

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

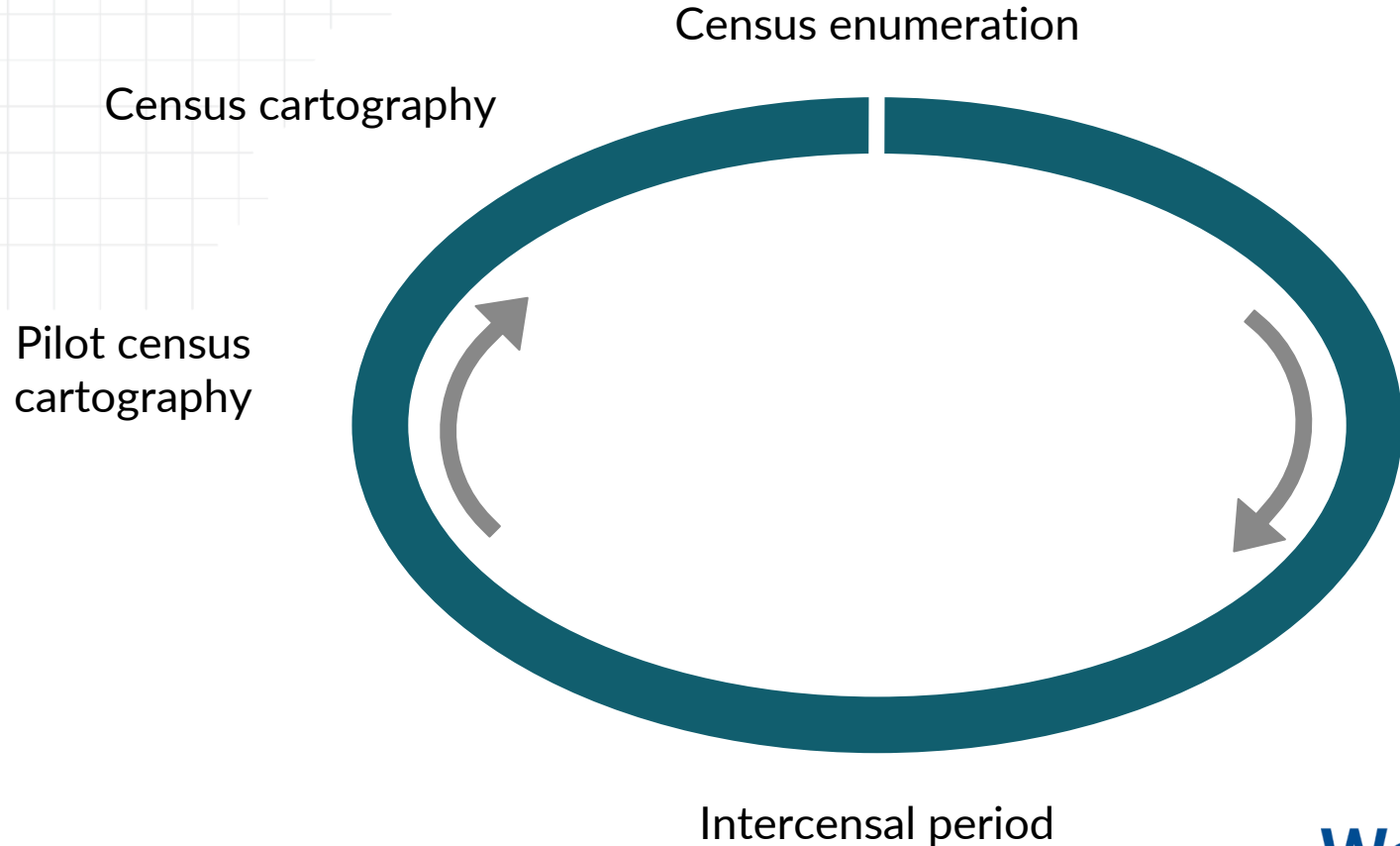
Day 2 – Case Study: Zambia

WorldPop

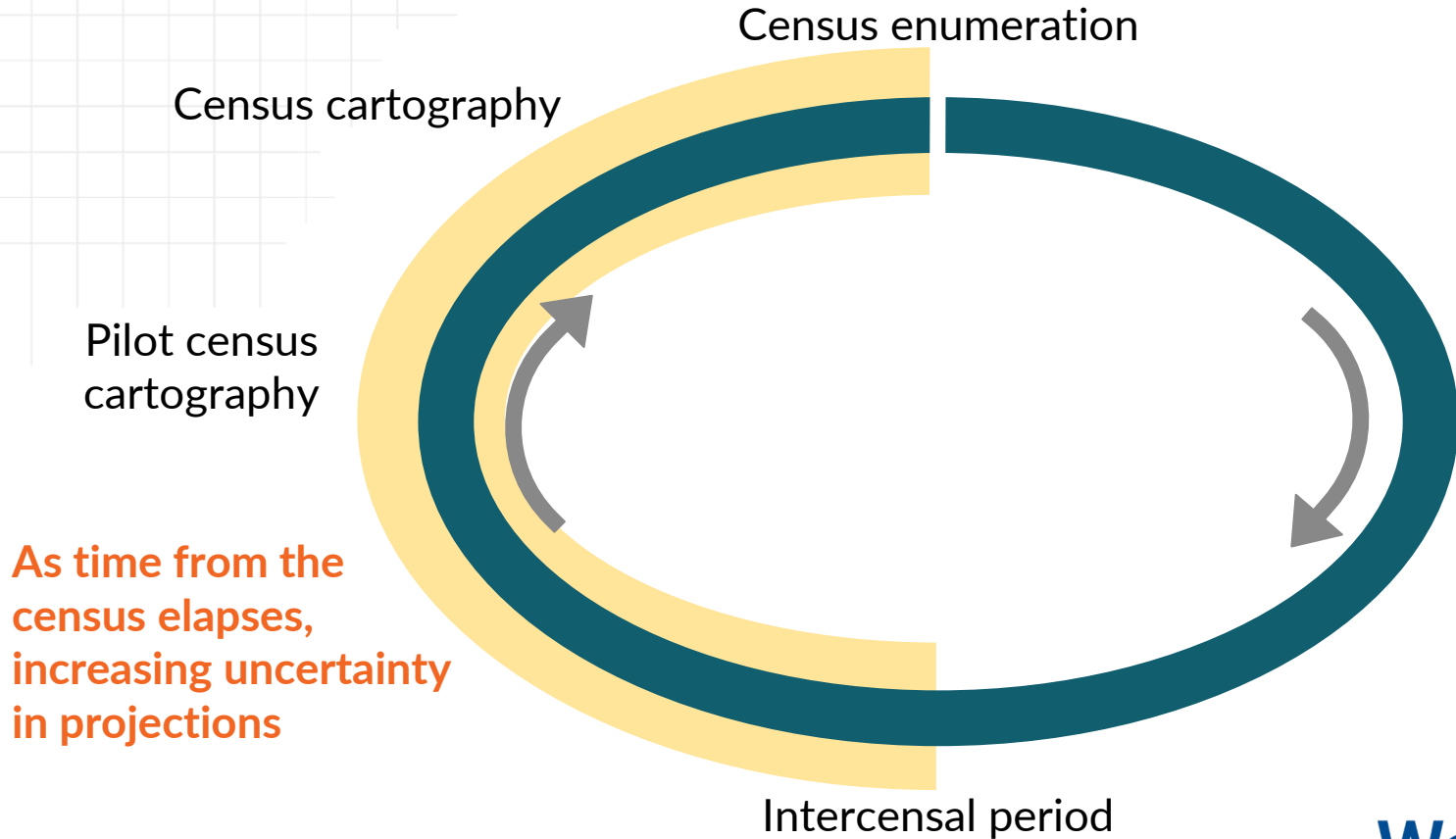
Background: Zambia

- 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

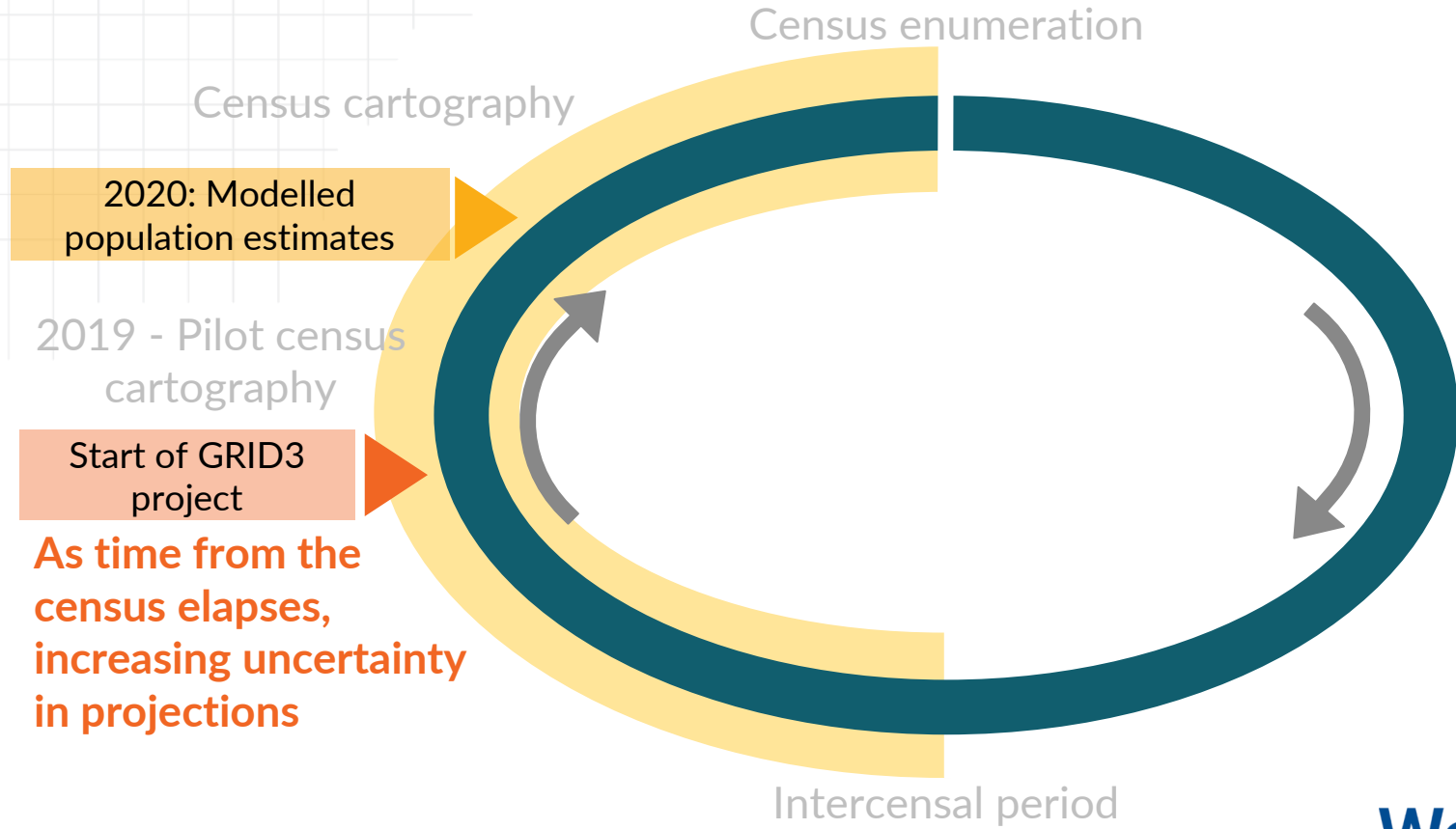
Stages of the census cycle



Stages of the census cycle



Stages of the census cycle: Zambia



Input datasets

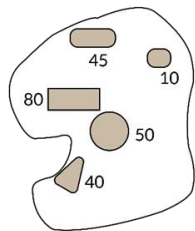
Population counts

3 sources of population counts:

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

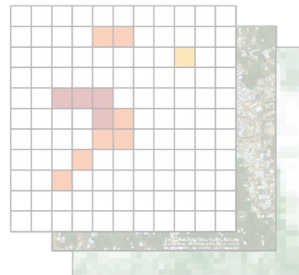
All survey datasets included **counts of population** enumerated for **each household**, with household or building locations (**latitude and longitude**) recorded using **GPS-enabled devices**.

Population counts from household surveys



+

Geospatial covariates

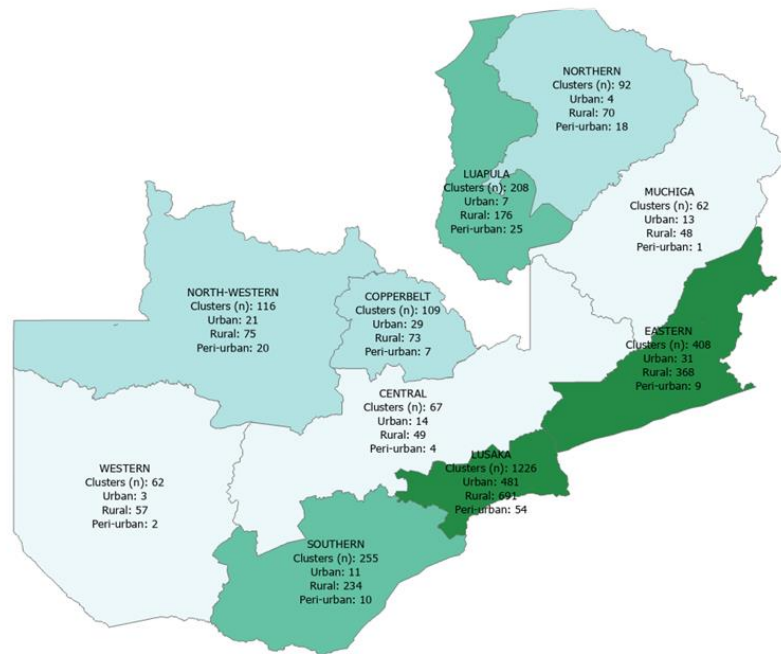
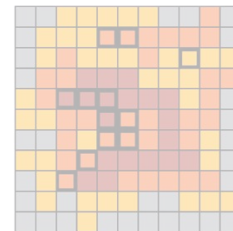


