Package ‘censusindia’

October 18, 2026

Title Census of India Data, 1901-2011, with Boundaries and Projections

Version 0.2.1

Description Tables from every decennial Census of India from 1901 to 2011: population time series, Primary Census Abstracts at state, district and sub-district level, mother tongue tables, Scheduled Caste and Scheduled Tribe populations, and the 1931 caste tables by religion, together with Ministry of Health and Family Welfare population projections to 2036. Historical series are drawn from Jolad and Singh (2026) <doi:10.7910/DVN/ON8CP8>. Census variables can be returned with state or district boundaries of the matching census year attached for mapping. Boundary files and the largest village-level tables are downloaded on first use and cached locally.

License MIT + file LICENSE

Copyright See file COPYRIGHTS for the data included in and downloaded by this package.

URL https://censusindia.saketlab.org/, https://github.com/saketlab/censusindia

BugReports https://github.com/saketlab/censusindia/issues

Depends R (>= 4.1.0)

Imports cli, dplyr, rlang (>= 1.0.0), sf, stats, stringr, tibble, tools, utils

Suggests curl, ggplot2, jsonlite, patchwork, plumber, s2, scales, testthat (>= 3.0.0)

Config/Needs/website biscale, DT, gganimate, ggrepel, ggridges, ggtern, gifski, knitr, rmarkdown, tidyr

Config/testthat/edition 3

Encoding UTF-8

LazyData true

LazyDataCompression xz

Config/roxygen2/version 8.1.1

NeedsCompilation no

Author Saket Choudhary [aut, cre, cph] (ORCID: <https://orcid.org/0000-0001-5202-7633>)

Maintainer Saket Choudhary <saketc@iitb.ac.in>

Repository CRAN

Date/Publication 2026-10-18 11:00:26 UTC

Contents

attach_geometry . . . 3
at_level . . . 4
census_1931_castes . . . 5
census_1931_exterior_castes . . . 6
census_1961_literacy . . . 6
census_1971 . . . 7
census_1981 . . . 8
census_2001_pca . . . 10
census_2001_pca_full . . . 11
census_2011_demographics . . . 12
census_2011_district_languages . . . 13
census_2011_linguistic_diversity . . . 14
census_2011_marginal_detail . . . 15
census_2011_mother_tongue . . . 16
census_2011_pca . . . 17
census_2011_sc_st . . . 18
census_2011_subdistrict_languages . . . 19
census_2011_tribes . . . 20
census_2011_workers . . . 21
census_base_layer . . . 22
census_cache_dir . . . 23
census_download . . . 23
census_languages . . . 24
census_population_time_series . . . 25
census_sc_st . . . 26
census_subdistricts_2011 . . . 27
census_variables . . . 28
census_years . . . 28
clear_geometry_cache . . . 29
compare_maps . . . 29
get_census . . . 30
get_census_boundaries . . . 31
get_palette . . . 32
get_population . . . 33
india_states . . . 34
list_census_geographies . . . 35
list_census_variables . . . 35

list_states 36
plotting 36
plot_map 37
population_projections_district 39
population_projections_district_age 40
population_projections_district_lgd 41
population_projections_district_lgd_age 42
population_projections_state 43
population_projections_state_age 44
search_census_variables 45
serve 45
sex_ratio 46
share 47
theme_census 48

Index **49**

`attach_geometry` *Attach geographic boundaries to census data*

Description

Adds geometry for mapping. Auto-detects state or district level from available columns.

Usage

```
attach_geometry(
  data,
  year,
  geography = NULL,
  unmatched = c("warn", "error", "ignore")
)
```

Arguments

`data`	A data frame containing census data with state/district names.
`year`	Census year for boundaries. Available: 1941, 1951, 1961, 1971, 1981, 1991, 2001, 2011.
`geography`	Geographic level: "state" or "district". If NULL (default), auto-detects based on available columns.
`unmatched`	What to do when rows get no geometry, or when two rows are given the same polygon: `"warn"` (default), `"error"`, or `"ignore"`. Name matching falls back to substring and edit distance, so a district can be handed a wrong-but-similar shape; this is how the check surfaces.

Value

An sf object with geometry attached.

Boundary files are downloaded on first use; see `census_cache_dir()`.

Examples

```
# Attach district boundaries
census_2011_pca |>
  dplyr::mutate(st_pct = 100 * st_population / population_total) |>
  attach_geometry(2011)
```

---

at_level — *Census rows at one geographic level*

Description

The village-level 2011 tables (`census_2011_demographics()`, `census_2011_workers()`, `census_2011_marginal_detail` stack seven levels from India down to ward, and three sectors. Summing without filtering counts every population several times over. census_1981 stacks levels and sectors the same way.

Usage

```
at_level(data, level = "district", sector = "total")
```

Arguments

`data`	A table with a `level` column, such as one of the village-level 2011 tables or census_1981.
`level`	Geographic level, e.g. `"district"` or `"state"`.
`sector`	`"total"`, `"rural"` or `"urban"`.

Details

India, state, district, subdistrict and village are exact: each level sums to its parent for every numeric column, and rural + urban sums to total.

`"town"` and `"ward"` are not. The census files a municipality and that same municipality plus its outgrowth as separate town rows, so the two overlap. Summing towns to their subdistrict overcounts 325 subdistricts by 73,877,047 people in total; Bengaluru's `BBMP (M Corp.)` and `BBMP (M Corp. + OG)` each carry 8,443,675. Ward rows differ from their town by 3,738,272 across 295 towns. Treat both as a lookup, not an additive slice.

Value

`data` filtered to that level and sector.

Examples

```
census_1981 |>
  at_level("state") |>
  nrow()
```

`census_1931_castes` *Census 1931 castes by religion*

Description

Population of every caste and community listed in the caste tables of the 1931 Census, cross-classified by religion, for 24 provinces, princely states and agencies of British India.

Usage

```
census_1931_castes
```

Format

A tibble with 2,226 rows and 17 columns:

year Census year (1931)

province Province, princely state or agency, standardised

unit_type `"British Province"`, `"Princely State"` or `"Agency"`

caste Caste or community as printed

rank Position in the province's ranked caste list, 1 the largest

hindu,muslim,sikh,jain,buddhist,christian,tribal,parsi,jew Persons of the caste returned under each religion; `tribal` is the 1931 "Tribal" or "Primitive" religion return

total Sum of the religion columns

total_ranked_list Total printed in the province's ranked list; NA for the Andaman & Nicobar Islands

notes Compiler's flag, for example where `total` and `total_ranked_list` disagree

Details

A religion column is NA where the province's table has no column for that religion, and 0 where the column exists and no one was returned. Ajmer-Merwara has totals only, so its religion columns and `total` are NA. Residual entries such as "Caste Unspecified" and "Indian Christian" are rows of their own.

Source

Jolad, Shivakumar, Kalyani, Gaurav and Baharadia, Samarth (2026). "Database of castes." Harvard Dataverse. doi:10.7910/DVN/WT5VIY. Digitised from the provincial, princely state and agency reports of the Census of India, 1931.

census_1931_exterior_castes *Census 1931 exterior castes*

Description

Castes and tribes the 1931 Census treated as exterior, with their population, by province or state section of Appendix I of the Census of India 1931, Volume I. "Exterior castes" is the report's term.

Usage

```
census_1931_exterior_castes
```

Format

A tibble with 533 rows and 5 columns:

year Census year (1931)

section Province or state section of the appendix

caste Caste or tribe as printed, with bracketed qualifiers

population Persons treated as exterior

note Compiler's remark

Source

Jolad, Shivakumar, Kalyani, Gaurav and Baharadia, Samarth (2026). "Database of castes." Harvard Dataverse. doi:10.7910/DVN/WT5VIY. Census of India, 1931, Volume I, Appendix I: Exterior Castes.

census_1961_literacy *Census 1961 literacy data*

Description

District-level literacy rates from the 1961 Census.

