



# BeginneR Session

## -- Nested data handling --

@kilometer00

Who ! ?

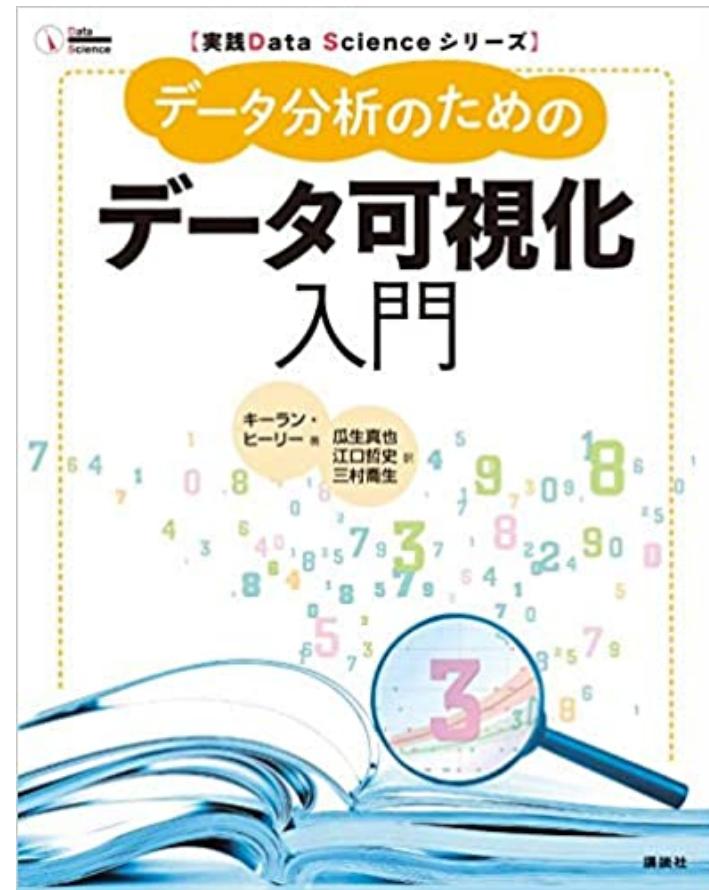
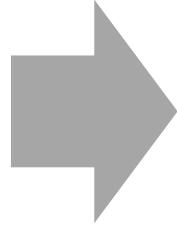
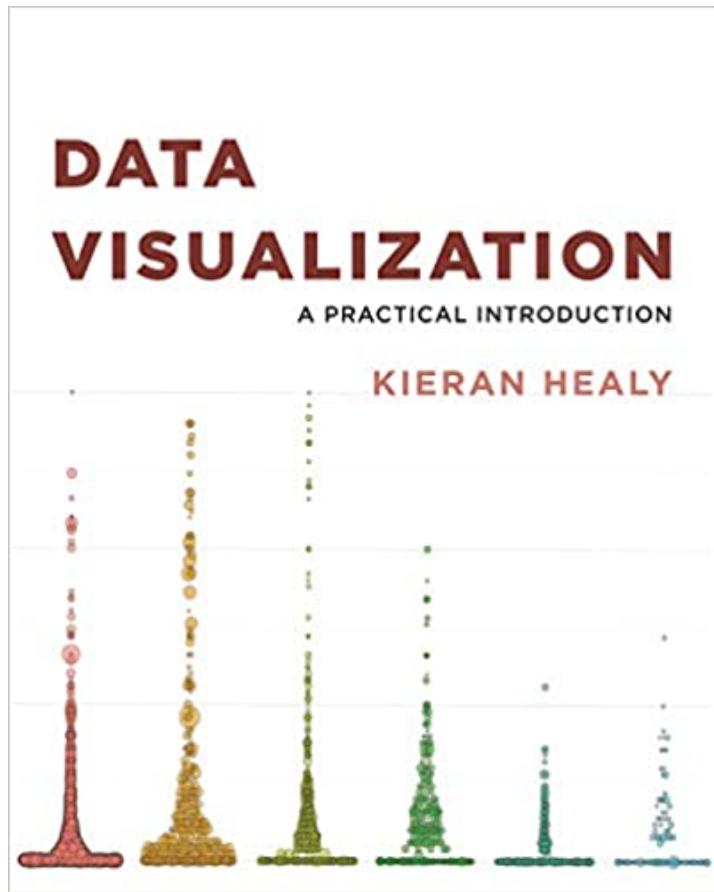


# Who ! ?

- @kilometer
- Postdoc Researcher (Ph.D. Eng.)
- Neuroscience
- Computational Behavior
- Functional brain imaging
- R : ~ 10 years



宣伝!! (書籍の翻訳に参加しました。)



絶賛販売中！



# BeginneR Session



BeginneR



BeginneR



Advanced



Hoxo\_m

If I have seen further it is by standing on the shoulders of Giants.

