

# Baloto

Andres Rodriguez

2023-06-25

## Extraccion de datos

Los datos fueron obtenidos de la pagina oficial de Baloto “<https://www.balotoresultados.co/historico>” a partir del año 2017

El dataset compuesto de :

Variable	Tipo	Descripción
Fecha	Date	Fecha del sorteo
Balota_1	Int	Resultado de la balota 1
Balota_2	Int	Resultado de la balota 2
Balota_3	Int	Resultado de la balota 3
Balota_4	Int	Resultado de la balota 4
Balota_5	Int	Resultado de la balota 5
Super_balota	Int	Resultado de la superbalota

Se cargan los datos y ajustan formatos de variables

```
baloto <- read.csv("Data/baloto.csv")
names(baloto) <- c("Fecha", "Balota_1", "Balota_2", "Balota_3", "Balota_4", "Balota_5", "Super_balota")
baloto$Fecha <- as.Date(baloto$Fecha, format = '%m/%d/%Y')
baloto$anio = year(baloto$Fecha)
# Se muestran los primeros 5 registros del dataset
head(baloto)
```

Fecha	Balota_1	Balota_2	Balota_3	Balota_4	Balota_5	Super_balota	anio
2017-04-22	15	28	36	37	41	10	2017
2017-04-26	1	2	25	31	33	14	2017
2017-04-29	6	19	25	28	36	1	2017
2017-05-03	5	6	7	12	37	6	2017
2017-05-06	1	20	36	39	43	10	2017
2017-05-10	12	19	22	35	38	1	2017

```
# Se comprueba formato de las variables
str(baloto)
```

```
## 'data.frame':    603 obs. of  8 variables:
## $ Fecha      : Date, format: "2017-04-22" "2017-04-26" ...
## $ Balota_1   : int  15 1 6 5 1 12 11 15 1 7 ...
## $ Balota_2   : int  28 2 19 6 20 19 19 21 11 22 ...
## $ Balota_3   : int  36 25 25 7 36 22 27 26 28 29 ...
## $ Balota_4   : int  37 31 28 12 39 35 28 28 31 32 ...
## $ Balota_5   : int  41 33 36 37 43 38 37 36 41 41 ...
## $ Super_balota: int  10 14 1 6 10 1 8 3 9 14 ...
## $ anio       : num  2017 2017 2017 2017 2017 ...
```

## Analisis descriptivo

Se revisan estadísticos básicos y aquí algunas particularidades que cuando menos resultan interesantes:

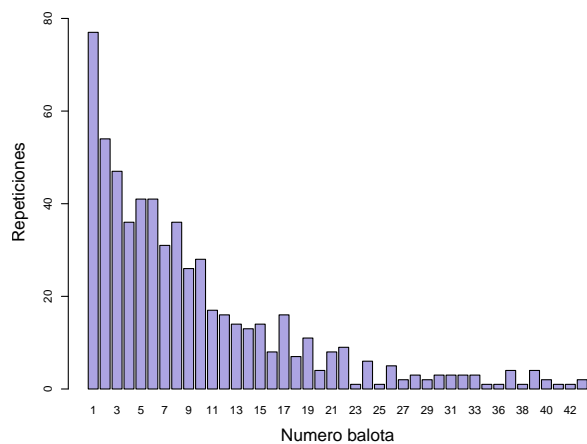
- Parece que la balota N° 4 no ha sacado valor 1 por lo menos desde la muestra de los datos.

Siendo un juego de azar quizá se podría esperar una distribución de frecuencias uniforme, sin embargo:

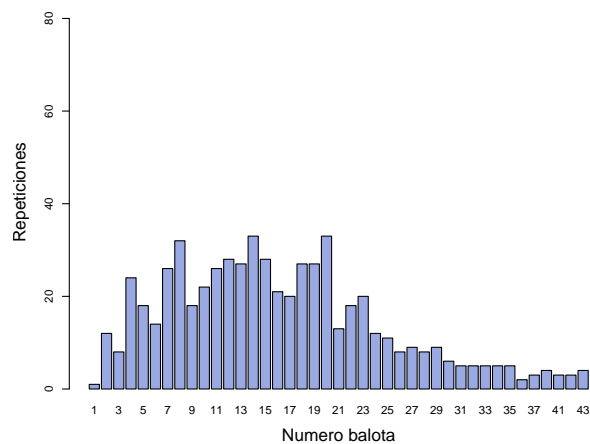
1. La balota N° 1 concentra el 75% de sus resultados con valores  $\leq 13$
2. Lo contrario sucede con la balota N° 5 en la que sus resultados (50%) están concentrados en 13% de sus valores.
3. El resto de balotas parecen tener un comportamiento un poco mas “uniforme” aunque igualmente se pueden denotar algunas tendencias.

