

# Package ‘usdata’

June 2, 2024

**Title** Data on the States and Counties of the United States

**Version** 0.3.1

**Description**

Demographic data on the United States at the county and state levels spanning multiple years.

**License** GPL-3

**Encoding** UTF-8

**LazyData** true

**RoxygenNote** 7.3.1

**URL** <https://github.com/OpenIntroStat/usdata>,  
<https://openintrostat.github.io/usdata/>

**BugReports** <https://github.com/OpenIntroStat/usdata/issues>

**Suggests** dplyr, ggplot2, maps, lubridate, sf, testthat

**Imports** tibble

**Depends** R (>= 2.10)

**NeedsCompilation** no

**Author** Mine Çetinkaya-Rundel [aut, cre]  
(<<https://orcid.org/0000-0001-6452-2420>>),  
David Diez [aut],  
Leah Dorazio [aut]

**Maintainer** Mine Çetinkaya-Rundel <[cetinkaya.mine@gmail.com](mailto:cetinkaya.mine@gmail.com)>

**Repository** CRAN

**Date/Publication** 2024-06-02 09:40:02 UTC

## R topics documented:

abbr2state . . . . .	2
airline_delay . . . . .	3
county . . . . .	4
county_2019 . . . . .	5

county_complete . . . . .	9
fatal_police_shootings . . . . .	16
gerrymander . . . . .	18
govrace10 . . . . .	19
houserace10 . . . . .	20
pierce_county_house_sales . . . . .	22
pop_age_2019 . . . . .	23
pop_race_2019 . . . . .	24
prez_pwr . . . . .	25
prrace08 . . . . .	26
senaterace10 . . . . .	27
state2abbr . . . . .	28
state_stats . . . . .	29
urban_owner . . . . .	31
urban_rural_pop . . . . .	32
us_crime_rates . . . . .	33
us_temp . . . . .	34
us_time_survey . . . . .	35
voter_count . . . . .	36
vote_nsa . . . . .	37

<b>Index</b>	<b>39</b>
--------------	-----------

---

abbr2state	<i>Convert state abbreviations to names</i>
------------	---

---

## Description

Two utility functions. One converts state names to the state abbreviations, and the second does the opposite.

## Usage

```
abbr2state(abbr)
```

## Arguments

abbr	A vector of state abbreviation.
------	---------------------------------

## Value

Returns a vector of the same length with the corresponding state names or abbreviations.

## Author(s)

David Diez

**See Also**

[state2abbr](#), [county](#), [county\\_complete](#)

**Examples**

```
abbr2state("MN")
```

---

airline_delay	<i>Airline Delays for December 2019 and 2020.</i>
---------------	---

---

**Description**

Summary Data counts for airline per carrier per US City.

**Usage**

```
airline_delay
```

**Format**

A data frame with 3351 rows and 21 variables.

**year** Year data collected

**month** Numeric representation of the month

**carrier** Carrier.

**carrier\_name** Carrier Name.

**airport** Airport code.

**airport\_name** Name of airport.

**arr\_flights** Number of flights arriving at airport

**arr\_del15** Number of flights more than 15 minutes late

**carrier\_ct** Number of flights delayed due to air carrier. (e.g. no crew)

**weather\_ct** Number of flights due to weather.

**nas\_ct** Number of flights delayed due to National Aviation System (e.g. heavy air traffic).

**security\_ct** Number of flights canceled due to a security breach.

**late\_aircraft\_ct** Number of flights delayed as a result of another flight on the same aircraft delayed

**arr\_cancelled** Number of cancelled flights

**arr\_diverted** Number of flights that were diverted

**arr\_delay** Total time (minutes) of delayed flight.

**carrier\_delay** Total time (minutes) of delay due to air carrier

**weather\_delay** Total time (minutes) of delay due to inclement weather.

**nas\_delay** Total time (minutes) of delay due to National Aviation System.

**security\_delay** Total time (minutes) of delay as a result of a security issue .

**late\_aircraft\_delay** Total time (minutes) of delay flights as a result of a previous flight on the same airplane being late.

**Source**

Bureau of Transportation Statistics

**Examples**

```
library(ggplot2)
ggplot(airline_delay, aes(arr_flights, arr_del15, color = as.factor(year))) +
  geom_point(alpha = 0.3) +
  labs(
    x = "Total Number of inbound flights",
    y = "Number of flights delayed by more than 15 mins",
    title = "Inbound vs delayed flights by year",
    color = "Year"
  )
```

---

county

*United States Counties*

---

**Description**

Data for 3142 counties in the United States. See the [county\\_complete](#) data set for additional variables.

**Usage**

county

**Format**

A data frame with 3142 observations on the following 14 variables.

**name** County names.

**state** State names.

**pop2000** Population in 2000.

**pop2010** Population in 2010.

**pop2017** Population in 2017.

**pop\_change** Population change from 2010 to 2017.

**poverty** Percent of population in poverty in 2017.

**homeownership** Home ownership rate, 2006-2010.

**multi\_unit** Percent of housing units in multi-unit structures, 2006-2010.

**unemployment\_rate** Unemployment rate in 2017.

**metro** Whether the county contains a metropolitan area.

**median\_edu** Median education level (2013-2017).

**per\_capita\_income** Per capita (per person) income (2013-2017).

**median\_hh\_income** Median household income.

**smoking\_ban** Describes whether the type of county-level smoking ban in place in 2010, taking one of the values "none", "partial", or "comprehensive".

**Source**

These data were collected from Census Quick Facts (no longer available as of 2020) and its accompanying pages. Smoking ban data were from a variety of sources.

**See Also**

[county\\_complete](#)

**Examples**

```
library(ggplot2)

ggplot(county, aes(x = median_edu, y = median_hh_income)) +
  geom_boxplot()
```

---

county\_2019

*American Community Survey 2019*

---

**Description**

Data for 3142 counties in the United States with many variables of the 2019 American Community Survey.

**Usage**

```
county_2019
```

**Format**

A data frame with 3142 observations on the following 95 variables.