Usage

```
census_1961_literacy
```

Format

A tibble with 347 rows and 8 columns:

year Census year (1961)

state Name of the state

state_name_harmonized Harmonized state name for joining across datasets

district Name of the district

state_district_code Combined state-district code

literacy_total Total literacy rate (percent)

literacy_male Male literacy rate (percent)

literacy_female Female literacy rate (percent)

Source

Jolad, Shivakumar and Singh, Madhav (2026). "Indian Census Data Collection, 1901-2026: Digitised Subnational Population and Administrative Datasets." Harvard Dataverse. doi:10.7910/DVN/ON8CP8.

---

census_1971 *Census 1971 Primary Census Abstract*

Description

State and district population from the 1971 Census with rural/urban breakdown and SC/ST populations.

Usage

```
census_1971
```

Format

A tibble with 370 rows and 21 columns:

year Census year (1971)

geography Geographic level: "state" or "district"

state Name of the state

state_name_harmonized Harmonized state name for joining across datasets

name Name of the state or district

area_km2 Area in square kilometers

population_total Total population

population_rural Rural population

population_urban Urban population

males_total Total male population

males_rural Rural male population

males_urban Urban male population

females_total Total female population

females_rural Rural female population

females_urban Urban female population

sc_population_total Total Scheduled Caste population

sc_population_rural Rural Scheduled Caste population

sc_population_urban Urban Scheduled Caste population

st_population_total Total Scheduled Tribe population

st_population_rural Rural Scheduled Tribe population

st_population_urban Urban Scheduled Tribe population

Known defects

Population, males and females partition exactly into rural and urban. The SC/ST columns do not: one `sc_population` row and five `st_population` rows fail `rural + urban == total`, the largest gaps being 493,272 (Hooghly, SC) and 472,910 (Jhabua, ST). District coverage is also short of the state total for Mysore and Tamil Nadu.

Source

Jolad, Shivakumar and Singh, Madhav (2026). "Indian Census Data Collection, 1901-2026: Digitised Subnational Population and Administrative Datasets." Harvard Dataverse. doi:10.7910/DVN/ON8CP8.

`census_1981` *Census 1981 Primary Census Abstract*

Description

State and district data from the 1981 Census with literacy and worker classification.

Usage

```
census_1981
```

Format

A tibble with 2,364 rows and 39 columns including:

year Census year (1981)

geo_id Unique geographic identifier

geo_name Name of the geographic unit

level Geographic level (india, state, district)

state Name of the state

state_name_harmonized Harmonized state name for joining across datasets

district Name of the district (if applicable)

sector Sector: "total", "rural", or "urban"

area_km2 Area in square kilometers

total_persons Total population

total_males Male population

total_females Female population

literate_persons Total literate population

main_workers_persons Total main workers

cultivators_persons Total cultivators

agri_labour_persons Total agricultural labourers

hh_industry_persons Household industry workers

other_workers_persons Other workers

non_workers_persons Non-workers

Known defects

The source is OCR'd and carries transcription errors. Of 420 rows with a rural and urban counterpart, 83 fail `rural + urban == total`; the largest gap is 101,809,324 in Himachal Pradesh `non_workers_males`, a powers-of-ten slip rather than rounding. District sums disagree with the state row in 19 of 30 states. Check any column you rely on before trusting it.

Source

Jolad, Shivakumar and Singh, Madhav (2026). "Indian Census Data Collection, 1901-2026: Digitised Subnational Population and Administrative Datasets." Harvard Dataverse. doi:10.7910/DVN/ON8CP8.

census_2001_pca — *Census 2001 Primary Census Abstract (PCA)*

Description

District-level population, SC/ST, literacy, and worker statistics from the 2001 Census. Covers 593 districts; `population_total` sums to 1,028,610,328, the 2001 India total.

Usage

```
census_2001_pca
```

Format

A tibble with 593 rows and 19 columns:

year Census year (2001)

geography Geographic level ("district")

state_code Numeric state code

state_name Name of the state

state_name_harmonized Harmonized state name for joining across datasets

district_code Numeric district code

name Name of the district

households Number of households

population_total Total population

population_male Male population

population_female Female population

population_0_6 Population aged 0-6 years

sc_population Scheduled Caste population

st_population Scheduled Tribe population

literate_total Total literate population

workers_total Total workers

main_workers Main workers

marginal_workers Marginal workers

non_workers Non-workers

Details

Use 2001 boundaries with `attach_geometry()`. Districts are 2001 vintage, so the 2011 boundary file has no polygon for several of them and will mismatch others: `North Cachar Hills` (renamed Dima Hasao in 2010) and an undivided `Medinipur` (split into Purba and Paschim in 2002) both resolve to the wrong 2011 district. See census_2011_pca for the 2011 round.

Until version 0.0.0.9000 this table shipped as `census_2011_pca` with a hard-coded `year` of 2011.

Source

Census of India 2001, Primary Census Abstract.

`census_2001_pca_full` *Census 2001 Primary Census Abstract, full indicator set*

Description

The complete published 2001 PCA at district level: every indicator split by sex, and each district given for total, rural and urban. 593 districts x 3 sectors = 1,779 rows.

Usage

```
census_2001_pca_full
```

Format

A tibble with 1,779 rows and 47 columns. Beyond the identifiers (`year`, `geography`, `state_code`, `state_name`, `district_code`, `name`, `sector`) each indicator appears with a `_total`, `_male` and `_female` suffix:

households Number of households

population_total, population_male, population_female Total population

pop_0_6_total, pop_0_6_male, pop_0_6_female Population aged 0-6 years

sc_total, sc_male, sc_female Scheduled Caste population

st_total, st_male, st_female Scheduled Tribe population

literate_total, literate_male, literate_female Literate population

illiterate_total, illiterate_male, illiterate_female Illiterate population

workers_total, workers_male, workers_female All workers

main_workers_total, main_workers_male, main_workers_female Main workers

marginal_workers_total, marginal_workers_male, marginal_workers_female Marginal workers

main_cultivators_total, main_agri_labour_total, main_hh_industry_total, main_other_total Main workers by category

marginal_cultivators_total, marginal_agri_labour_total, marginal_hh_industry_total, marginal_other_total Marginal workers by category

non_workers_total, non_workers_male, non_workers_female Non-workers

Details

census_2001_pca is the narrower 19-column view of the same data, kept for the columns it shares with census_2011_pca.

Source

Census of India 2001, Primary Census Abstract (series `PC01_PCA_TOT`).

census_2011_demographics

Census 2011 demographics (PCA)

Description

Population, SC/ST, and literacy counts from the 2011 Primary Census Abstract at all geographic levels (India, state, district, subdistrict, town/village, ward), split by total/rural/urban sector. Downloaded on first use; see `census_cache_dir()`. Use `at_level()` before summing.

Usage

```
census_2011_demographics()
```

Value

A tibble with 751,594 rows and 28 columns:

state_code Numeric state code

district_code Numeric district code

subdistrict_code Numeric subdistrict code

town_village_code Town or village code

ward_code Ward code

level Geographic level (india, state, district, subdistrict, town, village, ward)

name Name of the geographic unit

state_name_harmonized Harmonized state name for joining across datasets

sector Sector: "total", "rural", or "urban"

households Number of households

population_total,population_male,population_female Total, male, and female population

pop_0_6_total,pop_0_6_male,pop_0_6_female Population aged 0-6 years

sc_total,sc_male,sc_female Scheduled Caste population

st_total,st_male,st_female Scheduled Tribe population

literate_total,literate_male,literate_female Literate population

illiterate_total,illiterate_male,illiterate_female Illiterate population

Source

Census of India 2011, Primary Census Abstract.

