Package 'censusindia'

November 5, 2026

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

Version 0.3.0

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>),
Atharva Shetwe [aut] (ORCID: <https://orcid.org/0009-0001-0409-3334>)

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

Repository CRAN

Date/Publication 2026-11-05 08:10:02 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 . . . 19
census_2011_subdistrict_languages . . . 20
census_2011_tribes . . . 21
census_2011_workers . . . 22
census_base_layer . . . 23
census_cache_dir . . . 24
census_download . . . 24
census_languages . . . 25
census_population_time_series . . . 26
census_sc_st . . . 27
census_subdistricts_2011 . . . 28
census_variables . . . 29
census_years . . . 29
clear_geometry_cache . . . 30
compare_maps . . . 30
get_census . . . 31
get_census_boundaries . . . 33
get_palette . . . 33
get_population . . . 34
india_states . . . 36
list_census_geographies . . . 36

list_census_variables . . . 37
list_states . . . 37
plotting . . . 38
plot_map . . . 38
population_projections_district . . . 40
population_projections_district_age . . . 41
population_projections_district_lgd . . . 42
population_projections_district_lgd_age . . . 43
population_projections_state . . . 44
population_projections_state_age . . . 46
search_census_variables . . . 47
serve . . . 47
sex_ratio . . . 48
share . . . 49
theme_census . . . 50

Index **51**

`attach_geometry` *Attach geographic boundaries to census data*

Description

Joins state or district boundary polygons to a census data frame for mapping. When `geography` is `NULL`, the geographic level is inferred from the columns of `data`.

Usage

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

Arguments

`data`	A data frame of census data with state names and, for districts, district names. `state_name_harmonized` and `district` join exactly; other name columns fall back to fuzzy matching.
`year`	Census year of the boundaries. Available: 1941, 1951, 1961, 1971, 1981, 1991, 2001, 2011.
`geography`	Geographic level: `"state"` or `"district"`. If `NULL` (default), inferred from the columns of `data`.
`unmatched`	Action when rows get no geometry or when two distinct units claim the same polygon: `"warn"` (default), `"error"`, or `"ignore"`. Because fuzzy matching uses substrings and edit distance as a fallback, a district whose true polygon is missing can otherwise be assigned a neighbor's shape silently.

Value

An sf object with geometry attached.

Boundary files are downloaded on first use and cached locally; see census_cache_dir().

Examples

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

---

at_level *Filter a stacked census table to one geographic level and sector*

Description

The village-level 2011 tables (census_2011_demographics(), census_2011_workers(), census_2011_marginal_detail and census_1981 stack multiple geographic levels (from India down to village or ward) and three sectors ("total", "rural", "urban") in the same table. Summing rows without filtering both geography and sector counts people multiple times. at_level() filters geography and, when a sector column is present, sector.

Usage

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

Arguments

data	A data frame with a geography column (e.g., census_1981 or one of the village-level 2011 tables).
level	Geographic level, such as "district" or "state". Accepts "sub-district" as a synonym for "subdistrict".
sector	"total" (default), "rural", or "urban".

Details

The levels "india", "state", "district", "subdistrict", and "village" are additively exact: each level sums to its parent, and rural + urban == total.

The levels "town" and "ward" are **not** additive: the Census lists a municipality and the same municipality with its outgrowths as separate town rows (for example, BBMP (M Corp.) and BBMP (M Corp. + OG) in Bengaluru), and ward rows do not always sum to the parent town row. Use town and ward rows for lookups, not aggregation.

Value

data filtered to level and sector.

Examples

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

---

census_1931_castes *Census 1931 castes by religion*

Description

Population of each caste and community in the caste tables of the 1931 Census, by religion. The table has 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, with a standard name

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

caste Caste or community from the source

rank Position in the ranked caste list of the province. 1 is the largest.

hindu,muslim,sikh,jain,buddhist,christian,tribal,parsi,jew Persons of the caste in each religion. `tribal` is the 1931 "Tribal" or "Primitive" religion.

total Sum of the religion columns

total_ranked_list Total in the ranked list of the province. NA for the Andaman & Nicobar Islands.

notes Note from the compiler, for example where `total` and `total_ranked_list` are different

Details

A religion column is NA if the table of the province has no column for that religion. The value is 0 if the column is present and has no persons. Ajmer-Merwara has only totals. Its religion columns and `total` are NA. Remainder entries, for example "Caste Unspecified" and "Indian Christian", are separate rows.

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 that the 1931 Census put in the "exterior castes" group, with their population. Each row comes from a province or state section of Appendix I of the Census of India 1931, Volume I. "Exterior castes" is the term of the report.

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 from the source, with qualifiers in brackets

population Persons in the exterior castes group

note Note from the compiler

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

State and district literacy rates from the 1961 Census. The literacy columns are percentages of the population. The table has no counts. You cannot add the rates across areas.

Usage

```
census_1961_literacy
```

Format

A tibble with 347 rows (26 state, 321 district) and 10 columns:

year Census year (1961)

geography Geographic level: "state" or "district"

state_code Integer 2011 census code of the harmonized state

state_name_harmonized Harmonized state name, used to join datasets

state_name State name from the source

district_code Integer district code from the source. NA on state rows.

district Name of the district. NA on state rows.

literacy_rate Literacy rate, persons (percent)

literacy_rate_male Male literacy rate (percent)

literacy_rate_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 and urban values and the SC/ST population.

Usage

```
census_1971
```

Format

A tibble with 1,140 rows (380 areas x 3 sectors) and 13 columns:

year Census year (1971)

geography Geographic level: "state" or "district"

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

state_code Integer 2011 census code of the harmonized state

state_name_harmonized Harmonized state name, used to join datasets

state_name State name from the source

district Name of the district. NA on state rows.

area_km2 Area in square kilometers

population Population

males Male population

females Female population

sc_population Scheduled Caste population

st_population Scheduled Tribe population

Details

The table stores each sector as a row. Each of the 28 states and 352 districts has one row each for `"total"`, `"rural"`, and `"urban"`. Filter to `sector == "total"` before summing across rows; otherwise the sum double-counts every person.