**state** State.

**name** County name.

**fips** FIPS code.

**median\_individual\_income** Median individual income (2019).

**median\_individual\_income\_moe** Margin of error for median\_individual\_income.

**pop** 2019 population.

**pop\_moe** Margin of error for pop.

**white** Percent of population that is white alone (2015-2019).

**white\_moe** Margin of error for white.

**black** Percent of population that is black alone (2015-2019).

**black\_moe** Margin of error for black.

**native** Percent of population that is Native American alone (2015-2019).

**native\_moe** Margin of error for native.

**asian** Percent of population that is Asian alone (2015-2019).

**asian\_moe** Margin of error for asian.

**pac\_isl** Percent of population that is Native Hawaiian or other Pacific Islander alone (2015-2019).

**pac\_isl\_moe** Margin of error for pac\_isl.

**other\_single\_race** Percent of population that is some other race alone (2015-2019).

**other\_single\_race\_moe** Margin of error for other\_single\_race.

**two\_plus\_races** Percent of population that is two or more races (2015-2019).

**two\_plus\_races\_moe** Margin of error for two\_plus\_races.

**hispanic** Percent of population that identifies as Hispanic or Latino (2015-2019).

**hispanic\_moe** Margin of error for hispanic.

**white\_not\_hispanic** Percent of population that is white alone, not Hispanic or Latino (2015-2019).

**white\_not\_hispanic\_moe** Margin of error for white\_not\_hispanic.

**median\_age** Median age (2015-2019).

**median\_age\_moe** Margin of error for median\_age.

**age\_under\_5** Percent of population under 5 (2015-2019).

**age\_under\_5\_moe** Margin of error for age\_under\_5.

**age\_over\_85** Percent of population 85 and over (2015-2019).

**age\_over\_85\_moe** Margin of error for age\_over\_85.

**age\_over\_18** Percent of population 18 and over (2015-2019).

**age\_over\_18\_moe** Margin of error for age\_over\_18.

**age\_over\_65** Percent of population 65 and over (2015-2019).

**age\_over\_65\_moe** Margin of error for age\_over\_65.

**mean\_work\_travel** Mean travel time to work (2015-2019).

**mean\_work\_travel\_moe** Margin of error for mean\_work\_travel.

**persons\_per\_household** Persons per household (2015-2019)

**persons\_per\_household\_moe** Margin of error for persons\_per\_household.

**avg\_family\_size** Average family size (2015-2019).

**avg\_family\_size\_moe** Margin of error for avg\_family\_size.

**housing\_one\_unit\_structures** Percent of housing units in 1-unit structures (2015-2019).

**housing\_one\_unit\_structures\_moe** Margin of error for housing\_one\_unit\_structures.

**housing\_two\_unit\_structures** Percent of housing units in multi-unit structures (2015-2019).

**housing\_two\_unit\_structures\_moe** Margin of error for housing\_two\_unit\_structures.

**housing\_mobile\_homes** Percent of housing units in mobile homes and other types of units (2015-2019).

**housing\_mobile\_homes\_moe** Margin of error for housing\_mobile\_homes.

**median\_individual\_income\_age\_25plus** Median individual income (2019 dollars, 2015-2019).

**median\_individual\_income\_age\_25plus\_moe** Margin of error for median\_individual\_income\_age\_25plus.

**hs\_grad** Percent of population 25 and older that is a high school graduate (2015-2019).

**hs\_grad\_moe** Margin of error for hs\_grad.

**bachelors** Percent of population 25 and older that earned a Bachelor's degree or higher (2015-2019).

**bachelors\_moe** Margin of error for bachelors.

**households** Total households (2015-2019).

**households\_moe** Margin of error for households.

**households\_speak\_spanish** Percent of households speaking Spanish (2015-2019).

**households\_speak\_spanish\_moe** Margin of error for households\_speak\_spanish.

**households\_speak\_other\_indo\_euro\_lang** Percent of households speaking other Indo-European language (2015-2019).

**households\_speak\_other\_indo\_euro\_lang\_moe** Margin of error for households\_speak\_other\_indo\_euro\_lang.

**households\_speak\_asian\_or\_pac\_isl** Percent of households speaking Asian and Pacific Island language (2015-2019).

**households\_speak\_asian\_or\_pac\_isl\_moe** Margin of error for households\_speak\_asian\_or\_pac\_isl.

**households\_speak\_other** Percent of households speaking non European or Asian/Pacific Island language (2015-2019).

**households\_speak\_other\_moe** Margin of error for households\_speak\_other.

**households\_speak\_limited\_english** Percent of limited English-speaking households (2015-2019).

**households\_speak\_limited\_english\_moe** Margin of error for households\_speak\_limited\_english.

**poverty** Percent of population below the poverty level (2015-2019).

**poverty\_moe** Margin of error for poverty.

**poverty\_under\_18** Percent of population under 18 below the poverty level (2015-2019).

**poverty\_under\_18\_moe** Margin of error for poverty\_under\_18.

**poverty\_65\_and\_over** Percent of population 65 and over below the poverty level (2015-2019).

**poverty\_65\_and\_over\_moe** Margin of error for poverty\_65\_and\_over.

**mean\_household\_income** Mean household income (2019 dollars, 2015-2019).

**mean\_household\_income\_moe** Margin of error for mean\_household\_income.

**per\_capita\_income** Per capita money income in past 12 months (2019 dollars, 2015-2019).

**per\_capita\_income\_moe** Margin of error for per\_capita\_income.

**median\_household\_income** Median household income (2015-2019).

**median\_household\_income\_moe** Margin of error for median\_household\_income.

**veterans** Percent among civilian population 18 and over that are veterans (2015-2019).

**veterans\_moe** Margin of error for veterans.

**unemployment\_rate** Unemployment rate among those ages 20-64 (2015-2019).

**unemployment\_rate\_moe** Margin of error for unemployment\_rate.

**uninsured** Percent of civilian noninstitutionalized population that is uninsured (2015-2019).

**uninsured\_moe** Margin of error for uninsured.

**uninsured\_under\_6** Percent of population under 6 years that is uninsured (2015-2019).

**uninsured\_under\_6\_moe** Margin of error for uninsured\_under\_6.

**uninsured\_under\_19** Percent of population under 19 that is uninsured (2015-2019).

**uninsured\_under\_19\_moe** Margin of error for uninsured\_under\_19.

**uninsured\_65\_and\_older** Percent of population 65 and older that is uninsured (2015-2019).

**uninsured\_65\_and\_older\_moe** Margin of error for uninsured\_65\_and\_older.

**household\_has\_computer** Percent of households that have desktop or laptop computer (2015-2019).

**household\_has\_computer\_moe** Margin of error for household\_has\_computer.

**household\_has\_smartphone** Percent of households that have smartphone (2015-2019).

**household\_has\_smartphone\_moe** Margin of error for household\_has\_smartphone.

**household\_has\_broadband** Percent of households that have broadband internet subscription (2015-2019).

**household\_has\_broadband\_moe** Margin of error for household\_has\_broadband.

### Source

The data were downloaded via the tidycensus R package.

### See Also

[county](#), [county\\_complete](#)

### Examples

```
library(ggplot2)

ggplot(
  county_2019,
  aes(
    x = hs_grad, y = median_individual_income,
    size = sqrt(pop) / 1000
  )
) +
  geom_point(alpha = 0.5) +
  scale_color_discrete(na.translate = FALSE) +
  guides(size = FALSE) +
  labs(
    x = "Percentage of population graduated from high school",
    y = "Median individual income"
  )
)
```



---

county_complete	<i>United States Counties</i>
-----------------	-------------------------------

---

**Description**

Data for 3142 counties in the United States.

**Usage**

county\_complete

**Format**

A data frame with 3142 observations on the following 188 variables.

