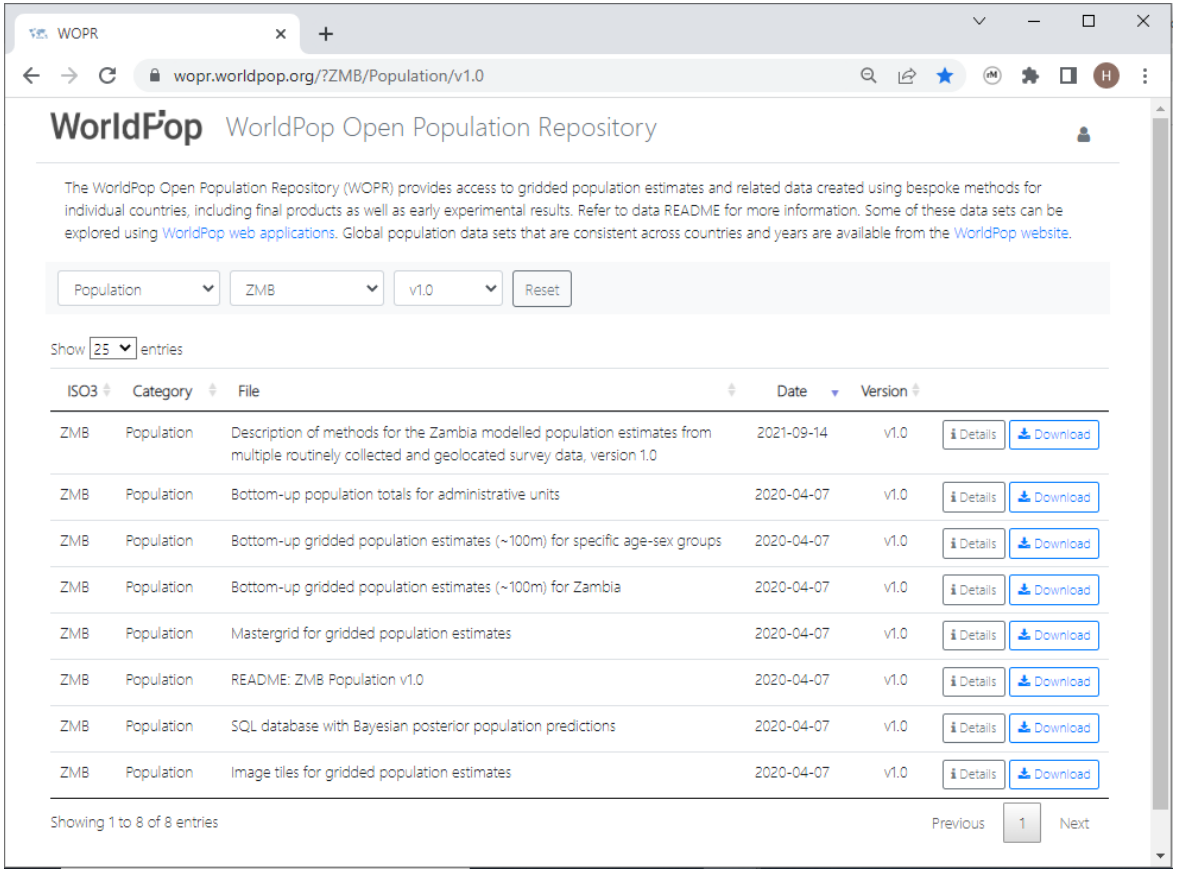
- WOPR: https://wopr.worldpop.org/
- GRID3 data hub: https://data.grid3.org/maps/grid3-zambia-gridded-population-estimates-version-1-0/

Aggregated population estimates can be calculated (with uncertainty) using woprVision interface:

- https://apps.worldpop.org/woprVision/

Data accessibility

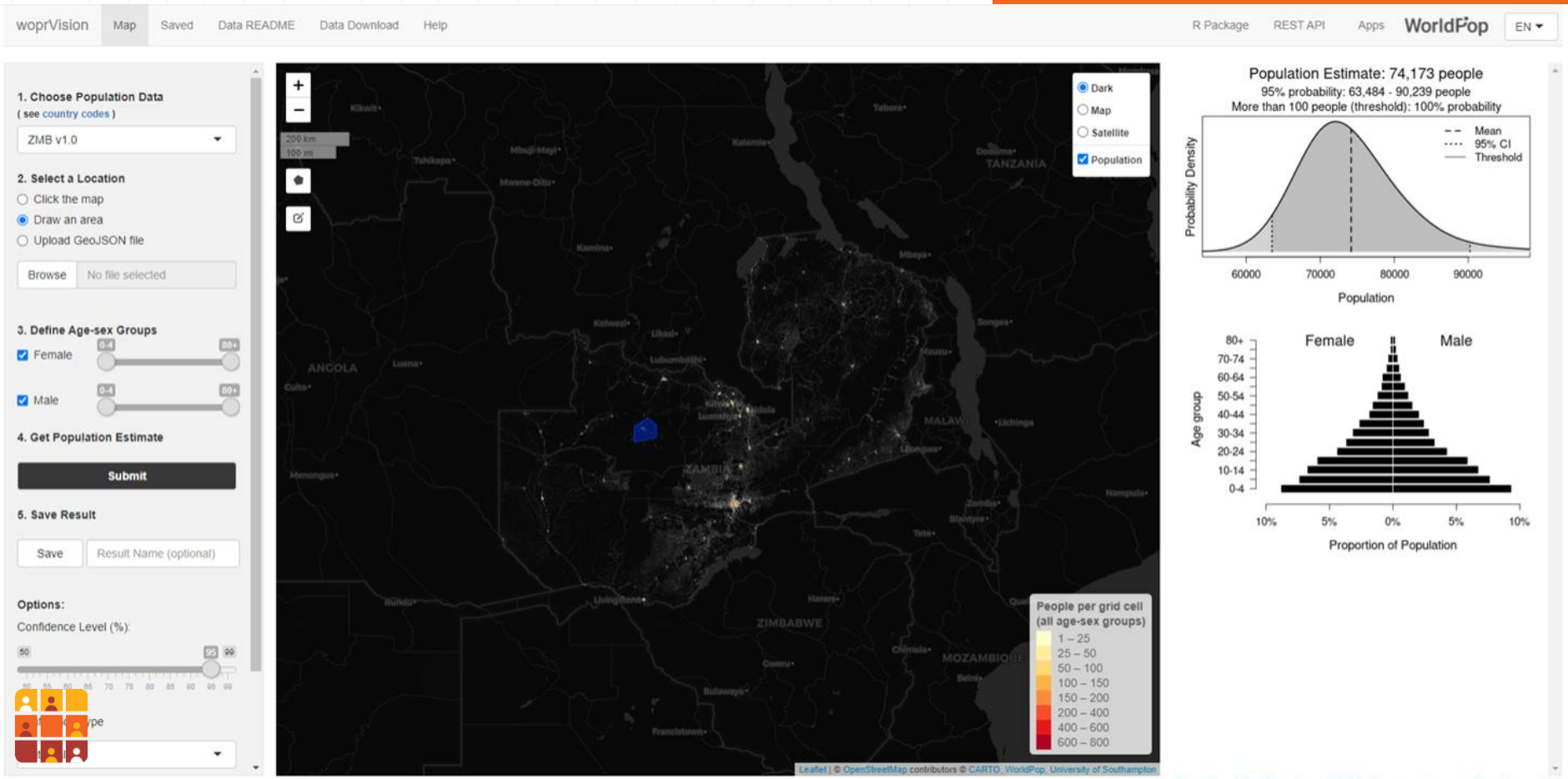
https://wopr.worldpop.org/?ZMB/Population/v1.0



