Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets

December 2018

Center for International Earth Science Information Network (CIESIN)
Columbia University

Abstract

This document outlines the basic methodology used to construct the Gridded Population of the World, Version 4 (GPWv4) data collection and describes the data sets included in the collection, all of which have been updated as Revision 11 for this release. The Introduction briefly describes the input data, the purpose of the collection, the main characteristics of GPWv4, and lists the data sets in the collection. Details of the methodology, including the sources of the input data used to produce the data sets, are outlined in the Data and Methodology section. The Data Set Descriptions section describes the purpose and content of each data set, as well as available resolutions and formats. Additional sections of this documentation provide guidance on the use of the data, examples of potential use cases, and information on limitations and use constraints.

Suggested citation for this document:

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets. Palisades NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H45Q4T5F Accessed DAY MONTH YEAR.

We appreciate feedback regarding these data sets, such as suggestions, discovery of errors, difficulties in using the data, and format preferences.

Please submit comments to SEDAC User Services (https://sedac.uservoice.com/knowledgebase/topics/110829-gpwv4) or contact:

NASA Socioeconomic Data and Applications Center (SEDAC)
Center for International Earth Science Information Network (CIESIN)
Columbia University
Phone: 1 (845) 365-8920
Email: ciesin.info@ciesin.columbia.edu

Contents

I. Introduction 3
II. Data and Methodology 4
III. Data Set Descriptions 12
1. Population Count, v4.11 (2000, 2005, 2010, 2015, 2020) 14
2. Population Density, v4.11 (2000, 2005, 2010, 2015, 2020) 14
3. UN WPP-Adjusted Population Count, v4.11 (2000, 2005, 2010, 2015, 2020) 15
4. UN WPP-Adjusted Population Density, v4.11 (2000, 2005, 2010, 2015, 2020) 15
5. Data Quality Indicators, v4.11 (2010) 16
6. Land and Water Area, v4.11 (2010) 17
7. Administrative Unit Center Points with Population Estimates, v4.11 (2000, 2005, 2010, 2015, 2020) 18
8. National Identifier Grid, v4.11 (2010) 20
9. Basic Demographic Characteristics, v4.11 (2010) 22
IV. How to Use the Data 23
V. Potential Use Cases 24
VI. Limitations 25
VII. Acknowledgments 27
VIII. Disclaimer 27
IX. Use Constraints 27
X. Recommended Citations 27
XI. Source Code 29
XII. References 29
XIII. Documentation Copyright and License 32
XIV. Appendices 32
Appendix A: Data Revision History 32
Appendix B: Contributing Authors & Documentation Revision History 34
Appendix C: Methodologies Used to Produce GPWv4 Ancillary Data Sets 36
Appendix D: Countries Not Adjusted to the Global Boundary Framework 38
Appendix E: Currency and Type of Population Data used in GPWv4 39
Appendix F: Administrative Levels of Primary Population, Sex, and Age Data 46
Appendix G: Countries for Which Growth Rates Were Calculated at Multiple Administrative Levels 52
Appendix H: Highest Age Group Class Available by Country 53

I. Introduction

The Gridded Population of the World (GPW) collection, now in its fourth version (GPWv4), models the distribution of human population on a continuous global raster surface. Since the release of the first version of this global population surface in 1995, the essential inputs to GPW have been population census tables and corresponding geographic boundaries. The purpose of GPW is to provide a spatially disaggregated population layer that is compatible with data sets from social, economic, and Earth science disciplines, and remote sensing. It provides globally consistent and spatially explicit data for use in research, policy-making, and communications.

The fourth version of GPW (GPWv4) is a raster data collection of globally integrated national population data from the 2010 round of Population and Housing Censuses, which occurred between 2005 and 2014. The input data are extrapolated to produce population estimates for the years 2000, 2005, 2010, 2015, and 2020. A set of estimates adjusted to national level, historic and future, population predictions from the United Nation's World Population Prospects report are also produced for the set same set of years. The raster data sets are constructed from the national or subnational input administrative units to which the estimates have been matched. GPWv4 is gridded with an output resolution of 30 arc-seconds, or approximately 1 kilometer at the equator.

The nine data sets of the current release are collectively referred to as the Revision 11 (or v4.11) data sets. In this release, several issues identified in the 4.10 release of November 2017 have been corrected as follows:

- The extent of the final gridded data has been updated to a full global extent.
- Erroneous no data pixels in all of the gridded data were recoded as 0 in cases where census reported known 0 values.
- The netCDF files were updated to include the Mean Administrative Unit Area layer, the Land Area and Water Area layers, and two layers indicating the administrative level(s) of the demographic characteristics input data.
- The National Identifier Grid was reprocessed to remove artifacts from inland water. In addition, two attributes were added to indicate the administrative levels of the demographic characteristics input data, and the data set zip files were corrected to include the National Identifier Polygons shapefile.
- Two new classes (Total Land Pixels and Ocean Pixels) were added to the Water Mask.
- The administrative level names of the Greece Administrative Unit Center Points were translated to English.

Separate rasters are available for population counts and population density consistent with national censuses and population registers, or alternative sources in rare cases where no census or register was available. All estimates of population counts and population density have also been nationally adjusted to population totals from the United Nation’s *World Population Prospects: The 2015 Revision* (United Nations, 2015). In addition, rasters are available for basic demographic characteristics (age and sex), data quality indicators, and land and water areas. A vector data set of the center point locations

(centroids) for each of the input administrative units and a raster of national level numeric identifiers are included in the collection to share information about the input data layers.

The raster data sets are now available in ASCII (text) format as well as in GeoTiff format. Five of the eight raster data sets are also available in netCDF format. In addition, the native 30 arc-second resolution data were aggregated to four lower resolutions (2.5 arc-minute, 15 arc-minute, 30 arc-minute, and 1 degree) to enable faster global processing and support of research communities that conduct analyses at these resolutions. The data are available at all five resolutions in ASCII and GeoTiff format. NetCDF files are available at all resolutions except 30 arc-second. Additional details can be found in Section III. Data Set Descriptions.

The nine Revision 11 (v4.11) data sets can be downloaded from the GPWv4 web page (http://sedac.ciesin.columbia.edu/data/collection/gpw-v4/sets/browse):

1. Population Count, v4.11 (2000, 2005, 2010, 2015, 2020)
2. Population Density, v4.11 (2000, 2005, 2010, 2015, 2020)
3. UN WPP-Adjusted Population Count, v4.11 (2000, 2005, 2010, 2015, 2020)
4. UN WPP-Adjusted Population Density, v4.11 (2000, 2005, 2010, 2015, 2020)
5. Data Quality Indicators, v4.11 (2010)
6. Land and Water Area, v4.11 (2010)
7. Administrative Unit Center Points with Population Estimates, v4.11 (2000, 2005, 2010, 2015, 2020)
8. National Identifier Grid, v4.11 (2010)
9. Basic Demographic Characteristics, v4.11 (2010)

Appendix A contains a log of changes to the data sets by version. Appendix B describes the version history of the GPWv4 collection documentation (i.e., this document).

For the latest data releases, collection information, and other important news, follow SEDAC on Twitter (http://www.twitter.com/ciesin), Facebook (https://www.facebook.com/socioeconomicdataandappsctr), and the CIESIN YouTube channel (https://www.youtube.com/channel/UCjUjAvV7M04SxxpM5wq4fMw).

II. Data and Methodology

GPWv4 is a minimally-modeled gridded population data collection. Population estimates are created by extrapolating the raw census estimates to a series of target years: 2000, 2005, 2010, 2015, and 2020. Additionally, a set of estimates that have been nationally adjusted to data from the United Nation's *World Population Prospects: The 2015 Revision* (United Nations, 2015) are included for each of the target years. Both sets of estimates are proportionally allocated to raster cells (i.e., pixels) using a uniform areal-weighting approach to produce the population surfaces.

There are a number of more highly-modeled methods, including dasymetric modeling and smart interpolation (Hay et al., 2005), that incorporate additional geographic data. These data are used to produce weight matrices for determining how to apportion population by pixel. Several global data products use ancillary data in their spatial modeling, incorporating remotely sensed data on land cover, urban extent, accessibility, or all of the above in order to generate population surfaces (Balk et al., 2006; Bhaduri et al., 2002; Freire et al., 2016; Tatem et al., 2007).

The main benefit of disaggregating demographic variables by areal-weighting is the maintenance of fidelity to the input data. Census information modeled with this approach may be freely and easily incorporated into global analyses that make use of ancillary data sets that might be endogenous to more highly-modeled population surfaces. The modeled census information is also suitable for use in dasymetric and other modeling approaches—the population counts can be reallocated based on other layers.

The following two sections describe the input data sets and specific methodology used to produce the GPWv4 Revision 11 data sets. For the methodology used to produce the ancillary data sets included in the GPWv4 data collection, please see Appendix C.

A description of the history of GPW and the development of version 4 is available in Doxsey-Whitfield et al. (2015).

Input Data

The GPWv4 data collection is produced using multiple input data sets, including geographic boundaries, census data, United Nations population estimates, and a global water mask. The sources and uses of the input data are described below.

Population and boundary data

The two basic inputs of GPW are non-spatial population data (i.e., tabular counts of population listed by administrative area) and spatially-explicit administrative boundary data (i.e., shapefiles or Esri feature classes). The complete list of sources of population and boundary data used for each country can be downloaded in Microsoft Excel format by following the link to "Country-level Information and Sources" on the GPWv4 "Methods" web page: http://sedac.ciesin.columbia.edu/data/collection/gpw-v4/methods/method1.

Global framework boundaries

A global framework of international boundaries was used to ensure consistent alignment between countries. The Global Administrative Areas, version 2.0 (GADMv2) (GADM, 2011) data set was selected as the global framework. In cases where the resolution of the census geography obtained from other sources far exceeded the GADMv2 boundaries, the higher resolution geographies were used. Countries not adjusted to this framework are listed in Appendix D.

United Nations population estimates

The population estimates provided in the United Nation's *World Population Prospects: The 2015 Revision* (United Nations, 2015) often correct for over- or under-reporting in the nationally-reported figures. Therefore, the GPWv4 national-level population estimates created for the target years 2015 and 2020 were adjusted to the medium-variant UN projections, and those created for the years 2000, 2005, and 2010 were adjusted to the historic UN estimates, to produce population rasters that match the UN country totals.

Water mask

The water mask was used to exclude areas of water and permanent ice from the population allocation. Features in the water mask originate from global water layers (Table 1), country water layers (Table 2), and water features included in any input administrative boundary data.

Table 1. Global sources of water body data.

Source Data Set Title	Reference
Shuttle Radar Topology Mission, Version 2 (SRTM) Water Body Data Files	(United States Geological Survey, 2003)
Global Lakes and Wetlands Database	(Lehner and Döll, 2004)
Global Land Ice Measurements from Space (GLIMS) Glacier Database, Version 1	(GLIMS and NSIDC, 2005, updated 2015)
Land Cover Project: Water Bodies, Version 4.0	(European Space Agency Climate Change Initiative, 2016)

Table 2. Country-level sources of water body data.

Country	Data Set Title	Reference
Canada	Lakes and Rivers	(Statistics Canada, 2011)
Finland and the Aland Islands	General Map, 1 to 1M: Vesi Alue.	(Maanmittauslaitos, 2010)
Iceland	IS 500V Vatn flákar	(Landmælingar Íslands, 2012)
Norway	N1000 Kartdata	(Kartverket, 2014)
Puerto Rico	Areal Hydrography	(United States Census Bureau Geography Division, 2016)
Svalbard and Jan Mayen Islands	S100 Kartdata (Svalbard 1:100,000 basemap themes)	(Norwegian Polar Institute, 2014)
Sweden	Hydrografi	(Lantmäteriet, 2016)
United States of America	Areal Hydrography	(United States Census Bureau Geography Division, 2016)

Methods

The development of GPWv4 builds upon previous versions of the data collection (Balk et al., 2006; Deichmann et al., 2001; Tobler et al., 1997) and follows the steps described below. The first two steps are generally completed manually, while the remaining steps are automated, where possible, via Python scripts (Python Software Foundation, 2014a; Python Software Foundation, 2014b), incorporating GDAL software (GDAL, 2014) and Esri ArcPy site package modules (Esri, 2014).