**state** State.

**name** County name.

**fips** FIPS code.

**pop2000** 2000 population.

**pop2010** 2010 population.

**pop2011** 2011 population.names

**pop2012** 2012 population.

**pop2013** 2013 population.

**pop2014** 2014 population.

**pop2015** 2015 population.

**pop2016** 2016 population.

**pop2017** 2017 population.

**age\_under\_5\_2010** Percent of population under 5 (2010).

**age\_under\_5\_2017** Percent of population under 5 (2017).

**age\_under\_18\_2010** Percent of population under 18 (2010).

**age\_over\_65\_2010** Percent of population over 65 (2010).

**age\_over\_65\_2017** Percent of population over 65 (2017).

**median\_age\_2017** Median age (2017).

**female\_2010** Percent of population that is female (2010).

**white\_2010** Percent of population that is white (2010).

**black\_2010** Percent of population that is black (2010).

**black\_2017** Percent of population that is black (2017).

**native\_2010** Percent of population that is a Native American (2010).

**native\_2017** Percent of population that is a Native American (2017).

**asian\_2010** Percent of population that is a Asian (2010).

**asian\_2017** Percent of population that is a Asian (2017).

**pac\_isl\_2010** Percent of population that is Hawaii or Pacific Islander (2010).

**pac\_isl\_2017** Percent of population that is Hawaii or Pacific Islander (2017).

**other\_single\_race\_2017** Percent of population that identifies as another single race (2017).

**two\_plus\_races\_2010** Percent of population that identifies as two or more races (2010).

**two\_plus\_races\_2017** Percent of population that identifies as two or more races (2017).

**hispanic\_2010** Percent of population that is Hispanic (2010).

**hispanic\_2017** Percent of population that is Hispanic (2017).

**white\_not\_hispanic\_2010** Percent of population that is white and not Hispanic (2010).

**white\_not\_hispanic\_2017** Percent of population that is white and not Hispanic (2017).

**speak\_english\_only\_2017** Percent of population that speaks English only (2017).

**no\_move\_in\_one\_plus\_year\_2010** Percent of population that has not moved in at least one year (2006-2010).

**foreign\_born\_2010** Percent of population that is foreign-born (2006-2010).

**foreign\_spoken\_at\_home\_2010** Percent of population that speaks a foreign language at home (2006-2010).

**women\_16\_to\_50\_birth\_rate\_2017** Birth rate for women ages 16 to 50 (2017).

**hs\_grad\_2010** Percent of population that is a high school graduate (2006-2010).

**hs\_grad\_2016** Percent of population that is a high school graduate (2012-2016).

**hs\_grad\_2017** Percent of population that is a high school graduate (2017).

**some\_college\_2016** Percent of population with some college education (2012-2016).

**some\_college\_2017** Percent of population with some college education (2017).

**bachelors\_2010** Percent of population that earned a bachelor's degree (2006-2010).

**bachelors\_2016** Percent of population that earned a bachelor's degree (2012-2016).

**bachelors\_2017** Percent of population that earned a bachelor's degree (2017).

**veterans\_2010** Percent of population that are veterans (2006-2010).

**veterans\_2017** Percent of population that are veterans (2017).

**mean\_work\_travel\_2010** Mean travel time to work (2006-2010).

**mean\_work\_travel\_2017** Mean travel time to work (2017).

**broadband\_2017** Percent of population who has access to broadband (2017).

**computer\_2017** Percent of population who has access to a computer (2017).

**housing\_units\_2010** Number of housing units (2010).

**homeownership\_2010** Home ownership rate (2006-2010).

**housing\_multi\_unit\_2010** Housing units in multi-unit structures (2006-2010).

**median\_val\_owner\_occupied\_2010** Median value of owner-occupied housing units (2006-2010).

**households\_2010** Households (2006-2010).

**households\_2017** Households (2017).

**persons\_per\_household\_2010** Persons per household (2006-2010).

**persons\_per\_household\_2017** Persons per household (2017).

**per\_capita\_income\_2010** Per capita money income in past 12 months (2010 dollars, 2006-2010)

**per\_capita\_income\_2017** Per capita money income in past 12 months (2017 dollars, 2017)

**metro\_2013** Whether the county contained a metropolitan area in 2013.

**median\_household\_income\_2010** Median household income (2006-2010).

**median\_household\_income\_2016** Median household income (2012-2016).

**median\_household\_income\_2017** Median household income (2017).

**private\_nonfarm\_establishments\_2009** Private nonfarm establishments (2009).

**private\_nonfarm\_employment\_2009** Private nonfarm employment (2009).

**percent\_change\_private\_nonfarm\_employment\_2009** Private nonfarm employment, percent change from 2000 to 2009.

**nonemployment\_establishments\_2009** Nonemployer establishments (2009).

**firms\_2007** Total number of firms (2007).

**black\_owned\_firms\_2007** Black-owned firms, percent (2007).

**native\_owned\_firms\_2007** Native American-owned firms, percent (2007).

**asian\_owned\_firms\_2007** Asian-owned firms, percent (2007).

**pac\_isl\_owned\_firms\_2007** Native Hawaiian and other Pacific Islander-owned firms, percent (2007).

**hispanic\_owned\_firms\_2007** Hispanic-owned firms, percent (2007).

**women\_owned\_firms\_2007** Women-owned firms, percent (2007).

**manufacturer\_shipments\_2007** Manufacturer shipments, 2007 (\$1000).

**mercent\_whole\_sales\_2007** Mercent wholesaler sales, 2007 (\$1000).

**sales\_2007** Retail sales, 2007 (\$1000).

**sales\_per\_capita\_2007** Retail sales per capita, 2007.

**accommodation\_food\_service\_2007** Accommodation and food services sales, 2007 (\$1000).

**building\_permits\_2010** Building permits (2010).

**fed\_spending\_2009** Federal spending, in thousands of dollars (2009).

**area\_2010** Land area in square miles (2010).

**density\_2010** Persons per square mile (2010).

**smoking\_ban\_2010** Describes whether the type of county-level smoking ban in place in 2010, taking one of the values "none", "partial", or "comprehensive".

**poverty\_2010** Percent of population below poverty level (2006-2010).

**poverty\_2016** Percent of population below poverty level (2012-2016).

**poverty\_2017** Percent of population below poverty level (2017).

**poverty\_age\_under\_5\_2017** Percent of population under age 5 below poverty level (2017).

**poverty\_age\_under\_18\_2017** Percent of population under age 18 below poverty level (2017).

**civilian\_labor\_force\_2007** Civilian labor force in 2007.

**employed\_2007** Number of civilians employed in 2007.

**unemployed\_2007** Number of civilians unemployed in 2007.

**unemployment\_rate\_2007** Unemployment rate in 2007.  
**civilian\_labor\_force\_2008** Civilian labor force in 2008.  
**employed\_2008** Number of civilians employed in 2008.  
**unemployed\_2008** Number of civilians unemployed in 2008.  
**unemployment\_rate\_2008** Unemployment rate in 2008.  
**civilian\_labor\_force\_2009** Civilian labor force in 2009.  
**employed\_2009** Number of civilians employed in 2009.  
**unemployed\_2009** Number of civilians unemployed in 2009.  
**unemployment\_rate\_2009** Unemployment rate in 2009.  
**civilian\_labor\_force\_2010** Civilian labor force in 2010.  
**employed\_2010** Number of civilians employed in 2010.  
**unemployed\_2010** Number of civilians unemployed in 2010.  
**unemployment\_rate\_2010** Unemployment rate in 2010.  
**civilian\_labor\_force\_2011** Civilian labor force in 2011.  
**employed\_2011** Number of civilians employed in 2011.  
**unemployed\_2011** Number of civilians unemployed in 2011.  
**unemployment\_rate\_2011** Unemployment rate in 2011.  
**civilian\_labor\_force\_2012** Civilian labor force in 2012.  
**employed\_2012** Number of civilians employed in 2012.  
**unemployed\_2012** Number of civilians unemployed in 2012.  
**unemployment\_rate\_2012** Unemployment rate in 2012.  
**civilian\_labor\_force\_2013** Civilian labor force in 2013.  
**employed\_2013** Number of civilians employed in 2013.  
**unemployed\_2013** Number of civilians unemployed in 2013.  
**unemployment\_rate\_2013** Unemployment rate in 2013.  
**civilian\_labor\_force\_2014** Civilian labor force in 2014.  
**employed\_2014** Number of civilians employed in 2014.  
**unemployed\_2014** Number of civilians unemployed in 2014.  
**unemployment\_rate\_2014** Unemployment rate in 2014.  
**civilian\_labor\_force\_2015** Civilian labor force in 2015.  
**employed\_2015** Number of civilians employed in 2015.  
**unemployed\_2015** Number of civilians unemployed in 2015.  
**unemployment\_rate\_2015** Unemployment rate in 2015.  
**civilian\_labor\_force\_2016** Civilian labor force in 2016.  
**employed\_2016** Number of civilians employed in 2016.  
**unemployed\_2016** Number of civilians unemployed in 2016.  
**unemployment\_rate\_2016** Unemployment rate in 2016.