Verified against the printed volume (Table II). Ten missing Mysore districts were added and OCR errors in SC/ST cells were corrected. Madras is entirely urban (rural = 0).

inst/ocr-corrections.csv records every correction and its source page. Within each column and sector, districts sum to their state and rural + urban equals total. Nil entries (`". ."`) are stored as `NA`.

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 categories.

Usage

```
census_1981
```

Format

A tibble with 2,365 rows and 38 columns:

year Census year (1981)

geography Geographic level: "india", "state", "district", "town" or "urban_agglomeration"

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

state_code Integer 2011 census code of the harmonized state. NA on the India rows.

state_name_harmonized Harmonized state name, used to join datasets

state_name State name from the source

district Name of the district. NA above district level and on urban agglomerations.

name Name of the unit from the source. Urban agglomerations have only this name.

area_km2 Area in square kilometers

occupied_houses Number of occupied residential houses

households Number of households

population, males, females Population: persons, males and females

literate, literate_male, literate_female Literate population

main_workers, main_workers_male, main_workers_female Main workers

cultivators, cultivators_male, cultivators_female Main workers who are cultivators

agri_labour, agri_labour_male, agri_labour_female Main workers who are agricultural laborers

hh_industry, hh_industry_male, hh_industry_female Main workers in household industry

other_workers, other_workers_male, other_workers_female Other main workers

marginal_workers, marginal_workers_male, marginal_workers_female Marginal workers

non_workers, non_workers_male, non_workers_female Non-workers

Details

The table stores geographic levels and sectors as rows. `geography` is `"india"`, `"state"`, `"district"`, `"town"`, or `"urban_agglomeration"`, with one row per sector. Call `at_level()` (or filter `geography` and `sector == "total"`) before summing across rows. Each of the 398 districts has a total, rural, and urban row; towns and urban agglomerations have only an urban row.

Verified against the printed volume (`India_1981_PCA_General.pdf`). In every state and district, persons = males + females and total = rural + urban for population, literacy, and each worker category, and districts sum to their state. OCR errors, leading digits clipped in the scan, and printing errors in the source volume were corrected.

inst/ocr-corrections.csv records every correction with its page and evidence. The urban row of Guna, printed in the volume but omitted from the digitized source, was restored. Assam held no census in 1981 and has no rows.

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 data from the 2001 Census. The table has 593 districts. `population` adds 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 Integer state code

state_name_harmonized Harmonized state name, used to join datasets

state_name State name from the source

district_code Integer district code. Each state has its own numbers. One code can occur in more than one state.

district Name of the district

households Number of households

population Total population

males Male population

females Female population

population_0_6 Population of age 0-6 years

sc_population Scheduled Caste population

st_population Scheduled Tribe population

literate Literate population

workers Total workers

main_workers Main workers

marginal_workers Marginal workers

non_workers Non-workers

Details

Use the 2001 boundaries with `attach_geometry()`. The 2011 boundary file has no polygon for some 2001 districts and the wrong polygon for others, for example `North Cachar Hills` and the undivided `Medinipur`. For 2011, refer to census_2011_pca.

Source

Census of India 2001, Primary Census Abstract.

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

Description

The full 2001 PCA at district level. Each indicator has values for persons, males and females. Each district has one row each 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. Male and female columns have a `_male` or `_female` suffix. A column with no suffix gives persons:

year Census year (2001)

geography Geographic level ("district")

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

state_code Integer state code

state_name_harmonized Harmonized state name, used to join datasets

state_name State name from the source

district_code Integer district code. Each state has its own numbers.

district Name of the district

households Number of households

population, males, females Population

population_0_6, population_0_6_male, population_0_6_female Population of age 0-6 years

sc_population, sc_population_male, sc_population_female Scheduled Caste population

st_population, st_population_male, st_population_female Scheduled Tribe population

literate, literate_male, literate_female Literate population

illiterate, illiterate_male, illiterate_female Illiterate population

workers, workers_male, workers_female All workers

main_workers, main_workers_male, main_workers_female Main workers

main_cultivators, main_agri_labour, main_hh_industry, main_other Main workers by category, persons

marginal_workers, marginal_workers_male, marginal_workers_female Marginal workers

marginal_cultivators, marginal_agri_labour, marginal_hh_industry, marginal_other Marginal workers by category, persons

non_workers, non_workers_male, non_workers_female Non-workers

Details

census_2001_pca is a 19-column subset of the same data. It has the columns that are also in census_2011_pca.

The table gives each sector as a row. Filter `sector == "total"` before you add rows.

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. The table has all geographic levels (India, state, district, subdistrict, town/village, ward) and the total, rural and urban sectors. The package downloads the table on first use (refer to `census_cache_dir()`). Use `at_level()` before you add rows.

Usage

```
census_2011_demographics()
```

Value

A tibble with 751,594 rows and 29 columns. Male and female columns have a `_male` or `_female` suffix. A column without a suffix gives persons:

year Census year (2011)

geography Geographic level: "india", "state", "district", "subdistrict", "town", "village" or "ward"

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

state_code,district_code,subdistrict_code,town_village_code,ward_code Integer geographic codes. NA above the level of the unit

state_name_harmonized Harmonized state name. Use it to join datasets

name Name of the geographic unit

households Number of households

population,males,females Total, male, and female population

population_0_6,population_0_6_male,population_0_6_female Population of age 0-6 years

sc_population,sc_population_male,sc_population_female Scheduled Caste population

st_population,st_population_male,st_population_female Scheduled Tribe population

literate,literate_male,literate_female Literate population

illiterate,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. The table has values by sex and by sector.

Usage

```
census_2011_district_languages
```

Format

A tibble with 116,979 rows and 12 columns:

year Census year (2011)

geography Geographic level ("district")

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

state_code Integer state code

state_name_harmonized Harmonized state name, used to join datasets

district_code Integer district code

district Name of the district

language_name Name of the language

language_group Integer language group code

speakers Speakers

speakers_male Male speakers. NA on rural and urban rows.

speakers_female Female speakers. NA on rural and urban rows.

Details

The table stores each sector as a row. Each district and language has one row each for `"total"`, `"rural"`, and `"urban"` (38,993 rows per sector). Filter `sector == "total"` before summing across rows, or use `census_languages()`, which applies this filter automatically. The census reports speakers by sex only for the total sector; `speakers_male` and `speakers_female` are `NA` on rural and urban rows.