1. *Locate tabular population counts*

Tabular population input data were collected at the highest resolution available from the results of the 2010 round of censuses, which occurred between 2005 and 2014. These were collected from hundreds of national statistics offices and other organizations. Where census results were unavailable or not yet released, official population estimates from national statistics offices or the United Nations were used. Appendix E lists the currency and type of data (e.g., census, population register, official estimates) used for each country. In some cases, multiple levels of administrative data were used for a given country. For example, data for Paris, France were available at a higher resolution than for the rest of the country, as were matching boundary data. Therefore, the higher level administrative data for the city were merged with the lower level data. Appendix F provides the administrative level(s) of the input data used for each country.

2. *Match population counts to geographic boundaries (census or administrative)*

Geographic boundary data were collected from a variety of national agencies (e.g., statistics offices, mapping agencies, planning agencies), as well as other organizations, such as humanitarian agencies (e.g., UN OCHA). Ideally, the boundaries available for a given country were those originating with the census. In the absence of official census boundaries, other administrative boundary data were used. The population census counts or official estimates were then matched to the digital geographic boundaries. Matching was based on common identifying codes or unit names used in the census and was accomplished in Microsoft Excel (Microsoft, 2013).

3. *If needed, adjust boundaries to the global framework*

A global framework of international boundaries was used to ensure consistent alignment between countries. The Global Administrative Areas, version 2.0 (GADM, 2011) data set was selected as the framework as it is publicly available and frequently used in the research community. The international boundaries of census geography data sets were adjusted to the GADMv2 framework, although in cases where the resolution of the census geography far exceeded the GADMv2 boundaries, the former were kept (e.g., New Zealand, the United Kingdom, and the United States) (see Appendix D).

Boundary adjustments were either performed manually in consultation with reference sources, or via an automated procedure (i.e., a Python script). To validate automated adjustments, the initial area of each geographical unit was compared with its area following the boundary adjustment. If the boundary adjustment significantly changed the area of a unit, or cut a unit out altogether, then the boundary of that unit was manually evaluated and corrected. On a case-by-case basis, one of these strategies was implemented: the input boundaries were spatially adjusted by rubbersheeting to better match the framework; the framework was partially or wholly abandoned in favor of higher resolution sources; or, when no other solution could be found, the affected unit was aggregated with a neighboring unit.

4. Estimate population for target years

To facilitate global comparison of countries that conduct their censuses at different times, annualized growth rates were used to estimate counts for the target years of 2000, 2005, 2010, 2015, and 2020. Growth rates were calculated for each administrative unit by matching to a previous census enumeration or estimate using Microsoft Excel. Annualized rates of change were calculated as follows:

$$r = \frac{\ln\left(\frac{P_2}{P_1}\right)}{t}$$

where r is the annualized growth rate, P_1 is the population count at the time of the earlier census, P_2 is the population count from the current census, and t is the number of years between population counts.

During a quality review of the data, units with a growth rate higher than 10% or lower than −10% were examined more closely. If it was determined that an administrative boundary change caused the large (positive or negative) growth rate, a coarser resolution growth rate was substituted. If it was determined that the large growth rate reflected actual population growth or decline (as per the source data), it was retained.

Population estimates were then calculated for the target years as follows:

$$P_x = P_2 e^{rt}$$

where P_x is the population estimate in the target year x, and P_2, r, and t are as defined above.

For some countries, it was not possible to match at the highest resolution between the two points in time for one of the following reasons: substantial reorganization of administrative units took place between the two enumeration periods; previous census data were not released at the same resolution as the current census data; or only coarser geographies were comparable because high-resolution enumeration areas were created anew for each census. For these countries, censuses were matched and

growth rates were calculated at a coarser resolution (e.g., state), and applied to each nested higher resolution unit (e.g., county). In some cases, we adopted a hybrid approach, matching the highest resolution where possible and coarsening where needed. Appendix G lists the countries for which a hybrid approach was used.

5. Adjust population to United Nations estimates

The National-level estimates for 2000, 2005, 2010, 2015, and 2020 were then adjusted to the estimates of the United Nation's *World Population Prospects: The 2015 Revision* (United Nations, 2015). The medium-variant UN projections were used for the years 2015 and 2020, and historic estimates were used for years 2000, 2005, and 2010. These adjusted estimates are hereafter referred to as UN WPP-adjusted estimates.

Adjustment factors for matching national estimates to UN estimates were calculated as follows:

$$a = \frac{P_{UN}}{\sum P_x}$$

where a is the adjustment factor, P_x is the population estimate in the target year, and P_{UN} is the UN national estimate for the target year.

Adjustment factors were applied at the sub-national level as follows:

$$P_{adj} = P_x * a$$

where P_{adj} is the sub-national UN WPP-adjusted estimate, and P_x and a are as defined above.

For all years, a data quality check was completed wherein the national sums of the population estimates were compared to the national sums of the UN WPP-adjusted population estimates to identify countries with differences larger than 20%. These were examined more closely to determine the cause of the disparity. In some cases, errors in calculating growth rates or in tabular processing were uncovered and corrected. In other cases, no errors were found, and it was concluded that our source data might differ from that of the U.N..

6. Estimate population by age and sex

The estimation of the demographic variables, age and sex, for the year 2010 was accomplished through a process of applying age and sex proportions to the 2010 population estimates. This procedure was necessary to achieve full global coverage of cross-tabulated age by sex population count estimates for the year 2010 and to ensure that the sum of population across the year 2010 rasters for each variable equals the total population represented in the 2010 global population raster.

The demographic source data are predominantly census based and at the same geographic scale as the total population count data; however, for a number of countries, cross-tabulated variables were only available from alternative sources (see Appendix E) or at coarser geographic scales (see Appendix F). In these cases, proportions were applied to the 2010 population estimates in each nested higher resolution unit. For example, proportions calculated at the Province level were applied to each District within a given Province.

The procedure for estimating the demographic variables is detailed below.

First, estimates of the male and female population in 2010 were created by calculating the proportions of males and females in each geographic unit for the year of the input data, and then applying those proportions to the 2010 estimates of total population for each geographic unit, as follows:

$$P_m = \frac{c_m}{c_m + c_f} * P_t \qquad P_f = \frac{c_f}{c_m + c_f} * P_t$$

where *P* is the 2010 estimated population, *c* is the census population, and the subscripts *m*, *f*, and *t* refer to male, female, and total, respectively.

Second, if single-year age data were available for a given country, the data were aggregated into 5-year age groups. Next, estimates of the population by age and sex in 2010 were created by calculating the proportions of males and females in each 5-year age group for each geographic unit for the year of the input data, and then applying those proportions to the 2010 estimates of male and female population calculated in step one, as follows:

$$P_{m,i} = \frac{c_{m,i}}{\sum_i c_m} * P_m \qquad P_{f,i} = \frac{c_{f,i}}{\sum_i c_f} * P_f$$

where the subscript *i* refers to any age group in the set of all age groups, and *P*, *c*, *m*, and *f* are as defined above.

Third, five maximum age group classes were calculated from the age estimates: 65+, 70+, 75+, 80+, and 85+. Each of these classes was only calculated for countries with available data. The only maximum age group class with global coverage is the 65+ age group. Appendix H provides the highest age group class available by country.

Finally, the corresponding male and female age groups were summed to produce the estimated total population in that age group for year 2010:

$$P_{t,i} = P_{m,i} + P_{f,i}$$

For four countries (Benin, Laos, Malaysia, and Sri Lanka), age and sex data were available separately at high resolution, but cross-tabulated data were only available at

coarser resolution. In order to produce high resolution cross-tabulated variable estimates for these countries, three pre-processing steps were required.

First, the estimated female population by age group was calculated by applying the coarser resolution proportion of females in each 5-year age group to the population in the corresponding age group at the higher resolution:

$$e_{f,i} = \frac{b_{f,i}}{b_{m,i} + b_{f,i}} * c_{t,i}$$

where *e* is the estimated population, *b* is the coarser resolution census population, *c* is the higher resolution census population, the subscripts *m*, *f*, and *t* refer to male, female, and total, respectively, and the subscript *i* refers to any age group in the set of all age groups.

Second, the estimates were adjusted to ensure that the sum of females across the age groups was equal to the number of females reported in each unit:

$$d_{f,i} = \frac{c_f}{\sum_i e_{f,i}} * e_{f,i}$$

where *d* is the adjusted estimated population and *c*, *e*, *f*, and *i* are as defined above.

Third, the male population in each age group at the higher resolution was calculated by subtracting the adjusted estimate of female population from the total population in each high resolution age group:

$$d_{m,i} = c_{t,i} - d_{f,i}$$

These data were then processed using the same previously described procedures as for all other countries.

7. Transform to raster

To create the raster population count data sets, the population estimates were distributed to a 30 arc-second (~1 km at the equator) grid using an areal-weighting method. This method, also known as uniform distribution or proportional allocation, does not make use of any other geographic data in order to spatially disaggregate the census population. Population was allocated to the raster pixels (i.e., grid cells) through the simple assumption that the population of a pixel is the exclusive function of the land area within that pixel. For pixels that intersect sub-national or national boundaries, population was allocated based on the proportion of the area of the pixel occupied by each unit. A water mask was applied to the data to prevent lakes, rivers, and ice-covered areas from distorting the actual population distribution. Finally, the native 30 arc-second resolution data were aggregated to four lower resolutions (2.5 arc-minute, 15 arc-minute, 30 arc-minute, and 1 degree) to enable faster global

processing and support research communities that conduct analyses at these resolutions.

As a final validation, the population count rasters were summarized by country and those sums compared to the national totals in the tabular data to ensure no errors were made in processing.

III. Data Set Descriptions

All GPWv4 data sets are stored in geographic coordinate system (latitude/longitude). The raster data sets are available as GeoTiff and ASCII (text) files at five resolutions (Table 3). Five of the eight raster data sets are also available in netCDF format at all resolutions shown in Table 3 except 30 arc-second. NetCDF is an array-oriented open file format developed for the earth sciences. It allows for very rapid manipulation of very large data sets and is commonly used in climatology and meteorology. Several layers can be included in a netCDF file making them useful for both spatial and temporal analyses.

Table 3. Resolution equivalents in seconds, minutes, degrees, and kilometers

Seconds (sec) and Minutes (min)	Degrees (deg)	Kilometers (km)[1]
30 sec	0.008333 deg	~ 1 km
2.5 min	0.041667 deg	~ 5 km
15 min	0.25 deg	~ 30 km
30 min	0.5 deg	~ 55 km
60 min	1 deg	~ 110 km

Due to the file size of the 30 arc-second ASCIIs, each raster was split into eight tiles as depicted in Figure 1. Tiling of 30 arc-second ASCII rasters..

Each Revision 11 netCDF file includes data layers of the selected data set as well as fifteen data quality layers (Table 4). The file is packaged with a CSV file describing the contents of the netCDF, and four lookup tables in txt format that are needed to interpret the values in the corresponding rasters (Table 5).

[1]Values in kilometers are approximate at the equator. The cell size in kilometers from the equator to the North Pole varies due to the continual change in the width of longitudinal lines.

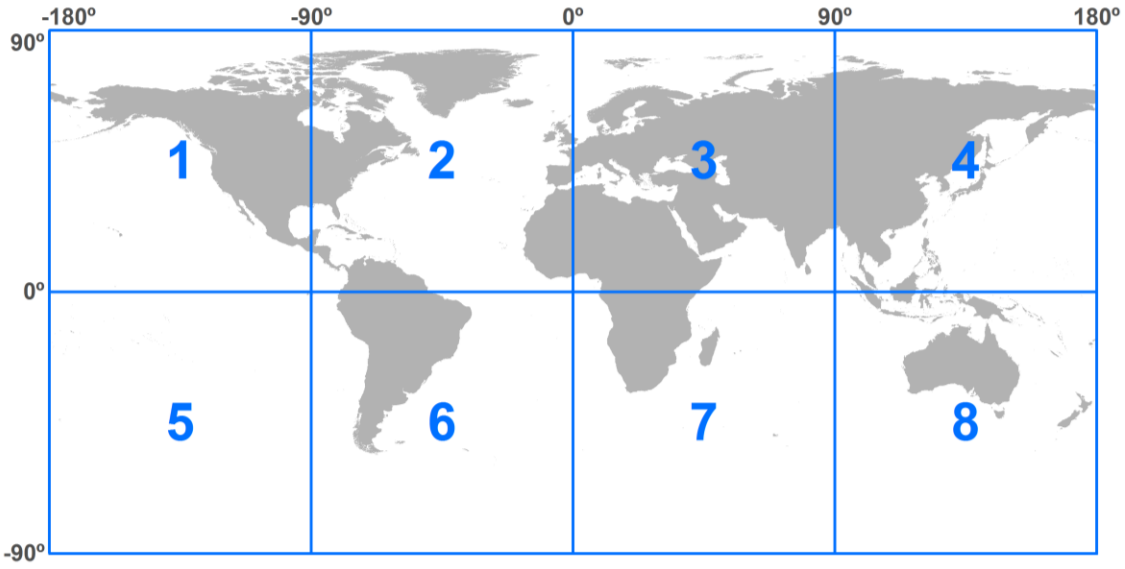

Figure 1. Tiling of 30 arc-second ASCII rasters.