-- Sir Isaac Newton, 1676

# BeginneR Session



BeginneR

Before

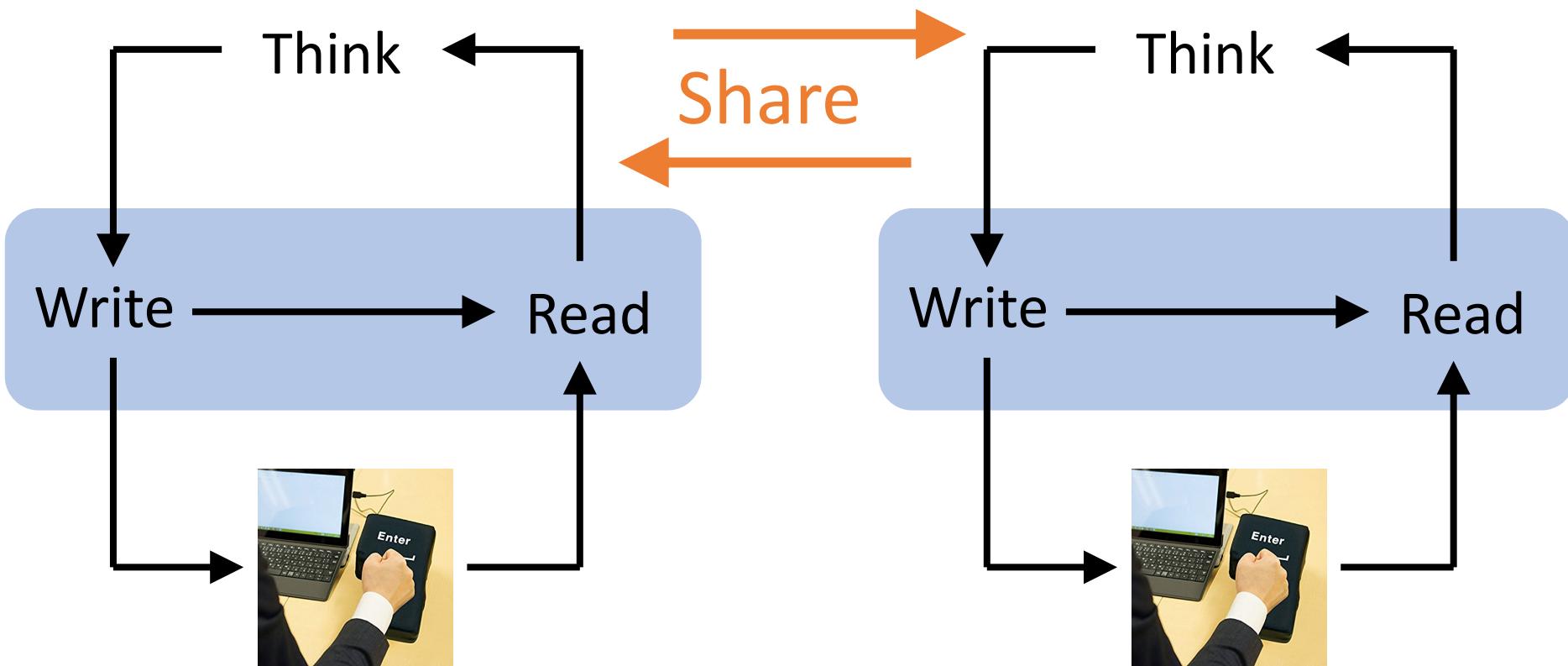


BeginneR

After

# Programming

Communicate





# BeginneR Session

## -- Nested data handling --

@kilometer00



you



packages



# Install

```
install.packages("tidyverse")  
install.packages("palmerpenguins")
```

# Attach

```
library("tidyverse")  
library("palmerpenguins")
```



# Pipe algebra %>%

{magrittr}

$X \text{ %}>\% f$	$\longleftrightarrow$	$f(X)$
$X \text{ %}>\% f(y)$	$\longleftrightarrow$	$f(X, y)$
$X \text{ %}>\% f \text{ %}>\% g$	$\longleftrightarrow$	$g(f(X))$
$X \text{ %}>\% f(y, .)$	$\longleftrightarrow$	$f(y, X)$

# data.frame

Data table

Figures

`read_csv`

`write_csv`

`pivot_longer`

Wide

`Long`

`ggplot`

`plot`

`pivot_wider`

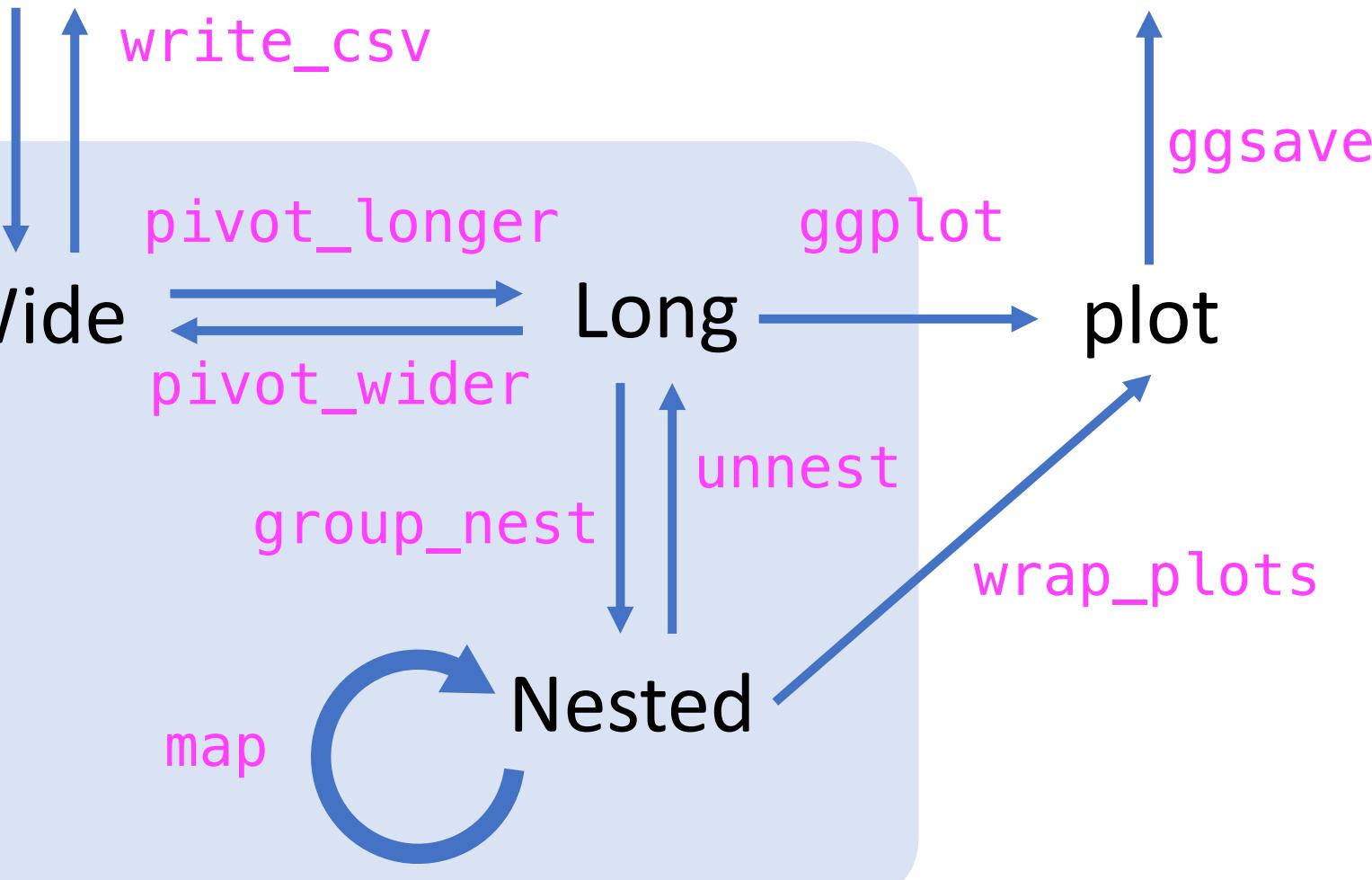
`group_nest`

`unnest`

`wrap_plots`

`map`

Nested



?data.frame

```
data.frame(  
  x = c(1:3),  
  y = letters[1:3],  
  z = seq(3, 5, by = 1) )
```

```
##      x y z  
## 1 1 a 3  
## 2 2 b 4  
## 3 3 c 5
```

```
?data.frame
```

```
data.frame(  
  x = c(1:3),  
  y = letters[1:3],  
  z = seq(3, 5, by = 1))
```

	x	y	z
## 1	1	a	3
## 2	2	b	4
## 3	3	c	5

## observation

variable

```
?data.frame
```

```
a <-  
  data.frame(  
    x = c(1:3),  
    y = letters[1:3],  
    z = seq(3, 5, by = 1) )
```

```
a$x
```

```
## [1] 1 2 3
```

```
?data.frame
```

```
a %>% mutate(new = x + 1)
```

