

## Ejercicio 2.17

Nahuel Bizoso, Santiago Robatto, Sofia Terra

2025-08-25

**Exercise 2.17 (Sick trees redux)** Repeat Exercise 2.5 utilizing simulation to *approximate* the posterior probability that a randomly selected maple tree has mold. Specifically, simulate data for 10,000 trees and remember to set your random number seed.

*Primero definiremos los estados del árbol. Es decir, creamos los estados “infectado” y “no infectado” y le asignamos su respectiva probabilidad. La misma es la probabilidad a priori. .*

```
library(tidyverse)
library(bayesrules)

infectados <- data.frame(infectado= c('infectado', 'no_infectado')) # dataframe con las posibilidades

priori <- c(0.18, 0.82) # probabilidades a priori
```

*Ahora realizaremos la simulación de árboles infectados y no infectados en base a las probabilidades a priori que cargamos anteriormente. De esta manera aproximamos la distribución de arboles infectados y no infectados, aún sin conocer la especie.*

```
set.seed(1233)

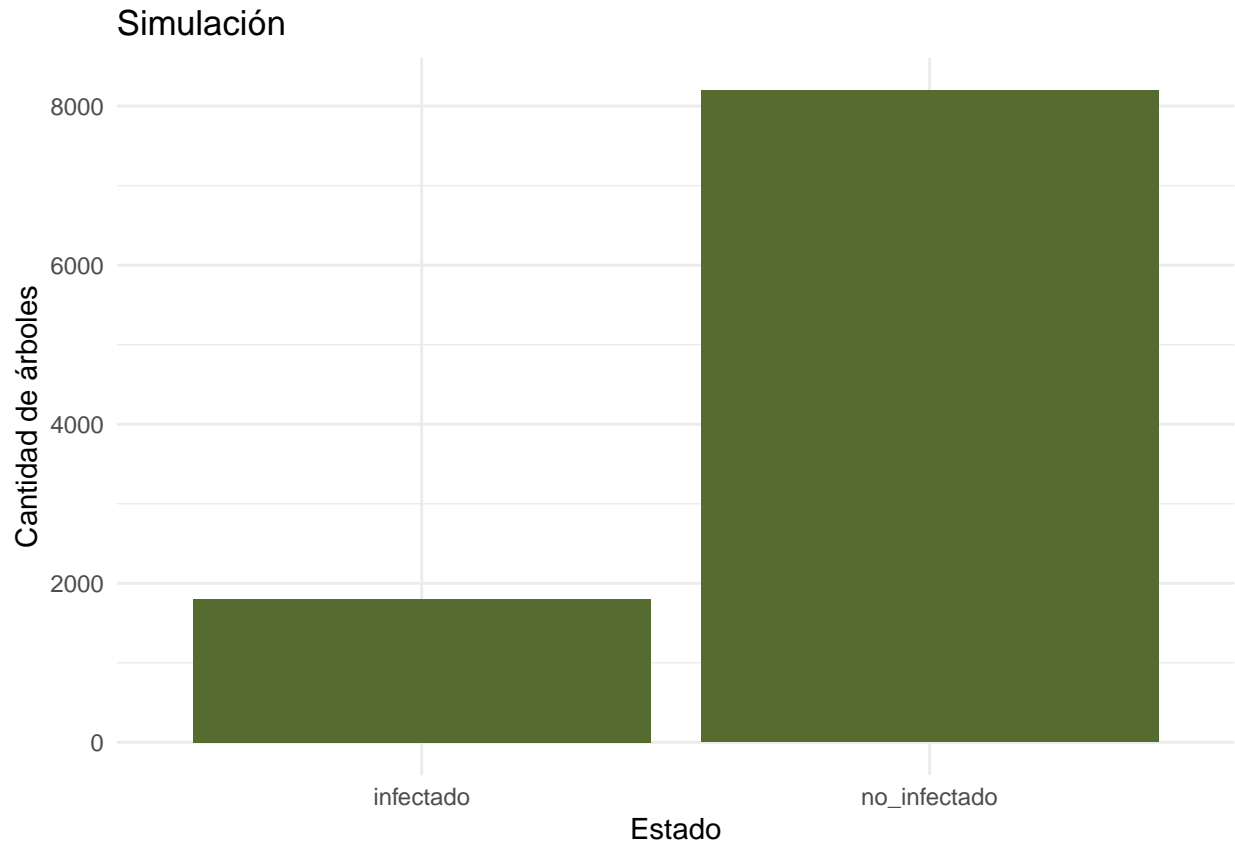
Simulacion <- sample_n(infectados, size= 10000, weight= priori, replace= TRUE )

#Realizamos la simulación con las probabilidades que le asignamos
#a cada suceso, La misma es realizada con reemplazo,
#para que las probabilidades se mantengan constantes.
```