Examples

```
census_2011_demographics() |> at_level("state")
```

census_2011_district_languages

Census 2011 district languages (C-16)

Description

Mother tongue speakers by language at district level from the 2011 Census C-16 tables, with male/female and rural/urban breakdowns.

Usage

```
census_2011_district_languages
```

Format

A tibble with 38,993 rows and 11 columns:

state_code Numeric state code

state_name_harmonized Harmonized state name for joining across datasets

district_code Numeric district code

language_name Name of the language

language_group Numeric language group code

total_speakers Total speakers

male_speakers Male speakers

female_speakers Female speakers

rural_speakers Rural speakers

urban_speakers Urban speakers

district_name Name of the district

Source

Census of India 2011, C-16 Mother Tongue Tables.

`census_2011_linguistic_diversity`
Census 2011 linguistic diversity

Description

District-level linguistic diversity metrics derived from the 2011 Census C-16 mother tongue tables, counting named mother tongues only. C-16 files the unscheduled languages of the northeast under per-group "Others" residuals, which carry no language name and are excluded; `coverage` reports the share of the district the metrics are computed from, and falls below 0.9 in 34 districts and below 0.5 in 7.

Usage

```
census_2011_linguistic_diversity
```

Format

A tibble with 640 rows and 11 columns:

state_name_harmonized Harmonized state name for joining across datasets

state_code Numeric state code

district_code Numeric district code

n_languages Number of distinct languages spoken

total_speakers Total speakers across all languages

shannon_entropy Shannon entropy of the language distribution

effective_languages Effective number of languages (2^shannon_entropy)

dominant_language Most-spoken language

dominant_share Share of speakers of the dominant language

district_name Name of the district

coverage Share of district population in named mother tongues

Source

Census of India 2011, C-16 Mother Tongue Tables (derived).

census_2011_marginal_detail

Census 2011 marginal worker detail (PCA)

Description

Marginal workers from the 2011 Primary Census Abstract split by duration of work (3-6 months and 0-3 months) and by activity, plus non-workers, at all geographic levels and total/rural/urban sector. Downloaded on first use; see `census_cache_dir()`. Use `at_level()` before summing.

Usage

```
census_2011_marginal_detail()
```

Value

A tibble with 751,594 rows and 42 columns:

state_code,district_code,subdistrict_code,town_village_code,ward_code Geographic codes

level Geographic level (india, state, district, subdistrict, town, village, ward)

name Name of the geographic unit

state_name_harmonized Harmonized state name for joining across datasets

sector Sector: "total", "rural", or "urban"

marginal_workers_3_6_total,marginal_workers_3_6_male,marginal_workers_3_6_female Marginal workers employed 3-6 months

marg_cultivators_3_6_total,marg_cultivators_3_6_male,marg_cultivators_3_6_female Marginal cultivators, 3-6 months

marg_agri_labour_3_6_total,marg_agri_labour_3_6_male,marg_agri_labour_3_6_female Marginal agricultural labourers, 3-6 months

marg_hh_industry_3_6_total,marg_hh_industry_3_6_male,marg_hh_industry_3_6_female Marginal household industry workers, 3-6 months

marg_other_3_6_total,marg_other_3_6_male,marg_other_3_6_female Marginal other workers, 3-6 months

marginal_workers_0_3_total,marginal_workers_0_3_male,marginal_workers_0_3_female Marginal workers employed 0-3 months

marg_cultivators_0_3_total,marg_cultivators_0_3_male,marg_cultivators_0_3_female Marginal cultivators, 0-3 months

marg_agri_labour_0_3_total,marg_agri_labour_0_3_male,marg_agri_labour_0_3_female Marginal agricultural labourers, 0-3 months

marg_hh_industry_0_3_total,marg_hh_industry_0_3_male,marg_hh_industry_0_3_female Marginal household industry workers, 0-3 months

marg_other_0_3_total,marg_other_0_3_male,marg_other_0_3_female Marginal other workers, 0-3 months

non_workers_total,non_workers_male,non_workers_female Non-workers

Source

Census of India 2011, Primary Census Abstract.

Examples

```
census_2011_marginal_detail() |> at_level("state")
```

---

census_2011_mother_tongue

Census 2011 mother tongue data (C-16)

Description

Mother tongue speakers by language at state and district levels from the 2011 Census C-16 tables, with rural/urban and male/female breakdowns. Downloaded on first use; see `census_cache_dir()`. For per-district or per-subdistrict counts that sum correctly, use `census_languages()`.

Usage

```
census_2011_mother_tongue()
```

Value

A tibble with 350,157 rows and 18 columns:

state_code Numeric state code

state_name Name of the state

state_name_harmonized Harmonized state name for joining across datasets

district_code District code ("000" for state total)

area_name Name of the state or district

language_code Census language code

language_name Name of the language or dialect

language_level L1 for main languages, L2 for dialects

language_group Numeric language group code

total_persons Total speakers

total_males Male speakers

total_females Female speakers

rural_persons Rural speakers

rural_males Rural male speakers

rural_females Rural female speakers

urban_persons Urban speakers

urban_males Urban male speakers

urban_females Urban female speakers

Source

Census of India 2011, C-16 Mother Tongue Tables.

Examples

```
census_2011_mother_tongue() |> head()
```

---

census_2011_pca *Census 2011 Primary Census Abstract (PCA)*

Description

District-level population, SC/ST, literacy, and worker statistics from the 2011 Census. Covers all 640 districts; `population_total` sums to 1,210,854,977, the 2011 India total.

Usage

```
census_2011_pca
```

Format

A tibble with 640 rows and 19 columns:

year Census year (2011)

geography Geographic level ("district")

state_code Numeric state code

state_name Name of the state

state_name_harmonized Harmonized state name for joining across datasets

district_code Numeric district code

name Name of the district

households Number of households

population_total Total population

population_male Male population

population_female Female population

population_0_6 Population aged 0-6 years

sc_population Scheduled Caste population

st_population Scheduled Tribe population

literate_total Total literate population

workers_total Total workers

main_workers Main workers

marginal_workers Marginal workers

non_workers Non-workers

Details

Aggregated at district level from the village-level 2011 PCA that also backs `census_2011_demographics()`, `census_2011_workers()` and `census_2011_marginal_detail()`. For the 2001 round see census_2001_pca.

Source

Jolad, Shivakumar and Singh, Madhav (2026). "Indian Census Data Collection, 1901-2026: Digitised Subnational Population and Administrative Datasets." Harvard Dataverse. doi:10.7910/DVN/ON8CP8.

`census_2011_sc_st` *Census 2011 Scheduled Castes and Scheduled Tribes (A-10/A-11)*

Description

District-level Scheduled Caste and Scheduled Tribe population from the 2011 Census A-10 and A-11 tables.

Usage

```
census_2011_sc_st
```

Format

A tibble with 14,028 rows and 9 columns:

year Census year (2011)

state_name_harmonized Harmonized state name for joining across datasets

district_name Name of the district

category `"Scheduled Caste"` or `"Scheduled Tribe"`, spelled out

category_code `"SC"` or `"ST"`, for filtering by the short form

caste_tribe_name Name of the caste or tribe

population Population of the caste or tribe

percentage Share of district population

district_total_population Total district population

Details

The two categories are not symmetric. Scheduled Tribes are enumerated by name, 581 of them. Scheduled Castes are not: all 582 SC rows carry the single label `"All Scheduled Castes"`, one per district. Nothing in this package resolves caste below that aggregate.

Filter with either spelling: `category` holds the long form, `category_code` holds `SC`/`ST`. `census_sc_st()` accepts both.

