

8

Wrangling

6. verb

1) arrange

2) filter

3) select

4) mutate

5) summarize

6) group-by

→

1) abc

2 NA 5

2 4 8

1 3 5

1 4 NA

4) mutate :d-axis 3

abc d

1 3 5 3

1 4 NA 3

2 NA 5 6

2 4 8 6

2) filter a == 1

abc

1 3 5

1 4 NA

1 NA

2 7

3) select a, c

a c

1 5

1 NA

2 8

5) 6) group-by (a == 1, 2)

mutate s = sum(b)

ab c s

1 3 5 7

1 4 NA 7

2 NA 5 4

2 4 8 4

原形 : abc
1 3 5
1 4 NA
2 NA 5
2 4 8

* * : the order of operations matter

Symbols : == equal

!= not equal to

Storing Results.

>, >= greater, and greater or ==

<, <= ! : ---

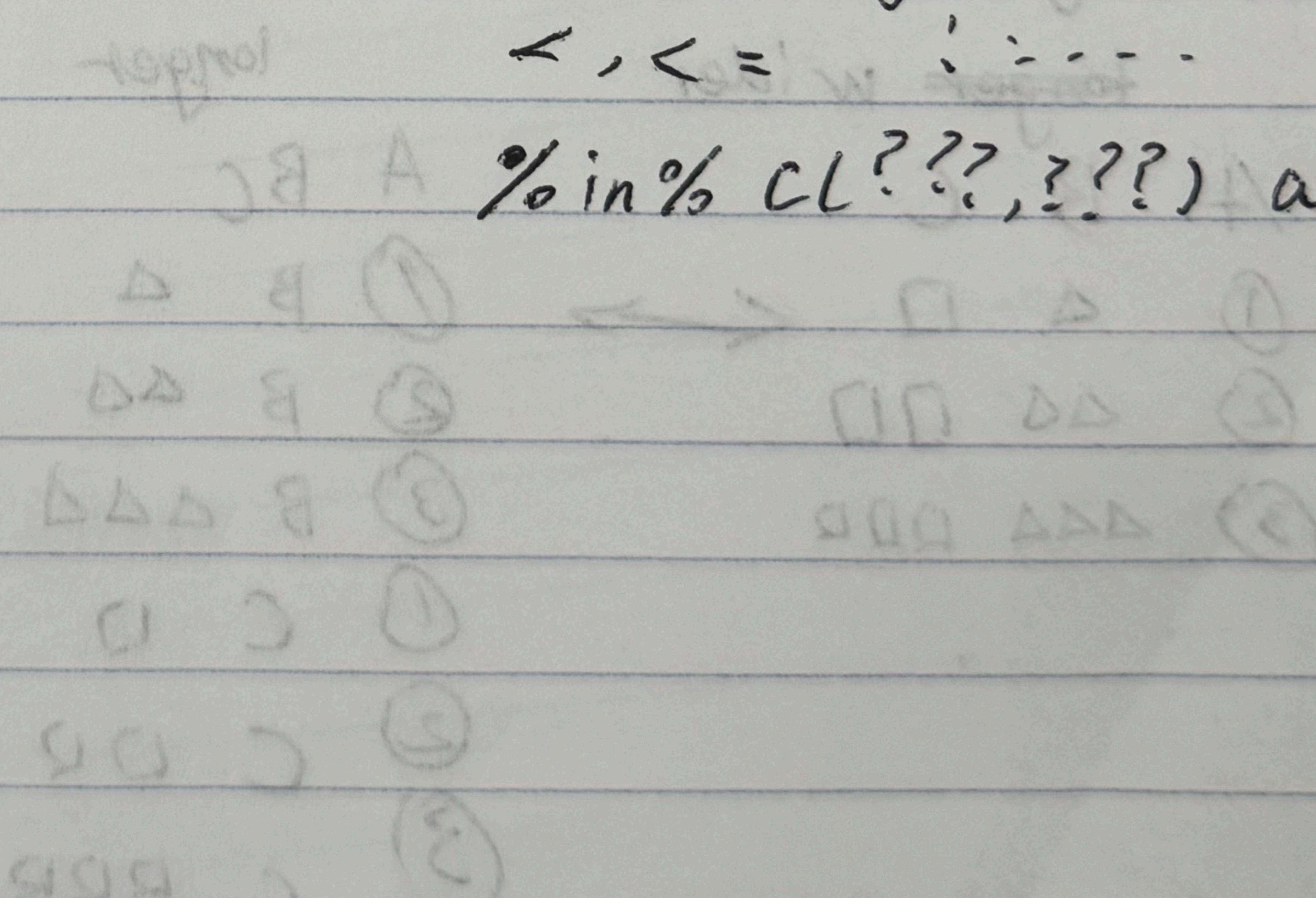
* *, <- * * (>)

