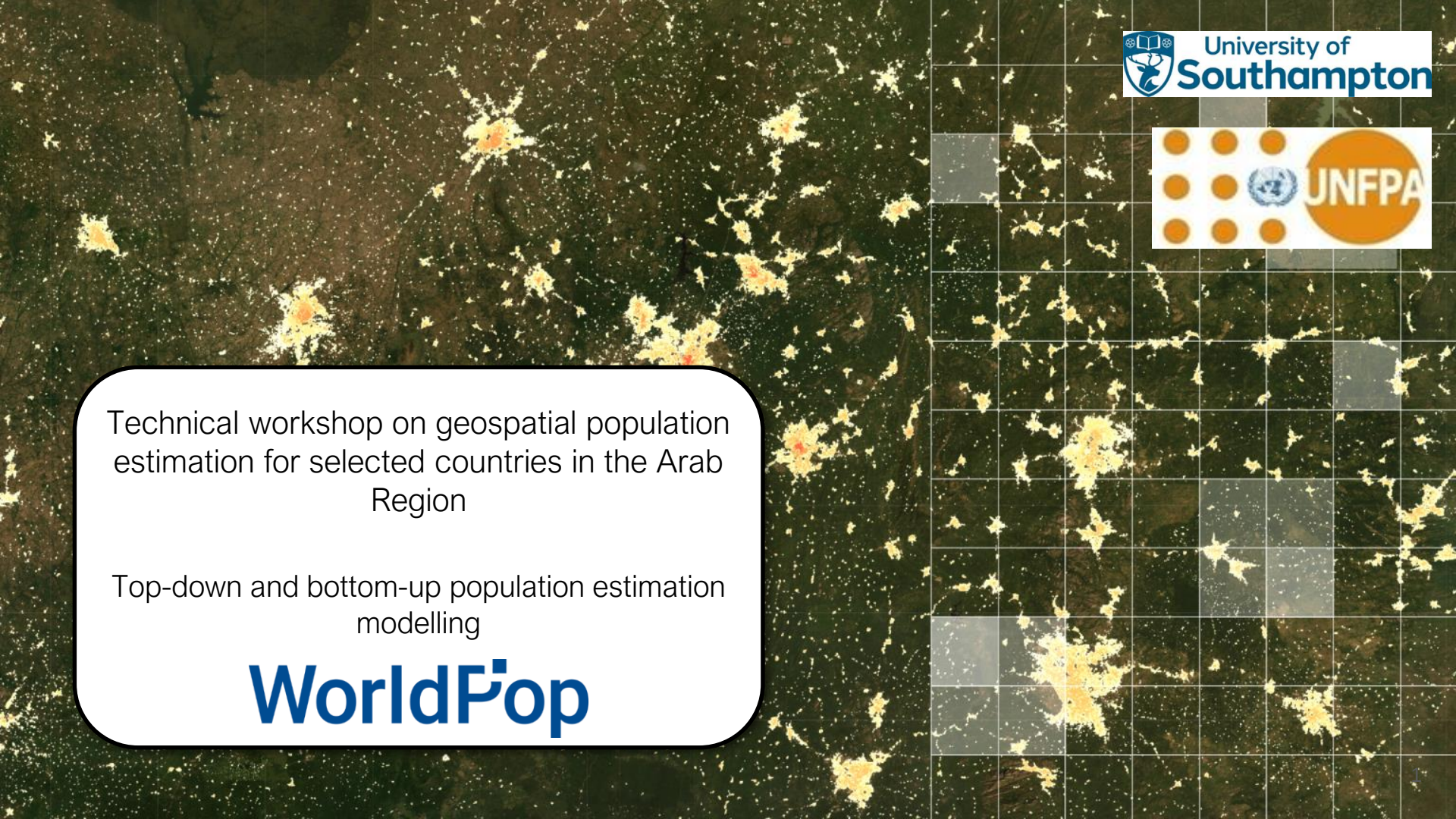


University of Southampton

UNFPA

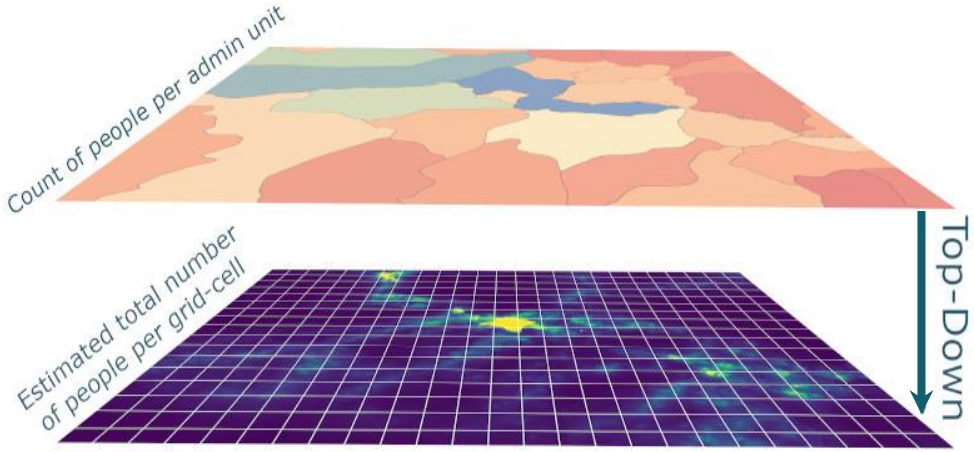
Technical workshop on geospatial population estimation for selected countries in the Arab Region

Top-down and bottom-up population estimation modelling

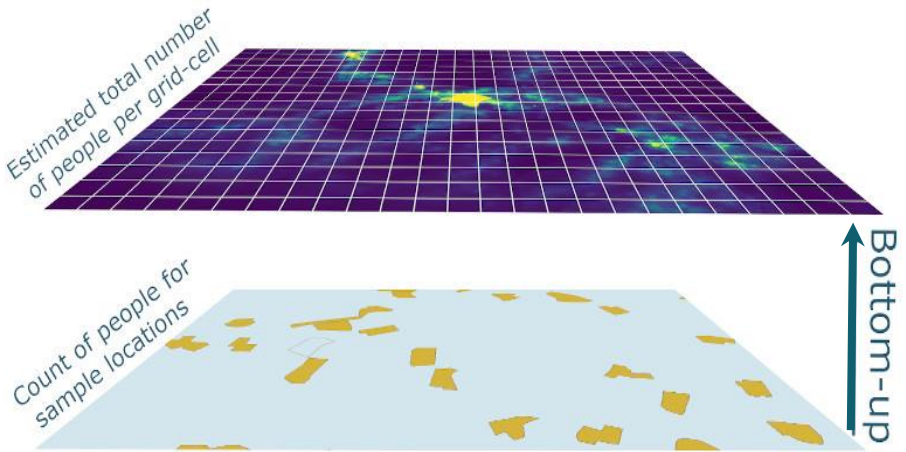
WorldPop

Small area population estimation approaches

“Top down” modelling approach

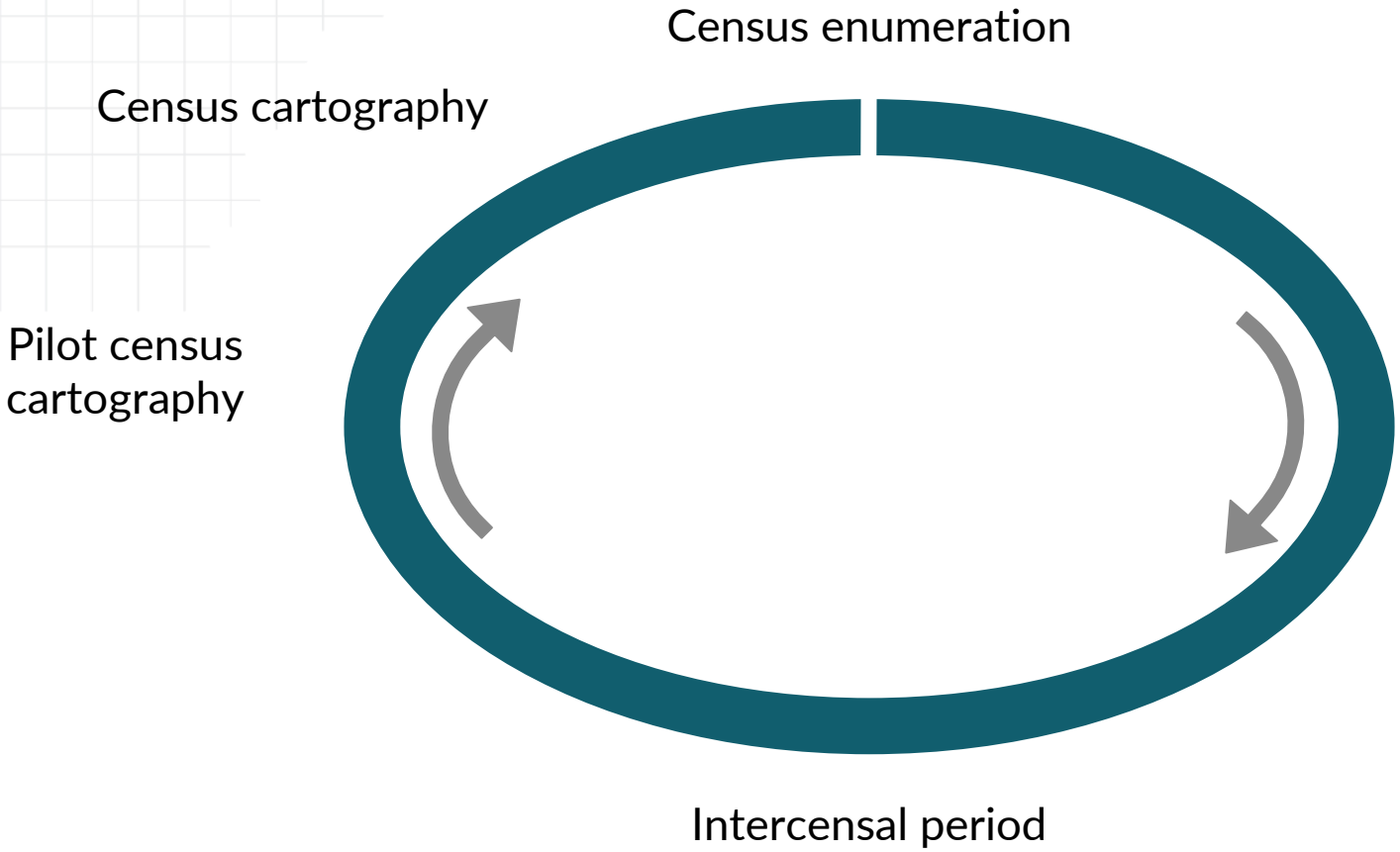


“Bottom-up” modelling approach

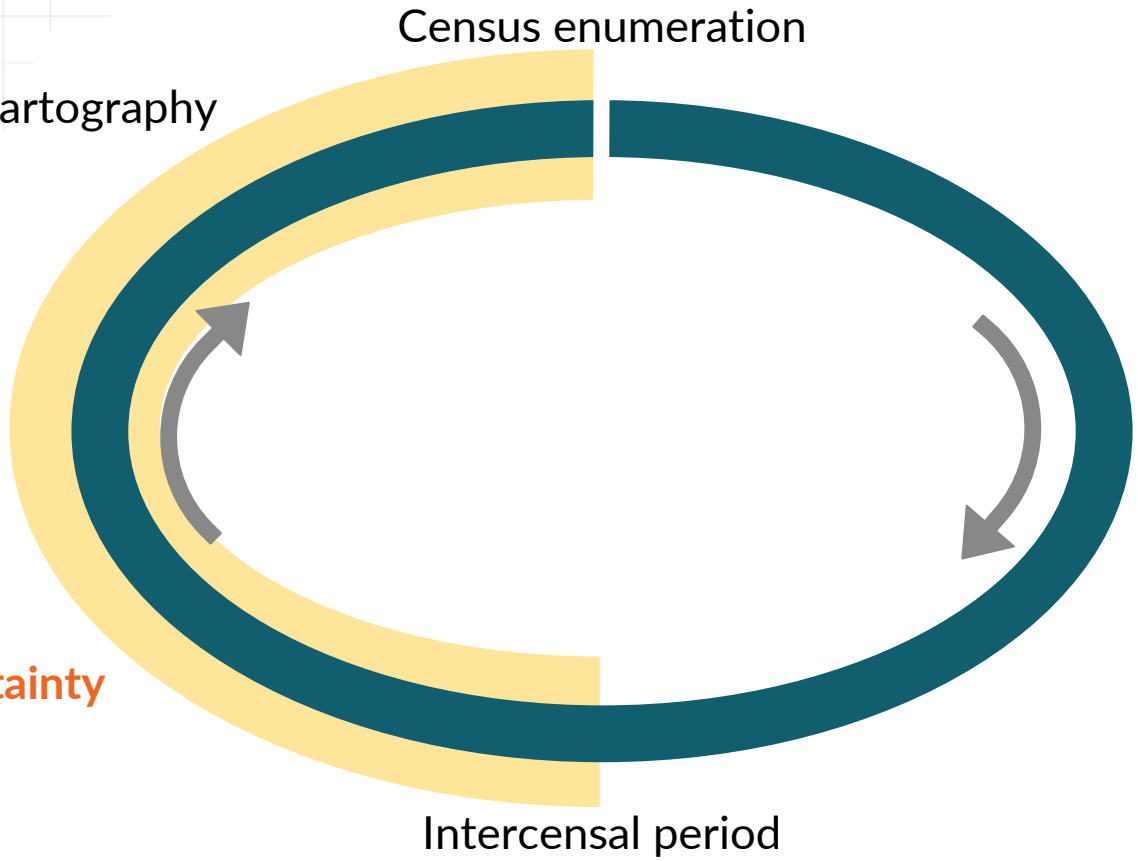


www.worldpop.org/methods/populations

Stages of the census cycle

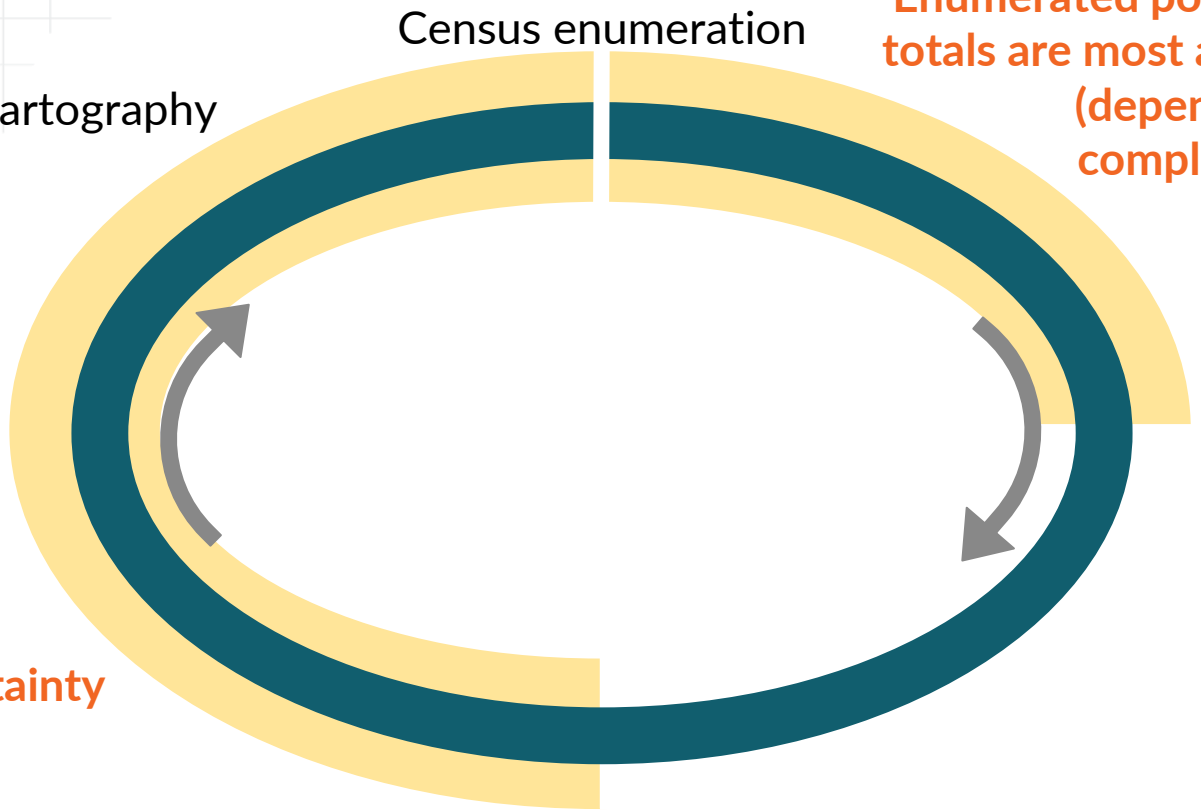


Stages of the census cycle



As time from the census elapses, increasing uncertainty in projections

Stages of the census cycle



Enumerated population totals are most accurate (dependent on completeness)

As time from the census elapses, increasing uncertainty in projections

Approaches for creating gridded population datasets

Data availability at different stages of the census cycle

Following census enumeration:

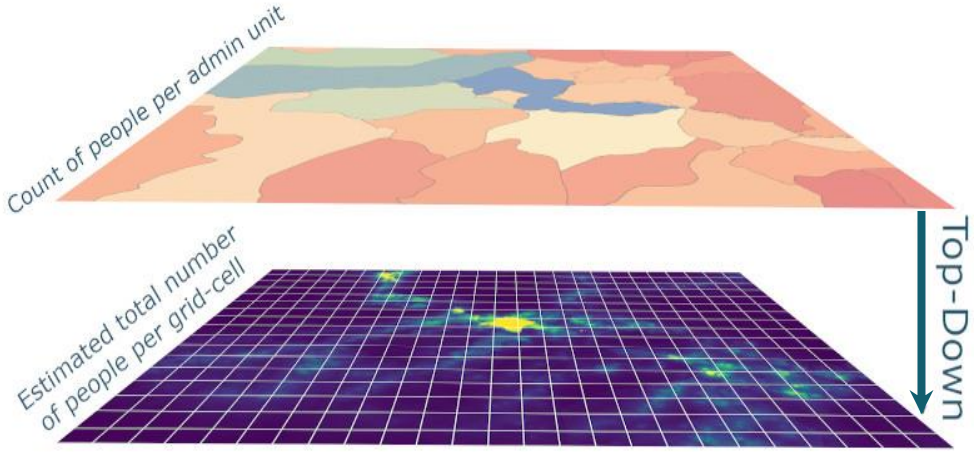
- **Top-down:** Population totals for administrative areas (e.g. Wards, Districts, EAs) can be spatially disaggregated, assuming complete spatial coverage
- **Bottom-up:** If spatial coverage is incomplete, population in enumerated areas could be used to predict in incomplete areas - “fill in the gaps”