Population estimation



Prediction of population in all grid cells (including unsurveyed areas)

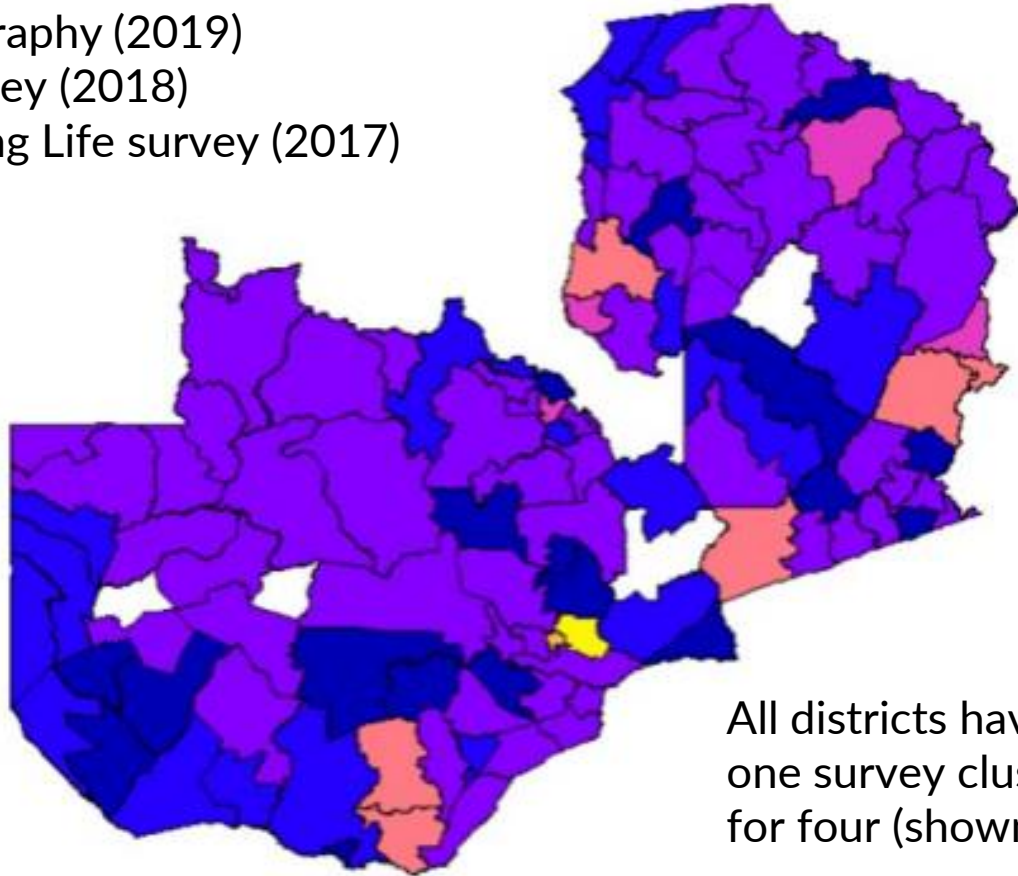
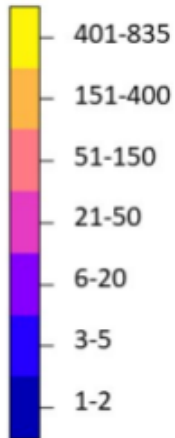
Gridded population estimates



3 sources of population counts:

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

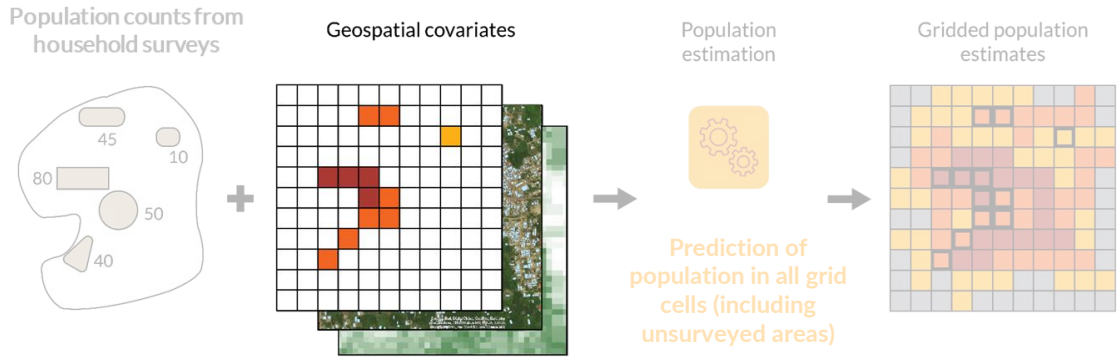
Counts of survey clusters per district



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

Input datasets

Geospatial covariates



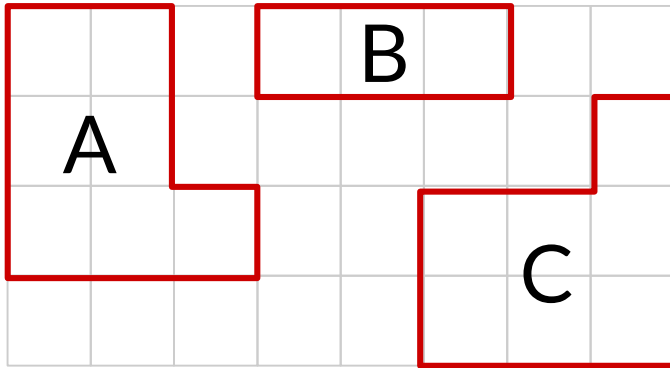
- Predictive variables related to population density, and associated with the **built** and **natural environment** are considered, such as building density and vegetation cover.
- Geospatial covariates are **spatially harmonised** so that all variables have the same spatial resolution and grid cell 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



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 ($>750\text{m}^2$) filtered out.