ST rows include `"Generic Tribes etc."`, the residual for members who returned no specific tribe: 585 rows, 2,804,148 people. census_2011_tribes is this table's ST rows with that residual removed and the names harmonised.

Source

Census of India 2011, A-10 (SC) and A-11 (ST) tables.

---

census_2011_subdistrict_languages

Census 2011 subdistrict languages (C-16)

Description

Mother tongue speakers by language at subdistrict level from the 2011 Census C-16 tables, with male/female and rural/urban breakdowns. The district rows that share each subdistrict's `district_code` are excluded, so these rows sum to the matching row of census_2011_district_languages. Downloaded on first use; see `census_cache_dir()`.

Usage

```
census_2011_subdistrict_languages()
```

Value

A tibble with 144,134 rows and 11 columns:

state_code Numeric state code

state_name_harmonized Harmonized state name for joining across datasets

district_code Numeric district code

area_name Name of the subdistrict

language_name Name of the language

language_group Numeric language group code

total_speakers Total speakers

male_speakers Male speakers

female_speakers Female speakers

rural_speakers Rural speakers

urban_speakers Urban speakers

Source

Census of India 2011, C-16 Mother Tongue Tables.

Examples

```
census_2011_subdistrict_languages() |> head()
```

census_2011_tribes *Census 2011 Scheduled Tribes (A-11)*

Description

District-level population of individual Scheduled Tribes from the 2011 Census A-11 tables. `tribe_name` is harmonised; `tribe_name_original` keeps the source string verbatim, including the full synonym list the census publishes ("Gond, Arakh, Arrakh, Agaria, ...") and any parenthetical district qualifier.

Usage

```
census_2011_tribes
```

Format

A tibble with 12,861 rows and 8 columns:

year Census year (2011)

state_name_harmonized Harmonized state name for joining across datasets

district_name Name of the district

tribe_name Harmonized tribe name

tribe_name_original Original tribe name as recorded in the source

population Tribe population

percentage Share of district population

district_total_population Total district population

Details

Harmonisation takes the first synonym and drops the qualifier, so `"Thoti (in Adilabad, Hyderabad, ...)"` becomes `"Thoti"`. That collapses 581 source strings to 483 tribe names, and it is deliberate: the same tribe is written differently across states, sometimes differing only in whitespace.

These are the ST rows of census_2011_sc_st with `"Generic Tribes etc."` removed, which is why the population here sums to 103,409,280 rather than the 106,213,428 in that table. Neither figure equals the 104,545,716 in `census_2011_pca$st_population`; the A-11 tables and the Primary Census Abstract do not reconcile, so pick one source per analysis and say which.

Source

Census of India 2011, A-11 Scheduled Tribe tables.

census_2011_workers *Census 2011 workers (PCA)*

Description

Worker classification from the 2011 Primary Census Abstract at all geographic levels, split by total/rural/urban sector. Main and marginal workers are broken down by activity (cultivators, agricultural labourers, household industry, other workers). Downloaded on first use; see `census_cache_dir()`. Use `at_level()` before summing.

Usage

```
census_2011_workers()
```

Value

A tibble with 751,594 rows and 42 columns:

state_code,district_code,subdistrict_code,town_village_code,ward_code Geographic codes

level Geographic level (india, state, district, subdistrict, town, village, ward)

name Name of the geographic unit

state_name_harmonized Harmonized state name for joining across datasets

sector Sector: "total", "rural", or "urban"

total_workers_total,total_workers_male,total_workers_female All workers

main_workers_total,main_workers_male,main_workers_female Main workers

main_cultivators_total,main_cultivators_male,main_cultivators_female Main cultivators

main_agri_labour_total,main_agri_labour_male,main_agri_labour_female Main agricultural labourers

main_hh_industry_total,main_hh_industry_male,main_hh_industry_female Main household industry workers

main_other_total,main_other_male,main_other_female Main other workers

marginal_workers_total,marginal_workers_male,marginal_workers_female Marginal workers

marginal_cultivators_total,marginal_cultivators_male,marginal_cultivators_female Marginal cultivators

marginal_agri_labour_total,marginal_agri_labour_male,marginal_agri_labour_female Marginal agricultural labourers

marginal_hh_industry_total,marginal_hh_industry_male,marginal_hh_industry_female Marginal household industry workers

marginal_other_total,marginal_other_male,marginal_other_female Marginal other workers

Source

Census of India 2011, Primary Census Abstract.

Examples

```
census_2011_workers() |> at_level("state")
```

---

census_base_layer *Base layer carrying India's full claimed extent*

Description

A ggplot2 layer of the state polygons for a census year, drawn as flat fill. Put it under a district choropleth so the national outline stays whole: the census enumerates no districts in the territory India claims but does not administer, so a district layer alone truncates the north by roughly two degrees of latitude.

Usage

```
census_base_layer(year, fill = "grey80", ...)
```

Arguments

year	Census year whose boundaries to draw. One of 1941, 1951, 1961, 1971, 1981, 1991, 2001, 2011.
fill	Fill colour for the base. Defaults to the grey plot_map() uses for missing values.
...	Passed to ggplot2::geom_sf().

Details

plot_map() adds this automatically. Use it directly when hand-rolling a map with ggplot2::geom_sf().

Value

A ggplot2 layer.

Examples

```
ggplot2::ggplot() +
  census_base_layer(2011)
```

census_cache_dir *Downloaded data cache*

Description

Census boundaries and the seven largest tables are downloaded on first use, checked against a pinned MD5 checksum, and kept here for later sessions.

Usage

```
census_cache_dir()

clear_census_cache()
```

Details

The cache lives in `tools::R_user_dir()`. Set the option `censusindia.cache_dir` to use another directory.

Value

`census_cache_dir()` returns the cache path, which may not exist yet. `clear_census_cache()` deletes it and invisibly returns that path.

See Also

`census_download()` to fetch everything at once; `clear_geometry_cache()` to release what the session holds in memory.

Examples

```
census_cache_dir()
```

census_download *Download all remote data*

Description

Fetches every boundary file and large table not yet in `census_cache_dir()`, so later calls work offline.

Usage

```
census_download()
```

Value

The cache directory, invisibly.

Examples

```
census_download()
```

---

census_languages *Language speakers at a chosen geographic level*

Description

The C-16 mother tongue tables stack state, district and subdistrict rows, and a subdistrict carries its parent's `district_code`. Filtering on `district_code != "000"` therefore keeps subdistricts too, and summing the result double counts: 640 districts become 6,231 rows, and district totals inflate by a median factor of two.

Usage

```
census_languages(level = c("district", "subdistrict"), language_group = NULL)
```

Arguments

`level`	`"district"` (38,993 rows, 640 districts) or `"subdistrict"` (144,134 rows, 5,988 areas).
`language_group`	Optionally restrict to one language group, as listed in the `language_group` column. Default `NULL` keeps all.

Details

`area_name` does not separate the two levels either, because in 392 of the 640 districts the headquarters tehsil carries the district's own name. Subdistricts do partition their district, so the district value for any `(district_code, language_code)` is half the group sum. The tables this returns are that split, applied once at build time.

Value

A tibble at the requested level, one row per language per area.

See Also

census_2011_linguistic_diversity for one row per district with entropy already computed.

Examples

```
census_languages("district") |> head()
```

census_population_time_series

Census population time series (1901-2011)

Description

Decadal population at state and district levels, 1901-2011.