Can download:

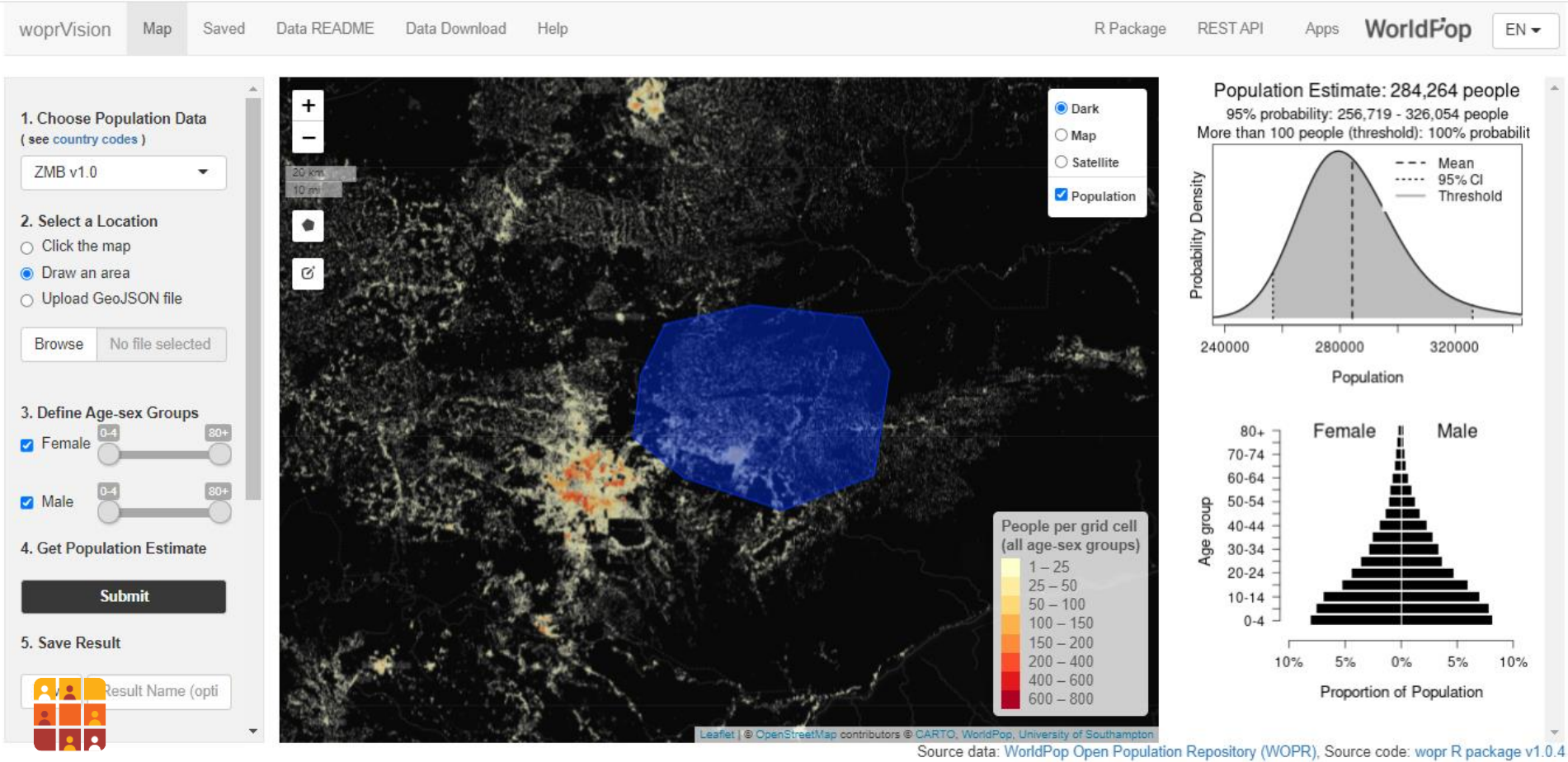
- Gridded population estimates (total population and age-sex structured)
- Aggregated population totals for administrative units
- Data README
- Methods report

https://apps.worldpop.org/woprVision/



Source data: WorldPop Open Population Repository (WOPR), Source code: wopr R package v1.0.4

https://apps.worldpop.org/woprVision/

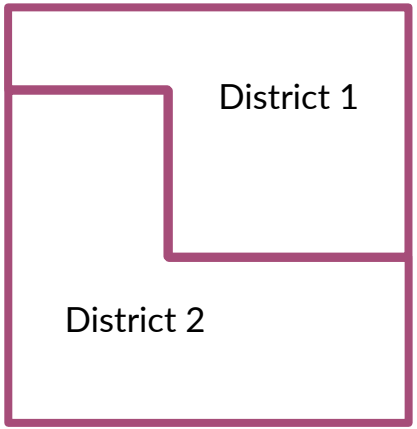


Source data: WorldPop Open Population Repository (WOPR), Source code: wopr R package v1.0.4

