

# Package ‘socviz’

July 23, 2025

**Type** Package

**Title** Utility Functions and Data Sets for Data Visualization

**Version** 1.2

**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 (>= 3.5)

**Imports** dplyr, fs, graphics, magrittr, rlang, tibble

**Suggests** ggplot2

**Encoding** UTF-8

**LazyData** true

**URL** <http://kjhealy.github.io/socviz/>

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

**RoxygenNote** 7.1.0

**NeedsCompilation** no

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

**Repository** CRAN

**Date/Publication** 2020-06-10 13:10:02 UTC

## Contents

asasec . . . . .	2
bad_date . . . . .	3
boomer . . . . .	3
center_df . . . . .	4
color_comp . . . . .	5
color_pal . . . . .	5
color_table . . . . .	6

county_data . . . . .	7
county_map . . . . .	8
edu . . . . .	9
election . . . . .	10
elections_historic . . . . .	11
fredts . . . . .	12
freq_tab . . . . .	12
gss_lon . . . . .	13
gss_sib . . . . .	14
gss_sm . . . . .	15
int_to_year . . . . .	16
lawschools . . . . .	17
lay_out . . . . .	18
maunaloa . . . . .	19
oecd_le . . . . .	19
oecd_sum . . . . .	20
opiates . . . . .	20
organdata . . . . .	21
prefix_replace . . . . .	22
prefix_strip . . . . .	23
preg . . . . .	24
preg2 . . . . .	25
round_df . . . . .	25
setup_course_notes . . . . .	26
studebt . . . . .	27
titanic . . . . .	27
tw_tab . . . . .	28
yahoo . . . . .	29
%nin% . . . . .	29
<b>Index</b>	<b>31</b>

---

asasec

---

*American Sociological Association Section Membership*


---

## Description

Membership (2005-2015) and some financial information for sections of the American Sociological Association

## Usage

asasec

## Format

A data frame with 572 rows and 9 columns.

**Source**

ASA Annual Report 2016

---

bad_date	<i>Dates in character form</i>
----------	--------------------------------

---

**Description**

A table of dates and observations with the date column stored as a character string.

**Usage**

```
bad_date
```

**Format**

A tibble with 10 rows and two columns.

**Source**

Chris Delcher.

---

boomer	<i>Monthly Births in the US, and England &amp; Wales</i>
--------	--

---

**Description**

Births by month, 1933-2015 (United States) and 1938-1991 (England & Wales)

**Usage**

```
boomer
```

**Format**

A tibble with 1,644 rows and 6 columns.

**Details**

The variables are as follows:

- date. Year and month. (Day is arbitrarily set to 01 for all observations, data are monthly.)
- month. Month of the year (1-12).
- n\_days. The number of days in a given month/year date.
- births. Total live births for that month.
- total\_pop. National population estimate for that month.
- country. United States or England & Wales.

**Source**

UK Office of National Statistics; US Census Bureau.

---

center_df	<i>center_df</i>
-----------	------------------

---

**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	<i>A table of hex color values related to types of color blindness</i>
-------------	--

---

**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

---

county\_data

*Census Data on US Counties*


---

**Description**

Selected county data (including state-level observations on some variables)

**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).
- pop\_dens4. Population density per square mile, 2014 estimate (quartiles)
- pop\_dens6. Population density per square mile, 2014 estimate (six categories)
- pct\_black. Percent black population, 2014 estimate (seven category factor)
- pop. Population, 2014 estimate
- female. Female persons, percent, 2013
- white. White alone, percent, 2013
- black. Black alone, percent, 2013
- travel\_time. Mean travel time to work (minutes), workers age 16+, 2009-2013
- land\_area. Land area in square miles, 2010
- hh\_income. Median household income, 2009-2013
- su\_gun4. Firearm-related suicides per 100,000 population, 1999-2015. Factor variable cut into quartiles. 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.
- 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.

- fips. FIPS code (integer).
- votes\_dem\_2016. Provisional count of Democratic votes in the 2016 Presidential election.
- votes\_gop\_2016. Provisional count of Republican votes in the 2016 Presidential election.
- total\_votes\_2016. Provisional count of votes cast in the 2016 Presidential election.
- per\_dem\_2016. Democratic Presidential vote, percent.
- per\_gop\_2016. Republican Presidential vote, percent.
- diff\_2016. Difference between Democratic and Republican Presidential vote.
- votes\_dem\_2012. Provisional count of Democratic votes in the 2012 Presidential election.
- votes\_gop\_2012. Provisional count of Republican votes in the 2012 Presidential election.
- total\_votes\_2012. Provisional count of votes cast in the 2012 Presidential election.
- per\_dem\_2012. Democratic Presidential vote, percent.
- per\_gop\_2012. Republican Presidential vote, percent.
- diff\_2012. Difference between Democratic and Republican Presidential vote.
- winner. Winning candidate, 2016 Presidential Election.
- partywinner16. Winning party, 2016 Presidential Election.
- winner12. Winning candidate, 2012 Presidential Election.
- partywinner12. Winning party, 2012 Presidential Election.
- flipped. Did the area flip parties from 2012 to 2016.