%in% CL???, ???) a list of values

---> w - to vi

select(...)
eg: filter(...)

head(* * ,)



at 1 to small at top : most common

2nd row samples get in 2nd row : most - esculer

9.

Date

as.Date(today()) # get today's date

today <- as.Date(today()) # store today's date
class(today)

year, month, ↑ same

month(today, label = TRUE)

week(today) # show the week

mday

10. Reshaping

1) aggregate data: group_by() and sum() have another column of data aggregate information

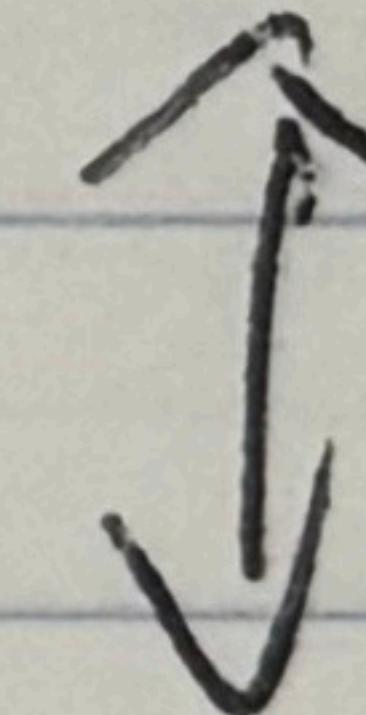
2) raw data, reshaped: retain all information

Wider and Longer formats

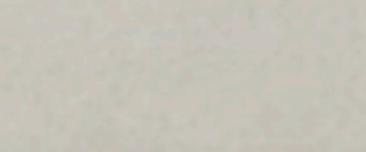
wider > longer

	A	B	C
①	△	□	
②	△△	□□	
③	△△△	□□□	

Pivot - longer (data ...)