En el siguiente gráfico visualizamos la distribución de árboles infectados en base a nuestros datos simulados:

```
library(RColorBrewer)

ggplot(Simulacion, aes(x= infectado)) +
  geom_bar(fill= '#556B2F' ) +
  labs(title= "Simulación", x= "Estado",
        y= "Cantidad de árboles") +
  theme_minimal()
```



En base a esto generaremos una tabla que nos muestre los valores porcentuales que se obtuvieron de la simulación.

```
library(janitor)
```

```
Simulacion %>%
  tabyl(infectado) %>%
  adorn_totals("row")
```

```
##      infectado      n percent
##      infectado  1802  0.1802
##    no_infectado  8198  0.8198
##           Total 10000  1.0000
```

Ya tenemos la simulación de los árboles infectados.

Ahora es de interés cargar la información sobre tipo de árbol. Como solo es de interés saber si es maple o no, generaremos una nueva variable Booleana que sea “Maple” o “No\_Maple”.

Por el ejercicio 2.5, sabemos que entre los infectados el 80% son maples y entre los no infectados 10% son maples.

En base a estas probabilidades simularemos la distribución de maples y no maples según árbol infectado o no.

```
Simulacion <- Simulacion %>%
  mutate(data_model = case_when(infectado == "infectado" ~ 0.8,
                                infectado == "no_infectado" ~ 0.1))
glimpse(Simulacion)
```

```
## Rows: 10,000
## Columns: 2
## $ infectado <chr> "no_infectado", "no_infectado", "no_infectado", "no_infecta~
## $ data_model <dbl> 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.1, 0.8,~
```

*#Se realiza la simulacion donde incluimos la probabilidad condicionada  
#de los maples dado si está infectado o no.*

Para definir la variable donde indica si es o no un maple crearemos un vector.

Teniendo dicha variable donde nos indica si es o no un Maple, simularemos la especie teniendo en cuenta las probabilidades condicionales cargadas anteriormente.

```
data<- c("Maple", "No Maple")
set.seed(123)

Simulacion <- Simulacion %>%
group_by(1:n()) %>%
mutate(maple = sample(data, size = 1, prob = c(data_model, 1 - data_model)))

#Lo que hicimos fue realizar la simulacion y añadimos una columna con 2
#categorias (Maple, No_Maple) donde alli se encuentra el resultado de la simulación
```

A partir de la simulación se obtiene la siguiente tabla de valores:

```
Simulacion %>%
tabyl(maple, infectado) %>%
adorn_totals(c("col", "row"))
```

```
##      maple infectado no_infectado Total
##      Maple      1420          768  2188
##      No Maple      382         7430  7812
##      Total      1802         8198 10000
```

Teniendo esta tabla de valores, filtraremos la información ya que nos interesa solamente los maples que están infectados.

Es decir queremos calcular la probabilidad a posteriori de que el árbol esté infectado dado que es un maple.

