# Package ‘socviz’

August 21, 2026

**Type** Package

**Title** Utilities and Data Sets for Data Visualization

**Version** 2.0.0

**Maintainer** Kieran Healy <kjhealy@gmail.com>

**Description** Supporting materials for a course and book on data visualization. It contains utility functions for graphs and several sample data sets. See Healy (2019) <ISBN 978-0691181622>.

**License** MIT + file LICENSE

**Depends** R (>= 4.5)

**Imports** ggplot2 (>= 4.0.0), graphics, grid, magrittr, systemfonts, tibble

**Encoding** UTF-8

**LazyData** true

**LazyDataCompression** xz

**URL** https://kjhealy.github.io/socviz/, https://github.com/kjhealy/socviz

**BugReports** https://github.com/kjhealy/socviz/issues

**NeedsCompilation** no

**Config/roxygen2/version** 8.1.0

**Author** Kieran Healy [aut, cre, cph]

**Repository** CRAN

**Date/Publication** 2026-08-21 12:40:02 UTC

## Contents


acs_poverty . . . . . 2
acs_poverty_lon . . . . . 3
asasec . . . . . 4
center_df . . . . . 5
color_comp . . . . . 5

color_pal . . . . . 6
color_table . . . . . 7
counties_sf . . . . . 7
county_comp . . . . . 8
county_data . . . . . 9
county_map . . . . . 10
edu . . . . . 11
election . . . . . 12
election24 . . . . . 13
election24_county_df . . . . . 14
elections_historic . . . . . 15
farsinvolved . . . . . 16
fredts . . . . . 17
gss_lon . . . . . 17
gss_sm . . . . . 18
int_to_year . . . . . 20
lawschools . . . . . 20
maunaloa . . . . . 21
oecd_le . . . . . 22
oecd_sum . . . . . 22
okboomer . . . . . 23
opiates . . . . . 24
organdata . . . . . 25
prefix_replace . . . . . 26
prefix_strip . . . . . 27
round_df . . . . . 28
states_sf . . . . . 28
studebt . . . . . 29
theme_socviz . . . . . 30
theme_socviz_map . . . . . 31
titanic . . . . . 32
tw_tab . . . . . 32
yahoo . . . . . 33
%nin% . . . . . 34

**Index** **35**

---

`acs_poverty` *US County Poverty Rates by Age Group*

---

## Description

A dataset of US poverty rates by selected age groups within counties.

## Usage

```
acs_poverty
```

### Format

A tibble with 9,666 rows and 4 columns.

### Details

- fips. County FIPS code.
- age_group. Adults 18-64, Children <18, Seniors 65+.
- age_rate. Poverty rate (percent) for ‘age_group‘ in county.
- total_rate. Poverty rate (percent) for all ages in county.

### Source

American Community Survey 2023 ACS 5-year estimates, Table B17018.

---

## `acs_poverty_lon` *US County Poverty Rates by Age Group, Longer Version*

---

### Description

A dataset of US poverty rates by selected age groups within counties.

### Usage

```
acs_poverty_lon
```

### Format

A tibble with 12,888 rows and 3 columns.

### Details

- geoid. County FIPS code.
- age_group. Adults 18-64, Children <18, Seniors 65+, All Ages.
- prop_poor. Proportion (0-1) of ‘age_group‘ in poverty within county.

### Source

American Community Survey 2023 ACS 5-year estimates, Table B17018.

`asasec` *American Sociological Association Section Membership*

## Description

Membership and some financial information for sections of the American Sociological Association in 2014/15.

## Usage

```
asasec
```

## Format

## ‘asasec‘ A data frame with 52 rows and 9 columns:

**Section** Section name.

**Sname** Short name.

**Beginning** Cash on hand at beginning of year (2015).

**Revenues** Membership revenues.

**Expenses** Section expenses.

**Ending** Cash on hand at end of year (2015).

**Journal** Does the Section run a journal?

**Year** Membership year.

**Members** Number of members (2014).

## Details

Data from the American Sociological Association.

## Author(s)

Kieran Healy

## Source

ASA Annual Report 2016.