Source

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

`census_2011_linguistic_diversity`
Census 2011 linguistic diversity

Description

District-level metrics of linguistic diversity, calculated from the 2011 Census C-16 mother tongue tables. The metrics use only named mother tongues. C-16 puts the unscheduled languages of the northeast in "Others" rows for each language group. These rows have no language name, and the metrics do not include them. `coverage` gives the fraction of the district population that the metrics use.

Usage

```
census_2011_linguistic_diversity
```

Format

A tibble with 640 rows and 13 columns:

year Census year (2011)

geography Geographic level ("district")

state_code Integer state code

state_name_harmonized Harmonized state name, used to join datasets

district_code Integer district code

district Name of the district

n_languages Number of different languages

speakers Speakers of all named languages

shannon_entropy Shannon entropy of the language distribution

effective_languages Effective number of languages (2^shannon_entropy)

dominant_language Language with the most speakers

dominant_share Fraction of the speakers that speak `dominant_language`

coverage Fraction of the district population with a named mother tongue

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, divided by the duration of work (3-6 months and 0-3 months) and by activity. The table also gives non-workers. It has all geographic levels and the total, rural and urban sectors. The package downloads the table on first use. Refer to `census_cache_dir()`. Use `at_level()` before you add rows.

Usage

```
census_2011_marginal_detail()
```

Value

A tibble with 751,594 rows and 43 columns. Male and female columns have a `_male` or `_female` suffix. A column without a suffix gives persons:

year Census year (2011)

geography Geographic level: "india", "state", "district", "subdistrict", "town", "village" or "ward"

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

state_code,district_code,subdistrict_code,town_village_code,ward_code Integer geographic codes. NA above the level of the unit

state_name_harmonized Harmonized state name. Use it to join datasets

name Name of the geographic unit

marginal_workers_3_6,marginal_workers_3_6_male,marginal_workers_3_6_female Marginal workers with 3-6 months of work

marginal_cultivators_3_6,marginal_cultivators_3_6_male,marginal_cultivators_3_6_female Marginal cultivators, 3-6 months

marginal_agri_labour_3_6,marginal_agri_labour_3_6_male,marginal_agri_labour_3_6_female Marginal agricultural labourers, 3-6 months

marginal_hh_industry_3_6,marginal_hh_industry_3_6_male,marginal_hh_industry_3_6_female Marginal household industry workers, 3-6 months

marginal_other_3_6,marginal_other_3_6_male,marginal_other_3_6_female Marginal other workers, 3-6 months

marginal_workers_0_3,marginal_workers_0_3_male,marginal_workers_0_3_female Marginal workers with 0-3 months of work

marginal_cultivators_0_3,marginal_cultivators_0_3_male,marginal_cultivators_0_3_female Marginal cultivators, 0-3 months

marginal_agri_labour_0_3,marginal_agri_labour_0_3_male,marginal_agri_labour_0_3_female Marginal agricultural labourers, 0-3 months

marginal_hh_industry_0_3,marginal_hh_industry_0_3_male,marginal_hh_industry_0_3_female Marginal household industry workers, 0-3 months

marginal_other_0_3,marginal_other_0_3_male,marginal_other_0_3_female Marginal other workers, 0-3 months

non_workers,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 from the 2011 Census C-16 tables, at state, district and subdistrict level. The counts are by sector and by sex. The package downloads the table on first use. Refer to `census_cache_dir()`. `census_languages()` gives the same counts for one level at a time, without the "Others" residual rows.

Usage

```
census_2011_mother_tongue()
```

Details

The source file has no subdistrict code. A district and its subdistricts have the same `district_code`. The build sets `geography` from the row order: the rows of the district come first.

The district rows, the subdistrict rows and the census_2011_pca population of a district have the same total. Filter to one `geography` before you add rows. The table also stores sector as rows, so filter to `sector == "total"` too.

In Shajapur district (Madhya Pradesh), five tehsils (Agar, Shajapur, Gulana, Moman Badodiya, Shujalpur) have only one Hindi row each. The L1 row of Kalapipal holds the Hindi speakers of all six tehsils. The district rows are complete.

Value

A tibble with 1,050,471 rows (350,157 for each sector) and 16 columns:

year Census year (2011)

geography "state", "district" or "subdistrict"

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

state_code Integer state code

state_name_harmonized Harmonized state name. Use it to join datasets

state_name State name from the source

district_code Integer district code. NA on state rows

district District name as in census_2011_pca. NA on state rows

subdistrict Subdistrict name from the source. NA above subdistrict level

language_code Census language code

language_name Name of the language or dialect

language_level L1 for main languages, L2 for dialects

language_group Integer language group code

speakers Speakers

speakers_male Male speakers

speakers_female 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 data from the 2011 Census. The table has all 640 districts. `population` adds 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 Integer state code

state_name_harmonized Harmonized state name, used to join datasets

state_name State name from the source

district_code Integer district code, unique in India (1-640)

district Name of the district

households Number of households

population Total population

males Male population

females Female population

population_0_6 Population of age 0-6 years

sc_population Scheduled Caste population

st_population Scheduled Tribe population

literate Literate population

workers Total workers

main_workers Main workers

marginal_workers Marginal workers

non_workers Non-workers

Details

The values are district sums of the village-level 2011 PCA. The same source gives `census_2011_demographics()`, `census_2011_workers()` and `census_2011_marginal_detail()`. For 2001, refer to 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 12 columns:

year Census year (2011)

geography Geographic level ("district")

state_code Integer state code

state_name_harmonized Harmonized state name, used to join datasets

district District name, as census_2011_pca writes it

category `"Scheduled Caste"` or `"Scheduled Tribe"` (full name)

category_code `"SC"` or `"ST"` (short form)

caste_tribe_name Name of the caste or tribe

parent_tribe The tribe that contains the tribe of this row ("Naga" for the (ST) rows of Nagaland). NA on other rows.

population Population of the caste or tribe

percent `population / district_total_population * 100`, rounded to two decimal places

district_total_population District population: the `population` column of census_2011_pca for the same district

Details

The table gives each Scheduled Tribe by name: 581 names. The table does not give Scheduled Castes by name. All 582 SC rows have the label `"All Scheduled Castes"`, one row for each district.

