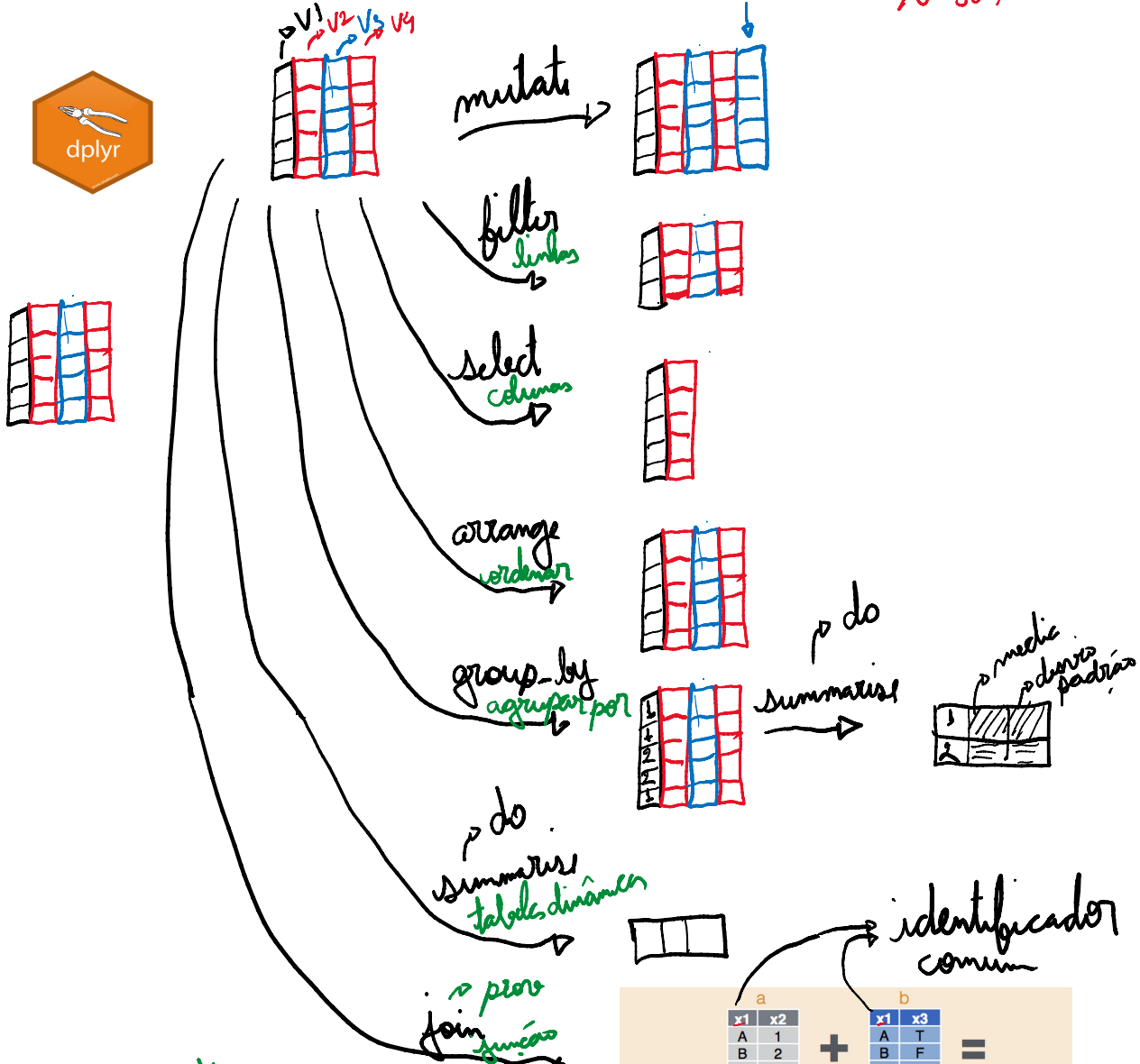


Introdução ao software R

↳ pacote dplyr ~ Hadley Wickham

organizar - arrumar - corrigir - limpar } arrumar os dados
 60-90%
 análise - resumos - processar } análise de dados
 90-95%



sum, mean, sd, lag, lead
 slice, first, last, nth
 filtro por posição / índice
 acesso primeiro
 acesso último
 acesso enésimo

identificador comum

x1	x2	x3
A	1	T
B	2	F
C	3	NA

+

x1	x3	x2
A	T	1
B	F	2
D	T	NA

=

mutating Joins

dplyr::left_join(a, b, by = "x1")
 Join matching rows from b to a.

dplyr::right_join(a, b, by = "x1")
 Join matching rows from a to b.

dplyr::inner_join(a, b, by = "x1")
 Join data. Retain only rows in both sets.

dplyr::full_join(a, b, by = "x1")
 Join data. Retain all values, all rows.

dialecto tidyverse ^{o pipe} %>%

resultados ← dados %>%

mutate(...) %>%

filter(...) %>%

summarise(...) %>%

select(...) %>%

mutate(...)

} → sequência de
procedimentos
em cadeia