## center_df — *center_df*

### Description

Scale and/or center the numeric columns of a data frame or tibble

### Usage

```
center_df(data, sc = FALSE, cen = TRUE)
```

### Arguments

| | |
|---|---|
| data | A data frame or tibble |
| sc | Scale the variables (default FALSE) |
| cen | Center the variables on their means (default TRUE) |

### Details

Takes a data frame or tibble as input and scales and/or centers the numeric columns. By default, centers but doesn't scale

### Value

An object of the same class as ‘data‘, with the numeric columns scaled or centered as requested

### Author(s)

Kieran Healy

### Examples

```
head(center_df(organdata))
```

## color_comp — *Plot a table of color hex values as a table of colors*

### Description

Plot a table of color hex values as a table of colors

### Usage

```
color_comp(df)
```

### Arguments

| | |
|---|---|
| `df` | data frame of color hex values |

### Details

Given a data frame of color values, plot them as swatches

### Value

Plot of table of colors

### Author(s)

Kieran Healy

### Examples

```
color_table
color_comp(color_table)
```

---

## `color_pal` *Draw a palette of colors*

---

### Description

Draw a palette of colors

### Usage

```
color_pal(col, border = "gray70", ...)
```

### Arguments

| | |
|---|---|
| `col` | vector of colors |
| `border` | border |
| `...` | other arguments |

### Details

Borrowed from the colorspace library

### Value

Plot of a color palette

### Author(s)

colorspace library authors

### Examples

```
color_pal(c("#66C2A5", "#FC8D62", "#8DA0CB"))
```

---

## color_table — *A table of hex color values related to types of color blindness*

---

### Description

Hex values for five default ggplot colors, with corresponding approximations for three kinds of color blindness. Produced by the ‘dichromat‘ package.

### Usage

```
color_table
```

### Format

A tibble with five rows and four columns.

### Source

Kieran Healy

---

## counties_sf — *US County geometries and demographic data*

---

### Description

US County map data layer with selected Census Bureau demographic variables.

### Usage

```
counties_sf
```

### Format

## ‘counties_sf‘ A simple features object 3,144 rows and 16 columns:

**fips** FIPS code.

**name** County name.

**area_sqmi** Area in square miles.

**white** N White population.

**black** N Black population.

**asian** N Asian population.

**nh_white** N Non-Hispanic White population.

**hispanic** N Hispanic population.

**pop** Total population.

**black_disc** Percent Black, discretized.

**hisp_disc** Percent Hispanic, discretized.

**nhwhite_disc** Percent Non-Hispanic White, discretized.

**asian_disc** Percent Asian, discretized.

**pop_dens** Population density per square mile.

**pop_dens_disc** Population density per square mile, discretized.

**su_gun6** Firearm-related suicides per 100,000 population, 1999-2015. Factor variable cut into six categories. Note that the values in this variable contain an inaccurate bottom-quartile coding by construction. Do not present this variable as an accurate measure of the firearm-related suicide rate.

**pop_dens6** Population density per square mile, discretized into six categories, 2014 estimates.

**geometry** Geometry.

## Details

A simple features object. Load the ‘sf‘ package before using. Alaska and Hawaii have had their geometries scaled and shifted to the bottom left of the map area. Alaska’s Aleutian islands are not included. Except where noted, population counts and other demographic information are from the 2024 5-year ACS estimates.

## Author(s)

Kieran Healy

## Source

US Census Bureau.

---

`county_comp` *US County Components of Population Change*

---

## Description

A dataset of components of population change (rates only) for US Counties in 2023.

## Usage

```
county_comp
```

## Format

A tibble with 3,144 rows and 8 columns.

### Details

- fips. County FIPS code.
- county. County name.
- state. State abbreviation.
- rbirth. Birth rate.
- rdeath. Death rate.
- rnatchg. Natural change rate.
- rintl. International migration rate.
- rdom. Domestic migration rate.
- rnetmig. Net migration rate.

### Source

US Census Bureau Components of Population Change 2023 estimates.

---

## county_data *Census Data on US Counties*

---

### Description

Selected county data (including US and state-level observations on some variables). Preserved for use with the first edition of the book only.

### Usage

```
county_data
```

### Format

A data frame with 3195 rows and 13 columns.

### Details

The variables are as follows:

- id. FIPS State and County code (character)
- name. State or County Name
- state. State abbreviation
- census_region. Census region
- pop_dens. Population density per square mile, 2014 estimate (seven categories).
- pct_black. Percent black population, 2014 estimate (seven category factor)
- pop_dens6. Population density per square mile, 2014 estimate (six categories)
- su_gun6. Firearm-related suicides per 100,000 population, 1999-2015. Factor variable cut into six categories. Note that the values in this variable contain an inaccurate bottom-quartile coding by construction. Do not present this variable as an accurate measure of the firearm-related suicide rate.