Pivot - wider



* name_from : get the name of Data

values_from : Values of the column values

Joining

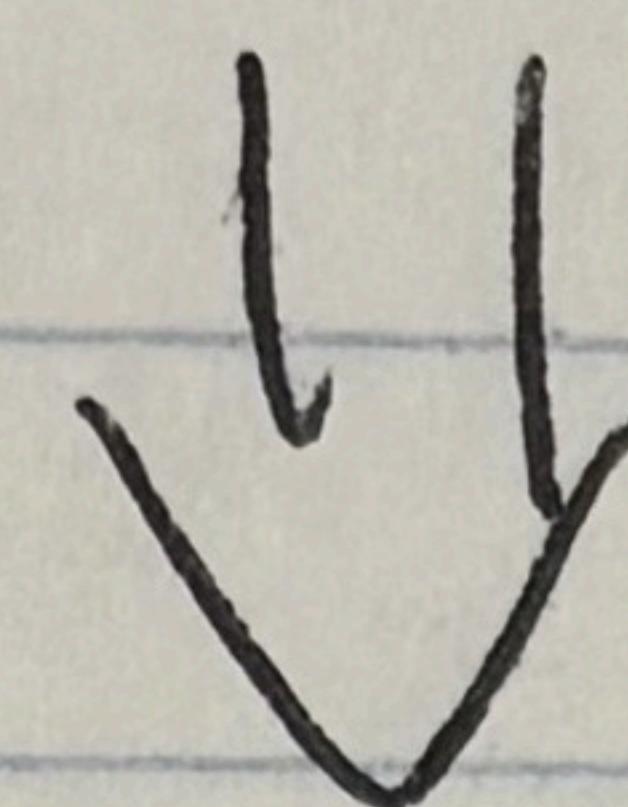
1) Left-join

students - 1

	s.	class
1.	A	①
2.	B	②
3.	C	③

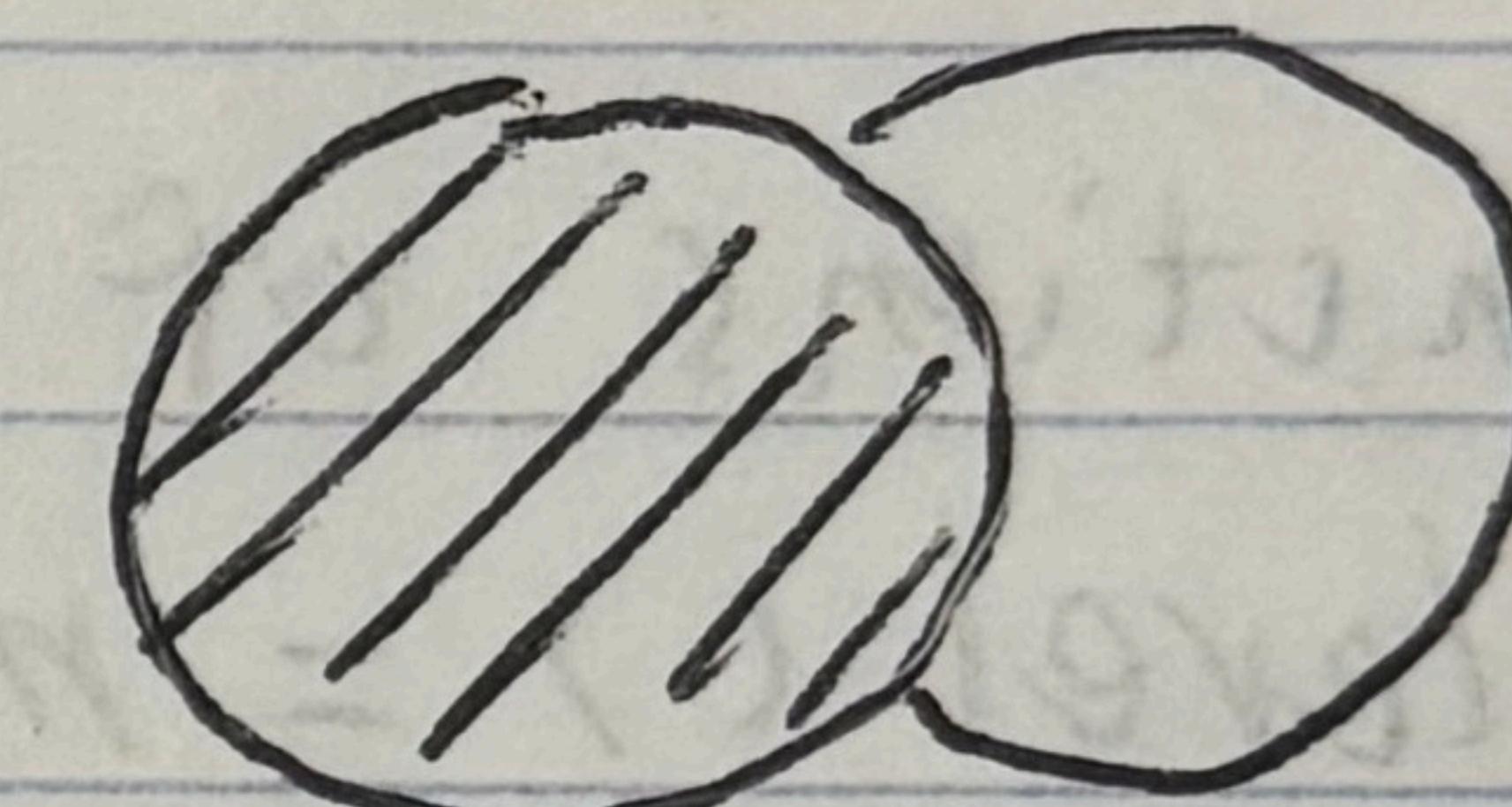
enrollment - 1

	class	enrollment
	①	18
	②	17
	③	24



students - 1 |>

left-join (enrollments - 1)



