Package 'usa'

May 14, 2026

Title Updated US State Facts and Figures

Version 1.0.0

Description Updated versions of the 1970s ``US State Facts and Figures'' objects from the 'datasets' package included with R. The new data is compiled from a number of sources, primarily from the United States Census Bureau or the relevant federal agency. Modern tidy tibbles provide richer state-level data including identifiers, geography, capitals, demographics, and socioeconomic statistics. Convenience vectors parallel the base 'datasets' state objects but extend coverage to all 51 jurisdictions: the 50 states and the District of Columbia.

License CC BY 4.0

URL https://k5cents.github.io/usa/, https://github.com/k5cents/usa

BugReports https://github.com/k5cents/usa/issues

Depends R (>= 3.5)

Imports tibble (>= 2.1.3)

Suggests covr (>= 3.3.2), pkgdown, testthat (>= 2.1.0)

Config/Needs/website pkgdown

Config/Needs/coverage covr

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

NeedsCompilation no

Author Kiernan Nicholls [aut, cre, cph] (ORCID: <https://orcid.org/0000-0002-9229-7897>)

Maintainer Kiernan Nicholls <k5cents@gmail.com>

Repository CRAN

Date/Publication 2026-05-14 13:50:02 UTC

Contents

`city_names` *US ZIP Cities*

Description

The United States Postal Service's official names for the cities in which ZIP codes are contained. This vector contains unique values, sorted alphabetically; because of this, they do not line up the other vectors in the way zip_codes and zip_centers do.

Usage

```
city_names
```

Format

A character vector of length 19108.

Source

Daniel Coven's web site and the CivicSpace US ZIP Code Database written by Schuyler Erle schuyler@geocoder.us, 5 August 2004.

counties *US Counties*

Description

The county subdivisions of the US states and territories.

Usage

```
counties
```

Format

A tibble with 3,235 rows and 3 variables:

fips Five-digit FIPS code (state FIPS + county FIPS)

name County name (type suffix such as "County", "Parish", "Borough" removed)

state USPS state/territory abbreviation

Source

Census TIGER 2020 national county reference file, https://www2.census.gov/geo/docs/reference/codes2020/national_county2020.txt

county_names *US County Names*

Description

The name of distinct US counties.

Usage

```
county_names
```

Format

A character vector of length 1,925.

Source

Census TIGER 2020 national county reference file, https://www2.census.gov/geo/docs/reference/codes2020/national_county2020.txt

`people` *Synthetic Sample of US population*

Description

A statistically representative synthetic sample of 20,000 Americans. Each record is a simulated survey respondent.

Usage

```
people
```

Format

A tibble with 20,000 rows and 40 variables:

id Sequential unique ID

fname Random first name, see details

lname Random last name, see details

gender Gender (male/female)

age Age capped at 85

race Race and Ethnicity

edu Educational attainment

div Census regional division

married Marital status

house_size Household size

children Has children

us_citizen Is a US citizen

us_born Was born in the US

house_income Family income

emp_status Employment status

emp_sector Employment sector

hours_work Hours worked per week

hours_vary Hours vary week to week

mil Has served in the military

house_own Home ownership

metro Lives in metropolitan area

internet Household has internet access

foodstamp Receives food stamps

house_moved Moved in the last year

pub_contact Contacted or visited a public official

boycott Participated in a product boycott

hood_group Participated in a community association

hood_talks Talked with neighbors

hood_trust Trusts neighbors

tablet Uses a tablet or e-reader

texting Uses text messaging

social Uses social media

volunteer Volunteered

register Is registered to vote

vote Voted in the most recent midterm election

party Political party

religion Religious (evangelical) affiliation

ideology Political ideology

govt Follows government and public affairs

guns Owns a gun

Details

This dataset was originally produced by the Pew Research center for their paper entitled *For Weighting Online Opt-In Samples, What Matters Most?* The synthetic population dataset was created to serve as a reference for making online opt-in surveys more representative of the overall population.