### Source

US Census Bureau, Centers for Disease Control

---

## county_map *US County map file*

---

### Description

US county map data

### Usage

```
county_map
```

### Format

A data frame with 191,372 rows and 7 columns.

### Details

- long. Longitude
- lat. Latitude
- order. Order
- hole. Hole (true/false)
- piece. Piece
- group. Group
- id. FIPS code

### Source

Eric Celeste

edu *Years of school completed by people 25 years and over in the US.*

## Description

Counts of educational attainment (in thousands) from 1940 to 2016

## Usage

```
edu
```

## Format

A tibble with 366 rows and 11 columns.

## Details

The variables are as follows:

- age Character. Cut into 25-34, 35-54, 55>
- sex Character. Male, Female.
- year Integer.
- total Integer. Total in thousands.
- elem4 Double. 0 to 4 years of Elementary School completed.
- elem8 Double. 5 to 8 years of Elementary School completed.
- hs3 Double. 1 to 3 years of High School completed.
- hs4 Double. 4 years of High School completed.
- coll3 Double. 1 to 3 years of College completed.
- coll4 Double. 4 or more years of College completed.
- median Double. Median years of education.

## Source

US Census Bureau

`election` *US Presidential Election 2016, State-level results*

## Description

State-level vote totals and shares for the 2016 US Presidential election. The variables are as follows:

- state. State name.
- st. State abbreviation.
- fips. State FIPS code
- total_vote. Total votes cast.
- vote_margin. Winner's vote margin
- winner. Winning candidate.
- party. Winning party.
- pct_margin. Winner's percentage margin (proportion of total vote)
- r_points. Percentage point difference between Trump share and Clinton
- d_points. Percentage point difference between Clinton share and Trump
- pct_clinton. Clinton vote share (proportion)
- pct_trump. Trump vote share (proportion)
- pct_johnson. Johnson vote share (proportion)
- pct_other. Other vote share (proportion)
- clinton_vote. Clinton vote total
- trump_vote. Trump vote total
- johnson_vote. Johnson vote total
- other_vote. Other vote total
- ev_dem. Electoral votes for Clinton
- ev_rep. Electoral votes for Trump
- ev_oth. Electoral votes for Other
- census. Census region.

## Usage

```
election
```

## Format

A (tibble) data frame with 51 rows and 22 columns.

## Source

Vote data from Dave Leip, US Election Atlas, http://uselectionatlas.org.

`election24` *US Presidential Election 2024, State-level results*

## Description

State-level vote totals and shares for the 2024 US Presidential election.

## Usage

```
election24
```

## Format

## ‘election24‘ A data frame with 51 rows and 20 columns:

**state** State name.

**st** State abbreviation.

**fips** State FIPS code (character).

**total_vote** Total votes case.

**vote_margin** Vote margin (Trump positive values; Harris negative.)

**winner** Winning candidate.

**party** Winning party.

**pct_margin** Winner’s percentage margin (proportion of total vote)

**r_points** Percentage point difference between Trump vote percent and Harris vote percent

**d_points** Percentage point difference between Harris vote percent and Trump vote percent

**pct_harris** Harris vote share (proportion)

**pct_trump** Trump vote share (proportion)

**pct_other** Other vote share (proportion)

**harris_vote** Harris vote total

**trump_vote** Trump vote total

**other_vote** Other vote total

**ev_dem** Electoral votes for Harris

**ev_rep** Electoral votes for Trump

**ev_other** Electoral votes for Others

**census** Census region

## Author(s)

Kieran Healy

## Source

Vote data from Wikipedia, https://en.wikipedia.org/wiki/2024_United_States_presidential_election

`election24_county_df` *US County-level Presidential Election data, 2024*

## Description

A tibble with US presidential election data

## Usage

```
election24_county_df
```

## Format

## ‘election24_county_df‘ A tibble object with 3,153 rows and 7 columns:

**fips** County FIPS code.

**st** State name abbreviation

**votes_dem** Votes for Harris/Walz ticket.

**votes_gop** Votes for Trump/Vance ticket.

**total_votes** Total votes cast.

**winner** Winning party.

**flipped** Did the party winner change from the winner in 2020? (Yes/No)

## Details

A tibble.

## Author(s)

Kieran Healy

## Source

Election data derived from https://doi.org/10.7910/DVN/VOQCHQ

## elections_historic *US Presidential Election vote shares*

### Description

A dataset of US presidential elections from 1824 to 2024, with information on the winner, runner up, and various measures of vote share. The variables are as follows:

### Usage

```
elections_historic
```

### Format

A (tibble) data frame with 51 rows and 19 columns.