During the intercensal period:

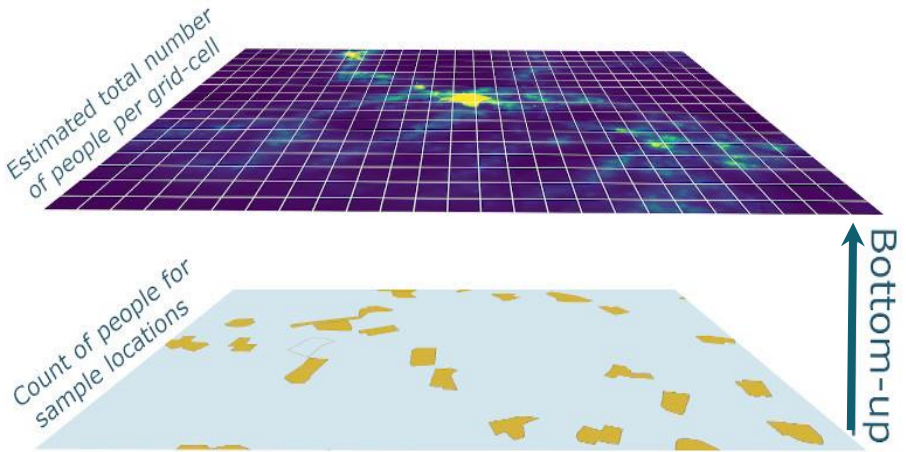
- **Top-down:** Projected population totals can be spatially disaggregated, however projections become increasingly uncertain with time elapsed
- **Bottom-up:** Activities which include complete (geolocated) listing of households or collect population count information can be used for bottom-up modelling, such as:
  - Listings from household surveys
  - Census cartography

Small area population estimation approaches

“Top down” modelling approach



“Bottom-up” modelling approach



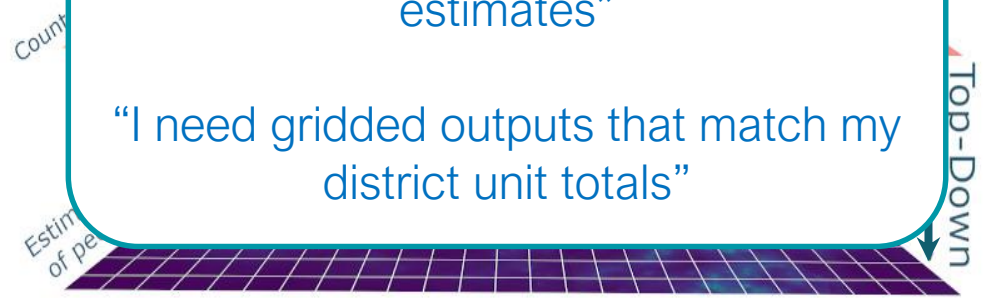
www.worldpop.org/methods/populations

Small area population estimation approaches

"Top down" modelling approach

"I trust these subnational province-level projections, but need finer scale estimates"

"I need gridded outputs that match my district unit totals"

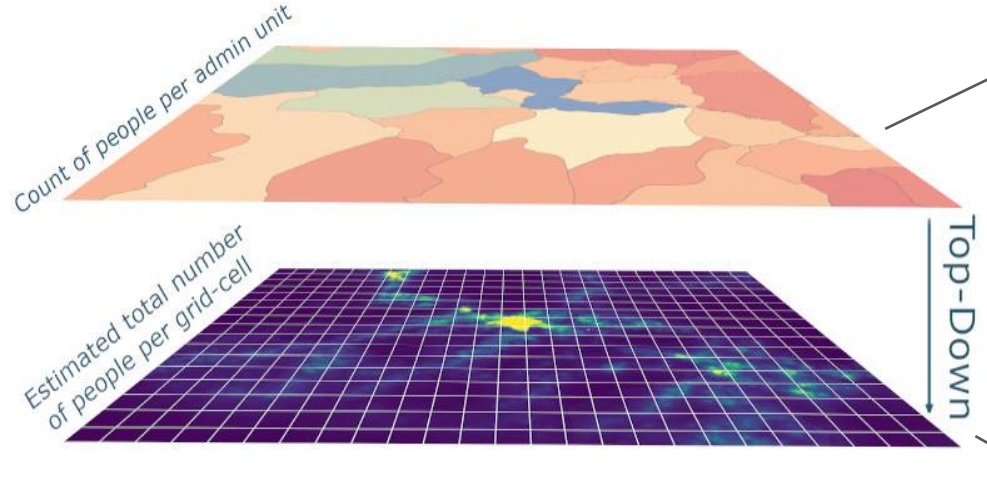


"Bottom-up" modelling approach

"The last national census was 1984 so don't trust that data, but I need small area population data and do have some recent sample enumeration data"

"We have conducted our census, but could not access certain areas and need estimates for these"

“Top down” modelling approach

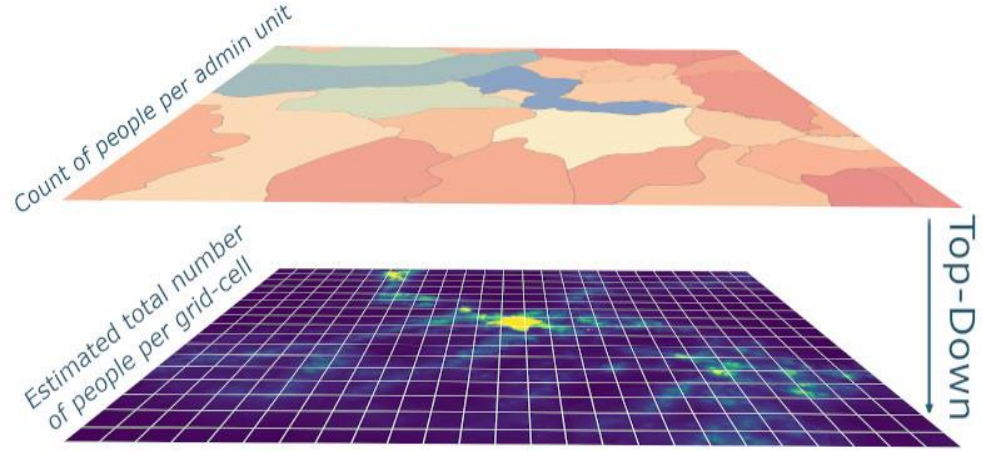


Count of people per admin unit, e.g.:

- 2024 projected District population
- 2022 enumerated Ward population

Estimated total number of people per grid cell

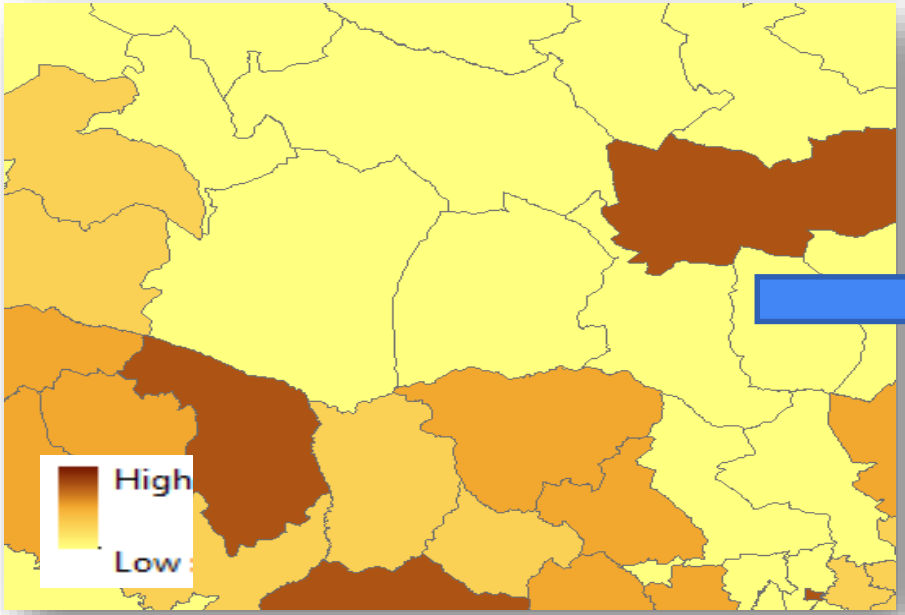
“Top down” modelling approach

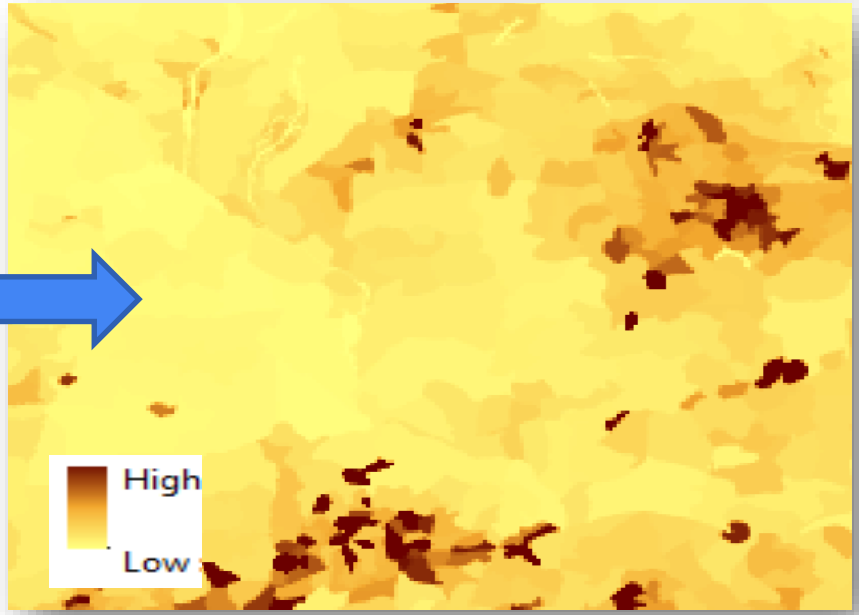


Population counts per administrative unit can mask small area variations in population. **Spatial disaggregation approaches** can be used to create gridded datasets from population counts for administrative units.

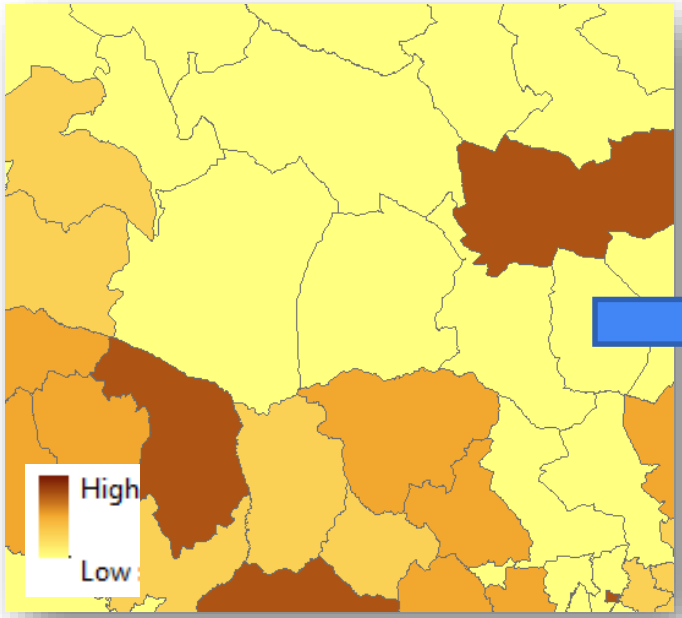
There are several disaggregation methods that can be used. Each of which creates a **gridded “weighting” layer** to assign total population count to the grid cells within the administrative unit

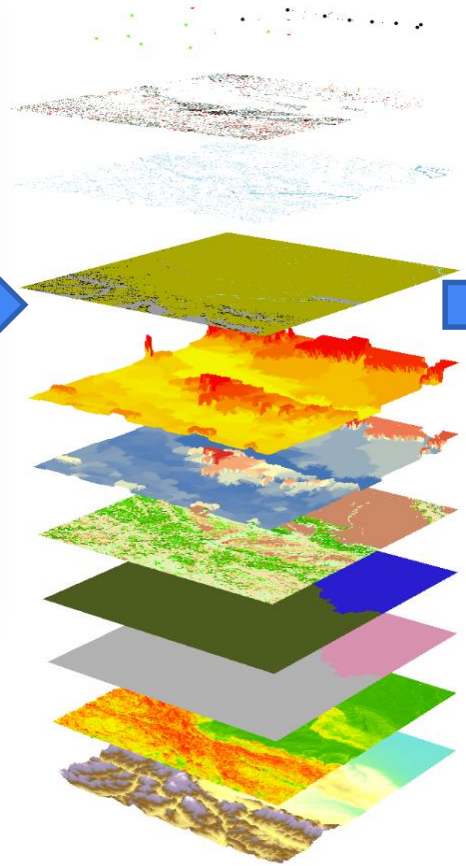
Top-down: How can we go from aggregated counts to gridded surfaces?