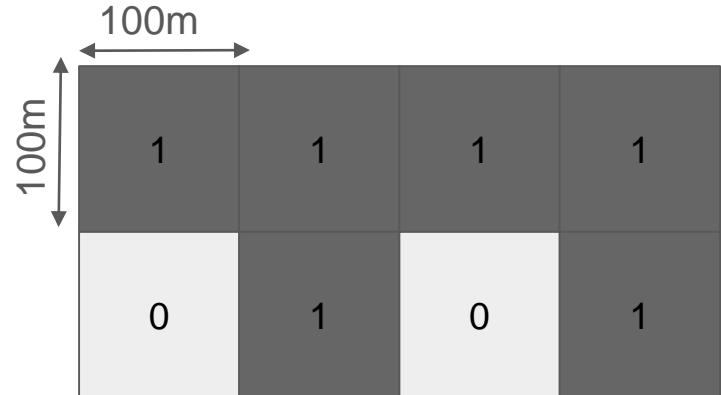
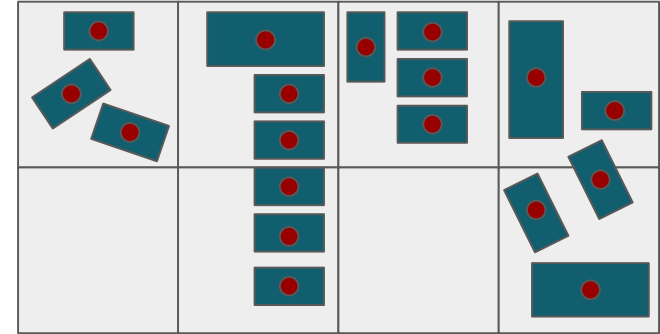
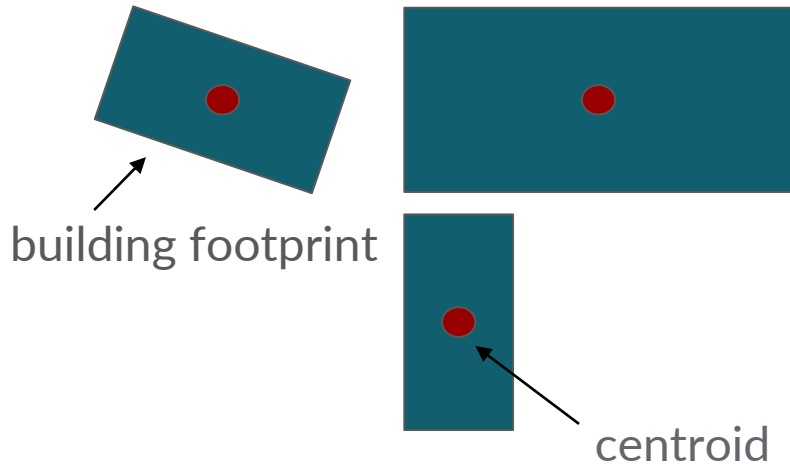


Input datasets

Geospatial covariates - settlement

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

- Based on building footprint centroid



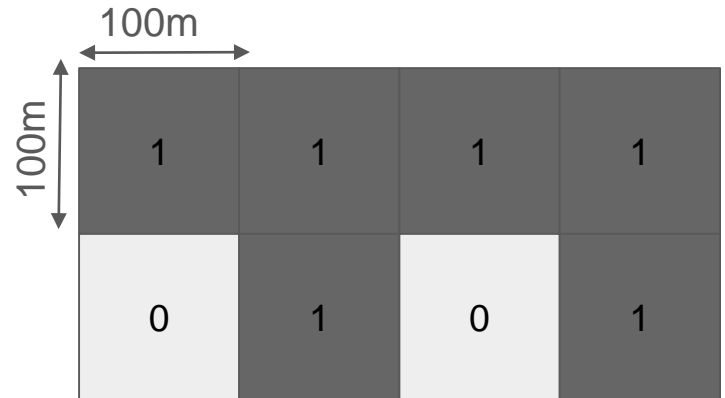
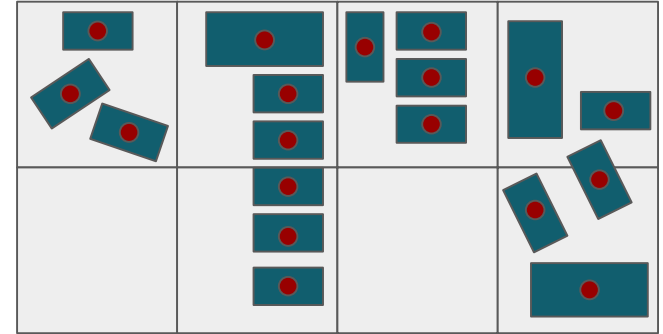
Input datasets

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:

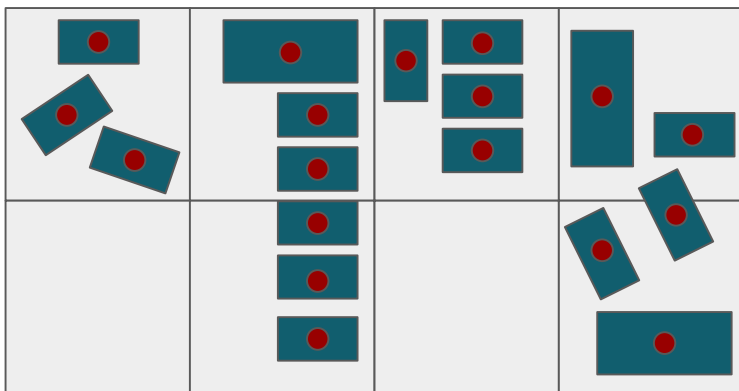
- Rural: clumps of <500 grid cells
- Small urban: 500-1500 contiguous grid cells
- Large urban: - >1500 contiguous grid cells



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

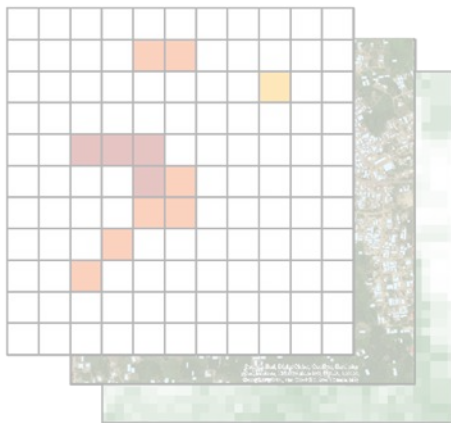
Maximum
building
perimeter

The “bottom-up” approach

Population counts from household surveys



Geospatial covariates



+



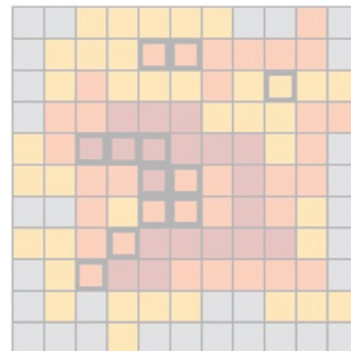
Population estimation



Prediction of population in all grid cells (including unsurveyed areas)



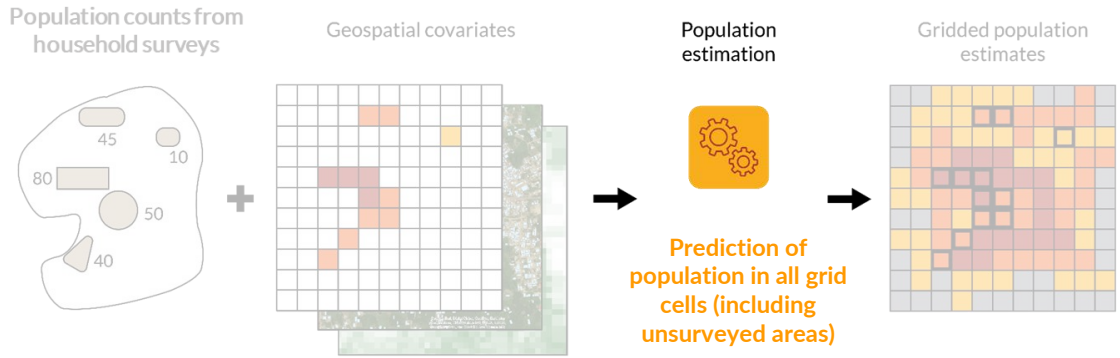
Gridded population estimates



Input datasets

Population estimation

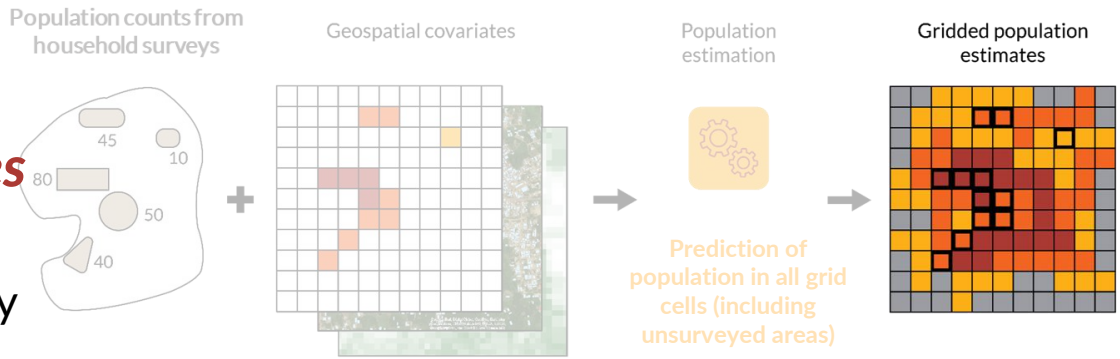
Bayesian statistical model



- **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

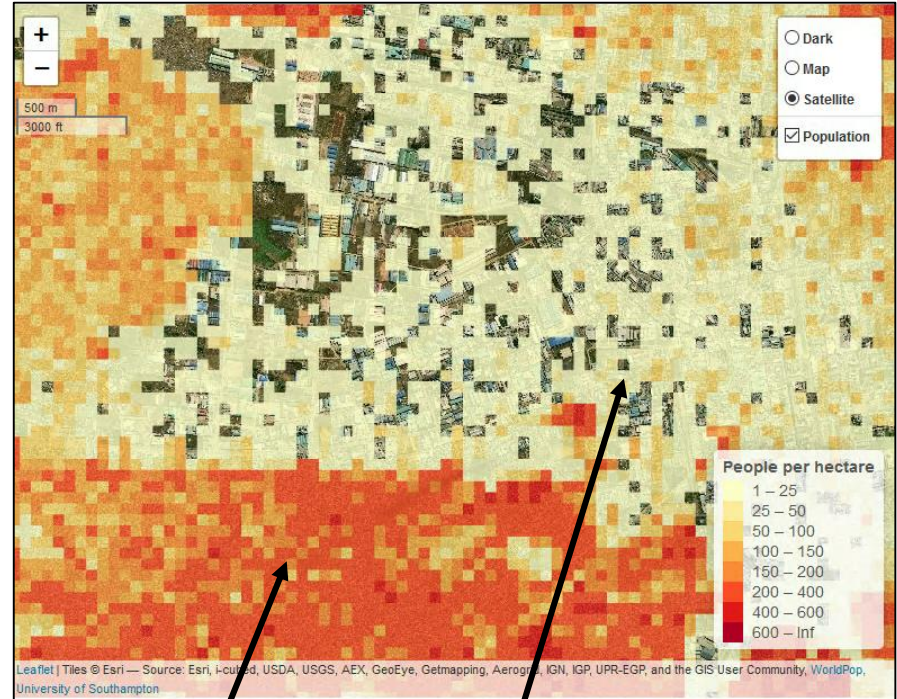


- Output from model for every ~100m x 100m grid cell
- For every grid cell, we calculated the **probability distribution of population counts** (i.e. posterior predictions), using a grid cell's covariate values and the parameterised model
- The posterior predictions provide, for every grid cell, a **mean 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:

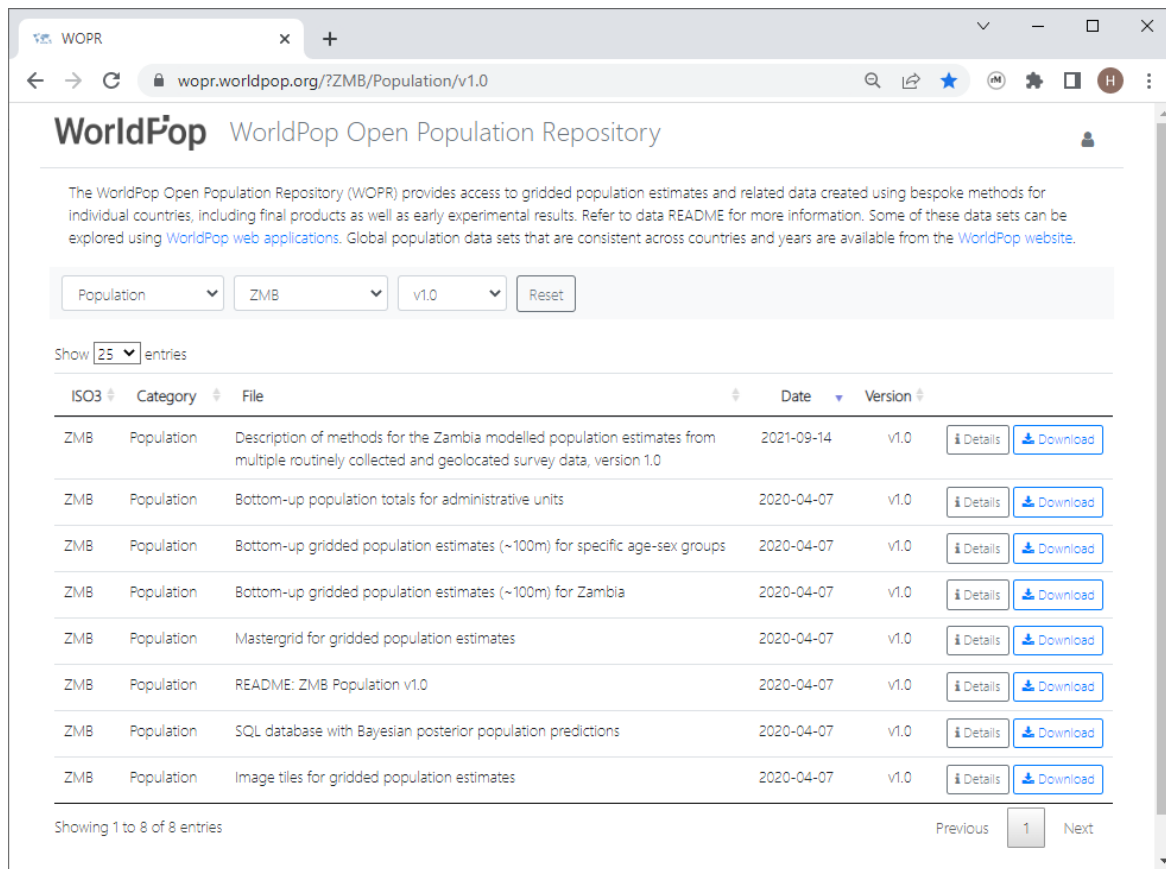
- WOPR: <https://wopr.worldpop.org/>
- GRID3 data hub:
<https://data.grid3.org/maps/grid3-zambia-gridded-population-estimates-version-1-0/>

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>



The screenshot shows the WorldPop Open Population Repository website. The page title is "WorldPop WorldPop Open Population Repository". Below the title, there is a description of the repository. A search bar contains the filters "Population", "ZMB", and "v1.0", with a "Reset" button. Below the search bar, it says "Show 25 entries". A table lists 8 entries with columns for ISO3, Category, File, Date, and Version. Each entry has "Details" and "Download" buttons.

ISO3	Category	File	Date	Version		
ZMB	Population	Description of methods for the Zambia modelled population estimates from multiple routinely collected and geolocated survey data, version 1.0	2021-09-14	v1.0	Details	Download
ZMB	Population	Bottom-up population totals for administrative units	2020-04-07	v1.0	Details	Download
ZMB	Population	Bottom-up gridded population estimates (~100m) for specific age-sex groups	2020-04-07	v1.0	Details	Download
ZMB	Population	Bottom-up gridded population estimates (~100m) for Zambia	2020-04-07	v1.0	Details	Download
ZMB	Population	Mastergrid for gridded population estimates	2020-04-07	v1.0	Details	Download
ZMB	Population	README: ZMB Population v1.0	2020-04-07	v1.0	Details	Download
ZMB	Population	SQL database with Bayesian posterior population predictions	2020-04-07	v1.0	Details	Download
ZMB	Population	Image tiles for gridded population estimates	2020-04-07	v1.0	Details	Download

Showing 1 to 8 of 8 entries

Previous 1 Next

Can download:

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

1. Choose Population Data
(see country codes)

ZMB v1.0

2. Select a Location

- Click the map
- Draw an area
- Upload GeoJSON file

Browse No file selected

3. Define Age-sex Groups

- Female 0-4 80+
- Male 0-4 80+

4. Get Population Estimate

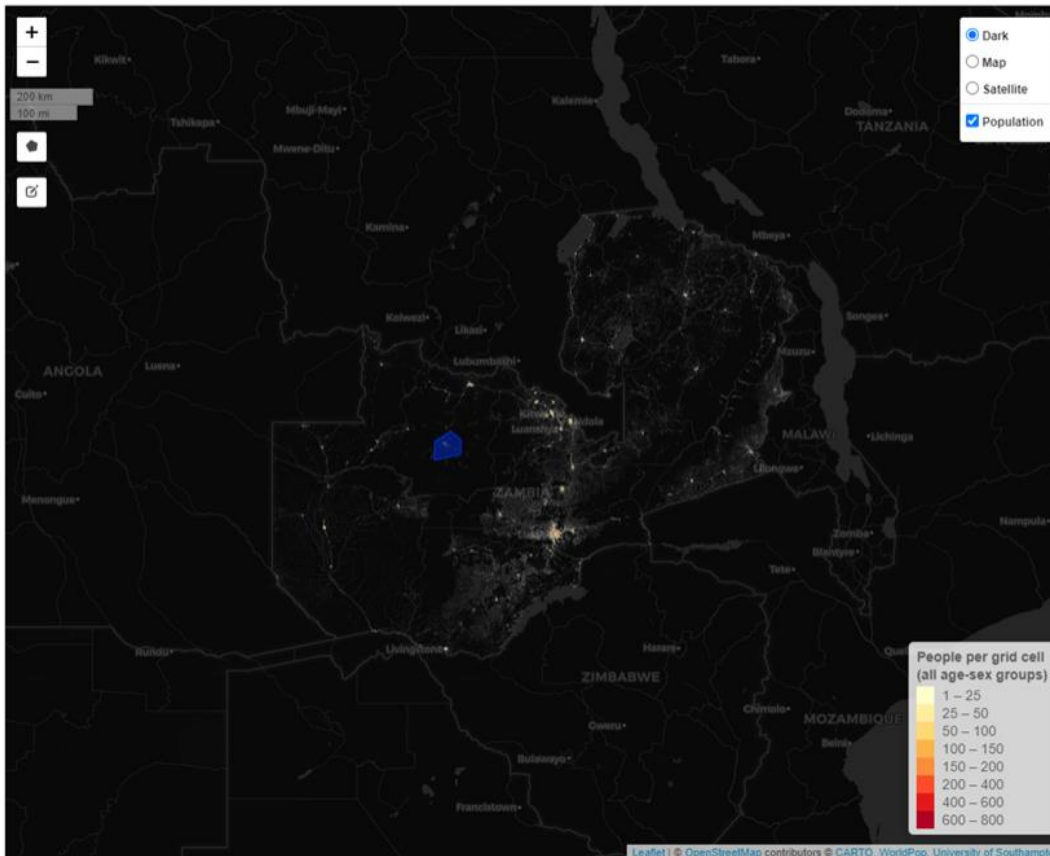
Submit

5. Save Result

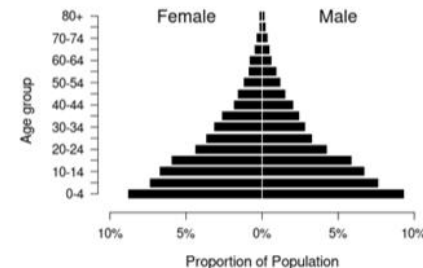
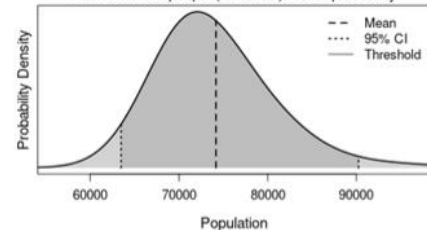
Save Result Name (optional)

