

Population Enumerations Database

Version 2.0, 16 November 2019

Dominik Paprotny, GFZ German Research Centre for Geosciences

Changes compared to version 1.0:

- Data updated and revised using information available up to 1 November 2019.
- A discussion paper was added.
- Some typos in the datasets and documentation were corrected.

1. Introduction

The Population Enumerations Database lists all modern population censuses since 1703 for all territories (independent or not) in a standardized and readily-reusable format, indicating the population count, census type, date, territorial coverage and other relevant information on the enumeration procedure or accuracy. It is supplemented by additional datasets with lists of polities and population estimates, and Matlab code, to support the analysis of the data.

2. Contents

The database includes the following files:

- PED.xls – the Population Enumerations Database, with 16 variables as described in section 3, and an additional sheet containing a list of sources.
- PED.csv – as above, but in semicolon-delimited CSV format and without the list of sources.
- PED.txt – as above, but in tab-delimited text file format.
- MATLAB.zip – a ZIP archive containing files for use in Matlab software:
 - Census_analyzer.m – a code for analysing the database;
 - Census_data.mat – the PED as a Matlab data file;
 - Countries.m – a code generating a list of all polities for a given year together with the population from the latest census;
 - Countries_1800_2019.mat – a data file with a list of polities;
 - Underenumeration.mat – a data file for uncertainty analysis of population censuses (code not provided).
- Supplement.zip – a ZIP archive containing files supporting the analysis of census data:
 - Polities_crossreference.xls – a cross-reference file between different databases of polities, allowing quicker integration between polity coding systems.
 - World_summary.xls – contains world and continental population estimates and total enumerated population.
 - World_estimates.xls – contains historical estimates of world population from variety of sources.
 - Census_errors.xls – contains a selection of censuses with estimates of under-/overenumeration and related GDP per capita estimates for uncertainty analysis.

3. Variables of the Population Enumerations Database

Variable	Description	Mandatory field
Country_code	Numeric code used for uniquely distinguishing all polities in the database	yes
Country_RSS_code	Numeric code based on Russett et al. (1968), updated with Correlates of War project files "State System Membership (v2016)" and "Territorial Change (v5)"	no

Country_ISO_code	Three-letter code from the International Organization for Standardization's (2017) standard ISO 3166, including discontinued codes for historical polities	no
Country_name_EN	Name of the polity. A single, most recent name is used throughout, hence it is not necessarily valid at each census year	yes
Census_year	Year of the census	yes
Census_month	Month of the census. If the month is unknown, "0" value is assigned	yes
Census_day	Day of the census. If the day is unknown, "0" value is assigned	yes
Census_population	Population enumerated during the census. The figure does not include adjustment for underenumeration unless only an adjusted figure was available (such cases are indicated in the 'Census_note' field)	no
Census_area	Area (in km ²) of the census. In a few cases (indicated in the 'Census_note' field) this value is not equal to the total area of the polity. This figure is generally based on the 2016, 1990, 1970 and 1955 UN yearbooks, and national sources, and often changes due to remeasurements, rather than actual change in size	no
Census_type	Type of the census: <ul style="list-style-type: none"> DJ – de jure. DF – de facto. RDJ – de jure, register-based. RDF – de facto, register-based. 	no
Census_territory	Information on the territorial coverage of the census in general, excluding special cases for undercoverage of certain parts of the polity (which is indicated in the 'Census_note' field). Recurring explanations include: <ul style="list-style-type: none"> Current – the territory during the census was the same as the current (as of 2019) territory of the polity. Territory of the census – refers to the territory within borders of the polity as they were during the census year. It is used in cases where the exact description of the territory is not available or not feasible, or the territory was changing very frequently. Pre-WW1 territory – refers to the territory as it was before the start of World War I in 1914. Interwar territory – refers to the territory as it was as the result of redrawing of borders after the end of World War I in 1918, but before the start of World War II in 1939. Post-WW2 – refers to the territory as it was as the result of redrawing of borders after the end of World War II in 1945. 	yes
Census_note	Other relevant information on the census, particularly on: <ul style="list-style-type: none"> the dates of field enumeration, if it was particularly long, thus affecting the accuracy of the population total; estimates of census underenumeration or overenumeration; which major groups of population, or parts of the polity, were excluded from the census; whether population of other polities was included in the population figure; relevant major changes in the territory since the last census not described in 'Census_territory'; major deviations from the de facto/de jure concepts; different census dates for parts of the polity; whether the population figure is provisional, rounded, adjusted or based on incomplete processing of census returns; whether the census was only for population (no housing enumeration) or a housing/agricultural census with a population count; whether the census used other sources of information apart from field enumeration (except censuses indicated as register-based); methodological differences affecting the census area figure; population of certain enumerated population groups or parts of 	no

