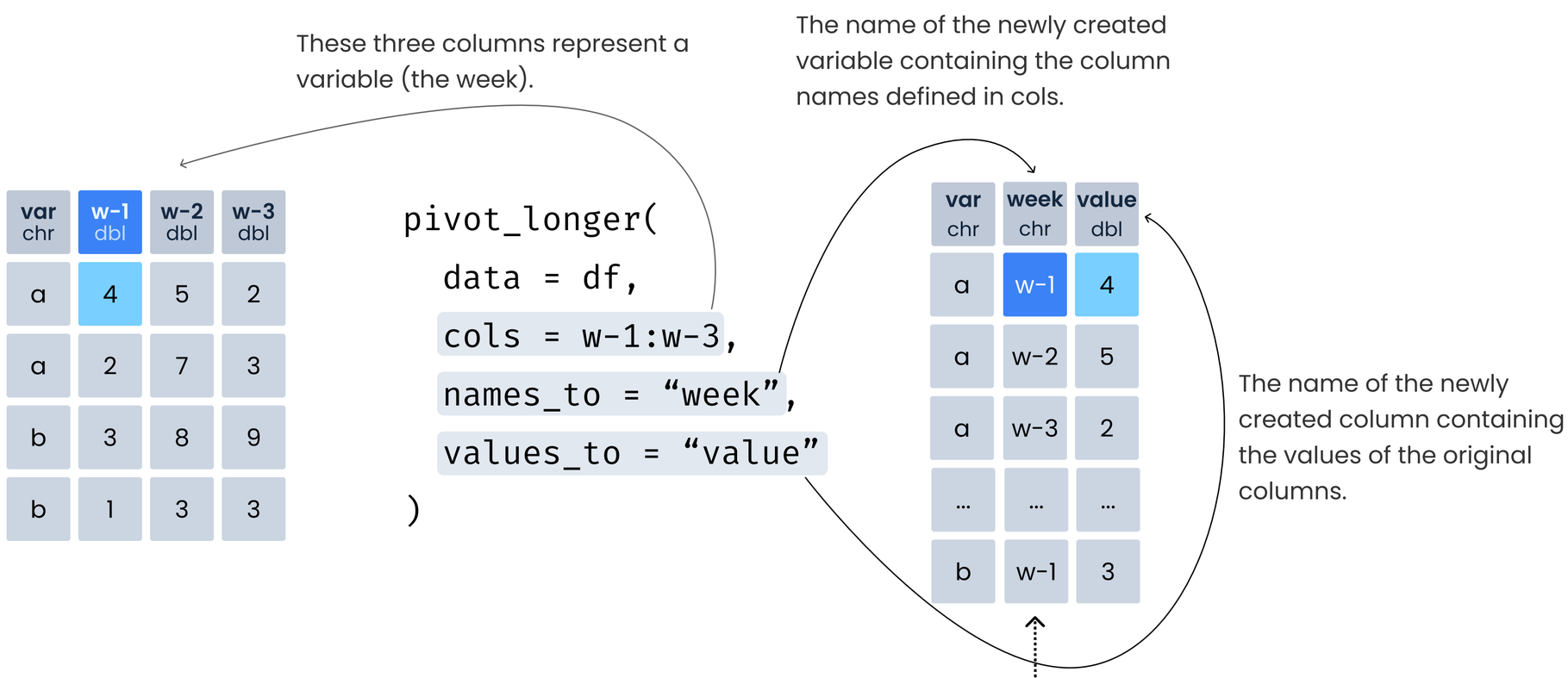
THE ULTIMATE GUIDE TO pivot_longer

TIDYVERSE CHEATSHEETS

Dataframes are often created to facilitate data entry or data comparison. This often leads to untidy data where the columns do not represent variables. pivot_longer helps to tidy such dataframes. Ultimately, the function makes a dataframe longer by increasing its number of rows and decreasing its number of columns.

Tidy a dataframe by making it longer

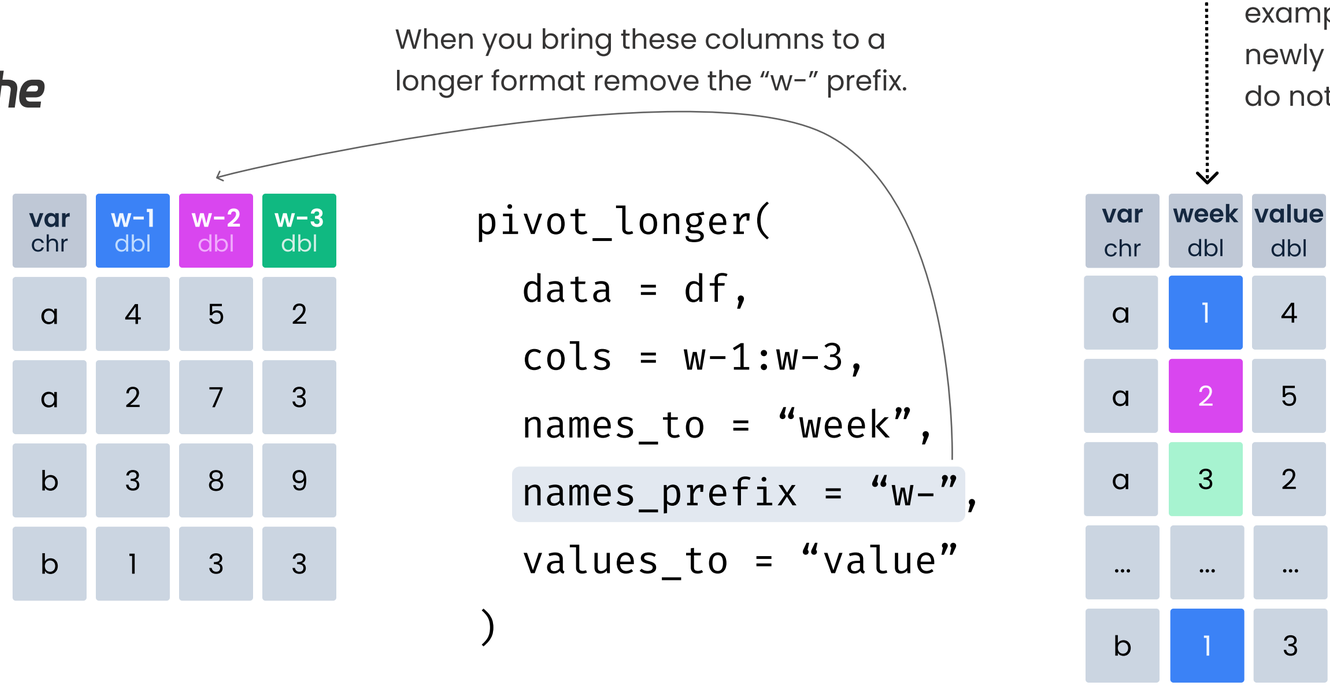
Sometimes the values of a variable are stored in columns. For example, this data set has three variables: var, the week and the value of the week. We can see these three variables if we make the dataframe longer (more rows, less columns).



