

# Taller de R básico. Día 3.

*Viacheslav Shalisko*

*19 de octubre de 2016*

## A. Cargar la tabla de datos

Estructura de datos (tabla `Datos_del_censo.csv`):

1. *Centro* - código del CU
2. *Especie* - nombre científico
3. *Codigo* - identificador único del árbol
4. *AB* - área basal del árbol (dm<sup>2</sup>)
5. *DTr* - diametro del tronco equivalente (cm)
6. *Alt* - estatura del árbol (m)
7. *DCop* - diametro promedio de la copa (m)
8. *ExcCop* - excentricidad de la copa

```
arbolado <- read.csv("Materiales/Datos_del_censo.csv")
dim(arbolado)
```

```
## [1] 4785    8
```

```
head(arbolado)
```

```
##      Centro      Especie      Codigo      AB  DTr  Alt  DCop  ExcCop
## 1 CUCOSTA  Pithecellobium dulce  C10-3-1 13.21 41.0 9.8 15.4   0.30
## 2 CUCOSTA   Acacia macracantha  C10-3-10  1.91 15.6 4.4  4.7   0.63
## 3 CUCOSTA Pithecellobium lanceolatum C10-3-100  0.95 11.0 6.4  9.5   0.91
## 4 CUCOSTA   Salix bonplandiana C10-3-1000  0.83 10.3 7.7  4.9   0.20
## 5 CUCOSTA   Tabebuia rosea C10-3-1001  0.28  6.0 2.4  2.0   0.63
## 6 CUCOSTA Pithecellobium lanceolatum C10-3-1002  0.51  8.1 4.1  4.3   0.71
```

```
tail(arbolado)
```

```
##      Centro      Especie      Codigo      AB  DTr  Alt  DCop  ExcCop
## 4780 CUCSUR Araucaria heterophylla C9-3-94  6.97 29.8 16.2  5.5   0.50
## 4781 CUCSUR   Ficus benjamina C9-3-95 17.60 47.3 11.5 10.2   0.38
## 4782 CUCSUR   Simarouba glauca C9-3-96  2.63 18.3 11.5  9.4   0.58
## 4783 CUCSUR  Jacaranda mimosifolia C9-3-97 76.95 99.0 20.6 13.0   0.57
## 4784 CUCSUR   Jatropha cordata C9-3-98  2.11 16.4  2.7  3.6   0.23
## 4785 CUCSUR  Jacaranda mimosifolia C9-3-99  7.55 31.0 18.2  5.2   0.77
```

```
str(arbolado)
```

```
## 'data.frame': 4785 obs. of 8 variables:
## $ Centro : Factor w/ 4 levels "CUALTOS","CUCIENEGA",...: 3 3 3 3 3 3 3 3 3 3 ...
## $ Especie: Factor w/ 150 levels "Acacia farnesiana",...: 104 4 105 127 139 105 105 105 139 12
7 ...
## $Codigo : Factor w/ 4780 levels "C10-3-1","C10-3-10",...: 1 2 3 4 5 6 7 8 9 10 ...
## $ AB : num 13.21 1.91 0.95 0.83 0.28 ...
## $ DTr : num 41 15.6 11 10.3 6 8.1 11.6 9 6 15.8 ...
## $ Alt : num 9.8 4.4 6.4 7.7 2.4 4.1 4.4 4.5 4.3 11.6 ...
## $ DCop : num 15.4 4.7 9.5 4.9 2 4.3 5.2 4.3 2.5 5.1 ...
## $ ExcCop : num 0.3 0.63 0.91 0.2 0.63 0.71 0.84 0.84 0.75 0.42 ...
```

```
levels(arbolado$Centro)
```

```
## [1] "CUALTOS" "CUCIENEGA" "CUCOSTA" "CUCSUR"
```