Options:

Confidence Level (%)



Population Estimate: 74,173 people
95% probability: 63,484 - 90,239 people
More than 100 people (threshold): 100% probability



1. Choose Population Data
(see country codes)

ZMB v1.0

2. Select a Location

- Click the map
- Draw an area
- Upload GeoJSON file

Browse No file selected

3. Define Age-sex Groups

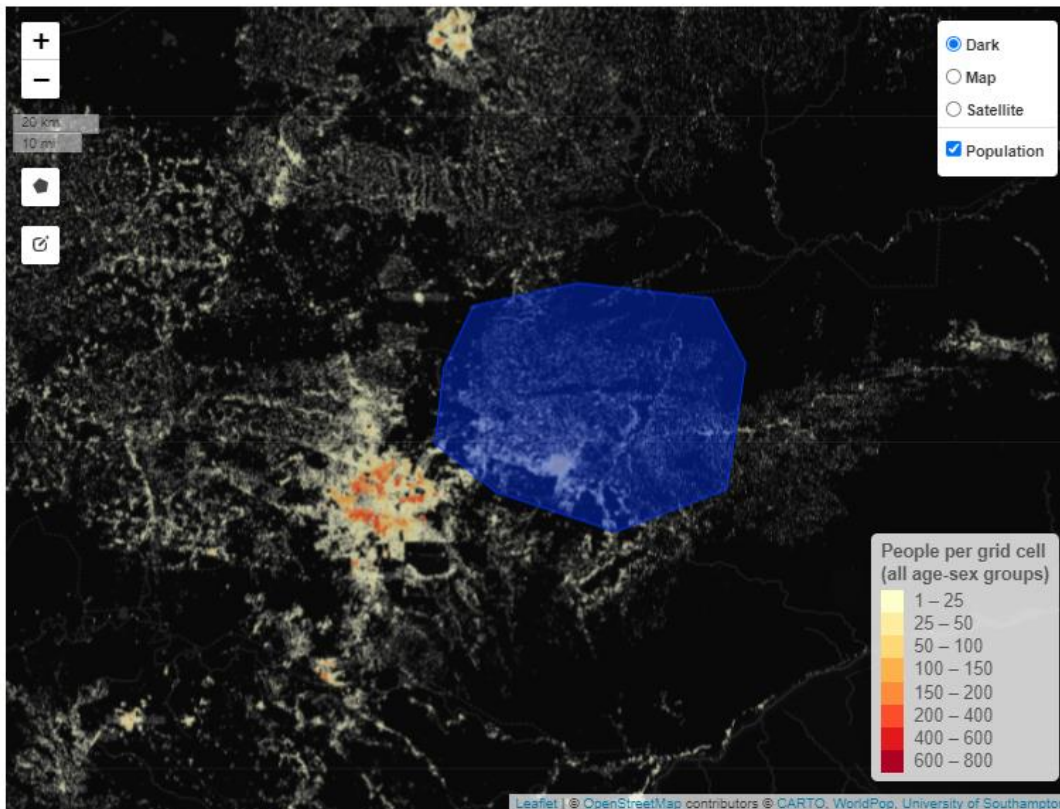
- Female 0-4 80+
- Male 0-4 80+

4. Get Population Estimate

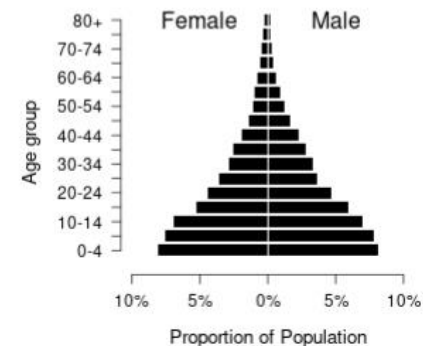
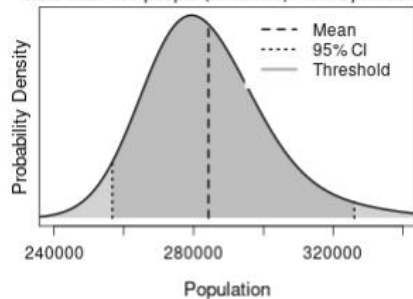
Submit

5. Save Result

Result Name (opti)



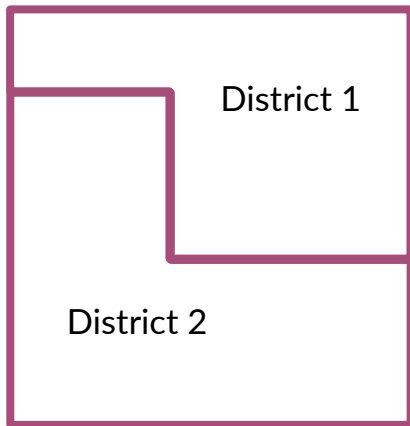
Population Estimate: 284,264 people
95% probability: 256,719 - 326,054 people
More than 100 people (threshold): 100% probability



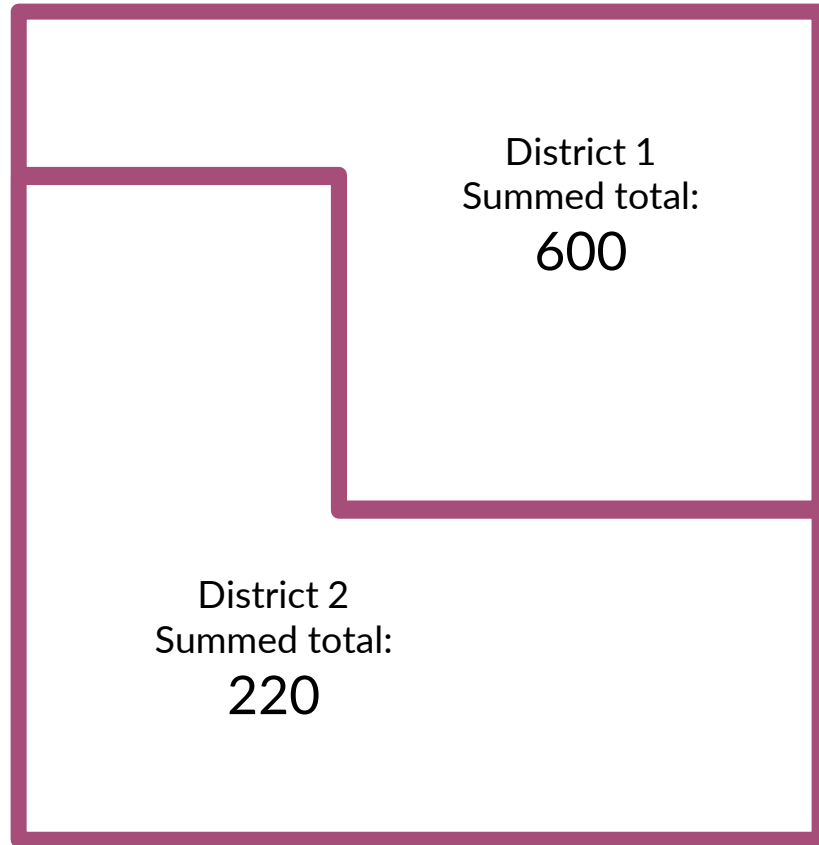
Calculating aggregated population estimates

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:

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

1. Choose Population Data
(see country codes)

ZMB v1.0

2. Select a Location

- Click the map
- Draw an area
- Upload GeoJSON file

Browse No file selected

3. Define Age-sex Groups

- Female 0-4 80+
- Male 0-4 80+

4. Get Population Estimate

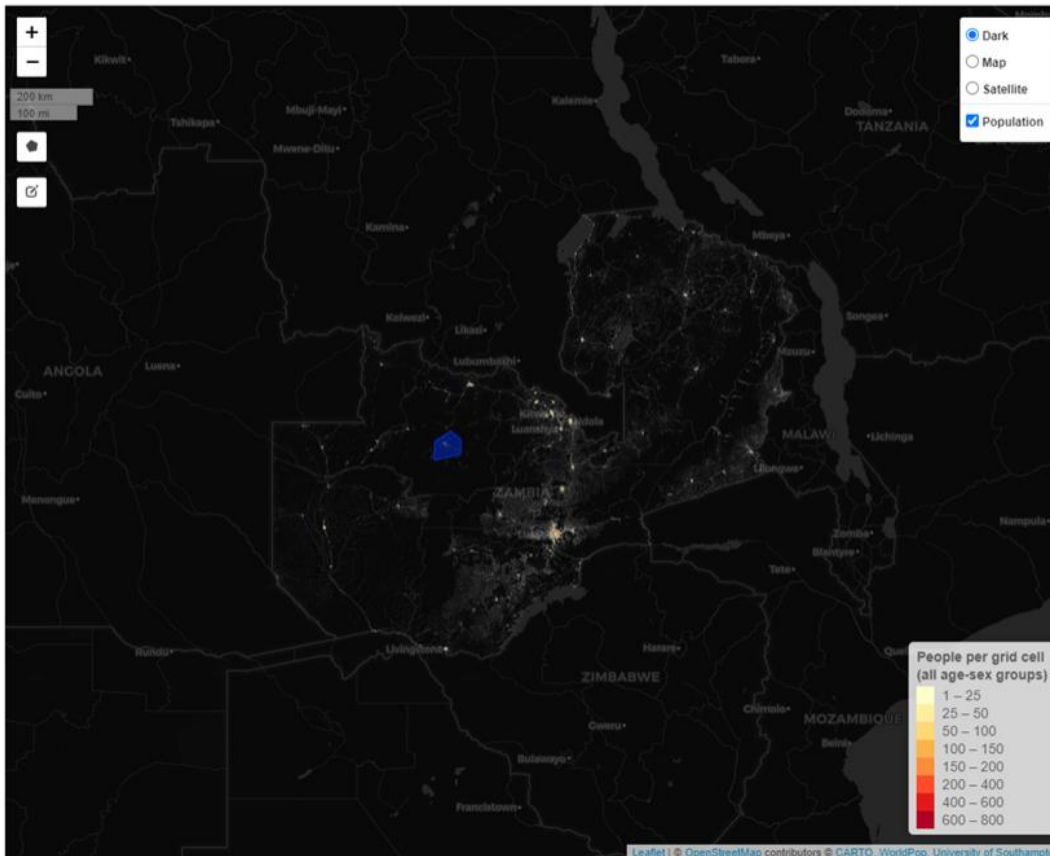
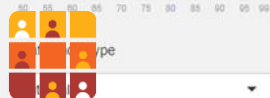
Submit