You can filter on two columns. `category` has the full name, and `category_code` has `SC` or `ST`. `census_sc_st()` accepts the two forms.

The ST rows include `"Generic Tribes etc."`. These rows give the persons who did not give a specific tribe: 585 rows, 2,804,148 persons. census_2011_tribes contains the ST rows of this table, without these rows and with harmonized names.

For Nagaland, the census tabulates the Scheduled Tribe `"Naga"` alongside its constituent tribes marked `"(ST)"` (`Angami(ST)`, `Ao(ST)`, ...). Those 193 rows carry `parent_tribe = "Naga"`, and in each district the `"Naga"` row equals their sum. To count each person once, sum only rows where `is.na(parent_tribe)`; the ST rows then match `census_2011_pca$st_population` in every district (104,545,716 in total).

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, by sector and by sex. The table does not include the district rows that have the same `district_code` as the subdistricts. The rows of a district add to its row in census_2011_district_languages. The package downloads the table on first use. Refer to `census_cache_dir()`.

Usage

```
census_2011_subdistrict_languages()
```

Details

The table stores sector as rows (144,134 rows for each sector). Filter to `sector == "total"` before you add rows, or use `census_languages()`, which applies this filter. `speakers_male` and `speakers_female` are NA on rural and urban rows.

Value

A tibble with 432,402 rows and 12 columns:

year Census year (2011)

geography Geographic level ("subdistrict")

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

state_code Integer state code

state_name_harmonized Harmonized state name. Use it to join datasets

district_code Integer code of the parent district

subdistrict Name of the subdistrict

language_name Name of the language

language_group Integer language group code

speakers Speakers

speakers_male Male speakers. NA on rural and urban rows

speakers_female Female speakers. NA on rural and urban rows

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 each Scheduled Tribe, from the 2011 Census A-11 tables. `tribe_name` is the harmonized name. `tribe_name_original` is the text from the source. This text includes the full list of synonyms from the census ("Gond, Arakh, Arrakh, Agaria, ...") and any district qualifier in parentheses.

Usage

```
census_2011_tribes
```

Format

A tibble with 12,861 rows and 11 columns:

year Census year (2011)

geography Geographic level ("district")

state_code Integer state code

state_name_harmonized Harmonized state name, used to join datasets

district District name, as census_2011_pca writes it

tribe_name Harmonized tribe name

tribe_name_original Tribe name from the source

parent_tribe The tribe that contains the tribe of this row ("Naga" for the (ST) rows of Nagaland). NA on other rows.

population Tribe population

percent `population / district_total_population * 100`, rounded to two decimal places

district_total_population District population: the `population` column of census_2011_pca for the same district

Details

To harmonize a name, the package keeps the first synonym and removes the qualifier: `"Thoti (in Adilabad, Hyderabad, ...)"` becomes `"Thoti"`, and `"Angami(ST)"` becomes `"Angami"`. The 581 source names become 481 tribe names. Different states write the same tribe in different ways, and some names differ only in spaces.

This table contains the ST rows of census_2011_sc_st excluding `"Generic Tribes etc."`. The constituent tribes of Nagaland carry `parent_tribe = "Naga"`, as in census_2011_sc_st. Before summing within a district, exclude either those rows or the `"Naga"` umbrella rows so each Naga person is counted only once. The `"Naga (ST)"` row (Naga with no specific tribe) is harmonized to `"Naga (unspecified)"`.

Source

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

`census_2011_workers` *Census 2011 workers (PCA)*

Description

Workers by type from the 2011 Primary Census Abstract. The table has all geographic levels and the total, rural and urban sectors. It divides main and marginal workers by activity: cultivators, agricultural labourers, household industry and other workers. The package downloads the table on first use (refer to `census_cache_dir()`). Use `at_level()` before you add rows.

Usage

```
census_2011_workers()
```

Value

A tibble with 751,594 rows and 43 columns. Male and female columns have a `_male` or `_female` suffix. A column without a suffix gives persons:

year Census year (2011)

geography Geographic level: "india", "state", "district", "subdistrict", "town", "village" or "ward"

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

state_code,district_code,subdistrict_code,town_village_code,ward_code Integer geographic codes. NA above the level of the unit

state_name_harmonized Harmonized state name. Use it to join datasets

name Name of the geographic unit

workers,workers_male,workers_female All workers

main_workers,main_workers_male,main_workers_female Main workers

main_cultivators,main_cultivators_male,main_cultivators_female Main cultivators

main_agri_labour,main_agri_labour_male,main_agri_labour_female Main agricultural labourers

main_hh_industry,main_hh_industry_male,main_hh_industry_female Main household industry workers

main_other,main_other_male,main_other_female Main other workers

marginal_workers,marginal_workers_male,marginal_workers_female Marginal workers

marginal_cultivators,marginal_cultivators_male,marginal_cultivators_female Marginal cultivators

marginal_agri_labour,marginal_agri_labour_male,marginal_agri_labour_female Marginal agricultural labourers

marginal_hh_industry,marginal_hh_industry_male,marginal_hh_industry_female Marginal household industry workers

marginal_other,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 with the full claimed extent of India*

Description

A ggplot2 layer of the state polygons for a census year, with a flat fill. Put it below a district choropleth to keep the national outline complete. The census has no districts in the territory that India claims but does not administer.

Usage

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

Arguments

`year`	Census year of the boundaries. One of 1941, 1951, 1961, 1971, 1981, 1991, 2001, 2011.
`fill`	Fill color for the base layer. The default is the grey that `plot_map()` uses for missing values.
`...`	Arguments for `ggplot2::geom_sf()`.

Details

`plot_map()` adds this layer automatically. Use it directly when you make 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 boundary files and the seven largest tables are downloaded on first use, verified against pinned MD5 checksums, and cached on disk across R sessions.

Usage

```
census_cache_dir()

clear_census_cache()
```

Details

By default the cache lives in `tools::R_user_dir()`. Set `options(censusindia.cache_dir = ...)` to use a different directory.

Value

`census_cache_dir()` returns the cache directory path (which may not yet exist on disk). `clear_census_cache()` deletes the cache directory, clears the in-memory session cache, and returns the path invisibly.