### Details

- election. Number of the election counting from the first US presidential election. 1824 is the 10th election.
- year. Year.
- winner. Full name of winner.
- win_party. Party affiliation of winner.
- ec_votes. Electoral college votes for winner.
- ec_denom. Number of votes in the electoral college.
- ec_pct. Winner's share of electoral college vote. (A proportion. Range is 0 to 1.)
- popular_pct. Winner's share of popular vote. (A proportion. Range is 0 to 1.)
- popular_margin. Winner's margin of the popular vote, expressed as a proportion. Can be positive or negative.
- votes. Total votes cast in the election.
- margin. Winner's vote margin in the popular vote.
- runner_up. Runner up candidate.
- ru_part. Party affiliation of runner up candidate.
- turnout_pct. Voter turnout as a proportion of eligible voters. (A proportion. Range is 0 to 1.)
- winner_lname Last name of winner.
- winner_label Winner's last name and election year.
- ru_lastname. Runner up's last name.
- ru_label. Runner up's last name and election year.
- two_term. Is this a two term presidency? (TRUE/FALSE.) Note that F.D. Roosevelt was elected four times.

### Source

https://en.wikipedia.org/wiki/List_of_United_States_presidential_elections_by_popular_vote_margin.

`farsinvolved` *Child Pedestrians involved in Fatal Motor Vehicle Crashes, 2009-2023*

## Description

Daily data on child pedestrians (aged 0-17 years) involved in a motor vehicle crash that resulted in a fatality.

## Usage

```
farsinvolved
```

## Format

## ‘farsinvolved‘ A data frame with 5,490 rows and 4 columns:

**month** Month (character)

**day** Day of the month (character)

**year** Year (character)

**n** Number of pedestrians

## Details

Each row is a day of the year between January 1st 2009 and December 31st 2023. The ‘n‘ column is the number of pedestrians in the United States who were involved in a motor vehicle crash that day, where the event resulted in a fatality and where the pedestrian was aged between 0 and 17 years old. The person killed is not necessarily the pedestrian.

## Author(s)

Kieran Healy

## Source

National Highway Traffic Safety Administration (NHTSA) Motor Vehicle Crash Data Querying and Reporting

## fredts — *Monetary Base and S&P 500 series*

### Description

Two time series of financial data from FRED, the _i means indexed to 100 in the base observation.

### Usage

```
fredts
```

### Format

A data frame with 5 columns and 357 rows.

### Source

FRED data.

## gss_lon — *General Social Survey data, 1972-2024*

### Description

A dataset containing an extract from the General Social Survey. See http://gss.norc.org/Get-Documentation for full documentation of the variables. This data contains many of the same variables as ‘gss_sm‘, but for all available years from 1972-2024.

### Usage

```
gss_lon
```

### Format

A data frame with 75,699 rows and 25 columns.

### Details

- year. GSS year for this respondent.
- id. Respondent id number.
- ballot. Ballot used for interview.
- age. Age of respondent.
- degree. R’s highest degree.
- race. Race of respondent.

- sex. Respondent's sex.
- siblings. Number of brothers and sisters (recoded from SIBS).
- kids. Number of children (recoded from CHILDS).
- bigregion. Region of interview (identical with REGION).
- region. Region of interview.
- income16. Total family income.
- religion. R's religious preference (recoded from RELIGION)
- marital. Marital status.
- padeg. Father's highest degree.
- madeg. Mother's highest degree.
- partyid. Political party affiliation.
- polviews. Think of self as liberal or conservative.
- happy. General happiness.
- partners_rc. How many sex partners r had in last year. (Recoded from PARTNERS)
- grass. Should marijuana be made legal.
- zodiac. Respondent's astrological sign.
- wtssall. Person weight variable (1972-2018).
- wtssps. Person weight variable (1972-2024).
- vpsu. Sampling unit
- vstrat. Stratification unit

## Source

National Opinion Research Center, http://gss.norc.org.

---

`gss_sm` *General Social Survey data, 2016*

---

## Description

A dataset containing an extract from the 2016 General Social Survey. See http://gss.norc.org/Get-Documentation for full documentation of the variables.

## Usage

```
gss_sm
```

## Format

A data frame with 2538 rows and 26 columns.