5. Save Result

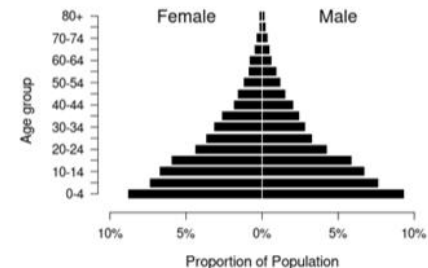
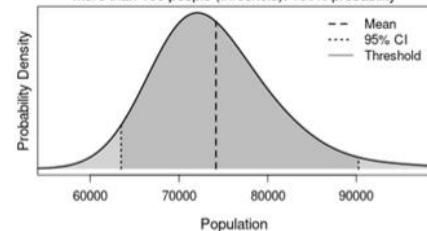
Save Result Name (optional)

Options:

Confidence Level (%)



Population Estimate: 74,173 people
95% probability: 63,484 - 90,239 people
More than 100 people (threshold): 100% probability



1. Choose Population Data
(see country codes)

ZMB v1.0

2. Select a Location

- Click the map
- Draw an area
- Upload GeoJSON file

Browse No file selected

3. Define Age-sex Groups

- Female 0-4 80+
- Male 0-4 80+

4. Get Population Estimate

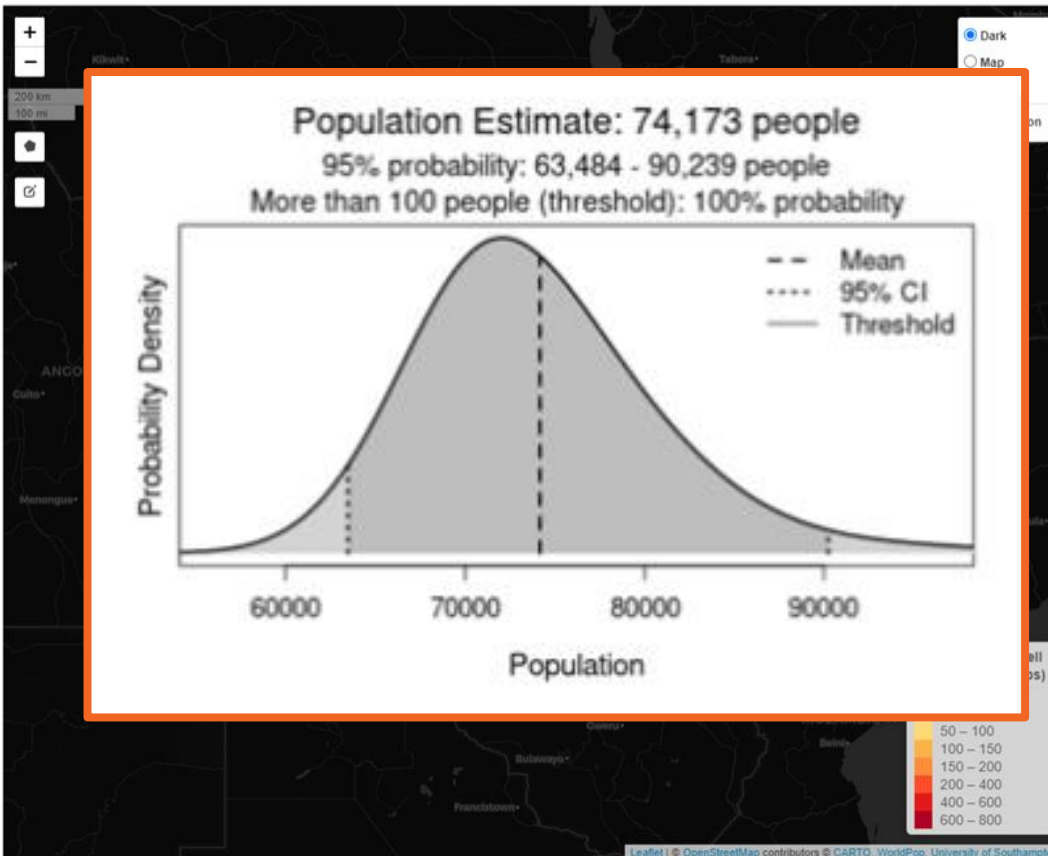
Submit

5. Save Result

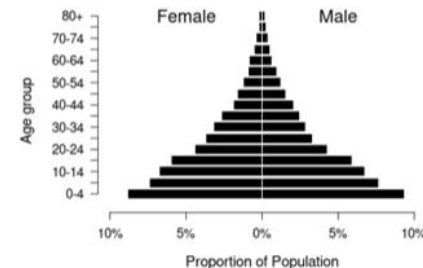
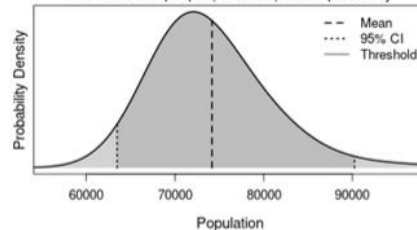
Save Result Name (optional)

Options:

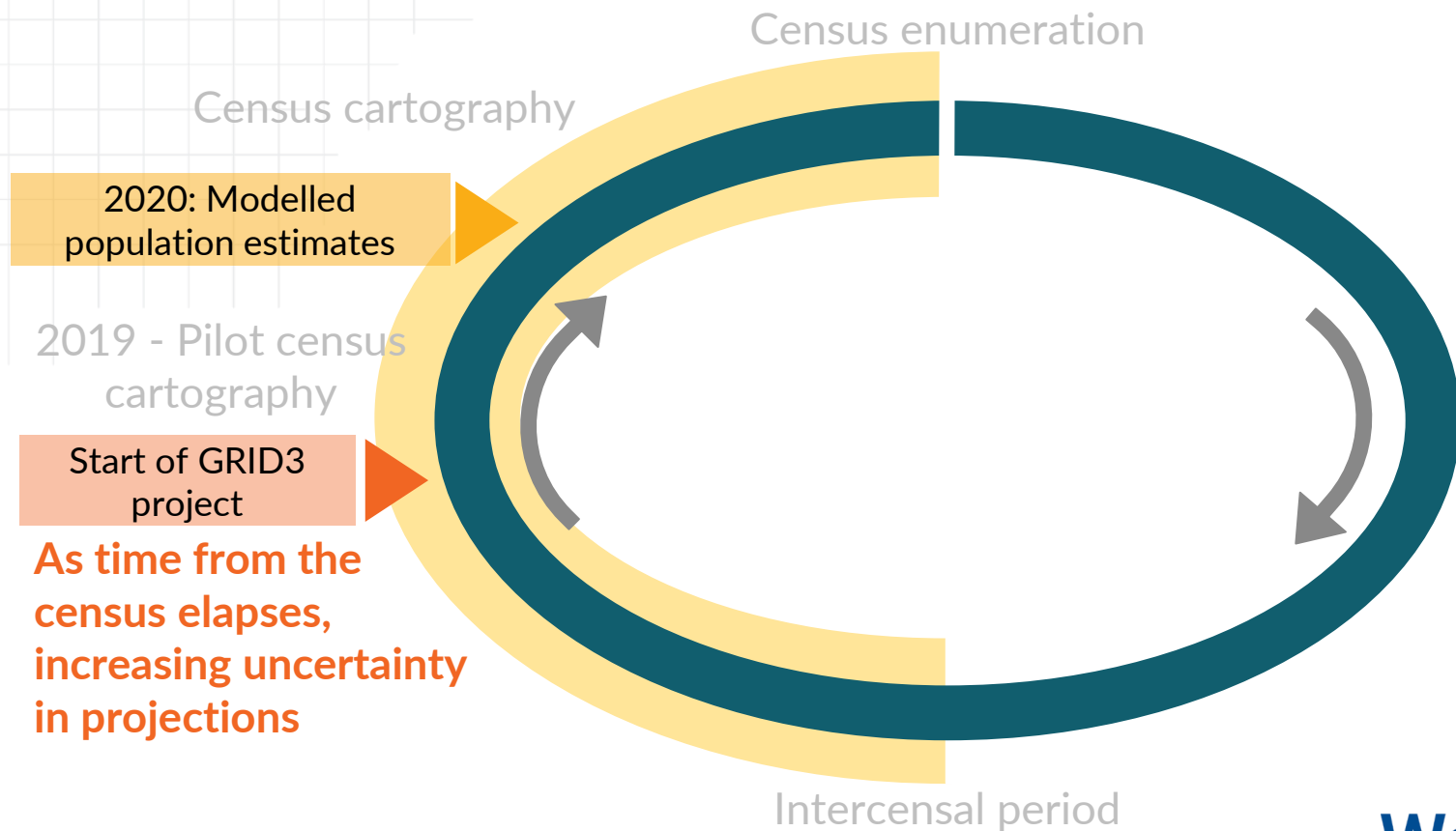
Confidence Level (%)



