

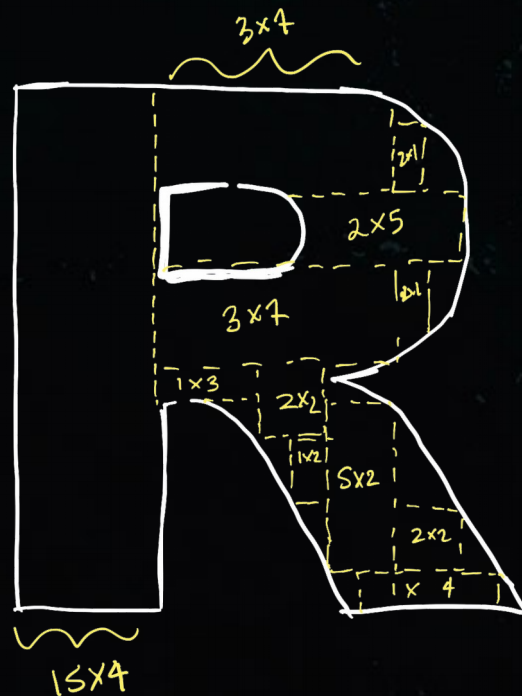
Reorganización de datos

Martín Amodeo

DBBF, UNS

IADO CONICET-UNS

CCT Bahía Blanca

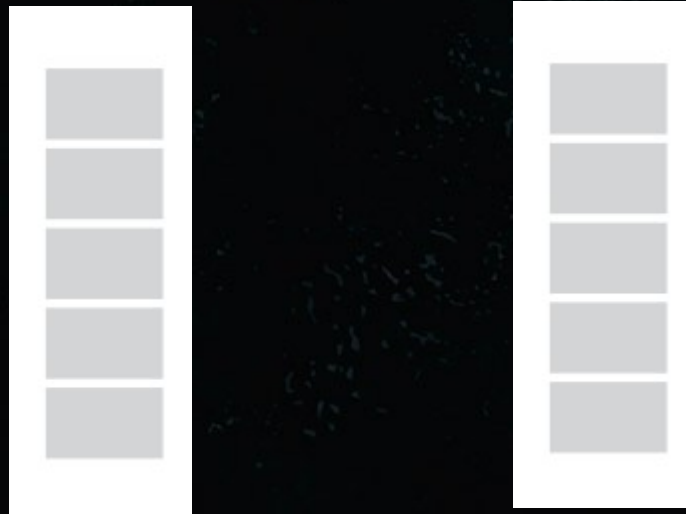


Funciones simples (valores)



log() sqrt() exp()

Funciones simples (vectorización)