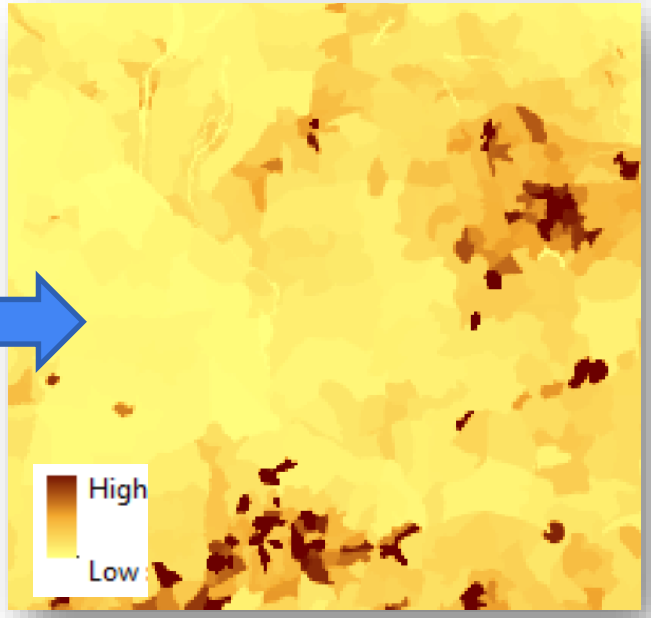




Top-down modelling



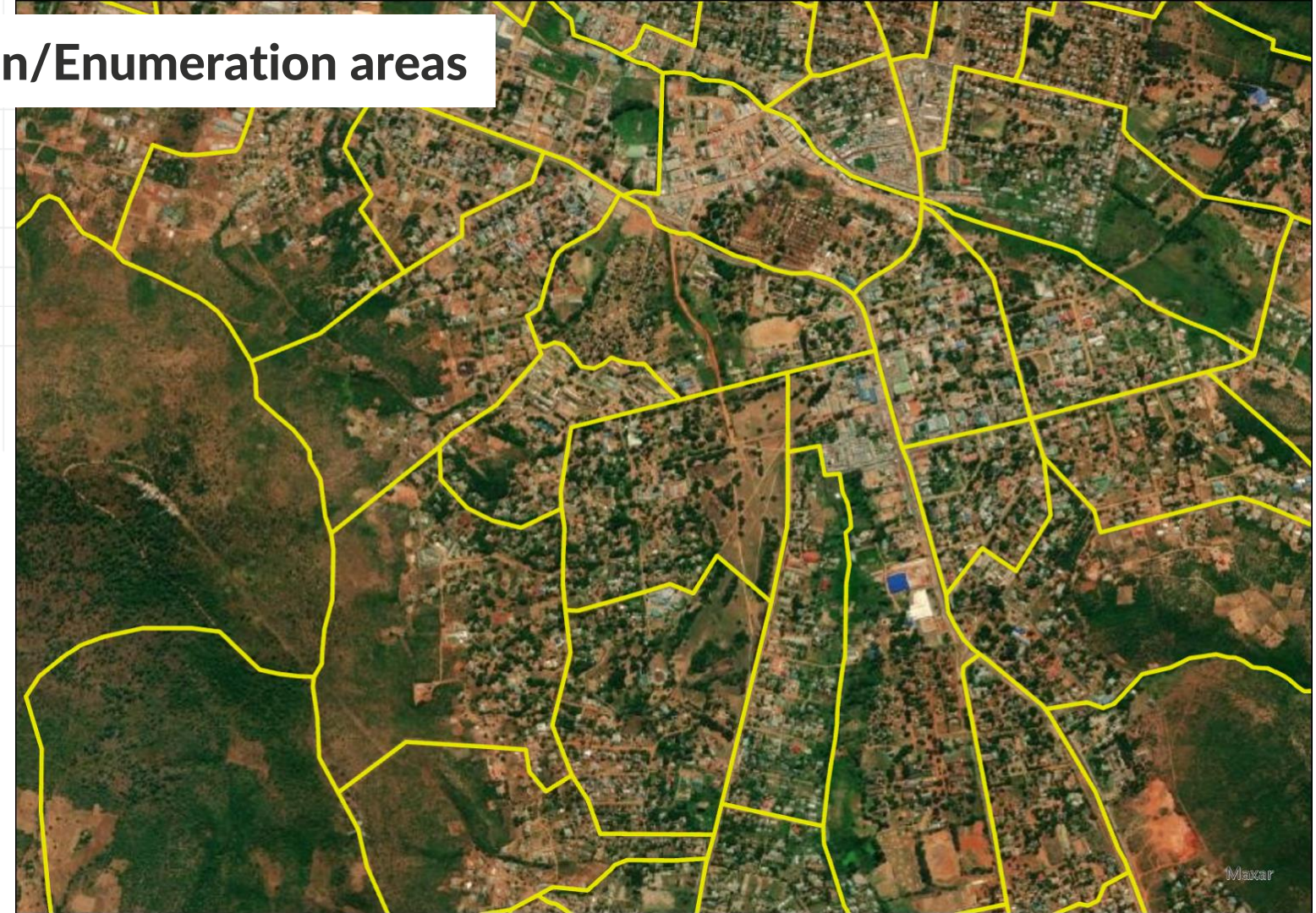




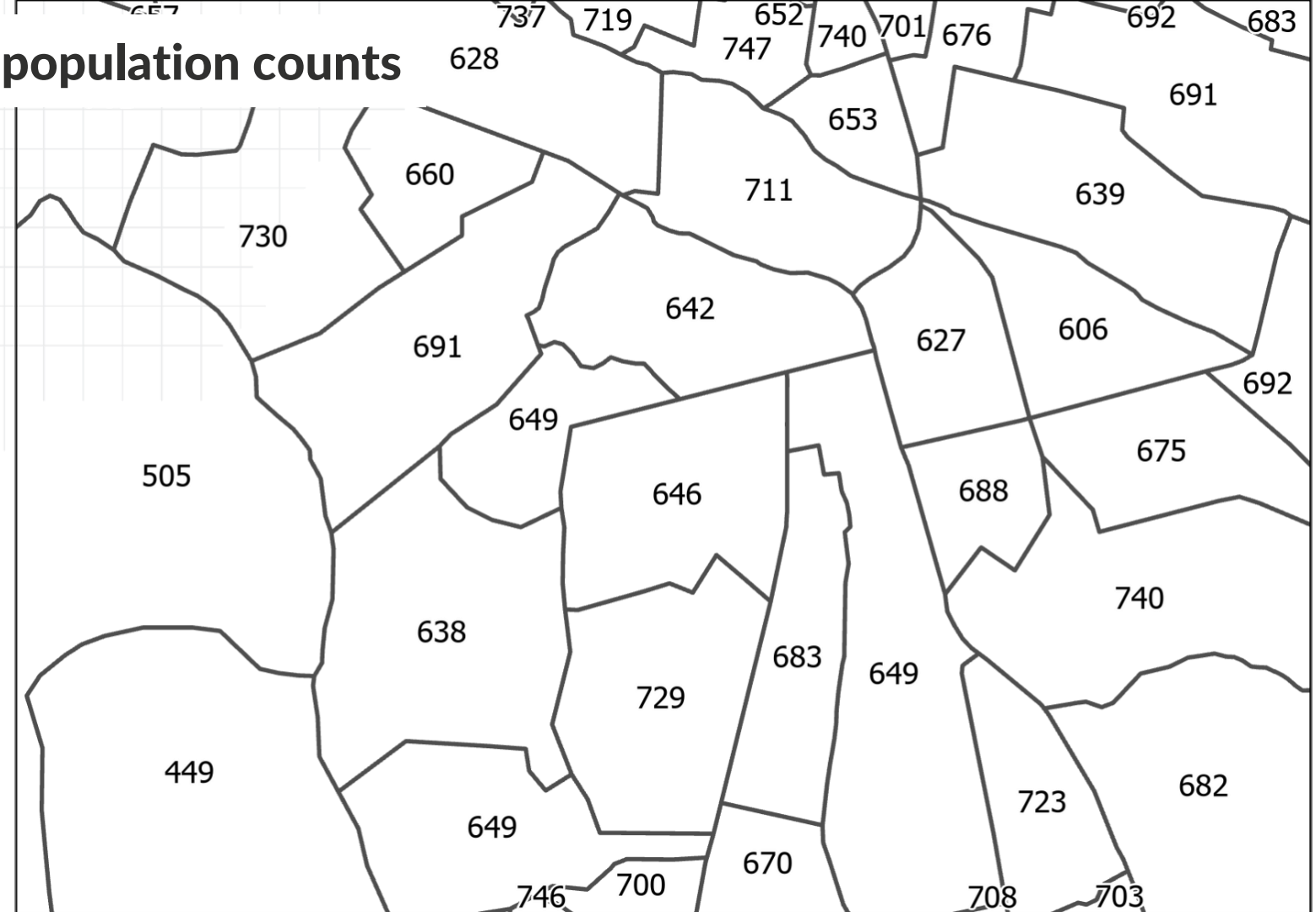
Population counts aggregated at coarse, irregular administrative unit level, making integration and comparisons with other data challenging

Integration with satellite/GIS data related to human population distribution patterns to map counts to regular grids

Admin/Enumeration areas

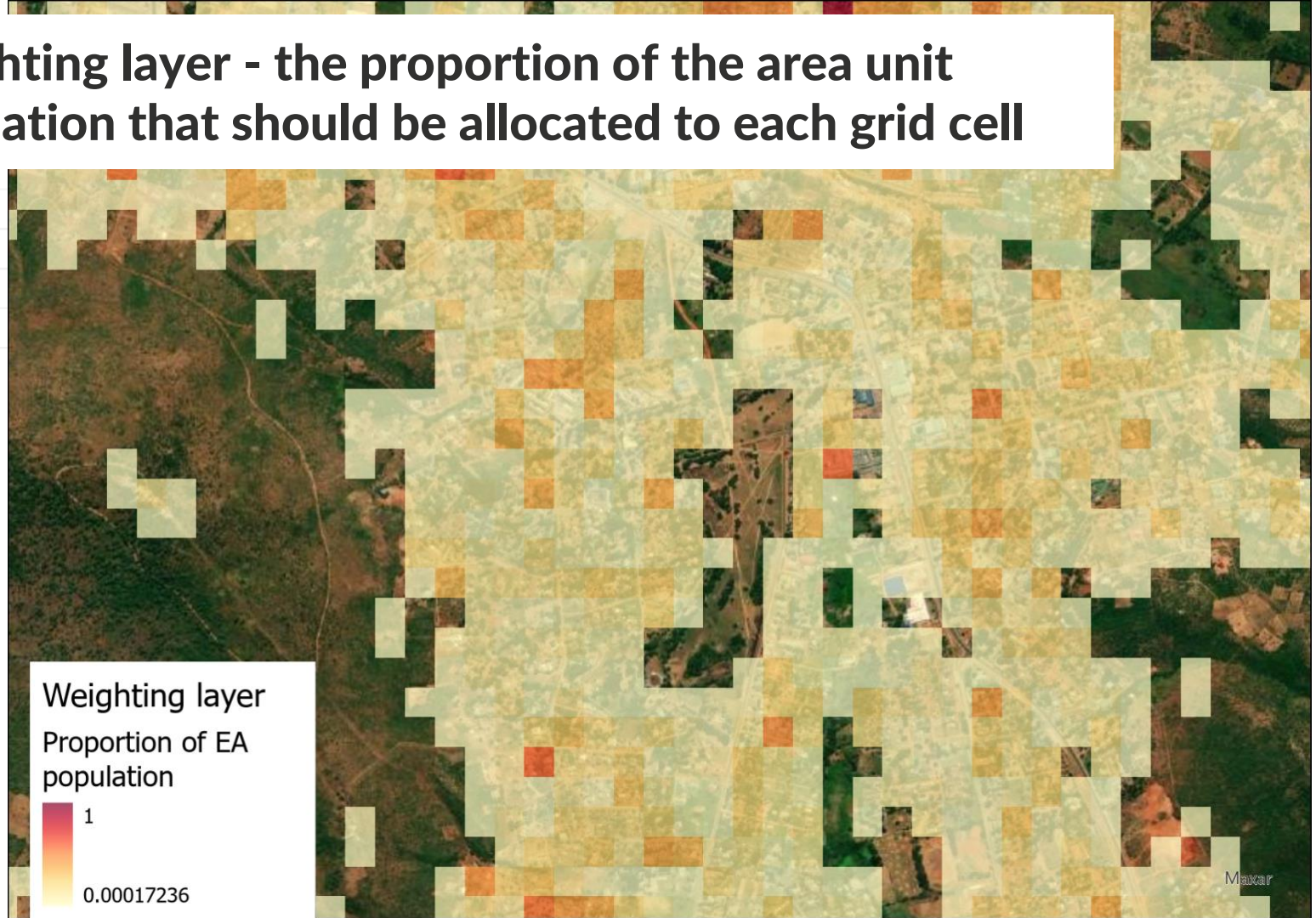


Area population counts

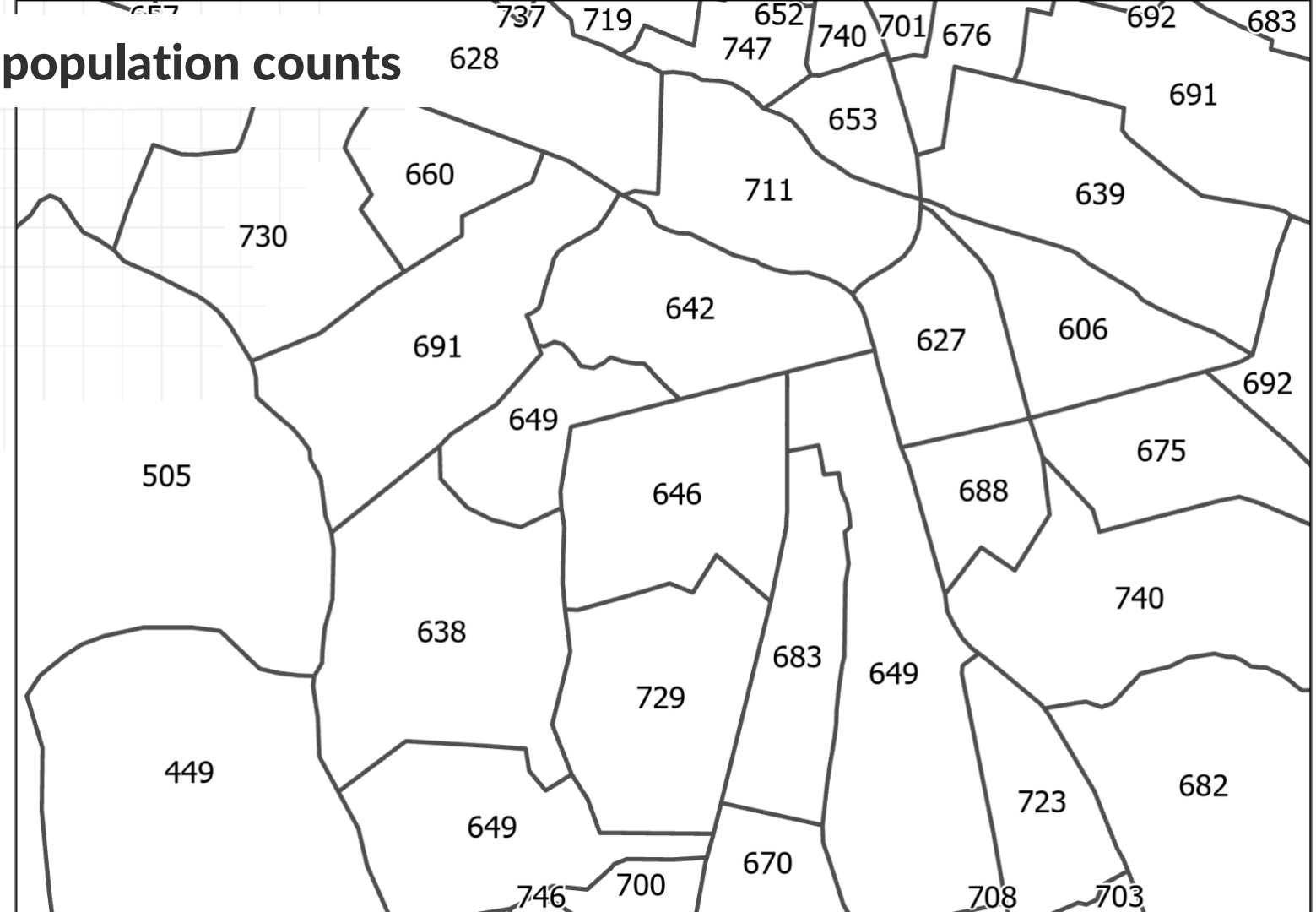


Constrain to areas of settlement - e.g. from building footprints

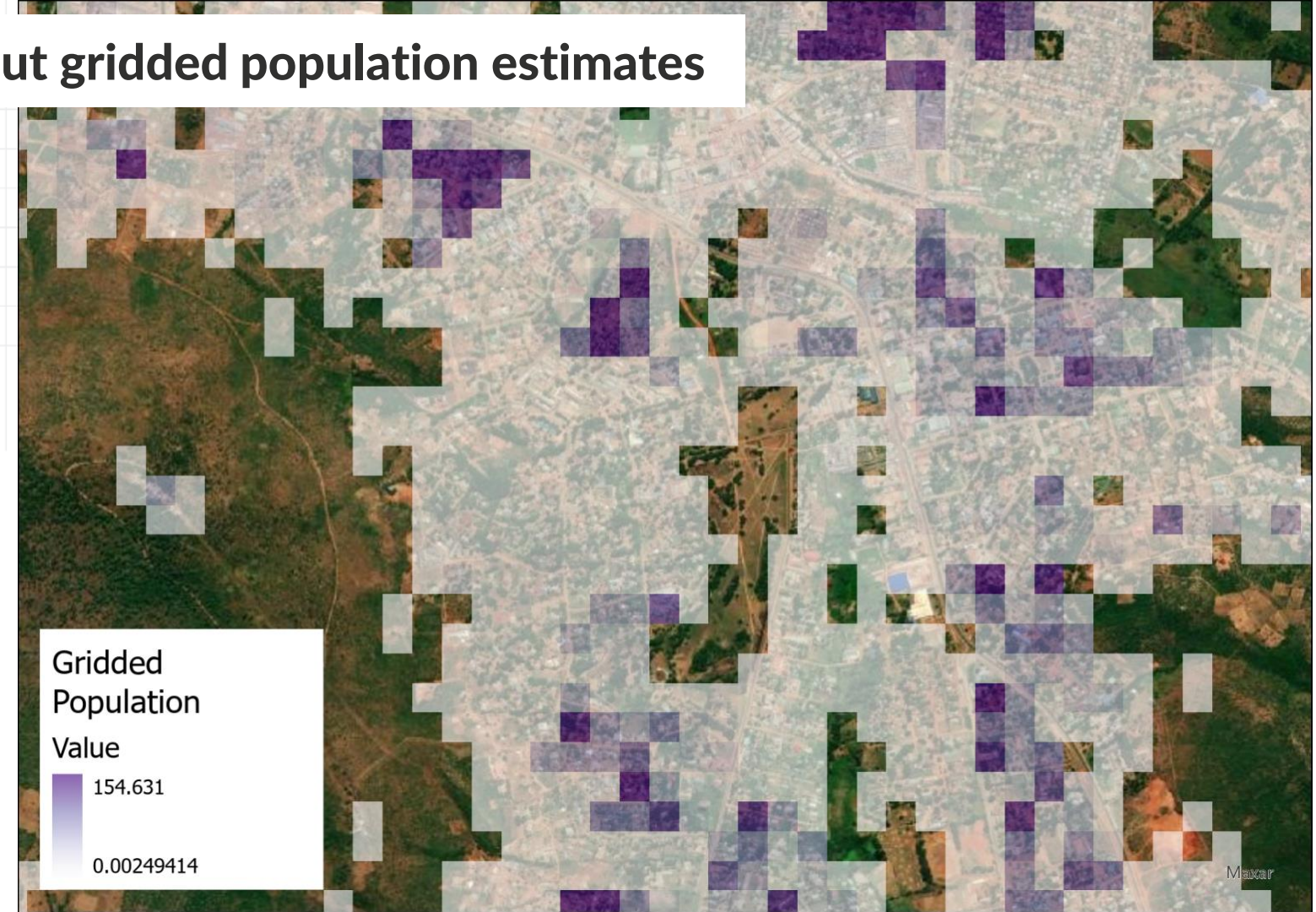
Weighting layer - the proportion of the area unit population that should be allocated to each grid cell