## Details

- year. gss year for this respondent.
- id. respondent id number.
- ballot. ballot used for interview.
- age. age of respondent.
- childs. number of children.
- sibs. number of brothers and sisters.
- degree. Rs highest degree.
- race. race of respondent.
- sex. respondent's sex.
- region. region of interview.
- income16. total family income.
- relig. rs religious preference.
- marital. marital status.
- padeg. fathers highest degree.
- madeg. mothers highest degree.
- partyid. political party affiliation.
- polviews. think of self as liberal or conservative.
- happy. general happiness.
- partners. how many sex partners r had in last year.
- grass. should marijuana be made legal.
- zodiac. respondents astrological sign.
- pres12. raw variable for whether the Respondent voted for Obama. Recoded to obama in this dataset.
- wtssall. weight variable.
- income_rc. Recoded income variable.
- agegrp. Age variable recoded into age categories
- ageq. Age recoded into quartiles.
- siblings. Top-coded sibs variable.
- kids. Top-coded childs variable.
- bigregion. Region variable (Census divisions) recoded to four Census regions.
- religion. relig variable recoded to six categories.
- partners_rc. partners variable recoded to five categories.
- obama. Respondent says the voted for Obama in 2012. 1 = yes; 0 = all other non-design options (Romney, other candidate, did not vote, refused, etc.)

## Source

National Opinion Research Center, http://gss.norc.org.

## int_to_year *int_to_year*

### Description

Convert an integer to a date.

### Usage

```
int_to_year(x, month = "06", day = "15")
```

### Arguments

| | |
|---|---|
| x | An integer or vector integers. |
| month | The month to be added to the year. Months 1 to 9 should be given as character strings, i.e. "01", "02", etc, and not 1 or 2, etc. |
| day | The day to be added to the year. Days should be given as character strings, i.e., "01" or "02", etc, and not 1 or 2, etc. |

### Value

A vector of dates where the input integer forms the year component. The day and month components added will by default be the 15th of June, so that tick marks will appear in the middle of the series on plots. For input, only years 0:9999 are accepted.

### Author(s)

Kieran Healy

### Examples

```
int_to_year(1960)
class(int_to_year(1960))
int_to_year(1960:1965)
int_to_year(1990, month = "01", day = "30")
```

## lawschools *US Law School Enrollments 1963-2015*

### Description

Annual enrollments in US Law Schools.

### Usage

```
lawschools
```

### Format

A tibble with 53 rows and 11 columns.

### Details

The variables are as follows:

- ay. Academic year. character.
- year. Year. integer.
- n_schools. Number of law schools. integer.
- fy_enrollment. First year enrollment. integer.
- fy_male. First year enrollment, men. integer.
- fy_female. First year enrollment, women. integer.
- jd_total. Total JD enrollment. integer.
- jd_male. Total JD enrollment, men. integer.
- jd_female. Total JD enrollment, women. integer.
- tot_enrolled. Total enrolled. integer.
- jd_llb_awarded. JD/LLB degrees awarded. integer.

### Source

American Bar Association

---

## `maunaloa` *Mauna Loa Atmospheric CO2 Concentration*

---

### Description

A subset of the co2 data in base R's [datasets] package, in a ggplot2-friendly format.

### Usage

```
maunaloa
```

### Format

A data frame with 4 columns and 271 rows.

### Source

R base datasets; Cleveland (1993).

## oecd_le — *Life Expectancy in the OECD, 1960-2023.*

### Description

Life expectancy data for individual countries.

### Usage

```
oecd_le
```

### Format

A tibble with 2,203 rows and 4 columns.

### Details

The variables are as follows:

- country. Country. (Character)
- year. Year. (Integer.)
- lifeexp. Life Expectancy at Birth, measured in years.
- is_usa. Indicator for USA or Other country.

### Source

OECD

## oecd_sum — *Life Expectancy in the OECD, 1960-2023*

### Description

Life expectancy data summary table.

### Usage

```
oecd_sum
```

### Format

A tibble with 64 rows and 5 columns.

### Details

The variables are as follows:

- year. Year. (Integer.)
- other. Life Expectancy at birth in OECD countries excluding the USA. Measured in years.
- usa. Life Expectancy at birth in the USA. Measured in years.
- diff. Difference between usa and other.
- hi_lo. Is usa above or below the oecd average?

### Source

OECD

---

## okboomer *Monthly Births in the U.S., 1933-2015*

---

### Description

Births by month, 1933-2015, with decomposition components.