**uninsured\_2017** Percent of population who are uninsured (2017).

**uninsured\_age\_under\_6\_2017** Percent of population under 6 who are uninsured (2017).

**uninsured\_age\_under\_19\_2017** Percent of population under 19 who are uninsured (2017).

**uninsured\_age\_over\_74\_2017** Percent of population under 74 who are uninsured (2017).

**civilian\_labor\_force\_2017** Civilian labor force in 2017.

**employed\_2017** Number of civilians employed in 2017.

**unemployed\_2017** Number of civilians unemployed in 2017.

**unemployment\_rate\_2017** Unemployment rate in 2017.

**median\_individual\_income\_2019** Median individual income (2019).

**pop\_2019** 2019 population.

**white\_2019** Percent of population that is white alone (2015-2019).

**black\_2019** Percent of population that is black alone (2015-2019).

**native\_2019** Percent of population that is Native American alone (2015-2019).

**asian\_2019** Percent of population that is Asian alone (2015-2019).

**pac\_isl\_2019** Percent of population that is Native Hawaiian or other Pacific Islander alone (2015-2019).

**other\_single\_race\_2019** Percent of population that is some other race alone (2015-2019).

**two\_plus\_races\_2019** Percent of population that is two or more races (2015-2019).

**hispanic\_2019** Percent of population that identifies as Hispanic or Latino (2015-2019).

**white\_not\_hispanic\_2019** Percent of population that is white alone, not Hispanic or Latino (2015-2019).

**median\_age\_2019** Median age (2015-2019).

**age\_under\_5\_2019** Percent of population under 5 (2015-2019).

**age\_over\_85\_2019** Percent of population 85 and over (2015-2019).

**age\_over\_18\_2019** Percent of population 18 and over (2015-2019).

**age\_over\_65\_2019** Percent of population 65 and over (2015-2019).

**mean\_work\_travel\_2019** Mean travel time to work (2015-2019).

**persons\_per\_household\_2019** Persons per household (2015-2019)

**avg\_family\_size\_2019** Average family size (2015-2019).

**housing\_one\_unit\_structures\_2019** Percent of housing units in 1-unit structures (2015-2019).

**housing\_two\_unit\_structures\_2019** Percent of housing units in multi-unit structures (2015-2019).

**housing\_mobile\_homes\_2019** Percent of housing units in mobile homes and other types of units (2015-2019).

**median\_individual\_income\_age\_25plus\_2019** Median individual income (2019 dollars, 2015-2019).

**hs\_grad\_2019** Percent of population 25 and older that is a high school graduate (2015-2019).

**bachelors\_2019** Percent of population 25 and older that earned a Bachelor's degree or higher (2015-2019).

**households\_2019** Total households (2015-2019).

**households\_speak\_spanish\_2019** Percent of households speaking Spanish (2015-2019).

**households\_speak\_other\_indo\_euro\_lang\_2019** Percent of households speaking other Indo-European language (2015-2019).

**households\_speak\_asian\_or\_pac\_isl\_2019** Percent of households speaking Asian and Pacific Island language (2015-2019).

**households\_speak\_other\_2019** Percent of households speaking non European or Asian/Pacific Island language (2015-2019).

**households\_speak\_limited\_english\_2019** Percent of limited English-speaking households (2015-2019).

**poverty\_2019** Percent of population below the poverty level (2015-2019).

**poverty\_under\_18\_2019** Percent of population under 18 below the poverty level (2015-2019).

**poverty\_65\_and\_over\_2019** Percent of population 65 and over below the poverty level (2015-2019).

**mean\_household\_income\_2019** Mean household income (2019 dollars, 2015-2019).

**per\_capita\_income\_2019** Per capita money income in past 12 months (2019 dollars, 2015-2019).

**median\_household\_income\_2019** Median household income (2015-2019).

**veterans\_2019** Percent among civilian population 18 and over that are veterans (2015-2019).

**unemployment\_rate\_2019** Unemployment rate among those ages 20-64 (2015-2019).

**uninsured\_2019** Percent of civilian noninstitutionalized population that is uninsured (2015-2019).

**uninsured\_under\_6\_2019** Percent of population under 6 years that is uninsured (2015-2019).

**uninsured\_under\_19\_2019** Percent of population under 19 that is uninsured (2015-2019).

**uninsured\_65\_and\_older\_2019** Percent of population 65 and older that is uninsured (2015-2019).

**household\_has\_computer\_2019** Percent of households that have desktop or laptop computer (2015-2019).

**household\_has\_smartphone\_2019** Percent of households that have smartphone (2015-2019).

**household\_has\_broadband\_2019** Percent of households that have broadband internet subscription (2015-2019).

## Source

The data prior to 2011 was from <http://census.gov>, though the exact page it came from is no longer available.

More recent data comes from the following sources.

- Downloaded via the `tidycensus` R package.
- Download links for spreadsheets were found on <https://www.ers.usda.gov/data-products/country-level-data-sets/download-data>
- Unemployment - Bureau of Labor Statistics - LAUS data - <https://www.bls.gov/lau/>.
- Median Household Income - Census Bureau - Small Area Income and Poverty Estimates (SAIPE) data.

- The original data table was prepared by USDA, Economic Research Service.
- Census Bureau.
- 2012-16 American Community Survey 5-yr average.
- The original data table was prepared by USDA, Economic Research Service.
- Tim Parker (tparker at ers.usda.gov) is the contact for much of the new data incorporated into this data set.

### See Also

[county](#)

### Examples

```
library(dplyr)
library(ggplot2)

county_complete |>
  mutate(
    pop_change = 100 * ((pop2017 / pop2013) - 1),
    metro_area = if_else(metro_2013 == 1, TRUE, FALSE)
  ) |>
  ggplot(aes(
    x = poverty_2016,
    y = pop_change,
    color = metro_area,
    size = sqrt(pop2017) / 1e3
  )) +
  geom_point(alpha = 0.5) +
  scale_color_discrete(na.translate = FALSE) +
  guides(size = FALSE) +
  labs(
    x = "Percentage of population in poverty (2016)",
    y = "Percentage population change between 2013 to 2017",
    color = "Metropolitan area",
    title = "Population change and poverty"
  )

# Counties with high population change
county_complete |>
  mutate(pop_change = 100 * ((pop2017 / pop2013) - 1)) |>
  filter(pop_change < -10 | pop_change > 25) |>
  select(state, name, fips, pop_change)

# Population by metro area
county_complete |>
  mutate(metro_area = if_else(metro_2013 == 1, TRUE, FALSE)) |>
  filter(!is.na(metro_area)) |>
  ggplot(aes(x = metro_area, y = log(pop2017))) +
  geom_violin() +
  labs(
```

```

    x = "Metro area",
    y = "Log of population in 2017",
    title = "Population by metro area"
  )

# Poverty and median household income
county_complete |>
  mutate(metro_area = if_else(metro_2013 == 1, TRUE, FALSE)) |>
  ggplot(aes(
    x = poverty_2016,
    y = median_household_income_2016,
    color = metro_area,
    size = sqrt(pop2017) / 1e3
  )) +
  geom_point(alpha = 0.5) +
  scale_color_discrete(na.translate = FALSE) +
  guides(size = FALSE) +
  labs(
    x = "Percentage of population in poverty (2016)",
    y = "Median household income (2016)",
    color = "Metropolitan area",
    title = "Poverty and median household income"
  )

# Unemployment rate and poverty
county_complete |>
  mutate(metro_area = if_else(metro_2013 == 1, TRUE, FALSE)) |>
  ggplot(aes(
    x = unemployment_rate_2017,
    y = poverty_2016,
    color = metro_area,
    size = sqrt(pop2017) / 1e3
  )) +
  geom_point(alpha = 0.5) +
  scale_color_discrete(na.translate = FALSE) +
  guides(size = FALSE) +
  labs(
    x = "Unemployment rate (2017)",
    y = "Percentage of population in poverty (2016)",
    color = "Metropolitan area",
    title = "Unemployment rate and poverty"
  )

```

---

fatal\_police\_shootings

*Fatal Police Shootings data.*

---

### Description

A subset of the Washington Post database. Contains records of every fatal police shooting by an on-duty officer since January 1, 2015.



**Usage**

```
fatal_police_shootings
```

**Format**

A data frame with 6421 rows and 12 variables.

**date** date of fatal shooting.

**manner\_of\_death** shot or shot and Tasered.

**armed** Indicates if the victim was armed with some sort of implement that a police officer believed could inflict harm.

**age** the age of the victim.

**gender** The gender of the victim. The Post identifies victims by the gender they identify with if reports indicate that it differs from their biological sex.