Area population counts

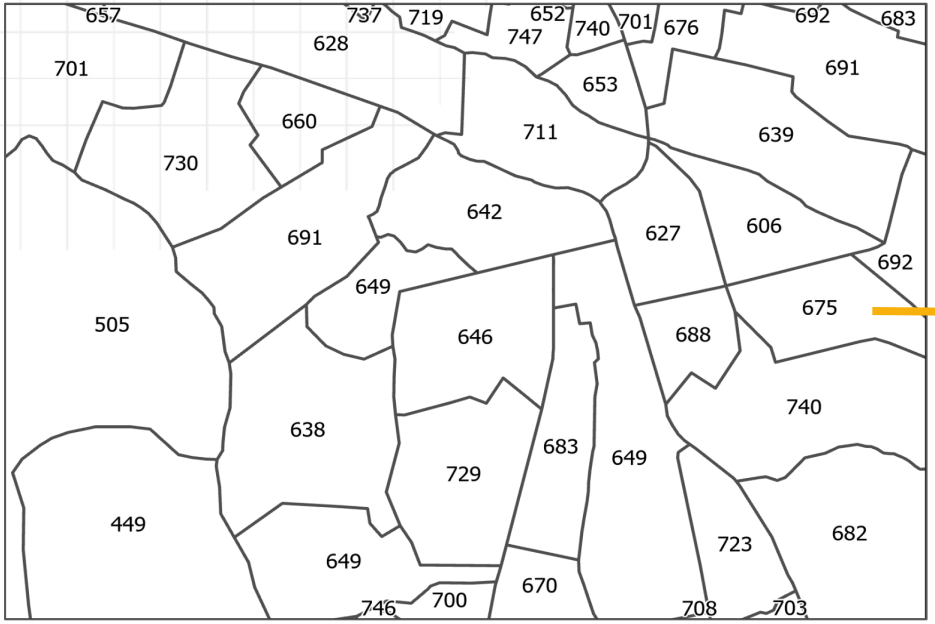


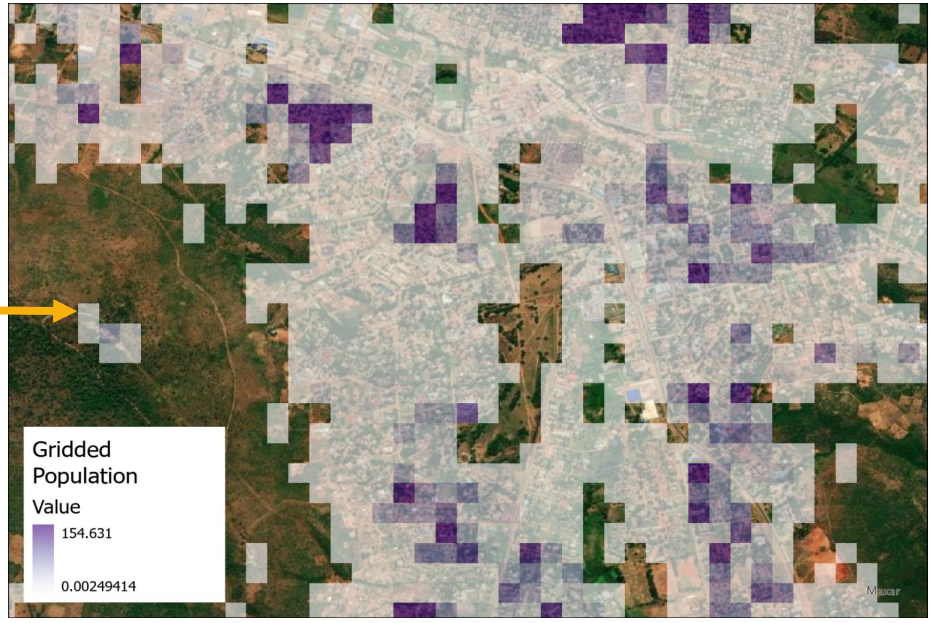
Output gridded population estimates



Gridded population datasets

From area counts to gridded estimates

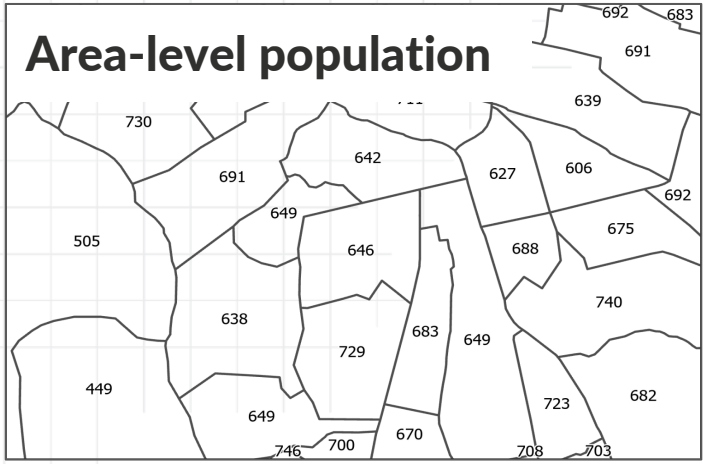


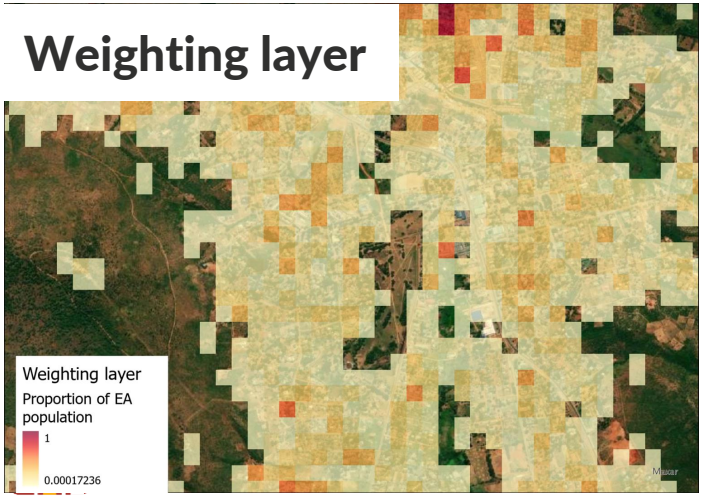


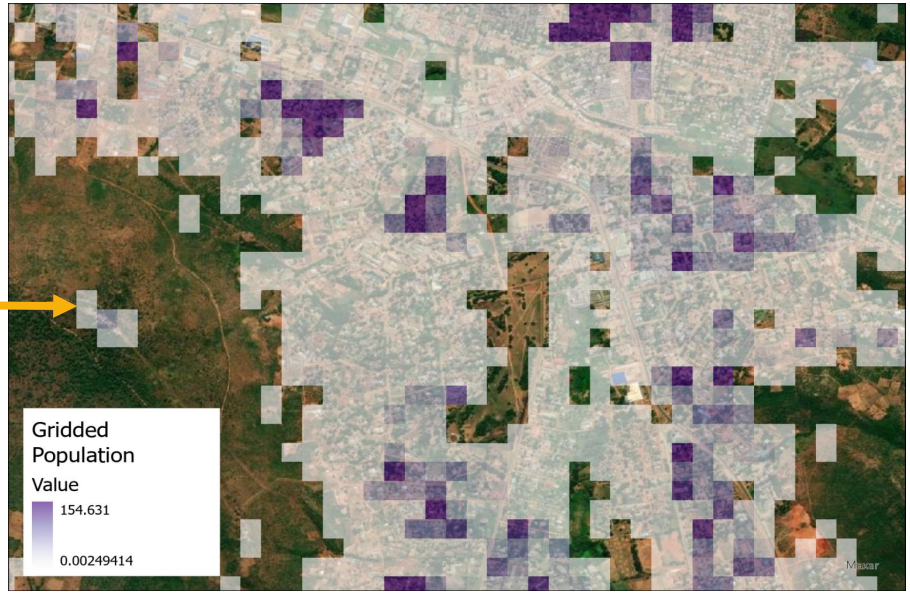
Area population

Gridded population estimates

Top-down estimation

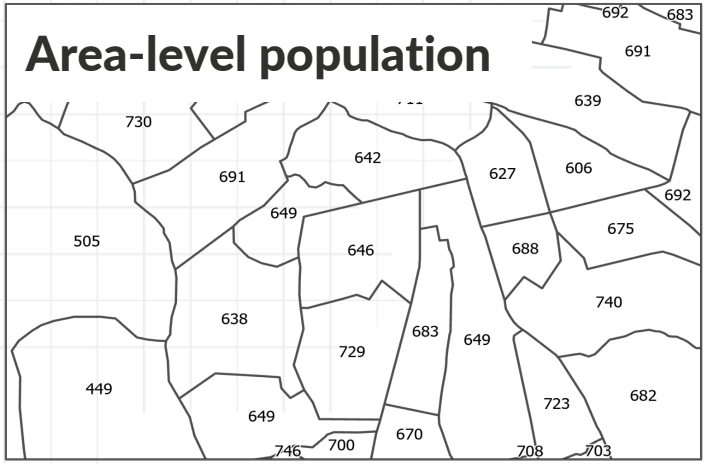


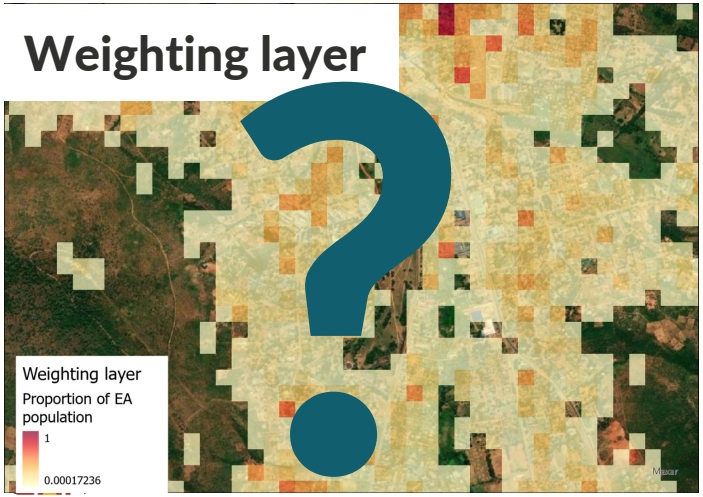


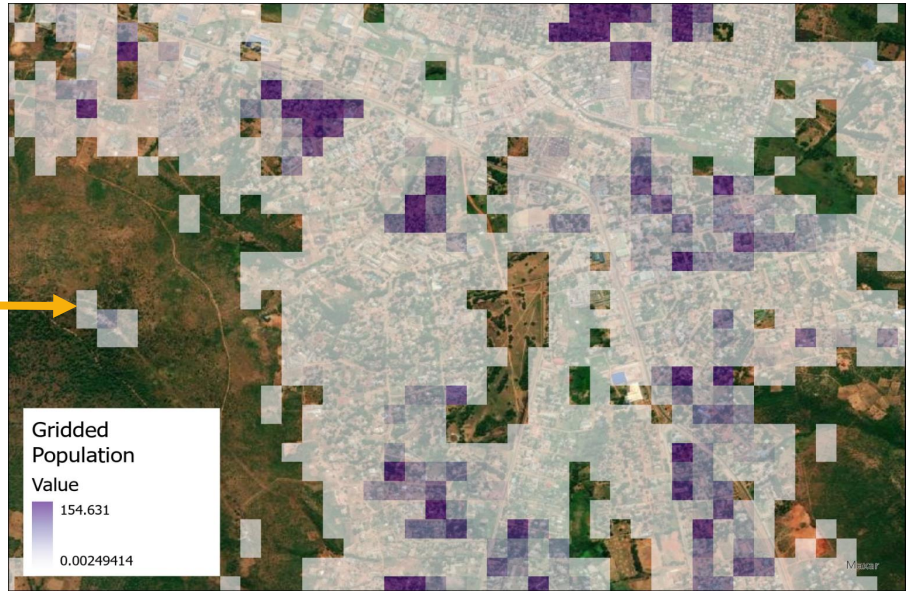


Gridded population estimates

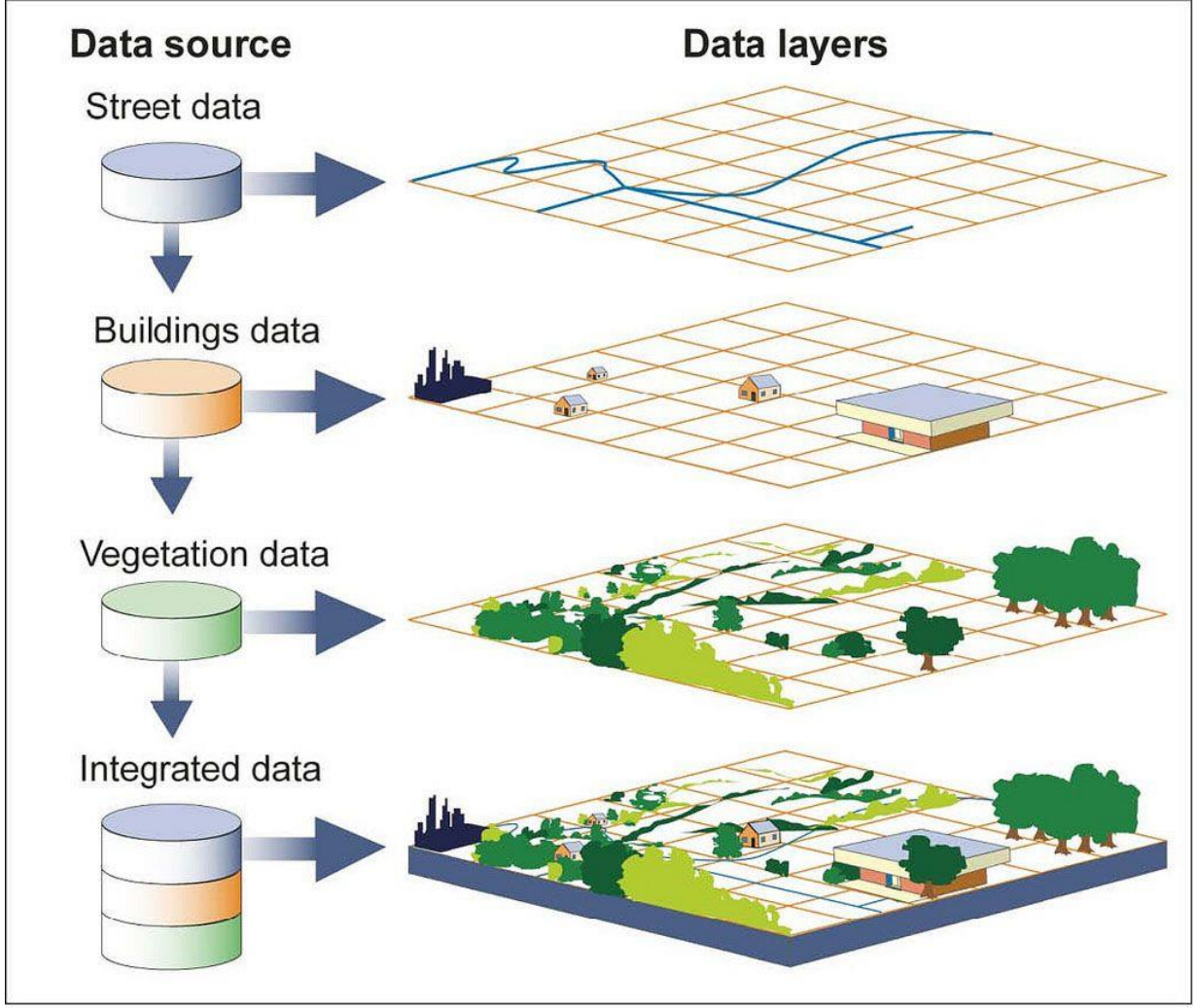
Top-down estimation



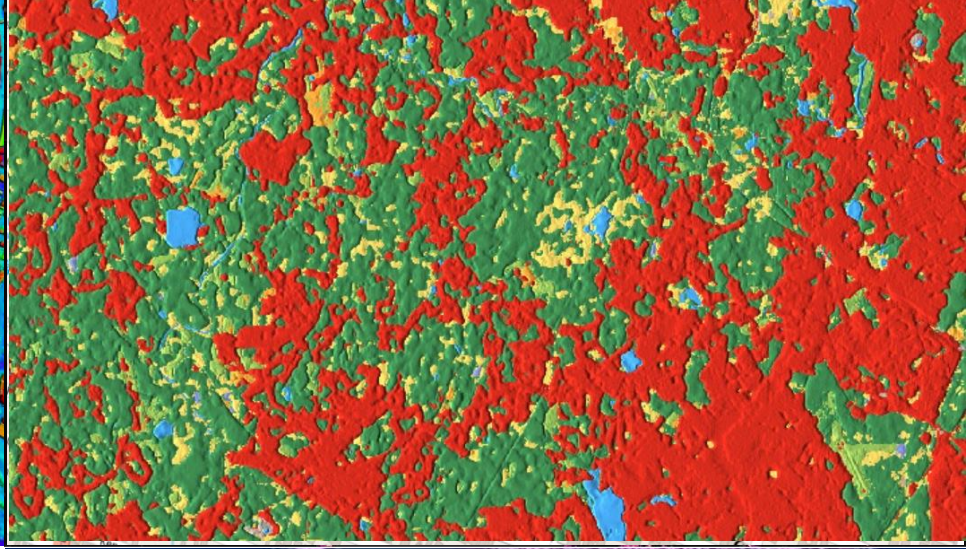


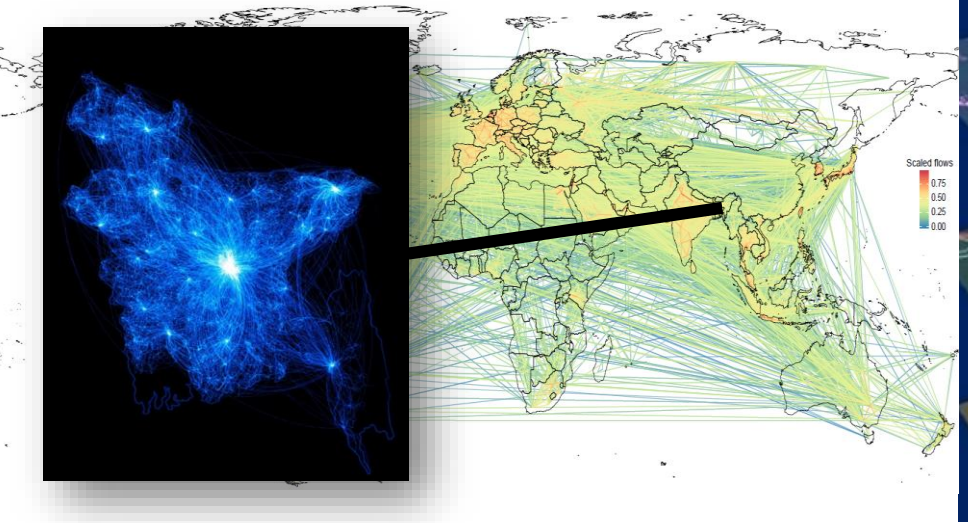


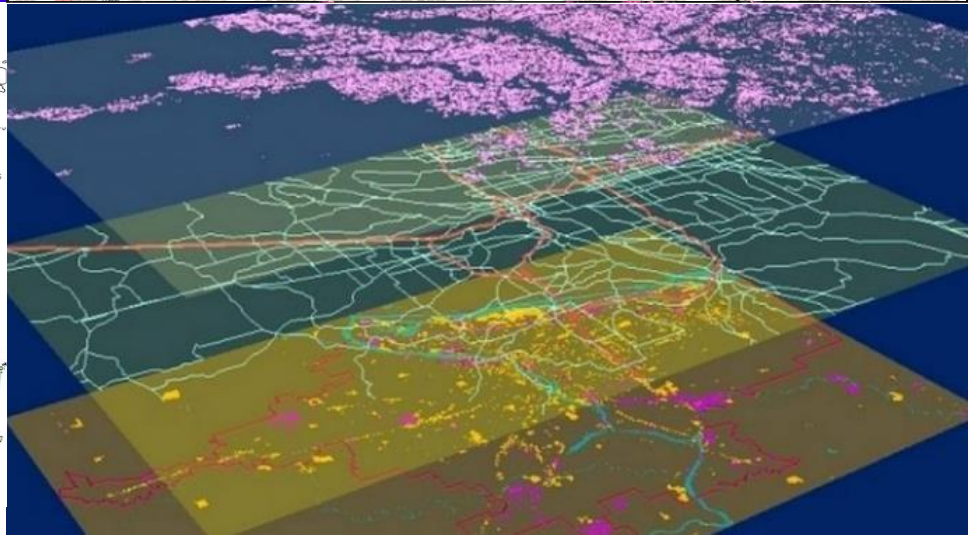
Gridded population estimates


Data source
Data layers
Street data
Buildings data
Vegetation data
Integrated data

Source: GAO.

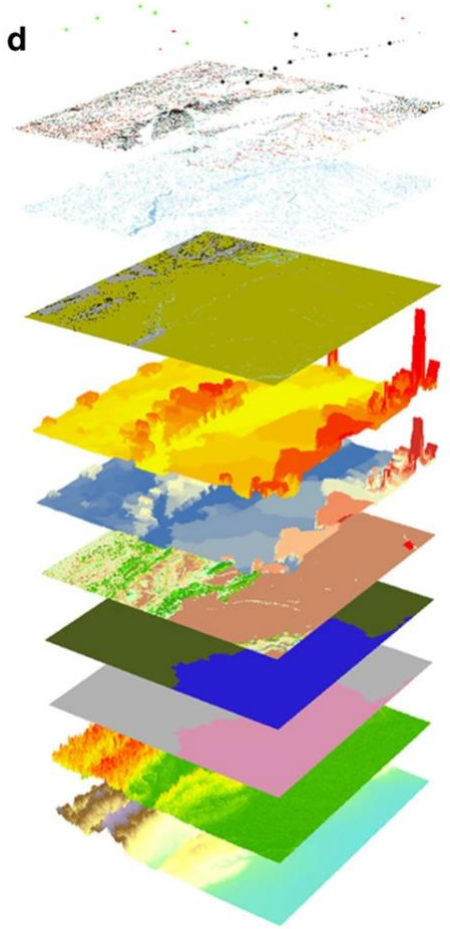



Scaled flows
0.75
0.50
0.25
0.00



Making a weighting layer

- The weighting layer needs to be created from predictive variables (geospatial data) related to population density, and associated with the built and natural environment
- Geospatial data are harmonised so that all variables have the same spatial resolution and grid cell alignment
- Select the subset of geospatial covariates that are most strongly correlated with population density



Geospatial data stack

Why and when to use top-down?

