

Unrangling Verbs

verb	action	code Example
Arrange	Arrange rows according to some column	elections > arrange(desc(votes))
filter	filter out subset of rows	elections_small > filter(state_name %in% c("Hawaii"))
Select	Select a subset of columns	elections > select(state_name)
Mutate	Mutate / create column	elections > mutate(diff = repub - dem)
Summarize	Calculate numerical summary of column	elections > summarize(median_rep = median(repub))
Group by	group rows by a specified column	election > group_by(state_name) > summarize(total_votes = sum(votes))

Dealing with NAs

Penguins |> summarize(mean = mean(mass, na.rm=TRUE))

Penguins |> filter(!is.na(body-mass))

Penguins |> na.omit } removes na from any variable

Dates

Syntax

today <- as.Date(today())

Year(today)

Month(today) month(today, label=TRUE)

Week(today)

Day(today) day of the month

Yday(today) day of the year

Wday(today) day of the week

Reshaping

Make data longer/wider

Country	1999	2000
A	0	2
B	1	80
C	2	200

collapsing columns to rows

first_longer(data, cols=2:3, names_to="year", values_to="case")

Country	year	type	Count
A	1999	case	1
A	1999	pop	2
A	2000	case	3
B	1999	pop	4
B	1999	case	5
B	2000	pop	6
C	2000	case	7
C	1999	pop	8
C	1999	case	9

spread across new variables

first_wider(data, names_from="type", values_from="Count")

Joining

id	x
1	a
2	b

id	y
2	1000
3	1500

left-join

id	x	y
1	a	NA
2	b	1000

inner-join

id	x	y
2	b	1000

full-join

id	x	y
1	a	NA
2	b	1000
3	NA	1500

Data Wrangling

factors

fct_relevel() = reorder levels

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

fct_interc() = order levels from highest to lowest freq

fct_rev() = reverse the current order

fct_recode() = manually change levels

fct_hmpt() = group together common levels

strings

function	argument	returns
str_replace()	x, pattern, replacement	modified string
str_replace_all()	x, pattern, replacement	"
str_lower()	x	"
str_sub()	x, start, end	"
str_length()	x	#
str_detect()	x, pattern	True/False

C	2	1
5	NA	3
8	4	NA
NA	NA	4

Pivot wider

a	b	C
2	NA	5
2	4	8
1	3	5
1	4	NA

arrange(desc(C))

a	C
1	5
1	NA
2	7
2	8

select(a,c)

a	b	C
1	3	5
1	4	NA
2	4	8

mutate(a=2a)

a	b	C
2	3	5
2	4	8

na.omit

groupby

a	b	C
1	3	5
1	4	NA
2	4	8

mutate()

a	b	C
1	3	5
1	4	NA
2	4	8

filter(a=1)

a	b	C
1	3	5
1	4	NA

groupby(b)

a	b	C
1	3	5
1	4	NA
2	4	8

summarize(sum(b))

a	b	C
1	7	
2	4	

first_longer

C	n	v
5	2	a
8	3	b
NA	1	c