



DATA DICTIONARY - UN DEMOGRAPHICS BEFORE/AFTER COVID-19 ANALYSIS PROJECT

R G Nattrass



AUGUST 1, 2025

RGN

Table of Contents

Data Source	2
Data Information.....	2
Column Descriptions	3
Ethical Considerations.....	4
Data Relevance	5
Data Profile	5

Data Source

UN- DEMOGRAPHICAL DATA

<https://population.un.org/wpp/downloads?folder=Standard%20Projections&group=CSV%20format>

Data Citation - “United Nations, Department of Economic and Social Affairs, Population Division(2024). World Population Prospects 2024, Online Edition.”

Data Information

The data sourced above is open access data from the UN made available under a Creative Commons license CC BY 3.0 IGO: <http://creativecommons.org/licenses/by/3.0/igo/> .

- This data is survey data from census records from countries around the world. Its collection run up until 2023/24, after which it continues and supplies probability estimates for the future. Descriptions of columns are below:

SortOrder (numeric): record counter

LocID (numeric): numeric code for the location; for countries and areas, it follows the ISO 3166-1 numeric standard

Notes (string): symbol linked to location notes file (available for download below)

ISO3_code (string): ISO 3166-1 alpha-3 location codes

ISO2_code (string): ISO 3166-1 alpha-2 location codes

SDMX_code (string): SDMX Global Registry, cross domain code list for geographical areas (version 2.0) represents a combination of reference area codes in M49 and ISO-3166 classification for international data exchange and interoperability https://sdmx.org/?page_id=3215 and https://registry.sdmx.org/ws/public/sdmxapi/rest/codelist/SDMX/CL_AREA/2.0

LocTypeID (numeric): code for location type

LocTypeName (string): type of location

ParentID (numeric): numeric code of the parent location

Location (string): name of the region, subregion, country or area

VarID (numeric): numeric code for the scenario

Variant (string): projection scenario name (Medium is the most used); for more information see Definition of Projection Scenarios

Time (numeric): year the data refers to

MidPeriod (numeric): numeric value identifying the mid period of the data, with the decimal representing the month (e.g. 1950.5 for July of 1950)

For files with data separated by age, the following columns are available:

AgeGrp (string): label identifying the single age (e.g. 15) or age group (e.g. 15-19)

AgeGrpStart (numeric): initial age of the age group

AgeGrpSpan (numeric): length of the age group, in years

Column Descriptions

Indicator	IndicatorName	Unit
TPopulation1Jan	Total Population, as of 1 January	thousands
TPopulation1July	Total Population, as of 1 July	thousands
TPopulationMale1July	Male Population, as of 1 July	thousands
TPopulationFemale1July	Female Population, as of 1 July	thousands
PopDensity	Population Density, as of 1 July	persons per square km
PopSexRatio	Population Sex Ratio, as of 1 July	males per 100 females
MedianAgePop	Median Age, as of 1 July	years
NatChange	Natural Change, Births minus Deaths	thousands
NatChangeRT	Rate of Natural Change	per 1,000 population
PopChange	Population Change	thousands
PopGrowthRate	Population Growth Rate	percentage
DoublingTime	Population Annual Doubling Time	years
Births	Births	thousands
BirthsMale	Male Births	thousands
BirthsFemale	Females Births	thousands
Births1519	Births by women aged 15 to 19	thousands
CBR	Crude Birth Rate	births per 1,000 population
TFR	Total Fertility Rate	live births per woman
NRR	Net Reproduction Rate	surviving daughters per woman
MAC	Mean Age Childbearing	years
SRB	Sex Ratio at Birth	males per 100 female births
Deaths	Total Deaths	thousands
DeathsMale	Male Deaths	thousands
DeathsFemale	Female Deaths	thousands
CDR	Crude Death Rate	deaths per 1,000 population
LEx	Life Expectancy at Birth, both sexes	years
LExMale	Male Life Expectancy at Birth	years
LExFemale	Female Life Expectancy at Birth	years
LE15	Life Expectancy at Age 15, both sexes	years
LE15Male	Male Life Expectancy at Age 15	years
LE15Female	Female Life Expectancy at Age 15	years
LE60	Life Expectancy at Age 60, both sexes	years
LE60Male	Male Life Expectancy at Age 60	years