Table 4. Rasters included in every netCDF download

Data Set (*one of the following*)	**Quality Layers (15)***
Population Count, v4.11	Data Context, v4.11 Mean Administrative Unit Area, v4.11 Water Mask, v4.11 Land Area, v4.11 Water Area, v4.11 National Identifier Grid, v4.11 **Data Code, v4.11** **Input Data Year, v4.11** **Input Data Level, v4.11** **Input Sex Data Level, v4.11** **Input Age Data Level, v4.11** **Growth Rate Start Year, v4.11** **Growth Rate End Year, v4.11** **Growth Rate Administrative Level, v4.11** **Year of Most Recent Census, v4.11**
Population Density, v4.11	
UN WPP-Adjusted Population Count, v4.11	
UN WPP-Adjusted Population Density, v4.11	
Basic Demographic Characteristics v4.11	
**Quality layers in bold were created from attributes of the National Identifier Grid, v4.11 data set and are only available for download as part of the netCDF files.*	

Table 5. Ancillary files included in every netCDF download

CSV describing contents of the netCDF	**Lookup Tables (4)**
Four Columns: *file_name* – file name of netCDF *order* – numerical order of rasters in the netCDF *raster_name* – name of the rasters in the netCDF *raster_description* – description of the rasters in the netCDF	gpw_v4_data_quality_indicators_rev11_context_lookup.txt gpw_v4_national_identifier_grid_rev11_datacode_lookup.txt gpw_v4_national_identifier_grid_rev11_lookup.txt gpw_v4_data_quality_indicators_rev11_watermask_lookup.txt

1. Population Count, v4.11 (2000, 2005, 2010, 2015, 2020)

Data set description:
The population count rasters consist of estimates of human population (number of persons per pixel), consistent with national censuses and population registers, for the years 2000, 2005, 2010, 2015, and 2020. The files for this data set were produced as global rasters at 30 arc-second horizontal resolution (approximately 1 km at the equator). To enable faster global processing, and in support of research communities, the 30 arc-second data were aggregated to 2.5 arc-minute, 15 arc-minute, 30 arc-minute and 1 degree resolutions.

Data set web page:
http://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-rev11

Data set format:
The files for this data set are available as global rasters in GeoTiff, ASCII (text), and netCDF format. Each downloadable is a compressed zip file containing either the global GeoTiff or ASCII for the year and resolution of the estimate, or the netCDF containing all years of the estimate at a selected resolution, the data quality layers (Table 4) and ancillary files (Table 5). A separate documentation zip file contains PDF documentation, a Microsoft Excel file (.xlsx) with country-level information and sources, and a text file (.txt) with a log of changes to the data set by version.

2. Population Density, v4.11 (2000, 2005, 2010, 2015, 2020)

Data set description:
The population density rasters consist of estimates of human population (number of persons per square kilometer), consistent with national censuses and population registers, for the years 2000, 2005, 2010, 2015, and 2020. The rasters were created by dividing the population count raster for a given target year by the land area raster. The files for this data set were produced as global rasters at 30 arc-second horizontal resolution (approximately 1 km at the equator). To enable faster global processing, and in support of research communities, the 30 arc-second data count data were aggregated to 2.5 arc-

minute, 15 arc-minute, 30 arc-minute and 1 degree resolutions to produce density rasters at these resolutions.

Data set web page:
http://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-rev11

Data set format:
The files for this data set are available as global rasters in GeoTiff, ASCII (text), and netCDF format. Each downloadable is a compressed zip file containing either the global GeoTiff or ASCII for the year and resolution of the estimate, or the netCDF containing all years of the estimate at a selected resolution, the data quality layers (Table 4) and ancillary files (Table 5). A separate documentation zip file contains PDF documentation, a Microsoft Excel file (.xlsx) with country-level information and sources, and a text file (.txt) with a log of changes to the data set by version.

3. UN WPP-Adjusted Population Count, v4.11 (2000, 2005, 2010, 2015, 2020)

Data set description:
The UN WPP-adjusted population count rasters consist of estimates of human population (number of persons per pixel), consistent with national censuses and population registers with respect to relative spatial distribution, but adjusted to match the 2015 Revision of the United Nation's *World Population Prospects* country totals for the years 2000, 2005, 2010, 2015, and 2020. The files for this data set were produced as global rasters at 30 arc-second horizontal resolution (approximately 1 km at the equator). To enable faster global processing, and in support of research communities, the 30 arc-second data were aggregated to 2.5 arc-minute, 15 arc-minute, 30 arc-minute and 1 degree resolutions.

Data set web page:
http://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-count-adjusted-to-2015-unwpp-country-totals-rev11

Data set format:
The files for this data set are available as global rasters in GeoTiff, ASCII (text), and netCDF format. Each downloadable is a compressed zip file containing either the global GeoTiff or ASCII for the year and resolution of the estimate, or the netCDF containing all years of the estimate at a selected resolution, the data quality layers (Table 4) and ancillary files (Table 5). A separate documentation zip file contains PDF documentation, a Microsoft Excel file (.xlsx) with country-level information and sources, and a text file (.txt) with a log of changes to the data set by version.

4. UN WPP-Adjusted Population Density, v4.11 (2000, 2005, 2010, 2015, 2020)

Data set description:
The UN WPP-adjusted population density rasters consist of estimates of human population density (number of persons per square kilometer), based on counts consistent with national censuses and population registers with respect to relative spatial

distribution, but adjusted to match the 2015 revision of the United Nation's *World Population Prospects* country totals, for the years 2000, 2005, 2010, 2015, and 2020. The rasters were created by dividing the UN WPP-adjusted population count raster for a given target year by the land area raster. The files for this data set were produced as global rasters at 30 arc-second horizontal resolution (approximately 1 km at the equator). To enable faster global processing, and in support of research communities, the 30 arc-second adjusted count data were aggregated to 2.5 arc-minute, 15 arc-minute, 30 arc-minute and 1 degree resolutions to produce density rasters at these resolutions..

Data set web page:
http://sedac.ciesin.columbia.edu/data/set/gpw-v4-population-density-adjusted-to-2015-unwpp-country-totals-rev11

Data set format:
The files for this data set are available as global rasters in GeoTiff, ASCII (text), and netCDF format. Each downloadable is a compressed zip file containing either the global GeoTiff or ASCII for the year and resolution of the estimate, or the netCDF containing all years of the estimate at a selected resolution, the data quality layers (Table 4) and ancillary files (Table 5). A separate documentation zip file contains PDF documentation, a Microsoft Excel file (.xlsx) with country-level information and sources, and a text file (.txt) with a log of changes to the data set by version.

5. Data Quality Indicators, v4.11 (2010)

Data set description:
The Data Quality Indicators data set consists of three data layers created to provide context for the population count and density rasters, and explicit information on the spatial precision of the input boundary data.

The Data Context raster explains pixels with "0" population estimate in the population count and density rasters, based on information included in the census documents. Pixels are categorized into eight categories as shown in Table 6.

Table 6. Data context code values and category descriptions

Data Context Value	Data Context Category	Population Value
0	Not applicable	> 0
201	Park or protected area	NoData
202	Military district, airport zone, or other infrastructure	NoData
203	Not enumerated or not reported in census	NoData
204	No households	0
205	Uninhabited	0
206	Population not gridded	NoData
207	Missing age and/or sex data	> 0 (BDC* NoData)

** BDC = Basic Demographic Characteristics data set.*

Pixels in categories 201-204 were explicitly identified as such in the census documentation or boundaries.

Pixels in category 205 ('Uninhabited') belong to units where the reported total population in the census was '0', but no further information was provided. A pixel in category 205 may be uninhabited because it is a national park or a military district, or for some other reason; however, due to the large number of units in GPWv4, we were unable to research each unit to explain why the population count was '0'.

Pixels categorized as 206 ('Population not gridded') belong to administrative units where the data could not be gridded due to a lack of information or data integration issues. The units are currently being researched and will be updated in future releases if possible.

The Water Mask raster distinguishes between pixels that are completely water and/or ice (Total Water Pixels), pixels that contain water and land (Partial Water Pixels), pixels that are completely land (Total Land Pixels), and pixels that are completely ocean water (Ocean Pixels).

The Mean Administrative Unit Area raster measures the mean input unit size in square kilometers. It provides a quantitative surface that indicates the size of the input unit(s) from which the population count and density rasters were created.

The files for this data set were produced as global rasters at 30 arc-second horizontal resolution (approximately 1 km at the equator). To enable faster global processing, and in support of research communities, the 30 arc-second data were aggregated to 2.5 arc-minute, 15 arc-minute, 30 arc-minute and 1 degree resolutions.

The methodology used to produce this ancillary data set is outlined in Appendix C.

Data set web page:
http://sedac.ciesin.columbia.edu/data/set/gpw-v4-data-quality-indicators-rev11

Data set format:
The files for this data set are available as global rasters in both GeoTiff and ASCII (text) format. Each downloadable is a compressed zip file containing one of the three data layers at the desired resolution and file format. The zip files for the Data Context and Water Mask layers in ASCII format also include a lookup table (.txt) that explains the meanings of the data values. A separate documentation zip file contains PDF documentation, a Microsoft Excel file (.xlsx) with country-level information and sources, and a text file (.txt) with a log of changes to the data set by version.

6. Land and Water Area, v4.11 (2010)

Data set description:
The Land and Water Area data set includes two rasters that represent surface areas of land and water in square kilometers per pixel. The Land Area raster provides estimates of

the land area, excluding permanent ice and water, within each pixel, and was used to calculate the population density rasters. The Water Area raster provides estimates of the water area (permanent ice and water), within each pixel. The sum of land area and water area of a pixel equals the total surface area of that pixel. The files for this data set were produced as global rasters at 30 arc-second horizontal resolution (approximately 1 km at the equator). To enable faster global processing, and in support of research communities, the 30 arc-second data were aggregated to 2.5 arc-minute, 15 arc-minute, 30 arc-minute and 1 degree resolutions.

The methodology used to produce this ancillary data set is outlined in Appendix C.

Data set web page:
http://sedac.ciesin.columbia.edu/data/set/gpw-v4-land-water-area-rev11

Data set format:
The files for this data set are available as global rasters in both GeoTiff and ASCII (text) format. Each downloadable is a compressed zip file containing one of the two data layers at the desired resolution and file format. A separate documentation zip file contains PDF documentation, a Microsoft Excel file (.xlsx) with country-level information and sources, and a text file (.txt) with a log of changes to the data set by version.

7. Administrative Unit Center Points with Population Estimates, v4.11 (2000, 2005, 2010, 2015, 2020)

Data set description:
The Administrative Unit Center Points (i.e., Centroids) data set consists of UN WPP-adjusted population estimates and densities for the years 2000, 2005, 2010, 2015 and 2020, the 2010 basic demographic characteristics (age and sex), and administrative name, land and water area, and data context by administrative unit center point (centroid) location (Table 5). The center points are based on approximately 13.5 million input administrative units used in GPWv4; therefore, these files require computers and software that can read large amounts of data into memory.

There are two types of centroids; true centroids and inside centroids. True centroids are the representative center of an input feature. However, this centroid may fall outside of its input feature (e.g., the true centroid of a unit with two islands will lie at a point in between the islands). Inside centroids are forced to be inside the input feature. Although the point locations in this data set represent the true centroid, the coordinates of both the true centroids (Table 7, CENTROID_X and CENTROID_Y) and the inside centroids (Table 7, INSIDE_X and INSIDE_Y) are included as attributes.

The methodology used to produce this ancillary data set is outlined in Appendix C.

