

5. Effective vis

1) Facet : split data into smaller subsets for visualization

```
ggplot(data, aes(...)) +
```

```
geom_... +
```

```
facet_wrap(~ group-variable)
```

vis.

→ lab(x=..., y=...,
c=...)

Basic principles: 1.) professionalism : meaningful axis is table

2.) Accessibility : have alt text → #/ fig-cap: ?

3.) Design details

4.) Ethics

#/ fig-alt: "

Data Wrangling

1. Filter: rows based on conditions

```
closed_data <- data |>
```

```
filter(V, )
```

2. Mutate : create new v.

```
mutate(new-variable = Vx, ...)
```

3. arrange

4. select

5. summarize

6. group-by

= = equal to

!= not equal

> greater

>=

<

<=

%in% c(?, ?)

a list of multiple values

Date: V-today's date

as.Date(today())

year

month

month(today, label = Tmp)

8

Wrangling

6. verb

1) arrange

2) filter

3) select

4) mutate

5) summarize

6) group-by

 \rightarrow

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

a c

1 5

1 NA

2 7

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 - favor

toprow

D B ① → D A ①

D D B ② → D D D ③

D D D B ③ → D D D D ④

D D D D B ④ → D D D D D ⑤

D D D D D B ⑤ → D D D D D D ⑥

D D D D D D B ⑥ → D D D D D D D ⑦

D D D D D D D B ⑦ → D D D D D D D D ⑧

D D D D D D D D B ⑧ → D D D D D D D D D ⑨

D D D D D D D D D B ⑨ → D D D D D D D D D D ⑩

head (* *,)

90 3 ⑤

900 3 ⑥

9000 3 ⑦

at 8 to small out top : most common

2nd by samples out in 2nd : most - samples

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

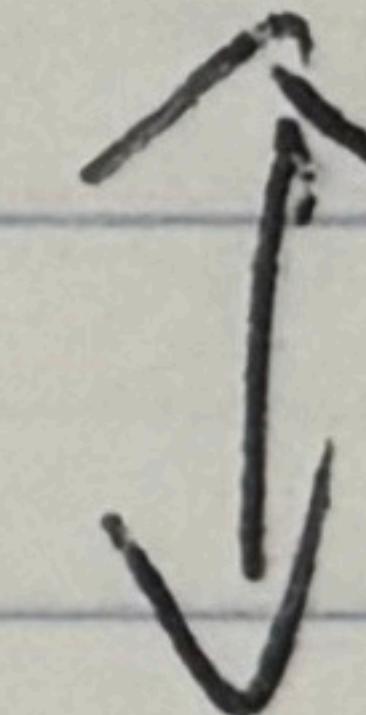
2) raw data, reshaped: retain all information

Wider and Longer formats

wider > longer

	A	B	C
①	△	□	
②	△△	□□	
③	△△△	□□□	

Pivot - longer (data ...)



