

Análisis de #múltiples epidemias y prevalencias con R y purrr

Andree Valle Campos



[@avallecam](https://twitter.com/avallecam)



slides: <http://bit.ly/bbslisepi>



taller: <https://rstudio.cloud/project/1747813>



código:

https://github.com/avallecam/LIS_brownbag/blob/master/20201010-analisis-epidemia-prevalencia/taller.Rmd

```
1 class Repeat
2   def print_message
3     puts "I Will Not Repeat My Code"
4     puts "I Will Not Repeat My Code"
5     puts "I Will Not Repeat My Code"
6     puts "I Will Not Repeat My Code"
7     puts "I Will Not Repeat My Code"
8     puts "I Will Not Repeat My Code"
9     puts "I Will Not Repeat My Code"
10  end
11 end
```



“R es una herramienta de flujos de trabajo (y que también hace estadística)”

[@dgkeyes](#)

- (F) Identificables
- (A) Accesibles
- (I) Interoperables
- (R) Reusables



FAIR principles: <https://www.go-fair.org/fair-principles/>



Un paquete organiza, documenta y facilita la distribución de R scripts / functions

[@tinyheero](#) / [@kadamatiharish](#) / [@drob](#)





Experiencia: Flujos -> Paquetes (en desarrollo)



Inmunómica:
Microarreglo de
proteínas
[biostats2019](https://biostats2019.org)

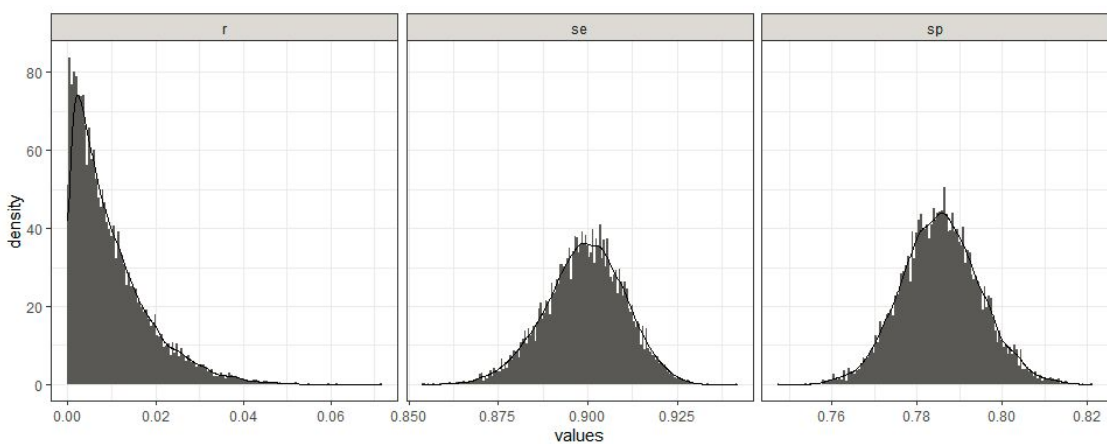
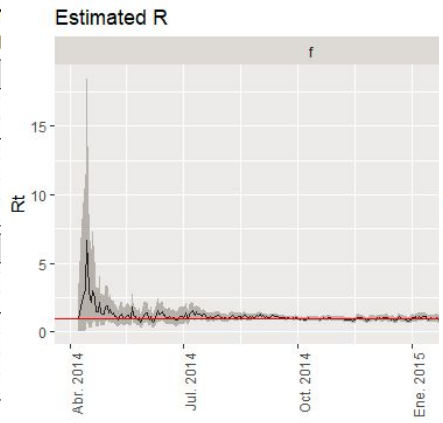
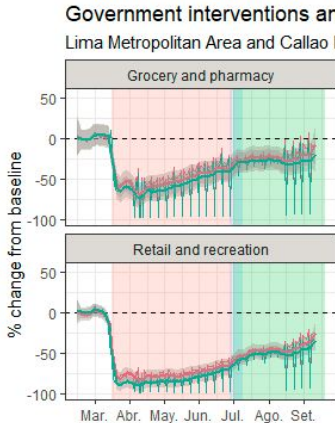
Flujo en Epi:
OR, RR, PR
[epitidy](https://epitidy.org)

Poder Estadístico
y Tamaño de
Muestra
[powder](https://powder.org)

Canales
Endémicos
[epichannel](https://epichannel.org)