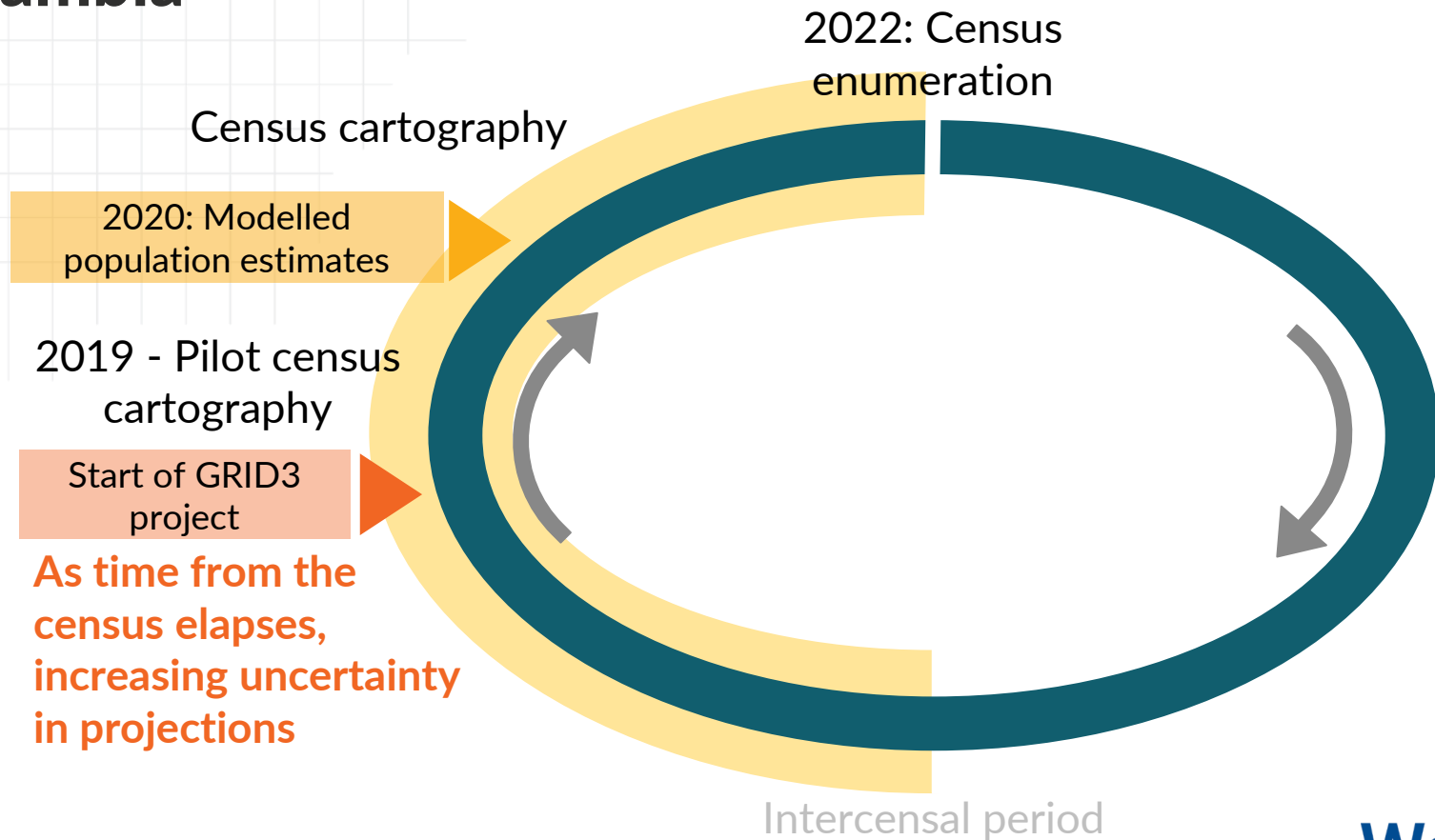
Population Estimate: 74,173 people
95% probability: 63,484 - 90,239 people
More than 100 people (threshold): 100% probability



Zambia



Zambia



Zambia 2022 Population and Housing Census

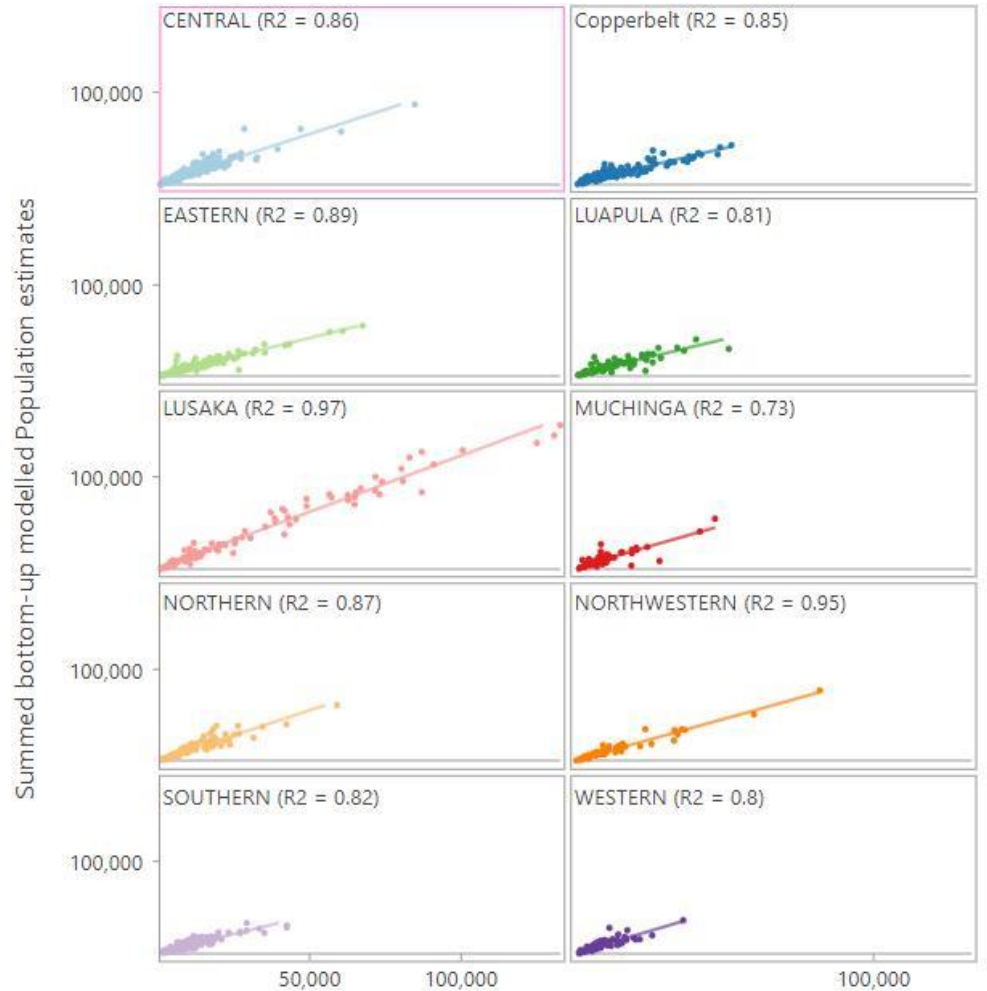
- Conducted August 18th – September 14th 2022
- National population count (preliminary): 19,610,769
- Hierarchy of administrative units in Zambia:

<u>Admin level</u>	<u>Unit</u>	<u>Count</u>
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....

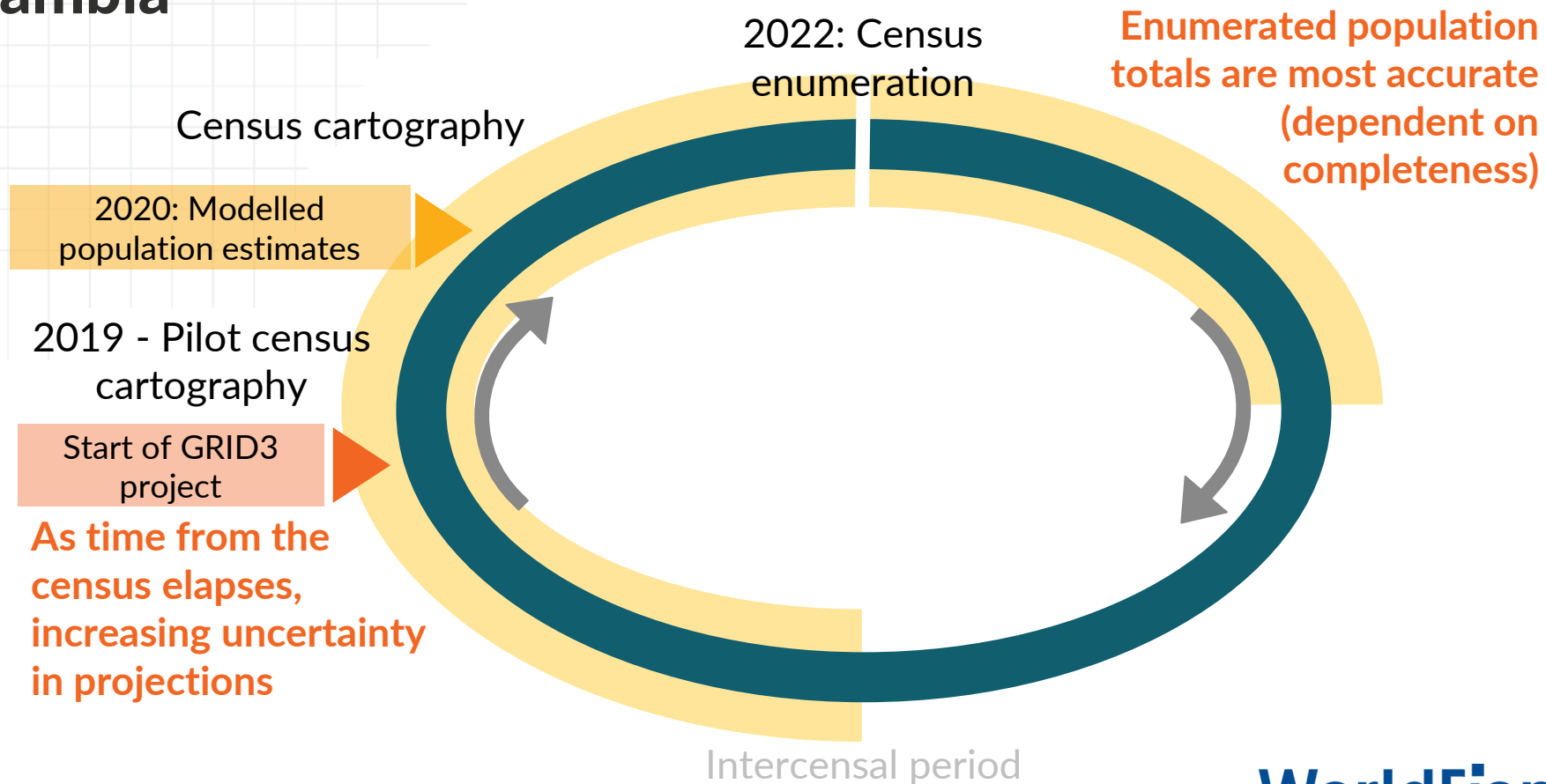


Bottom-up modelled population estimates vs. 2022 census

- 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

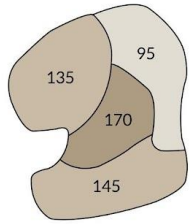


Zambia: gridded population estimates from 2022 census

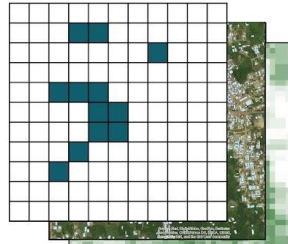
- Aim: produce gridded population estimates from 2022 census
 - Gridded output at 3 arc second (~100m) spatial resolution
 - Input population counts at ward-level (n=1858 units nationally, median population=7,249)
 - Spatially constrained, based on census household point dataset
 - "Validated" against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