Usage

```
census_population_time_series
```

Format

A tibble with 7,901 rows and 12 columns:

year Census year (1901, 1911, ..., 2011)

geography Geographic level: "state" or "district"

state_code Numeric state code

state_name Name of the state

state_name_harmonized Harmonized state name for joining across datasets

district_code Numeric district code (0 for state-level)

name Name of the state or district

population Total population

males Male population

females Female population

variation_absolute Absolute change from previous census

variation_percent Percentage change from previous census

Coverage limits

Sikkim, Mizoram and Daman & Diu are absent from every year, so 2011 sums to 1,208,903,947 across 626 districts rather than 1,210,854,977 across 640. `get_census()` uses the Primary Census Abstract for 2001 and 2011 instead. District rows also fall short of their state totals before 1961 (57 districts carry no population in 1901-1941) and in 1981, where all 27 Assam districts are missing while the state estimate is present. The `year` column further carries 1900, 1910, 1940, 1948, 1950, 1960, 1962 and 2021, which are not censuses; use `census_years()` before any `lag()`.

Source

Jolad, Shivakumar and Singh, Madhav (2026). "Indian Census Data Collection, 1901-2026: Digitised Subnational Population and Administrative Datasets." Harvard Dataverse. doi:10.7910/DVN/ON8CP8.

census_sc_st — *Scheduled Caste and Scheduled Tribe rows, by either name*

Description

census_2011_sc_st stores `category` as `"Scheduled Caste"` and `"Scheduled Tribe"`. `filter(category == "SC")` therefore matches nothing and raises no error. This accessor takes either spelling, in any case, and the table itself now carries a `category_code` column holding `SC`/`ST` so plain `dplyr::filter()` works both ways too.

Usage

```
census_sc_st(category = NULL, state = NULL)
```

Arguments

`category`	`"SC"`, `"ST"`, `"Scheduled Caste"`, `"Scheduled Tribe"`, or `NULL` (default) to keep both. Case-insensitive; a vector selects several.
`state`	Optionally restrict to one or more states, by canonical name or two-letter code, as in `get_census()`.

Details

The two categories are not symmetric. Scheduled Tribes are named, 581 of them. All 582 Scheduled Caste rows carry the single label `"All Scheduled Castes"`, one per district, so `"SC"` returns district aggregates rather than individual castes.

Value

A tibble of census_2011_sc_st rows.

See Also

census_2011_tribes for the ST rows with names harmonised and the `"Generic Tribes etc."` residual removed.

Examples

```
nrow(census_sc_st("SC"))
nrow(census_sc_st("Scheduled Caste"))
```

census_subdistricts_2011

Subdistrict directory (2011)

Description

Subdistrict (tehsil/taluka) population from the 2011 Census.

Usage

```
census_subdistricts_2011
```

Format

A tibble with 5,988 rows and 15 columns:

year Census year for population data (2011)

geography Geographic level ("subdistrict")

state Name of the state

state_name_harmonized Harmonized state name for joining across datasets

district Name of the district

subdistrict Name of the subdistrict (tehsil/taluka)

population Total population

males Male population

females Female population

households Number of households

inhabited_villages Number of inhabited villages

uninhabited_villages Number of uninhabited villages

towns Number of towns

area_km2 Area in square kilometers

density Population density per square kilometer

Details

A subdistrict crossing a district line is listed once per district, marked "(Part)". Population, males, females and households come from `census_2011_demographics()` and sum to census_2011_pca. Villages, towns, area and density come from the Dataverse source and are NA where it has no match.

Source

Jolad, Shivakumar and Singh, Madhav (2026). "Indian Census Data Collection, 1901-2026: Digitised Subnational Population and Administrative Datasets." Harvard Dataverse. doi:10.7910/DVN/ON8CP8.

census_variables *Census variables lookup*

Description

Available census variables with labels, years, geographic levels, and categories.

Usage

```
census_variables
```

Format

A tibble with 23 rows and 5 columns:

variable Variable name used in package functions

label Human-readable label

years Comma-separated list of available years

geographies Comma-separated list of available geographic levels

category Variable category (population, literacy, workers, etc.)

census_years *Census years only*

Description

census_population_time_series carries rows for 1900, 1910, 1940, 1948, 1950, 1960, 1962 and 2021 alongside the twelve decennial censuses. A `lag()` over the unfiltered table silently produces one- and thirteen-year intervals labelled as decadal.

Usage

```
census_years(data)
```

Arguments

`data`	A frame with a `year` column.

Value

`data` keeping only the twelve census years, 1901 to 2011.

Examples

```
census_population_time_series |>
  census_years() |>
  dplyr::distinct(year)
```

clear_geometry_cache *Clear geometry cache*

Description

Releases the boundaries and downloaded tables held in memory for the session.

Usage

```
clear_geometry_cache()
```

Value

Invisible NULL

See Also

`clear_census_cache()` to delete downloaded files from disk.

Examples

```
clear_geometry_cache()
```

compare_maps *Compare maps across years*

Description

Compare maps across years

Usage

```
compare_maps(
  data_list,
  fill_var,
  title = NULL,
  ncol = NULL,
  palette = "viridis",
  common_scale = TRUE,
  show_state_boundaries = TRUE,
  ...
)
```

Arguments

data_list	A named list of sf data frames, with names as years
fill_var	The variable to map (string)
title	Overall title for the combined plot
ncol	Number of columns for arrangement
palette	Color palette (see plot_map for options)
common_scale	Logical. Use the same color scale across all maps?
show_state_boundaries	Logical. Overlay state boundaries?
...	Additional arguments passed to plot_map

Value

A patchwork object

Examples

```
# Compare ST percentage across years
library(dplyr)

st_1971 <- census_1971 |>
  filter(geography == "district") |>
  mutate(st_pct = 100 * st_population_total / population_total) |>
  attach_geometry(1971)

st_2011 <- census_2011_pca |>
  mutate(st_pct = 100 * st_population / population_total) |>
  attach_geometry(2011)

# Compare
compare_maps(
  list("1971" = st_1971, "2011" = st_2011),
  fill_var = "st_pct",
  title = "ST Population % by District",
  palette = "red_blue"
)
```

---

get_census — *Get India census data*

Description

Census data for India at state, district, or subdistrict level.

Usage

```
get_census(
  year,
  geography = c("state", "district", "subdistrict"),
  variables = NULL,
  state = NULL,
  geometry = FALSE,
  sector = c("total", "rural", "urban")
)
```

Arguments

year	Census year. One of: 1901, 1911, 1921, 1931, 1941, 1951, 1961, 1971, 1981, 1991, 2001, 2011.
geography	Geographic level: "state", "district", or "subdistrict". Not all geographies are available for all years.
variables	Character vector of variables to retrieve. Use `list_census_variables()` to see available variables. If NULL (default), returns all available variables for the specified year and geography.
state	Optional. Filter to specific state(s). Can be state name or abbreviation (e.g., "Maharashtra" or "MH").
geometry	If TRUE, returns an sf object with geographic boundaries.
sector	For 1971/1981, filter to "total", "rural", or "urban". Other years have no sector split, so "rural" or "urban" is an error there.

Value

A tibble (or sf object if geometry = TRUE).

Examples

```
get_census(2011, "state")
get_census(1971, "district", state = "Maharashtra")
get_census(1981, "district", variables = c("population", "literacy_rate"), geometry = TRUE)
```

---

get_census_boundaries *Get census boundaries*

Description

State or district boundaries for a given census year.

Usage

```
get_census_boundaries(year, geography = c("state", "district"))
```

Arguments

year	Census year. Available: 1941, 1951, 1961, 1971, 1981, 1991, 2001, 2011.
geography	Geographic level: "state" or "district".

Value

An sf object containing the geographic boundaries with harmonized state names.

Boundary files are downloaded on first use; see `census_cache_dir()`.

Examples

```
boundaries <- get_census_boundaries(1971, "district")
ggplot2::ggplot(boundaries) +
  ggplot2::geom_sf()
```

---

get_palette *Get color palette*

Description

Get color palettes for maps. Available palettes: "viridis", "red_blue", "blue_red", "greens", "oranges", "reds", "blues", "purple_green".