## B. Estadística descriptiva

### Resumen

```
summary(arbolado[,c(1,2,4:8)])
```

```
##          Centro          Especie          AB
## CUALTOS :1828 Quercus resinosa : 530 Min. : 0.200
## CUCIENEGA: 445 Pithecellobium lanceolatum: 501 1st Qu.: 0.950
## CUCOSTA :2143 Guazuma ulmifolia : 329 Median : 2.780
## CUCSUR : 369 Fraxinus uhdei : 278 Mean : 5.817
## Eysenhardtia polystachya : 245 3rd Qu.: 6.610
## Acacia macracantha : 176 Max. :456.340
## (Other) :2726
##          DTr          Alt          DCop          ExcCop
## Min. : 5.00 Min. : 1.500 Min. : 0.200 Min. :0.0000
## 1st Qu.: 11.00 1st Qu.: 4.200 1st Qu.: 3.800 1st Qu.:0.3700
## Median : 18.80 Median : 6.200 Median : 5.500 Median :0.5100
## Mean : 22.22 Mean : 7.121 Mean : 6.165 Mean :0.5026
## 3rd Qu.: 29.00 3rd Qu.: 9.100 3rd Qu.: 7.900 3rd Qu.:0.6500
## Max. :241.00 Max. :26.600 Max. :64.500 Max. :0.9900
##          NA's :46
```

Las funciones básicas de la estadística descriptiva incluyen:

1. Promedio y mediana - `mean()` , `median()`
2. Mínimo, máximo y rango - `min()` , `max()` , `range()`
3. Desviación estandar y varianza - `sd()` , `var()`
4. Cuantiles y rango intercuartilico - `quantile()` , `IQR()`
5. Suma y número de elementos - `sum()` , `length()`

```
summary(arbolado$AB)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.200   0.950   2.780   5.817   6.610 456.300
```

```
range(arbolado$AB)
```

```
## [1]    0.20 456.34
```

```
quantile(arbolado$AB)
```

```
##      0%      25%      50%      75%     100%
##    0.20    0.95    2.78    6.61 456.34
```

```
IQR(arbolado$AB)
```

```
## [1] 5.66
```

```
var(arbolado$AB)
```

```
## [1] 173.8403
```

```
sqrt(var(arbolado$AB))
```

```
## [1] 13.18485
```

```
sd(arbolado$AB)
```

```
## [1] 13.18485
```

```
probabilidades <- c(0.01,0.05,0.1,0.5,0.9,0.95,0.99)
quantile(arbolado$AB, probs = probabilidades)
```

```
##      1%      5%     10%     50%     90%     95%     99%
##    0.200  0.270  0.410  2.780 12.956 19.630 44.436
```

```
boxplot.stats(arbolado$AB)$stats
```

```
## [1]  0.20  0.95  2.78  6.61 15.00
```

```
boxplot.stats(arbolado$AB, coef = 3)$stats
```

```
## [1] 0.20 0.95 2.78 6.61 23.52
```

## C. Algunas funciones de control

### Estadística por grupos

```
aggregate(arbolado$AB ~ arbolado$Centro, FUN = sum)
```

```
##   arbolado$Centro arbolado$AB
## 1      CUALTOS      8521.96
## 2     CUCIENEGA      3478.53
## 3      CUCOSTA     12647.99
## 4      CUCSUR      3184.82
```

```
aggregate(arbolado$AB ~ arbolado$Centro, FUN = length)
```

```
##   arbolado$Centro arbolado$AB
## 1      CUALTOS      1828
## 2     CUCIENEGA       445
## 3      CUCOSTA     2143
## 4      CUCSUR       369
```

```
aggregate(arbolado$AB ~ arbolado$Centro, FUN = mean)
```

```
##   arbolado$Centro arbolado$AB
## 1      CUALTOS    4.661904
## 2     CUCIENEGA    7.816921
## 3      CUCOSTA    5.902002
## 4      CUCSUR     8.630949
```

```
aggregate(arbolado$AB ~ arbolado$Centro, FUN = range)
```

```
##   arbolado$Centro arbolado$AB.1 arbolado$AB.2
## 1      CUALTOS      0.20      114.36
## 2     CUCIENEGA      0.23       77.44
## 3      CUCOSTA      0.20     456.34
## 4      CUCSUR      0.21     419.84
```