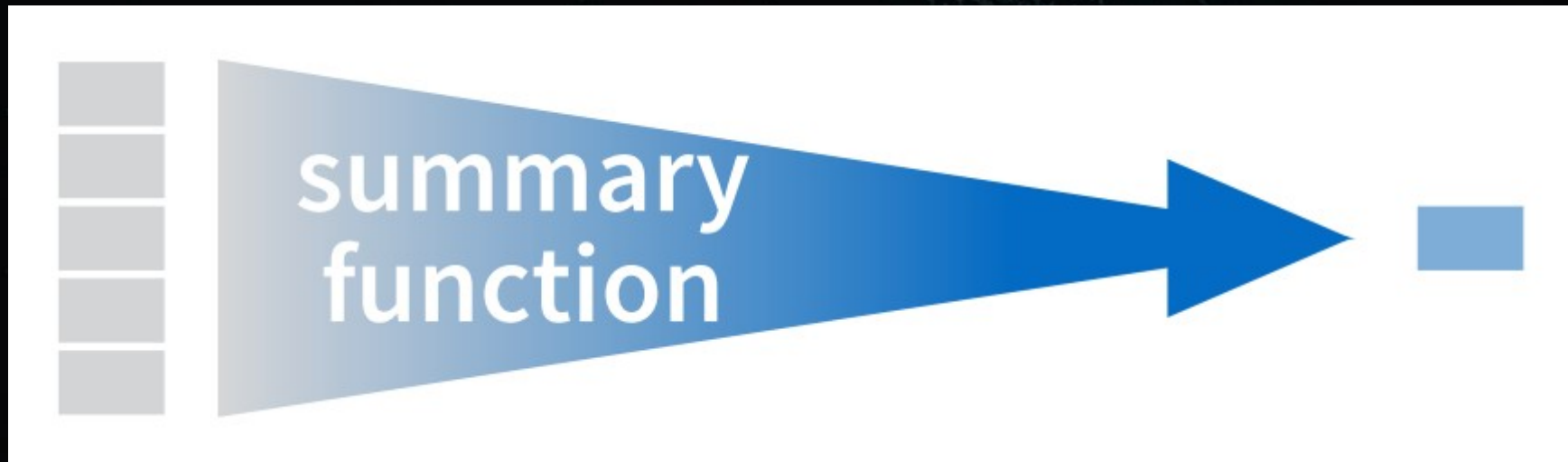
`log()` `sqrt()` `exp()`

Vectorización

- R opera de forma “vectorizada”
- Si una función que normalmente opera sobre un valor la aplico sobre un vector, va a realizar la operación sobre cada elemento y devolver un vector

`log(c(2, 45, 86))`

Funciones de Resumen



`mean()`

`median()`

`sum()`

`min()`

`max()`

`sd()`

`var()`

Recategorizar una variable categórica