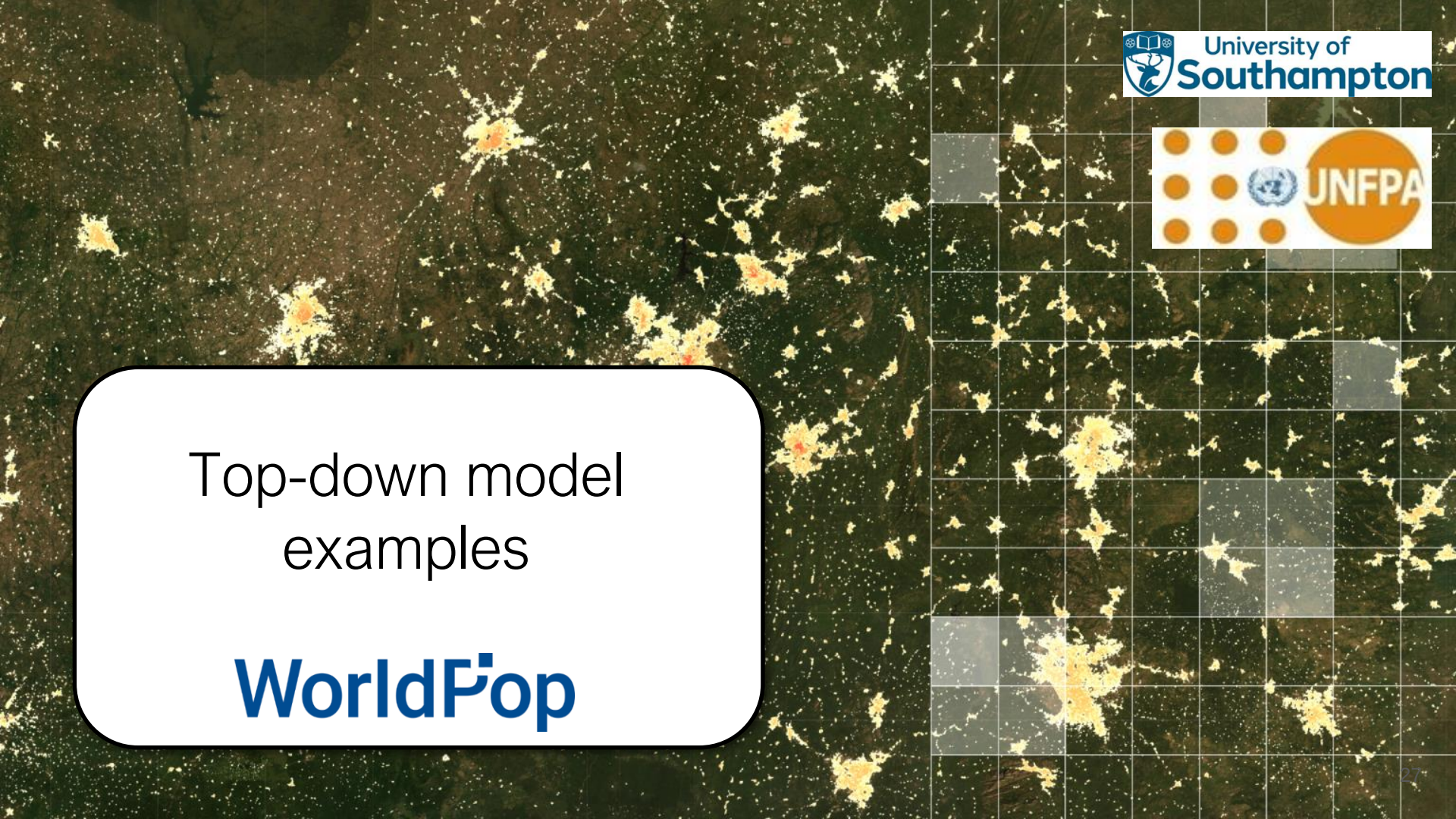
- Need for small area estimates that match official/other estimates at larger levels of aggregation
- Need for rapid small area estimates built from projections
- Need for estimates that can be distributed openly
- Need for estimates that can be integrated with other datasets

WorldPop

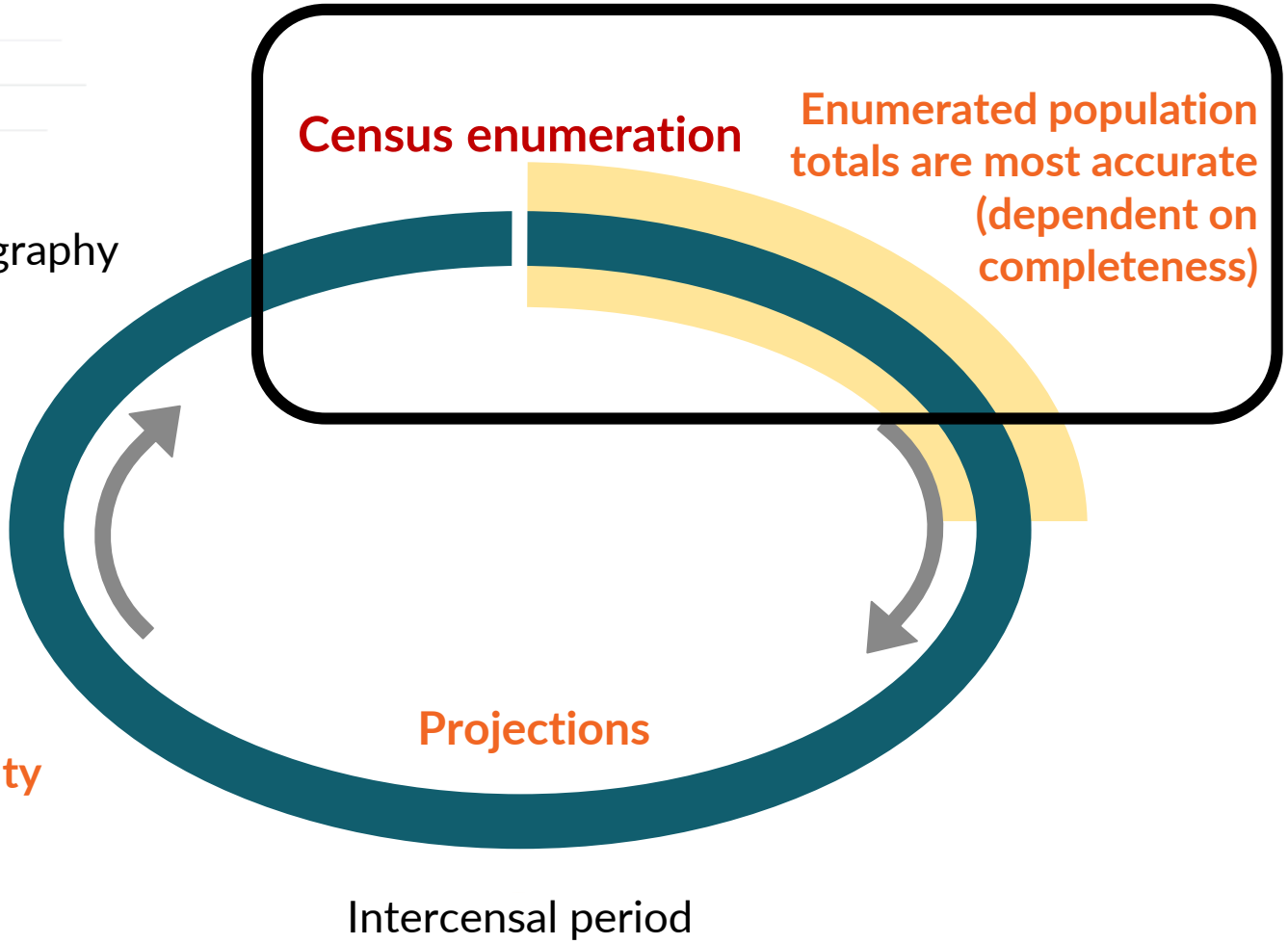
Limitations

- If the population counts that we use as input to our top-down model are inaccurate, then our population map will be too
- We often cannot capture all relevant determinants of population density variations with our stack of geospatial data
- The model is only as good as the data it is trained on – if we have too few input units, and/or the range of population densities represented within them is narrow, then the output modelled estimates will be poor

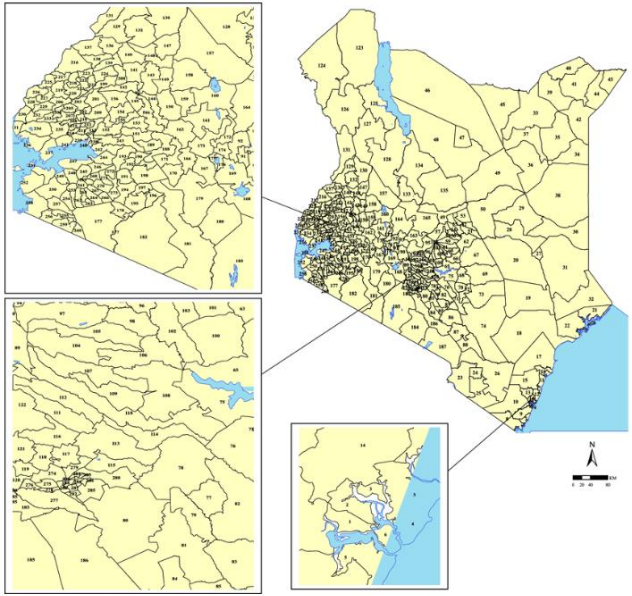
WorldPop


University of Southampton
UNFPA
Top-down model examples
WorldPop

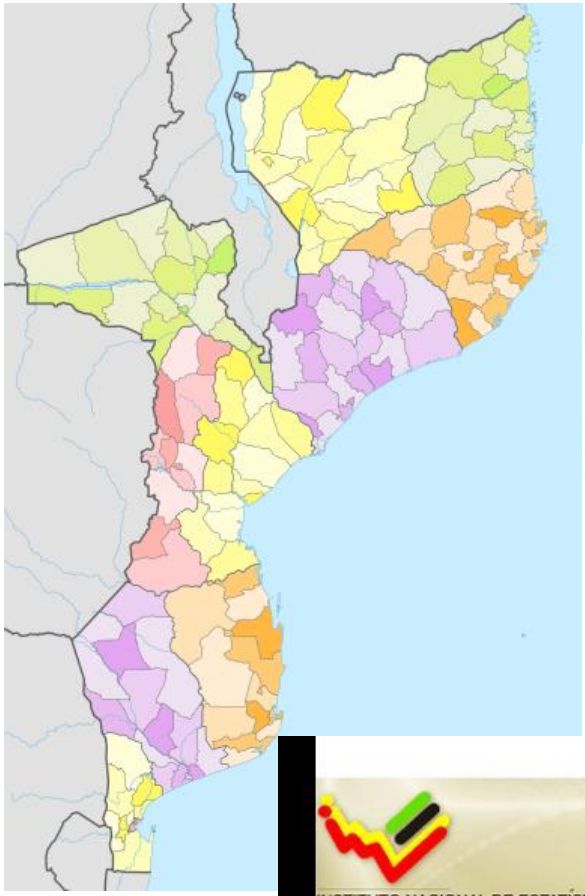
Stages of the census cycle



As time from the census elapses, increasing uncertainty in projections

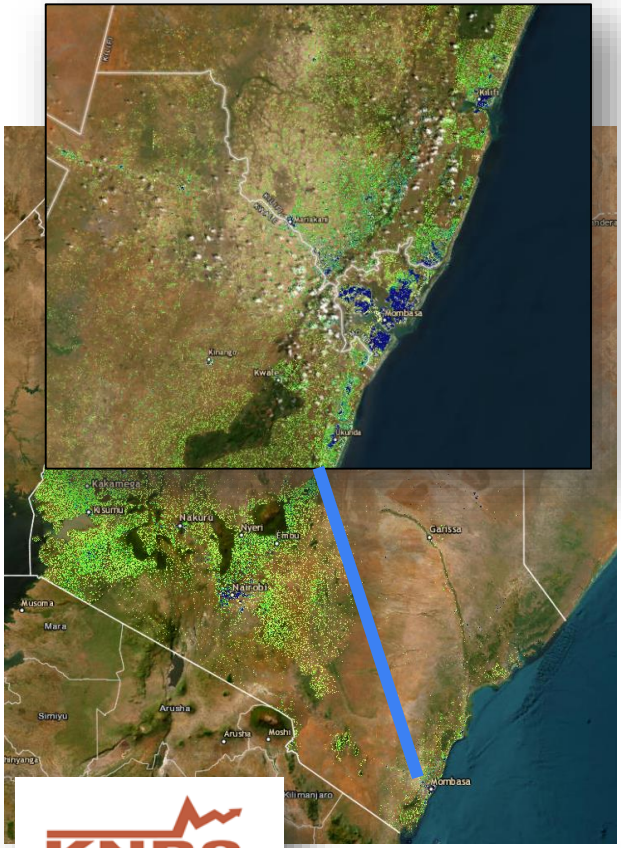


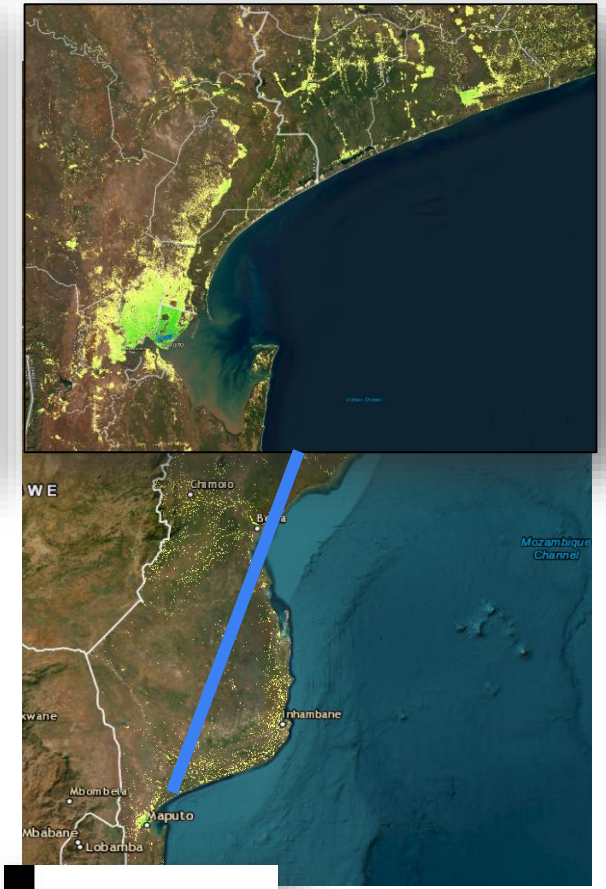
KNBS
KENYA NATIONAL
BUREAU OF STATISTICS
Keeping you informed

