Package ‘usdata’

July 22, 2025

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] (ORCID:
<https://orcid.org/0000-0001-6452-2420>),
David Diez [aut],
Leah Dorazio [aut]

Maintainer Mine Çetinkaya-Rundel <cetinkaya.mine@gmail.com>

Repository CRAN

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

Contents

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

Index **39**

`abbr2state` *Convert state abbreviations to names*

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 *Airline Delays for December 2019 and 2020.*

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` *United States Counties*

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/county-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 winnner.

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 Represenatives races*

Description

Election results for the 2010 U.S. House of Represenatives 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 *Population Race 2019 Data.*

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$p_obama[match(houserace10$abbr, pr$state)]
y1 <- (houserace10$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 houserace10)

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 *State-level data*

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` *State summary info*

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 *US Crime Rates*

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 larcency 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