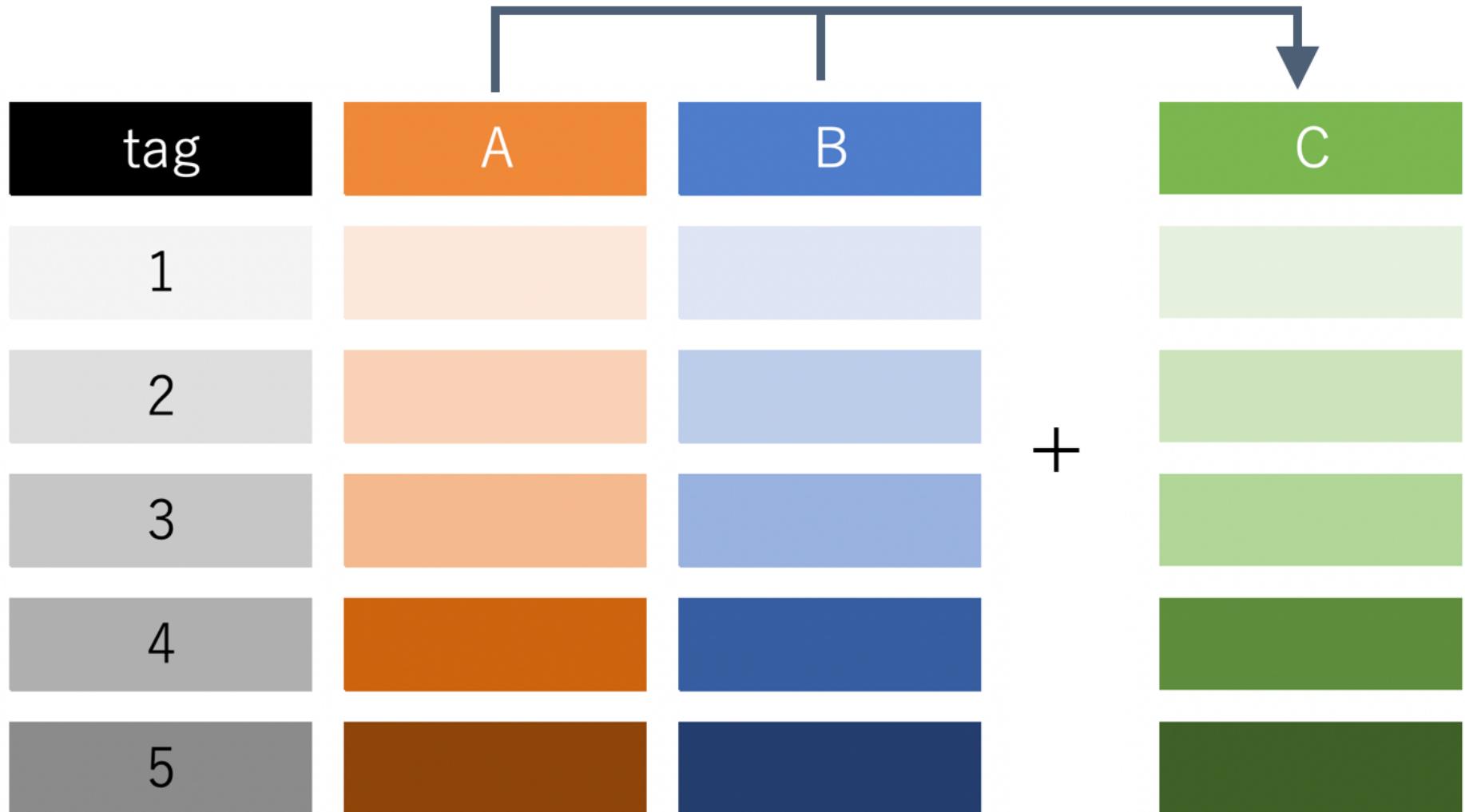
```
##   x y z new
## 1 1 a 3  2
## 2 2 b 4  3
## 3 3 c 5  4
```

```
a %>% mutate(new = x + y)
```

```
##   x y z new
## 1 1 a 3  4
## 2 2 b 4  6
## 3 3 c 5  8
```