See Appendix B: Synthetic population dataset for a more detailed description of the method for and rationale behind creating this dataset.

In short, the dataset was created to overcome the limitations of using large, federal benchmark survey datasets such as the American Community Survey (ACS) or Current Population Survey (CPS). These surveys often do not contain the exact questions asked in online-opt in surveys, keeping them from being used for proper adjustment.

This *synthetic* dataset was created by combining nine separate benchmark datasets. Each had a set of common demographic variables but many added unique variables such as gun ownership or voter registration. The surveys were combined, stratified, sampled, combined, and imputed to fill missing values from each. From this large dataset, the original 20,000 surveys from the ACS were kept to ensure accurate demographic distribution.

The names were randomly assigned to respondents to better simulate a synthetic sample of the population. First names were taken from the `babynames` dataset which contains the Social Security Administration's record of baby names from 1880 to 2017 along with gender and proportion. First names were proportionally randomly assigned by birth year and sex. Last names were taken from the Census Bureau, who provides the 162,254 most common last names in the 2010 Census, covering over 90% of the population. For a given surname, the proportion of that name belonging to members of each race and ethnicity is provided. The last names were proportionally randomly assigned by race.

Source

"For Weighting Online Opt-In Samples, What Matters Most?" Pew Research Center, Washington, D.C. (January 26, 2018) https://www.pewresearch.org/methods/2018/01/26/for-weighting-online-opt-in-samp

---

`state_abbs` *US State Abbreviations*

Description

The 2-letter USPS abbreviations for the 50 states and District of Columbia. Parallel to state_names.

Usage

```
state_abbs
```

Format

A character vector of length 51.

Source

https://www2.census.gov/geo/docs/reference/state.txt

`state_areas` *US State Land Areas*

Description

Land area in square miles for the 50 states and District of Columbia. Parallel to state_names.

Usage

```
state_areas
```

Format

A numeric vector of length 51.

Source

TIGER/Web REST API (State_County layer)

state_capitals *US State Capitals*

Description

Capital cities for the 50 states and District of Columbia, with coordinates and 2020 Census population.

Usage

```
state_capitals
```

Format

A tibble with 51 rows and 5 variables:

abb 2-letter USPS abbreviation (join key)

capital Capital city name

lat Latitudinal coordinate of the capital

long Longitudinal coordinate of the capital

population Capital city population (2020 Decennial Census, city proper)

Source

https://www.census.gov/quickfacts/

state_centers *US State Geographic Centers*

Description

A list with components named `x` and `y` giving the approximate geographic centroid of each state in longitude and latitude. Parallel to state_names.

Usage

```
state_centers
```

Format

A list of length two, each element a numeric vector of length 51.

x Centroid longitudinal coordinate

y Centroid latitudinal coordinate

Source

TIGER/Web REST API (State_County layer)

state_convert *Convert state identifiers*

Description

Take a vector of state identifiers and convert to a common format. Supports all five identifier types in state_ids: USPS abbreviation, full name, FIPS code, AP style abbreviation, and ISO 3166-2 code.

Usage

```
state_convert(x, to = c("abb", "name", "fips", "ap", "iso"))
```

Arguments

x	A character vector of state identifiers in any supported format.
to	The format returned: "abb", "name", "fips", "ap", or "iso". Defaults to "abb".

Value

A character vector of single format state identifiers.

Examples

```
state_convert(c("AL", "Vermont", "06"))
state_convert(c("AL", "Vermont", "06"), to = "name")
state_convert(c("AL", "Vermont", "06"), to = "fips")
state_convert(c("AL", "Vermont", "06"), to = "ap")
state_convert(c("AL", "Vermont", "06"), to = "iso")
```

state_divisions *US State Census Divisions*

Description

The Census division to which each state belongs, one of nine. Parallel to state_names.

Usage

```
state_divisions
```

Format

A factor vector of length 51.