Acceso a
ENAH (contrib.)
ENDES.PE

Desarrollo Epidemic Channel, Javiera, Perú, 2008/2009

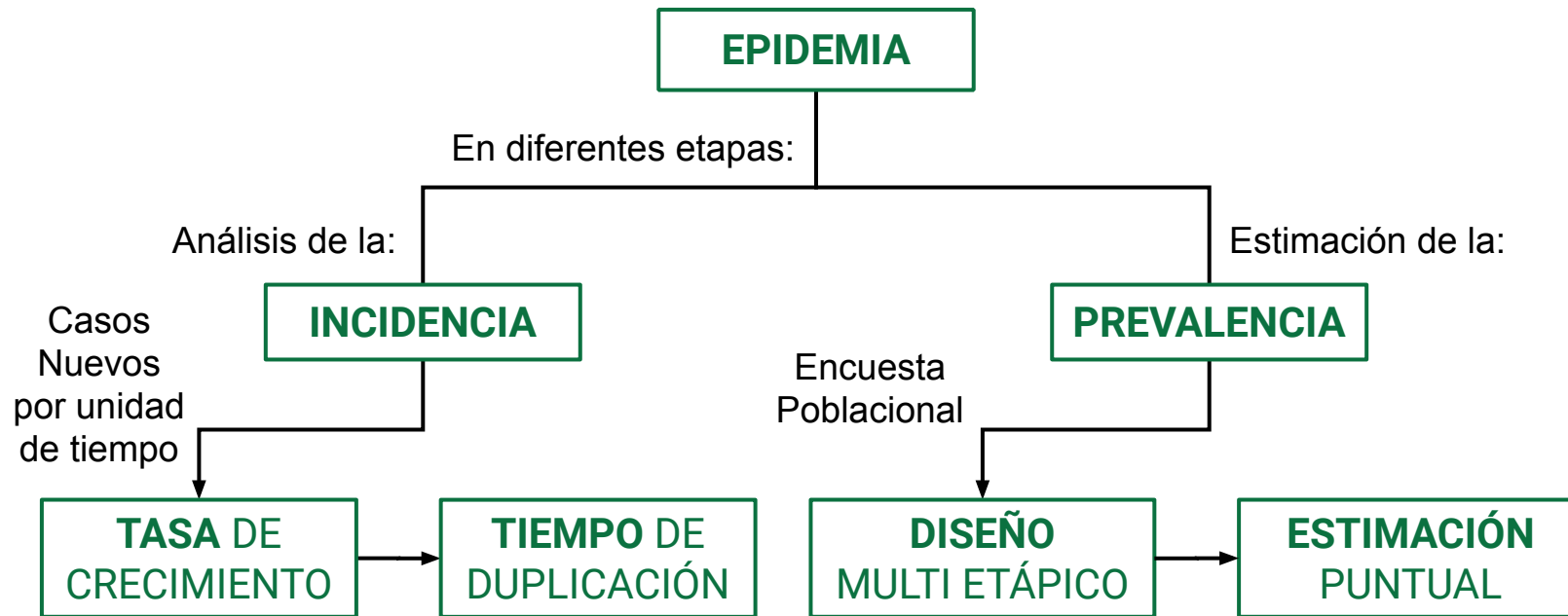


Descargar y Visualizar
datos abiertos de
COVID-19
[covid19viz](https://covid19viz.org)

Flujo: describir
Curva Epidémica y
Estimar Rt
[incidenceflow](https://incidenceflow.org)

Reproducibilidad en
análisis de Encuestas
Serológicas
[serosurvey](https://serosurvey.org)