### Estadística por columnas y/o grupos

El grupo de funciones apply permite realizar operaciones de forma cíclica sin necesidad de definir un ciclo de manera formal:

1. `lapply()` - aplicar una función a cada elemento de una lista, resultado es una lista de mismo largo que lista de entrada
2. `sapply()` - lo mismo que `lapply()`, pero con resultado simplificado (vector, matriz o array)

3. `apply()` - aplicar una función sobre dimensiones de un matriz de datos
4. `tapply()` - aplicar una función sobre grupos definidos en un vector, similar a `aggregate()`
5. `mapply()` - versión multivariante de `tapply()`

### Compara `tapply()` y `aggregate()`

```
tapply(arbolado$AB, arbolado$Centro, FUN = mean)
```

```
##    CUALTOS CUCIENEGA  CUCOSTA  CUCSUR
## 4.661904  7.816921  5.902002  8.630949
```

```
aggregate(arbolado$AB ~ arbolado$Centro, FUN = mean)
```

```
##   arbolado$Centro arbolado$AB
## 1          CUALTOS  4.661904
## 2        CUCIENEGA  7.816921
## 3          CUCOSTA  5.902002
## 4          CUCSUR  8.630949
```

### Compara `lapply()` y `sapply()`

```
lapply(arbolado[,4:8],sd,na.rm = TRUE)
```

```
## $AB
## [1] 13.18485
##
## $DTr
## [1] 15.71243
##
## $Alt
## [1] 3.847538
##
## $DCop
## [1] 3.352365
##
## $ExcCop
## [1] 0.2009095
```

```
sapply(arbolado[,4:8],sd,na.rm = TRUE)
```

```
##      AB      DTr      Alt      DCop      ExcCop
## 13.1848514 15.7124261 3.8475377 3.3523652 0.2009095
```

```
sapply(arbolado[,4:8],range,na.rm = TRUE)
```

```
##          AB DTr  Alt DCop ExcCop
## [1,]    0.20  5   1.5  0.2   0.00
## [2,] 456.34 241 26.6 64.5   0.99
```

```
sapply(arbolado[,4:8],function(x) length(x[!is.na(x)]))
```

```
##      AB      DTr      Alt      DCop ExcCop
## 4785  4785  4739  4785  4785
```

Nota que el parámetro `na.rm = TRUE` permite omitir valores sin datos en las funciones

```
sd(arbolado$Alt, na.rm = TRUE)
```