Pivot - wider

	A	BC
①	B	△
②	B	△△
③	B	△△△

* name_from : get the name of Data

values_from : Values of the column values

Joining

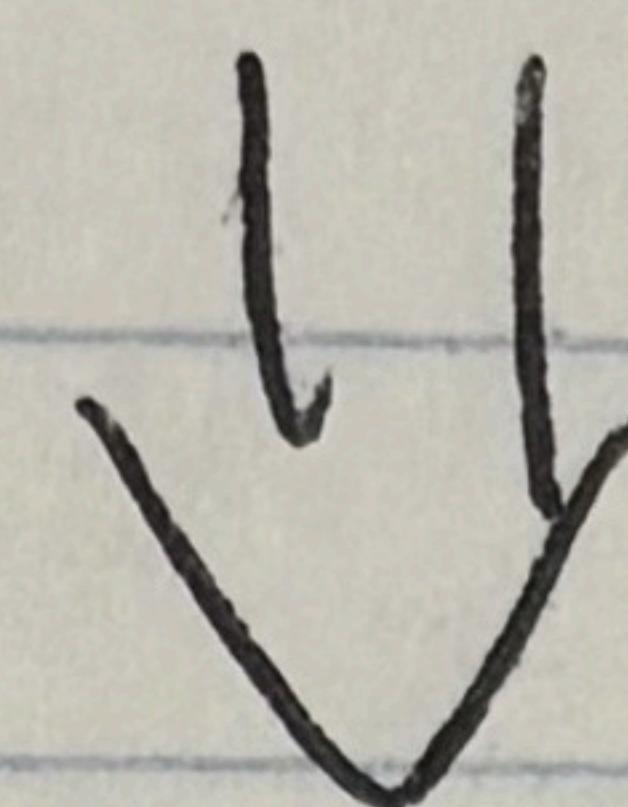
1) Left-join

students - 1

	s.	class
1.	A	①
2.	B	②
3.	C	③

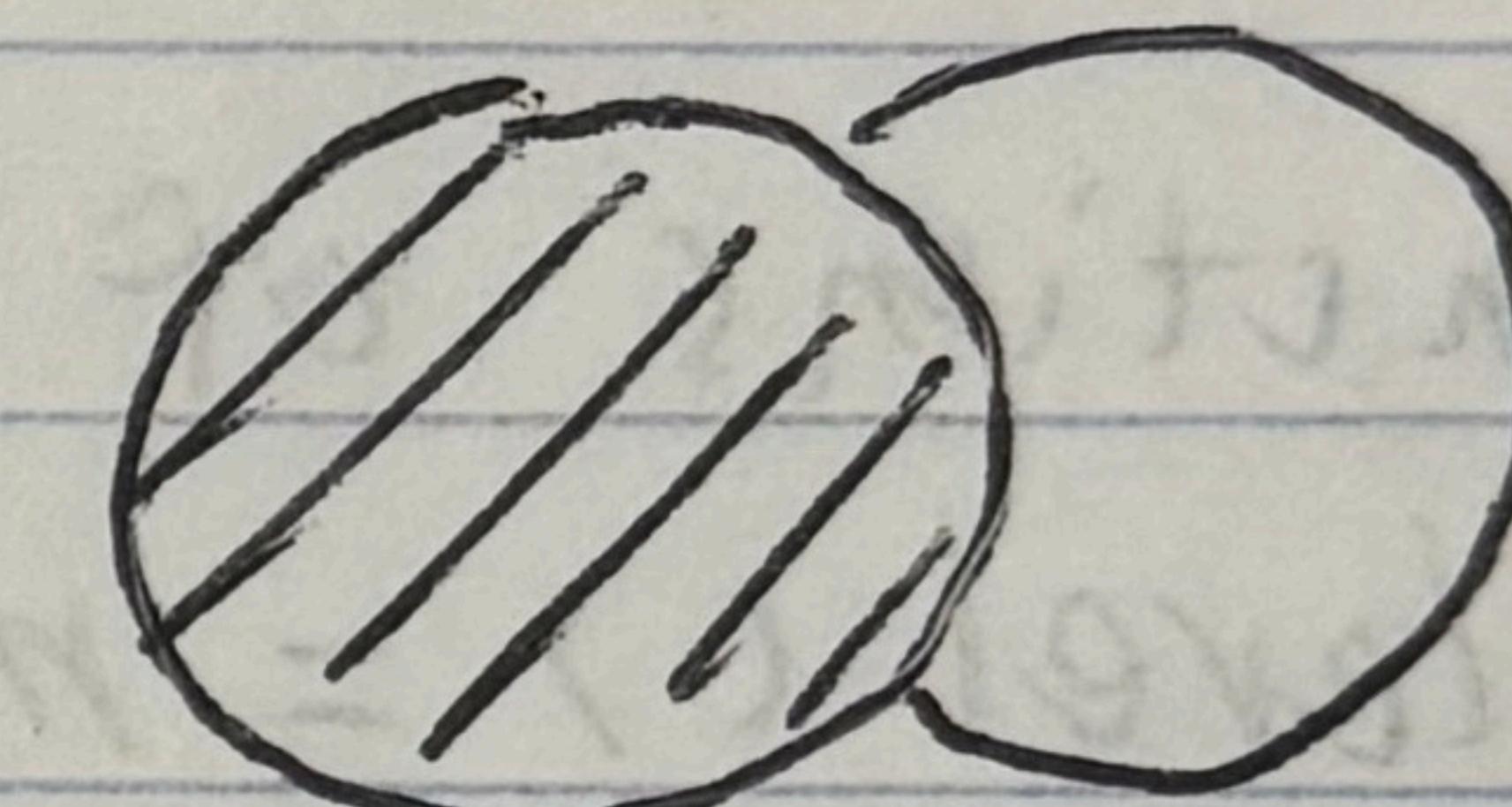
enrollment - 1

	class	enrollment
	①	18
	②	17
	③	24



students - 1 |>

left-join (enrollments - 1)