`mutate(dat, C = fun(A, B))`

`dat %>% mutate(C = fun(A, B))`



# data.frame



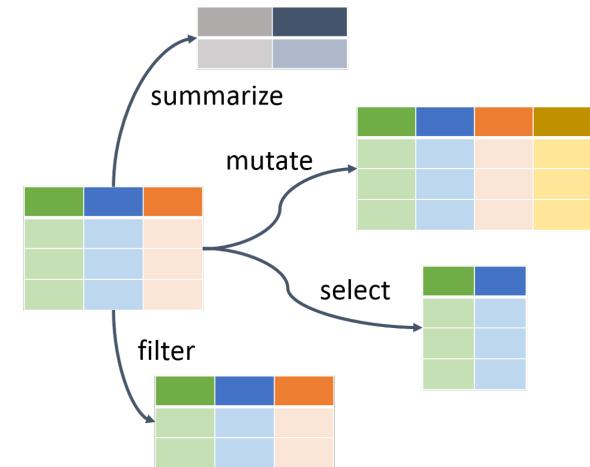
`library(tidyverse)`



## Verbs



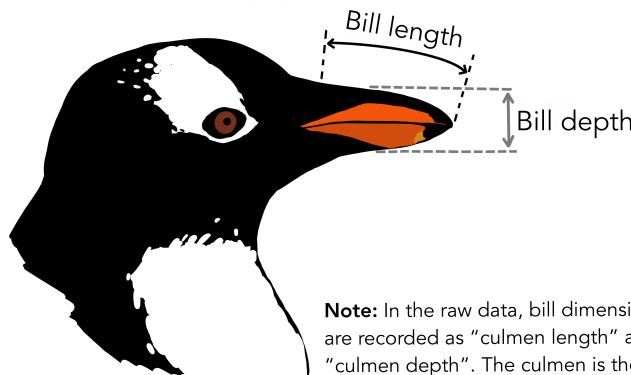
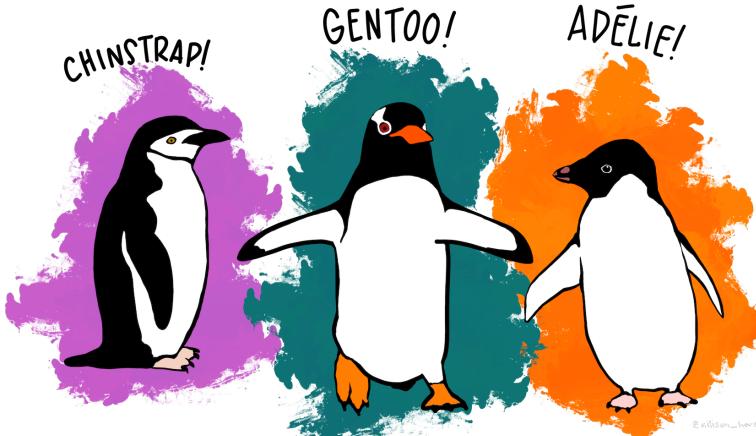
`mutate`  
`filter`  
`select`  
`rename`  
`summarize`



```
library(palmerpenguins)
penguins %>% head()
```



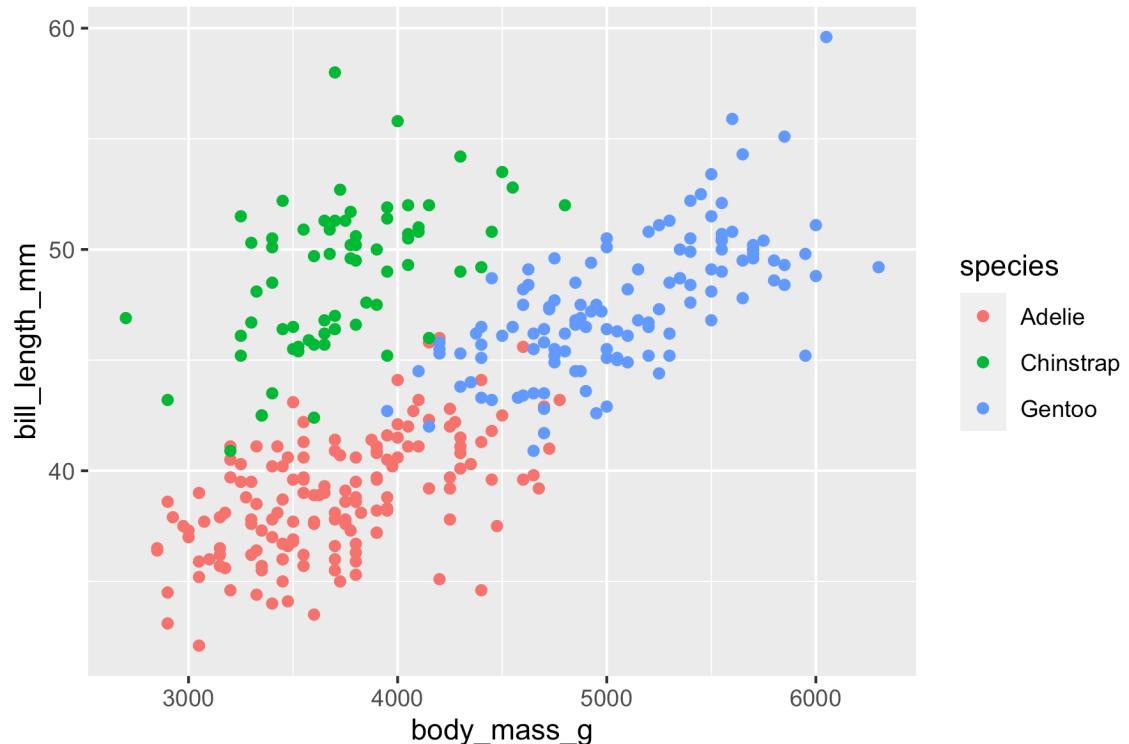
```
# A tibble: 6 x 8
  species island bill_length_mm bill_depth_mm flipper_length...
  <fct>   <fct>          <dbl>           <dbl>           <dbl> ...
1 Adelie  Torg...        39.1            18.7          181
2 Adelie  Torg...        39.5            17.4          186
3 Adelie  Torg...        40.3            18             195
4 Adelie  Torg...         NA              NA             NA
5 Adelie  Torg...        36.7            19.3          193
6 Adelie  Torg...        39.3            20.6          190
# ... with 3 more variables: body_mass_g <int>, sex <fct>,
#   year <int>
```



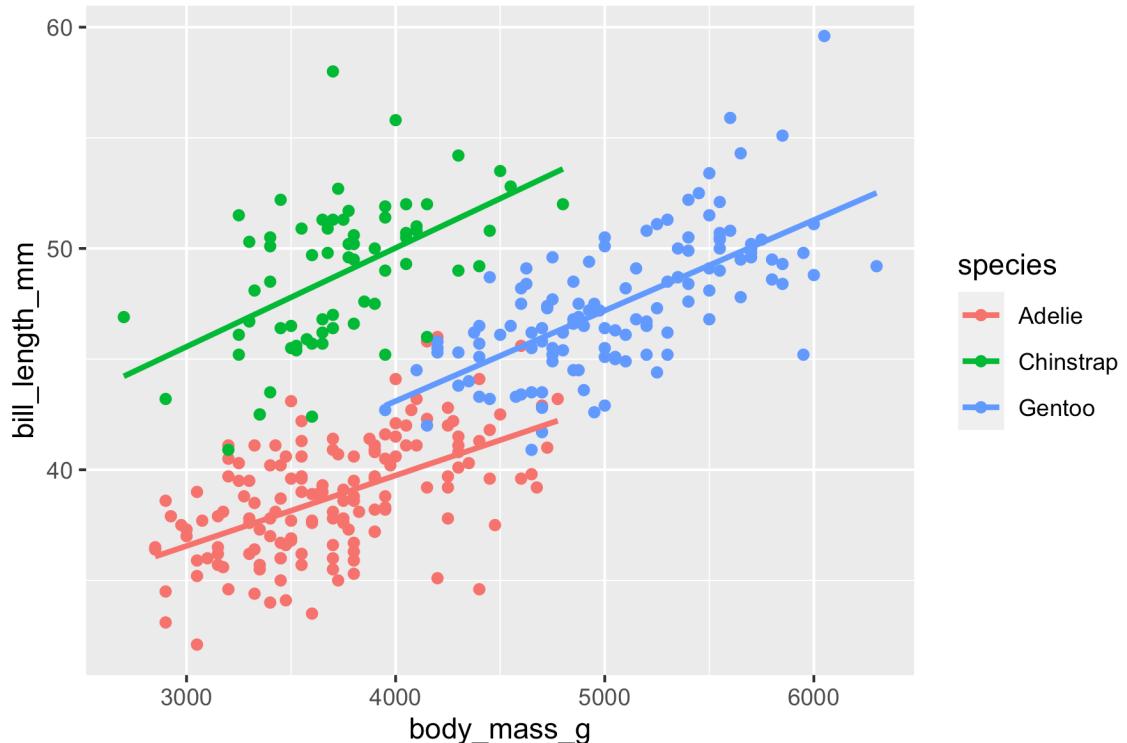
**Note:** In the raw data, bill dimensions are recorded as "culmen length" and "culmen depth". The culmen is the dorsal ridge atop the bill.