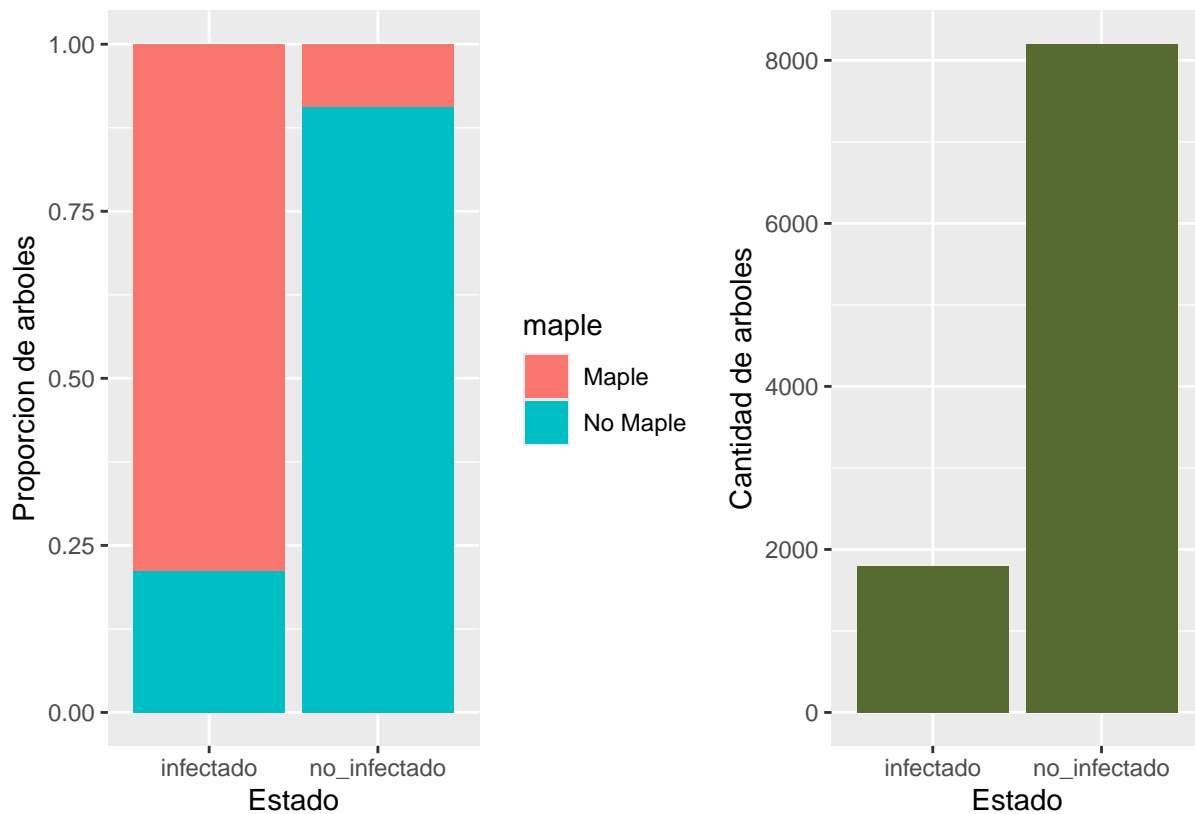
```
Maples_Inf <- Simulacion %>%
  filter(maple== "Maple") %>%
  tabyl(infectado) %>%
  adorn_totals("row")
Maples_Inf
```

```
##      infectado    n  percent
##      infectado 1420 0.6489945
##      no_infectado 768 0.3510055
##      Total    2188 1.0000000
```

En el gráfico de la izquierda, observamos la proporción de maples por infectado y no infectado. Claramente se observa que los maples representan alrededor del 80% de los árboles infectados. Esto es consistente con las probabilidades que cargamos para realizar la simulación. En el gráfico de la derecha se visualiza la cantidad de árboles infectados y no infectados.

```
library(patchwork)
p1<- ggplot(Simulacion, aes(x = infectado, fill = maple)) +
  geom_bar(position = "fill") +
  labs(x= "Estado",
       y= "Proporcion de arboles")

p2 <- ggplot(Simulacion, aes(x = infectado)) +
  geom_bar(fill= '#556B2F' )+
  labs(x= "Estado",
       y= "Cantidad de arboles")
p1+p2
```



Si bien los anteriores gráficos son útiles, entendemos que una manera alternativa de observar la información puede ser la siguiente, donde se visualiza la cantidad de infectados y no infectados separados por si es maple o no.

```
ggplot(Simulacion, aes(x = infectado)) +
  geom_bar(fill= '#556B2F') +
  facet_wrap(~ maple) +
  labs(x= "Estado",
       y= "Cantidad de arboles")
```