See Also

`census_download()` downloads all remote files upfront; `clear_geometry_cache()` evicts objects cached in session memory.

Examples

```
census_cache_dir()
```

census_download *Download all remote data*

Description

Downloads every boundary file and remote table not yet present in `census_cache_dir()`. Once complete, all package functions work offline.

Usage

```
census_download()
```

Value

The cache directory path, invisibly.

Examples

```
census_download()
```

`census_languages` *Language speakers at a chosen geographic level*

Description

The C-16 mother tongue table (`census_2011_mother_tongue()`) stacks state, district, and subdistrict rows in a single table where subdistricts share their parent district's `district_code`. Summing by `district_code` without filtering `geography` double-counts every speaker. `census_languages()` returns a single geographic level and sector at a time, excluding the `"Others"` residual rows.

Usage

```
census_languages(
  level = c("district", "subdistrict"),
  language_group = NULL,
  sector = c("total", "rural", "urban"),
  state = NULL
)
```

Arguments

`level`	`"district"` (38,993 rows per sector across 640 districts) or `"subdistrict"` (144,134 rows per sector across 5,988 subdistricts).
`language_group`	Optional integer vector of `language_group` codes to keep, or `NULL` (default) for all groups.
`sector`	`"total"` (default), `"rural"`, or `"urban"`. Because sector is stored as rows, summing without filtering `sector` double-counts every speaker.
`state`	Optional state filter: one or more canonical state names or two-letter codes, as in `get_census()`.

Value

A tibble for the requested level and sector, with one row per language per area. Columns match census_2011_district_languages or `census_2011_subdistrict_languages()`.

See Also

census_2011_linguistic_diversity for district-level diversity metrics (Shannon entropy, effective number of languages).

Examples

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

`census_population_time_series`
Census population time series (1901-2011)

Description

Population at state and district level for each census from 1901 to 2011.

Usage

```
census_population_time_series
```

Format

A tibble with 7,901 rows and 12 columns:

year Census year (integer: 1901, 1911, ..., 2011). The column also has years that are not census years. Refer to the section below.

geography Geographic level: "state" or "district"

state_code Integer 2011 census code of the harmonized state

state_name_harmonized Harmonized state name, used to join datasets

state_name State name from the source

district_code Integer district code. NA on state rows.

district Name of the district. NA on state rows.

population Total population

males Male population

females Female population

variation_absolute Absolute change from the previous census

variation_percent Change from the previous census, in percent

Coverage limits

Sikkim, Mizoram and Daman & Diu are not in the table for any year. For 2001 and 2011, `get_census()` uses the Primary Census Abstract.

The district rows do not add to their state totals before 1961 and in 1981. In 1901-1941, 57 districts have no population. In 1981, all 27 Assam districts are missing, but the state row is present.

The `year` column also contains 1900, 1910, 1940, 1948, 1950, 1960, 1962 and 2021. These are not census years. Use `census_years()` to remove them before you use `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*

Description

census_2011_sc_st stores full names in `category` (`"Scheduled Caste"`, `"Scheduled Tribe"`) and short codes in `category_code` (`"SC"`, `"ST"`), so `dplyr::filter(category == "SC")` silently returns zero rows. `census_sc_st()` accepts either form, case-insensitively.

Usage

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

Arguments

`category`	`"SC"`, `"ST"`, `"Scheduled Caste"`, `"Scheduled Tribe"`, or `NULL` (default) to keep both categories. Case-insensitive; accepts a vector.
`state`	Optional state filter: one or more canonical state names or two-letter codes, as in `get_census()`.

Details

Individual Scheduled Tribes are listed by name (581 tribes). Scheduled Castes are reported only as a single district aggregate (`"All Scheduled Castes"`, 582 districts).

Value

A tibble of census_2011_sc_st rows.

See Also

census_2011_tribes for Scheduled Tribe rows with harmonized names and without the `"Generic Tribes etc."` residual.

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 16 columns:

year Census year of the population data (2011)

geography Geographic level ("subdistrict")

state_code Integer 2011 census state code

state_name_harmonized Harmonized state name, used to join datasets

state_name State name from the source

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 Persons per square kilometer

Details

If a subdistrict is in two or more districts, the table has one row for each district, with the mark "(Part)". Population, males, females and households come from `census_2011_demographics()` and add to census_2011_pca. Villages, towns, area and density come from the Dataverse source. They are NA where the source 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

Variables resolved by `get_census()`, with one row per variable and geographic level. `years` lists the census years in which the variable is available at that level. Generated by calling `get_census()` across every combination in `list_census_geographies()`.

Usage

```
census_variables
```

Format

A tibble with 50 rows and 5 columns:

variable Variable name in the package functions

label Label for display

geography Geographic level: `"state"`, `"district"` or `"subdistrict"`

years Comma-separated census years for that geography

category Variable category (population, literacy, workers and others)

census_years *Keep only decennial census years*

Description

census_population_time_series includes non-decennial years (1900, 1910, 1940, 1948, 1950, 1960, 1962, and 2021) alongside the twelve decennial censuses. Computing `lag()` on the unfiltered table silently produces 1-year or 13-year intervals instead of 10-year intervals.

Usage

```
census_years(data)
```

Arguments

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

Value

`data` restricted to the twelve decennial census years (1901 to 2011).

Examples

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

`clear_geometry_cache` *Clear geometry cache*

Description

Evicts cached boundary geometries and downloaded tables from session memory.

Usage

```
clear_geometry_cache()
```

Value

Invisible `NULL`.