Usage

```
get_palette(name, reverse = FALSE)
```

Arguments

name	Name of the palette
reverse	Logical. Reverse direction?

Value

A character vector of colors

Examples

```
get_palette("blues")
get_palette("red_blue", reverse = TRUE)
```

get_population — *Get population projections*

Description

MOHFW population projections (2011-2036) at state or district level, based on the 2011 Census. District projections are available using either Census 2011 boundaries (640 districts, 5-year intervals) or current LGD boundaries (785 districts, annual).

Usage

```
get_population(
  year = NULL,
  geography = c("state", "district"),
  state = NULL,
  boundary = c("census2011", "lgd"),
  geometry = FALSE,
  by_age = FALSE
)
```

Arguments

`year`	Integer or integer vector of projection years. If NULL (default), returns all available years.
`geography`	Geographic level: `"state"` or `"district"`.
`state`	Optional. Filter to specific state(s) by name or abbreviation (e.g., `"Maharashtra"` or `"MH"`).
`boundary`	For district geography, which boundary system: `"census2011"` (default, 640 districts) or `"lgd"` (785 districts). Ignored for state geography.
`geometry`	If `TRUE`, attaches geographic boundaries (Census 2011). Only works with `boundary = "census2011"` for district level.
`by_age`	If `TRUE`, returns one row per age group. States come in 5-year bands from "0-4" to "80+"; districts in single years of age 0-14, then 5-year bands 15-19 through 80+. See population_projections_state_age, `population_projections_district_age()` and `population_projections_district_lgd_age()`. State coverage by age is 23 major states plus an "India" row and one North-East aggregate row, "North East States Excluding Assam"; any other state is an error.

Details

State-level data covers 2011-2036 and has 39 entries: 28 states, the 9 UTs that MOHFW gives, the merged "Dadra & Nagar Haveli and Daman & Diu", and "India". MOHFW gives "Dadra & Nagar Haveli" and "Daman & Diu" as two UTs. This package adds the merged UT as the sum of the two. To get the sum of the states, remove "India" and one of the two Dadra/Daman forms. Refer to population_projections_state. Values are numbers of persons. This is also true for the state rows by age, but population_projections_state_age stores them in thousands.

The source data rounds Persons, Male, and Female independently (in thousands), so `population` may differ from `males` + `females` by up to 1000 at the state level.

District-level Census 2011 data has projections at 5-year intervals (2011, 2016, 2021, 2026, 2031). LGD district data has annual projections from 2012 to 2031.

A `year` absent from the table is estimated from the two nearest table years by geometric (exponential) growth, applied separately to `males`, `females` and `population`:

$$P_t = P_1(P_2/P_1)^{(t-t_1)/(t_2-t_1)}$$

Estimated districts need not sum exactly to the estimated state.

Years after the last table year extrapolate with the last interval's rate, up to five years out: 2041 for states, 2036 for districts. Years before 2011 or past that limit are an error.

Telangana and Ladakh are included as separate entries even though they did not exist as states at the time of the 2011 Census.

Value

A tibble (or sf object if `geometry = TRUE`) with columns: `year`, `state_name_harmonized`, `males`, `females`, `population`, `estimate`, `district` for district-level data, and `age_group` if `by_age = TRUE`. `estimate` is `"published"` for a year printed in the MOHFW tables, `"interpolated"` or `"extrapolated"` otherwise.

Examples

```
# State-level projections for 2021
get_population(2021, "state")

# District-level for Kerala
get_population(2021, "district", state = "Kerala")

# LGD boundary districts
get_population(2021, "district", boundary = "lgd")

# Age-stratified state projections
get_population(2021, "state", by_age = TRUE)
```

`india_states` *Indian states lookup*

Description

States and union territories with codes, abbreviations, and regions.

Usage

```
india_states
```

Format

A tibble with 36 rows and 5 columns:

state_code Numeric state code (Census 2011 codes)

state_name Official state name

state_name_harmonized Harmonized state name for joining across datasets

state_abbr Two-letter state abbreviation

region Geographic region (North, South, East, West, Central, Northeast, Islands)

---

list_census_geographies *List available census geographies*

Description

Geographic levels available for each census year.

Usage

```
list_census_geographies()
```

Value

A tibble showing available year-geography combinations.

Examples

```
list_census_geographies()
```

---

list_census_variables *List census variables*

Description

Available census variables, optionally filtered by year, geography, or category.

Usage

```
list_census_variables(year = NULL, geography = NULL, category = NULL)
```

Arguments

year	Filter to variables available for specific census year.
geography	Filter to "state", "district", or "subdistrict".
category	Filter to "population", "literacy", "workers", "sc_st", "housing", "geography", or "change".

Value

A tibble of available variables.

Examples

```
list_census_variables()
list_census_variables(year = 1981)
list_census_variables(category = "literacy", geography = "district")
```

---

list_states *List Indian states*

Description

States and union territories with codes, abbreviations, and regions.

Usage

```
list_states(region = NULL)
```

Arguments

`region`	Filter by region: "North", "South", "East", "West", "Central", "Northeast", "Islands".

Value

A tibble of states with codes, names, abbreviations, and regions.

Examples

```
list_states()
list_states(region = "South")
```

---

plotting *Plotting functions for census data*

Description

Choropleth maps and comparison plots from census data.

plot_map *Create a choropleth map from census data.*

Description

Create a choropleth map from census data.

Usage

```
plot_map(
  data,
  fill_var,
  title = NULL,
  subtitle = NULL,
  legend_title = NULL,
  palette = "viridis",
  reverse = FALSE,
  direction = NULL,
  na_color = "grey80",
  boundary_year = NULL,
  base_layer = TRUE,
  show_state_boundaries = FALSE,
  state_boundary_color = "black",
  state_boundary_width = 0.3,
  trans = "identity",
  midpoint = NULL,
  oob = NULL,
  limits = NULL,
  breaks = NULL,
  labels = NULL,
  font_family = NULL,
  legend_position = "right"
)
```

Arguments

data	An sf object or data frame with geometry attached via attach_geometry
fill_var	The variable to map to fill color (unquoted or string)
title	Optional title for the map
subtitle	Optional subtitle for the map
legend_title	Title for the legend
palette	Color palette. Options include: • "viridis" - Perceptually uniform, colorblind-friendly (default) • "red_blue" - Diverging red to blue (red = high) • "blue_red" - Diverging blue to red (blue = high)

- "greens" - Sequential green palette
- "oranges" - Sequential orange palette
- "reds" - Sequential red palette
- "blues" - Sequential blue palette
- "purple_green" - Diverging purple to green

`reverse` Logical. Reverse the color palette direction? Default FALSE.

`direction` Numeric. Alternative to reverse: use -1 to reverse, 1 for default. If both reverse and direction are specified, direction takes precedence.

`na_color` Color for missing values. Default is "grey80"

`boundary_year` Census year whose state boundaries to draw beneath the data, keeping India's full claimed extent on the map even where the census enumerates no districts. Defaults to the `year` column if present.

`base_layer` Logical. Draw India's full state outline beneath the data? Default TRUE

`show_state_boundaries` Logical. Overlay state boundaries?

`state_boundary_color` Color for state boundaries. Default is "black"

`state_boundary_width` Width of state boundary lines. Default is 0.3

`trans` Transformation for the color scale (e.g., "log10", "sqrt"). Default is "identity" (no transformation)

`midpoint` Value to centre a diverging palette on, e.g. `0` for growth or `1000` for a sex ratio. Without it the palette centres on the middle of `limits`, which is rarely the meaningful value.

`oob` How to treat values outside `limits`. The ggplot2 default censors them to `na_color`, making them indistinguishable from genuinely missing data; pass `scales::squish` to clamp them to the nearest end instead.