Contenido



`incidence -> nest -> map`

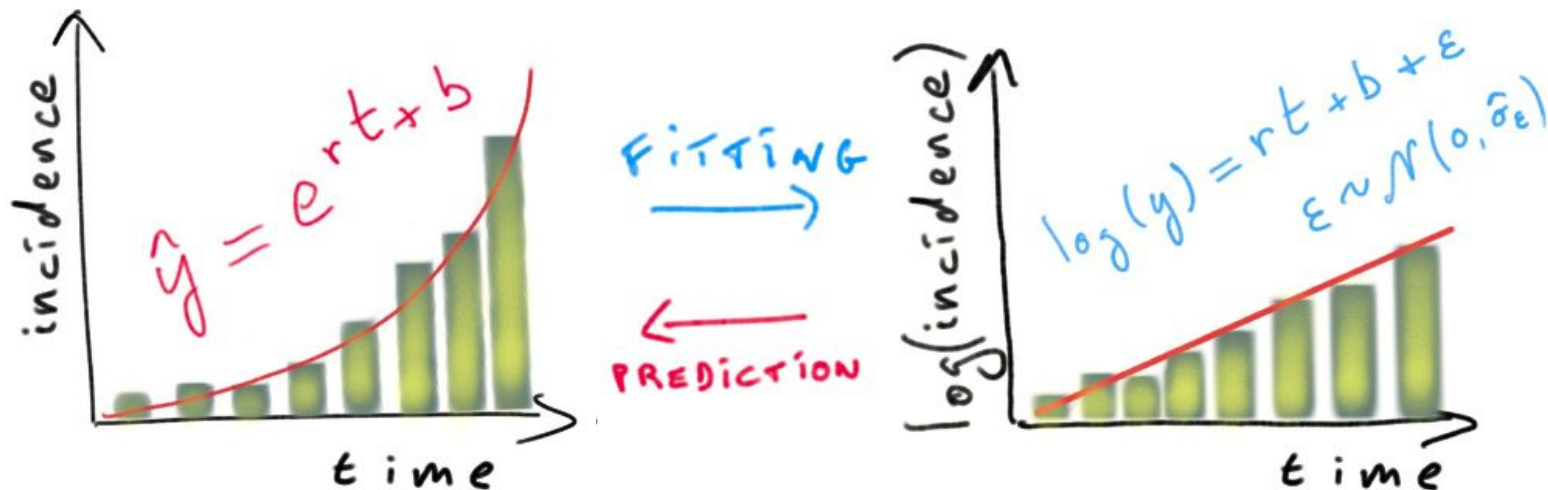
`expand_grid -> sym -> pmap`

INTRODUCCIÓN

CASO SINGULAR

RETO MÚLTIPLE

Incidencia: Introducción (1/)



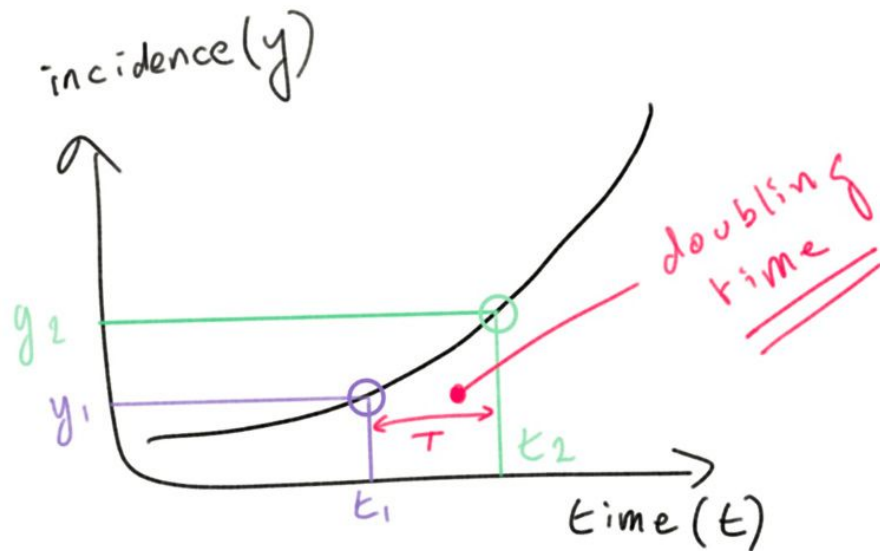
y : Número de casos incidentes

r : Tasa de Crecimiento

β_0 : Intercepto

$$\left. \begin{array}{l} y : \text{Número de casos incidentes} \\ r : \text{Tasa de Crecimiento} \\ \beta_0 : \text{Intercepto} \end{array} \right\} \log(y) = \beta_0 + rt + \epsilon \Rightarrow \hat{y} = \exp^{rt+\beta_0}$$

Incidencia: Introducción (1/)



y : Número de casos incidentes

r : Tasa de Crecimiento

β_0 : Intercepto

$$\left. \begin{array}{l} y : \text{Número de casos incidentes} \\ r : \text{Tasa de Crecimiento} \\ \beta_0 : \text{Intercepto} \end{array} \right\} \hat{y} = \exp^{rt + \beta_0}$$

Con la incidencia en escala original:

T es el tiempo que le toma a la curva epidémica para duplicarse

$$\hat{y}_2 = 2\hat{y}_1$$

$$2 = \frac{\hat{y}_2}{\hat{y}_1} = \frac{\exp^{rt_2 + \beta_0}}{\exp^{rt_1 + \beta_0}}$$

$$2 = \exp^{rt_2 + \beta_0 - (rt_1 + \beta_0)}$$

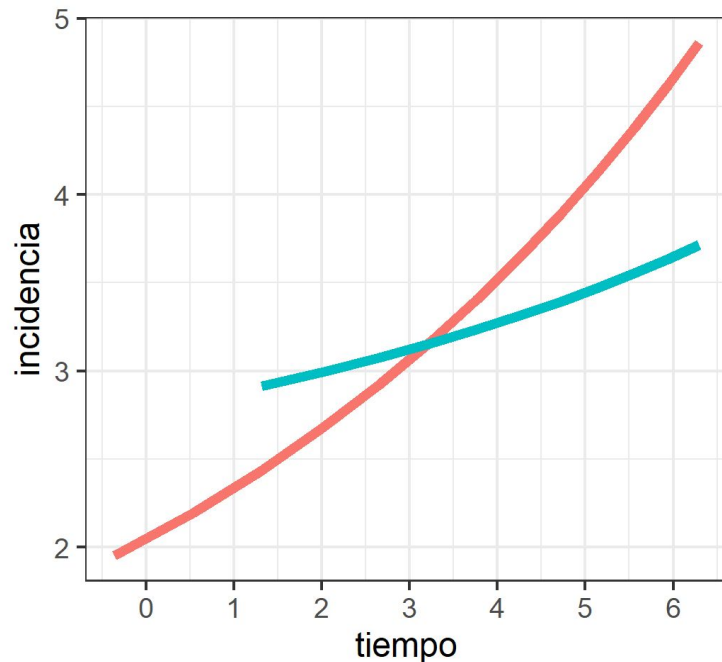
$$2 = \exp^{r(t_2 - t_1)}$$

$$2 = \exp^{rT}$$

$$T = \frac{\log(2)}{r}$$

Incidencia: Introducción (1/)

$$T = \frac{\log(2)}{r}$$



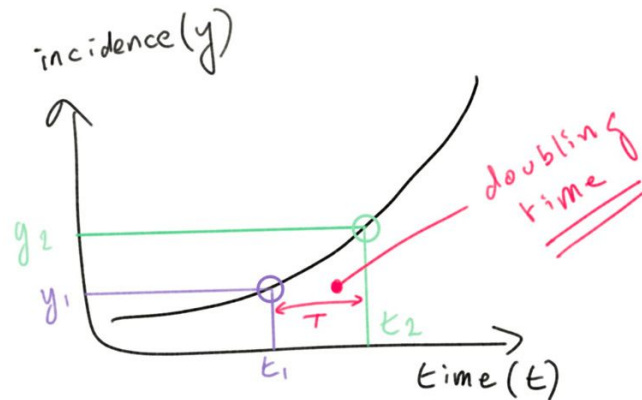
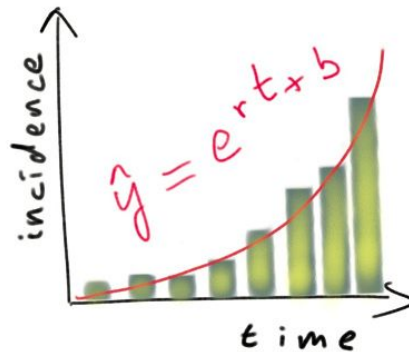
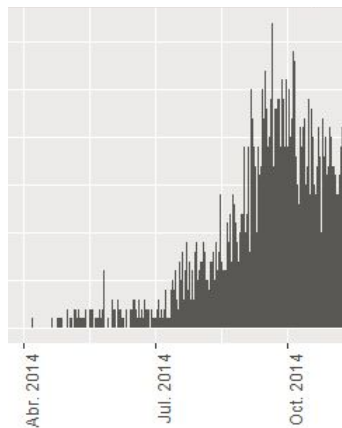
1
2