- `levels(datos$ZONA) <- c("A", "B", "C")`
- `factor(datos$ZONA, labels= c("A","B","C"))`
- `factor(datos$ZONA, levels=c("B", "A", "C"))`

Categorizar una variable numérica



```
cut(datos$temperatura,  
breaks=c(0,10,50,100),  
labels=c("frío","templado","cálido"))
```

Categorizar una variable numérica



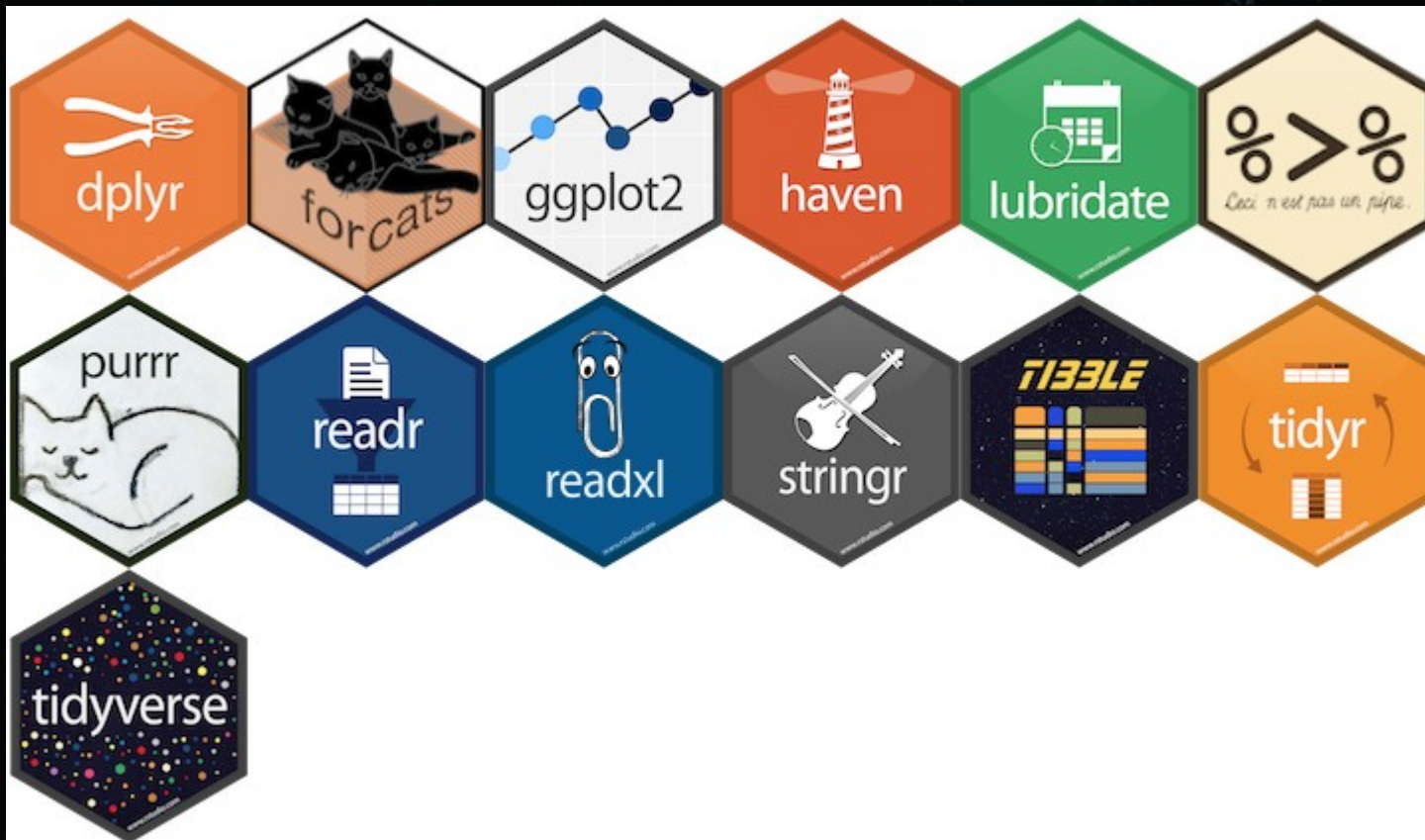
- `cut(datos$temperatura, breaks=c(), labels= c("", "", ""))`