### 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 (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 (
- pct\_trump. Trump vote share (
- pct\_johnson. Johnson vote share (
- pct\_other. Other vote share (
- 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 variables.

**Source**

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

---

elections_historic	<i>US Presidential Election vote shares</i>
--------------------	---

---

**Description**

A dataset of US presidential elections from 1824 to 2016, with information on the winner, runner up, and various measures of vote share. Data for 2016 are provisional as of early December 2016. The variables are as follows:

**Usage**

```
elections_historic
```

**Format**

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

**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\_pct. Winner's share of electoral college vote. (Range is 0 to 1.)
- popular\_pct. Winner's share of popular vote. (Range is 0 to 1.)
- popular\_margin. Winner's point margin in the popular vote. 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. (Rate 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.
- ec\_votes. Electoral college votes cast for winner.
- ec\_denom. Total number of electoral college votes.

**Source**

[https://en.wikipedia.org/wiki/List\\_of\\_United\\_States\\_presidential\\_elections\\_by\\_popular\\_vote\\_margin](https://en.wikipedia.org/wiki/List_of_United_States_presidential_elections_by_popular_vote_margin).

---

fredts	<i>Monetary Base and S&amp;P 500 series</i>
--------	---

---

**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.

---

freq_tab	<i>Generate a tidy n-way frequency table</i>
----------	--

---

**Description**

Generate a tidy n-way frequency table

**Usage**

`freq_tab(df, ...)`

**Arguments**

<code>df</code>	tibble or data frame (implicit within pipeline)
<code>...</code>	grouping, as with <code>group_by()</code>

**Details**

Tidyverse, pipeline, and dplyr-friendly frequency tables

**Value**

A tibble with the grouping variables, the N (`'n'`) per group, and the proportion (`'prop'`) of each group, calculated with respect to the outermost grouping variable.

**Author(s)**

Kieran Healy

**Examples**

```
mtcars %>% freq_tab(vs, gear, carb)
```

gss\_lon

*General Social Survey data, 1972-2016***Description**

A dataset containing an extract from the General Social Survey. See <http://gss.norc.umd.edu/Get-Documentation> for full documentation of the variables. This data contains the same variables as 'gss\_sm', but for all available years from 1972-2016.

**Usage**

```
gss_lon
```

**Format**

A data frame with 62,366 rows and 26 variables.

**Details**

- year. gss year for this respondent.
- id. respondent id number.
- ballot. ballot used for interview.
- age. age of respondent.
- degree. Rs 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 (recoded from REGION).
- income16. total family income.
- religion. rs religious preference (recoded from RELIGION)
- 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\_rc. how many sex partners r had in last year. (Recoded from PARTNERS)
- grass. should marijuana be made legal.
- zodiac. respondents astrological sign.
- pres12. R's stated vote in the 2012 Presidential election
- wtssall. weight variable.
- vpsu. Sampling unit
- vstrat. Stratification unit

### Source

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

---

gss\_sib

*General Social Survey data, 1972-2016*

---

### 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 seven variables from 'gss\_lon' with all NA values omitted.

### Usage

gss\_sib

### Format

A data frame with 60,423 rows and 7 variables.

### Details

- year. gss year for this respondent.
- id. respondent id number.
- age. age of respondent.
- race. race of respondent.
- sex. respondent's sex.
- siblings. Number of brothers and sisters (recoded from SIBS).
- kids. Number of children (recoded from CHILDS).

### 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 variables.

**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 2o, 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

---

`lay_out`*Arrange ggplot2 plots in an arbitrary grid*

---

**Description**

Arrange ggplot2 plots in an arbitrary grid

**Usage**

```
lay_out(...)
```

**Arguments**

...                    A series lists of of ggplot objects

**Details**

The function takes arguments of the form ‘list(plot, row(s), column(s))’ where ‘plot’ is a ggplot2 plot object, and the rows and columns identify an area of the grid that you want that plot object to occupy. See <http://stackoverflow.com/questions/18427455/multiple-ggplots-of-different-sizes>

**Value**

A grid of ggplot2 plots

**Author(s)**

Extracted from the [wq] package

**Examples**

```
library(ggplot2)
p1 <- qplot(x=wt,y=mpg,geom="point",main="Scatterplot of wt vs.
  mpg", data=mtcars)
p2 <- qplot(x=wt,y=disp,geom="point",main="Scatterplot of wt vs
  disp", data=mtcars)
p3 <- qplot(wt,data=mtcars)
lay_out(list(p1, 1:2, 1:4),
  list(p2, 3:4, 1:2),
  list(p3, 3:4, 3:4))
```

---

maunaloa	<i>Mauna Loa Atmospheric CO2 Concentration</i>
----------	--

---

**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	<i>Life Expectancy in the OECD, 1960-2015.</i>
---------	--

---

**Description**

Life expectancy data for individual countries.

**Usage**

```
oecd_le
```

**Format**

A tibble with 1,746 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	<i>Life Expectancy in the OECD, 1960-2015</i>
----------	---

---

**Description**

Life expectancy data summary table.