Fecha	Balota_1	Balota_2	Balota_3	Balota_4	Balota_5	Super_balota	anio
Min.	Min. :	Min. :	Min. :	Min. :	Min. :	Min. :	Min.
:2017-04-22	1.00	1.00	1.00	2.00	1.0	1.000	:2017
1st Qu.:2018-10-01	1st Qu.:3.00	1st Qu.:9.00	1st Qu.:15.00	1st Qu.:23.00	1st Qu.:31.0	1st Qu.:4.000	1st Qu.:2018
Median	Median :	Median :	Median :	Median :	Median :	Median :	Median
:2020-03-11	7.00	:15.00	:22.00	:30.00	:37.0	9.000	:2020
Mean	Mean :	Mean :	Mean :	Mean :	Mean :	Mean :	Mean
:2020-03-28	9.34	:16.13	:22.18	:27.92	:34.3	8.476	:2020
3rd Qu.:2021-09-23	3rd Qu.:13.00	3rd Qu.:21.00	3rd Qu.:28.00	3rd Qu.:35.00	3rd Qu.:41.0	3rd Qu.:12.000	3rd Qu.:2021
Max.	Max. :	Max. :	Max. :	Max. :	Max. :	Max. :	Max.
:2023-03-08	:43.00	:43.00	:43.00	:43.00	:43.0	:16.000	:2023