### Usage

```
okboomer
```

### Format

## ‘okboomer‘ A data frame with 996 rows and 11 columns:

**date** Date in date format

**year_fct** Year as ordered factor

**month_fct** Month as ordered factor

**n_days** N of days in this month

**births** Total births in this month

**total_pop** Population

**births_pct** Births as a proportion of total population

**births_pct_day** Average daily births per million population

**seasonal** Seasonal component from an STL decomposition of ‘births_pct_day‘

**trend** Trend component from an STL decomposition of ‘births_pct_day‘

**remainder** Remainder component from an STL decomposition of ‘births_pct_day‘

### Details

Dataset originally constructed to reproduce a visualization exercise by Aaron Penne.

### Author(s)

Kieran Healy

### Source

U.S. Census Bureau.

---

`opiates` *Opiate-Related Deaths in the United States, 1999-2020*

---

### Description

State-level data on opiate related deaths in the US, from the CDC WONDER database.

### Usage

```
opiates
```

### Format

## ‘opiates‘ A tibble frame with 1,122 rows and 8 columns:

**fips** State FIPS code.

**st** State abbreviation.

**year** Year.

**deaths** N opiate-related deaths.

**crude** Crude death rate per 100,000 population.

**adjusted** Adjusted death rate.

**region** Census region.

**division_name** Census division.

### Details

Dataset is Multiple Cause of Death, 1999-2020. Standard Population: 2000 U.S. Std. Population. Rates per 100,000. Default intercensal populations for years 2001-2009. MCD ICD-10 Codes selected: T40.0 (Opium), T40.1 (Heroin), T40.2 (Other opioids), T40.3 (Methadone), T40.4 (Other synthetic narcotics), T40.6 (Other and unspecified narcotics). UCD ICD-10 Codes selected: X40-X44, X60-X64, X85, Y10-Y14.

### Author(s)

Kieran Healy

### Source

CDC WONDER, http://wonder.cdc.gov/mcd-icd10.html

`organdata` *Organ donation in the OECD*

## Description

A dataset containing data on rates of organ donation for seventeen OECD countries between 1991 and 2002. The variables are as follows:

## Usage

```
organdata
```

## Format

A (tibble) data frame with 237 rows and 21 columns.

## Details

- country. Country name.
- year. Year.
- donors. Organ Donation rate per million population.
- pop. Population in thousands.
- pop_dens. Population density per square mile.
- gdp. Gross Domestic Product in thousands of PPP dollars.
- gdp_lag. Lagged Gross Domestic Product in thousands of PPP dollars.
- health. Health spending, thousands of PPP dollars per capita.
- health_lag Lagged health spending, thousands of PPP dollars per capita.
- pubhealth. Public health spending as a percentage of total expenditure.
- roads. Road accident fatalities per 100,000 population.
- cerebvas. Cerebrovascular deaths per 100,000 population (rounded).
- assault. Assault deaths per 100,000 population (rounded).
- external. Deaths due to external causes per 100,000 population.
- txp_pop. Transplant programs per million population.
- world. Welfare state world (Esping Andersen.)
- opt. Opt-in policy or Opt-out policy.
- consent_law. Consent law, informed or presumed.
- consent_practice. Consent practice, informed or presumed.
- consistent. Law consistent with practice, yes or no.
- ccode. Abbreviated country code.

## Source

Macro-economic and spending data: OECD. Other data: Kieran Healy.

`prefix_replace` *prefix_replace*

## Description

Replace series of characters (usually variable names) at the beginning of a character vector.

## Usage

```
prefix_replace(var_names, prefixes, replacements, toTitle = TRUE, ...)
```

## Arguments

| | |
|---|---|
| `var_names` | A character vector, usually variable names |
| `prefixes` | A character vector, usually variable prefixes |
| `replacements` | A character vector of replacements for the ‘prefixes‘, in the same order as them. |
| `toTitle` | Convert results to Title Case? Defaults to TRUE. |
| `...` | Other arguments to ‘gsub‘ |

## Details

Takes a character vector (usually vector of variable names from a summarized or tidied model object), along with a vector of character terms (usually the prefix of a dummy or categorical variable added by R when creating model terms) and strips the latter away from the former. Useful for quickly cleaning variable names for a plot.

## Value

A character vector with ‘prefixes‘ terms in ‘var_names‘ replaced with the content of the ‘replacement‘ terms.

## Author(s)

Kieran Healy

## Examples

```
prefix_replace(iris$Species, c("set", "ver", "vir"), c("sat",
    "ber", "bar"))
```

`prefix_strip` *prefix_strip*

## Description

Strip a series of characters from the beginning of a character vector.