Co-produced between ZamStats and WorldPop

Census population counts



Geospatial covariates

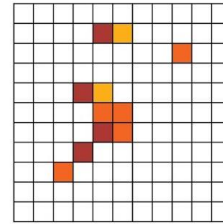


Population disaggregation



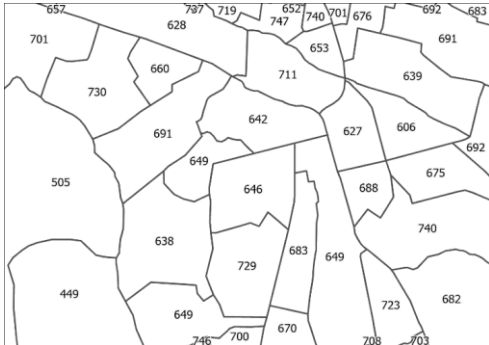
Spatial weighting layer created based on covariates, using dasymetric mapping

Gridded population estimates



Gridded population estimates, constrained to grid cells with households

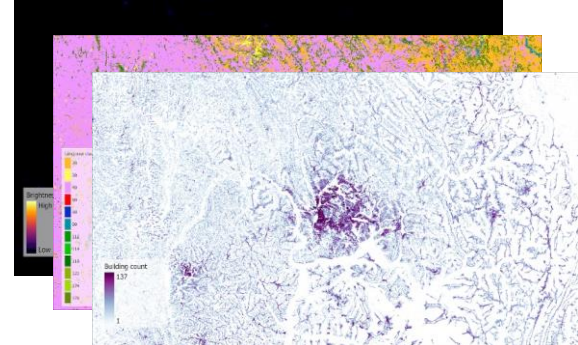
Ward boundaries with population counts



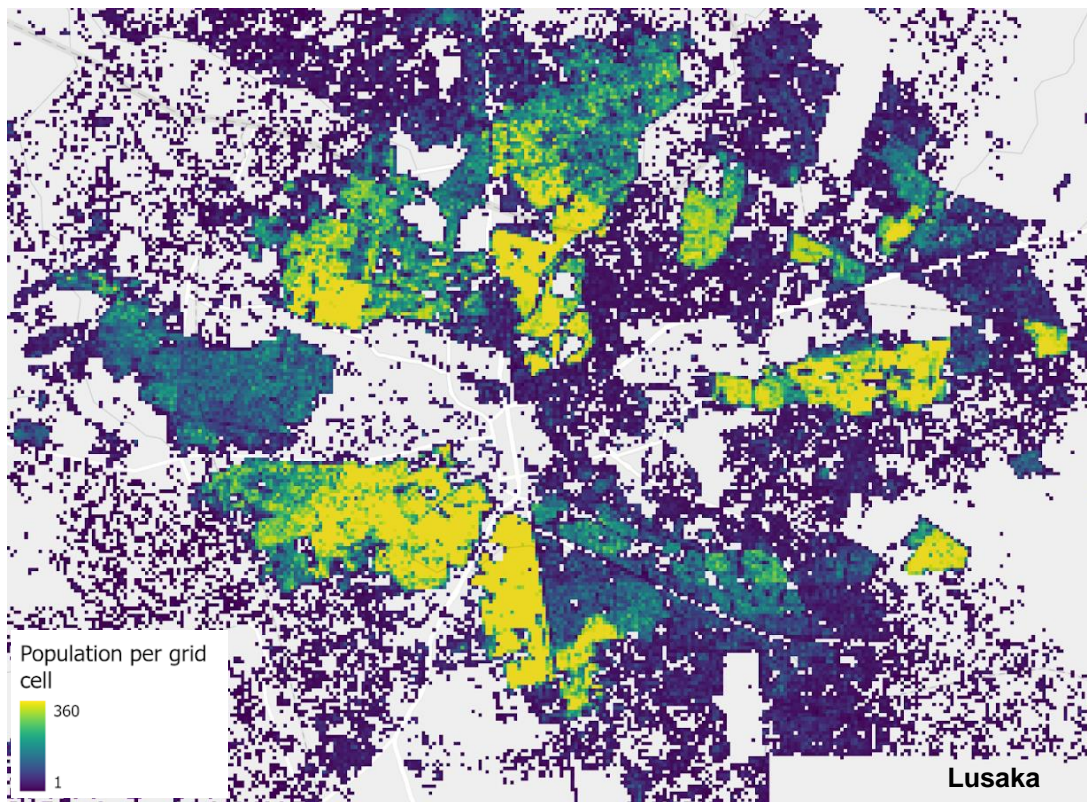
Household points (from census)



Geospatial covariates – including building footprint metrics



Gridded estimates – total population

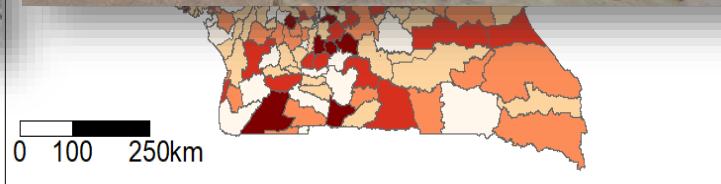
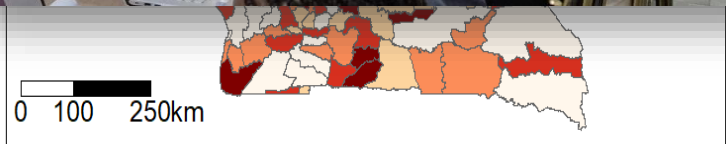
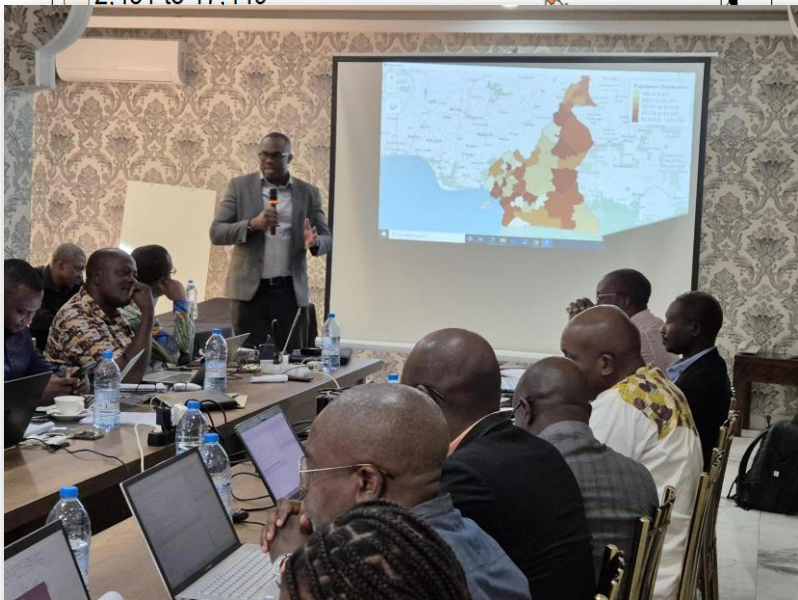


Population per grid cell (~100m)	Count of grid cells
1 - 10	1,150,392
11 - 20	224,813
21 - 50	98,089
51 - 100	30,983
101 - 200	15,490
>200	4,396
TOTAL	1,525,866

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....

Cameroon: National household survey listings



Papua New Guinea: Malaria survey household listings

nso.gov.pg/statistics/population/

NATIONAL STATISTICAL OFFICE

Home Statistics Censuses & Surveys Documents About us

POPULATION

Topics

- Economy
- Education
- Health
- Population
- Social

Users will find population data collected and produced by the National Statistical Office, such as documents, tables, graphs other other formats available.

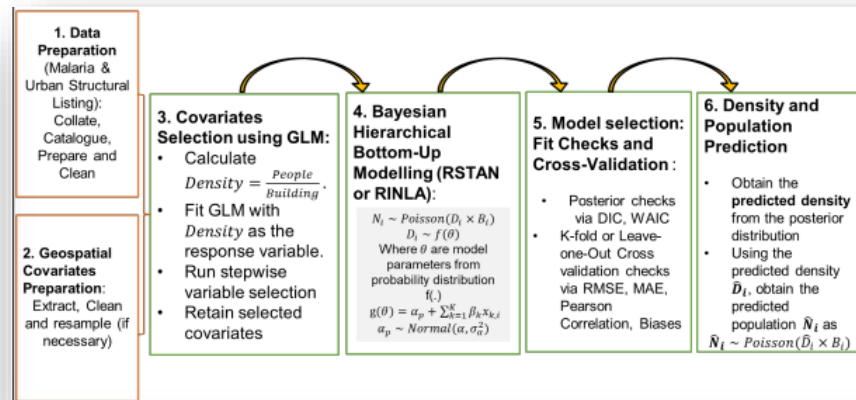
Population Estimates 2021

Total Population	Males	Females
11,781,559	6,142,585	5,638,974
2021	2021	2021

Population Key Indicators

Key Indicators	Estimates
Sex ratio ((no. of males/no. of females) X 100)	108.9
Percentage of working age population (15 – 64 years old)	61.6
Percentage of older population (65 years old and over)	2.7
Percentage of youth (15 to 24 years old)	21.6
Total dependency ratio (0 - 14 + 65 years old)/ 15-64 years old	62.4
Children dependency ratio (0-14/15-64 years old)	58.1
Old-age dependency ratio (65 years old and over/ 15-64 years old)	4.3
Median age, both sexes	21.2
Median age, Male	20.8
	21.7
	57.2
	55.3

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



Metric	In-Sample				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