**Usage**

oecd\_sum

**Format**

A tibble with 57 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

---

opiates	<i>Opiate-Related Deaths in the United States, 1999-2014</i>
---------	--

---

**Description**

State-level data on opiate related deaths in the US, from the CDC Wonder database. Query details: Dataset is Multiple causes of death, 1999-2014; 2006 Urbanization; Autopsy, Gender, Place of Death, States, 10-year age groups, and Hispanic Origin, Weekday, Year/Month set to ALL. Standard Population 2000 US Std Population. Default intercensal populations for years 2001-2009 except Infant age groups. Rates per 100,000 population. 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.

**Usage**

opiates

**Format**

A tibble with 800 rows and 10 columns.

**Details**

The variables are as follows:

- year. Year
- state. State name.
- fips. State FIPS code.
- deaths. Number of opiate-related deaths.
- population. Population.
- crude. Crude death rate.
- adjusted. Adjusted death rate.
- adjusted.se. Standard error of Adjusted death rate.
- region. Census region. (Stored as an ordered factor.)
- abbr. Abbreviated state name.
- division\_name. Census Division. (Character.)

**Source**

Centers for Disease Control CDC WONDER data

---

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 variables.

**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	<i>prefix_strip</i>
--------------	---------------------

---

**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"))
```

---

preg

*An untidy table of data*

---

**Description**

A table of data from Wickham (2014).

**Usage**

```
preg
```

**Format**

A `tbl_df` with 3 rows and 3 columns.

**Source**

Hadley Wickham (2014).



---

preg2	<i>A wider table of untidy data</i>
-------	-------------------------------------

---

**Description**

A second table of data from Wickham (2014).

**Usage**

```
preg2
```

**Format**

An object of class `\codetbl_df` (inherits from `\codetbl`, `\codedata.frame`) with 2 rows and 4 columns.

**Source**

Hadley Wickham (2014).

---

round_df	<i>round_df</i>
----------	-----------------

---

**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))
```

---

setup_course_notes	<i>setup_course_notes</i>
--------------------	---------------------------

---

**Description**

Copy and expand course notes to the desktop

**Usage**

```
setup_course_notes(  
  folder,  
  zipfile = "dataviz_course_notes.zip",  
  packet = "dataviz_course_notes"  
)
```

**Arguments**

folder	The destination to copy to within the user's home. This must be supplied by the user.
zipfile	The name of the zipped course materials file in the socviz library.
packet	The name of the course packet folder to be created

**Details**

Transfers a zip file containing course materials from the socviz library to the Desktop.

**Value**

The 'zipfile' is copied to 'folder' and its contents expanded into a directory, the 'packet'.

**Author(s)**

Kieran Healy

**Examples**

```
setup_course_notes()
```

---

studebt	<i>Student debt data</i>
---------	--------------------------

---

**Description**

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

**Usage**

studebt

**Format**

A tibble with 16 rows and 4 columns.

**Source**

Federal Reserve Bank of New York

---

titanic	<i>A table of survival rates from the Titanic</i>
---------	---

---

**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	<i>Yahoo Revenue and Employees</i>
-------	------------------------------------

---

**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

---

<code>%nin%</code>	<i><code>%nin%</code></i>
--------------------	---------------------------

---

**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 vector 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

- asasec, [2](#)
- bad\_date, [3](#)
- boomer, [3](#)
- color\_table, [6](#)
- county\_data, [7](#)
- county\_map, [8](#)
- edu, [9](#)
- election, [10](#)
- elections\_historic, [11](#)
- fredts, [12](#)
- gss\_lon, [13](#)
- gss\_sib, [14](#)
- gss\_sm, [15](#)
- lawschools, [17](#)
- maunaloa, [19](#)
- oecd\_le, [19](#)
- oecd\_sum, [20](#)
- opiates, [20](#)
- organdata, [21](#)
- preg, [24](#)
- preg2, [25](#)
- studebt, [27](#)
- titanic, [27](#)
- yahoo, [29](#)

[%nin%, 29](#)

asasec, [2](#)

bad\_date, [3](#)

boomer, [3](#)

center\_df, [4](#)

color\_comp, [5](#)

color\_pal, [5](#)

color\_table, [6](#)

county\_data, [7](#)

county\_map, [8](#)

edu, [9](#)

election, [10](#)

elections\_historic, [11](#)

fredts, [12](#)

freq\_tab, [12](#)

gss\_lon, [13](#)

gss\_sib, [14](#)

gss\_sm, [15](#)

int\_to\_year, [16](#)

lawschools, [17](#)

lay\_out, [18](#)

maunaloa, [19](#)

oecd\_le, [19](#)

oecd\_sum, [20](#)

opiates, [20](#)

organdata, [21](#)

prefix\_replace, [22](#)

prefix\_strip, [23](#)

preg, [24](#)

preg2, [25](#)

round\_df, [25](#)

setup\_course\_notes, [26](#)

studebt, [27](#)

titanic, [27](#)

tw\_tab, [28](#)

yahoo, [29](#)