## Usage

```
prefix_strip(var_string, prefixes, toTitle = TRUE, ...)
```

## Arguments

| | |
|---|---|
| `var_string` | A character vector, usually variable names |
| `prefixes` | A character vector, usually variable prefixes |
| `toTitle` | Convert results to Title Case? Defaults to TRUE. |
| `...` | Other arguments to ‘gsub‘ |

## Details

Takes a character vector (usually vector of variable names from a summarized or tidied model object), along with a vector of character terms (usually the prefix of a dummy or categorical variable added by R when creating model terms) and strips the latter away from the former. Useful for quickly cleaning variable names for a plot.

## Value

A character vector with ‘prefixes‘ terms stripped from the beginning of ‘var_name‘ terms.

## Author(s)

Kieran Healy

## Examples

```
prefix_strip(iris$Species, c("set", "v"))
```

## round_df *round_df*

### Description

Round numeric columns of a data frame or tibble

### Usage

```
round_df(data, dig = 2)
```

### Arguments

| | |
|---|---|
| data | A data frame or tibble |
| dig | The number of digits to round to |

### Details

Takes a data frame or tibble as input, rounds the numeric columns to the specified number of digits.

### Value

An object of the same class as ‘data‘, with the numeric columns rounded off to ‘dig‘

### Author(s)

Kieran Healy

### Examples

```
head(round_df(iris, 0))
```

## states_sf *US State geometries*

### Description

US State map data layer

### Usage

```
states_sf
```

### Format

## ‘states_sf‘ A simple features object with 51 rows and 5 columns:

**fips** State FIPS code

**st** State name abbreviation

**state** State name

**census** Census region

**geometry** Geometry

### Details

A simple features object. Load the ‘sf‘ package before using. Alaska and Hawaii have had their geometries scaled and shifted to the bottom left of the map area. Alaska’s Aleutian islands are not included.

### Author(s)

Kieran Healy

### Source

US Census Bureau.

---

## studebt *Student debt data*

---

### Description

Outstanding student debts in 2016 across 8 income categories, by percent of all borrowers and percent of all balances.

### Usage

```
studebt
```

### Format

## ‘studebt‘ A data frame with 16 rows and 4 columns:

**Debt** Debt categories (character)

**type** Pct in terms of Borrowers or Balances

**pct** Percentage of all type

**Debtrc** Debt categories (ordered factor)

### Source

Federal Reserve Bank of New York.

theme_socviz *A ggplot2 theme for socviz*

## Description

A ggplot theme with defaults for axis styling, legends, panels, strips, and plot chrome. Requires ggplot2 >= 4.0.0.

## Usage

```
theme_socviz(
  base_size = 12,
  base_family = "Source Sans 3",
  header_family = "Source Sans 3",
  base_line_size = base_size/24,
  base_rect_size = base_size/24,
  ink = "black",
  paper = "white",
  accent = "#0072B2"
)
```

## Arguments

| | |
|---|---|
| base_size | Base font size in points. Default is 12. |
| base_family | Base font family. Default is ‘"Source Sans 3"‘. |
| header_family | Font family for plot titles. Default is ‘"Source Sans 3"‘. |
| base_line_size | Base line width, scaled from ‘base_size‘. |
| base_rect_size | Base rect border width, scaled from ‘base_size‘. |
| ink | Color used for text, lines, and foreground elements. Default is ‘"black"‘. |
| paper | Color used for backgrounds. Default is ‘"white"‘. |
| accent | Accent color for geom defaults. Default is ‘"#0072B2"‘. |

## Details

The theme uses Source Sans 3 (regular weight) as the base font family and Source Sans 3 Semibold as the header family. If the fonts are not installed, they will be downloaded automatically from Google Fonts via [systemfonts::require_font].

## Value

A ggplot2 theme object.

### Examples

```
## Not run:
library(ggplot2)
ggplot(mtcars, aes(wt, mpg)) +
  geom_point() +
  theme_socviz()

## End(Not run)
```

---

## theme_socviz_map *A map theme for socviz*

---

### Description

A theme based on [theme_socviz] with all axes, grids, and borders removed, suitable for plotting maps.