Artwork by @allison\_horst

```
ggplot(data = penguins) +  
  aes(x = body_mass_g,  
      y = bill_length_mm,  
      color = species) +  
  geom_point()
```



```
ggplot(data = penguins) +  
  aes(x = body_mass_g,  
      y = bill_length_mm,  
      color = species) +  
  geom_point() +  
  geom_smooth(method = "lm", se = F)
```



```
penguins_xy <-  
penguins %>%  
mutate(x = body_mass_g,  
y = bill_length_mm)
```

```
penguins_xy %>%  
filter(species == "Adelie") %>%  
lm(y ~ x, data = .)
```

Call:

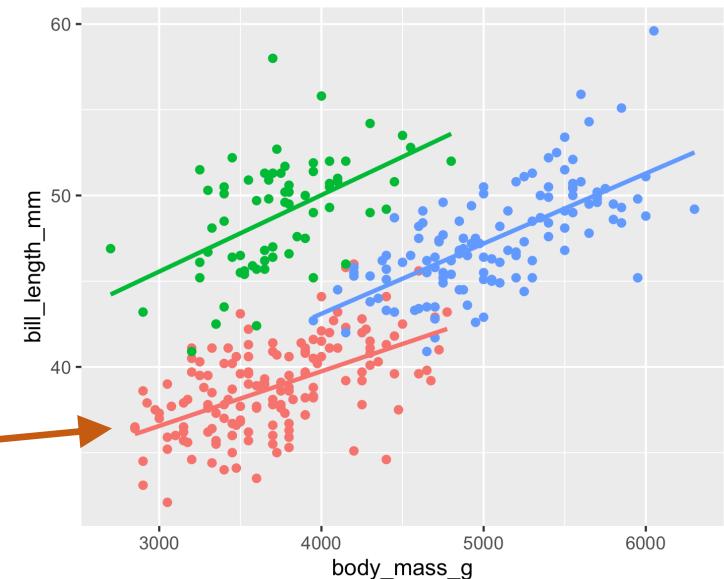
```
lm(formula = y ~ x, data = .)
```

Coefficients:

(Intercept)

26.994139

x  
0.003188



```
penguins_xy %>%  
  filter(species == "Adelie") %>%  
  lm(y ~ x, data = .) %>%  
  summary()
```

Call:

```
lm(formula = y ~ x, data = .)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.4208	-1.3690	0.1874	1.4825	5.6168

Coefficients:

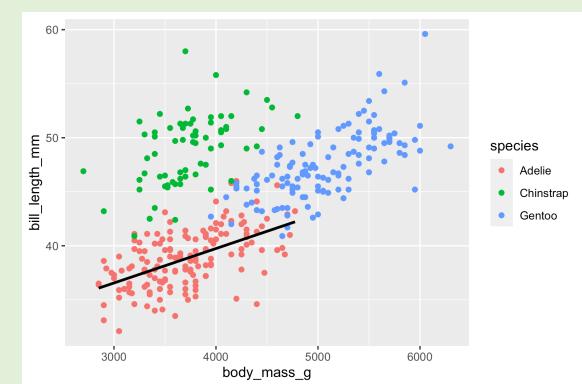
	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.699e+01	1.483e+00	18.201	< 2e-16 ***
x	3.188e-03	3.977e-04	8.015	2.95e-13 ***

---

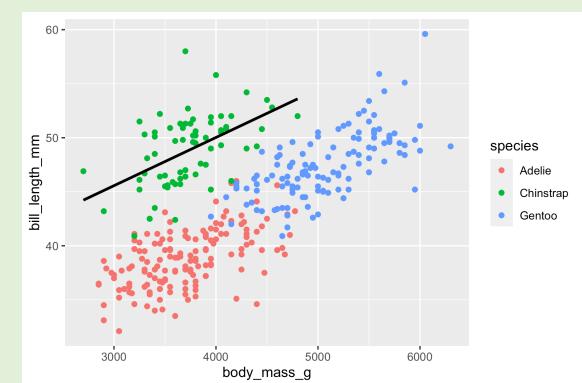
Residual standard error: 2.234 on 149 degrees of freedom  
(1 observation deleted due to missingness)

Multiple R-squared: 0.3013, Adjusted R-squared: 0.2966  
F-statistic: 64.24 on 1 and 149 DF, p-value: 2.955e-13

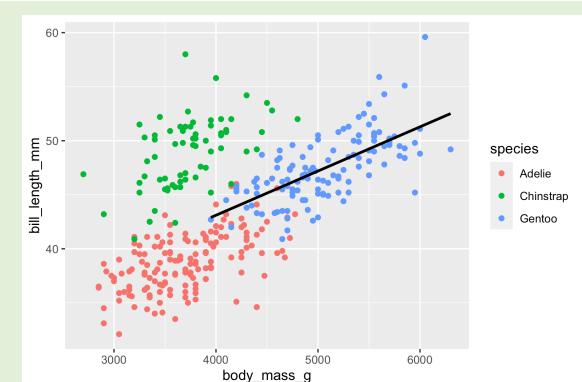
```
penguins_xy %>%  
  filter(species == "Adelie") %>%  
  lm(y ~ x, data = .) %>%  
  summary()
```



```
penguins_xy %>%  
  filter(species == "Chinstrap") %>%  
  lm(y ~ x, data = .) %>%  
  summary()
```



```
penguins_xy %>%  
  filter(species == "Gentoo") %>%  
  lm(y ~ x, data = .) %>%  
  summary()
```



```
penguins_xy %>% group_nest(species)
```

```
# A tibble: 3 x 2
```

	species	data
	<fct>	<list<tbl_df[,9]>>
1	Adelie	[152 × 9]
2	Chinstrap	[68 × 9]
3	Gentoo	[124 × 9]