**race** W White non-Hispanic; B Black non-Hispanic; A Asian; N Native American; H Hispanic; O Other None unknown.

**city** The municipality where the fatal shooting took place. Note that in some cases this field may contain a county name if a more specific municipality is unavailable or unknown.

**state** two-letter postal code abbreviation.

**signs\_of\_mental\_illness** If news reports have indicated the victim had a history of mental health issues, expressed suicidal intentions or was experiencing mental distress at the time of the shooting.

**threat\_level** The general criteria for the attack label was that there was the most direct and immediate threat to life that would include incidents where officers or others were shot at, threatened with a gun, attacked with other weapons or physical force, etc. ; the attack category is meant to flag the highest level of threat; the other and undetermined categories represent all remaining cases; other includes many incidents where officers or others faced significant threats.

**flee** If news reports have indicated the victim was moving away from officers by Foot, by Car, or Not fleeing.

**body\_camera** If news reports have indicated an officer was wearing a body camera and it may have recorded some portion of the incident.

**Source**

[Washington Post](#)

**Examples**

```
library(dplyr)

# List race frequency and percentage
fatal_police_shootings |>
  group_by(race) |>
  summarize(n = n()) |>
  mutate(freq = n / sum(n) * 100)
# List different weapons that victims were armed with
fatal_police_shootings |>
  distinct(armed)
```

---

 gerrymander

*Gerrymander*


---

### Description

A dataset on gerrymandering and its influence on House elections. The data set was originally built by Jeff Whitmer.

### Usage

```
gerrymander
```

### Format

A data frame with 435 rows and 12 variables:

**district** Congressional district.

**last\_name** Last name of 2016 election winner.

**first\_name** First name of 2016 election winner.

**party16** Political party of 2016 election winner.

**clinton16** Percent of vote received by Clinton in 2016 Presidential Election.

**trump16** Percent of vote received by Trump in 2016 Presidential Election.

**dem16** Did a Democrat win the 2016 House election. Levels of 1 (yes) and 0 (no).

**state** State the Representative is from.

**party18** Political Party of the 2018 election winner.

**dem18** Did a Democrat win the 2018 House election. Levels of 1 (yes) and 0 (no).

**flip18** Did a Democrat flip the seat in the 2018 election? Levels of 1 (yes) and 0 (no).

**gerry** Categorical variable for prevalence of gerrymandering with levels of low, mid and high.

### Source

[Washington Post](#)

### Examples

```
library(ggplot2)
library(dplyr)
ggplot(gerrymander |> filter(gerry != "mid"), aes(clinton16, dem16, color = gerry)) +
  geom_jitter(height = 0.05, size = 3, shape = 1) +
  geom_smooth(method = "glm", method.args = list(family = "binomial"), se = FALSE) +
  scale_color_manual(values = c("purple", "orange")) +
  labs(
    title = "Logistic Regression of 2016 House Elections",
    subtitle = "by Congressional District",
    x = "Percent of Presidential Vote Won by Clinton",
```

```

    y = "Seat Won by Democrat Candidate",
    color = "Gerrymandering"
  )

```

---

govrace10

*Election results for 2010 Governor races in the U.S.*


---

## Description

Election results for 2010 Governor races in the U.S.

## Usage

```
govrace10
```

## Format

A data frame with 37 observations on the following 23 variables.

**id** Unique identifier for the race, which does not overlap with other 2010 races (see [houserace10](#) and [senaterace10](#))

**state** State name

**abbr** State name abbreviation

**name1** Name of the winning candidate

**perc1** Percentage of vote for winning candidate (if more than one candidate)

**party1** Party of winning candidate

**votes1** Number of votes for winning candidate

**name2** Name of candidate with second most votes

**perc2** Percentage of vote for candidate who came in second

**party2** Party of candidate with second most votes

**votes2** Number of votes for candidate who came in second

**name3** Name of candidate with third most votes

**perc3** Percentage of vote for candidate who came in third

**party3** Party of candidate with third most votes

**votes3** Number of votes for candidate who came in third

**name4** Name of candidate with fourth most votes

**perc4** Percentage of vote for candidate who came in fourth

**party4** Party of candidate with fourth most votes

**votes4** Number of votes for candidate who came in fourth

**name5** Name of candidate with fifth most votes

**perc5** Percentage of vote for candidate who came in fifth

**party5** Party of candidate with fifth most votes

**votes5** Number of votes for candidate who came in fifth

**Source**

MSNBC.com, retrieved 2010-11-09.

**Examples**

```
table(govrace10$party1, govrace10$party2)
```

---

houserace10

*Election results for the 2010 U.S. House of Representatives races*

---

**Description**

Election results for the 2010 U.S. House of Representatives races

**Usage**

```
houserace10
```

**Format**

A data frame with 435 observations on the following 24 variables.

**id** Unique identifier for the race, which does not overlap with other 2010 races (see [govrace10](#) and [senaterace10](#))

**state** State name

**abbr** State name abbreviation

**num** District number for the state

**name1** Name of the winning candidate

**perc1** Percentage of vote for winning candidate (if more than one candidate)

**party1** Party of winning candidate

**votes1** Number of votes for winning candidate

**name2** Name of candidate with second most votes

**perc2** Percentage of vote for candidate who came in second

**party2** Party of candidate with second most votes

**votes2** Number of votes for candidate who came in second

**name3** Name of candidate with third most votes

**perc3** Percentage of vote for candidate who came in third

**party3** Party of candidate with third most votes

**votes3** Number of votes for candidate who came in third

**name4** Name of candidate with fourth most votes

**perc4** Percentage of vote for candidate who came in fourth

**party4** Party of candidate with fourth most votes  
**votes4** Number of votes for candidate who came in fourth  
**name5** Name of candidate with fifth most votes  
**perc5** Percentage of vote for candidate who came in fifth  
**party5** Party of candidate with fifth most votes  
**votes5** Number of votes for candidate who came in fifth

### Details

This analysis in the Examples section was inspired by and is similar to that of Nate Silver's district-level analysis on the FiveThirtyEight blog in the New York Times: <https://fivethirtyeight.com/features/2010-an-aligning-election/>

### Source

MSNBC.com, retrieved 2010-11-09.

### Examples

```
hr <- table(houserace10[, c("abbr", "party1")])
nr <- apply(hr, 1, sum)

pr <- prrace08[prrace08$state != "DC", c("state", "p_obama")]
hr <- hr[as.character(pr$state), ]
(fit <- glm(hr ~ pr$p_obama, family = binomial))

x1 <- pr$p_obama[match(houserace10$abbr, pr$state)]
y1 <- (houserace10$party1 == "Democrat") + 0
g <- glm(y1 ~ x1, family = binomial)

x <- pr$p_obama[pr$state != "DC"]
nr <- apply(hr, 1, sum)
plot(x, hr[, "Democrat"] / nr,
     pch = 19, cex = sqrt(nr), col = "#22558844",
     xlim = c(20, 80), ylim = c(0, 1),
     xlab = "Percent vote for Obama in 2008",
     ylab = "Probability of Democrat winning House seat"
    )
X <- seq(0, 100, 0.1)
lo <- -5.6079 + 0.1009 * X
p <- exp(lo) / (1 + exp(lo))
lines(X, p)
abline(h = 0:1, lty = 2, col = "#888888")
```

---

pierce\_county\_house\_sales

*Pierce County House Sales Data for 2020*

---

### Description

Real estate sales for Pierce County, WA in 2020.

### Usage

pierce\_county\_house\_sales

### Format

A data frame with 16814 rows and 19 variables.

**sale\_date** Date the legal document (deed) was executed.

**sale\_price** Dollar amount recorded for the sale.

**house\_square\_feet** Sum of the square feet for the building.

**attic\_finished\_square\_feet** Finished living area in the attic.

**basement\_square\_feet** Total square footage of the basement..

**attached\_garage\_square\_feet** Total square footage of the attached or built in garage(s).

**detached\_garage\_square\_feet** Total detached garage(s) square footage.

**fireplaces** Total count of single, double or PreFab stoves.

**hvac\_description** Text description associated with the predominant heating source for the built-as structure i.e. Forced Air, Electric Baseboard, Steam, etc. .