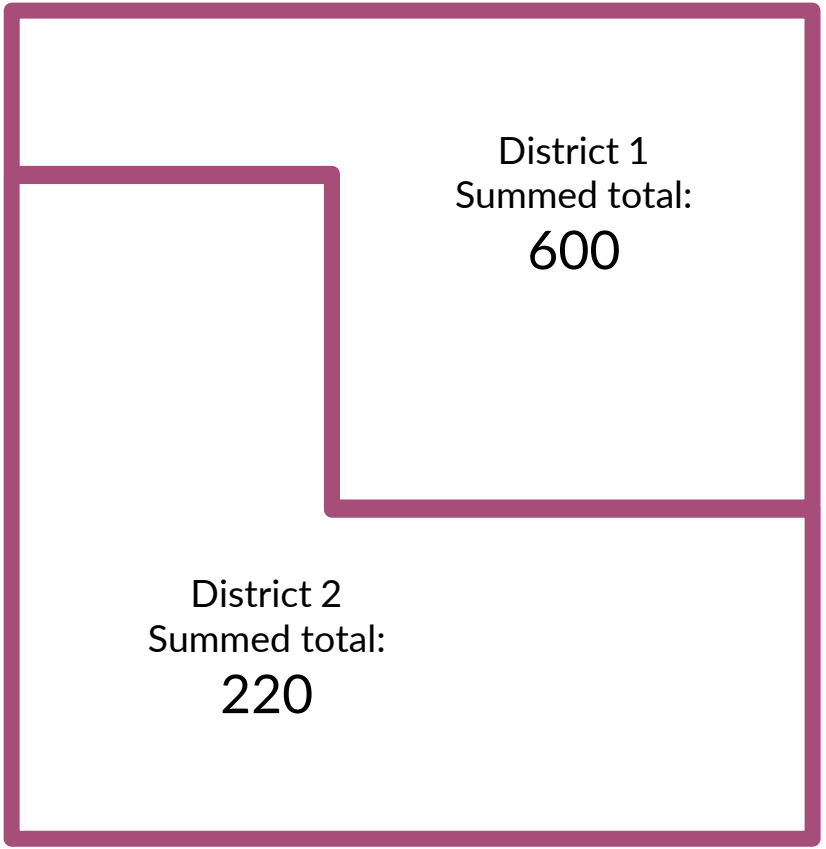
Calculating aggregated population estimates

60.6	121.2	60.6	18.0	12.0
0.9	43.1	121.2	60.6	60.6
0	8.6	60.6	12.0	12.0
0.9	8.6	12.9	0.9	12.9
0	129.4	0	0.9	0.9

Gridded population estimates (raster)



District boundaries (vector)



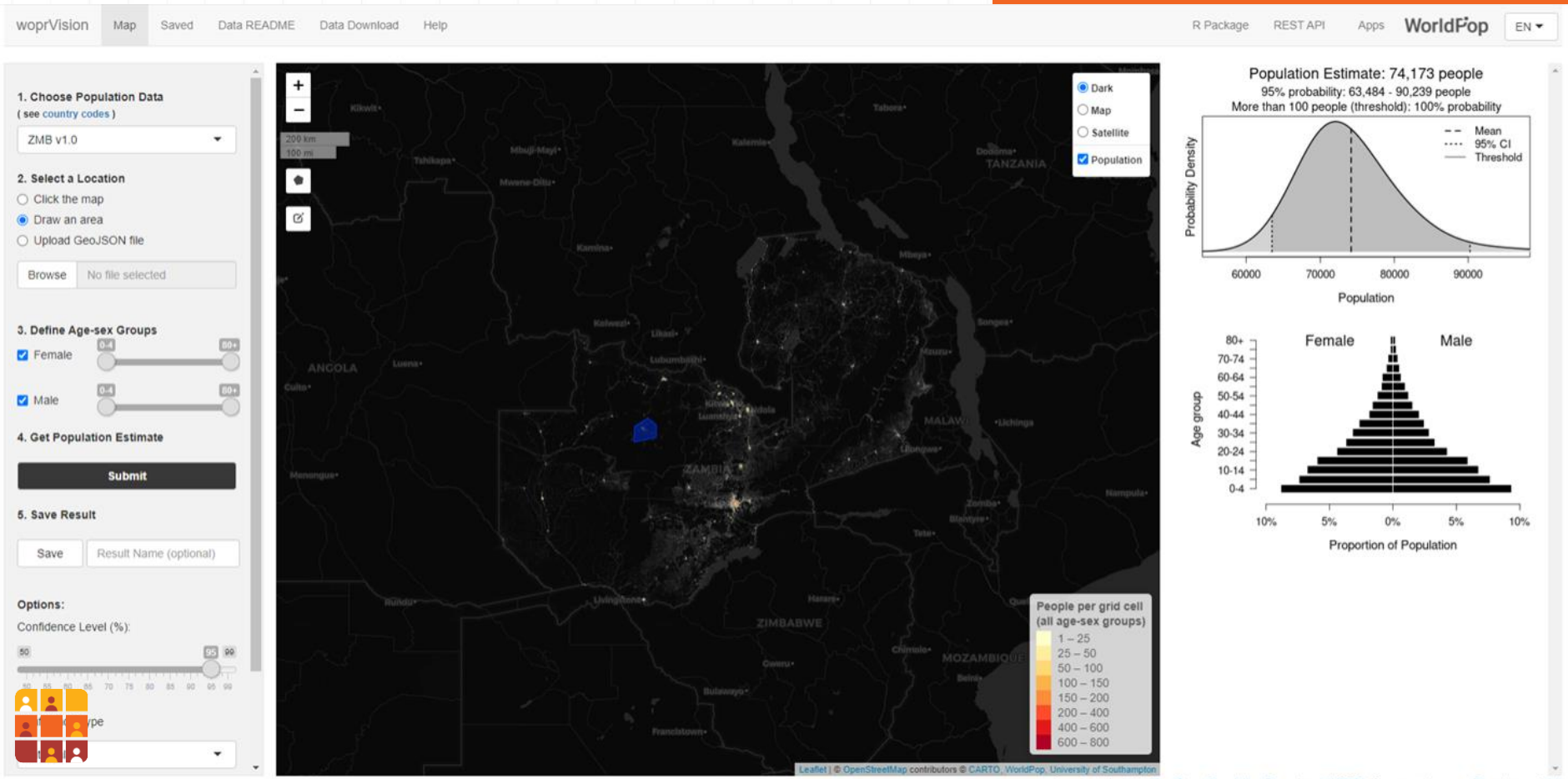
District 1:

60.6+121.2+60.6+18+12+121.2+60.6+60.6+60.6+12+12 = 600

District 2:

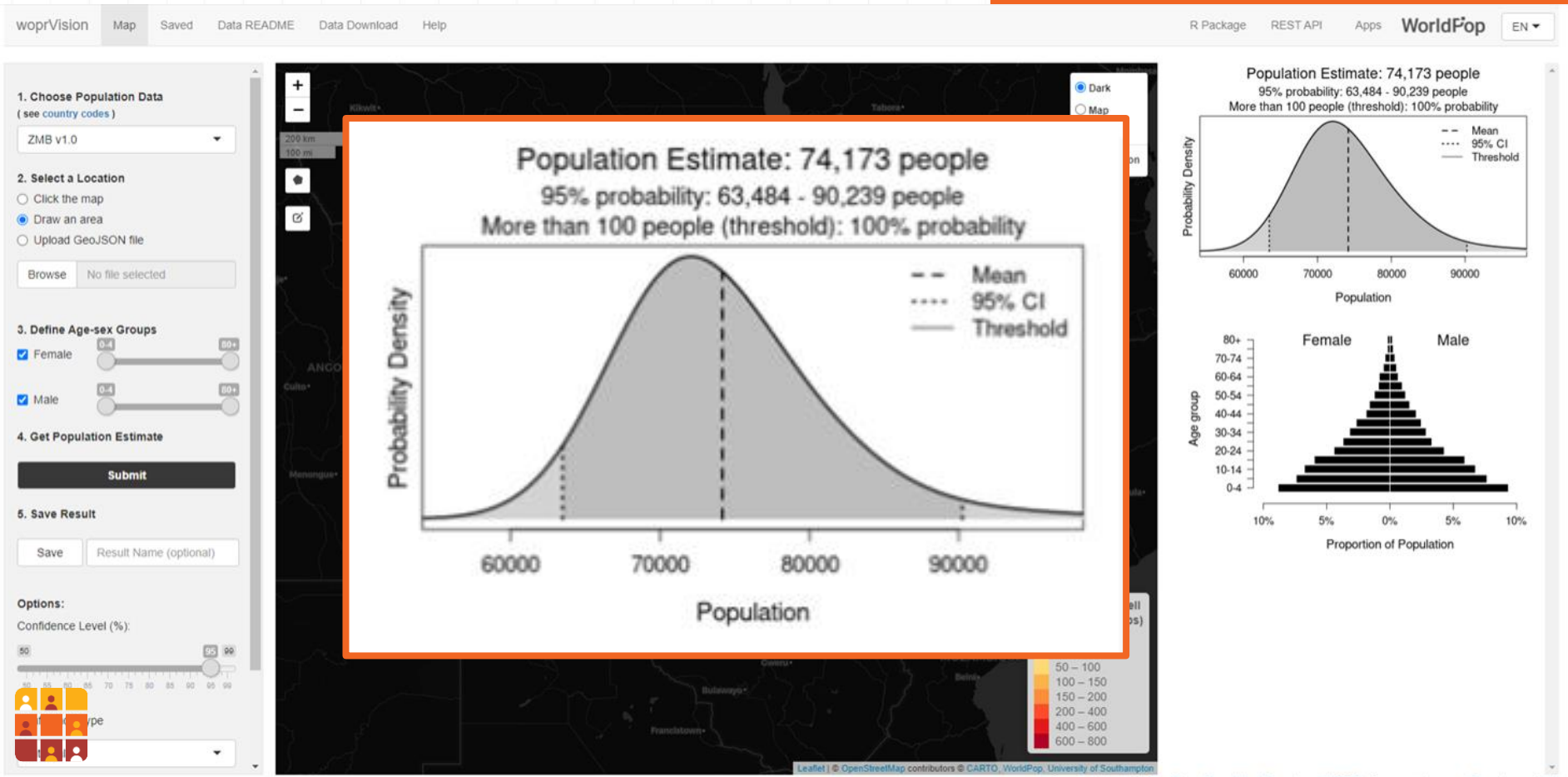
0.9+43.1+8.6+0.9+8.6+12.9+0.9+12.9+129.4+0+0.9+0.9