```
penguins_xy %>% group_nest(species)
```

```
# A tibble: 3 × 2
```

	species	data
1	Adelie	[152 × 9]
2	Chinstrap	[68 × 9]
3	Gentoo	[124 × 9]

```
penguins_xy %>%
```

```
  group_nest(species, island)
```

```
# A tibble: 5 × 3
```

	species	island	data
1	Adelie	Biscoe	[44 × 8]
2	Adelie	Dream	[56 × 8]
3	Adelie	Torgersen	[52 × 8]
4	Chinstrap	Dream	[68 × 8]
5	Gentoo	Biscoe	[124 × 8]

```
penguins_xy %>% group_nest(species)
```

```
# A tibble: 3 x 2
```

	species	data
1	Adelie	[152 x 9]
2	Chinstrap	[68 x 9]
3	Gentoo	[124 x 9]

```
penguins_xy %>% group_nest(species) %>%
  .$data %>% .[[1]]
```

```
# A tibble: 152 x 9
```

	island	bill_length_mm	bill_depth_mm	flipper_length_mm	body_mass_g	species	sex	year
1	Torgersen	39.1	18.7	181		Adelie	Female	2009
2	Torgersen	39.5	17.4	186		Adelie	Female	2009
3	Torgersen	40.3	18	195		Adelie	Female	2009
4	Torgersen	NA	NA	NA		Adelie	Female	2009
5	Torgersen	36.7	19.3	193		Adelie	Female	2009
6	Torgersen	39.3	20.6	190		Adelie	Female	2009

```
penguins_lm <-  
  penguins_xy %>%  
  group_nest(species) %>%  
  mutate(fit = map(data, ~ lm(y ~ x, data = .)),  
         summary = map(fit, summary))
```

# A tibble: 3 × 4

	species	data	fit	summary
1	Adelie	[152 × 9]	<lm>	<smmry.lm>
2	Chinstrap	[68 × 9]	<lm>	<smmry.lm>
3	Gentoo	[124 × 9]	<lm>	<smmry.lm>

```
penguins_lm <-  
  penguins_xy %>%  
  group_nest(species) %>%  
  mutate(fit = map(data, ~ lm(y ~ x, data = .)),  
         summary = map(fit, summary))
```

# A tibble: 3 × 4

	species	data	fit	summary
1	Adelie	[152 × 9]	<lm>	<smmry.lm>
2	Chinstrap	[68 × 9]	<lm>	<smmry.lm>
3	Gentoo	[124 × 9]	<lm>	<smmry.lm>

```
penguins_xy %>%  
  filter(species == "Adelie") %>%  
  lm(y ~ x, data = .) %>%  
  summary()
```

```
penguins_lm <-  
  penguins_xy %>%  
  group_nest(species) %>%  
  mutate(fit = map(data, ~ lm(y ~ x, data = .))),  
         summary = map(fit, summary))
```

```
penguins_lm$summary[[1]]
```

Call:

```
lm(formula = y ~ x, data = .)
```

Residuals:

Min	1Q	Median	3Q	Max
-6.4208	-1.3690	0.1874	1.4825	5.6168

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.699e+01	1.483e+00	18.201	< 2e-16 ***
x	3.188e-03	3.977e-04	8.015	2.95e-13 ***

```
penguins_lm <-  
  penguins_xy %>%  
  group_nest(species) %>%  
  mutate(fit = map(data, ~ lm(y ~ x, data = .)),  
         summary = map(fit, summary),  
         a = map_dbl(fit, ~ .$coefficients[2]),  
         R2 = map_dbl(summary, ~ .$r.squared))  
  
# A tibble: 3 x 6  
#> species     data       fit      summary      a      R2  
#> <fct> <list<tbl_df[,9]>>> <list>      <list>    <dbl>    <dbl>  
#> 1 Adelie     [152 × 9]   <lm>    <smmry.lm> 0.00319  0.301  
#> 2 Chinstrap   [68 × 9]    <lm>    <smmry.lm> 0.00446  0.264  
#> 3 Gentoo     [124 × 9]   <lm>    <smmry.lm> 0.00409  0.448
```

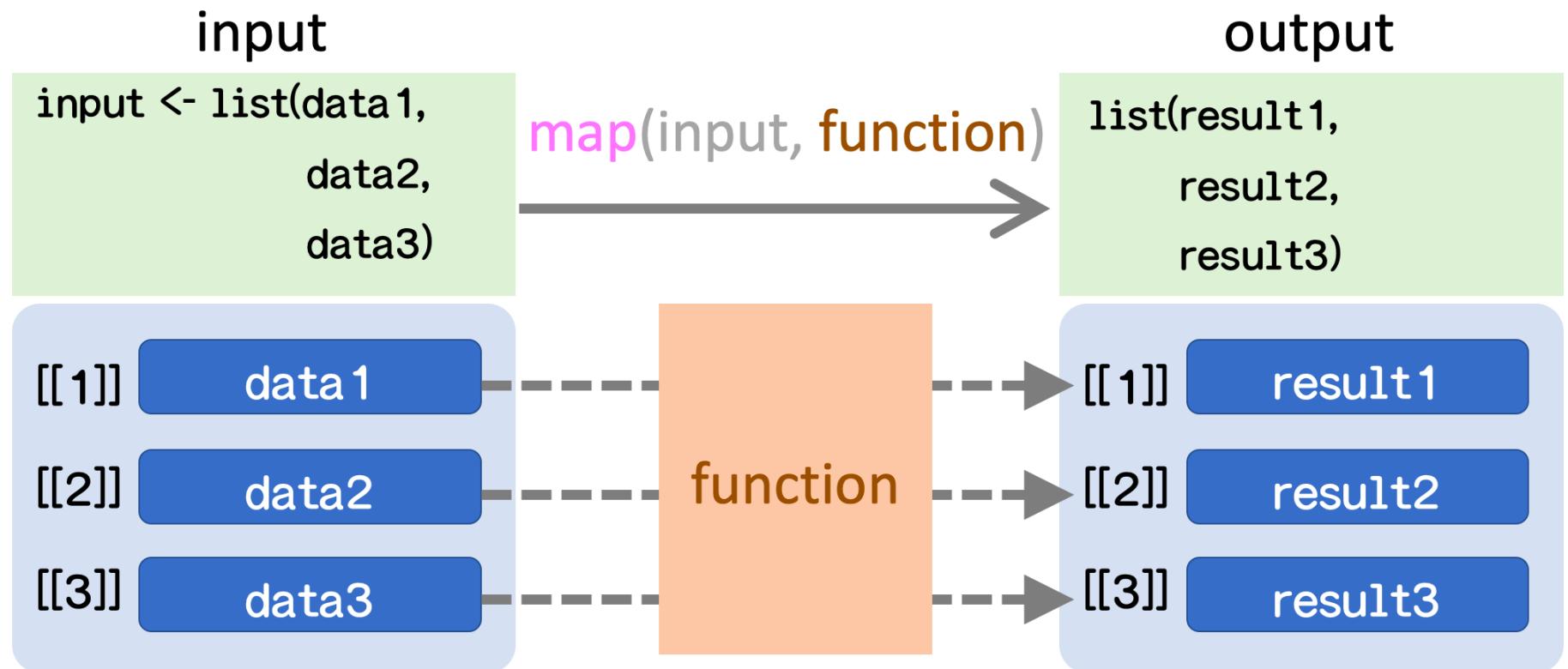
## Wrapper functions

`map_dbl()`, `map_chr()`

`map_dfc()`, `map_dfr()`

1. group\_nest()
2. mutate()
3. map()

# ?map



?map