s . c . e

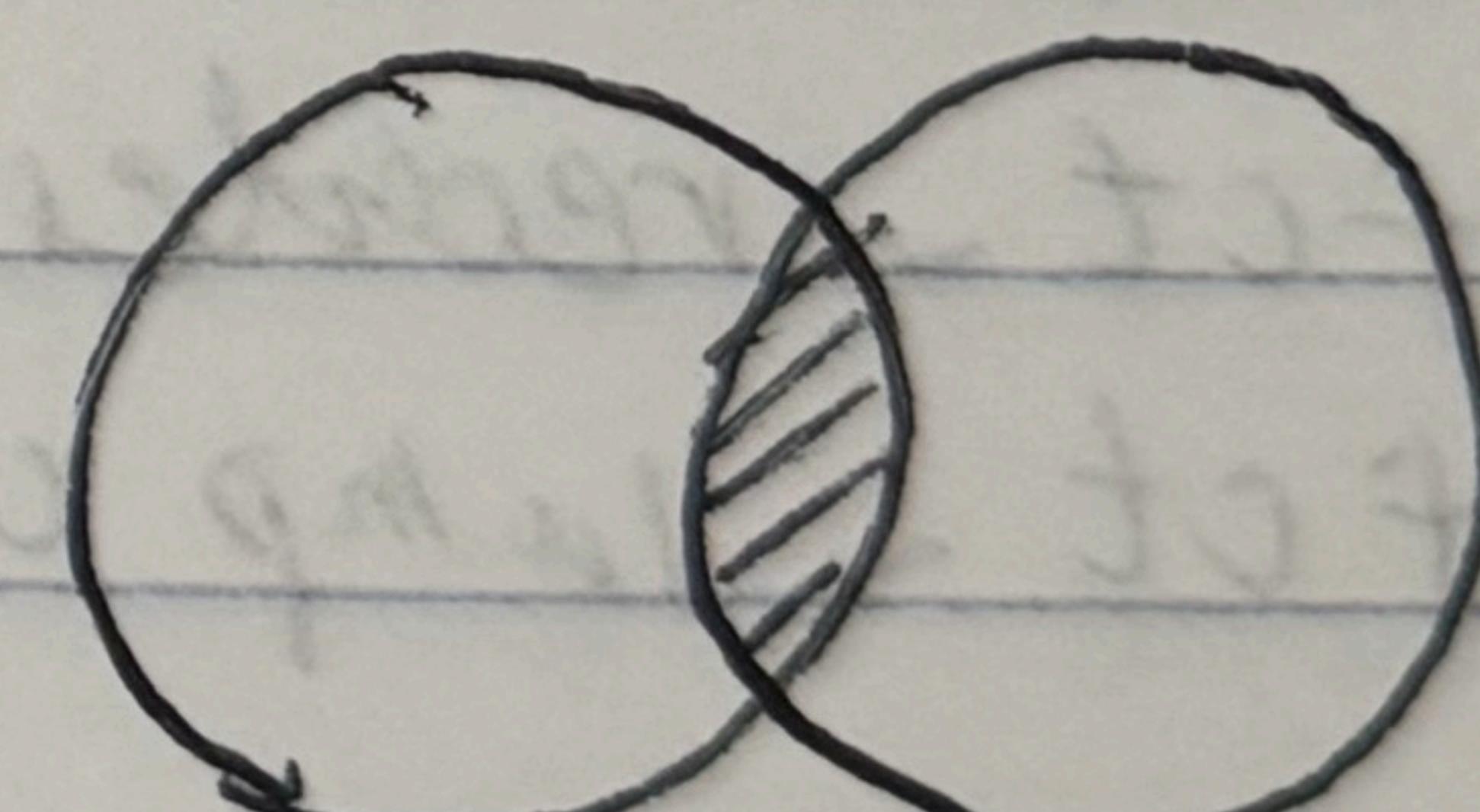
	s	c	e
1.	A	①	18
2.	B	②	17
3.	C	③	NA

(这个 inner-join，只会合并有的数据)

2) inner-join : inner-join (enrollments - 1)

s . c . e

1	A	①	18
2	B	②	24



3) full-join : full-join (enrollments)