Compared to the previous example, the values of the newly created variable week do not have a prefix "w-".

Remove a prefix from the column names

When you put a dataframe into a longer format, you sometimes want to remove a prefix from the column names.



THE ULTIMATE GUIDE TO pivot_longer

TIDYVERSE CHEATSHEETS

Remove a suffix from the column names

When you put a dataframe into a longer format, you sometimes want to remove a suffix from the column names.

var chr	1-w dbl	2-w dbl	3-w dbl
a	4	5	2
a	2	7	3
b	3	8	9
b	1	3	3

When you put these columns into the longer format, remove the "-w" suffix.

To understand what's going on here, you need to learn about regular expressions. Here is a good website for self-study: https://stringr.tidyverse.org/articles/regular-expressions.html

```
pivot_longer(
  data = df,
  cols = w-1:w-3,
  names_to = “week”,
  names_pattern = “(.*)-w$”,
  values_to = “value”
)
```

var chr	week dbl	value dbl
a	1	4
a	2	5
a	3	2
…	…	…
b	1	3

Separate the new column into multiple columns

The column names define multiple variables that should be separated in the longer format of the dataframe.

var chr	w-1 dbl	w-2 dbl	w-3 dbl
a	4	5	2
a	2	7	3
b	3	8	9
b	1	3	3

Create two columns based on these columns names.

Separate the two columns based on this separator: “-”

```
pivot_longer(
  cols = w-1:w-3,
  names_to = c(“prefix”,
               “week”),
  names_sep = “-”,
  values_to = "value"
)
```

var chr	prefix chr	week dbl	value dbl
a	w	1	4
a	w	2	5
a	w	3	2
…	…	…	…
b	w	1	3

Remove rows where the value is NA

Lines with NAs should not be included in the longer format of the dataframe.

var chr	w-1 dbl	w-2 dbl	w-3 dbl
a	4	5	2
a	2	7	3
b	NA	8	9
b	1	3	3

```
pivot_longer(
  cols = one:three,
  names_to = “week”,
  values_to = “value”,
  values_drop_na = TRUE
)
```

var chr	week chr	value dbl
a	w-1	4
a	w-2	5
a	w-3	2
…	…	…
b	w-1	3

Remove the row in the long format where this value is NA.

THE ULTIMATE GUIDE TO pivot_longer

TIDYVERSE CHEATSHEETS

The columns define more than one variable

Sometimes column names define several variables that should be kept separate in the longer format of the dataframe.

var chr	one1 dbl	one2 dbl	two1 dbl	two2 dbl
a	4	5	2	1
a	2	7	3	4
b	3	8	9	7
b	1	3	3	9

The brackets indicate the groups you want to keep as columns. Between the brackets you specify regular expressions.

`.value` is a placeholder for the columns to be created based on the pattern in `names_pattern`.

```
pivot_longer(
  cols = one1:two2,
  names_to = c(".value",
               "week"),
  names_pattern = "(.{3})(.{1})"
)
```

var chr	week dbl	one dbl	two dbl
a	1	4	2
a	2	5	1
a	1	2	3
...	...	...	...
b	1	3	9

Convert the data type of the variable that holds the values

The values of the new variable should be converted to a specific data type.

var chr	w-1 dbl	w-2 dbl	w-3 dbl
a	4	5	2
a	2	7	3
b	3	8	9
b	1	3	3

```
pivot_longer(
  cols = one:three,
  names_to = "week",
  values_to = "value",
  values_transform = as.character
)
```

Change the data type of the new column that holds the values.

var chr	week dbl	value chr
a	w-1	4
a	w-2	5
a	w-3	2
...	...	...
b	w-1	3

Convert the data type of the variable that holds the names

The values of the new name variable should be converted to a specific data type.

var chr	w-1 dbl	w-2 dbl	w-3 dbl
a	4	5	2
a	2	7	3
b	3	8	9
b	1	3	3

```
pivot_longer(
  cols = one:three,
  names_to = "week",
  names_transform = as.factor,
  values_to = "value"
)
```

Change the data type of the new column that holds the name of the variable.

var chr	week fct	value chr
a	w-1	4
a	w-2	5
a	w-3	2
...	...	...
b	w-1	3