```
f <- function(num){ num * 4 }
```

```
dat <- 1:4
```

```
f(num = dat)
```

```
[1] 4 8 12 16
```

?map

```
f <- function(num){ num * 4 }
```

```
dat <- 1:4
```

```
f(num = dat)
```

```
[1] 4 8 12 16
```

```
dat <- list(1:4, 7:4)
```

```
f(num = dat)
```

Error in num\*4 : non-numeric argument to binary operator

?map

```
f <- function(num){ num * 4 }
```

```
dat <- list(1:4, 7:4)
```

```
f(num = dat)
```

```
map(.x = dat, .f = f)
```

```
[[1]]
```

```
[1] 4 8 12 16
```

```
[[2]]
```

```
[1] 28 24 20 16
```

?map

```
f <- function(num){ num * 4 }

dat <- list(1:4, 7:4)

map(.x = dat, .f = f)
```

by using **for**

```
result <- NULL
for(i in 1:length(dat)){
  result[[i]] <- f(dat[[i]])
}
```

# `data.frame`



Verbs

`mutate`  
`filter`  
`select`  
`rename`  
`summarize`

?map

```
f <- function(num){ num * 4 }
```

```
dat <- list(1:4, 7:4)
```

```
map(.x = dat, .f = f)
```

```
map(dat, f)
```

```
map(.x = dat, ~ f(num = .x))
```

```
map(.x = dat, function(num){num * 4})
```

```
map(dat, ~ {.x * 4})
```

```
map(dat, ~ {.* 4})
```

# group\_nest → mutate → map

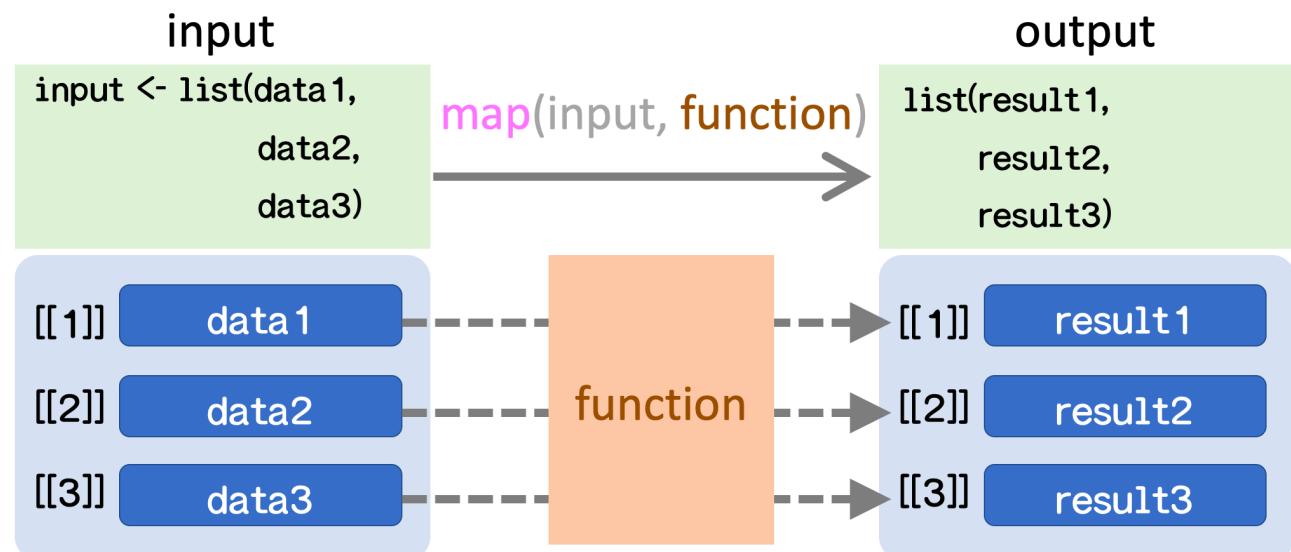
```
penguins_xy %>%
```

```
  group_nest(species) %>%
```

```
  mutate(fit = map(data, ~ lm(y ~ x, data = .)))
```

```
# A tibble: 3 × 4
```

	species	data	fit
1	Adelie	<list<tbl_df[,9]>> [152 × 9] → lm(y ~ x) → <lm>	<list>
2	Chinstrap	[68 × 9] → lm(y ~ x) → <lm>	
3	Gentoo	[124 × 9] → lm(y ~ x) → <lm>	



# group\_nest → mutate → map

```
penguins_xy %>%  
  group_nest(species) %>%  
  mutate(fit = map(data, ~ lm(y ~ x, data = .)))
```