Table 7. Field names and descriptions for the centroids data set

Field Name	Field Description
GUBID	Unique random (text) id
ISOALPHA	Three-letter country/state code
COUNTRYNM	English country/state name
NAME1 [a]	First administrative level name
NAME2 [a]	Second administrative level name
NAME3 [a]	Third administrative level name
NAME4 [a]	Fourth administrative level name
NAME5 [a]	Fifth administrative level name
NAME6 [a]	Sixth administrative level name
CENTROID_X	Longitude of the administrative unit center point in decimal degrees
CENTROID_Y	Latitude of the administrative unit center point in decimal degrees
INSIDE_X	Longitude of the administrative unit inside center point in decimal degrees
INSIDE_Y	Latitude of the administrative unit inside center point in decimal degrees
CONTEXT [b]	Data context value
CONTEXT_NM [b]	Data context category
WATER_CODE	Unit type code (L = Land unit, IW = Inland Water unit)
TOTAL_A_KM	Total area of the administrative unit in square km
WATER_A_KM	Water area of the administrative unit in square km
LAND_A_KM	Land area of the administrative unit in square km; this area field is used to calculate population density
UN_2000_E	UN WPP-adjusted population estimates 2000
UN_2005_E	UN WPP-adjusted population estimates, 2005
UN_2010_E	UN WPP-adjusted population estimates, 2010
UN_2015_E	UN WPP-adjusted population estimates, 2015
UN_2020_E	UN WPP-adjusted population estimates, 2020
UN_2000_DS	UN WPP-adjusted population density, 2000
UN_2005_DS	UN WPP-adjusted population density, 2005
UN_2010_DS	UN WPP-adjusted population density, 2010
UN_2015_DS	UN WPP-adjusted population density, 2015
UN_2020_DS	UN WPP-adjusted population density, 2020
B_2010_E	Population estimates, both sexes, 2010
F_2010_E	Population estimates, female, 2010
M_2010_E	Population estimates, male, 2010
AXX_YYB [c]	Estimated population ages XX to YY, both sexes, 2010
AZZPLUSB [c]	Estimated population ages ZZ and over, both sexes, 2010
AXX_YYF [c]	Estimated population ages XX to YY, female, 2010
AZZPLUSF [c]	Estimated population ages ZZ and over, female, 2010
AXX_YYYM [c]	Estimated population ages XX to YY, male, 2010
AZZPLUSM [c]	Estimated population ages ZZ and over, male, 2010

[a] Many countries are only available at coarser (first or second) administrative levels. A value of N.A. in the data indicates that a name is not available for that level.
[b] These fields refer to the data context layer in the Data Quality Indicators, v4.11 (2010) data set. For the full list of values for these fields, please see the details for this data set.
[c] For example, A05_09F refers to the 2010 estimated population for females ages 5–9.

Data set web page:
http://sedac.ciesin.columbia.edu/data/set/gpw-v4-admin-unit-center-points-population-estimates-rev11

Data set format:
The files for this data set are available at the global, continental and national levels. Global versions of the data are available in Esri Geodatabase, OGC GeoPackage (SQLite), and CSV formats and are packaged as five tiles: U.S.A. data by census regions (four tiles) and non-U.S.A. data (one tile). Continental versions of the data are also available in ESRI Geodatabase, OGC GeoPackage (SQLite), and CSV formats. As an alternative, the data are also available by country in shapefile, OGC GeoPackage (SQLite), and CSV formats. Due to file format size limitations, the U.S.A. data are packaged by state. Each downloadable is a compressed zip file containing either the shapefile, Esri Geodatabase, OGC GeoPackage (SQLite), or CSV for the desired geography. A separate documentation zip file contains PDF documentation, a Microsoft Excel file (.xlsx) with country-level information and sources, and a text file (.txt) with a log of changes to the data set by version.

8. National Identifier Grid, v4.11 (2010)

Data set description:
The National Identifier Grid is a raster representation of nation-states in GPWv4 for use in aggregating population data. This data set was produced from the input census units which are used to create a raster surface where pixels that cover the same census data source (most often a country or territory) have the same value. Note that these data are not official representations of country boundaries; rather, they represent the area covered by the input data. In cases where multiple countries overlapped a given pixel (e.g., on national borders), the pixels were assigned the country code of the input data set which made up the majority of the land area. This data set was produced as a global raster at 30 arc-second horizontal resolution (approximately 1 km at the equator). To enable faster global processing, and in support of research communities, the 30 arc-second data were aggregated to 2.5 arc-minute, 15 arc-minute, 30 arc-minute and 1 degree resolutions. Each level of aggregation results in the loss of one or more countries with areas smaller than the cell size of the final raster (Table 8). Rasters of all resolutions were also converted to polygon shapefiles with the attributes shown in Table 9.

The methodology used to produce this ancillary data set is outlined in Appendix C.

Data set web page:
http://sedac.ciesin.columbia.edu/data/set/gpw-v4-national-identifier-grid-rev11

Data set format:
The files for this data set are available as global rasters in both GeoTiff and ASCII (text) format. Each downloadable is a compressed zip file containing the GeoTiff or ASCII file and the polygon shapefile at the desired resolution. The zip files for ASCII format also

include a lookup table (.txt) that explains the meanings of the data values for country-level information. A separate documentation zip file contains PDF documentation, a Microsoft Excel file (.xlsx) with country-level information and sources, and a text file (.txt) with a log of changes to the data set by version.

Table 8. Countries lost during aggregation to lower resolution rasters

Country Lost		**Raster Resolution**			
Country Code	**Country Name**	**2.5 min**	**15 min**	**30 min**	**1 deg**
AND	Andorra		x	x	x
ABW	Aruba				x
BHR	Bahrain				x
HKG	China Hong Kong Special Administrative Region				x
MAC	China Macao Special Administrative Region		x	x	x
CUW	Curacao				x
GMB	Gambia				x
GIB	Gibraltar		x	x	x
GGY	Guernsey				x
VAT	Holy See	x	x	x	x
IMN	Isle of Man				x
LIE	Liechtenstein		x	x	x
LUX	Luxembourg				x
MCO	Monaco		x	x	x
BLM	Saint-Barthelemy				x
MAF	Saint-Martin (French part)		x	x	x
SMR	San Marino		x	x	x
SGP	Singapore				x
SXM	Sint Maarten (Dutch part)			x	x
PSE	State of Palestine				x

Table 9. Field names and descriptions for the National Identifier Grid polygons

Field Name	Field Description
GRIDCODE	CIESIN standard country or area code used for gridding
ISOCODE	Three-letter country/state code
UNSDCODE	United Nations standard country or area code for statistical use (M49)
NAME0	English country/state name
CIESINCODE	CIESIN standard country or area code for countries or areas without UNSD code. Identical to GRIDCODE
DATATYPE	Type of population data used: Final Census; Final Census and Population Register; Final Census, Population Estimate/Projection, Population Register; Preliminary/Provisional Census; Population Register; Population Estimate/Projection.
DATACODE	Code referencing a type of population data used: 1=Final Census; 2=Final Census and Population Register; 3=Final Census, Population Estimate/Projection; 4=Population Register; Preliminary/Provisional Census; 5=Population Register; 6=Population Estimate/Projection.
DATAYEAR	Year of the population data used
DATALEVEL	Highest level of administrative unit used for GPWv4 Revision 11 gridding; the administrative unit levels are categorized as follows: 0=country, 1=state/province, 2=county (or equivalent), and levels 3 through 6 representing progressively smaller units.
SEXLEVEL	Administrative level of input sex data used for GPWv4 Revision 11 gridding
AGELEVEL	Administrative level of input age data used for GPWv4 Revision 11 gridding
GRSTART	Earliest year used to calculate the annual exponential growth rate
GREND	Latest year used to calculate the annual exponential growth rate
GRLEVEL	Administrative level(s) used to calculate the annual exponential growth rate
LASTCENSUS	Year of the most recent census conducted in country/state
MEANUNITKM	Mean administrative unit area in square kilometers

9. Basic Demographic Characteristics, v4.11 (2010)

Data set description:
The Basic Demographic Characteristics rasters provide estimates of human population by age and sex as counts (number of persons per pixel) and densities (number of persons per square kilometer), consistent with national censuses and population registers, for the year 2010. To estimate the male and female populations by age in 2010, the proportions of males and females in each 5-year age group from ages 0–4 to ages 85+ for the given census year were calculated. These proportions were then applied to the 2010 estimates of the total population to obtain 2010 estimates of male and female populations by age. In some cases, the spatial resolution of the age and sex proportions was coarser than the

resolution of the total population estimates to which they were applied. The population density rasters were created by dividing the population count rasters by the land area raster. The files for this data set were produced as global rasters at 30 arc-second horizontal resolution (approximately 1 km at the equator). To enable faster global processing, and in support of research communities, the 30 arc-second data were aggregated to 2.5 arc-minute, 15 arc-minute, 30 arc-minute and 1 degree resolutions.

Data set web page:
http://sedac.ciesin.columbia.edu/data/set/gpw-v4-basic-demographic-characteristics-rev11

Data set format:
The files for this data set are available as global rasters in GeoTiff, ASCII (text), and netCDF format. Each downloadable is a compressed zip file whose contents vary depending on the chosen file format. For GeoTiff and ASCII, the data are packaged by age group at the desired resolution and include rasters of male, female, and total counts and densities. NetCDFs contain male, female or total population counts or densities at the desired resolution for all 5-year age groups up to 65+, as well as eleven data quality layers (Table 4). In addition, the zip files include five ancillary files (Table 5). A separate documentation zip file contains PDF documentation, a Microsoft Excel file (.xlsx) with country-level information and sources, and a text file (.txt) with a log of changes to the data set by version.

IV. How to Use the Data

The mean administrative unit area data layer in the Data Quality Indicators data set can help a user determine if GPWv4 is appropriate to use for their study. It provides the average size of the input unit from which population count and density rasters were created. Smaller input units provide a more detailed look at the population distribution throughout a country. The smaller the mean administrative unit area is, the more confident a user can be in using the population estimates within that area. If the unit of analysis is smaller than the mean administrative unit area values contained within the units of analysis, any population estimates made are much less likely to be precise enough for meaningful analysis.

Users should have the highest confidence using GPWv4 data in cases where their study site is larger than the mean administrative unit area for that location. Mean administrative unit area varies by country and within countries. Because the uniform areal weighting method does not include other layers to estimate a population distribution, in cases where a study site is smaller than the mean administrative unit area, the accuracy of a population estimate for that site must be regarded with lower confidence. Mean administrative unit area can ultimately be used as a metric for understanding the size of the input geographic units at a given location. A user can download this data set, run zonal statistics, and determine if their study area contains population data from multiple input units as opposed to a single unit.

The population count data are most useful for estimating the number of people within non-administrative geographies, such as land cover, elevation zones, watersheds, and ecologic units. By using a zonal sum function, it is possible to derive population estimates for any irregular geography where the unit of analysis is of a reasonable size. An important caveat is that the zonal sum function must have its resolution set to match that of GPWv4. If the cell size used by the zonal sum function is smaller than that of GPWv4, individual GPWv4 pixels will be included in the sum more than once – once for each smaller pixel within the GPWv4 pixel – and the population total will be an overestimate. If the zonal sum function uses a cell size larger than that of GPWv4, some GPWv4 pixels will be missed and the population total will be an underestimate.

The population count data are divided by the land area data to create the population density rasters for each respective year. The pixels used in GPWv4 are quadrilaterals and therefore the area of each pixel decreases with increasing latitude. Additionally, the pixels in the land area data that include water reflect the net land area (total area of pixel – area of water within the pixel). This means that the population density data cannot be multiplied by a fixed land area size to estimate population totals without introducing error in some or all of the pixels. Multiplying the population density values by pixel area will overestimate population in pixels that contain water. If a fixed value for pixel area is used, it will also over or underestimate population totals for pixels that are all land as the area of these pixels varies by latitude. The population density data are best used to determine statistics (minimum, maximum, mean, etc.) within varying geographies. Users who need population totals should use the GPWv4 population count data.

V. Potential Use Cases

The GPW data collection is a beneficial and valued asset for the research community. A number of research disciplines have used previous versions of GPW with a focus on human and environmental research. The data have been used in projects ranging from vulnerability mapping (Abson et al., 2012; Döll, 2009) and disaster impacts (Badal et al., 2005; Guha-Sapir et al., 2011) to health dimensions of environmental change (Caminade et al., 2014; Ermert et al., 2013). GPW's gridded format allows researchers to easily integrate the population data with a variety of other data sets, especially those produced through remote sensing.

GPW is different from other population data products because it is produced without using other geographic data to spatially disaggregate the census population. This allows it to be incorporated into global analyses that make use of ancillary data sets that might be endogenous to more highly-modeled population surfaces. It is also suitable for use in dasymetric and other modeling approaches that reallocate the population counts or densities based on other layers. For more information on the areal-weighting methods see Section II.