https://apps.worldpop.org/woprVision/



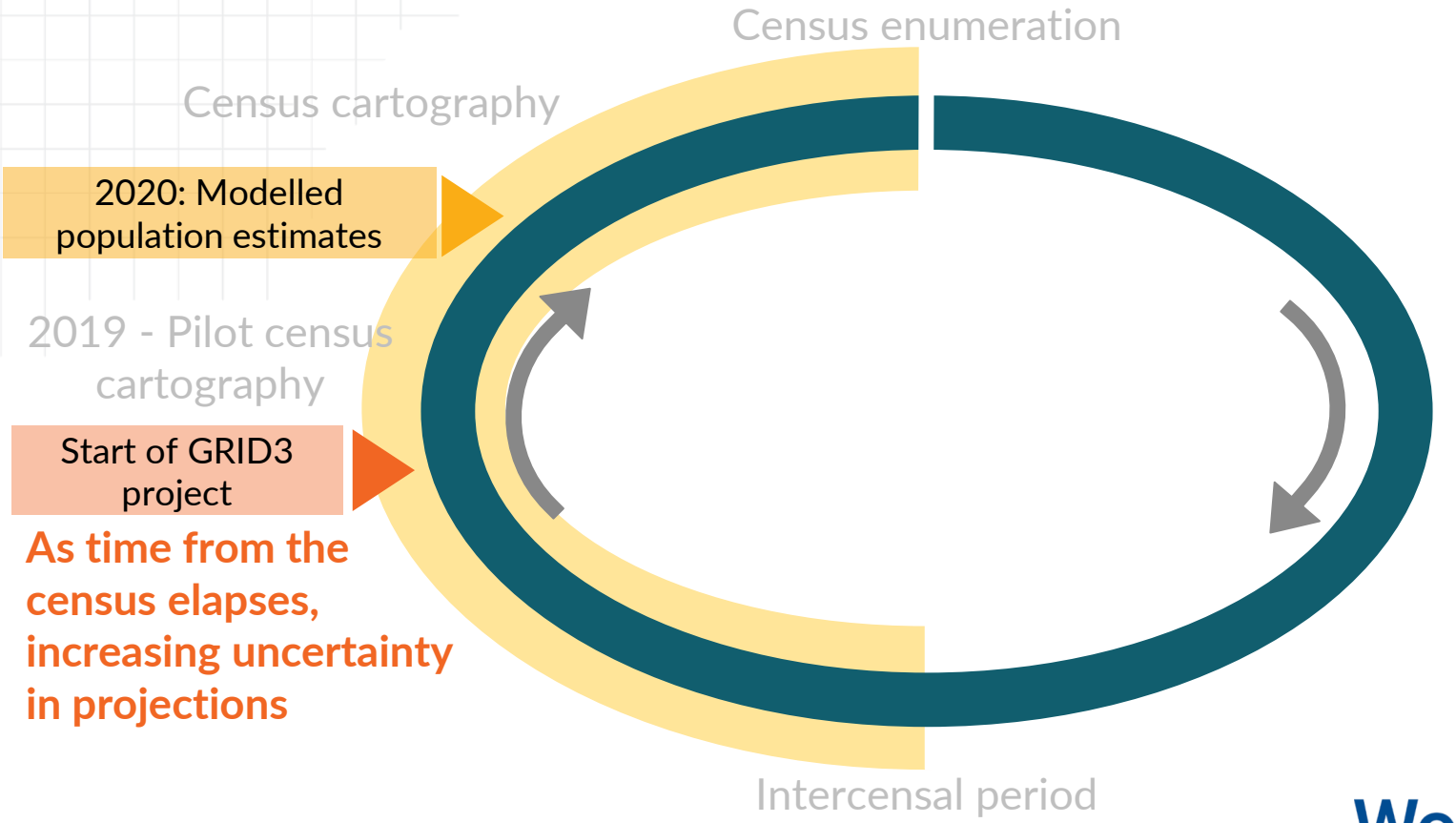
Source data: WorldPop Open Population Repository (WOPR), Source code: wopr R package v1.0.4

https://apps.worldpop.org/woprVision/



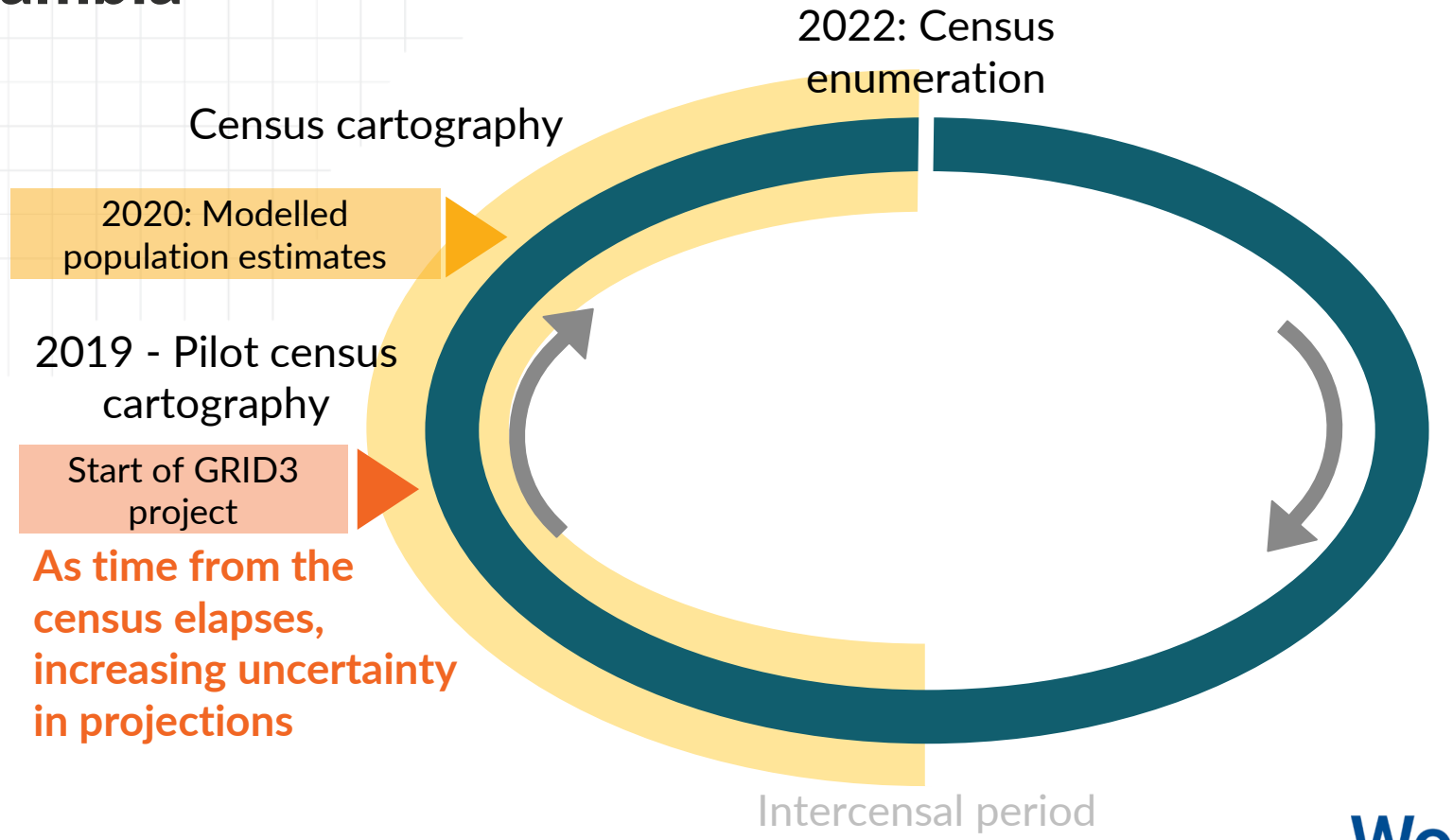
Source data: WorldPop Open Population Repository (WOPR), Source code: wopr R package v1.0.4