**exterior** Predominant type of construction materials used for the exterior siding on Residential Buildings.

**interior** Predominant type of materials used on the interior walls. i.e. Sheetrock or Paneling.

**stories** Number of floors/building levels above grade. Stories do not include attic or basement areas.

**roof\_cover** Material used for the roof. I.e. Composition Shingles, Wood Shake, Concrete Tile, etc.

**year\_built** Year the building was built, as stated by the building permit or a historical record.

**bedrooms** Number of bedrooms listed for a residential property.

**bathrooms** Number of baths listed for a residential property. The number is listed as a decimal, i.e. 2.75 = two full and one three-quarter baths. A tub/sink/toilet combination (plus any additional fixtures) is considered 1.0 bath. A shower/sink/toilet combination (plus any additional fixtures) is 0.75 bath. A sink/toilet combination is .5 bath.

**waterfront\_type** Describes the type of waterfront the property adjoins or has legal access to.

**view\_quality** Assigned to reflect the market appeal of the overall view available from the dwelling or property.

**utility\_sewer** Identifies if sewer/septic is installed, available or not available or if the property does not support an on site sewage disposal system.

**Source**

[Pierce County, Washington](#)

**Examples**

```
library(dplyr)
library(lubridate)

# List house sales frequency and average price grouped by month
pierce_county_house_sales |>
  mutate(month_sale = month(sale_date)) |>
  group_by(month_sale) |>
  summarize(freq = n(), mean_price = mean(sale_price)) |>
  arrange(desc(freq))

# List house sales frequency and average price group by waterfront type
pierce_county_house_sales |>
  group_by(waterfront_type) |>
  summarize(freq = n(), mean_price = mean(sale_price)) |>
  arrange(desc(mean_price))
```

---

pop\_age\_2019

*Population Age 2019 Data.*

---

**Description**

State level data on population by age.

**Usage**

```
pop_age_2019
```

**Format**

A data frame with 2820 rows and 4 variables.

**state** State as 2 letter abbreviation.

**state\_name** State name.

**age** Age cohort for population.

**population** Population of age cohort.

**state\_total\_population** total estimated state population in 2019

**Source**

[Centers for Disease Control and Prevention](#)

## Examples

```
library(dplyr)

# List age population for each state with percent of total
pop_age_2019 |>
  group_by(state_name, age) |>
  mutate(percent = population / state_total_population * 100) |>
  select(state_name, age, population, percent)

pop_age_2019 |>
  select(state_name, state_total_population) |>
  distinct() |>
  arrange(desc(state_total_population))
```

---

pop_race_2019	<i>Population Race 2019 Data.</i>
---------------	-----------------------------------

---

## Description

State level data on population by race.

## Usage

```
pop_race_2019
```

## Format

A data frame with 2820 rows and 4 variables.

**state** State as 2 letter abbreviation.

**state\_name** State name.

**race** race cohort for population.

**hispanic** indicates whether population is Hispanic or Latino

**population** Population of race cohort.

**state\_total\_population** total estimated state population in 2019

## Source

[Centers for Disease Control and Prevention](#)



**Examples**

```
library(dplyr)

# List race population for each state with percent of total
pop_race_2019 |>
  group_by(state_name, race, hispanic) |>
  mutate(percent = population / state_total_population * 100) |>
  select(state_name, race, hispanic, population, percent)

pop_race_2019 |>
  select(state_name, state_total_population) |>
  distinct() |>
  arrange(desc(state_total_population))
```

---

```
prez_pwr          Presidential Power.
```

---

**Description**

Data from a Pew Research Center poll about Presidential power/control over gas prices.

**Usage**

```
prez_pwr
```

**Format**

A data frame with 365 rows and 3 variables.

**president** Sitting President at time of the poll.

**party** Political party of the respondent with levels d(emocrat) and r(epublican).

**has\_pwr** Respondent answer to the question: "Is the price of gasoline something the president can do alot about, or is that beyond the president's control?"

**Source**

[Pew Research Center, May 2006 & March 2012.](#)

**Examples**

```
library(ggplot2)
ggplot(prez_pwr, aes(has_pwr, fill = party)) +
  geom_bar() +
  labs(
    title = "Is the price of gasoline something the president can do alot about?",
    x = "",
    y = "Number of respondents",
    fill = "Respondent Party"
  ) +
  facet_wrap(~president)
```

prrace08

*Election results for the 2008 U.S. Presidential race***Description**

Election results for the 2008 U.S. Presidential race

**Usage**

```
prrace08
```

**Format**

A data frame with 51 observations on the following 7 variables.

**state** State name abbreviation

**state\_full** Full state name

**n\_obama** Number of votes for Barack Obama

**p\_obama** Proportion of votes for Barack Obama

**n\_mc\_cain** Number of votes for John McCain

**p\_mc\_cain** Proportion of votes for John McCain

**el\_votes** Number of electoral votes for a state

**Details**

In Nebraska, 4 electoral votes went to McCain and 1 to Obama. Otherwise the electoral votes were a winner-take-all.

**Source**

[Presidential Election of 2008, Electoral and Popular Vote Summary](#), retrieved 2011-04-21.

**Examples**

```
# ==> Obtain 2010 US House Election Data <==#
hr <- table(houserace10[, c("abbr", "party1")])
nr <- apply(hr, 1, sum)

# ==> Obtain 2008 President Election Data <==#
pr <- prrace08[prrace08$state != "DC", c("state", "p_obama")]
hr <- hr[as.character(pr$state), ]
(fit <- glm(hr ~ pr$p_obama, family = binomial))

# ==> Visualizing Binomial outcomes <==#
x <- pr$p_obama[pr$state != "DC"]
nr <- apply(hr, 1, sum)
```

```

plot(x, hr[, "Democrat"] / nr,
     pch = 19, cex = sqrt(nr), col = "#22558844",
     xlim = c(20, 80), ylim = c(0, 1), xlab = "Percent vote for Obama in 2008",
     ylab = "Probability of Democrat winning House seat"
)

# ==> Logistic Regression <==#
x1 <- pr$p_obama[match(housetrace10$abbr, pr$state)]
y1 <- (housetrace10$party1 == "Democrat") + 0
g <- glm(y1 ~ x1, family = binomial)
X <- seq(0, 100, 0.1)
lo <- -5.6079 + 0.1009 * X
p <- exp(lo) / (1 + exp(lo))
lines(X, p)
abline(h = 0:1, lty = 2, col = "#888888")

```

---

senaterace10

*Election results for the 2010 U.S. Senate races*


---

## Description

Election results for the 2010 U.S. Senate races

## Usage

```
senaterace10
```

## Format

A data frame with 38 observations on the following 23 variables.

**id** Unique identifier for the race, which does not overlap with other 2010 races (see [govrace10](#) and [housetrace10](#))

**state** State name

**abbr** State name abbreviation

**name1** Name of the winning candidate

**perc1** Percentage of vote for winning candidate (if more than one candidate)

**party1** Party of winning candidate

**votes1** Number of votes for winning candidate

**name2** Name of candidate with second most votes

**perc2** Percentage of vote for candidate who came in second

**party2** Party of candidate with second most votes

**votes2** Number of votes for candidate who came in second

**name3** Name of candidate with third most votes

**perc3** Percentage of vote for candidate who came in third

**party3** Party of candidate with third most votes  
**votes3** Number of votes for candidate who came in third  
**name4** Name of candidate with fourth most votes  
**perc4** Percentage of vote for candidate who came in fourth  
**party4** Party of candidate with fourth most votes  
**votes4** Number of votes for candidate who came in fourth  
**name5** Name of candidate with fifth most votes  
**perc5** Percentage of vote for candidate who came in fifth  
**party5** Party of candidate with fifth most votes  
**votes5** Number of votes for candidate who came in fifth

### Source

MSNBC.com, retrieved 2010-11-09.