Details

1. New England
2. Middle Atlantic
3. East North Central
4. West North Central
5. South Atlantic
6. East South Central
7. West South Central
8. Mountain
9. Pacific

Source

https://www2.census.gov/programs-surveys/popest/geographies/2018/state-geocodes-v2018.xlsx

---

state_facts *US State Facts*

Description

Updated version of the datasets::state.x77 matrix, which provided eight statistics from the 1970s. This version is a modern tibble with updated statistics.

Usage

```
state_facts
```

Format

A tibble with 51 rows and 9 variables:

name Full state name

population Resident population (2020 Decennial Census, April 1, 2020)

electors Votes in the Electoral College (2020 Census reapportionment, applies 2022–2032)

admission The date on which the state was admitted to the union

income Per capita income in dollars (2022 ACS 1-year)

life_exp Life expectancy at birth in years, both sexes (2021 NCHS)

murder Homicide rate per 100,000 population (2022 FBI NIBRS)

college Proportion of population 25+ with a bachelor's degree or higher (2022 ACS 1-year)

frost Mean number of days per year with minimum temperature below freezing (1991-2020 NCEI Climate Normals)

Details

See also state_ids for state identifiers and state_geo for geography.

Source

- Population: 2020 Decennial Census PL 94-171 file, variable `P1_001N` via tidycensus
- Electoral College: 2020 Census reapportionment (NARA https://www.archives.gov/electoral-college/allocation)
- Income: 2022 ACS 1-year, variable `B19301_001` (per capita income) via tidycensus
- Life Expectancy: NCHS 2021 state life tables via https://data.cdc.gov/api/views/it4f-frdc/rows.csv
- Murder: FBI Crime Data Explorer API (2022 NIBRS)
- Education: 2022 ACS 1-year Subject Table S1501, variable `S1501_C02_015` via tidycensus
- Frost: NCEI 1991-2020 Climate Normals, variable `ANN-TMIN-AVGNDS-LSTH032`, https://www.ncei.noaa.gov/data/normals-annualseasonal/1991-2020/

`state_geo` *US State Geography*

Description

Geographic and classificatory properties for the 50 states and District of Columbia. Keyed by `abb` to join with state_ids.

Usage

```
state_geo
```

Format

A tibble with 51 rows and 10 variables:

abb 2-letter USPS abbreviation (join key)

region Census Bureau region

division Census Bureau division

area_land Land area in square miles

area_water Water area in square miles

lat Centroid latitudinal coordinate

long Centroid longitudinal coordinate

contiguous `TRUE` for the 48 contiguous states and DC; `FALSE` for Alaska and Hawaii

landlocked `TRUE` for states with no coastline on an ocean, gulf, or Great Lake (21 states including DC)

peak_elev Elevation of the state high point in feet

Source

- Regions and divisions: https://www2.census.gov/programs-surveys/popest/geographies/2018/state-geocodes-v2018.xlsx
- Area and centroids: TIGER/Web REST API (State_County layer)
- Peak elevations: USGS state high point records

`state_ids` *US State Identifiers*

Description

The 50 states and District of Columbia — all naming and coding systems used to refer to each state. The backing data for `state_convert()`.

Usage

```
state_ids
```

Format

A tibble with 51 rows and 6 variables:

name Full legal name

abb 2-letter USPS abbreviation

fips Federal Information Processing Standard Publication 5-2 code

icp IPUMS Integrated Census Project (STATEICP) code, zero-padded 2-digit string

ap AP style abbreviation; the 8 states with no AP abbreviation (Alaska, Hawaii, Idaho, Iowa, Maine, Ohio, Texas, Utah) use the full state name per AP style

iso ISO 3166-2 code (e.g. `"US-AL"`)

Details

Naming convention: underscore objects (`state_ids`, `state_facts`, `state_geo`) are modern purpose-built tibbles. Convenience vectors (`state_abbs`, `state_names`, etc.) mirror the base R `datasets::state.*` vectors but cover all 51 rows (50 states + DC).