Zambia



WorldPop

Zambia



WorldPop

Zambia 2022 Population and Housing Census

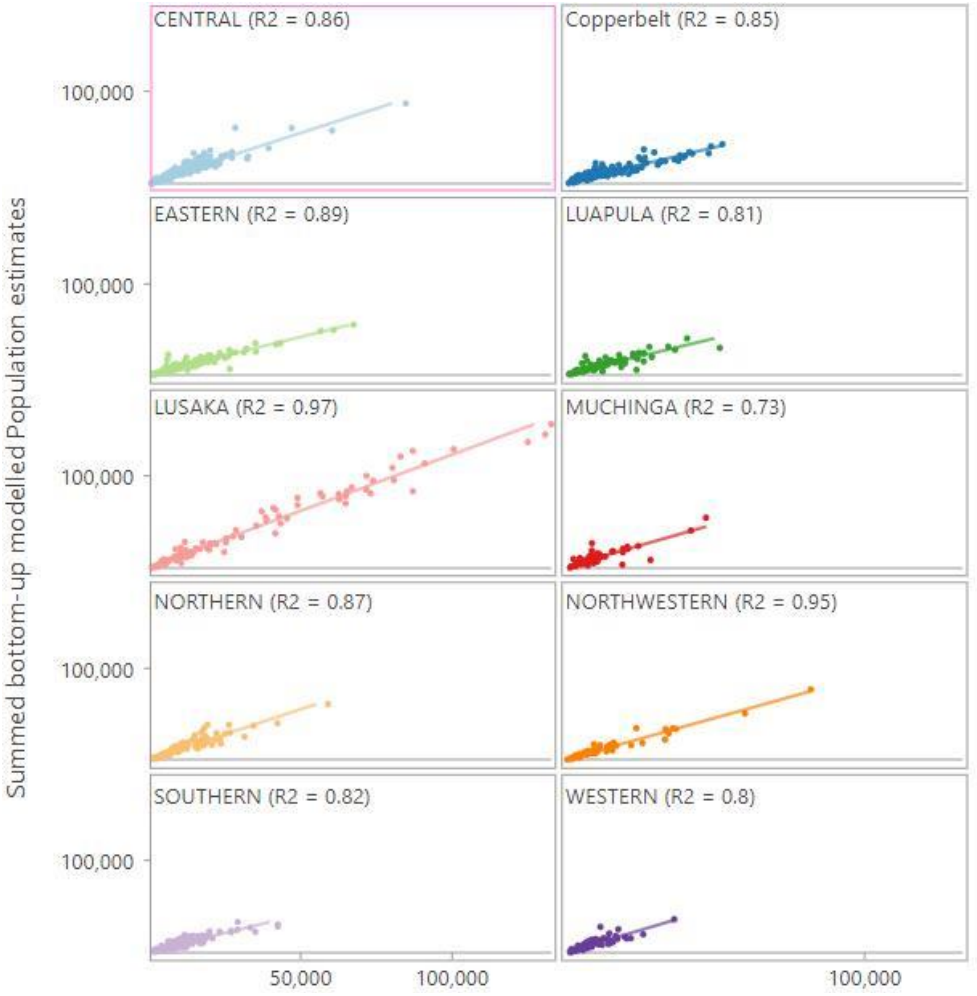
- Conducted August 18th – September 14th 2022
- National population count (preliminary): 19,610,769
- Hierarchy of administrative units in Zambia:

Admin level	Unit	Count
Admin level 1	Province	n = 10
Admin level 2	District	n = 116
Admin level 3	Constituency	n = 156
Admin level 4	Ward	n = 1858
Admin level 5	EAs	a lot….