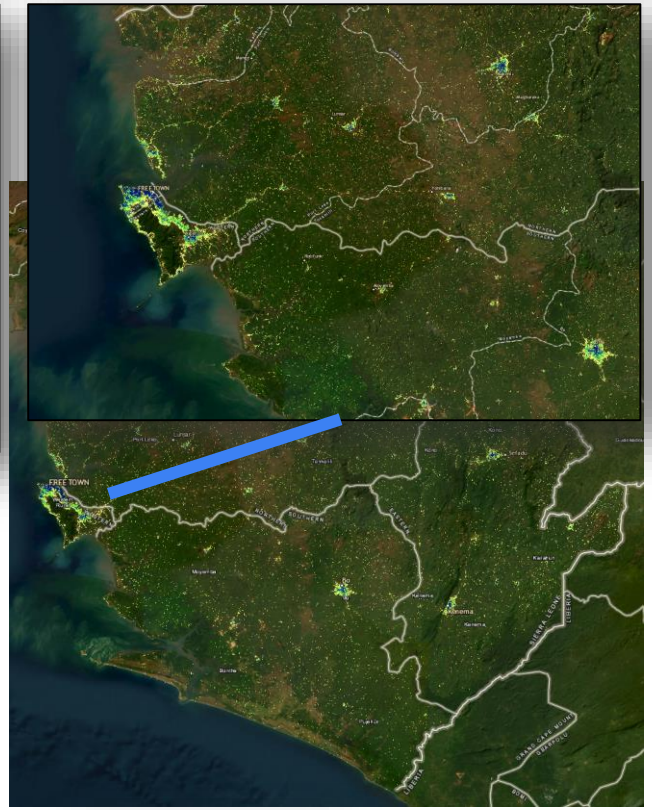


INSTITUTO NACIONAL DE ESTATÍSTICA

Statistics Sierra Leone
Stats SL







KNBS
KENYA NATIONAL BUREAU OF STATISTICS
Keeping you informed

Statistics Sierra Leone
Stats SL

unicef

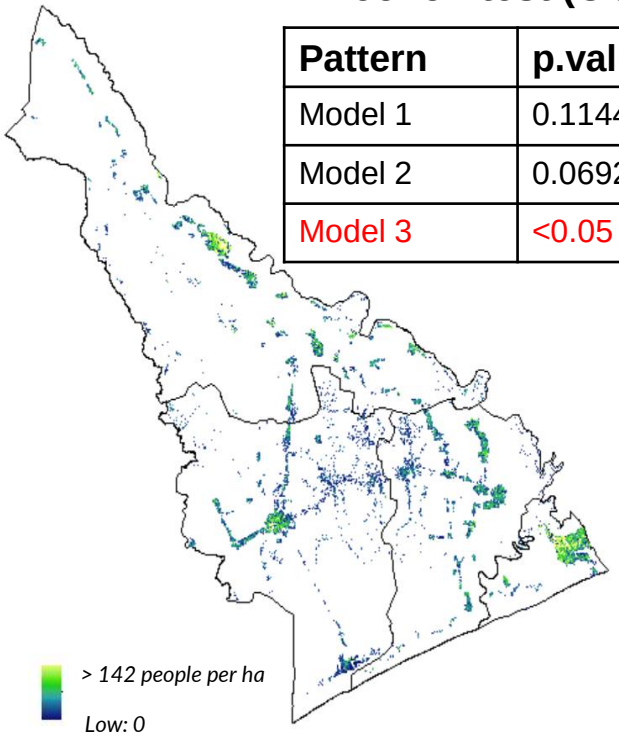
GRID3
GEO-REFERENCED INFRASTRUCTURE AND DEMOGRAPHIC DATA FOR DEVELOPMENT

INSTITUTO NACIONAL DE ESTATÍSTICA

https://wopr.worldpop.org/?/Population

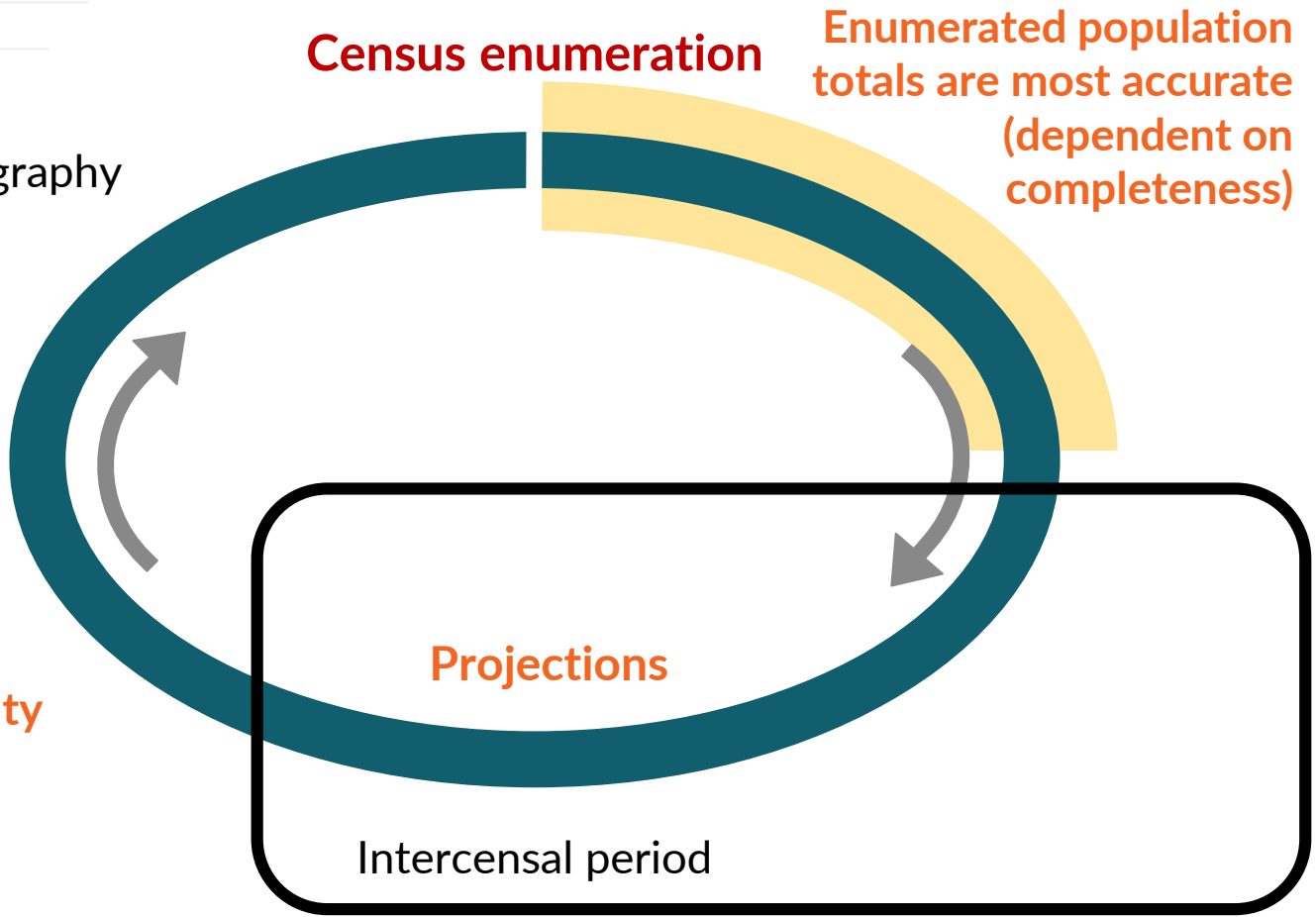
Wilcoxon test (Observed vs predicted)

Pattern	p.value	Conclusion
Model 1	0.1144	Significant differences
Model 2	0.06925	Slightly different distributions
Model 3	<0.05	No significant differences



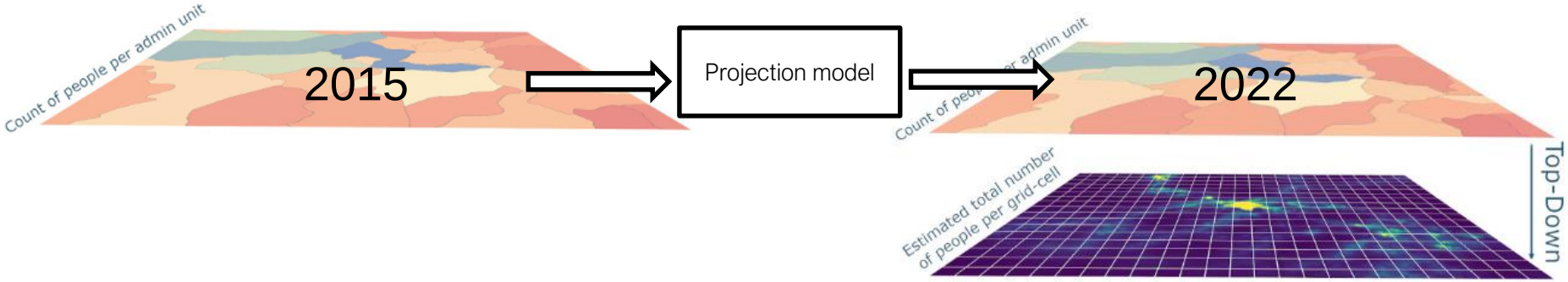
Field validation

Stages of the census cycle

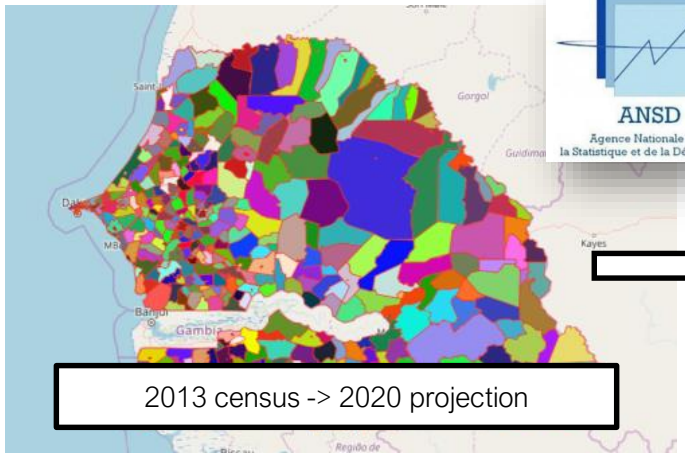


As time from the census elapses, increasing uncertainty in projections

Demographic projections, disaggregation modelling

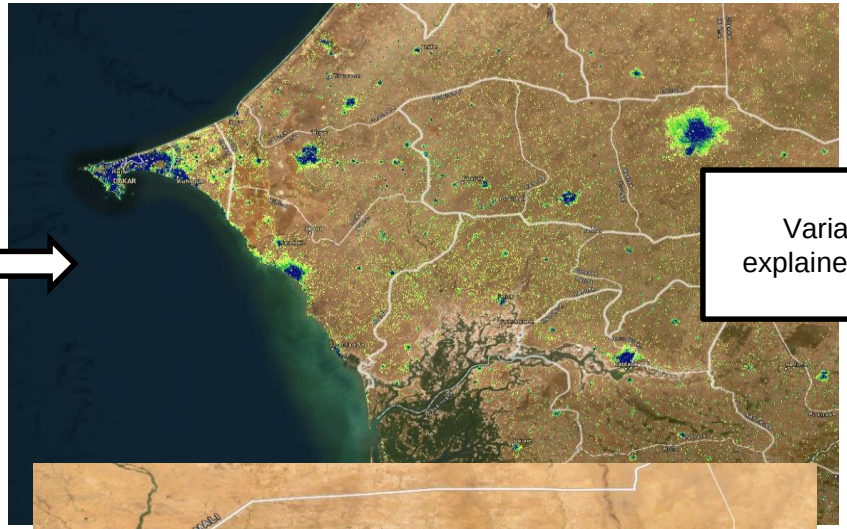


- Interest in small area estimates that match official projections, UN estimates
- Projections using cohort component, mathematical, economic methods
- Census-based, use of administrative records, displacement surveys etc, followed by top-down disaggregation

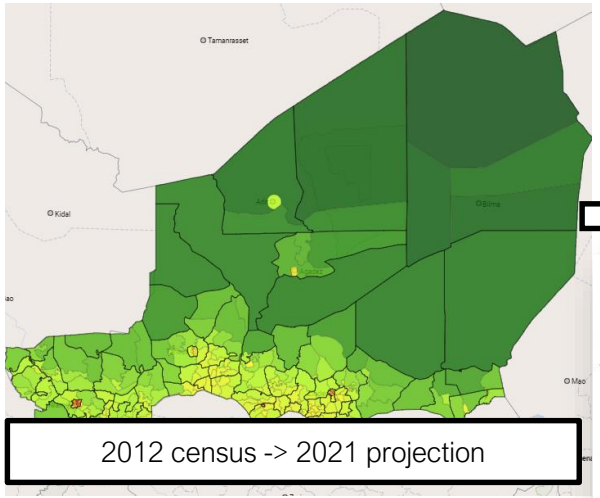


2013 census -> 2020 projection

ANSD
Agence Nationale de la Statistique et de la Démographie

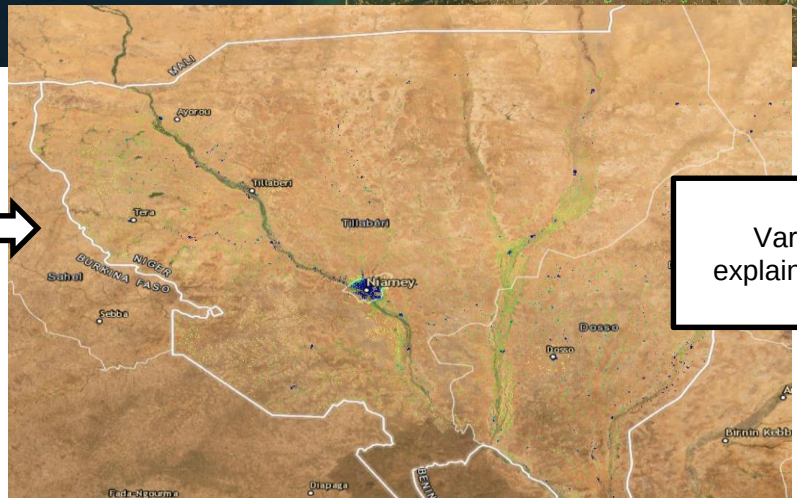


Variance explained: 98%



2012 census -> 2021 projection

Institut National de la Statistique NIGER

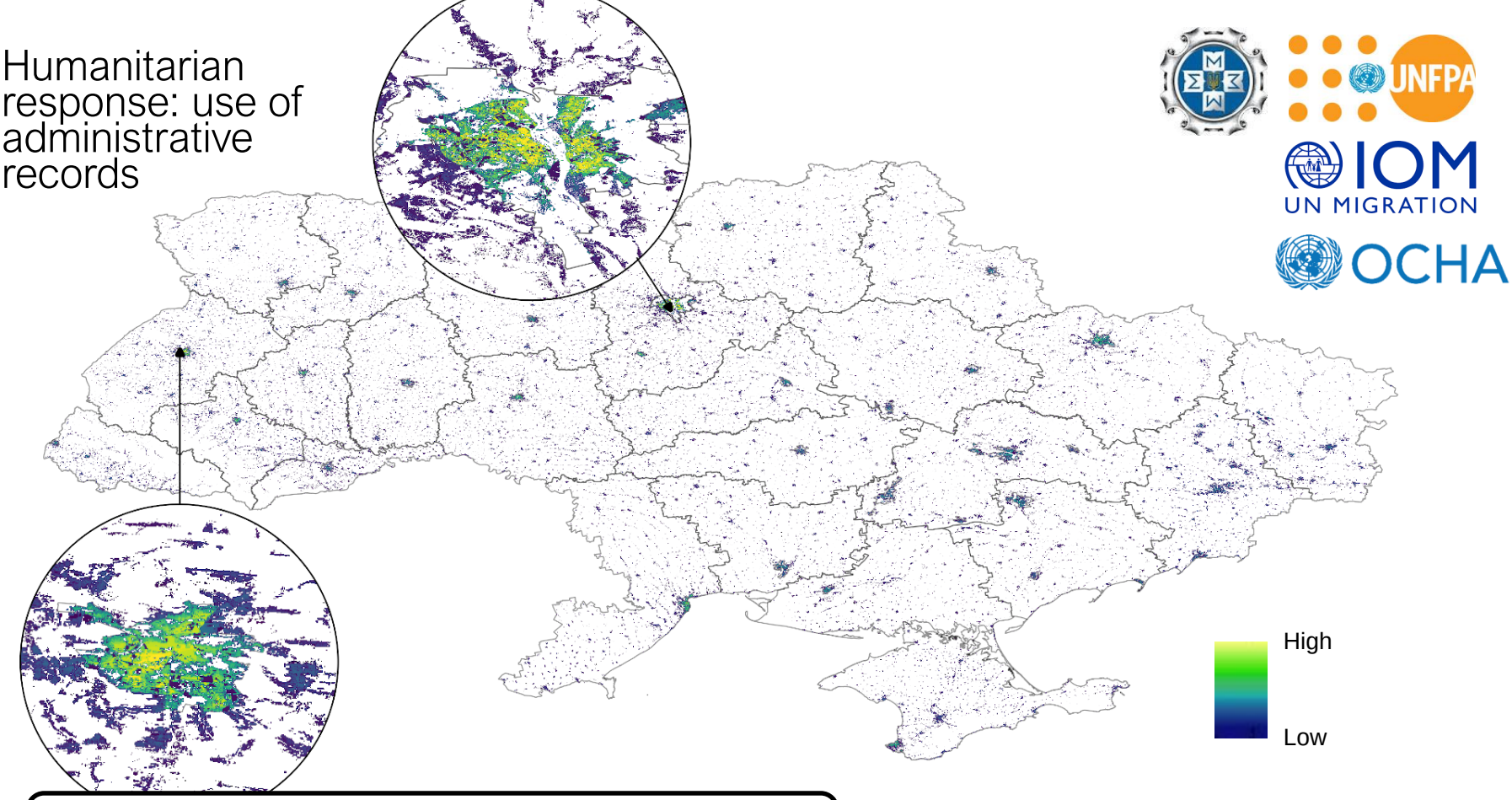


Variance explained: 92%

https://wopr.worldpop.org/?/Population

GRID3
GEO-REFERENCED INFRASTRUCTURE AND DEMOGRAPHIC DATA FOR DEVELOPMENT

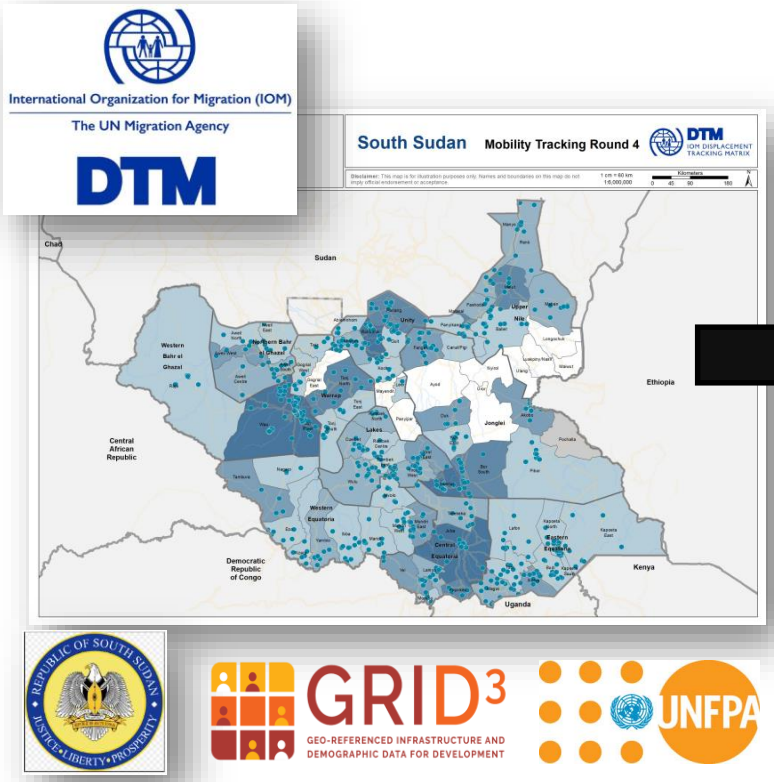
Humanitarian response: use of administrative records



Age/sex structured 2020 baseline residential population estimates per 100 x 100m grid square

https://data.humdata.org/dataset/gridded-population-estimates-for-ukraine-using-un-cod-ps-estimates-2020-version-2-0