### Usage

```
theme_socviz_map(
  base_size = 12,
  base_family = "Source Sans 3",
  header_family = "Source Sans 3",
  base_line_size = base_size/24,
  base_rect_size = base_size/24,
  ink = "black",
  paper = "white",
  accent = "#0072B2"
)
```

### Arguments

| | |
|---|---|
| base_size | Base font size in points. Default is 12. |
| base_family | Base font family. Default is ‘"Source Sans 3"‘. |
| header_family | Font family for plot titles. Default is ‘"Source Sans 3"‘. |
| base_line_size | Base line width, scaled from ‘base_size‘. |
| base_rect_size | Base rect border width, scaled from ‘base_size‘. |
| ink | Color used for text, lines, and foreground elements. Default is ‘"black"‘. |
| paper | Color used for backgrounds. Default is ‘"white"‘. |
| accent | Accent color for geom defaults. Default is ‘"#0072B2"‘. |

### Value

A ggplot2 theme object.

## Examples

```
## Not run:
library(ggplot2)
ggplot(map_data("state"), aes(long, lat, group = group)) +
  geom_polygon(fill = "gray90", colour = "white") +
  coord_map() +
  theme_socviz_map()

## End(Not run)
```

---

`titanic` *A table of survival rates from the Titanic*

---

## Description

A small table of survival rates from the Titanic, by sex

## Usage

```
titanic
```

## Format

A data frame with four rows and four columns.

## Source

Titanic data

---

`tw_tab` *Quickly make a two-way table of proportions (percentages)*

---

## Description

Quickly make a two-way table of proportions (percentages)

## Usage

```
tw_tab(x, y, margin = NULL, digs = 1, dnn = NULL, ...)
```

### Arguments

| | |
|---|---|
| x | Row variable |
| y | Column variable |
| margin | See ‘prop.table‘. Default is joint distribution (all cells sum to 100), 1 for row margins (rows sum to 1), 2 for column margins (columns sum to 1) |
| digs | Number of digits to round percentages to. Defaults to 1. |
| dnn | See ‘table‘. the names to be given to the dimensions in the result (the dimnames names). Defaults to NULL for none. |
| ... | Other arguments to be passed to ‘table‘. |

### Details

A wrapper for ‘table‘ and ‘prop.table‘ with the margin labels set by default to NULL and the cells rounded to percents at 1 decimal place.

### Value

A contingency table of percentage values.

### Author(s)

Kieran Healy

### Examples

```
with(gss_sm, tw_tab(bigregion, religion, useNA = "ifany", digs = 1))

with(gss_sm, tw_tab(bigregion, religion, margin = 2, useNA =
    "ifany", digs = 1))
```

---

## yahoo *Yahoo Revenue and Employees*

---

### Description

Data on Revenue and Employees at Yahoo before and during Marissa Mayer’s tenure as CEO.

### Usage

```
yahoo
```

### Format

A tibble with 4 columns and 12 rows.

## Source

QZ.com

---

`%nin%` *%nin%*

---

## Description

Convenience 'not-in' operator

## Usage

```
x %nin% y
```

## Arguments

| | |
|---|---|
| x | vector of items |
| y | vector of all values |

## Details

Complement of the built-in operator `%in%`. Returns the elements of `x` that are not in `y`.

## Value

logical vecotor of items in x not in y

## Author(s)

Kieran Healy

## Examples

```
fruit <- c("apples", "oranges", "banana")
"apples" %nin% fruit
"pears" %nin% fruit
```

# Index


∗ **datasets**
- acs_poverty, 2
- acs_poverty_lon, 3
- asasec, 4
- color_table, 7
- counties_sf, 7
- county_comp, 8
- county_data, 9
- county_map, 10
- edu, 11
- election, 12
- election24, 13
- election24_county_df, 14
- elections_historic, 15
- farsinvolved, 16
- fredts, 17
- gss_lon, 17
- gss_sm, 18
- lawschools, 20
- maunaloa, 21
- oecd_le, 22
- oecd_sum, 22
- okboomer, 23
- opiates, 24
- organdata, 25
- states_sf, 28
- studebt, 29
- titanic, 32
- yahoo, 33

%nin%, 34

acs_poverty, 2
acs_poverty_lon, 3
asasec, 4

center_df, 5
color_comp, 5
color_pal, 6
color_table, 7
counties_sf, 7
county_comp, 8
county_data, 9
county_map, 10

edu, 11
election, 12
election24, 13
election24_county_df, 14
elections_historic, 15

farsinvolved, 16
fredts, 17

gss_lon, 17
gss_sm, 18

int_to_year, 20

lawschools, 20

maunaloa, 21

oecd_le, 22
oecd_sum, 22
okboomer, 23
opiates, 24
organdata, 25

prefix_replace, 26
prefix_strip, 27

round_df, 28

states_sf, 28
studebt, 29

theme_socviz, 30
theme_socviz_map, 31
titanic, 32
tw_tab, 32

yahoo, 33