

# Package ‘usdata’

June 30, 2020

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

**Version** 0.1.0

**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** 7.1.0

**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]

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

**Repository** CRAN

**Date/Publication** 2020-06-30 09:20:03 UTC

## R topics documented:

abbr2state . . . . .	2
county . . . . .	3
county_complete . . . . .	4
govrace10 . . . . .	10
houserace10 . . . . .	11
prrace08 . . . . .	13
senaterace10 . . . . .	14

state2abbr . . . . .	16
state_stats . . . . .	16
urban_owner . . . . .	18
urban_rural_pop . . . . .	20
vote_nsa . . . . .	20
<b>Index</b>	<b>22</b>

---

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

---

### Description

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

### Usage

```
abbr2state(abbr)
```

### Arguments

abbr            A vector of state abbreviation.

### Value

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

### Author(s)

David Diez

### See Also

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

### Examples

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

---

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

---

## Description

Data for 3143 counties in the United States.

## Usage

```
county_complete
```

## Format

A data frame with 3143 observations on the following 111 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.

### 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.

- 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"
  )

```

---

govrace10

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


---

## Description

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

## Usage

```
govrace10
```

## Format

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

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

**state** State name

**abbr** State name abbreviation

**name1** Name of the winning candidate

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

**party1** Party of winning candidate

**votes1** Number of votes for winning candidate

**name2** Name of candidate with second most votes  
**perc2** Percentage of vote for candidate who came in second  
**party2** Party of candidate with second most votes  
**votes2** Number of votes for candidate who came in second  
**name3** Name of candidate with third most votes  
**perc3** Percentage of vote for candidate who came in third  
**party3** Party of candidate with third most votes  
**votes3** Number of votes for candidate who came in third  
**name4** Name of candidate with fourth most votes  
**perc4** Percentage of vote for candidate who came in fourth  
**party4** Party of candidate with fourth most votes  
**votes4** Number of votes for candidate who came in fourth  
**name5** Name of candidate with fifth most votes  
**perc5** Percentage of vote for candidate who came in fifth  
**party5** Party of candidate with fifth most votes  
**votes5** Number of votes for candidate who came in fifth

### Source

MSNBC.com, retrieved 2010-11-09.

### Examples

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

---

houserace10

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

---

### Description

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

### Usage

```
houserace10
```

**Format**

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

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

**state** State name

**abbr** State name abbreviation

**num** District number for the state

**name1** Name of the winning candidate

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

**party1** Party of winning candidate

**votes1** Number of votes for winning candidate

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

**Details**

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

**Source**

MSNBC.com, retrieved 2010-11-09.

**Examples**

```

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

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

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

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

```

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

---

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](#), [13](#)
  - \* **Abbreviation**
    - [abbr2state](#), [2](#)
    - [state2abbr](#), [16](#)
  - \* **States**
    - [prrace08](#), [13](#)
    - [state\\_stats](#), [16](#)
  - \* **State**
    - [abbr2state](#), [2](#)
    - [state2abbr](#), [16](#)
  - \* **United**
    - [prrace08](#), [13](#)
    - [state\\_stats](#), [16](#)
  - \* **crime**
    - [state\\_stats](#), [16](#)
  - \* **datasets**
    - [county](#), [3](#)
    - [county\\_complete](#), [4](#)
    - [govrace10](#), [10](#)
    - [houserace10](#), [11](#)
    - [prrace08](#), [13](#)
    - [senaterace10](#), [14](#)
    - [state\\_stats](#), [16](#)
    - [urban\\_owner](#), [18](#)
    - [urban\\_rural\\_pop](#), [20](#)
    - [vote\\_nsa](#), [20](#)
  - \* **election**
    - [prrace08](#), [13](#)
  - \* **energy**
    - [state\\_stats](#), [16](#)
  - \* **president**
    - [prrace08](#), [13](#)
  - \* **state**
    - [state\\_stats](#), [16](#)
- [abbr2state](#), [2](#), [16](#)
- [county](#), [2](#), [3](#), [8](#), [16](#)
- [county\\_complete](#), [2](#), [3](#), [4](#), [16](#)
- [govrace10](#), [10](#), [12](#), [15](#)
- [houserace10](#), [10](#), [11](#), [15](#)
- [prrace08](#), [13](#)
- [senaterace10](#), [10](#), [12](#), [14](#)
- [state2abbr](#), [2](#), [16](#)
- [state\\_stats](#), [16](#)
- [urban.owner \(urban\\_owner\)](#), [18](#)
- [urban\\_owner](#), [18](#)
- [urban\\_rural\\_pop](#), [20](#)
- [vote\\_nsa](#), [20](#)