Accounting for displacement and refugee camps



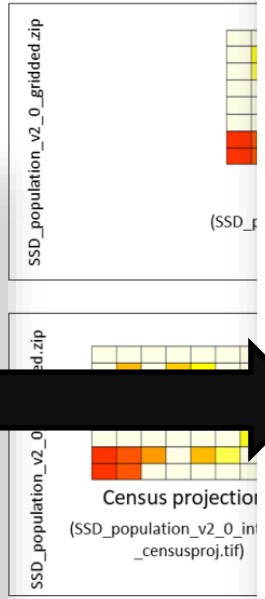
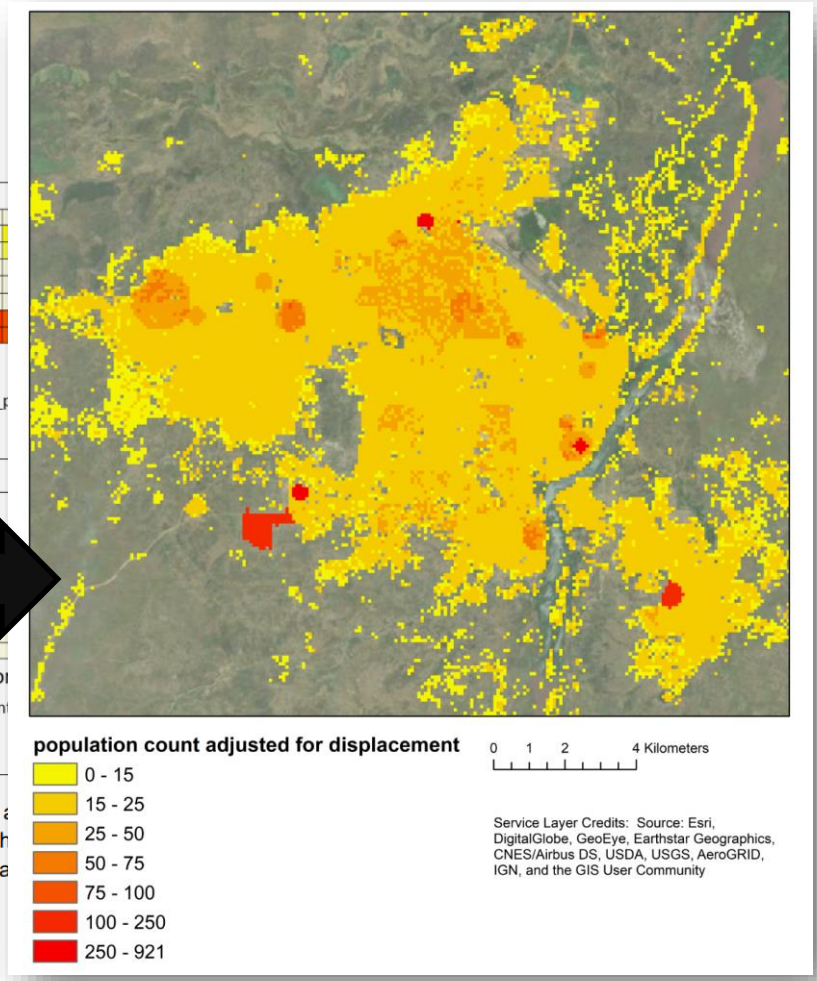


Figure 1. Schematic of the a
for displacement, for South
a specific location in the da



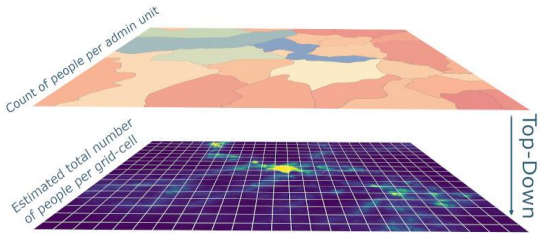
Tatem et al (2023) IOM Handbook; Dooley et al (2021) https://wopr.worldpop.org/?SSD/Population/v2.0

Summary: Top-down models

- If we have administrative unit level population counts that we trust, but need smaller area estimates, the top-down disaggregation approach can be used

- Relationships between population density and high resolution geospatial covariates are used to ‘disaggregate’ population counts from large areas to small area estimates, ensuring that large area totals are maintained

- The approach can be used with population projections, demographic sub-groups or key populations of interest



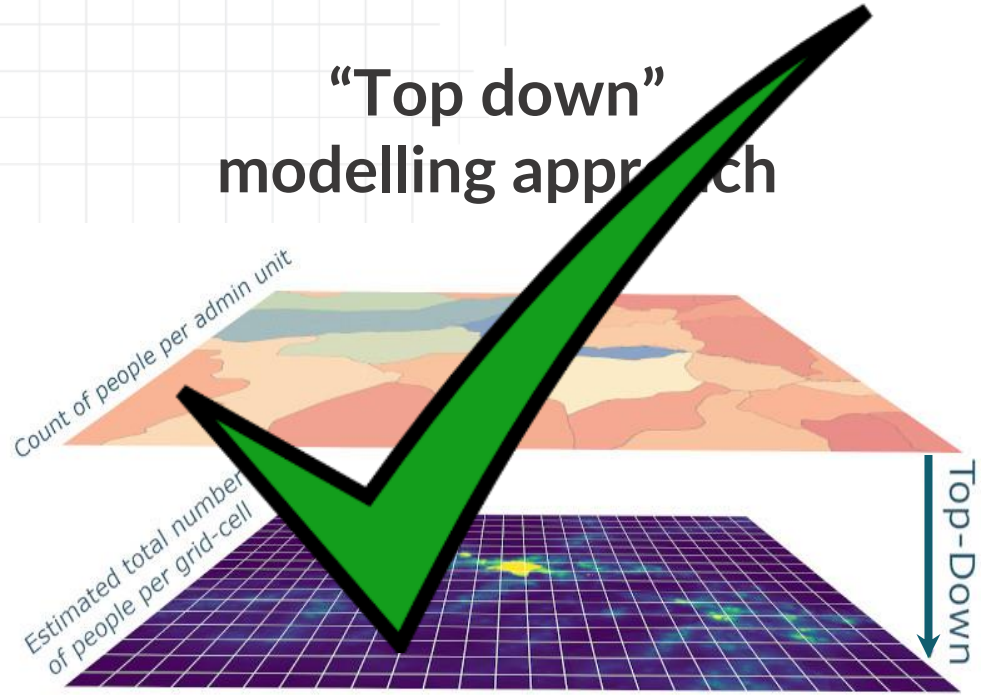
Questions?

To think about:

-Are there situations in your country where top-down modelled estimates could be valuable?

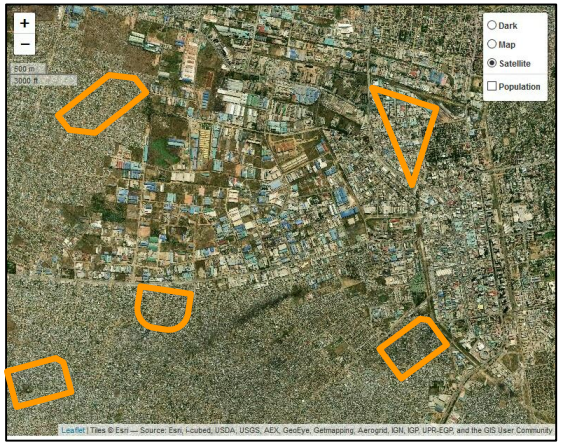
-What types of geospatial datasets would be important to use? (e.g. building damage maps, refugee camps, waterpoint locations….?)