See Also

`clear_census_cache()` deletes the 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,
  limits = NULL,
  show_state_boundaries = TRUE,
  ...
)
```

Arguments

data_list	A named list of sf data frames, keyed by year or panel label.
fill_var	Variable to map (unquoted column name or string).
title	Title for the combined plot.
ncol	Number of columns in the layout.
palette	Color palette; see plot_map for options.
common_scale	Logical. If TRUE (default), all panels share the same color scale limits.
limits	Optional numeric vector c(min, max) for the shared color scale. Overrides the range computed when common_scale = TRUE.
show_state_boundaries	Logical. If TRUE (default), overlays state boundaries on each panel.
...	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", sector == "total") |>
  mutate(st_pct = 100 * st_population / population) |>
  attach_geometry(1971)

st_2011 <- census_2011_pca |>
  mutate(st_pct = 100 * st_population / population) |>
  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

Queries Census of India data 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 every census year; see list_census_geographies().
variables	Character vector of variable names to return, or NULL (default) for all variables available at that year and geography. See list_census_variables().
state	Optional state filter: one or more canonical state names or two-letter codes (e.g., "Maharashtra" or "MH").
geometry	If TRUE, attaches census boundaries and returns an sf object. Available for "state" and "district".
sector	Sector filter: "total" (default), "rural", or "urban". Rural and urban splits are available in get_census() for 1971 and 1981 only; passing "rural" or "urban" for other years raises an error.

Value

A tibble (or an sf object when geometry = TRUE). Every year begins with the same identifier columns: year, geography, sector, state_code, state_name_harmonized, state_name, followed by district_code, district, and subdistrict where applicable. State frames carry no district columns. sector is "total" for years without a rural/urban split. For 1961, only literacy rates (literacy_rate, literacy_rate_male, literacy_rate_female) are available, not population counts.

Examples

```
get_census(2011, "state")
get_census(1971, "district", state = "Maharashtra")
get_census(1971, "state", sector = "rural")
get_census(1961, "state", variables = "literacy_rate")

get_census(1981, "district", variables = c("population", "literacy_rate"), geometry = TRUE)
```

get_census_boundaries *Get census boundaries*

Description

Retrieve state or district boundary polygons for a 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 with columns year, geography, state_code, state_name_harmonized, state_name, district (district level only), id, and geometry.

Boundary files are downloaded on first use and cached locally; see census_cache_dir().

Examples

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

get_palette *Get color palette*

Description

Gives a color palette for maps. The palettes are "viridis", "red_blue", "blue_red", "greens", "oranges", "reds", "blues" and "purple_green".

Usage

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

Arguments

name	Name of the palette
reverse	Logical. If TRUE, the direction is reversed.

Value

A character vector of colors

Examples

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

`get_population` *Get population projections*

Description

Queries Ministry of Health and Family Welfare (MoHFW) population projections (2011-2036, Census 2011 base) at state or district level. District projections are available on two boundary sets: Census 2011 districts (640 districts at 5-year intervals) and Local Government Directory (LGD) districts (783 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, or `NULL` (default) for all published years in the table.
`geography`	Geographic level: `"state"` or `"district"`.
`state`	Optional state filter: one or more canonical state names or two-letter codes (e.g., `"Maharashtra"` or `"MH"`).
`boundary`	District boundary system: `"census2011"` (default, 640 districts) or `"lgd"` (783 districts). Ignored when `geography = "state"`.
`geometry`	If `TRUE`, attaches Census 2011 boundaries and returns an `sf` object. At district level, supported only when `boundary = "census2011"`.
`by_age`	If `TRUE`, returns age-stratified rows. State projections use 5-year age bands from `"0-4"` to `"80+"`; district projections use single years for ages 0-14 and 5-year bands from `"15-19"` to `"80+"`. See population_projections_state_age, `population_projections_district_age()`, and `population_projections_district_lgd_age()`. State age projections cover 23 major states, `"India"`, and `"North East States Excluding Assam"`; requesting other states with `by_age = TRUE` raises an error.

Details

State projections cover 2011-2036 across 39 entries: 28 states, the 9 union territories printed by MoHFW (including separate `"Dadra & Nagar Haveli"` and `"Daman & Diu"` rows), the merged `"Dadra & Nagar Haveli and Daman & Diu"` UT (the sum of the two), and `"India"`. To sum across states without double-counting, exclude `"India"` and either the merged UT or the two constituent UTs; see population_projections_state.

Because the state report rounds Persons, Male, and Female independently to the nearest thousand, `population` and `males + females` can differ by up to 1,000 at state level (2,000 for the summed UT).

Census 2011 district projections are published at 5-year intervals (2011, 2016, 2021, 2026, 2031); LGD district projections are annual from 2012 to 2031.

When a requested `year` is between published table years, `get_population()` interpolates `males`, `females`, and `population` independently using geometric (exponential) growth between the two nearest published years:

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

Because each series is interpolated independently, summed district estimates need not equal the interpolated state estimate.

Years after the last published year are extrapolated using the final interval's growth rate for up to five years (to 2041 for states, 2036 for districts). Years before the start of the table or more than five years past the end raise an error.

Telangana and Ladakh are reported as separate entries even though both were created after the 2011 Census.

Value

A tibble (or an `sf` object if `geometry = TRUE`) with columns `year`, `geography`, `state_code`, `state_name_harmonized`, `district` (district level only), `age_group` (when `by_age = TRUE`), `males`, `females`, `population`, and `estimate`. `geography` is `"india"` on national rows. All counts are in persons. `estimate` is `"published"` for years printed in the MoHFW/IIPS tables, `"interpolated"` for years between printed years, and `"extrapolated"` for years after the last printed year.

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 Integer state code (Census 2011 codes)

state_name Official state name

state_abbr Two-letter state abbreviation

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

state_name_harmonized Harmonized state name, used to join datasets

list_census_geographies *List available census geographies*

Description

Geographic levels available for each census year.

Usage

```
list_census_geographies()
```

Value

A tibble of available year and 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	Census year(s) to filter by.
geography	"state", "district", or "subdistrict".
category	"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	Region(s) to filter by: "North", "South", "East", "West", "Central", "Northeast", or "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 of 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 a data frame with geometry from attach_geometry.
fill_var	Variable for the fill color (unquoted column name or string).
title	Optional title for the map.
subtitle	Optional subtitle for the map.
legend_title	Title for the legend.
palette	Color palette. Available options: • "viridis" - Perceptually uniform and colorblind-safe (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. If TRUE, reverses the palette direction. Default FALSE.
direction	Numeric. Alternative to reverse: -1 reverses the palette and 1 keeps the default. Takes precedence over reverse if both are set.
na_color	Color for missing values. Default "grey80".
boundary_year	Census year of the state outline drawn beneath the data. This background layer preserves India's full claimed northern extent where district polygons are absent. If NULL, inferred from data.
base_layer	Logical. If TRUE (default), draws the full state outline of India beneath the data.
show_state_boundaries	Logical. If TRUE, overlays state boundary lines on top of the data.
state_boundary_color	Color for state boundary lines. Default "black".
state_boundary_width	Width of state boundary lines. Default 0.3.
trans	Transformation for the color scale (e.g., "log10", "sqrt"). Default "identity".
midpoint	Center value for a diverging palette (e.g., 0 for growth or 1000 for sex ratio). If NULL, defaults to the midpoint of limits.
oob	Out-of-bounds handler for values outside limits. By default ggplot2 maps out-of-bounds values to na_color, making them indistinguishable from missing data; pass scales::squish to clamp them to the nearest limit instead.
limits	Optional numeric vector c(min, max) for the color scale.

breaks	Optional breaks for the legend.
labels	Optional labels for the legend breaks.
font_family	Font family for all map text (e.g., "Arial"). Default NULL uses the system font.
legend_position	One of "right" (default), "left", "bottom", "top", or "none". "bottom" and "top" orient the colorbar horizontally.

Value

A ggplot2 object.

Examples