Transformar una variable numérica



- `datos$temperatura <- log(datos$temperatura)`
- `datos$log_temperatura <- log(datos$temperatura)`

Paquetes: tidyverse



<https://www.rstudio.com/resources/cheatsheets/>

v1	v2	v3	v4
:	:	:	:
:	:	:	:
:	:	:	:
:	:	:	:
:	:	:	:
:	:	:	:
:	:	:	:
:	:	:	:
:	:	:	:

Dataframe



Vectores



Variables

Observaciones

dplyr

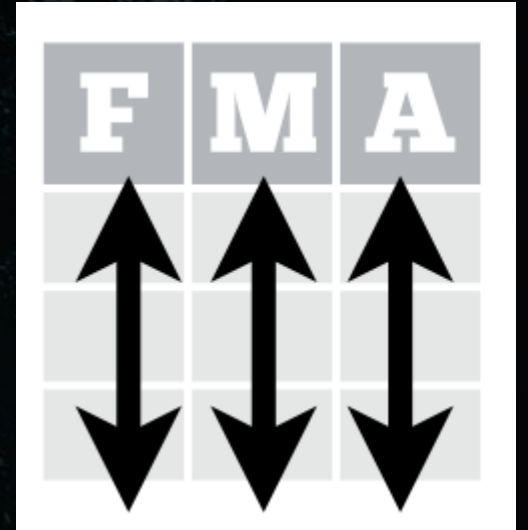


<https://dplyr.tidyverse.org/>

```
install.packages("dplyr")
```

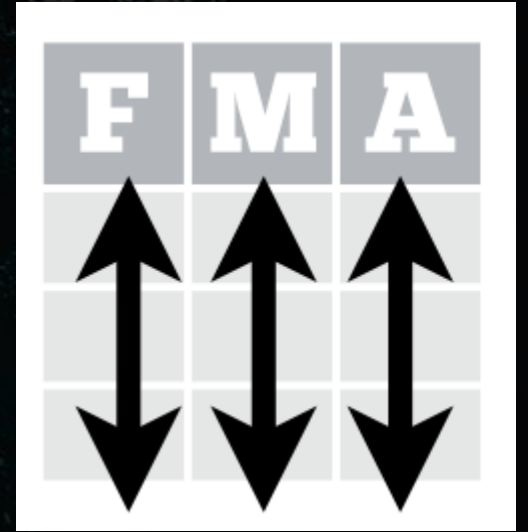
```
library(dplyr)
```


Selecciono variables



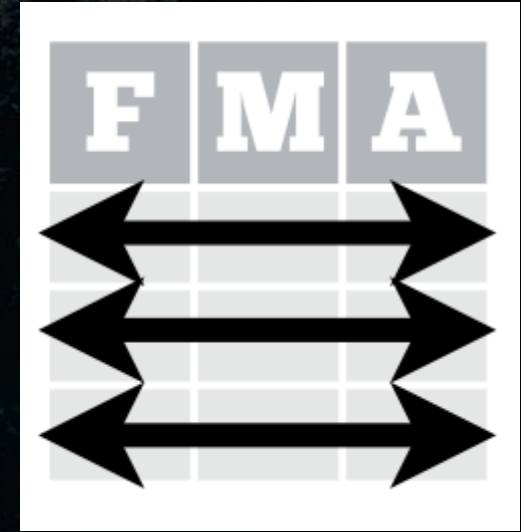
- `select(misdatos, F, M)`

Calculo nueva variable



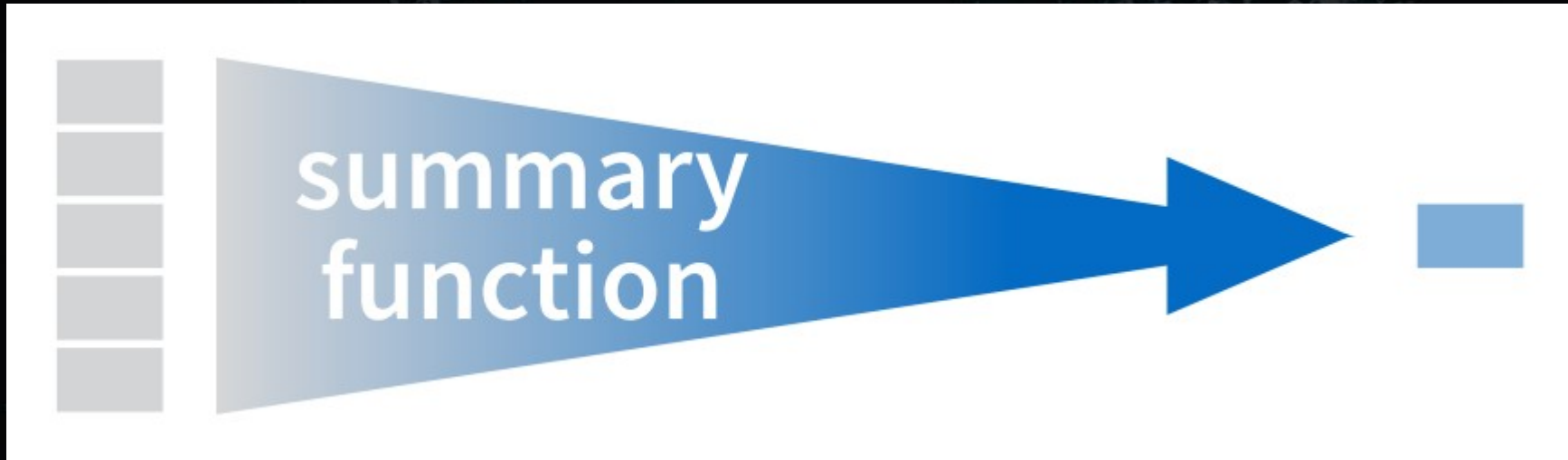
- `mutate(misdatos, T = F + M)`

Filtro según observaciones



- `filter(misdatos, F < 10.2)`

Resumen



- `summarise(misdatos, altura_promedio = mean(altura))`

`mean()`

`median()`

`sum()`

`min()`

`max()`

`sd()`

`var()`

tidyr



<https://tidyr.tidyverse.org/>

```
install.packages("tidyr")
```

```
library(tidyr)
```


Pivoteo

country	1999	2000
A	0.7K	2K
B	37K	80K
C	212K	213K



country	year	cases
A	1999	0.7K
B	1999	37K
C	1999	212K
A	2000	2K
B	2000	80K
C	2000	213K

Formato Ancho

Formato Largo

Pivoteo

country	1999	2000
A	0.7K	2K
B	37K	80K
C	212K	213K



country	year	cases
A	1999	0.7K
B	1999	37K
C	1999	212K
A	2000	2K
B	2000	80K
C	2000	213K

`pivot_longer()`



`pivot_wider()`

Resumen en grupos



- `agreggate(misdatos, y ~ grupo, FUN = mean)`
- `group_by()` {dplyr}

`mean()` `median()` `sum()` `min()` `max()` `sd()` `var()`