To broaden the applicability of GPW, version 4 was expanded to include the additional census variables: age (in five-year groups) and sex. The incorporation of basic demographic characteristics extend the usefulness of GPW into more social applications. For example, the degree of vulnerability to climate change, natural hazards, or infectious diseases is strongly influenced by the spatial distribution of a population's age and sex structure.

VI. Limitations

The GPWv4 data collection has a number of characteristics that may limit its applicability.

For example, variability of the precision of pixel-level estimates is inherent in the use of areal-weighting as the disaggregation method. The precision and accuracy of a given pixel is a direct function of the size of the input areal unit. In countries, or locations within countries, where the input units are quite large, the precision of individual pixels within those units is degraded. There are clear implications for the data user. To produce the best estimates, the data user must be aware of the size of the input areal units and select a study area that is larger than the average size of those units. Therefore, the data is only applicable for local analysis in certain locations and may not be suitable for some rural areas that are comprised of large areal units. To assist users in making determinations on data suitability, GPWv4 includes the mean administrative unit area ancillary data layer. This data layer should be referenced to determine the appropriate scale of analysis in a particular region or study area. See Section IV for more information on the mean administrative unit area ancillary data layer.

Additional limitations arise related to the currency of the input population and boundary data. For most countries, GPWv4 used the 2010 round of censuses, which occurred between the years 2005 and 2014. In cases where a census was delayed or did not occur, older estimates from previous census rounds, or current estimates from alternative sources were used instead. For those impacted countries the UN WPP-adjusted population estimates may be more appropriate for use. Users should also consider how issues such as territorial disputes and areas undergoing conflict could impact the accuracy of the input data and ultimately that of the rasters as well.

Small undercounts in the gridded estimates for some areas may result from the decision to exclude geographically undefined populations (i.e., individuals counted in the census that do not have a geographic location, such as homeless population or those living on boats) from the data collection, rather than distributing them across a country.

Finally, along coastlines where the land area in a given pixel may be very small, artificially high population densities occur next to highly populated areas. These artifacts result from the intersection (i.e., the extent, cell size, and alignment) of the GPWv4 master grid and any given set of geographic boundaries. In the uniform distribution/proportional allocation methodology, pixel level population count estimates

are calculated by first determining the proportion of land area of each pixel belonging to (i.e., the portion inside of) a geographic unit compared to the land area in that unit as a whole. The geographic unit's population density (estimated population in the entire unit divided by the land area in the entire unit) is multiplied by the pixel level land area in order to produce a pixel level count.

Figure 2 shows a single 1 km pixel (red outline) located in Hong Kong. The white outlined units are administrative units which contribute to that pixel. These administrative units are labeled with their population count, land area, and population density. The resulting pixel level population density is very high at ~90,000 persons per sq. km. This high density is a result of the total land area in this pixel being very low at ~0.25 km and the inclusion of administrative units with large populations to produce the pixel level estimate.

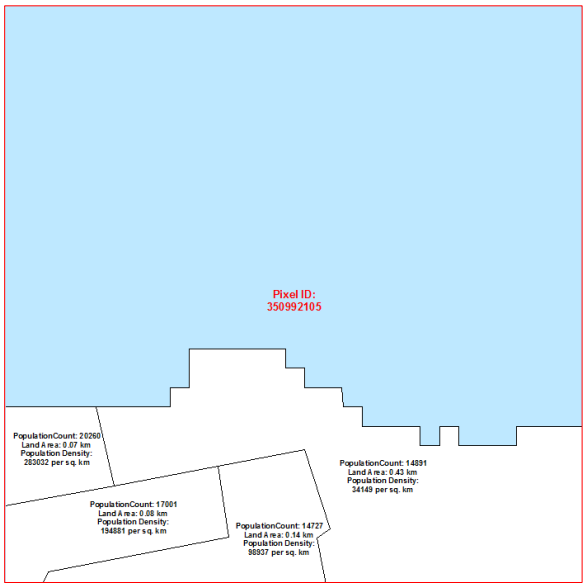

Figure 2: A single 1 km coastal pixel (red outline) located in Hong Kong shows very high population density.

The resulting very high population density can be considered a modifiable areal unit problem. The unit of analysis for GPWv4 is 30 arc-second pixels extracted from a global master grid. The alignment of those pixels with geographic boundaries from the input population and boundary data potentially produces coastal pixels with very small land areas, but with large population count estimates. The pixel level population density at such a location is then estimated to be very high, and according to the data and methodologies it actually is extremely high. In the real world it is unusual to find population densities as high as the data suggest directly on the coast. Additional dasymetric modeling has the potential to correct this issue since the use of covariate layers to inform weighted allocations of population can potentially draw the allocation away from the coastal pixels with very small land areas. That being said, dasymetric modeling approaches also may exhibit modifiable areal unit problems that are even more difficult to explain than the issue presented here.

Due to these known issues and limitations, the confidence of population information is not uniform across countries, but instead is a function of the currency, spatial resolution, and accuracy of the input sources. Users should reference the Microsoft Excel file (.xlsx) which includes country-level information and sources and is packaged with every data download.

VII. Acknowledgments

GPWv4 was produced by CIESIN at Columbia University with primary support from the National Aeronautics and Space Administration under Contract NNG13HQ04C for the Socioeconomic Data and Applications Distributed Active Archive Center (DAAC) for the EOS Data and Information System (EOSDIS).

VIII. Disclaimer

CIESIN follows procedures designed to ensure that data disseminated by CIESIN are of reasonable quality. If, despite these procedures, users encounter apparent errors or misstatements in the data, they should contact SEDAC User Services at https://sedac.uservoice.com/knowledgebase/topics/110829-gpwv4. Neither CIESIN nor NASA verifies or guarantees the accuracy, reliability, or completeness of any data provided. CIESIN provides this data without warranty of any kind whatsoever, either expressed or implied. CIESIN shall not be liable for incidental, consequential, or special damages arising out of the use of any data provided by CIESIN.

IX. Use Constraints

This work is licensed under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0). Users are free to use, copy, distribute, transmit, and adapt the work for commercial and non-commercial purposes, without restriction, as long as clear attribution of the source is provided.

X. Recommended Citations

Data Sets:

Administrative Unit Center Points with Population Estimates, v4.11 (2000, 2005, 2010, 2015, 2020)

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Administrative Unit Center Points with Population Estimates, Revision 11. Palisades, NY: NASA

Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4BC3WMT. Accessed DAY MONTH YEAR.

Basic Demographic Characteristics, v4.11 (2010)

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Basic Demographic Characteristics, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H46M34XX. Accessed DAY MONTH YEAR.

Data Quality Indicators, v4.11 (2010)

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Data Quality Indicators, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H42Z13KG. Accessed DAY MONTH YEAR.

Land and Water Area, v4.11 (2010)

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Land and Water Area, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4Z60M4Z. Accessed DAY MONTH YEAR.

National Identifier Grid, v4.11 (2010)

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): National Identifier Grid, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4TD9VDP. Accessed DAY MONTH YEAR.

Population Count, v4.11 (2000, 2005, 2010, 2015, 2020)

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4JW8BX5. Accessed DAY MONTH YEAR.

UN WPP-Adjusted Population Count, v4.11 (2000, 2005, 2010, 2015, 2020)

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Population Count Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11.

Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4PN93PB. Accessed DAY MONTH YEAR.

Population Density, v4.11 (2000, 2005, 2010, 2015, 2020)

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H49C6VHW. Accessed DAY MONTH YEAR.

UN WPP-Adjusted Population Density, v4.11 (2000, 2005, 2010, 2015, 2020)

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. Gridded Population of the World, Version 4 (GPWv4): Population Density Adjusted to Match 2015 Revision of UN WPP Country Totals, Revision 11. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4F47M65. Accessed DAY MONTH YEAR.

Scientific Publication:

Doxsey-Whitfield, E., K. MacManus, S.B. Adamo, L. Pistolesi, J. Squires, O. Borkovska and S.R. Baptista. 2015. Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. Papers in Applied Geography 1(3): 1-9. https://doi.org/10.1080/23754931.2015.1014272.

XI. Source Code

Please contact SEDAC User Services at ciesin.info@ciesin.columbia.edu for information about the code used to develop the data sets.

XII. References

ABSON, D. J., DOUGILL, A. J. & STRINGER, L. C. 2012. Using Principal Component Analysis for information-rich socio-ecological vulnerability mapping in Southern Africa. *Applied Geography,* 35**,** 515-524. http://www.sciencedirect.com/science/article/pii/S0143622812000859.

BADAL, J., VÁZQUEZ-PRADA, M. & GONZÁLEZ, Á. 2005. Preliminary quantitative assessment of earthquake casualties and damages. *Natural Hazards,* 34**,** 353-374. https://doi.org/10.1007/s11069-004-3656-6.

BALK, D. L., DEICHMANN, U., YETMAN, G., POZZI, F., HAY, S. I. & NELSON, A. 2006. Determining global population distribution: methods, applications and data.

Advances in Parasitology, 62**,** 119-156. https://doi.org/10.1016/S0065-308X(05)62004-0 [Accessed 05/11/2006].

BHADURI, B., BRIGHT, E., COLEMAN, P. & DOBSON, J. 2002. LandScan: Locating people is what matters. *Geoinformatics,* 5**,** 34-37.

CAMINADE, C., KOVATS, S., ROCKLOV, J., TOMPKINS, A. M., MORSE, A. P., COLÓN-GONZÁLEZ, F. J., STENLUND, H., MARTENS, P. & LLOYD, S. J. 2014. Impact of climate change on global malaria distribution. *Proceedings of the National Academy of Sciences,* 111**,** 3286-3291. http://www.pnas.org/content/111/9/3286.abstract.

DEICHMANN, U., BALK, D. & YETMAN, G. 2001. Transforming Population Data for Interdisciplinary Usages: From Census to Grid. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC), CIESIN, Columbia University.

DÖLL, P. 2009. Vulnerability to the impact of climate change on renewable groundwater resources: a global-scale assessment. *Environmental Research Letters,* 4**,** 035006. https://doi.org/10.1088/1748-9326/4/3/035006.

DOXSEY-WHITFIELD, E., MACMANUS, K., ADAMO, S. B., PISTOLESI, L., SQUIRES, J., BORKOVSKA, O. & BAPTISTA, S. R. 2015. Taking Advantage of the Improved Availability of Census Data: A First Look at the Gridded Population of the World, Version 4. *Papers in Applied Geography,* 1**,** 226-234. https://doi.org/10.1080/23754931.2015.1014272.

ERMERT, V., FINK, A. H. & PAETH, H. 2013. The potential effects of climate change on malaria transmission in Africa using bias-corrected regionalised climate projections and a simple malaria seasonality model. *Climatic Change,* 120**,** 741-754. https://doi.org/10.1007/s10584-013-0851-z.

ESRI 2014. ArcGIS Desktop: Release 10.3. Redlands, CA: Environmental Systems Research Institute.

EUROPEAN SPACE AGENCY CLIMATE CHANGE INITIATIVE 2016. Land Cover Project: Water Bodies, version 4.0.

FREIRE, S., MACMANUS, K., PESARESI, M., DOXSEY-WHITFIELD, E. & MILLS, J. 2016. Development of new open and free multi-temporal global population grids at 250 m resolution. 19th AGILE Conference on Geographic Information Science, June 14-17 2016 Helsinki, Finland.

GADM 2011. Global Administrative Areas, version 2.0. GADM.

GDAL 2014. GDAL - Geospatial Data Abstraction Library: Version 1.11.1. Open Source Geospatial Foundation.

GLIMS & NSIDC 2005, updated 2015. Global Land Ice Measurements from Space glacier database. Boulder, CO, USA: Compliled and made available by the international GLIMS community and the National Snow and Ice Data Center (NSDIC).

GUHA-SAPIR, D., RODRIGUEZ-LLANES, J. & JAKUBICKA, T. 2011. Using disaster footprints, population databases and GIS to overcome persistent problems for human impact assessment in flood events. *Natural Hazards,* 58**,** 845-852. https://doi.org/10.1007/s11069-011-9775-y.

HAY, S. I., NOOR, A. M., NELSON, A. & TATEM, A. J. 2005. The accuracy of human population maps for public health application. *Tropical Medicine and International Health,* 10**,** 1073-1086. https://doi.org/10.1111/j.1365-3156.2005.01487.x.

KARTVERKET 2014. N1000 Kartdata. Kartverket (Norwegian Mapping Authority).

