Data reshaping

The `separate` function

Structure of the `separate` function

separate takes one column and turns it into two or more columns.

```
name_of_dataset %>%
  separate(
    column_two,
    into = combine("category", "values"),
    sep = "/"  # sep must be changed for data at hand!
  )
```

- The first input is the name of the column that needs to be separated into multiple columns

- The `into =` input contains a vector, built using the `combine` function, that names the columns you end up with after everything is separated.

- The `sep =` input is used to specify the symbols that separate the data within the column.

Untidy data example

Problem: Untidy data frame stored in `table3`

country	year	rate
Afghanistan	1999	745/19987071
Afghanistan	2000	2666/20595360
Brazil	1999	37737/172006362
Brazil	2000	80488/174504898
China	1999	212258/1272915272
China	2000	213766/1280428583

Untidy data example

> **Goal:** Use `separate` to transform `table3` back to this:

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

`separate` schematic

`separate` takes one column and turns it into two or more columns.
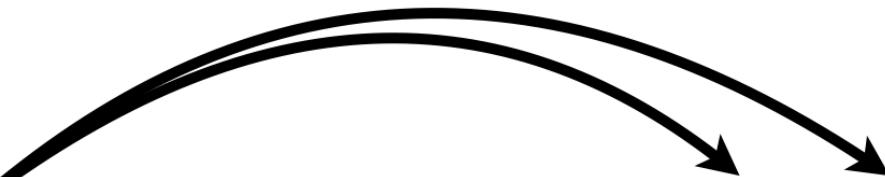
table3

Source: Figure 12.4 in *R for Data Science* by Garrett Grolemund and Hadley Wickham.

separate example

```
table3 %>%
  separate(
    col = rate,
    into = combine("cases", "population")
  )
```

separate example

```
table3 %>%
  separate(
    col = rate,
    into = combine("cases", "population")
  )
```

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

separate example

```
table3 %>%
  separate(
    col = rate,
    into = combine("cases", "population"),
    sep = "/",                   # Set the separating symbol
  )
```

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

separate example

```
table3 %>%
  separate(
    col = rate,
    into = combine("cases", "population"),
    sep = "/",                # Set the separating symbol
    convert = TRUE            # Convert data types, ensures that
                             #   cases and population cols are numeric
  )
```

country	year	cases	population
Afghanistan	1999	745	19987071
Afghanistan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

Credits

License

Acknowledgments

Content adapted from *R for Data Science* by Garrett Grolemund and Hadley Wickham, chapter 12, made available under the CC BY-NC-ND 3.0 license.