### Examples

```
library(ggplot2)

ggplot(senaterace10, aes(x = perc1)) +
  geom_histogram(binwidth = 5) +
  labs(x = "Winning candidate vote percentage")
```

---

state2abbr

*Convert state names to abbreviations*

---

### Description

Two utility functions. One converts state names to the state abbreviations, and the second does the opposite.

### Usage

```
state2abbr(state)
```

### Arguments

**state** A vector of state name, where there is a little fuzzy matching.

### Value

Returns a vector of the same length with the corresponding state names or abbreviations.

### Author(s)

David Diez

**See Also**

[abbr2state](#), [county](#), [county\\_complete](#)

**Examples**

```
state2abbr("Minnesota")

# Some spelling/capitalization errors okay
state2abbr("mINnesta")
```

---

state_stats	<i>State-level data</i>
-------------	-------------------------

---

**Description**

Information about each state collected from both the official US Census website and from various other sources.

**Usage**

```
state_stats
```

**Format**

A data frame with 51 observations on the following 23 variables.

**state** State name.  
**abbr** State abbreviation (e.g. "MN").  
**fips** FIPS code.  
**pop2010** Population in 2010.  
**pop2000** Population in 2000.  
**homeownership** Home ownership rate.  
**multiunit** Percent of living units that are in multi-unit structures.  
**income** Average income per capita.  
**med\_income** Median household income.  
**poverty** Poverty rate.  
**fed\_spend** Federal spending per capita.  
**land\_area** Land area.  
**smoke** Percent of population that smokes.  
**murder** Murders per 100,000 people.  
**robbery** Robberies per 100,000.  
**agg\_assault** Aggravated assaults per 100,000.

**larceny** Larcenies per 100,000.  
**motor\_theft** Vehicle theft per 100,000.  
**soc\_sec** Percent of individuals collecting social security.  
**nuclear** Percent of power coming from nuclear sources.  
**coal** Percent of power coming from coal sources.  
**tr\_deaths** Traffic deaths per 100,000.  
**tr\_deaths\_no\_alc** Traffic deaths per 100,000 where alcohol was not a factor.  
**unempl** Unemployment rate (February 2012, preliminary).

### Source

Census Quick Facts (no longer available as of 2020), InfoChimps (also no longer available as of 2020), [National Highway Traffic Safety Administration](#) (tr\_deaths, tr\_deaths\_no\_alc), [Bureau of Labor Statistics](#) (unempl).

### Examples

```
library(ggplot2)
library(dplyr)
library(maps)

states_selected <- state_stats |>
  mutate(region = tolower(state)) |>
  select(region, unempl, murder, nuclear)

states_map <- map_data("state") |>
  inner_join(states_selected)

# Unemployment map
ggplot(states_map, aes(map_id = region)) +
  geom_map(aes(fill = unempl), map = states_map) +
  expand_limits(x = states_map$long, y = states_map$lat) +
  scale_fill_viridis_c() +
  labs(x = "", y = "", fill = "Unemployment\n(%)")

# Murder rate map
states_map |>
  filter(region != "district of columbia") |>
  ggplot(aes(map_id = region)) +
  geom_map(aes(fill = murder), map = states_map) +
  expand_limits(x = states_map$long, y = states_map$lat) +
  scale_fill_viridis_c() +
  labs(x = "", y = "", fill = "Murders\nper 100k")

# Nuclear energy map
ggplot(states_map, aes(map_id = region)) +
  geom_map(aes(fill = nuclear), map = states_map) +
  expand_limits(x = states_map$long, y = states_map$lat) +
  scale_fill_viridis_c() +
```

```
labs(x = "", y = "", fill = "Nuclear energy\n(%)")
```

---

 urban\_owner

*Summary of many state-level variables*


---

### Description

Census data for the 50 states plus DC and Puerto Rico.

### Usage

```
urban_owner
```

### Format

A data frame with 52 observations on the following 28 variables.

**state** State

**total\_housing\_units\_2000** Total housing units available in 2000.

**total\_housing\_units\_2010** Total housing units available in 2010.

**pct\_vacant** a numeric vector

**occupied** Occupied.

**pct\_owner\_occupied** a numeric vector

**pop\_st** a numeric vector

**area\_st** a numeric vector

**pop\_urban** a numeric vector

**poppct\_urban** a numeric vector

**area\_urban** a numeric vector

**areapct\_urban** a numeric vector

**popden\_urban** a numeric vector

**pop\_ua** a numeric vector

**poppct\_urban.1** a numeric vector

**area\_ua** a numeric vector

**areapct\_ua** a numeric vector

**popden\_ua** a numeric vector

**pop\_uc** a numeric vector

**poppct\_uc** a numeric vector

**area\_uc** a numeric vector

**areapct\_uc** a numeric vector

**popden\_uc** a numeric vector

**pop\_rural** a numeric vector  
**poppct\_rural** a numeric vector  
**area\_rural** a numeric vector  
**areapct\_rural** a numeric vector  
**popden\_rural** a numeric vector

### Source

US Census.

### Examples

```
urban_owner
```

---

urban_rural_pop	<i>State summary info</i>
-----------------	---------------------------

---

### Description

Census info for the 50 US states plus DC.

### Usage

```
urban_rural_pop
```

### Format

A data frame with 51 observations on the following 5 variables.

**state** US state.  
**urban\_in** a numeric vector  
**urban\_out** a numeric vector  
**rural\_farm** a numeric vector  
**rural\_nonfarm** a numeric vector

### Source

US census.

### Examples

```
urban_rural_pop
```



---

us_crime_rates	<i>US Crime Rates</i>
----------------	-----------------------

---

### Description

National data on the number of crimes committed in the US between 1960 and 2019.

### Usage

```
us_crime_rates
```

### Format

A data frame with 60 rows and 12 variables.

**year** Year data was collected.

**population** Population of the United States the year data was collected.

**total** Total number of violent and property crimes committed.

**violent** Total number of violent crimes committed.

**property** Total number of property crimes committed.

**murder** Number of murders committed. Counted in violent total.

**forcible\_rape** Number of forcible rapes committed. Counted in violent total.

**robbery** Number of robberies committed. Counted in violent total.

**aggravated\_assault** Number of aggravated assaults committed. Counted in violent total.

**burglary** Number of burglaries committed. Counted in property total.

**larceny\_theft** Number of larceny thefts committed. Counted in property total.

**vehicle\_theft** Number of vehicle thefts committed. Counted in property total.

### Source

[Disaster Center](#)

### Examples

```
library(ggplot2)

ggplot(us_crime_rates, aes(x = population, y = total)) +
  geom_point() +
  labs(
    title = "Crimes V Population",
    x = "Population",
    y = "Total Number of Crimes"
  )
```

```
ggplot(us_crime_rates, aes(x = murder)) +  
  geom_boxplot() +  
  labs(  
    title = "US Murders",  
    subtitle = "1960 - 2019",  
    x = "Number of Murders"  
  ) +  
  theme(axis.text.y = element_blank())
```

---

us\_temp

*US Temperature Data*

---

## Description

A representative set of monitoring locations were taken from NOAA data that had both years of interest (1950 and 2022). The information was collected so as to spread the measurements across the continental United States. Daily high and low temperatures are given for each of 24 weather stations.

## Usage

us\_temp

## Format

A data frame with 17250 observations on the following 9 variables.

**station** Station ID, measurements from 24 stations.

**name** Name of the station.

**latitude** Latitude of the station.

**longitude** Longitude of the station.

**elevation** Elevation of the station.

**date** Date of observed temperature.

**tmax** High temp for the observed day.

**tmin** Low temp for the observed day.

**year** Factor variable for year, levels: 1950 and 2022.

## Details

Please keep in mind that these are two annual snapshots from a few dozen arbitrarily selected weather stations. A complete analysis would consider more than two years of data and a more precise random sample uniformly distributed across the United States.