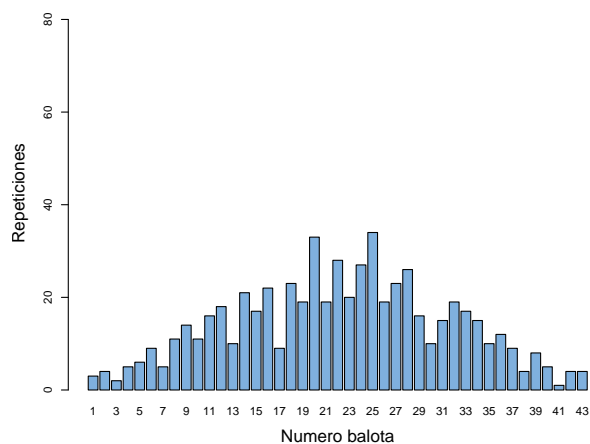
**Frecuencias balota 1**



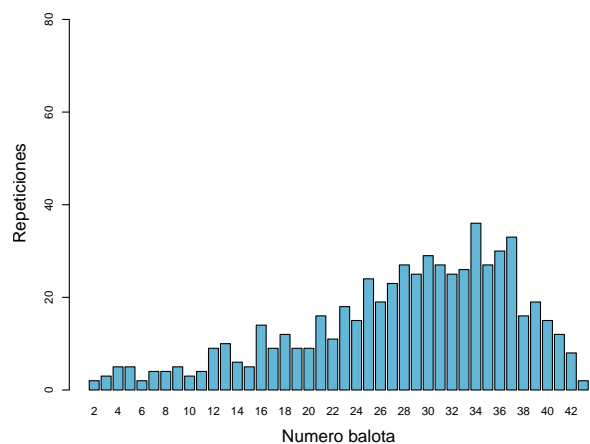
**Frecuencias balota 2**



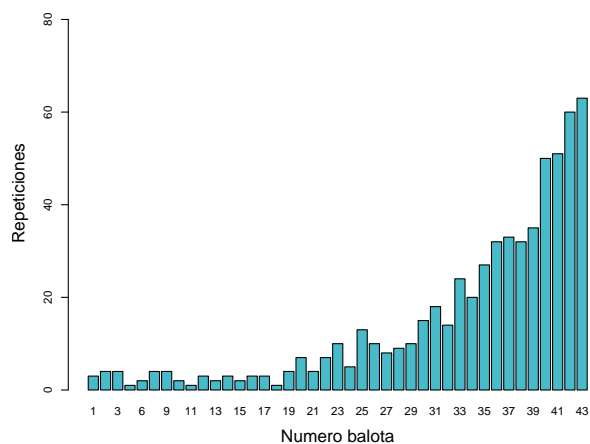
**Frecuencias balota 3**



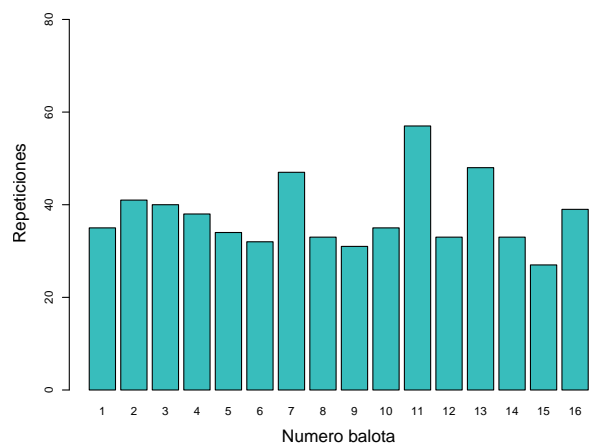
**Frecuencias balota 4**



**Frecuencias balota 5**



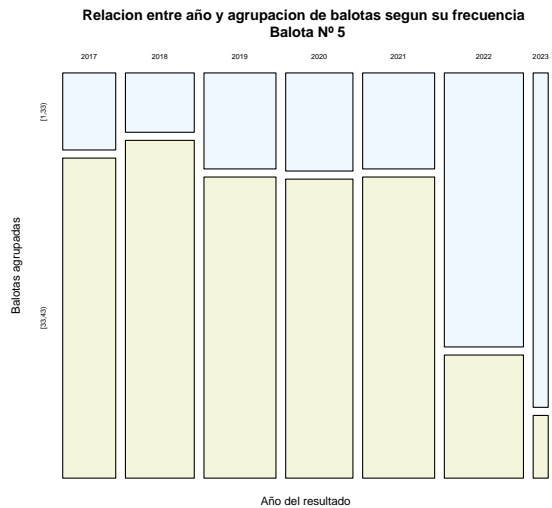
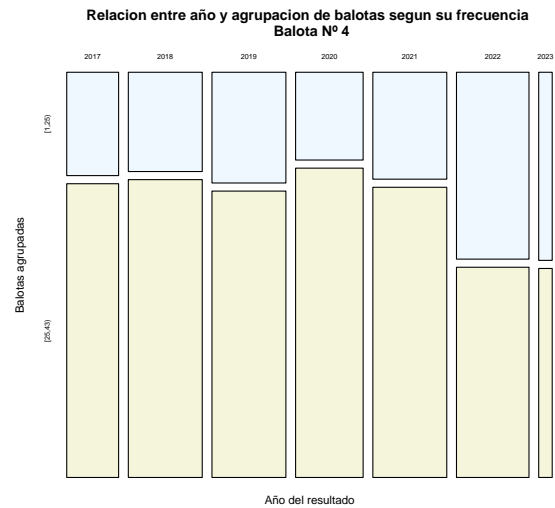
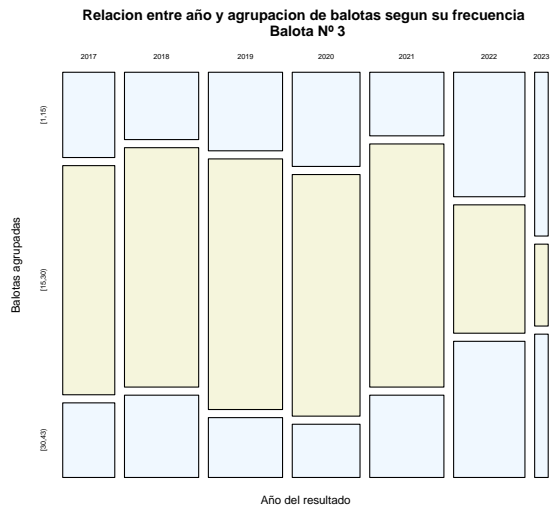
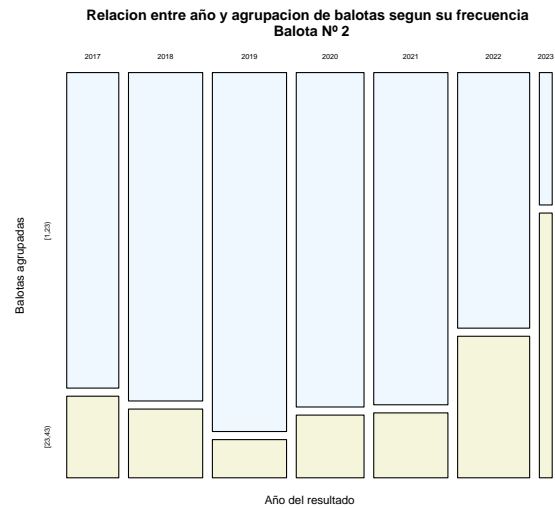
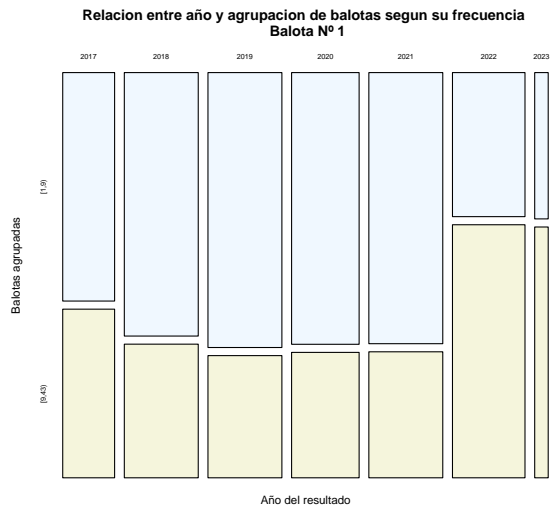
**Frecuencias super balota**



