Package ‘usa’

February 23, 2020

Title  Updated US State Facts and Figures
Version  0.1.0
Description  Updated versions of the 1970's ```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 United States Census Bureau or the relevant federal agency.
License  CC BY 4.0
BugReports  https://github.com/kiernann/usa/issues
Depends  R (>= 3.2)
Imports  tibble (>= 2.1.3)
Suggests  covr (>= 3.3.2), testthat (>= 2.1.0)
Encoding  UTF-8
LazyData  true
RoxygenNote  7.0.2
NeedsCompilation  no
Author  Kiernan Nicholls [aut, cre]
Maintainer  Kiernan Nicholls <kiernann@protonmail.com>
Repository  CRAN
Date/Publication  2020-02-23 09:30:03 UTC

R topics documented:

  city.name  .................................................. 2
  counties ....................................................... 3
  county.name .................................................. 3
  facts .......................................................... 4
  people ......................................................... 5
  state.abb ...................................................... 7
  state.area ..................................................... 7
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.code and zip.center do.

Usage

city.name

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

<table>
<thead>
<tr>
<th>counties</th>
<th>US Counties</th>
</tr>
</thead>
</table>

**Description**

The county subdivisions of the US states and territories.

**Usage**

counties

**Format**

A tibble with 3,232 rows and 3 variables:

- **fips** Federal Information Processing Standard Publication 5-2 code
- **name** Census county names
- **state** USPS official state, territory abbreviation code

**Source**

https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/home/?cid=nrcs143_013697

---

county.name

<table>
<thead>
<tr>
<th>county.name</th>
<th>US County Names</th>
</tr>
</thead>
</table>

**Description**

The name of distinct US counties.

**Usage**

county.name

**Format**

A character vector of length 19108.

**Source**

https://www.nrcs.usda.gov/wps/portal/nrcs/detail/national/home/?cid=nrcs143_013697
Description

Updated version of the datasets::state.x77 matrix, which provides eight statistics from the 1970’s. This version is a modern data frame format with updated (and alternative) statistics.

Usage

facts

Format

A tibble with 52 rows and 9 variables:

- **name** Full state name
- **population** Population estimate (September 26, 2019)
- **admission** The data which the state was admitted to the union
- **income** Per capita income (2018)
- **life_exp** Life expectancy in years (2017-18)
- **murder** Murder rate per 100,000 population (2018)
- **high** Percent adult population with at least a high school degree (2019)
- **bach** Percent adult population with at least a bachelor’s degree or greater (2019)
- **heat** Mean number of degree days (temperature requires heating) per year from 1981-2010

Source

- Education: https://data.census.gov/cedsci/table?q=S1501
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** Biological sex
- **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**

**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 2014 midterm elections

**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.
state.abb

Source

state.abb  US State Abbreviations

Description
The 2-letter abbreviations for the US state names.

Usage
state.abb

Format
A character vector of length 52.

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

state.area  US State Areas

Description
The area in square miles of the US states.

Usage
state.area

Format
A numeric vector of length 52.

Source
https://tigerweb.geo.census.gov/tigerwebmain/Files/acs19/tigerweb_acs19_state_us.html
state.center

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>A list with components named x and y giving the approximate geographic center of each state in negative longitude and latitude.</td>
</tr>
</tbody>
</table>

Usage

state.center

Format

A list of length two, each element a numeric vector of length 52.

- x Center longitudinal coordinate
- y Center latitudinal coordinate

Source

https://tigerweb.geo.census.gov/tigerwebmain/Files/acs19/tigerweb_acs19_state_us.html

state.division

<table>
<thead>
<tr>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Census division to which each state belongs, one of nine:</td>
</tr>
</tbody>
</table>

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

Usage

state.division
**state.name**

**Format**
A factor vector of length 52.

**Source**

<table>
<thead>
<tr>
<th>state.name</th>
<th>US State Names</th>
</tr>
</thead>
</table>

**Description**
The full names for the US states.

**Usage**

```
state.name
```

**Format**
A numeric vector of length 52.

**Source**
https://tigerweb.geo.census.gov/tigerwebmain/Files/acs19/tigerweb_acs19_state_us.html

<table>
<thead>
<tr>
<th>state.region</th>
<th>US State Regions</th>
</tr>
</thead>
</table>

**Description**
The Census region to which each state belongs, one of four:

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

**Usage**

```
state.region
```
Format

A factor vector of length 52.