LANDMÆLINGAR ÍSLANDS 2012. IS 500V Vatn flákar. Landmælingar Íslands (National Land Survey of Iceland).

LANTMÄTERIET 2016. Hydrografi. Lantmäteriet (Sweden Land Survey).

LEHNER, B. & DÖLL, P. 2004. Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology,* 296(1)**,** 1-22. Data available from: http://www.worldwildlife.org/pages/global-lakes-and-wetlands-database.

MAANMITTAUSLAITOS 2010. General Map, 1 to 1M: Vesi Alue. Maanmittauslaitos (National Land Survey of Finland).

MICROSOFT 2013. Microsoft Excel. Redmond, Washington: Microsoft.

NORWEGIAN POLAR INSTITUTE 2014. Kartdata Svalbard 1:100 000 (S100 Kartdata) Norwegian Polar Institute.

PYTHON SOFTWARE FOUNDATION 2014a. IDLE (Python's Integrated DeveLopment Environment) version 2.7.8.

PYTHON SOFTWARE FOUNDATION 2014b. Python programming language, version 2.7.8. http://www.python.org.

STATISTICS CANADA 2011. Lakes and Rivers.

TATEM, A. J., NOOR, A. M., VON HAGEN, C., DI GREGORIO, A. & HAY, S. I. 2007. High resolution population maps for low income nations: Combining land cover and census in East Africa. *PLoS One,* 2**,** e1298. https://doi.org/10.1371/journal.pone.0001298.

TOBLER, W., DEICHMANN, U., GOTTSEGEN, J. & MALOY, K. 1997. World population in a grid of spherical quadrilaterals. *International Journal of Population Geography,* 3**,** 203-225. http://www.geog.ucsb.edu/~kclarke/G232/tobler_1997.pdf.

UNITED NATIONS 2015. World Population Prospects: The 2015 Revision, DVD Edition. New York, NY: United Nations, Department of Economic and Social Affairs, Population Division.

UNITED STATES CENSUS BUREAU GEOGRAPHY DIVISION 2016. Areal Hydrography. United States Census Bureau, Geography Division.

UNITED STATES GEOLOGICAL SURVEY 2003. Shuttle Radar Topography Mission Water Body Data, version 2.0. U.S. Geological Survey, Earth Resources Observation and Sciences.

XIII. Documentation Copyright and License

Copyright © 2018. The Trustees of Columbia University in the City of New York. This document is licensed under the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0).

XIV. Appendices

Appendix A: Data Revision History

Revision Date	Data Set(s)	Revisions
December 2018	Population Count, v4.11 Population Density, v4.11 UN WPP-Adjusted Population Count, v4.11 UN WPP-Adjusted Population Density, v4.11 Data Quality Indicators, v4.11 Land and Water Area, v4.11 Basic Demographic Characteristics, v4.11	Erroneous no data pixels were recoded as zero and extent was updated to full global extent (90 to -90).
	Population Count, v4.11 Population Density, v4.11 UN WPP-Adjusted Population Count, v4.11 UN WPP-Adjusted Population Density, v4.11 Basic Demographic Characteristics, v4.11	NetCDF files were updated to include the following layers: Mean Administrative Unit Area; Land Area; Water Area; Input Sex Data Level; Input Age Data Level
	National Identifier Grid, v4.11	Raster was reprocessed to remove artifacts from inland water. Two attributes were added to indicate the administrative level of the demographic characteristics input data. Zip files were

		corrected to include the National Identifier Polygons shapefile.
	Data Quality Indicators, v4.11	Two new classes (Total Land Pixels and Ocean Pixels) were added to the Water Mask.
	Administrative Unit Center Points with Population Estimates, v4.11	Administrative level names of the Greece Administrative Unit Center Points were translated to English.
November 2017	Basic Demographic Characteristics, v4.10	New data set
	Population Count, v4.10 **Population Density, v4.10** **UN WPP-Adjusted Population Count, v4.10** **UN WPP-Adjusted Population Density, v4.10** Data Quality Indicators, v4.10 Land and Water Area, v4.10 National Identifier Grid, v4.10 **Basic Demographic Characteristics, v4.10**	Available at four additional resolutions: 2.5 minute, 15 minute, 30 minute, 1 degree. Available in ASCII format. Datasets in bold available as netCDF at all four additional resolutions. Not available at 30 arc-second resolution.
	Data Quality Indicators, v4.10	The global water mask was updated to include more recent glacier data and local water data sources for high latitude areas.
	Population Count, v4.10 Population Density, v4.10 UN WPP-Adjusted Population Count, v4.10 UN WPP-Adjusted Population Density, v4.10 Data Quality Indicators, v4.10 Land and Water Area, v4.10 Administrative Unit Center Points with Population Estimates, v4.10 National Identifier Grid, v4.10 Basic Demographic Characteristics, v4.10	Updated boundary or population data for 65 countries. See the log of changes to the data sets by version included with each data set download and online at: http://sedac.ciesin.columbia.edu/downloads/docs/gpw-v4/gpw-v4-change-log-rev11.txt.

Appendix B: Contributing Authors & Documentation Revision History

Revision Date	Contributors	Revisions

December 2018	Jane Mills, Kytt MacManus, Linda Pistolesi	All sections were reviewed for clarity and completeness.
		
		These sections were updated as follows:
		
		I. Introduction
		• Paragraph describing the changes made in this release was edited and presented in bullets
		
		III. Data Set Descriptions
		• Added Figure 1 "Tiling of 30 arc-second ASCII rasters" and text explaining why ASCIIs are tiled.
		• Updated Tables 4 and 5
		• In the section for the Data Quality Indicators data set:
		  ○ Updated Table 6
		  ○ Updated description of Water Mask
		• In the section for the National Identifier Grid data set:
		  ○ Updated Table 9
		
		X. Recommended Citations
		• All citations and DOIs updated
		
		XIV. Appendices
		• Appendix C: updated section 1b on Water Mask

November 2017	Alyssa Fico, Kytt MacManus, Jane Mills, Linda Pistolesi	All sections were reviewed for clarity and completeness and edited to reflect the updates to the eight previously released data sets (v4.0, July 2016), and the addition of the Basic Demographic Characteristics, v4.10 (2010) data set. The following sections were added: • IV. How to Use the Data • V. Potential Use Cases • VI. Limitations • XI. Source Code The Map Gallery section was removed (previously Section V) These sections were updated as follows: II. Data and Methodology • Added sub-section "Input Data" describing the inputs used to produce the GPWv4 data sets and the sources of the inputs (previously covered in IV. Data Sources). • Added Tables 1 and 2. • Described the steps for estimating population by age and sex in the "Methods" sub-section. • Described the addition of low resolution raster data products in the "Transform to raster" section. • Described QA/QC efforts in the methodology steps, where relevant.

		III. Data Set Descriptions • Updated descriptions to more closely mirror the metadata data set abstracts and reflect the additional resolutions and data formats available. • Added Table 3: Resolution equivalents in seconds, minutes, degrees, and kilometers; Table 4: Rasters included in every netCDF download; and Table 5: Ancillary files included in every netCDF download • Removed "Data set downloads" links. • In the section for the Data Quality Indicators data set: o Removed reference to the "Census quality grids" as this data layer is not yet ready for distribution and reference to it in the previously released documentation was an editorial error. o Updated the table of data context values and category descriptions (now Table 6) with two additional codes/descriptions. • In the section for the Administrative Unit Center Points with Population Estimates data set: o Updated the table of field names and field descriptions (now Table 7) with a new unique ID field (GUBID) and nine additional fields for the age and sex estimates. • In the section for the National Identifier Grid data set: o Added Table 8, which identifies the countries, lost during aggregation to lower resolutions. • Added a section for the Basic Demographic Characteristics, v4.10 (2010) data set. XII. References • Added 18 additional references XIV. Appendices • Added Appendix A: Data Revision History • Added Appendix B: Contributing Authors & Documentation Revision History • Added Appendix C: Methodologies Used to Produce GPWv4 Ancillary Data Sets • Appendix III changed to Appendix D • Appendix I changed to Appendix E • Appendix II changed to Appendix F, and renamed from "Countries with Multiple Levels of Census Data Inputs" to "Administrative Levels of Primary Population, Sex, and Age Data" • Appendix IV changed to Appendix G • Added Appendix H: Highest Age Group Class Available by Country

Appendix C: Methodologies Used to Produce GPWv4 Ancillary Data Sets

1. Data Quality Indicators, v4.11 (2010)

a) Data Context

- During the matching of census and boundary units:
  - Boundary units that matched census units with zero population or missing population were assigned a population context code.
  - Boundary units that failed to match to a census unit were researched and assigned a boundary context code to explain why they failed to match (e.g., parks, water features, areas that were not enumerated due to conflict, etc.).
- The boundary units with context codes were converted to raster to produce the Data Context raster.
- The 30 arc-second data were aggregated to 2.5 arc-minute, 15 arc-minute, 30 arc-minute and 1 degree resolutions. For each lower resolution raster, if a cell in the same resolution Population Count raster contained non-zero population, then the corresponding cell in the Data Context raster was assigned a context code of zero.

Note: This layer is not a complete representation of context features. For example, not all parks are characterized in the Data Context layer; only parks that were identified while researching a failed match, or those explicitly identified in input data, are represented.

b) Water Mask

- For each country, the water features from the input data sources were unioned (see Tables 1 and 2 in Section II).
- The unioned water features were intersected with the administrative boundaries to determine if any administrative units would be lost or significantly altered in terms of their land area by the application of the water mask.
  - Units with changes in area greater than 85% were manually evaluated with reference to Esri imagery basemaps, and water features were adjusted as appropriate.
- The final set of water features was converted to raster.
- Inland pixels that were not classified as Total or Partial Water were classified as Total Land Pixels. All other pixels were classified as Ocean Pixels.

c) Mean Administrative Unit Area

- A 1 kilometer fishnet was intersected with the input administrative boundaries.

- Because any given 1 kilometer pixel may intersect more than one administrative boundary, a simple mean of the areas of the intersected input boundaries was calculated at the pixel level.
- The fishnet data were converted to raster to produce the Mean Administrative Unit Area raster.

2. Land and Water Area, v4.11 (2010)

The Land and Water Area data set includes two rasters that represent surface areas of land and water in square kilometers per pixel. The total area of a surface represented by a given pixel in a uniform grid varies with latitude. To capture this spatial variation, surface areas were calculated using a locally-specified Mollweide Projection (EPSG:54009) on the input administrative vector units. The Land and Water Area rasters represent different portions of the calculated surface area.

- A 1 kilometer fishnet was intersected with the input administrative boundaries and water features.
- The resulting feature class was projected to a locally-specified Mollweide Projection.
- The areas of every polygon feature were calculated
- The areas were then summarized at the pixel level to get the total land area in a pixel and the total water area in a pixel.
- The final fishnet was then converted to two rasters using the pixel level land area and water area totals.

3. Administrative Unit Center Points with Population Estimates, v4.11 (2000, 2005, 2010, 2015, 2020)

- The polygon administrative boundaries were converted to point features as true centroids.
- Administrative and population data were joined to the centroids based on a common code.

4. National Identifier Grid, v4.11 (2010)

- A 1 kilometer fishnet was intersected with the input administrative boundaries.
- Each fishnet pixel was assigned the country code(s) of the intersecting administrative boundaries.
- The area of each polygon created by the intersection was calculated.
- The data were summarized at the pixel level, selecting the country code of the polygon with the largest area (i.e., Max).
- The final fishnet with country codes was converted to raster.

Appendix D: Countries Not Adjusted to the Global Boundary Framework

The following countries were not adjusted to the global boundary framework:

> Argentina, Australia, Austria, Bangladesh, Bosnia and Herzegovina, Brazil, Bulgaria, Canada, China, Croatia, Egypt, Faroe Islands, Fiji, France, French Polynesia, Greece, Greenland, Holy See, Hungary, India, Indonesia, Ireland, Italy, Japan, Kenya, Kosovo, Malawi, Malaysia, Marshall Islands, Mexico, Montenegro, Mozambique, Myanmar, New Zealand, Nigeria, Palau, Philippines, Portugal, Puerto Rico, Romania, Saint Helena, San Marino, Saudi Arabia, Serbia, Slovenia, Solomon Islands, South Africa, South Sudan, Sri Lanka, Sudan, Switzerland, Taiwan, The former Yugoslav Republic of Macedonia, Turkey, United Kingdom of Great Britain and Northern Ireland, United States of America, and Venezuela.