Valor medio resultado de cada balota con el rango de desviación muestral:

<b>Variable</b>	<b>Balota Media</b>	<b><math>\pm</math> Balotas</b>
Balota_1	9	9
Balota_2	16	9
Balota_3	22	9
Balota_4	28	9
Balota_5	34	9
Super_balota	8	5

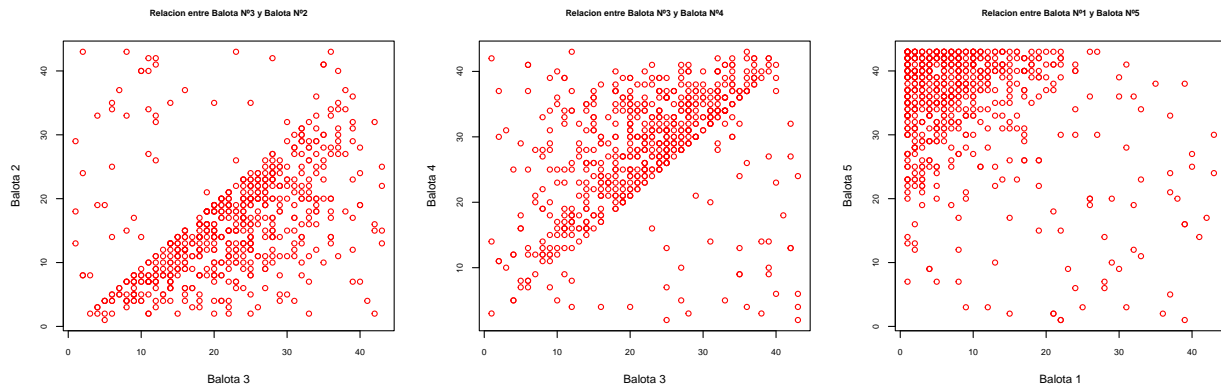
Agrupando variables



Una Particular relacion se encuentra entre la balota n° 3 y balota n°4 ya que pareciere indicar:

1. Correlacion directa entre balota N° 3 y N°2
2. COrrrelacion inversa entre balota N° 3 y N°4

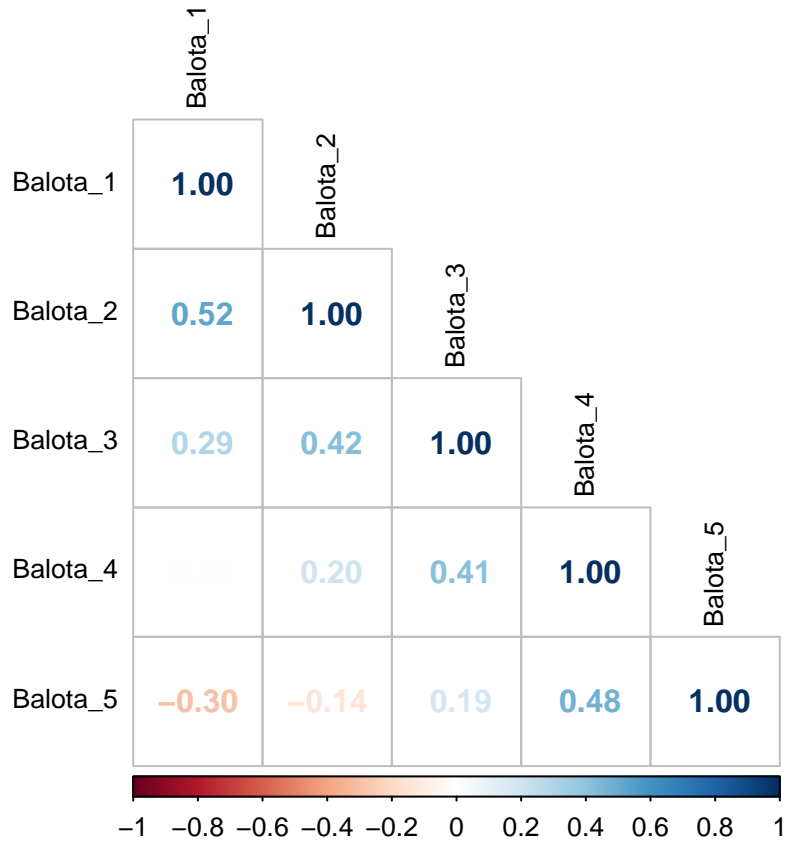
```
par(mfrow = c(1, 3))
plot(baloto$Balota_3, baloto$Balota_2, main = "Relacion entre Balota N°3 y Balota N°2", cex.main = 0.8,
     cex.axis = 0.8, ylab = 'Balota 2', xlab = 'Balota 3', col = 'red')
plot(baloto$Balota_3, baloto$Balota_4, main = "Relacion entre Balota N°3 y Balota N°4", cex.main = 0.8,
     cex.axis = 0.8, ylab = 'Balota 4', xlab = 'Balota 3', col = 'red')
plot(baloto$Balota_1, baloto$Balota_5, main = "Relacion entre Balota N°1 y Balota N°5", cex.main = 0.8,
     cex.axis = 0.8, ylab = 'Balota 5', xlab = 'Balota 1', col = 'red')
```



Para observar mejor esta relación, se agrega matriz de correlación entre las balotas:

```
## Warning: package 'corrplot' was built under R version 4.3.1
```

```
## corrplot 0.92 loaded
```



Mas alla de tendencias que se puedan observar y como estas cambian en el tiempo, al ser un juego de azar vemos como no se mantienen o relacionan variables de los resultados entre si, si bien algunos datos y valores que definitivamente siguen siendo vagos en función de buscar un patrón claro, para llegar a esto se deberían poder incluir mas variables y factores en la ejecución de cada sorteo, (temperatura, peso de las balotas, presión al interior de cada bolsa etc) la descriptiva de los resultados, quiza pueda ayudarte a decidir mejor que opciones tomar al momento de jugar.

Y no olvides que mas alla de alguna tendencia o valor repetitivo tu probabilidad de ganarte el baloto es del **0.0000006%**