Package ‘usdata’

October 12, 2022

Title     Data on the States and Counties of the United States
Version   0.2.0
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 7.1.1
URL       https://github.com/OpenIntroStat/usdata
BugReports https://github.com/OpenIntroStat/usdata/issues
Suggests  dplyr, ggplot2, maps, 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>
Repository CRAN
Date/Publication 2021-06-21 19:10:02 UTC

R topics documented:

  abbr2state ......................................................  2
  county ..........................................................  3
  county_2019 .....................................................  4
  county_complete .................................................  7
  govtrace10 ...................................................... 15
  houserace10 ..................................................... 16
**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")`
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.
pop_change  Population change from 2010 to 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.
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

```r
library(ggplot2)

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

<table>
<thead>
<tr>
<th>county_2019</th>
<th>American Community Survey 2019</th>
</tr>
</thead>
</table>

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_moe Margin of error for median_age.
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_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.
<table>
<thead>
<tr>
<th>Description</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>households_moe</td>
<td>Margin of error for households.</td>
</tr>
<tr>
<td>households_speak_spanish_moe</td>
<td>Margin of error for households_speak_spanish.</td>
</tr>
<tr>
<td>households_speak_other_indo_euro_lang_moe</td>
<td>Margin of error for households_speak_other_indo_euro_lang.</td>
</tr>
<tr>
<td>households_speak_asian_or_pac_isl</td>
<td>Percent of households speaking Asian and Pacific Island language (2015-2019).</td>
</tr>
<tr>
<td>households_speak_asian_or_pac_isl_moe</td>
<td>Margin of error for households_speak_asian_or_pac_isl.</td>
</tr>
<tr>
<td>households_speak_other</td>
<td>Percent of households speaking non European or Asian/Pacific Island language (2015-2019).</td>
</tr>
<tr>
<td>households_speak_other_moe</td>
<td>Margin of error for households_speak_other.</td>
</tr>
<tr>
<td>households_speak_limited_english_moe</td>
<td>Margin of error for households_speak_limited_english.</td>
</tr>
<tr>
<td>poverty</td>
<td>Percent of population below the poverty level (2015-2019).</td>
</tr>
<tr>
<td>poverty_moe</td>
<td>Margin of error for poverty.</td>
</tr>
<tr>
<td>poverty_under_18</td>
<td>Percent of population under 18 below the poverty level (2015-2019).</td>
</tr>
<tr>
<td>poverty_under_18_moe</td>
<td>Margin of error for poverty_under_18.</td>
</tr>
<tr>
<td>poverty_65_and_over</td>
<td>Percent of population 65 and over below the poverty level (2015-2019).</td>
</tr>
<tr>
<td>poverty_65_and_over_moe</td>
<td>Margin of error for poverty_65_and_over.</td>
</tr>
<tr>
<td>mean_household_income</td>
<td>Mean household income (2019 dollars, 2015-2019).</td>
</tr>
<tr>
<td>mean_household_income_moe</td>
<td>Margin of error for mean_household_income.</td>
</tr>
<tr>
<td>per_capita_income</td>
<td>Per capita money income in past 12 months (2019 dollars, 2015-2019).</td>
</tr>
<tr>
<td>per_capita_income_moe</td>
<td>Margin of error for per_capita_income.</td>
</tr>
<tr>
<td>median_household_income_moe</td>
<td>Margin of error for median_household_income.</td>
</tr>
<tr>
<td>veterans</td>
<td>Percent among civilian population 18 and over that are veterans (2015-2019).</td>
</tr>
<tr>
<td>veterans_moe</td>
<td>Margin of error for veterans.</td>
</tr>
<tr>
<td>unemployment_rate</td>
<td>Unemployment rate among those ages 20-64 (2015-2019).</td>
</tr>
<tr>
<td>unemployment_rate_moe</td>
<td>Margin of error for unemployment_rate.</td>
</tr>
<tr>
<td>uninsured</td>
<td>Percent of civilian noninstitutionalized population that is uninsured (2015-2019).</td>
</tr>
<tr>
<td>uninsured_moe</td>
<td>Margin of error for uninsured.</td>
</tr>
<tr>
<td>uninsured_under_6</td>
<td>Percent of population under 6 years that is uninsured (2015-2019).</td>
</tr>
<tr>
<td>uninsured_under_6_moe</td>
<td>Margin of error for uninsured_under_6.</td>
</tr>
<tr>
<td>uninsured_under_19</td>
<td>Percent of population under 19 that is uninsured (2015-2019).</td>
</tr>
<tr>
<td>uninsured_under_19_moe</td>
<td>Margin of error for uninsured_under_19.</td>
</tr>
<tr>
<td>uninsured_65_and_older</td>
<td>Percent of population 65 and older that is uninsured (2015-2019).</td>
</tr>
<tr>
<td>uninsured_65_and_older_moe</td>
<td>Margin of error for uninsured_65_and_older.</td>
</tr>
</tbody>
</table>
**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**