`limits` Optional limits for the color scale as c(min, max)

`breaks` Optional breaks for the legend

`labels` Optional labels for the legend breaks

`font_family` Font family for all map text. Default `NULL` uses the system default; e.g. `"Arial"`.

`legend_position` One of `"right"` (default), `"left"`, `"bottom"`, `"top"`, or `"none"`. `"bottom"`/`"top"` also turns the colorbar horizontal.

Value

A ggplot2 object

Examples

```
# Calculate ST percentage and attach geometry
st_data <- census_2011_pca |>
  dplyr::mutate(st_pct = 100 * st_population / population_total) |>
```

```
  attach_geometry(2011)

# Create a basic map
plot_map(st_data, st_pct, title = "ST Population %")

# With red-blue palette and state boundaries
plot_map(st_data, st_pct,
  title = "ST Population % (2011)",
  palette = "red_blue",
  show_state_boundaries = TRUE
)
```

`population_projections_district`

Population projections – district level, Census 2011 boundaries (2011-2031)

Description

MOHFW population projections at district level using Census 2011 boundaries (640 districts). Projections are at 5-year intervals: 2011, 2016, 2021, 2026, 2031. Values match the IIPS report's Table 8.

Usage

```
population_projections_district
```

Format

A tibble with 3,200 rows and 6 columns:

year Projection year (5-year intervals)

state_name_harmonized Harmonized state name for joining

district District name (MOHFW naming convention)

males Projected male population

females Projected female population

population Projected total population (males + females)

Details

The report's age tables (`population_projections_district_age()`) sum to different totals for Karnataka in 2016-2031 and for Uttar Pradesh and Nagaland in 2016.

Compatible with `attach_geometry()` using 2011 district boundaries.

Source

International Institute for Population Sciences. Projection of District Level Annual Population by Quinquennial Agegroup and Sex. https://www.iipsindia.ac.in/sites/default/files/1_3.pdf

---

`population_projections_district_age`

Population projections, district level, Census 2011 boundaries, by age group (2012-2031)

Description

Age-stratified detail behind population_projections_district, on Census 2011 district boundaries (no LGD spatial apportionment involved). Same age-group scheme as `population_projections_district_lgd_age()`: single years 0-14, then 5-year bands through 80+. Downloaded on first use; see `census_cache_dir()`.

Usage

```
population_projections_district_age()
```

Details

Values are as printed in the IIPS report. District names match population_projections_district.

Two blocks are derived because the PDF does not print them. Moga (Punjab) and Tapi (Gujarat) print no 80+ row for 2022-2026; it is the printed All ages minus the other printed age groups. Theni's (Tamil Nadu) 2017-2031 pages repeat Madurai's numbers; Karnataka's age tables sum to a different projection from population_projections_district.

Value

A tibble with 371,200 rows and 7 columns:

year Projection year (annual, 2012-2031)

state_name_harmonized Harmonized state name for joining

district District name (Census 2011 naming convention)

age_group Single years "0".."14", then 5-year bands "15-19" through "80+"

males Projected male population

females Projected female population

population Projected total population (males + females)

Source

Same as population_projections_district.

Examples

```
population_projections_district_age() |> head()
```

population_projections_district_lgd
Population projections – district level, LGD boundaries (2012-2031)

Description

MOHFW population projections scaled to current Local Government Directory (LGD) district boundaries. Annual projections from 2012 to 2031, summed across age groups from population_projections_district_l

Usage

```
population_projections_district_lgd
```

Format

A tibble with 15,660 rows and 6 columns:

year Projection year (annual, 2012-2031)

state_name_harmonized Harmonized state name for joining. Dadra & Nagar Haveli and Daman & Diu appear merged as one UT (their 2020 administrative merger); the LGD source has no district-level way to split them back apart.

district District name (LGD naming convention)

males Projected male population

females Projected female population

population Projected total population (males + females)

Known defects

Derived from population_projections_district_age() by area overlap, so district sums do not exactly reproduce population_projections_state, and Karnataka inherits the age tables' disagreement with population_projections_district. Theni (Tamil Nadu) 2017-2031 and the 80+ group of Moga (Punjab) and Tapi (Gujarat) in 2022-2026 are derived, as in population_projections_district_age().

Source

International Institute for Population Sciences. Projection of District Level Annual Population by Quinquennial Agegroup and Sex. https://www.iipsindia.ac.in/sites/default/files/1_3.pdf District scaling from Census 2011 to LGD boundaries via spatial overlap analysis.

population_projections_district_lgd_age

Population projections, district level, LGD boundaries, by age group (2012-2031)

Description

Age-stratified detail behind population_projections_district_lgd: single years of age for 0-14, five-year bands from 15-19 through 80+. Fixes the same double-counted "All ages" row bug documented on that table. Downloaded on first use; see `census_cache_dir()`.

Usage

```
population_projections_district_lgd_age()
```

Value

A tibble with 454,140 rows and 7 columns:

year Projection year (annual, 2012-2031)

state_name_harmonized Harmonized state name for joining. See population_projections_district_lgd for the Dadra & Nagar Haveli / Daman & Diu merged-UT caveat.

district District name (LGD naming convention)

age_group Single years "0".."14", then 5-year bands "15-19" through "80+". Non-overlapping: summing all rows for a district-year reproduces population_projections_district_lgd.

males Projected male population

females Projected female population

population Projected total population (males + females)

Source

Same as population_projections_district_lgd.

Examples

```
population_projections_district_lgd_age() |> head()
```

population_projections_state

Population projections - state level (2011-2036)

Description

MOHFW population projections for each state, from the 2011 Census. The table gives one projection for each year from 2011 to 2036.

Usage

```
population_projections_state
```

Format

A tibble with 1014 rows (39 entries x 26 years) and 5 columns:

year Projection year (2011-2036)

state_name_harmonized Harmonized state name for joining

males Projected male population

females Projected female population

population Projected total population

Details

India has 36 states and UTs. The table has 39 entries:

- 28 states.
- 9 UTs, with the names that the report uses: Andaman & Nicobar Islands, Chandigarh, Delhi, Jammu & Kashmir, Ladakh, Lakshadweep, Puducherry, "Dadra & Nagar Haveli" and "Daman & Diu". Jammu & Kashmir does not include Ladakh.
- "Dadra & Nagar Haveli and Daman & Diu". The two UTs became one UT in January 2020. This row is the sum of the two UTs. It has the same name as in the LGD district table, so data from 2020 and later can join.
- "India", the national total.

Thus the population of the two old UTs is in the table two times. To get the sum of the states, remove "India". Also remove the merged UT, or remove the two old UTs.

Values are numbers of persons. The source rounds persons, males and females to thousands separately. Thus `population` can be different from `males` + `females` by a maximum of 1000. For the merged UT, the maximum is 2000, because the row is the sum of two rounded rows.

Values are the same as Table 8 of the report. The India and Odisha rows use the July 2020 edition. That edition changed the male projection for Odisha. Thus the two rows are different from the November 2019 report at the link below.

Source

Ministry of Health and Family Welfare, Government of India. Population Projections for India and States 2011-2036. https://nhm.gov.in/New_Updates_2018/Report_Population_Projection_2019.pdf

---

`population_projections_state_age`

Population projections – state level, by age group (2011-2036)

Description

Age-stratified population reshaped directly from the MOHFW report's own Table 18, wide-by-year to tidy long. Summing to state-year reproduces population_projections_state to within report rounding (Uttar Pradesh 2016: 216,089 vs 216,087 thousand).

