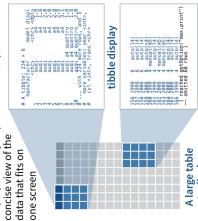
Fibbles - an enhanced data frame

ibble. Tibbles inherit the data frame S3 class for storing tabular data, the class, but improve three behaviors: The tibble package provides a new

- Subsetting [always returns a new tibble, [[and \$ always return a vector.
- No partial matching You must use full column names when subsetting
- Display When you print a tibble, R provides a concise view of the



- data frame display Control the default appearance with options: **options**(tibble.print_max = n,
 - tibble.print_min = m, tibble.width = Inf)
- · View full data set with View() or glimpse()
- Revert to data frame with as.data.frame()

CONSTRUCT A TIBBLE IN TWO WAYS

Handle Missing Values

Both make this tibble A tibble: 3 × <chr> *tibble*(x = 1:3, y = c("a", "b", "c")) Construct by columns. Construct by rows. ر", "(") tribble(~x,

as_tibble(x, ...) Convert data frame to tibble. enframe(x, name = "name", value = "value")

is_tibble(x) Test whether x is a tibble. Convert named vector to a tibble

R Studio

Fidy Data with tidyr

ridy data is a way to organize tabular data. It provides a consistent data structure across packages.

tidyr

Split Cells







× **4**

Preserves cases during vectorized operations

A table is tidy if:







case, is in its own row Each observation, or

its own column

Makes variables easy to access as vectors

separate(data, col, into, sep = "[^[:alnum:]] +", remove = TRUE, convert = FALSE, extra = "warn", fill = "warn", ...)

into individual, isolated Use these functions to

split or combine cells

*

Separate each cell in a column to make several columns.

1 2000 2K/20M 1999 37K/172M 2000 80K/174M country year rate 212K/1T 2000 213K/1T 1999 00 В Ш

spread(data, key, value, fill = NA, convert = FALSE,

drop = TRUE, sep = NULL

Use gather() and spread() to reorganize the values of a table into a new layout.

gather(data, key, value, ..., na.rm = FALSE,

convert = FALSE, factor_key = FALSE)

Reshape Data - change the layout of values in a table

*a*lues of a **value** column across the new columns.

column into the column names, spreading the

column, gathering the column values into a

single value column.

country 1999 2000

gather() moves column names into a key

spread() moves the unique values of a key

into = c("cases", "pop"))separate(table3, rate,

separate_rows(data, ..., sep = "[^[:alnum:].] +", convert = FALSE)

Separate each cell in a column to make several rows. Also separate_rows_().

212K 213K 80K

> 212K 213K

key

e m 00

₳

37K 212K

<u>ш</u> О

37K < m O

172M В O 37K/172M 80K/174M 2K/20M 212K/1T 2000 ωш

spread(table2, type, count)

key

gather(table4a, `1999`, `2000`,

key = "year", value = "cases")

separate_rows(table3, rate)

unite(data, col, ..., sep = "_", remove = TRUE)

Afghan 2000 Brazil 1999 Brazil 2000 sountry year China country century Afghan 19 Afghan Brazil Brazil China

unite(table5, century, year, China

col = "year", sep = "",

of the values of the variables listed in ... expand(data, ...) Adds to the data missing combinations of the

replace_na(data, replace = list(), ... fill(data, ..., .direction = c("down", "up"))

Fill in NA's in ... columns with most recent non-NA values.

Drop rows containing

drop_na(data, ...)

NA's in ... columns.

Replace NA's by column.

1

replace_na(x, list(x2 = 2))

fill(x, x2)

 $drop_na(x, x2)$

Collapse cells across several columns to make a single column.

Create new tibble with all possible combinations Expand Tables - quickly create tables with combinations of values

RStudio[®] is a trademark of RStudio, Inc. + CC BY SA RStudio • info@rstudio.com • 844 448-1212 • rstudio.com • Learn more at <u>tidyverse.org</u> • readr 1.1.0 • tibble 1.2.12 • tidyr 0.6.0 • Updated: 2017-01

expand(mtcars, cyl, gear, carb)

values of the variables listed in ... complete(mtcars, cyl, gear, carb)

complete(data, ..., fill = list())