$$r_1 > r_2$$
$$T_1 < T_2$$

r : Tasa de Crecimiento

T : Tiempo de Duplicación

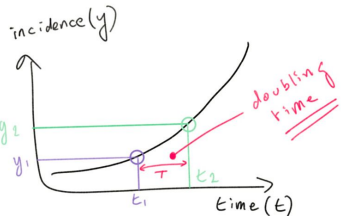
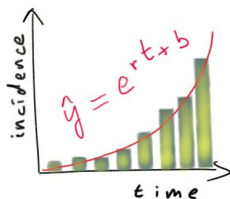
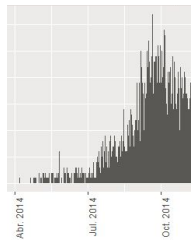
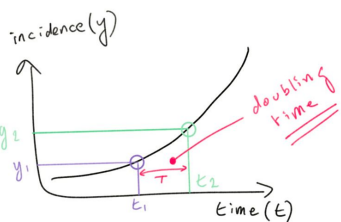
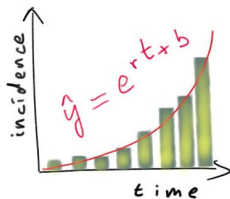
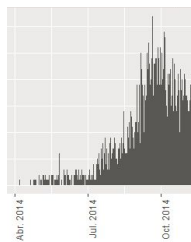
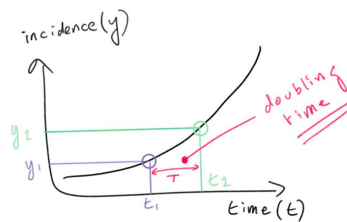
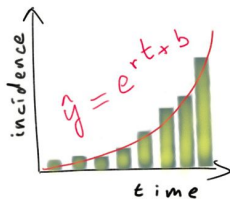
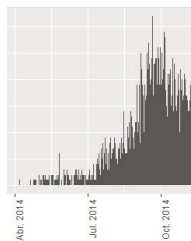
Incidencia: Caso Singular (3/)



```
dat <- ebola_sim$linelist$date_of_onset
i.7 <- incidence(dat, interval=7)
# plot(i.7)
f1 <- fit(i.7[1:20])
```

```
f1 %>% tidy_incidence()
#> # A tibble: 2 x 5
#>   mark parameter estimate conf_lower conf_upper
#>   <chr> <chr>      <dbl>      <dbl>      <dbl>
#> 1 1      r          0.0318      0.0260      0.0376
#> 2 1      doubling    21.8         18.5         26.7
```

Incidencia: Reto Múltiple (3/)



Problema:

¿Estimar r y t_d ...

... para 5 estratos de edad?

- copiar (repetir) el código!

... para 1852 distritos del Perú?

- ...!
- aplicas {purrr}

nest() + map()

nest() por: DISTRITOS (p.e.)

map(.x, .f, ...)

**Curva de
Incidencia**

**A function to apply to
each element of the list**

(elementos de la lista)

incidence::fit()

**Other
arguments to
pass to the
function**

map()

map(**curve** , **fit** , ...)

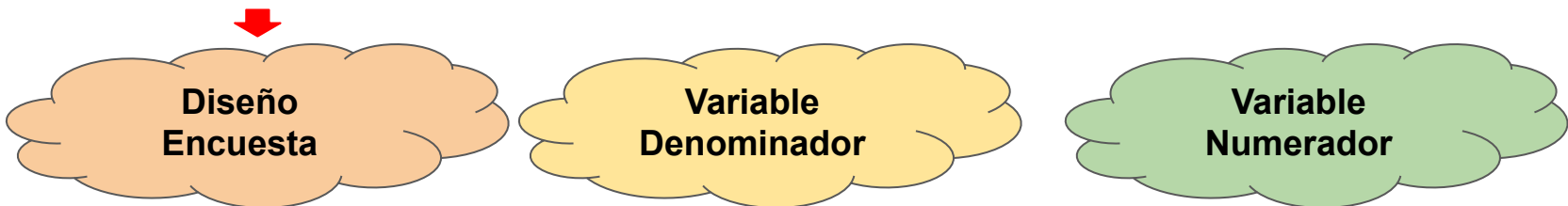
district 1
district 2
district 3
district 4
district 5



Prevalencia: Caso Singular (1/)

En Lima Metropolitana:

- Estrato: **Distrito**
- Unidad 1ra de Muestreo: **Conglomerados**
- Unidad 2ra de Muestreo: **Viviendas**



Dentro de los asintomáticos, ¿cuántos positivos hay?



Dentro de los positivos, ¿cuántos asintomáticos hay?