Source


Description

A matrix version of the facts tibble, used to more closely align with the datasets::state.x77 matrix included with R.

Usage

state.x19

Format

A tibble with 52 rows and 9 variables:

- **abb** 2-letter abbreviation
- **population** Population estimate as of September 26, 2019
- **income** Per capita income (2017)
- **life_exp** Life expectancy in years (2017-18)
- **murder** Murder rate per 100,000 population (2018)
- **high** Percent of population with at least a high school degree (2019)
- **bach** Percent of population with at least a bachelor’s degree (2019)
- **heat** Mean number of "degree days" per year from 1981-2010
<table>
<thead>
<tr>
<th>states</th>
<th>US State and Territories</th>
</tr>
</thead>
</table>

**Description**

The 50 states, District of Columbia, and Puerto Rico.

**Usage**

`states`

**Format**

A tibble with 52 rows and 8 variables:

- **abb** 2-letter abbreviation
- **name** Full legal name
- **fips** Federal Information Processing Standard Publication 5-2 code
- **region** Census Bureau region
- **division** Census Bureau division
- **area** Area in square miles
- **lat** Center latitudinal coordinate
- **long** Center longitudinal coordinate

<table>
<thead>
<tr>
<th>territory</th>
<th>US Territories</th>
</tr>
</thead>
</table>

**Description**

The 6 non-state territories and federal district.

**Usage**

`territory`

**Format**

A tibble with 7 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.abb**  
*US Territory Abbreviations*

**Description**

The 2-letter abbreviations for the US territory names.

**Usage**

`territory.abb`

**Format**

A character vector of length 52.

**Source**

[https://www2.census.gov/geo/docs/reference/state.txt](https://www2.census.gov/geo/docs/reference/state.txt)

---

**territory.area**  
*US State Areas*

**Description**

The area in square miles of the US territories.

**Usage**

`territory.area`

**Format**

A numeric vector of length 52.

**Source**

[https://tigerweb.geo.census.gov/tigerwebmain/Files/acs19/tigerweb_acs19_state_us.html](https://tigerweb.geo.census.gov/tigerwebmain/Files/acs19/tigerweb_acs19_state_us.html)
**territory.center**

**US Territory Centers**

**Description**
A list with components named \(x\) and \(y\) giving the approximate geographic center of each territory in negative longitude and latitude.

**Usage**

territory.center

**Format**
A list of length two, each element a numeric vector of length 5.

- \(x\)  Center longitudinal coordinate
- \(y\)  Center latitudinal coordinate

**Source**
https://tigerweb.geo.census.gov/tigerwebmain/Files/acs19/tigerweb_acs19_state_us.html

---

**territory.name**

**US Territory Names**

**Description**
The full names for the US territories.

**Usage**

territory.name

**Format**
A numeric vector of length 52.

**Source**
https://tigerweb.geo.census.gov/tigerwebmain/Files/acs19/tigerweb_acs19_state_us.html
zip.center

**Description**

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

**Usage**

zip.center

**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.code

**Description**

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

**Usage**

zip.code

**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.
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 52 rows and 9 variables:

- zip 5 digit ZIP code or military postal code (FPO/APO)
- city USPS official city name
- state USPS official state, territory abbreviation code
- latitude Decimal Latitude
- longitude 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.
Index

+ Topic datasets
  city.name, 2
counties, 3
county.name, 3
facts, 4
people, 5
state.abb, 7
state.area, 7
state.center, 8
state.division, 8
state.name, 9
state.region, 9
state.x19, 10
states, 11
territory, 11
territory.abb, 12
territory.area, 12
territory.center, 13
territory.name, 13
zip.center, 2, 14
zip.code, 2, 14
zipcodes, 15
city.name, 2
counties, 3
county.name, 3
datasets::state.x77, 4, 10
facts, 4, 10
people, 5
state.abb, 7
state.area, 7
state.center, 8
state.division, 8
state.name, 9
state.region, 9
state.x19, 10
states, 11