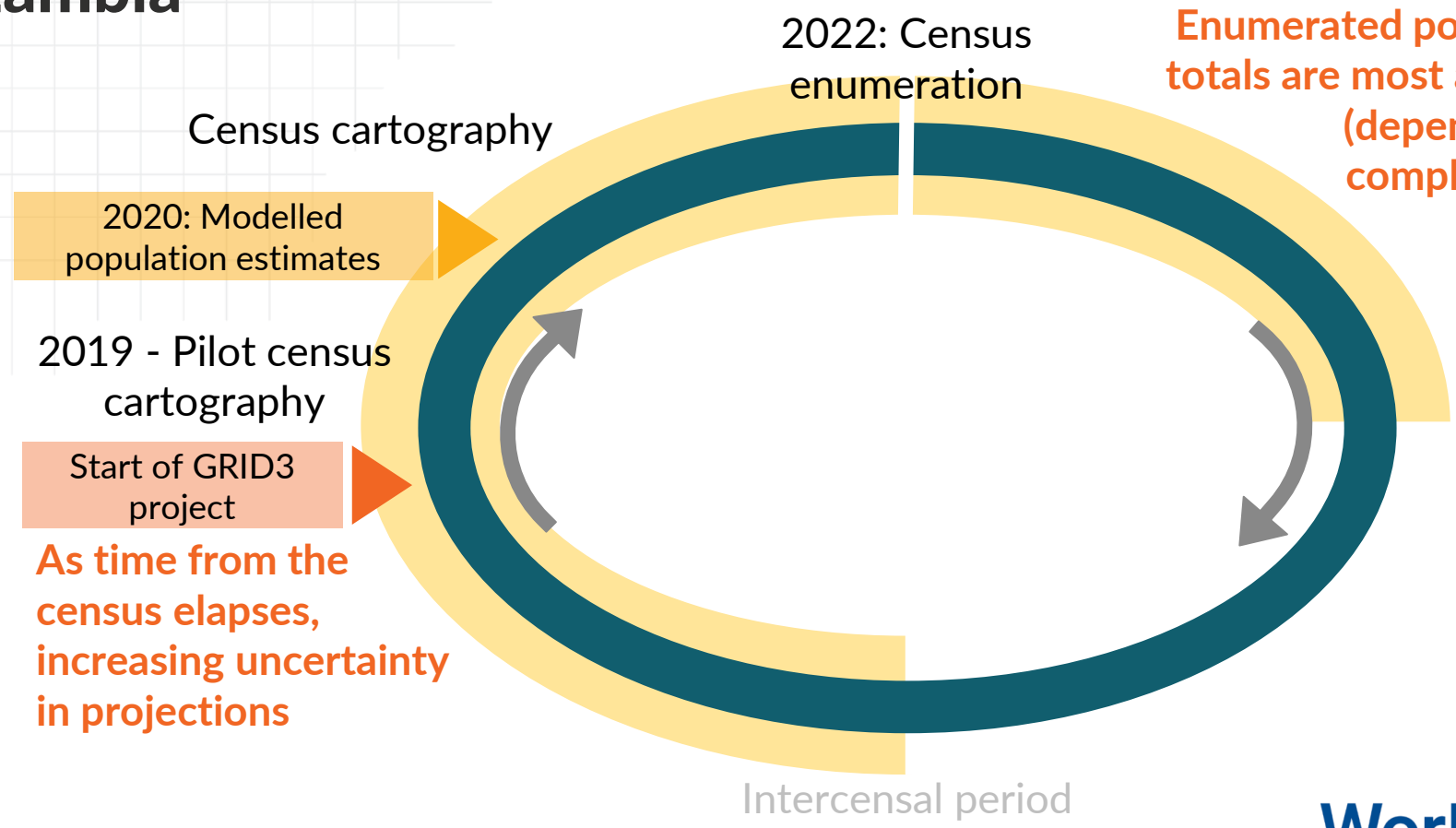
WorldPop

Bottom-up modelled population estimates vs. 2022 census

- Comparing population at ward level (1858 units nationally)
  - X-axis: 2022 enumerated population
  - Y-axis: summed grid cells in bottom-up modelled estimates – to provide ward-level estimates
  - One plot per province



Zambia



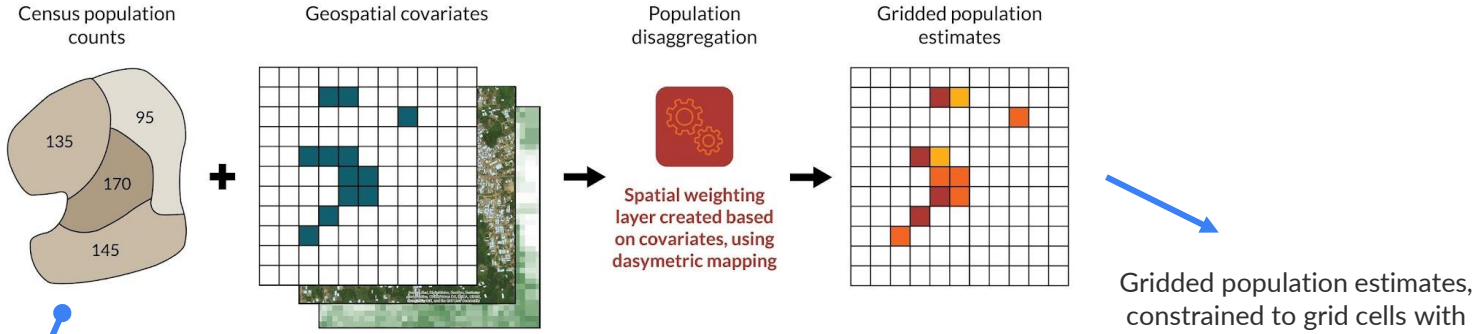
WorldPop

Zambia: gridded population estimates from 2022 census

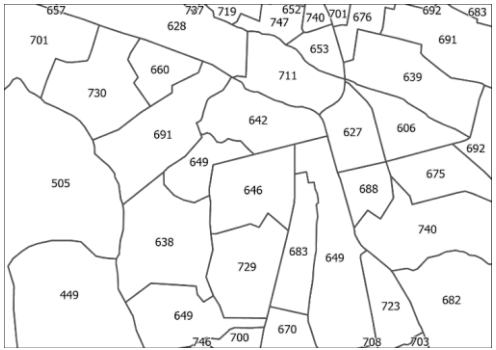
- Aim: produce gridded population estimates from 2022 census
  - Gridded output at 3 arc second (~100m) spatial resolution
  - Input population counts at ward-level (n=1858 units nationally, median population=7,249)
  - Spatially constrained, based on census household point dataset
  - "Validated" against household population counts from census

Co-produced between ZamStats and WorldPop

WorldPop

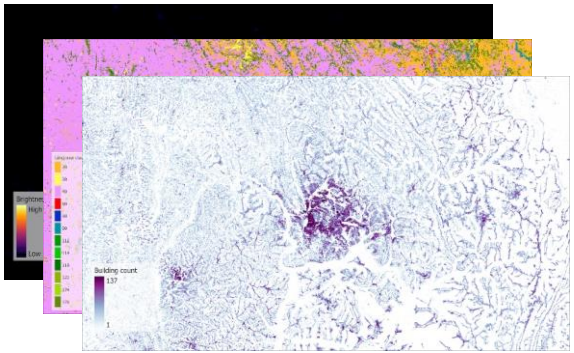


Ward boundaries with population counts



Household points (from census)