Appendix E: Currency and Type of Population Data used in GPWv4

Table E1 shows the year and type of the primary population data source used for the Revision 11 data sets for each country in GPWv4. For some countries, sex or age data were not included in the primary source and an alternate source was used. Tables E2 and E3 include the year and type of the alternate source of sex or age data for these countries.

Table E1: Currency and Type of Primary Population Data Source

Country or Territory Name	Data Year	Type of Data
Afghanistan	2011	Population Estimate/Projection
Åland Islands	2010	Population Register
Albania	2011	Final Census
Algeria	2008	Final Census
American Samoa	2010	Final Census
Andorra	2010	Population Register
Angola	2014	Final Census
Anguilla	2011	Preliminary/Provisional Census
Antigua and Barbuda	2011	Final Census
Argentina	2010	Final Census
Armenia	2011	Final Census
Aruba	2010	Final Census
Australia	2011	Final Census
Austria	2012	Population Register
Azerbaijan	2009	Final Census
Bahamas	2010	Final Census
Bahrain	2010	Final Census
Bangladesh	2011	Final Census
Barbados	2010	Final Census
Belarus	2009	Final Census
Belgium	2014	Population Register
Belize	2010	Final Census
Benin	2013	Final Census
Bermuda	2010	Final Census
Bhutan	2005	Final Census
Bolivia (Plurinational State of)	2012	Final Census
Bonaire Saint Eustatius and Saba	2010	Population Register
Bosnia and Herzegovina	2013	Final Census
Botswana	2011	Final Census
Brazil	2010	Final Census
British Virgin Islands	2010	Final Census
Brunei Darussalam	2011	Final Census
Bulgaria	2011	Final Census
Burkina Faso	2006	Final Census
Burundi	2008	Final Census
Cambodia	2008	Final Census

Cameroon	2005	Final Census
Canada	2011	Final Census
Cape Verde	2010	Final Census
Cayman Islands	2010	Final Census
Central African Republic	2012	Population Estimate/Projection
Chad	2009	Final Census
Chile	2010	Population Estimate/Projection
China	2010	Final Census
China Hong Kong Special Administrative Region	2011	Final Census
China Macao Special Administrative Region	2011	Final Census
Colombia	2005	Final Census
Comoros	2013	Population Estimate/Projection
Congo	2007	Final Census
Cook Islands	2011	Final Census
Costa Rica	2011	Final Census
Côte d'Ivoire	2014	Final Census
Croatia	2011	Final Census
Cuba	2012	Final Census
Curaçao	2011	Final Census
Cyprus	2011	Final Census
Czech Republic	2011	Final Census
Democratic People's Republic of Korea	2008	Final Census
Democratic Republic of the Congo	2008	Population Estimate/Projection
Denmark	2010	Population Register
Djibouti	2009	Final Census
Dominica	2011	Preliminary/Provisional Census
Dominican Republic	2010	Final Census
Ecuador	2010	Final Census
Egypt	2006	Final Census
El Salvador	2007	Final Census
Equatorial Guinea	2015	Preliminary/Provisional Census
Eritrea	2012	Population Estimate/Projection
Estonia	2011	Final Census
Ethiopia	2007	Final Census
Faroe Islands	2010	Population Register
Falkland Islands (Malvinas)	2012	Final Census
Fiji	2007	Final Census
Finland	2011	Population Register
France	2009	Final Census
French Guiana	2009	Population Register
French Polynesia	2007	Final Census
Gabon	2003	Final Census
Gambia	2013	Preliminary/Provisional Census
Georgia	2014	Final Census
Germany	2011	Population Register

Ghana	2010	Final Census
Gibraltar	2012	Final Census
Greece	2011	Final Census
Greenland	2010	Population Register
Grenada	2011	Preliminary/Provisional Census
Guadeloupe	2009	Population Register
Guam	2010	Final Census
Guatemala	2010	Population Estimate/Projection
Guernsey	2010, 2013, 2014	Final Census, Population Estimate/Projection, Population Register
Guinea	2014	Final Census
Guinea-Bissau	2009	Final Census
Guyana	2002	Final Census
Haiti	2003	Final Census
Holy See	2013	Population Estimate/Projection
Honduras	2013	Final Census
Hungary	2011	Final Census
Iceland	2010	Population Register
India	2011	Final Census
Indonesia	2010	Final Census
Iran (Islamic Republic of)	2011	Final Census
Iraq	2003, 2011	Population Estimate/Projection
Ireland	2011	Final Census
Isle of Man	2011	Final Census
Israel	2008	Final Census
Italy	2011	Final Census
Jamaica	2011	Final Census
Japan	2010	Final Census
Jersey	2011	Final Census
Jordan	2004	Final Census
Kazakhstan	2009	Final Census
Kenya	2009	Final Census
Kiribati	2010	Final Census
Kosovo	2011	Final Census
Kuwait	2005	Final Census
Kyrgyzstan	2009	Final Census
Lao People’s Democratic Republic	2005	Final Census
Latvia	2011	Final Census
Lebanon	2002	Population Estimate/Projection
Lesotho	2006	Final Census
Liberia	2008	Final Census
Libya	2006	Final Census
Liechtenstein	2010	Final Census
Lithuania	2011	Final Census
Luxembourg	2011	Final Census
Madagascar	2010	Population Estimate/Projection
Malawi	2008	Final Census

Malaysia	2010	Final Census
Maldives	2006	Final Census
Mali	2009	Final Census
Malta	2011	Final Census
Marshall Islands	2011	Final Census
Martinique	2009	Population Register
Mauritania	2013	Final Census
Mauritius	2011	Final Census
Mayotte	2012	Final Census
Mexico	2010	Final Census
Micronesia (Federated States of)	2010	Final Census
Monaco	2008	Population Estimate/Projection
Mongolia	2010	Final Census
Montenegro	2011	Final Census
Montserrat	2011	Final Census
Morocco	2014	Final Census
Mozambique	2007	Final Census
Myanmar	2014	Final Census
Namibia	2011	Final Census
Nauru	2011	Final Census
Nepal	2011	Final Census
Netherlands	2011	Population Register
New Caledonia	2009	Final Census
New Zealand	2013	Final Census
Nicaragua	2005	Final Census
Niger	2012	Final Census
Nigeria	2006	Final Census
Niue	2011	Final Census
Norfolk Island	2011	Final Census
Northern Mariana Islands	2010	Final Census
Norway	2011	Population Register
Oman	2010	Final Census
Pakistan	2010	Population Estimate/Projection
Palau	2005	Final Census
Panama	2010	Final Census
Papua New Guinea	2011	Final Census
Paraguay	2010	Population Estimate/Projection
Peru	2007	Final Census
Philippines	2010	Final Census
Pitcairn	2008	Final Census
Poland	2011	Final Census
Portugal	2011	Final Census
Puerto Rico	2010	Final Census
Qatar	2010	Final Census
Republic of Korea	2010	Final Census
Republic of Moldova	2004	Final Census
Réunion	2009	Population Register

Romania	2011	Final Census
Russian Federation	2010	Final Census
Rwanda	2012	Final Census
Saint Helena	2008, 2014	Final Census and Population Register
Saint Kitts and Nevis	2001	Final Census
Saint Lucia	2010	Final Census
Saint Pierre and Miquelon	2009	Final Census
Saint Vincent and the Grenadines	2012	Final Census
Saint-Barthelemy	2009	Population Register
Saint-Martin (French part)	2009	Population Register
San Marino	2010	Preliminary/Provisional Census
Sao Tome and Principe	2012	Final Census
Saudi Arabia	2010	Final Census
Senegal	2013	Final Census
Serbia	2011	Final Census
Seychelles	2010	Final Census
Sierra Leone	2004	Final Census
Singapore	2010	Final Census
Sint Maarten (Dutch part)	2011	Final Census
Slovakia	2011	Final Census
Slovenia	2010	Final Census
Solomon Islands	2009	Final Census
Somalia	2005	Population Estimate/Projection
South Africa	2011	Final Census
South Sudan	2008	Final Census
Spain	2011	Final Census
Sri Lanka	2012	Final Census
State of Palestine	2007	Preliminary/Provisional Census
Sudan	2008	Final Census
Suriname	2012	Final Census
Svalbard and Jan Mayen Islands	2010	Population Register
Swaziland	2007	Final Census
Sweden	2010	Population Register
Switzerland	2010	Final Census
Syrian Arab Republic	2004	Final Census
Taiwan	2010	Final Census
Tajikistan	2010	Final Census
Thailand	2010	Preliminary/Provisional Census
The former Yugoslav Republic of Macedonia	2010	Population Estimate/Projection
Timor-Leste	2010	Final Census
Togo	2010	Final Census
Tokelau	2011	Final Census
Tonga	2011	Final Census
Trinidad and Tobago	2011	Final Census
Tunisia	2014	Final Census
Turkey	2010	Population Register

Turkmenistan	2005	Population Estimate/Projection
Turks and Caicos Islands	2012	Final Census
Tuvalu	2012	Final Census
Uganda	2014	Preliminary/Provisional Census
Ukraine	2013	Population Estimate/Projection
United Arab Emirates	2005	Final Census
United Kingdom of Great Britain and Northern Ireland	2011	Final Census
United Republic of Tanzania	2012	Final Census
United States of America	2010	Final Census
United States Virgin Islands	2010	Final Census
Uruguay	2011	Final Census
Uzbekistan	2013	Population Register
Vanuatu	2009	Final Census
Venezuela (Bolivarian Republic of)	2011	Final Census
Viet Nam	2009	Final Census
Wallis and Futuna Islands	2008	Final Census
Western Sahara	2014	Final Census
Western Samoa	2011	Final Census
Yemen	2004	Final Census
Zambia	2010	Final Census
Zimbabwe	2012	Final Census

Table E2: Currency and Type of Sex Data Source

Country or Territory Name	Data Year	Type of Data
Angola	2014	Final Census
Central African Republic	2010	Population Estimate/Projection
Democratic Republic of the Congo	2010	Population Estimate/Projection
Eritrea	2010	Population Estimate/Projection
Gabon	2010	Population Estimate/Projection
Grenada	2010	Population Estimate/Projection
Guernsey	2013, 2014	Final Census and Population Register
Guyana	2012	Final Census
Iraq	2010	Population Estimate/Projection
Lebanon	2010	Population Estimate/Projection
Marshall Islands	2010	Population Estimate/Projection
Somalia	2010	Population Estimate/Projection
Turkmenistan	2010	Population Estimate/Projection

Table E3: Currency and Type of Age Data Source

Country or Territory Name	Data Year	Type of Data
Afghanistan	2010	Population Estimate/Projection
Angola	2014	Final Census
Central African Republic	2010	Population Estimate/Projection
Côte d'Ivoire	2010	Population Estimate/Projection
Democratic Republic of the Congo	2010	Population Estimate/Projection
Dominica	2006	Population Estimate/Projection
Equatorial Guinea	2010	Population Estimate/Projection
Eritrea	2010	Population Estimate/Projection
Gabon	2010	Population Estimate/Projection
Gambia	2010	Population Estimate/Projection
Grenada	2010	Population Estimate/Projection
Guernsey	2013, 2014	Final Census, Population Register
Guyana	2012	Final Census
Iraq	2010	Population Estimate/Projection
Lebanon	2010	Population Estimate/Projection
Marshall Islands	2010	Population Estimate/Projection
Oman	2014	Population Estimate/Projection
San Marino	2012	Official Statistics
Somalia	2010	Population Estimate/Projection
Syrian Arab Republic	2010	Population Estimate/Projection

Appendix F: Administrative Levels of Primary Population, Sex, and Age Data

The table below shows the administrative level(s) of the primary population, sex, and age data sources used for each country in GPWv4. It shows which countries were processed using population, sex, and age data inputs at multiple administrative levels, using sex data inputs at a lower administrative level than the population data inputs, and using age data inputs at a lower administrative level than the population data inputs.