```
# Calculate ST percentage and attach geometry
st_data <- census_2011_pca |>
  dplyr::mutate(st_pct = 100 * st_population / population) |>
  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 for the 640 districts of the 2011 Census boundaries. The table gives projections for 2011, 2016, 2021, 2026 and 2031.

Usage

```
population_projections_district
```

Format

A tibble with 3,200 rows and 8 columns:

year Projection year (5-year intervals)

geography Geographic level ("district")

state_code Integer 2011 census state code

state_name_harmonized Harmonized state name, used to join datasets

district District name (MOHFW names)

males Projected male population

females Projected female population

population Projected total population (males + females)

Details

Values are the same as Table 8 of the IIPS report, with one exception: Uttar Pradesh 2016. In Table 8, the 2016 column for Uttar Pradesh gives the 2017 value of each district. The Uttar Pradesh rows for 2016 use the 2016 sums of the age tables. `data-raw/projection_consistency_fixes.csv` gives these rows.

The age tables (`population_projections_district_age()`) add to these totals within rounding, except for Nagaland 2016.

Use `attach_geometry()` with the 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

The age-group detail of population_projections_district, on Census 2011 district boundaries. No LGD area apportionment applies. The age groups are the same as in `population_projections_district_lgd_age()`: single years 0-14, then 5-year bands to 80+. The package downloads the table on first use. Refer to `census_cache_dir()`.

Usage

```
population_projections_district_age()
```

Details

The values are from the printed IIPS report. The district names agree with population_projections_district.

The PDF does not print two blocks, so the package derives them. Moga (Punjab) and Tapi (Gujarat) have no 80+ row for 2022-2026. Their 80+ value is the printed All ages minus the other printed age groups. The 2017-2031 pages of Theni (Tamil Nadu) repeat the numbers of Madurai.

The printed age tables for Karnataka are a different projection from Table 8 of the same report. The package keeps their age-sex shares and scales them to Table 8. Between the five-year points of Table 8, it interpolates geometrically. The age groups of each district add to population_projections_district. The LGD tables come from these values.

Value

A tibble with 371,200 rows and 9 columns:

year Projection year (annual, 2012-2031)

geography Geographic level ("district")

state_code Integer 2011 census state code

state_name_harmonized Harmonized state name. Use it to join datasets

district District name as in Census 2011

age_group Single years "0".."14", then 5-year bands "15-19" to "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 the current Local Government Directory (LGD) district boundaries. The table gives one projection for each year from 2012 to 2031. Each value is the sum of the age groups in `population_projections_district_lgd_age()`.

Usage

```
population_projections_district_lgd
```

Format

A tibble with 15,660 rows and 8 columns:

year Projection year (each year, 2012-2031)

geography Geographic level ("district")

state_code Integer 2011 census state code

state_name_harmonized Harmonized state name, used to join datasets. Dadra & Nagar Haveli and Daman & Diu are one UT, as merged in 2020. The LGD source has no district-level data to divide them.

district District name (LGD names)

males Projected male population

females Projected female population

population Projected total population (males + females)

Known defects

The values come from `population_projections_district_age()`, scaled by area overlap. The district sums are not exactly equal to population_projections_state. Karnataka has the same difference from population_projections_district as the age tables.

The report does not print some values. The values for Theni (Tamil Nadu) 2017-2031 are calculated. The values for the 80+ group of Moga (Punjab) and Tapi (Gujarat) in 2022-2026 are also calculated. `population_projections_district_age()` has the same values.

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

The age-group detail of population_projections_district_lgd: single years of age for 0-14, then five-year bands from 15-19 to 80+. The package downloads the table on first use. Refer to `census_cache_dir()`.

Usage

```
population_projections_district_lgd_age()
```

Value

A tibble with 454,140 rows and 9 columns:

year Projection year (annual, 2012-2031)

geography Geographic level ("district")

state_code Integer 2011 census state code

state_name_harmonized Harmonized state name. Use it to join datasets. Refer to population_projections_district_lgd for the caveat about the merged UT Dadra & Nagar Haveli and Daman & Diu.

district District name as in the LGD

age_group Single years "0".."14", then 5-year bands "15-19" to "80+". The groups do not overlap. The sum of all rows for a district and year equals 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 7 columns:

year Projection year (2011-2036)

geography "india" on the India rows, "state" on all other rows

state_code Integer 2011 census state code. NA for India, Ladakh and the merged Dadra & Nagar Haveli and Daman & Diu.

state_name_harmonized Harmonized state name, used to join datasets

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.

The population of the two former UTs appears twice in the table. When summing across states, exclude `"India"` and either the merged UT or the two former UTs.

Values are counts of persons. Because the source report rounds persons, males, and females to thousands independently, `population` can differ from `males + females` by up to 1,000 (or up to 2,000 for the merged UT, which 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. 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

Population by age group, from Table 18 of the MOHFW report. For each state and year, the sum of the age groups is equal to population_projections_state within the rounding of the report. Values are numbers of persons.