## Source

<https://www.ncei.noaa.gov/cdo-web/>, retrieved 2023-09-23.

**Examples**

```

library(ggplot2)
library(maps)
library(sf)
library(dplyr)

# Summarize temperature by station and year for plotting
summarized_temp <- us_temp |>
  group_by(station, year, latitude, longitude) |>
  summarize(tmax_med = median(tmax, na.rm = TRUE), .groups = "drop") |>
  mutate(plot_shift = ifelse(year == "1950", 0, 2))

# Make a map of the US as a baseline
usa <- st_as_sf(maps::map("state", fill = TRUE, plot = FALSE))

# Layer the US map with summarized temperatures
ggplot(data = usa) +
  geom_sf() +
  geom_point(
    data = summarized_temp,
    aes(x = longitude + plot_shift, y = latitude, fill = tmax_med, shape = year),
    color = "black", size = 3
  ) +
  scale_fill_gradient(high = "red", low = "yellow") +
  scale_shape_manual(values = c(21, 24)) +
  labs(
    title = "Median high temperature, 1950 and 2022",
    x = "Longitude",
    y = "Latitude",
    fill = "Median\nhigh temp",
    shape = "Year"
  )

```

---

us\_time\_survey

*American Time Survey 2009 - 2019*


---

**Description**

Average Time Spent on Activities by Americans

**Usage**

```
us_time_survey
```

**Format**

A data frame with 11 rows and 8 variables.

**year** Year data collected

**household\_activities** Average hours per day spent on household activities - travel included

**eating\_and\_drinking** Average hours per day spent eating and drinking including travel.

**leisure\_and\_sports** Average hours per day spent on leisure and sports - including travel.

**sleeping** Average Hours spent sleeping.

**caring\_children** Average hours spent per day caring for and helping children under 18 years of age.

**working\_employed** Average hours spent working for those employed. (15 years and older)

**working\_employed\_days\_worked** Average hours per day spent working on days worked (15 years and older)

### Source

[US Bureau of Labor Statistics](#)

### Examples

```
library(ggplot2)
us_time_survey$year <- as.factor(us_time_survey$year)
ggplot(us_time_survey, aes(year, sleeping)) +
  geom_point(alpha = 0.3) +
  labs(
    x = "Year",
    y = "Average hours spent Sleeping",
    title = "US Average hours spent sleeping, 2009 - 2019"
  )
```

---

voter\_count

*US Voter Turnout Data.*

---

### Description

State-level data on federal elections held in November between 1980 and 2014.

### Usage

voter\_count

### Format

A data frame with 936 rows and 7 variables.

**year** Year election was held.

**region** Specifies if data is state or national total.

**voting\_eligible\_population** Number of citizens eligible to vote; does not count felons.

**total\_ballots\_counted** Number of ballots cast.

**highest\_office** Number of ballots that contained a vote for the highest office of that election.

**percent\_total\_ballots\_counted** Overall voter turnout percentage.

**percent\_highest\_office** Highest office voter turnout percentage.

## Source

[United States Election Project](#)

## Examples

```
library(ggplot2)

ggplot(voter_count, aes(x = percent_highest_office, y = percent_total_ballots_counted)) +
  geom_point() +
  labs(
    title = "Total Ballots V Highest Office",
    x = "Highest Office",
    y = "Total Ballots"
  )
```

---

vote\_nsa

*Predicting who would vote for NSA Mass Surveillance*

---

## Description

In 2013, the House of Representatives voted to not stop the National Security Agency's (NSA's) mass surveillance of phone behaviors. We look at two predictors for how a representative voted: their party and how much money they have received from the private defense industry.

## Usage

```
vote_nsa
```

## Format

A data frame with 434 observations on the following 5 variables.

**name** Name of the Congressional representative.

**party** The party of the representative: D for Democrat and R for Republican.

**state** State for the representative.

**money** Money received from the defense industry for their campaigns.

**phone\_spy\_vote** Voting to rein in the phone dragnet or continue allowing mass surveillance.

**Source**

MapLight. Available at <http://s3.documentcloud.org/documents/741074/amash-amendment-vote-maplight.pdf>.

**References**

Kravets, D., 2020. Lawmakers Who Upheld NSA Phone Spying Received Double The Defense Industry Cash. WIRED. Available at <https://www.wired.com/2013/07/money-nsa-vote/>.

**Examples**

```
table(vote_nsa$party, vote_nsa$phone_spy_vote)
boxplot(vote_nsa$money / 1000 ~ vote_nsa$phone_spy_vote,
        ylab = "$1000s Received from Defense Industry"
        )
```

# Index

- \* **2008**
  - [prrace08](#), [26](#)
- \* **Abbreviation**
  - [abbr2state](#), [2](#)
  - [state2abbr](#), [28](#)
- \* **States**
  - [prrace08](#), [26](#)
  - [state\\_stats](#), [29](#)
- \* **State**
  - [abbr2state](#), [2](#)
  - [state2abbr](#), [28](#)
- \* **United**
  - [prrace08](#), [26](#)
  - [state\\_stats](#), [29](#)
- \* **crime**
  - [state\\_stats](#), [29](#)
- \* **datasets**
  - [airline\\_delay](#), [3](#)
  - [county](#), [4](#)
  - [county\\_2019](#), [5](#)
  - [county\\_complete](#), [9](#)
  - [fatal\\_police\\_shootings](#), [16](#)
  - [gerrymander](#), [18](#)
  - [govrace10](#), [19](#)
  - [houserace10](#), [20](#)
  - [pierce\\_county\\_house\\_sales](#), [22](#)
  - [pop\\_age\\_2019](#), [23](#)
  - [pop\\_race\\_2019](#), [24](#)
  - [prez\\_pwr](#), [25](#)
  - [prrace08](#), [26](#)
  - [senaterace10](#), [27](#)
  - [state\\_stats](#), [29](#)
  - [urban\\_owner](#), [31](#)
  - [urban\\_rural\\_pop](#), [32](#)
  - [us\\_crime\\_rates](#), [33](#)
  - [us\\_temp](#), [34](#)
  - [us\\_time\\_survey](#), [35](#)
  - [vote\\_nsa](#), [37](#)
  - [voter\\_count](#), [36](#)
- \* **election**
  - [prrace08](#), [26](#)
- \* **energy**
  - [state\\_stats](#), [29](#)
- \* **president**
  - [prrace08](#), [26](#)
- \* **state**
  - [state\\_stats](#), [29](#)
- [abbr2state](#), [2](#), [29](#)
- [airline\\_delay](#), [3](#)
- [county](#), [3](#), [4](#), [8](#), [15](#), [29](#)
- [county\\_2019](#), [5](#)
- [county\\_complete](#), [3–5](#), [8](#), [9](#), [29](#)
- [fatal\\_police\\_shootings](#), [16](#)
- [gerrymander](#), [18](#)
- [govrace10](#), [19](#), [20](#), [27](#)
- [houserace10](#), [19](#), [20](#), [27](#)
- [pierce\\_county\\_house\\_sales](#), [22](#)
- [pop\\_age\\_2019](#), [23](#)
- [pop\\_race\\_2019](#), [24](#)
- [prez\\_pwr](#), [25](#)
- [prrace08](#), [26](#)
- [senaterace10](#), [19](#), [20](#), [27](#)
- [state2abbr](#), [3](#), [28](#)
- [state\\_stats](#), [29](#)
- [urban.owner \(urban\\_owner\)](#), [31](#)
- [urban\\_owner](#), [31](#)
- [urban\\_rural\\_pop](#), [32](#)
- [us\\_crime\\_rates](#), [33](#)
- [us\\_temp](#), [34](#)
- [us\\_time\\_survey](#), [35](#)
- [vote\\_nsa](#), [37](#)
- [voter\\_count](#), [36](#)