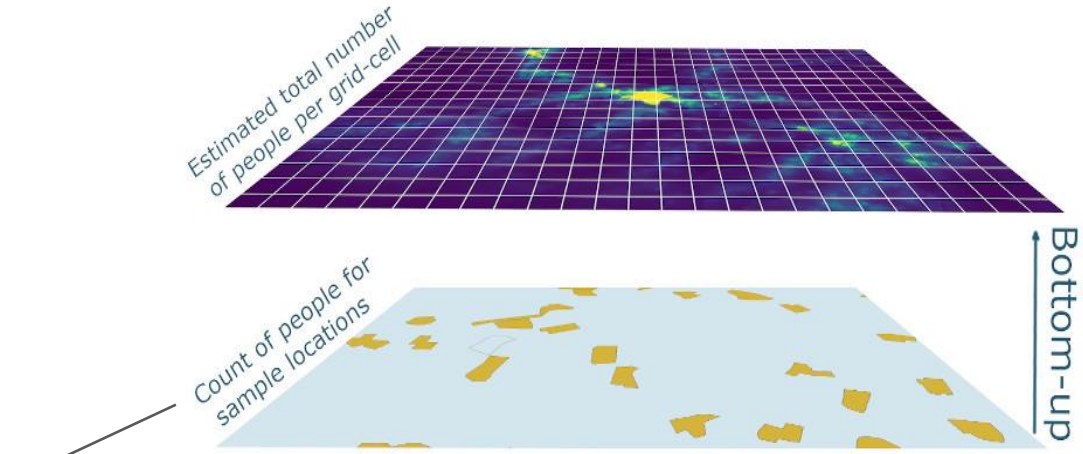
“Top down” modelling approach



www.worldpop.org/methods/populations

“Bottom-up” modelling approach

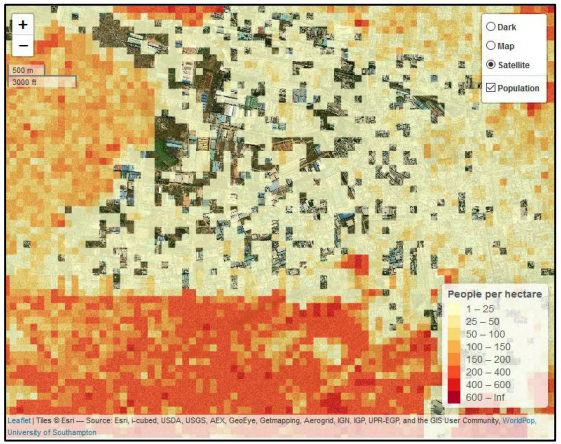


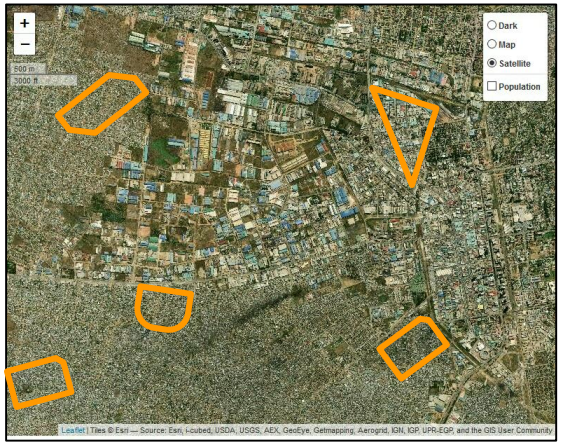


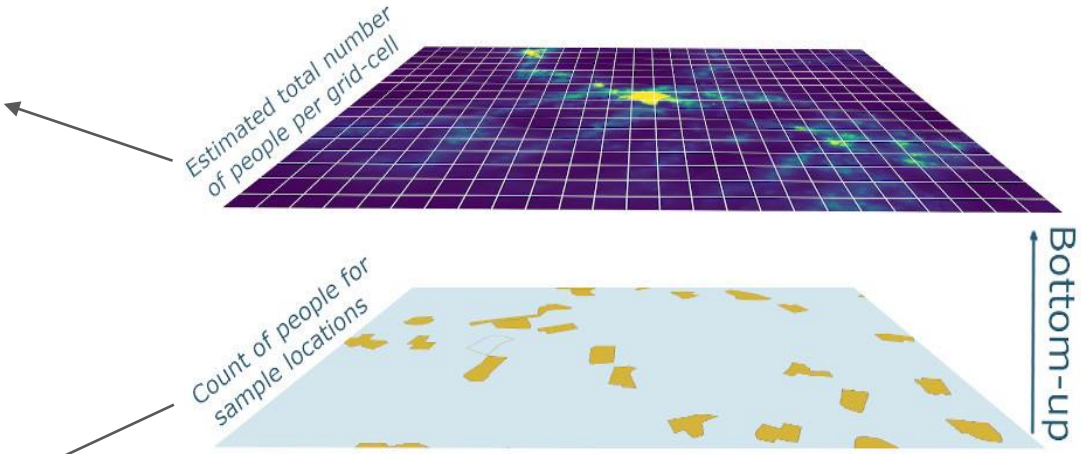
WorldPop

www.worldpop.org/methods/populations

“Bottom-up” modelling approach

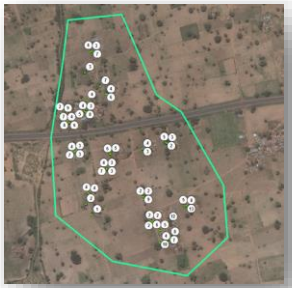


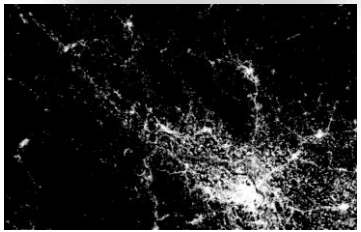


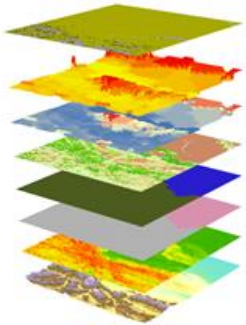


WorldPop

www.worldpop.org/methods/populations







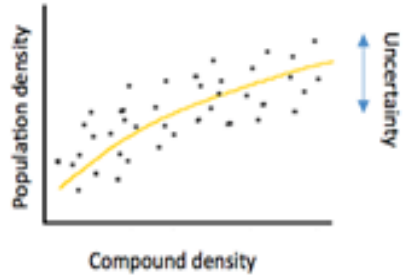
Population Data

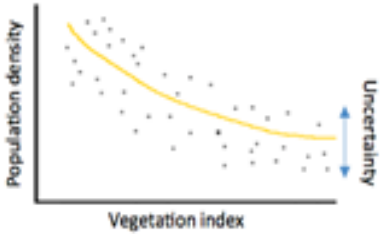
Buildings / Settlement

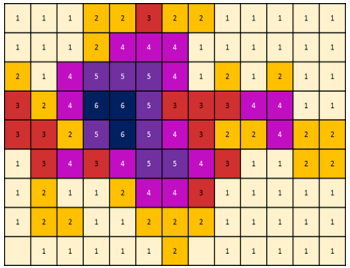
Geospatial data stack

Model

Population estimates







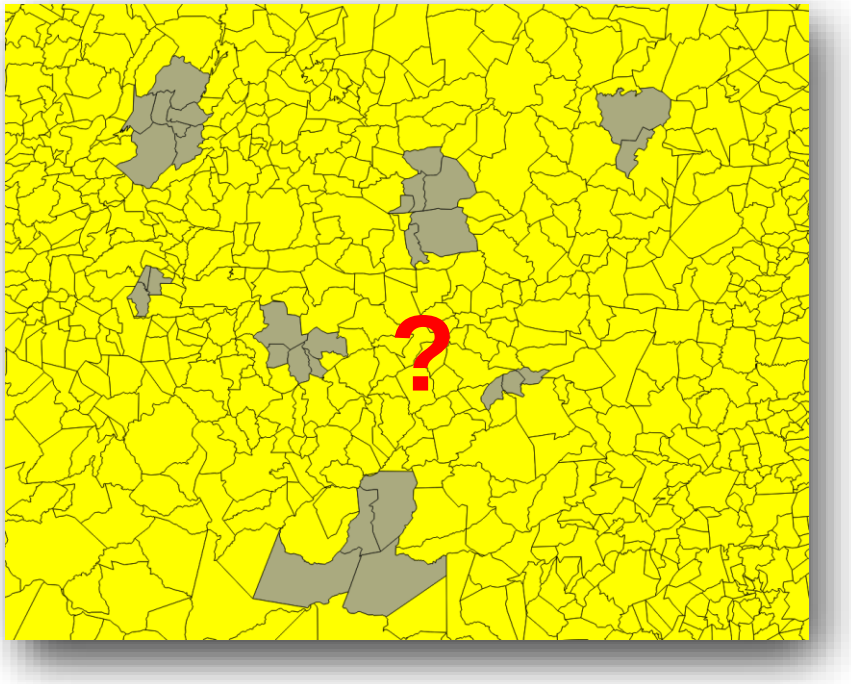
Wardrop et al (2018) PNAS

Bottom-up population estimates

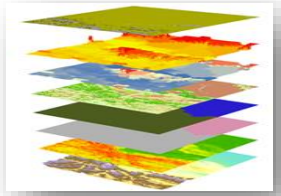
Population Data

Buildings / Settlement

Geospatial Data Stack



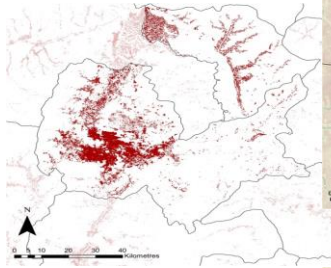
Bottom-up population estimates



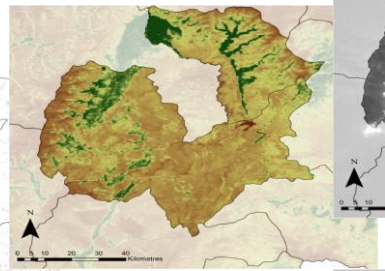
Population Data

Buildings / Settlement

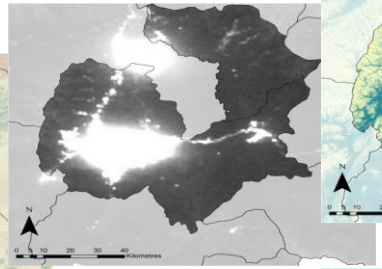
Geospatial Data Stack



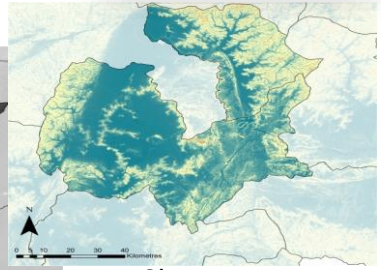
Number of compounds & area settled



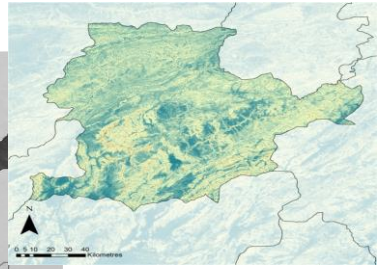
Vegetation index

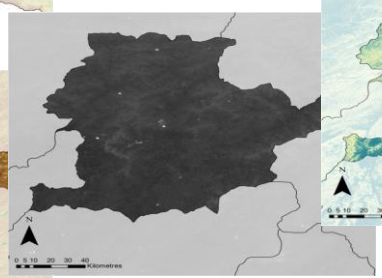


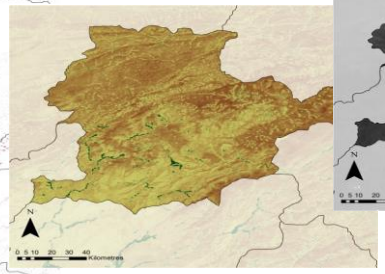
Night-time lights

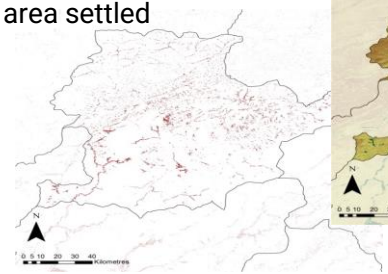


Slope

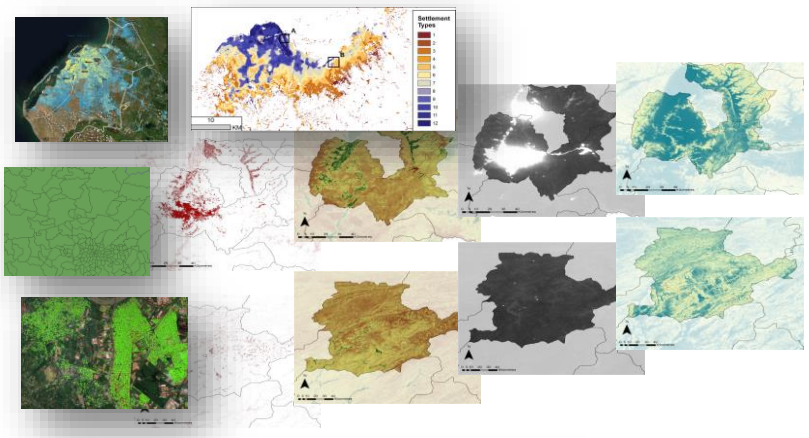


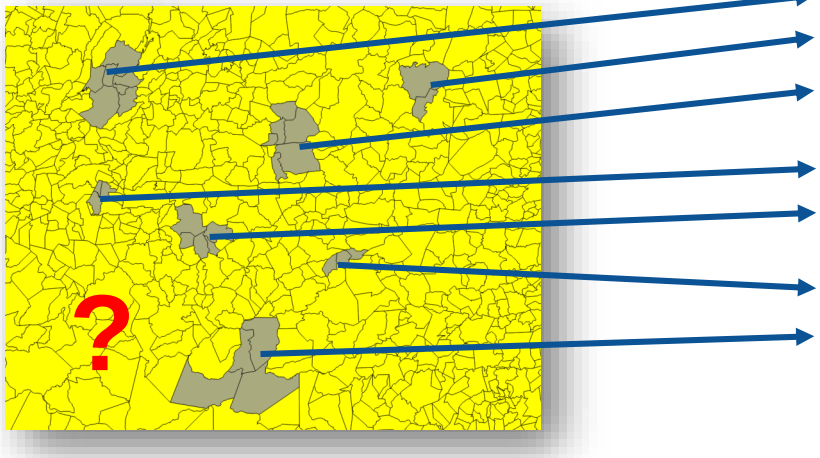


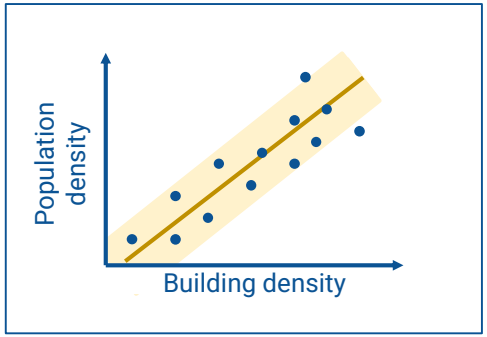


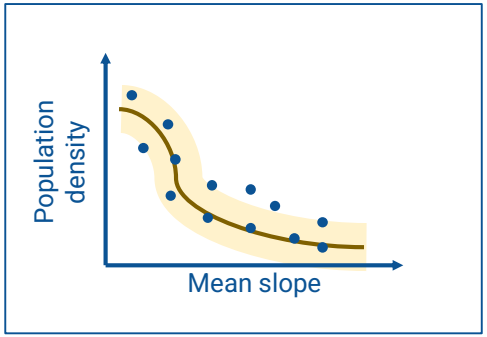


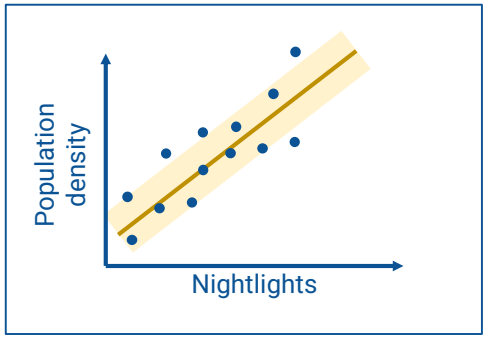
Bottom-up population estimates

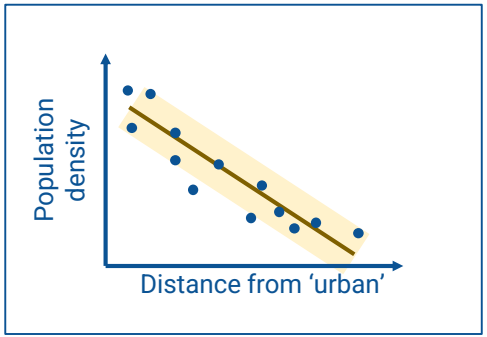










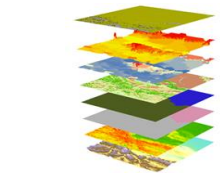


Bottom-up population estimates

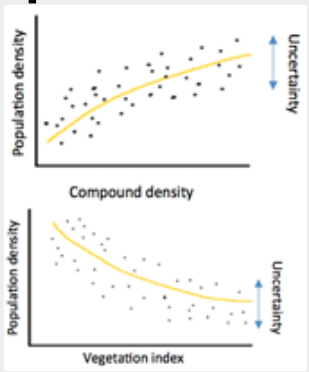
Population Data

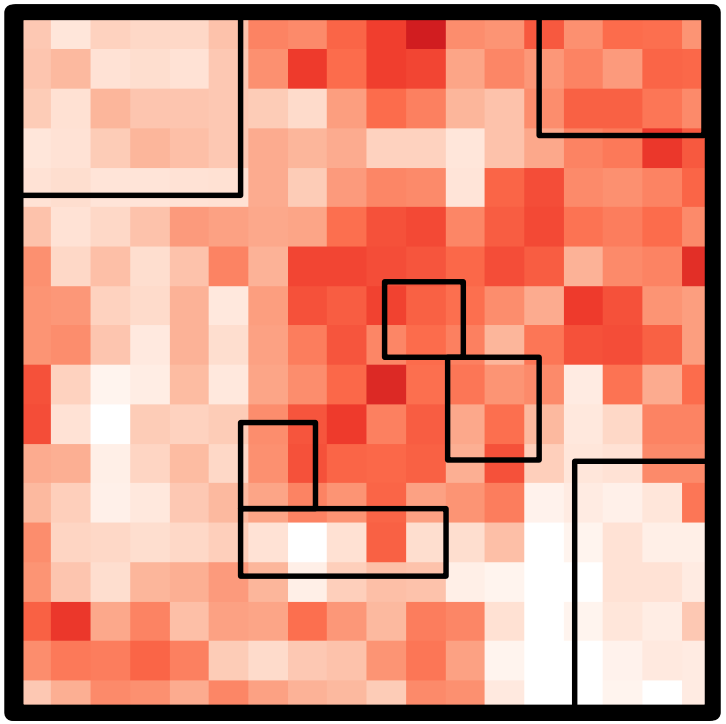
Buildings / Settlement

Geospatial Data Stack



Population model building and prediction

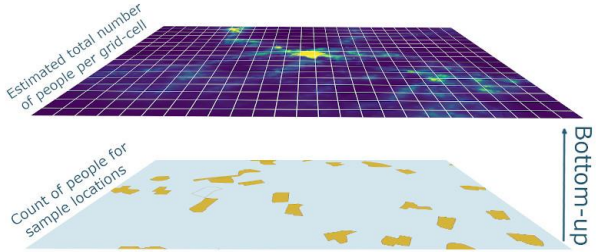




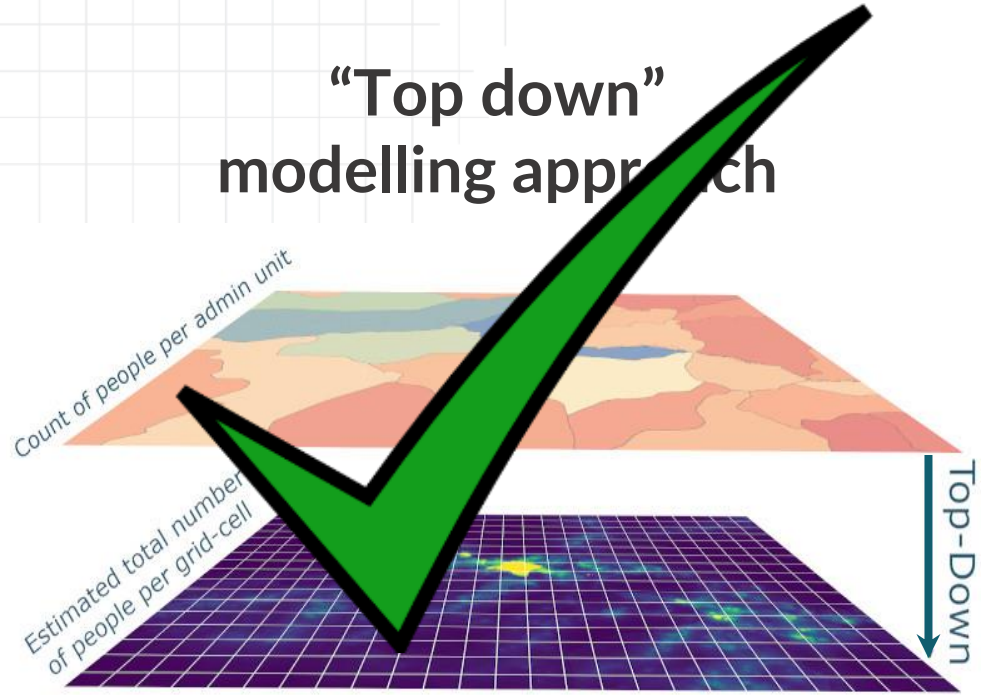
Gridded population estimates (100 m) that **fill gaps between surveyed areas**, with confidence intervals

Summary: Bottom-up models

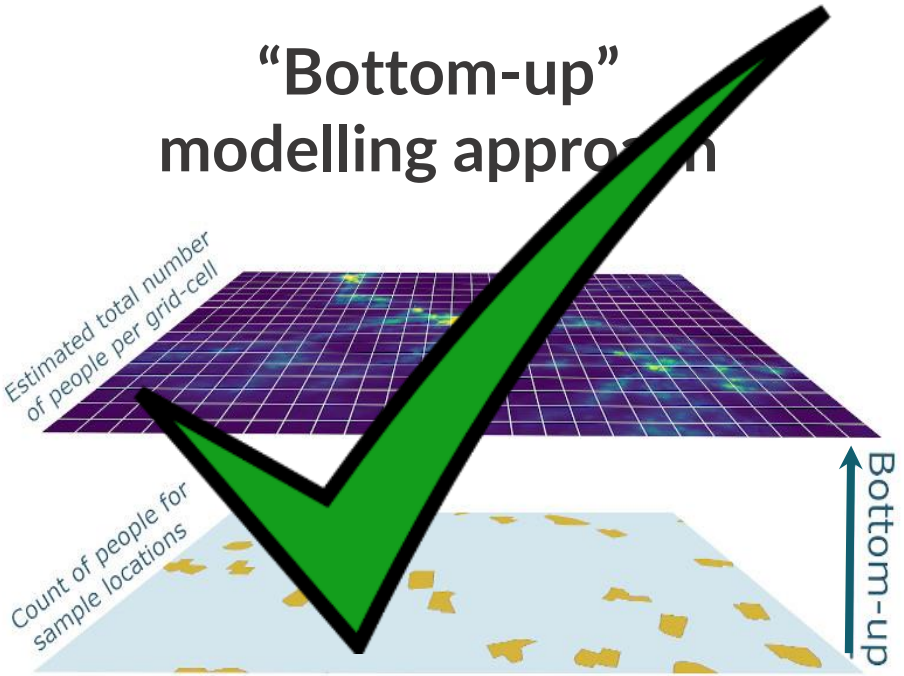
- If we have no recent population data that we trust, or only have incomplete data, then ‘bottom-up’ geospatial methods are an option for production of small area estimates
- Multiple types of models exist, and will likely need to be designed for the needs, situation and data available in a certain setting – often a slower and more intensive process than top-down modelling
- Important to quantify and communicate uncertainties and limitations
- Not a replacement for a census



“Top down” modelling approach



“Bottom-up” modelling approach



www.worldpop.org/methods/populations

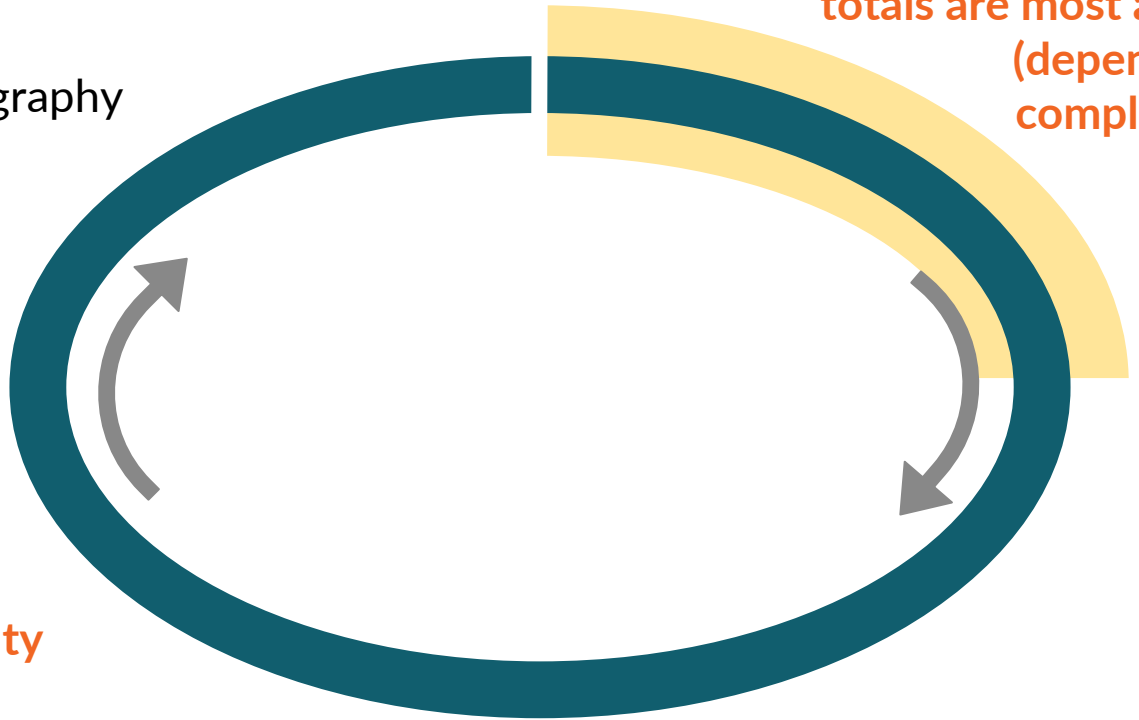
Stages of the census cycle

Census enumeration

Enumerated population totals are most accurate (dependent on completeness)

Census cartography

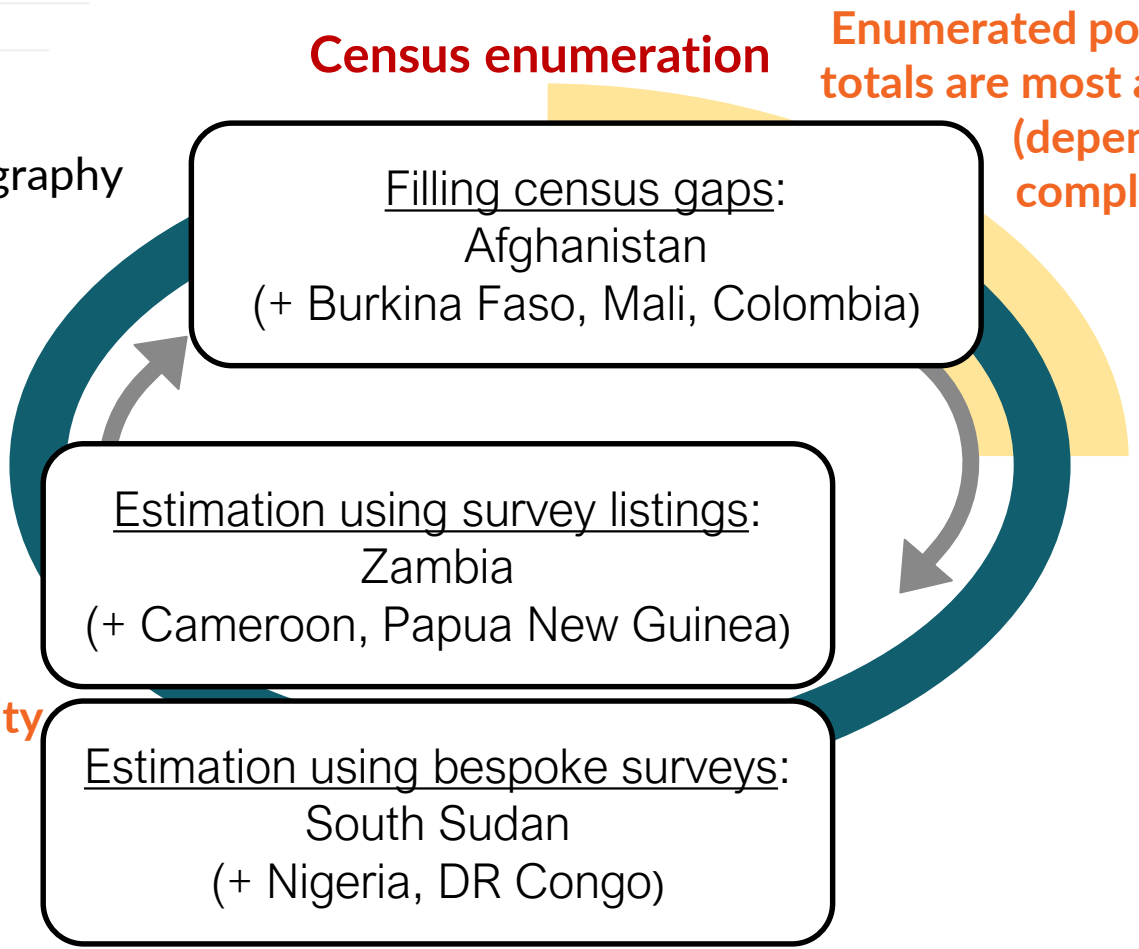
Pilot census cartography



As time from the census elapses, increasing uncertainty in projections

Intercensal period

Stages of the census cycle



Questions?

To think about (as you listen to case studies):

-Do the case studies presented have similarities to your own situations?

-If you see a need for it in your country, what types of population data might be available to support bottom-up estimation modelling?