```
f <- function(dat){ lm(y ~ x, data = dat) }
```

```
penguins_xy %>%  
  group_nest(species) %>%  
  mutate(fit = map(data, f))
```

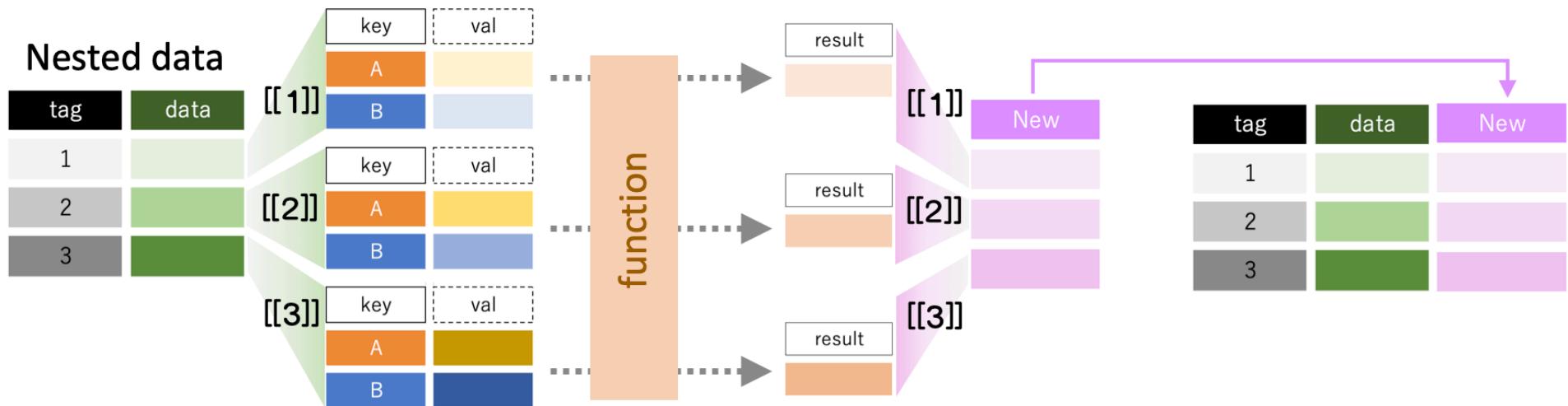
```
penguins_xy %>%  
  group_nest(species) %>%  
  mutate(fit = map(data,  
                  function(dat){  
                    lm(y ~ x, data = dat)}))
```

# 1. group\_nest()

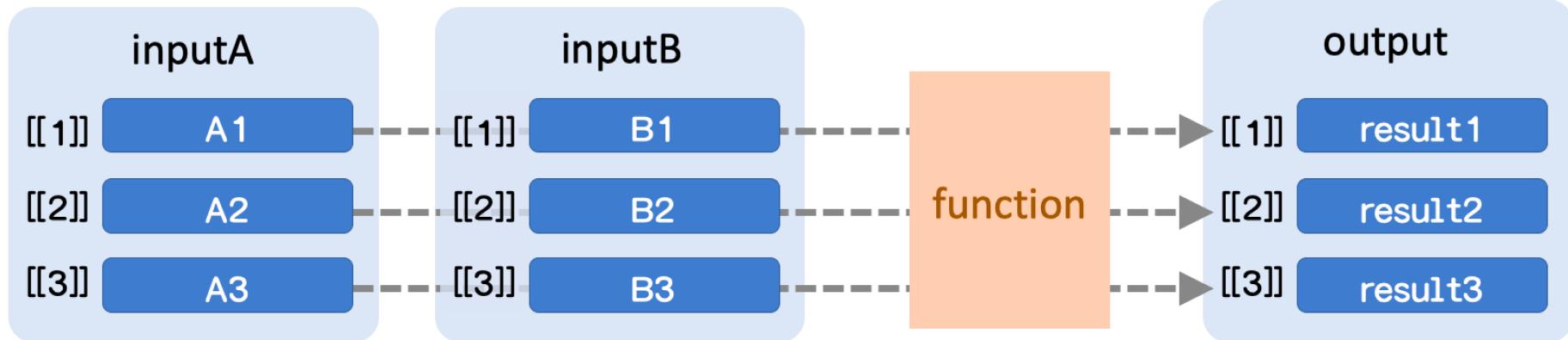
## 2. mutate()

### 3. map()

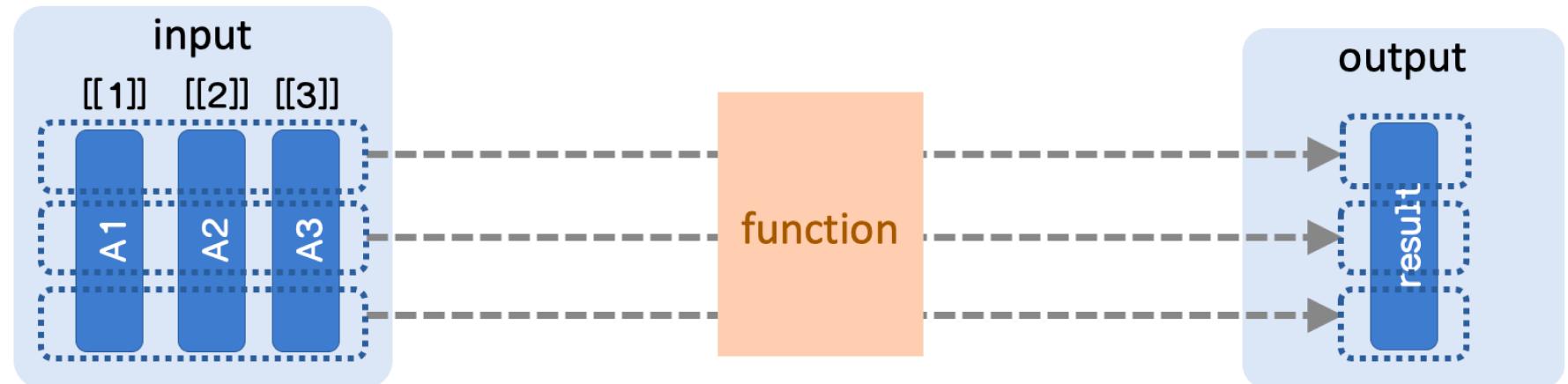
data\_nested %>% **mutate(New = map(data, function))**



`map2(.x = inputA, .y = inputB, function(.x, .y))`



`pmap(input, function(A1, A2, A3))`



 @kilometer が2020年05月09日に投稿

## nested dataでggplot

⌚ R, map, ggplot2, Nest

LG  
TM 13

---

 @kilometer が2019年11月19日に投稿

## map脳になろう、もしくはnested dataのハンドリング

⌚ R

LG  
TM 40

---

 @kilometer が2019年02月13日に投稿

## {purrr} mapを導入しよう。

⌚ R, map, purrr

LG  
TM 37

---

 @kilometer が2019年01月28日に投稿

## {tidyverse} nestしていこう。

⌚ R, purrr, tidyverse

LG  
TM 51

---

 @kilometer が2018年12月22日に投稿

## R言語入門（裏口） -- Landscape with R --

⌚ R, プログラミング言語, 裏口

LG  
TM 58

# data.frame

Data table

Figures

`read_csv`

`write_csv`

Wide

`pivot_longer`

Long

`ggplot`

`plot`

`pivot_wider`

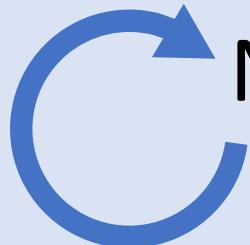
`group_nest`

`unnest`

`wrap_plots`



`map`



Nested



Enjoy!!