	the polity, if such information is helpful for interpreting the time series.	
Census_valid_start	First year of validity of the census population figure for calculating total enumerated population in the world. Value of “0” is assigned in the particular census is not used	yes
Census_valid_end	Last year of validity of the census population figure for calculating total enumerated population in the world. Value of “0” is assigned in the particular census is not used	yes
Census_count	Dummy variable for the purpose of counting the total number of censuses taken, with value of “1” identifying entries that should be counted (value of “0” is assigned otherwise)	yes
Source	List of sources of data used to generate the entry (see PED.xls file)	yes

4. Methods and concepts

The of census enumerations was constructed, on country-by-country basis. Various editions of the United Nations’ *Demographic Yearbook* were a fundamental source together with other UN datasets and publications (United Nations 2019). Additionally, a large number of country-specific sources was consulted including statistical yearbooks and websites, as well as publications by the US Census Bureau. Counting multiple editions of the same national statistical yearbook as a single source, the total number of sources included in the database is 73. In some cases, national and UN sources (and UN publications between themselves) disagreed on the census date, type or population figure. The national source was generally preferred, as were more recent UN publications over older ones. The database used information available up to 1 November 2019.

Inclusion of censuses in the database was subject to certain limitations. In principle all should satisfy the principles of modern censuses as laid out by United Nations (1955): universality (all members of community are covered), simultaneity (all facts refer to one point in time), individual units (of recording data), defined territory (of the census operation) and compilation (including publication of results). In practice, whether a given satisfies all conditions might be difficult to assess. As a consequence, some censuses do not have precise dates, which might imply that collection of data was not simultaneous; for others the spatial extent is not precisely known. They were nonetheless included if they were covered by UN yearbooks. The rule of universality was more strictly was enforced, therefore the database excludes:

- Colonial censuses where the non-European population was not enumerated, or only summary statistics were collected on the non-European population. It is unclear, however, in case of some censuses whether the non-European population was actually enumerated in the field; those censuses were generally included in the database. Also, all censuses of Australia and New Zealand are included despite omitting the indigenous population from most enumerations.
- Sample censuses, except situations where the census was carried out through a combination of sample field surveys and use of administrative registers. The French “rolling census”, which is in effect a large annual sample survey, is also included.

The censuses were assumed to be population and housing censuses, unless it was known that they were only population censuses. Housing or agricultural censuses were excluded from the database, except in situations where such census enumerated the population as well.

5. Disclaimer

The political division of the world into countries and their dependencies, and the territorial composition thereof, used in this database, both present and historical, was prepared purely

for purposes of statistical analysis of census data, and does not represent an official position of any person or organization. Also, the data is provided only for research purposes

References

- Russett, B. M., Singer, J. D., and Small, M. (1968). National Political Units in the Twentieth Century: A Standardized List. *The American Political Science Review*, 62(3), 932-951.
<http://www.jstor.org/stable/1953441>
- Correlates of War Project (2014). Territorial Change (v5). Available at:
<http://correlatesofwar.org>
- Correlates of War Project (2016). State System Membership List, v2016. Available at:
<http://correlatesofwar.org>
- International Organization for Standardization (2017). ISO 3166 "Codes for the representation of names of countries and their subdivisions". Available at
<https://www.iso.org/obp/ui/#search>
- United Nations (1955). *Demographic Yearbook 1955*. New York: Statistical Office of the United Nations.
- United Nations (2019). Demographic Yearbook System. Available at:
<https://unstats.un.org/unsd/demographic-social/products/dyb/#overview>.