s . c . e

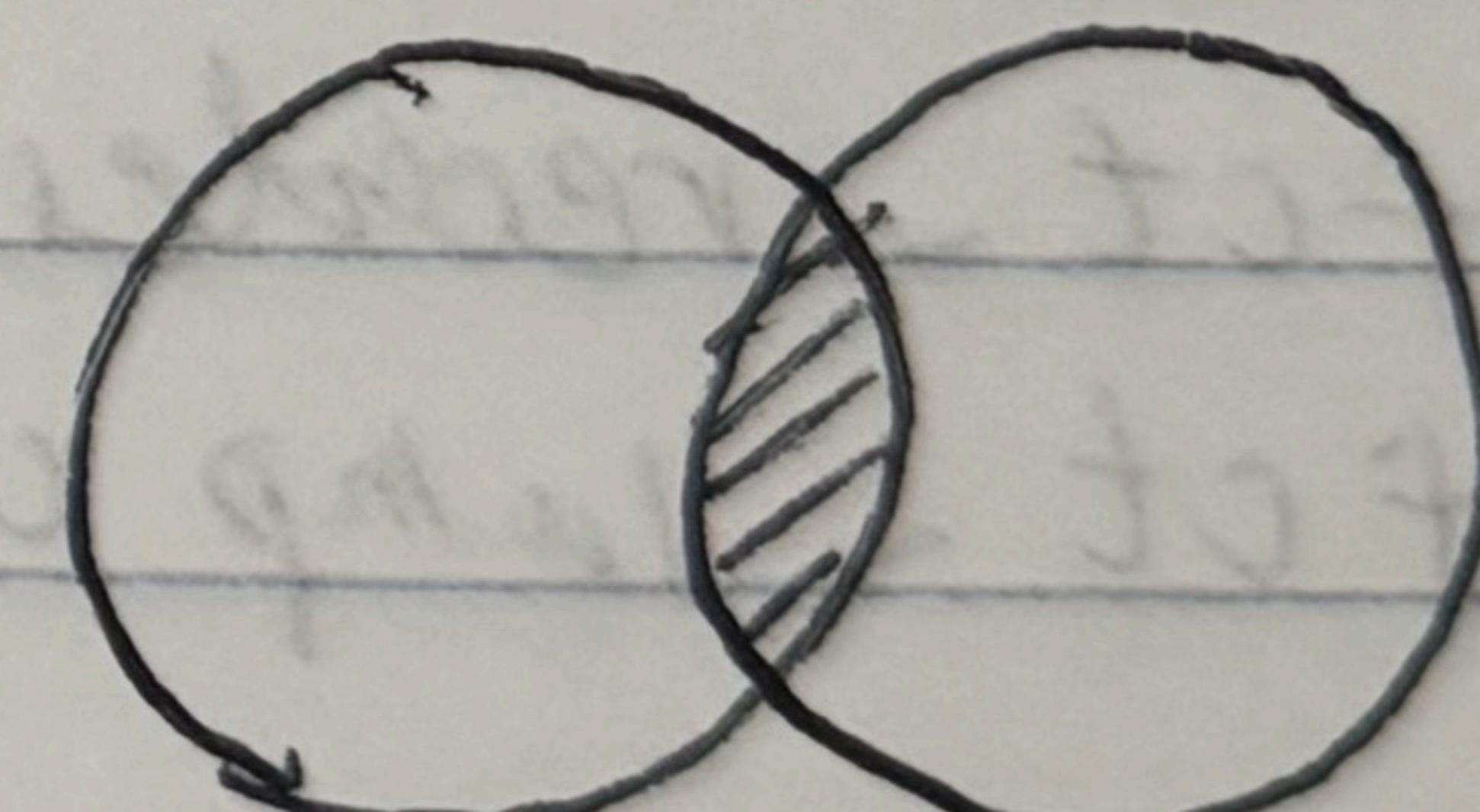
	s .	c .	e
1.	A	①	18
2.	B	②	17
3.	C	③	NA

(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