Country or Territory Name	Administrative Level of Data Source		
	Primary Population	Sex	Age
Afghanistan	2	2	0
Åland Islands	1	1	1
Albania	3	1	1
Algeria	2	2	2
American Samoa	3	3	3
Andorra	1	1	1
Angola	2	2	0
Anguilla	0	0	0
Antigua and Barbuda	1	1	1
Argentina	2	2	2
Armenia	2, 3	2, 3	1, 2
Aruba	2	2	2
Australia	5	5	5
Austria	4	4	4
Azerbaijan	1	1	1
Bahamas	1	1	1
Bahrain	1	1	0
Bangladesh	3	3	3
Barbados	1	1	1
Belarus	2	2	1
Belgium	4	4	4
Belize	1	1	1
Benin	2	2	2
Bermuda	1	1	1
Bhutan	2	2	2
Bolivia (Plurinational State of)	3	3	3
Bonaire Saint Eustatius and Saba	1	1	1
Bosnia and Herzegovina	3	3	3
Botswana	2	2	0
Brazil	5	5	5
British Virgin Islands	1	1	1
Brunei Darussalam	2	2	2
Bulgaria	2	2	2
Burkina Faso	3	3	1
Burundi	2	2	2

Cambodia	3	3	3
Cameroon	2	2	2
Canada	5	3	3
Cape Verde	2	2	2
Cayman Islands	1	1	1
Central African Republic	3	0	0
Chad	2	2	1
Chile	3	3	3
China	4	4	3
China Hong Kong Special Administrative Region	3	3	3
China Macao Special Administrative Region	3	3	3
Colombia	2	1	1
Comoros	1	1	1
Congo	1	1	1
Cook Islands	3	3	3
Costa Rica	3	3	3
Côte d'Ivoire	4	4	0
Croatia	2	2	2
Cuba	2	2	1
Curaçao	1	1	0
Cyprus	2	2	2
Czech Republic	3	3	3
Democratic People's Republic of Korea	2	2	1
Democratic Republic of the Congo	3	0	0
Denmark	3	3	3
Djibouti	1	0	0
Dominica	1	1	0
Dominican Republic	4	4	4
Ecuador	3	3	3
Egypt	2	2	2
El Salvador	2	2	2
Equatorial Guinea	2	2	0
Eritrea	1	0	0
Estonia	3	3	3
Ethiopia	3	3	3
Faroe Islands	3	3	3
Falkland Islands (Malvinas)	1	1	1
Fiji	3	3	3
Finland	2	2	2
France	4, 5	4, 5	4, 5
French Guiana	2	2	2
French Polynesia	3	3	3
Gabon	2	0	0
Gambia	2	2	0
Georgia	2	2	0
Germany	3	3	2

Ghana	2	2	1
Gibraltar	0	0	0
Greece	5	3	3
Greenland	3	3	3
Grenada	1	0	0
Guadeloupe	2	2	2
Guam	3	3	3
Guatemala	2	2	0
Guernsey	1	1	1
Guinea	3	3	0
Guinea-Bissau	2	2	1
Guyana	2	1	1
Haiti	4	4	0
Holy See[2]	0	N/A	N/A
Honduras	3	3	3
Hungary	4	4	4
Iceland	2	2	2
India	3	3	2
Indonesia	3	3	3
Iran (Islamic Republic of)	2	1	0
Iraq	2	0	0
Ireland	4	4	4
Isle of Man	1	1	1
Israel	3	3	3
Italy	4	4	4
Jamaica	1	1	1
Japan	2	2	2
Jersey	1	1	1
Jordan	3	3	1
Kazakhstan	2	1, 2	1, 2
Kenya	5	5	2
Kiribati	3	3	0
Kosovo	1	1	1
Kuwait	1	1	1
Kyrgyzstan	2	2	2
Lao People's Democratic Republic	3	3	3
Latvia	2	2	2
Lebanon	3	0	0
Lesotho	2	2	2
Liberia	2	2	1
Libya	1	0	0
Liechtenstein	1	1	1
Lithuania	2	2	2
Luxembourg	3, 4	3	3
Madagascar	4	0	0

[2] Age and Sex Data is not available for Holy See

Malawi	3	3	3
Malaysia	3	3	3
Maldives	2	2	2
Mali	3, 4	3, 4	1
Malta	2	2	2
Marshall Islands	1	0	0
Martinique	2	2	2
Mauritania	3	3	0
Mauritius	3	3	3
Mayotte	1	1	1
Mexico	3	3	3
Micronesia (Federated States of)	3	3	2
Monaco	2	2	0
Mongolia	2	2	1
Montenegro	1	1	1
Montserrat	2	2	2
Morocco	3	3	3
Mozambique	3	3	3
Myanmar	3	3	2
Namibia	3	2	2
Nauru	1	0	0
Nepal	4	4	4
Netherlands	2	2	2
New Caledonia	3	3	3
New Zealand	3	3	1
Nicaragua	2	2	2
Niger	2	2	2
Nigeria	2	2	2
Niue	1	1	0
Norfolk Island	1	1	1
Northern Mariana Islands	2	2	2
Norway	2	2	2
Oman	2	2	2
Pakistan	3	3	3
Palau	1	1	1
Panama	3	3	3
Papua New Guinea	2	2	0
Paraguay	2	2	2
Peru	3	3	3
Philippines	3	3	3
Pitcairn	0	0	0
Poland	4	4	4
Portugal	6	6	5
Puerto Rico	4	4	4
Qatar	2	2	0
Republic of Korea	2	2	2
Republic of Moldova	1	1	1

Réunion	2	2	2
Romania	4	4	4
Russian Federation	2, 3	1	1
Rwanda	3	3	3
Saint Helena	2	2	2
Saint Kitts and Nevis	2	2	2
Saint Lucia	1	1	1
Saint Pierre and Miquelon	1	1	1
Saint Vincent and the Grenadines	1	1	1
Saint-Barthelemy	0	0	0
Saint-Martin (French part)	0	0	0
San Marino	1	1	0
Sao Tome and Principe	2	2	2
Saudi Arabia	2	2	1
Senegal	2	2	0
Serbia	5	5	5
Seychelles	2	2	2
Sierra Leone	3	3	1
Singapore	2	2	2
Sint Maarten (Dutch part)	1	1	1
Slovakia	3	3	3
Slovenia	3	3	3
Solomon Islands	4	4	4
Somalia	3	0	0
South Africa	6	6	6
South Sudan	2	2	2
Spain	3	3	3
Sri Lanka	4	4	4
State of Palestine	2	2	2
Sudan	2	2	0
Suriname	2	2	2
Svalbard and Jan Mayen Islands	0	0	0
Swaziland	2	2	1
Sweden	3	3	2
Switzerland	3	3	3
Syrian Arab Republic	3	3	0
Taiwan	2	2	1
Tajikistan	2	2	1
Thailand	2	1	1
The former Yugoslav Republic of Macedonia	2	2	2
Timor-Leste	3	3	3
Togo	2	2	2
Tokelau	1	0	0
Tonga	1	1	1
Trinidad and Tobago	1	1	1
Tunisia	2	2	1
Turkey	2	2	2

Turkmenistan	1	0	0
Turks and Caicos Islands	2	2	0
Tuvalu	3	3	3
Uganda	3	3	0
Ukraine	2	2	1
United Arab Emirates	1	1	0
United Kingdom of Great Britain and Northern Ireland	3, 4, 6	3, 4, 6	3, 4, 6
United Republic of Tanzania	3	3	2
United States of America	5	5	5
United States Virgin Islands	2	2	2
Uruguay	3	3	3
Uzbekistan	1	1	0
Vanuatu	3	3	3
Venezuela (Bolivarian Republic of)	2	2	2
Viet Nam	3	3	3
Wallis and Futuna Islands	2	2	2
Western Sahara	3	3	3
Western Samoa	2	2	2
Yemen	2	2	2
Zambia	3	3	2
Zimbabwe	2	2	2

Appendix G: Countries for Which Growth Rates Were Calculated at Multiple Administrative Levels

The table below shows countries for which population estimates were produced using growth rates calculated at two or more administrative levels.

Country or Territory Name	Growth Rate Levels
Burundi	1, 2
Colombia	1, 2
Cook Islands	1, 2, 3
Curaçao	0, 1
Cyprus	1, 2
Dominican Republic	2, 3, 4
Faroe Islands	2, 3
France	4, 5
Iran (Islamic Republic of)	1, 2
Luxembourg	3, 4
Malawi	1, 2
Malaysia	1, 2
Maldives	1, 2
Mauritania	2, 3
Mauritius	1, 2
Morocco	1, 2
Mozambique	2, 3
Oman	1, 2
Philippines	1, 2
Poland	3, 4
Republic of Korea	1, 2
Singapore	1, 2
State of Palestine	1, 2
Turkey	1, 2
United Kingdom of Great Britain and Northern Ireland	2, 3
United Republic of Tanzania	1, 2
Western Sahara	1, 2
Zambia	0, 3

Appendix H: Highest Age Group Class Available by Country

Highest Age Group	Countries
65 plus	China Macao Special Administrative Region, Djibouti, Guinea, Kuwait, Maldives, Mauritius, Seychelles, Sierra Leone, Singapore, South Sudan
70 plus	Georgia, Montserrat, Niue, Poland, Svalbard and Jan Mayen Islands
75 plus	Anguilla, Bhutan, Cyprus, Egypt, Fiji, Italy, Lao People's Democratic Republic, Malaysia, Marshall Islands, Morocco, Nauru, Nepal, Qatar, Republic of Moldova, Timor-Leste, Tokelau, Western Sahara, Western Samoa, Zimbabwe
80 plus	Afghanistan, Bangladesh, Burundi, Central African Republic, Chad, Chile, Colombia, Comoros, Cook Islands, Côte d'Ivoire, Democratic People's Republic of Korea, Democratic Republic of the Congo, Equatorial Guinea, Eritrea, Falkland Islands (Malvinas), Gabon, Gambia, Grenada, Guatemala, Guernsey, Guinea-Bissau, Iraq, Jersey, Jordan, Kenya, Kosovo, Lebanon, Lithuania, Madagascar, Mali, Mauritania, Oman, Pakistan, Paraguay, Philippines, Pitcairn, Saint Helena, Saint Vincent and the Grenadines, Saudi Arabia, Senegal, Somalia, Sri Lanka, Syrian Arab Republic, Taiwan, Togo, Trinidad and Tobago, Tunisia, Turkmenistan, United Arab Emirates, United Republic of Tanzania, Uzbekistan
85 plus	Algeria, American Samoa, Armenia, Bahrain, Barbados, Belize, Bermuda, Bonaire Saint Eustatius and Saba, Bosnia and Herzegovina, Botswana, Brunei Darussalam, Bulgaria, Canada, China, China Hong Kong Special Administrative Region, Cuba, Czech Republic, Dominica, Greece, Guam, Guyana, Haiti, Hungary, Ireland, Isle of Man, Israel, Lesotho, Liberia, Libya, Mexico, Micronesia (Federated States of), New Zealand, Niger, Nigeria, Norfolk Island, Northern Mariana Islands, Palau, Puerto Rico, Republic of Korea, Romania, Russian Federation, Rwanda, Saint Kitts and Nevis, Serbia, Slovenia, Solomon Islands, State of Palestine, Suriname, The former Yugoslav Republic of Macedonia, Turks and Caicos Islands, Tuvalu, United States of America, United States Virgin Islands, Viet Nam, Yemen, Zambia
Over 85 plus	Åland Islands, Albania, Andorra, Angola, Antigua and Barbuda, Argentina, Aruba, Australia, Austria, Azerbaijan, Bahamas, Belarus, Belgium, Benin, Bolivia (Plurinational State of), Brazil, British Virgin Islands, Burkina Faso, Cambodia, Cameroon, Cape Verde, Cayman Islands, Congo, Costa Rica, Croatia, Curaçao, Denmark, Dominican Republic, Ecuador, El Salvador, Estonia, Ethiopia, Faroe Islands, Finland, France, French Guiana, French Polynesia, Germany, Ghana, Gibraltar, Greenland, Guadeloupe, Honduras, Iceland, India, Indonesia, Iran (Islamic Republic of), Jamaica, Japan, Kazakhstan, Kiribati, Kyrgyzstan, Latvia, Liechtenstein, Luxembourg, Malawi, Malta, Martinique, Mayotte, Monaco, Mongolia, Montenegro, Mozambique, Myanmar, Namibia, Netherlands, New Caledonia, Nicaragua, Norway, Panama, Papua New Guinea, Peru, Portugal, Réunion, Saint Lucia, Saint Pierre and Miquelon, Saint-Barthelemy, Saint-Martin (French part), San Marino, Sao Tome and Principe, Sint Maarten (Dutch part), Slovakia, South Africa, Spain, Sudan, Swaziland, Sweden, Switzerland, Tajikistan, Thailand, Tonga, Turkey, Uganda, Ukraine, United Kingdom of Great Britain and Northern Ireland, Uruguay, Vanuatu, Venezuela (Bolivarian Republic of), Wallis and Futuna Islands