Geospatial covariates – including building footprint metrics



Gridded estimates – total population

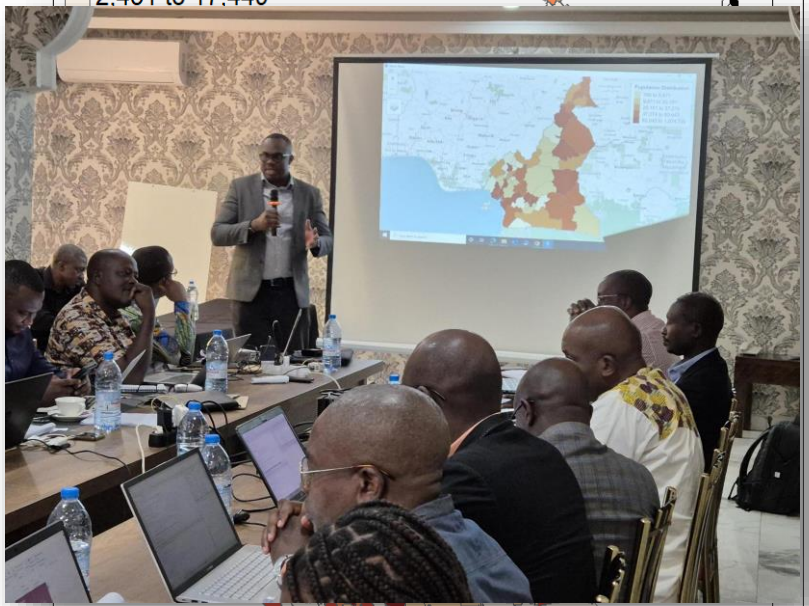
Population per grid cell (~100m)	Count of grid cells
1 - 10	1,150,392
11 - 20	224,813
21 - 50	98,089
51 - 100	30,983
101 - 200	15,490
>200	4,396
TOTAL	**1,525,866**

Summary statistic	Value
Minimum	1.4
Maximum	358.7
Mean	12.9
Median	7.3
Standard deviation	21.2
Sum	19,610,769

Additional examples....

WorldPop

Cameroon: National household survey listings

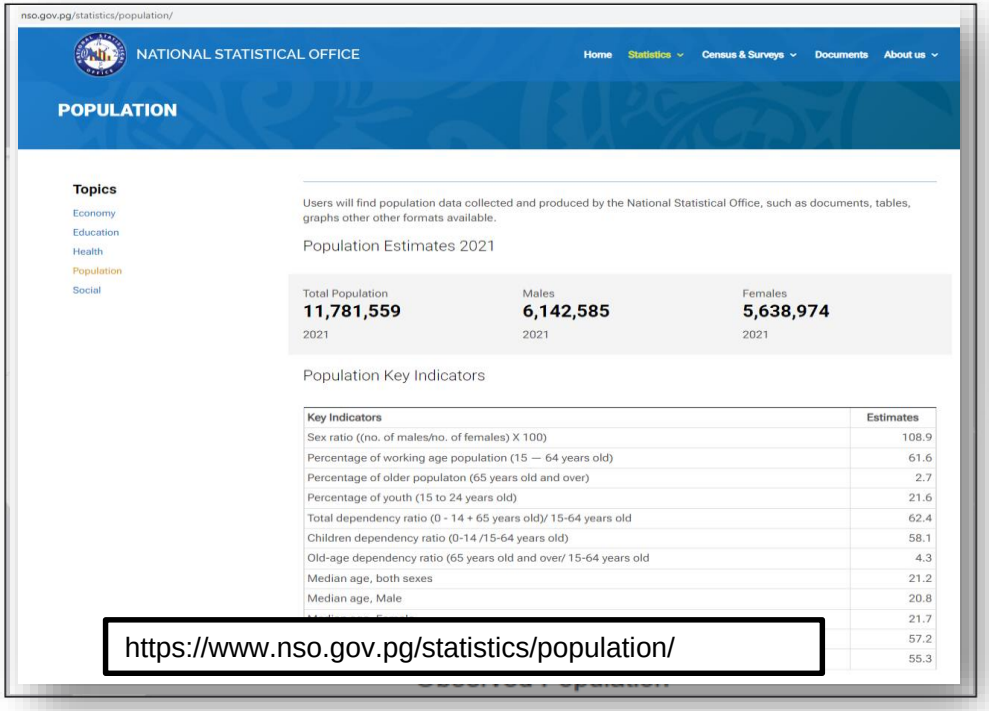


INSTITUT NATIONAL DE LA STATISTIQUE
NS
NATIONAL INSTITUTE OF STATISTICS

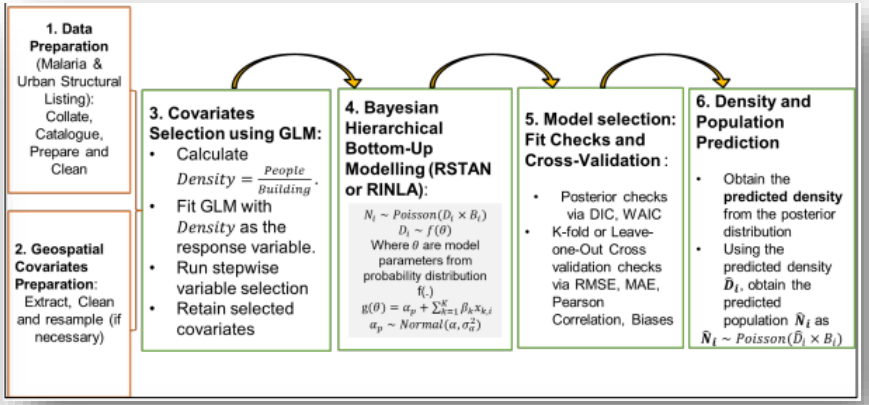
UNFPA

GRID3
GEO-REFERENCED INFRASTRUCTURE AND DEMOGRAPHIC DATA FOR DEVELOPMENT

Papua New Guinea: Malaria survey household listings



https://www.nso.gov.pg/statistics/population/



Metric	In-Sample Cross-Validation				Out-sample Cross-Validation	
	BHM (Count)	TSBHM (Count)	BHM (Density)	TSBHM (Density)	BHM (Density)	TSBHM (Density)
MAE	3.191	2.251	0.062	0.005	0.670	0.007
RMSE	5.734	4.213	1.891	0.220	14.716	0.172
Abias	3.142	2.123	0.054	0.003	0.665	0.003
CC	0.991	0.991	0.992	0.792	0.866	0.862

https://wopr.worldpop.org/?PNG/Population; *Nnanatu et al (2024) in review*