

Release Statement

Gridded population estimates for Yemen using UN COD-PS estimates 2019, 2023, version 1.0

25/07/2024

These data were produced by [WorldPop](#) at the University of Southampton and [World Bank Group](#). These data include gridded estimates of population at approximately 100m for 2019 and 2023 along with estimates of the number of people belonging to individual age-sex groups. These results were produced using subnational population estimates for Yemen provided in the Common Operational Dataset on Population Statistics (COD-PS) [1,2] and Subnational Administrative Boundaries for Yemen provided by OCHA [3].

These data may be distributed using a [Creative Commons Attribution 4.0 International \(CC BY 4.0\)](#) License, specified in legal code. Contact release@worldpop.org for more information.

WorldPop, University of Southampton, and their sponsors offer these data on a "where is, as is" basis; do not offer an express or implied warranty of any kind; do not guarantee the quality, applicability, accuracy, reliability or completeness of any data provided; and shall not be liable for incidental, consequential, or special damages arising out of the use of any data that they offer.

CITATION

WorldPop and World Bank Group. 2024 Gridded population estimates for Yemen using UN COD-PS estimates 2019 and 2023, version 1.0. https://data.worldpop.org/repo/prj/WP_WB/YEM/v1/

FILE DESCRIPTIONS

The projection for all GIS files is the geographic coordinate system WGS84 (World Geodetic System 1984).

yem_pack1_{2019/23}.zip

This zip file contains 14 rasters in geotiff format at a spatial resolution of 3 arc-seconds (approximately 100m at the equator) contains estimates of total population size per grid cell across Yemen, along with estimates of the number of people belonging to 15-49 years age-sex group and children under 5. For unconstrained, NA values represent areas that were mapped as water and for constrained NA values represent areas that were mapped as unsettled. These data are stored as floating-point numbers rather than integers to avoid rounding errors in aggregated populations for larger areas.

yem_pack2_{2019/23}.zip

This zip file contains 112 rasters in geotiff format at a spatial resolution of 3 arc-seconds (approximately 100m at the equator) contains estimates of total population size per grid cell across Yemen, along with estimates of the number of people belonging to individual age-sex groups. [Boys (0-17), Men (18+), Girls (0-17), Women (18+), Total /Women-Girl, Total MEN/Boys, 0 – 4... 75 – 79, 80+]. For unconstrained, NA values represent areas that were mapped as water and for constrained NA values represent areas that were mapped as unsettled. These data are stored as floating-point numbers rather than integers to avoid rounding errors in aggregated populations for larger areas.

METHODS

Pre-processing: Subnational Administrative Boundaries provided by OCHA were nibbled (i.e. cells with no data are replaced with the values of the nearest neighbors) to match the WorldPop mastergrid and avoid mismatch with the WorldPop covariates.

Modelling: Building height/area/volume per pixel were extracted from GHS [4] in addition to classifying pixels as residential or non-residential. The [constrained and unconstrained top-down](#) disaggregation method was used to produce the datasets. The Random Forest (RF)- based dasymetric mapping approach (Stevens et al., 2015 [5]) implemented in the popRF 'R' package [6] based on the Breiman (2001) [7] was used for the population modelling.

WORK CITED

1. Common Operational Dataset on Population Statistics for Yemen, 2019
<https://data.humdata.org/dataset/yemen-population-estimates>
2. Common Operational Dataset on Population Statistics for Yemen, 2023
<https://data.humdata.org/dataset/yemen-population-estimates>
3. Subnational Administrative Boundaries for Yemen provided by OCHA
<https://data.humdata.org/dataset/cod-ab-yem>
4. Pesaresi, Martino; Politis, Panagiotis (2022): GHS-BUILT-S R2022A - GHS built-up surface grid, derived from Sentinel-2 composite and Landsat, multitemporal (1975-2030). European Commission, Joint Research Centre (JRC) [Dataset] DOI:10.2905/D07D81B4-7680-4D28-B896-583745C27085 PID:<http://data.europa.eu/89h/d07d81b4-7680-4d28-b896-583745c27085>
5. Stevens FR, Gaughan AE, Linard C, Tatem AJ (2015) Disaggregating Census Data for Population Mapping Using Random Forests with Remotely-Sensed and Ancillary Data. PLoS ONE 10(2): e0107042. <https://doi.org/10.1371/journal.pone.0107042>
6. Bondarenko M., Nieves J.J., Forrest R.S., Andrea E.G., Jochem C., Kerr D., and Sorichetta A. (2021): popRF: Random Forest-informed Population Disaggregation R package, [_Comprehensive R Archive Network \(CRAN\)_url:https://cran.rproject.org/package=popRF](https://cran.rproject.org/package=popRF).
7. Breiman L. Random Forests. Mach Learn. 2001 [cited 2013 Jan 24].45(1):5–32. Available: <http://link.springer.com/article/10.1023/A:1010933404324>