Usage

```
population_projections_state_age
```

Format

A tibble with 2,448 rows and 8 columns:

year Projection year: 2011, 2016, 2021, 2026, 2031 and 2036

geography "india" on the India rows, "state" on all other rows, which include the North East row

state_code Integer 2011 census state code. NA on the India and North East rows.

state_name_harmonized Harmonized state name, used to join datasets. "India" and "North East States Excluding Assam" are the names that the report gives to these two total rows.

age_group Five-year age groups, "0-4" to "80+"

males Projected male population, persons

females Projected female population, persons

population Projected total population (the printed Person value), persons

Details

Values match the printed Table 18. `population` is the printed Person value, which the report rounds to thousands independently of `males` and `females`, so `population` can differ from `males + females` by up to 1,000. The India and Odisha rows use the July 2020 edition, as in population_projections_state, where `population` equals `males + females`.

This table has fewer areas than population_projections_state. The report gives 23 large states separately, and an "India" total. One row, "North East States Excluding Assam", gives the sum of Sikkim, Arunachal Pradesh, Nagaland, Manipur, Mizoram, Tripura and Meghalaya. Chandigarh, the small UTs and Ladakh are not in this table.

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

Search census variables by keyword across variable names and labels.

Usage

```
search_census_variables(pattern)
```

Arguments

pattern	Character string or regular expression pattern.

Value

A tibble of matching variables.

Examples

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

serve *Start the censusindia REST API*

Description

Starts the bundled plumber API on the host and port. Each endpoint takes GET query parameters that agree with the R function behind it. For vectors, use comma-separated values (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. Default "127.0.0.1" (only localhost). Use "0.0.0.0" to accept external connections.
port	Port number. Default 8895.
...	More arguments for the $run() method of plumber.

Details

Endpoint	Function
/query	get_census()
/population	get_population()
/boundaries	get_census_boundaries()
/languages	census_languages()
/sc-st	census_sc_st()
/datasets, /dataset/<name>	a dataset of the package, for example census_1931_castes
/variables, /search	list_census_variables(), search_census_variables()
/states, /geographies	list_states(), list_census_geographies()
/health	package version

The API returns JSON rows. It returns a GeoJSON FeatureCollection from /boundaries and when geometry=true. If the R function rejects the input, the API returns a 400 with the error message. A failed data download gives a 503. The interactive documentation is at /__docs__/.

A response has a maximum of 10,000 rows, or 100 features for GeoJSON. For larger results, use limit and offset to get the pages. The X-Total-Count header gives the full count.

Each request writes one JSON line to standard output. The line has the path, query, status, duration, client IP and a client label.

If the request has an X-Client header, the label comes from that header. An MCP server that wraps the API must send X-Client: mcp/<name>. For other requests, the label comes from the User-Agent: mcp, ai_agent, curl, wget, python, r, node, browser, other or unknown. Under systemd, these lines go to the journal.

Value

No value. In normal operation, the function does not stop.

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 Census of India uses this convention.

Usage

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

Arguments

females, males	Counts, of equal length.

Value

Numeric vector.

Examples

```
sex_ratio(census_2011_pca$females, census_2011_pca$males) |> 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$literate, census_2011_pca$population) |> head()
```

theme_census *Shared theme for censusindia figures*

Description

`theme_census()` is for plots with axes. `theme_census_map()` is 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, population)) +
  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, 19
census_2011_tribes, 21
census_population_time_series, 26
census_subdistricts_2011, 28
census_variables, 29
india_states, 36
population_projections_district, 40
population_projections_district_lgd, 42
population_projections_state, 44
population_projections_state_age, 46

at_level, 4
at_level(), *9*, *12*, *15*, *22*
attach_geometry, 3, *39*
attach_geometry(), *10*, *41*

census_1931_castes, 5, *48*
census_1931_exterior_castes, 6
census_1961_literacy, 6
census_1971, 7
census_1981, *4*, 8
census_2001_pca, 10, *12*, *18*
census_2001_pca_full, 11
census_2011_demographics, 12
census_2011_demographics(), *4*, *18*, *28*
census_2011_district_languages, 13, *20*, *25*
census_2011_linguistic_diversity, 14, *25*
census_2011_marginal_detail, 15
census_2011_marginal_detail(), *4*, *18*
census_2011_mother_tongue, 16
census_2011_mother_tongue(), *25*
census_2011_pca, *10*, *12*, *16, 17*, 17, *19*, *21*, *28*
census_2011_sc_st, 19, *21*, *27*
census_2011_subdistrict_languages, 20
census_2011_subdistrict_languages(), *25*
census_2011_tribes, *19*, 21, *27*
census_2011_workers, 22
census_2011_workers(), *4*, *18*
census_api *(*serve*)*, 47
census_base_layer, 23
census_cache_dir, 24
census_cache_dir(), *4*, *12*, *15, 16*, *20*, *22*, *24*, *33*, *41*, *43*
census_download, 24
census_download(), *24*
census_languages, 25
census_languages(), *13*, *16*, *20*, *48*
census_population_time_series, 26, *29*
census_sc_st, 27
census_sc_st(), *19*, *48*
census_subdistricts_2011, 28
census_variables, 29
census_years, 29
census_years(), *26*
clear_census_cache *(*census_cache_dir*)*, 24
clear_census_cache(), *30*
clear_geometry_cache, 30
clear_geometry_cache(), *24*
compare_maps, 30

get_census, 31
get_census(), *25–27*, *29*, *48*
get_census_boundaries, 33
get_census_boundaries(), *48*
get_palette, 33
get_population, 34
get_population(), *48*
ggplot2::geom_sf(), *23*

india_states, 36

list_census_geographies, 36
list_census_geographies(), *29*, *32*, *48*
list_census_variables, 37
list_census_variables(), *32*, *48*
list_states, 37
list_states(), *48*

plot_map, *31*, 38
plot_map(), *23*
plotting, 38
population_projections_district, 40, *41–43*
population_projections_district_age, 41
population_projections_district_age(), *34*, *41*, *43*
population_projections_district_lgd, 42, *43, 44*
population_projections_district_lgd_age, 43
population_projections_district_lgd_age(), *34*, *41, 42*
population_projections_state, *35*, *43*, 44, *46*
population_projections_state_age, *34*, 46

search_census_variables, 47
search_census_variables(), *48*
serve, 47
sex_ratio, 48
share, 49

theme_census, 50
theme_census_map *(*theme_census*)*, 50
tools::R_user_dir(), *24*