```r
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**

```r
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.
- **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 an Asian (2010).
- **asian_2017**: Percent of population that is an 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).


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).


percent_change_private_nonfarm_employment_2009  Private nonfarm employment, percent change from 2000 to 2009.
firms_2007  Total number of firms (2007).
manufacturer_shipments_2007  Manufacturer shipments, 2007 ($1000).
mercent_whole_sales_2007  Mercent wholesaler sales, 2007 ($1000).
sales_2007  Retail sales, 2007 ($1000).
accommodation_food_service_2007  Accommodation and food services sales, 2007 ($1000).
building_permits_2010  Building permits (2010).
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_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).
employed_2007  Number of civilians employed in 2007.
unemployed_2007  Number of civilians unemployed in 2007.
employed_2008  Number of civilians employed in 2008.
unemployed_2008  Number of civilians unemployed 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.
civlilan_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.
civlilan_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.
civlilan_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.
civlilan_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.
civlilan_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.
civlilan_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.
civlilan_labor_force_2016  Civilian labor force in 2016.
employed_2016  Number of civilians employed in 2016.
unemployed_2016  Number of civilians unemployed 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).
employed_2017  Number of civilians employed in 2017.
unemployed_2017  Number of civilians unemployed 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).


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).


mean_work_travel_2019 Mean travel time to work (2015-2019).

persons_per_household_2019 Persons per household (2015-2019)


housing_one_unit_structures_2019 Percent of housing units in 1-unit structures (2015-2019).


housing_mobile_homes_2019 Percent of housing units in mobile homes and other types of units (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_speak_other_2019 Percent of households speaking non European or Asian/Pacific Island language (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).
veterans_2019  Percent among civilian population 18 and over that are veterans (2015-2019).
uninsured_under_6_2019  Percent of population under 6 years 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
- 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

```r
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) +
```
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/](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"]
nyr <- apply(hr, 1, sum)
plot(x, hr[,"Democrat"] / nr,
    pch = 19, cex = sqrt(nyr), 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")

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


Examples

```r
#===> 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")
```

---

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

```r
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.


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 %>%
urban_owner

```r
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

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

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

```r
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

```r
vote_nsa
```
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

References

Examples
```r
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, 18
* Abbreviation
  abbr2state, 2
  state2abbr, 21
* States
  prrace08, 18
  state_stats, 21
* State
  abbr2state, 2
  state2abbr, 21
* United
  prrace08, 18
  state_stats, 21
* crime
  state_stats, 21
* datasets
  county, 3
  county_2019, 4
  county_complete, 7
  govrace10, 15
  houserace10, 16
  prrace08, 18
  senaterace10, 19
  state_stats, 21
  urban_owner, 23
  urban_rural_pop, 25
  vote_nsa, 26
  voter_count, 25
* election
  prrace08, 18
* energy
  state_stats, 21
* president
  prrace08, 18
* state
  state_stats, 21

abbr2state, 2, 21

county, 2, 3, 7, 13, 21
county_2019, 4
county_complete, 2, 3, 7, 7, 21
govrace10, 15, 17, 20
houserace10, 15, 16, 20
prrace08, 18
senaterace10, 15, 17, 19
state2abbr, 2, 21
state_stats, 21
urban_owner (urban_owner), 23
urban_owner, 23
urban_rural_pop, 25
vote_nsa, 26
voter_count, 25