LE60Female	Female Life Expectancy at Age 60	years
LE65	Life Expectancy at Age 65, both sexes	years
LE65Male	Male Life Expectancy at Age 65	years
LE65Female	Female Life Expectancy at Age 65	years
LE80	Life Expectancy at Age 80, both sexes	years
LE80Male	Male Life Expectancy at Age 80	years
LE80Female	Female Life Expectancy at Age 80	years
InfantDeaths	Infant Deaths, under age 1	thousands
IMR	Infant Mortality Rate	infant deaths per 1,000 live births
LBsurvivingAge1	Live births Surviving to Age 1	thousands
Under5Deaths	Deaths under age 5	thousands
Q5	Under-five Mortality Rate	deaths under age 5 per 1,000 live births
Q0040	Mortality before Age 40, both sexes	deaths under age 40 per 1,000 live births
Q0040Male	Male mortality before Age 40	deaths under age 40 per 1,000 male live births
Q0040Female	Female mortality before Age 40	deaths under age 40 per 1,000 female live births
Q0060	Mortality before Age 60, both sexes	deaths under age 60 per 1,000 live births
Q0060Male	Male mortality before Age 60	deaths under age 60 per 1,000 male live births
Q0060Female	Female mortality before Age 60	deaths under age 60 per 1,000 female live births
Q1550	Mortality between Age 15 and 50, both sexes	deaths under age 50 per 1,000 alive at age 15
Q1550Male	Male mortality between Age 15 and 50	deaths under age 50 per 1,000 males alive at age 15
Q1550Female	Female mortality between Age 15 and 50	deaths under age 50 per 1,000 females alive at age 15
Q1560	Mortality between Age 15 and 60, both sexes	deaths under age 60 per 1,000 alive at age 15
Q1560Male	Male mortality between Age 15 and 60	deaths under age 60 per 1,000 males alive at age 15
Q1560Female	Female mortality between Age 15 and 60	deaths under age 60 per 1,000 females alive at age 15
NetMigrations	Net Number of Migrants	thousands
CNMR	Net Migration Rate	per 1,000 population

Ethical Considerations

- While this data is likely trustworthy until the estimates begin, the UN does state that the COVID-19 pandemic did likely reduce the data that it had access to for previous years which should be taken into consideration. However, the fact that this was rectified in the next few years may help smoothen the impact. Some countries did supply administrative data instead of census data ([Country-specific data collection methods](#)).
- **Data Ethics**
 - o Citation required and documented – see above
 - o Privacy – no identifying information relevant or available

Data Relevance

- With regards to the data-relevance. I think this data could be very relevant in trying to answer the question posed. It will need to be cleaned, pruned, and the irrelevant information purged, especially with regards to the data past 2024, which is unwanted in this analysis. Questions asked in this dataset include social and economic questions such as:

Data Profile

Cleaning

- Removed previous analysis steps such as Grouping economic regions together under Location (i.e. AUKUS) so that only country names remain to allow geographic analyses that had an ISO3_code
- Removed historical data previous to 2015 and forecasted data after 2024
- Checked for useless blanks and duplicates – none found
- Exported as CSV to coding folder
- All data types acceptable

Data Statistics – Example

	SortOrder	LocID	Notes	SDMX_code	LocTypeID	ParentID
count	2320	2320	590	2320	2320	2320
mean	194.4482759	439.1336207	17.37288136	439.1336207	4	1197.74569
std	76.18935862	253.4832361	11.07759069	253.4832361	0	1090.720106
min	64	4	1	4	4	906
25%	128.75	221	5	221	4	914
50%	195	439	19	439	4	922
75%	259.25	659.25	29	659.25	4	926
max	326	894	36	894	4	5501

Data Merging

- Merged with data from WB, through Python package download, on
 - o KEY – 'ISO3_code'
- Initial Python exploratory steps use locational geo-data downloaded from Kaggle