Usage

```
population_projections_state_age
```

Format

A tibble with 2,448 rows and 5 columns:

year Projection year (2011-2036, at 5-year intervals: 2011, 2016, 2021, 2026, 2031, 2036)

state_name_harmonized Harmonized state name for joining where the report breaks the state out individually; "India" and "North East States Excluding Assam" pass through verbatim as aggregate rows

age_group Five-year bands "0-4" through "80+"

males Projected male population, in thousands

females Projected female population, in thousands

population Projected total population (the printed Person figure), in thousands

Details

Values match the printed Table 18. `population` is the printed Person figure, rounded separately from `males` and `females`, so it can differ from their sum by 1. India and Odisha follow the July 2020 edition, as in population_projections_state, and there `population` is `males + females`.

Coverage is narrower than population_projections_state: the report only breaks out 23 major states individually plus an "India" total, lumping the smaller North East states (Sikkim, Arunachal Pradesh, Nagaland, Manipur, Mizoram, Tripura, Meghalaya) into one "North East States Excluding Assam" row. Chandigarh, small UTs and Ladakh are not in this table at all.

Source

Ministry of Health and Family Welfare, Government of India. Population Projections for India and States 2011-2036, Table 18. https://nhm.gov.in/New_Updates_2018/Report_Population_Projection_2019.pdf

search_census_variables *Search census variables*

Description

Find census variables by keyword in names and labels.

Usage

```
search_census_variables(pattern)
```

Arguments

pattern	Character string or regex pattern to search.

Value

A tibble of matching variables.

Examples

```
search_census_variables("population")
search_census_variables("worker")
```

serve *Start the censusindia REST API*

Description

Launches the bundled plumber API on the given host and port. Every endpoint takes `GET` query parameters that mirror the R function behind it, with comma-separated values for vectors (`state=KL,TN`):

Usage

```
serve(host = "127.0.0.1", port = 8895L, ...)

census_api(host = "127.0.0.1", port = 8895L, ...)
```

Arguments

host	Host address to bind. Default `"127.0.0.1"` (localhost only). Use `"0.0.0.0"` to accept external connections.
port	Port number. Default `8895`.
...	Additional arguments passed to `plumber`'s `$run()` method.

Details

Endpoint	Function
/query	get_census()
/population	get_population()
/boundaries	get_census_boundaries()
/languages	census_languages()
/sc-st	census_sc_st()
/datasets, /dataset/<name>	any shipped dataset, e.g. census_1931_castes
/variables, /search	list_census_variables(), search_census_variables()
/states, /geographies	list_states(), list_census_geographies()
/health	package version

Results come back as JSON rows, or as a GeoJSON FeatureCollection when geometry=true and from /boundaries. Input the R function rejects is a 400 carrying its error message, and a failed data download is a 503. Interactive documentation is served at /__docs__/.

A response holds at most 10,000 rows, or 100 features for GeoJSON. Page through larger results with limit and offset; the X-Total-Count header gives the full count.

Each request writes one JSON line to standard output with its path, query, status, duration, client IP and a client label. The label comes from an X-Client header when one is sent (an MCP server wrapping the API should send X-Client: mcp/<name>), and otherwise from the User-Agent: mcp, ai_agent, curl, wget, python, r, node, browser, other or unknown. Under systemd these lines land in the journal.

Value

Called for its side effect; does not return under normal operation.

Examples

```
# Start locally on the default port
serve()

# Bind to all interfaces on a custom port
serve(host = "0.0.0.0", port = 9000)
```

---

sex_ratio *Sex ratio*

Description

Females per 1,000 males, the convention the Census of India reports.

Usage

```
sex_ratio(females, males)
```

Arguments

`females`, `males` Counts, of equal length.

Value

Numeric vector.

Examples

```
sex_ratio(census_2011_pca$population_female, census_2011_pca$population_male) |> head()
```

---

share *Share of a total, as a percentage*

Description

Share of a total, as a percentage

Usage

```
share(part, whole)
```

Arguments

`part`, `whole` Counts, of equal length.

Value

Numeric vector.

Examples

```
share(census_2011_pca$literate_total, census_2011_pca$population_total) |> head()
```

theme_census *Shared theme for censusindia figures*

Description

`theme_census()` for plots with axes, `theme_census_map()` for choropleths.

Usage

```
theme_census(base_size = 11)

theme_census_map(base_size = 11)
```

Arguments

`base_size`	Base font size in points.

Value

A `ggplot2` theme.

Examples

```
ggplot2::ggplot(census_2011_pca, ggplot2::aes(literate_total, population_total)) +
  ggplot2::geom_point() +
  theme_census()
```

Index

∗ **datasets**
- census_1931_castes, 5
- census_1931_exterior_castes, 6
- census_1961_literacy, 6
- census_1971, 7
- census_1981, 8
- census_2001_pca, 10
- census_2001_pca_full, 11
- census_2011_district_languages, 13
- census_2011_linguistic_diversity, 14
- census_2011_pca, 17
- census_2011_sc_st, 18
- census_2011_tribes, 20
- census_population_time_series, 25
- census_subdistricts_2011, 27
- census_variables, 28
- india_states, 34
- population_projections_district, 39
- population_projections_district_lgd, 41
- population_projections_state, 43
- population_projections_state_age, 44

at_level, 4
at_level(), *12*, *15*, *21*
attach_geometry, 3, *37*
attach_geometry(), *10*, *39*

census_1931_castes, 5, *46*
census_1931_exterior_castes, 6
census_1961_literacy, 6
census_1971, 7
census_1981, *4*, 8
census_2001_pca, 10, *11*, *18*
census_2001_pca_full, 11
census_2011_demographics, 12
census_2011_demographics(), *4*, *18*, *27*
census_2011_district_languages, 13, *19*
census_2011_linguistic_diversity, 14, *24*
census_2011_marginal_detail, 15
census_2011_marginal_detail(), *4*, *18*
census_2011_mother_tongue, 16
census_2011_pca, *10*, *11*, 17, *27*
census_2011_sc_st, 18, *20*, *26*
census_2011_subdistrict_languages, 19
census_2011_tribes, *18*, 20, *26*
census_2011_workers, 21
census_2011_workers(), *4*, *18*
census_api *(serve)*, 45
census_base_layer, 22
census_cache_dir, 23
census_cache_dir(), *4*, *12*, *15*, *16*, *19*, *21*, *23*, *32*, *40*, *42*
census_download, 23
census_download(), *23*
census_languages, 24
census_languages(), *16*, *46*
census_population_time_series, 25, *28*
census_sc_st, 26
census_sc_st(), *18*, *46*
census_subdistricts_2011, 27
census_variables, 28
census_years, 28
census_years(), *25*
clear_census_cache *(census_cache_dir)*, 23
clear_census_cache(), *29*
clear_geometry_cache, 29
clear_geometry_cache(), *23*
compare_maps, 29

get_census, 30
get_census(), *25*, *26*, *46*
get_census_boundaries, 31
get_census_boundaries(), *46*
get_palette, 32

get_population, 33
get_population(), *46*
ggplot2::geom_sf(), *22*

india_states, 34

list_census_geographies, 35
list_census_geographies(), *46*
list_census_variables, 35
list_census_variables(), *31*, *46*
list_states, 36
list_states(), *46*

plot_map, *30*, 37
plot_map(), *22*
plotting, 36
population_projections_district, 39, *40, 41*
population_projections_district_age, 40
population_projections_district_age(), *33*, *39*, *41*
population_projections_district_lgd, 41, *42*
population_projections_district_lgd_age, 42
population_projections_district_lgd_age(), *33*, *40, 41*
population_projections_state, *33*, *41*, 43, *44*
population_projections_state_age, *33*, 44

search_census_variables, 45
search_census_variables(), *46*
serve, 45
sex_ratio, 46
share, 47

theme_census, 48
theme_census_map *(*theme_census*)*, 48
tools::R_user_dir(), *23*