Prevalencia: Reto Múltiple (1/)

pmap()

Map over three or more lists. Put the lists into a list of list whose names match argument names in the function.

```
pmap(list(design, denom., numer.), fun, ...)
```

design	covar-01	outcome
design	covar-02	outcome
design	covar-03	outcome
design	outcome	covar-02
design	outcome	covar-03



Taller

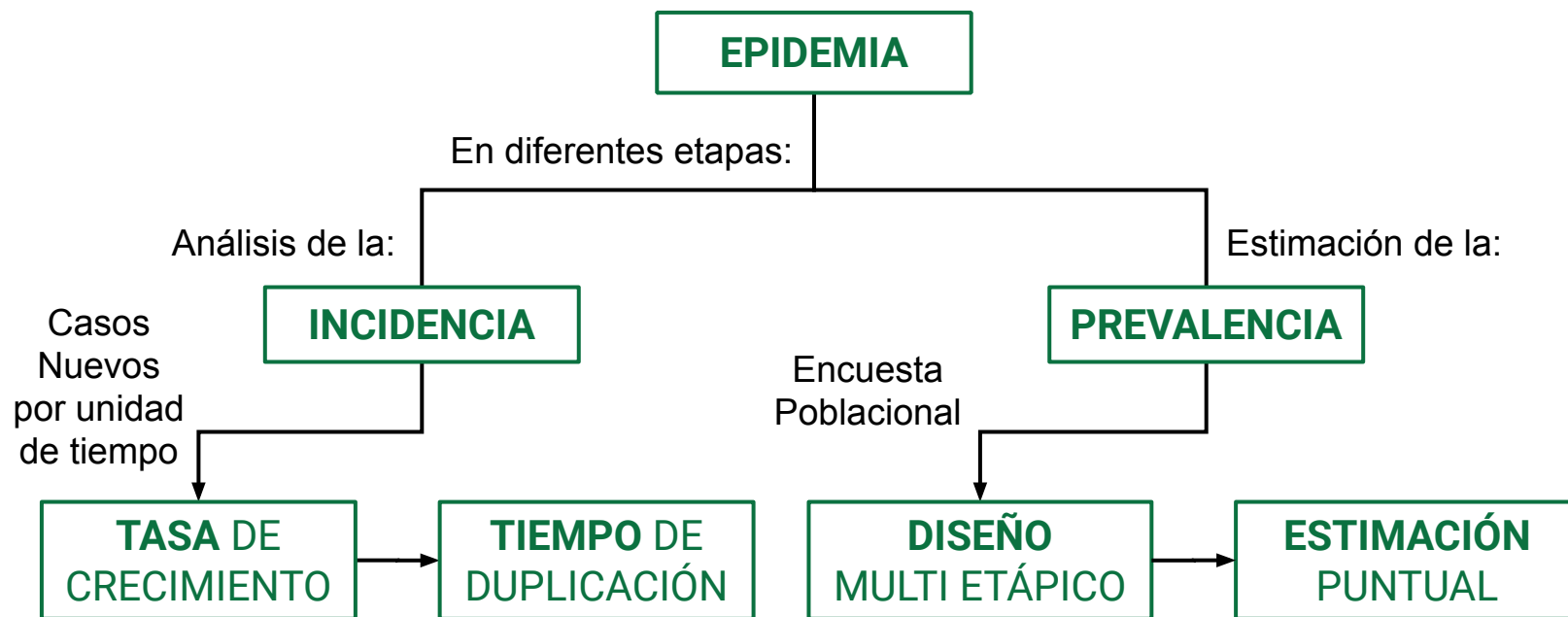


slides: <http://bit.ly/bbslisepi>



taller: <https://rstudio.cloud/project/1656284>

Retroalimentación



`incidence -> nest -> map`

`expand_grid -> sym -> pmap`

INTRODUCCIÓN

CASO SINGULAR

RETO MÚLTIPLE

Life hacks

 Más hacks [#reprex](#):

- <https://gist.github.com/avallecam>

 Más info:

- <https://trello.com/b/xtO9VP36/scibites>

¡Quedo atento a propuestas tuyas de mejora!

😎 **dplyr** al inicio es fácil... 💥 pero está evolucionando

👷 me adapté a **pivot_*** y **bang-bang** **{}}**

😞 pero los nuevos flujos con **(row|col)wise** aún no son intuitivos para mi

- <https://dplyr.tidyverse.org/articles/programming.html>
- <https://dplyr.tidyverse.org/articles/compatibility.html>
- <https://dplyr.tidyverse.org/articles/rowwise.html>
- <https://dplyr.tidyverse.org/articles/colwise.html>
- <https://columnwise-operations-dplyr.netlify.app/#1>

¡Gracias!

Andree Valle Campos

  @avallecam

 avallecam@gmail.com