Source

- Names, abbreviations, FIPS: https://www2.census.gov/geo/docs/reference/state.txt
- ICP codes: https://usa.ipums.org/usa-action/variables/STATEICP
- AP abbreviations: AP Stylebook
- ISO 3166-2: ISO Online Browsing Platform

state_names *US State Names*

Description

The full names for the 50 states and District of Columbia. Parallel to state_abbs.

Usage

```
state_names
```

Format

A character vector of length 51.

Source

https://www2.census.gov/geo/docs/reference/state.txt

state_regions *US State Census Regions*

Description

The Census region to which each state belongs, one of four. Parallel to state_names.

Usage

```
state_regions
```

Format

A factor vector of length 51.

Details

1. Northeast
2. Midwest
3. South
4. West

Source

https://www2.census.gov/programs-surveys/popest/geographies/2018/state-geocodes-v2018.xlsx

territory *US Territories*

Description

The 6 US territories: Puerto Rico (PR) and the 5 island territories (AS, GU, MP, UM, VI).

Usage

```
territory
```

Format

A tibble with 6 rows and 6 variables:

abb 2-letter abbreviation

name Full legal name

fips Federal Information Processing Standard Publication 5-2 code

area Area in square miles

lat Center latitudinal coordinate

long Center longitudinal coordinate

territory_abbs *US Territory Abbreviations*

Description

The 2-letter abbreviations for the US territories (PR, AS, GU, MP, UM, VI).

Usage

```
territory_abbs
```

Format

A character vector of length 6.

Source

https://www2.census.gov/geo/docs/reference/state.txt

territory_areas *US Territory Areas*

Description

The area in square miles of the US territories (PR, AS, GU, MP, UM, VI).

Usage

```
territory_areas
```

Format

A numeric vector of length 6.

Source

TIGER/Web REST API (State_County layer)

territory_centers *US Territory Geographic Centers*

Description

A list with components named `x` and `y` giving the approximate geographic center of each territory in longitude and latitude.

Usage

```
territory_centers
```

Format

A list of length two, each element a numeric vector of length 6.

x Center longitudinal coordinate

y Center latitudinal coordinate

Source

TIGER/Web REST API (State_County layer)

territory_names *US Territory Names*

Description

The full names for the US territories (PR, AS, GU, MP, UM, VI).

Usage

```
territory_names
```

Format

A character vector of length 6.

Source

https://www2.census.gov/geo/docs/reference/state.txt

zipcodes *US ZIP Code Locations*

Description

This tibble contains city, state, latitude, and longitude for U.S. ZIP codes from the CivicSpace Database (August 2004) augmented by Daniel Coven's web site (updated on January 22, 2012). The data was originally contained in the zipcode CRAN package, which was archived on January 1, 2020.

Usage

```
zipcodes
```

Format

A tibble with 44,336 rows and 5 variables:

zip 5 digit ZIP code or military postal code (FPO/APO)

city USPS official city name

state USPS official state, territory abbreviation code

lat Decimal latitude

long Decimal longitude

Source

Daniel Coven's web site and the CivicSpace US ZIP Code Database written by Schuyler Erle schuyler@geocoder.us, 5 August 2004.

zip_centers *US ZIP Centers*

Description

A list with components named x and y giving the approximate geographic center of each ZIP code in longitude and latitude.

Usage

```
zip_centers
```

Format

A list of length two, each element a numeric vector of length 44336.

x Center longitudinal coordinate

y Center latitudinal coordinate

Source

Daniel Coven's web site and the CivicSpace US ZIP Code Database written by Schuyler Erle schuyler@geocoder.us, 5 August 2004.

zip_codes *US ZIP Codes*

Description

The United States Postal Service's 5-digit codes used to identify a particular postal delivery area.

Usage

```
zip_codes
```

Format

A character vector of length 44336.

Source

Daniel Coven's web site and the CivicSpace US ZIP Code Database written by Schuyler Erle schuyler@geocoder.us, 5 August 2004.

Index