s . c . e

1.	A	①	18
2.	B	②	17
3.	C	③	NA
4.	NA	⑤	24



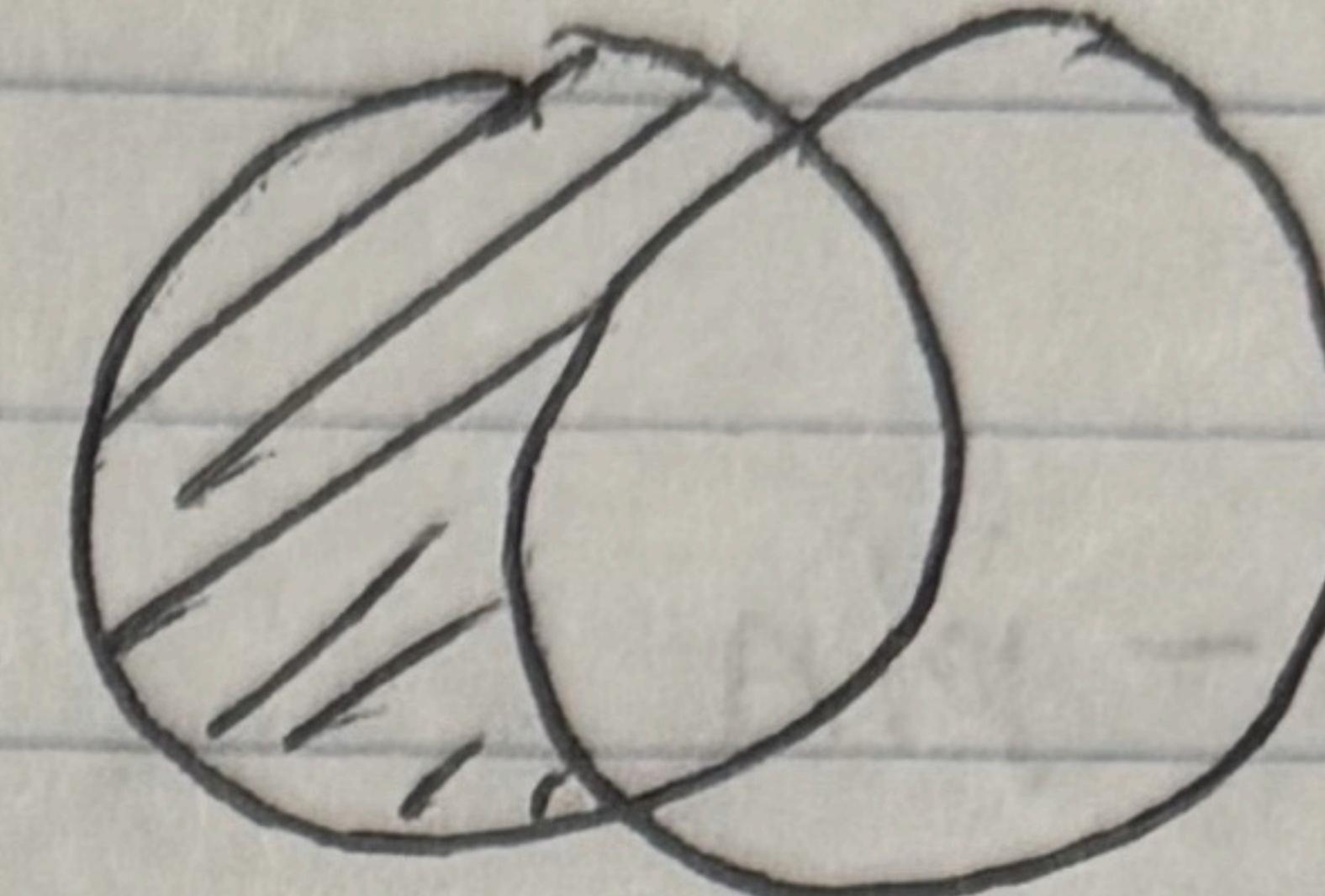
filtering join

semi-join

S . C
1. A ①
2. B ②

anti-join

S C
C ③



semi-join : Discards any observations in the left table that do not have a match in the right table

anti-join : Discards any observations in the left that don't have a match in the right table.

12. Factors

1) functions of changing the order of factor levels

fct_relevel() = manually reorder levels

fct_reorder() = reorder levels according to values of another variable

fct_infreq() = order levels from highest to lowest frequency

fct_rev() = reverse the current order

2) functions for changing the labels or values of factor levels

fct_recode() = manually change levels

fct_lump() = group together least common levels

13. STR

- 1) Str-replace(x, pattern, replacement) } the first x → replacement
- 2) Str-replace-all(x, pattern, replacement) } all
- 3) str-to-lower(x) convert all upper case letter x to lower
- 4) str-sub(x, start, end)
- 5) ... - length(x)
- 6) ... - detect(x, pattern) = # to x, True / False

14. Data import

2 - kinds of