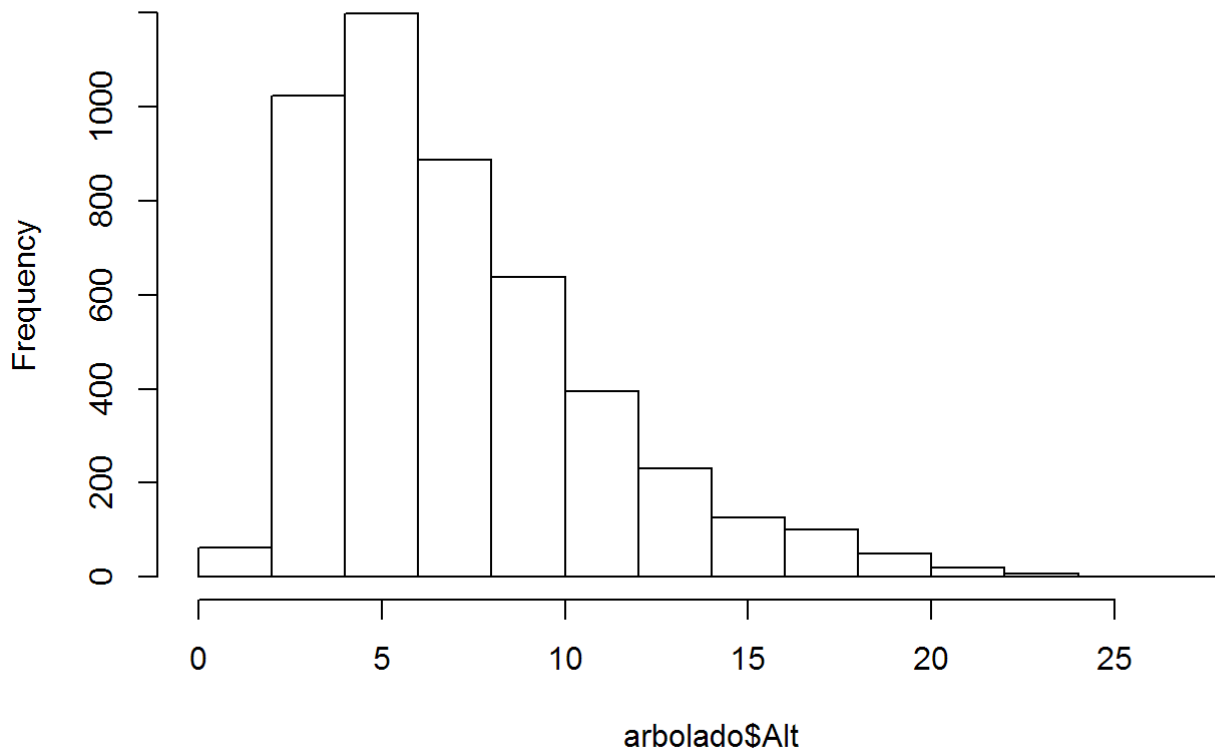
```
## [1] 3.847538
```

## D. Gráficas simples

### Histogramas

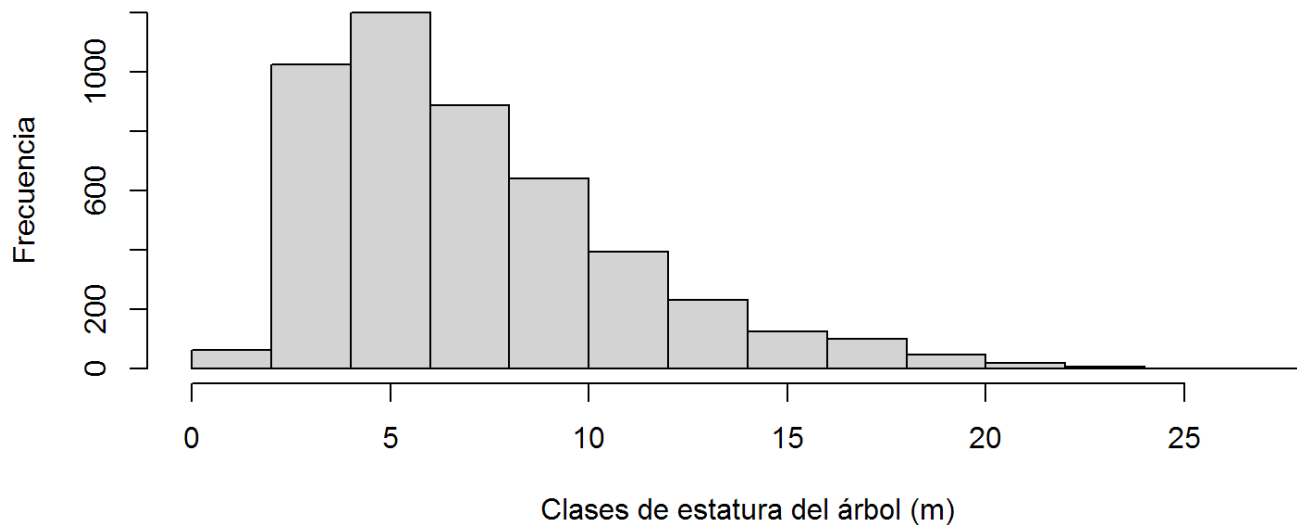
```
hist(arbolado$Alt)
```

**Histogram of arbolado\$Alt**

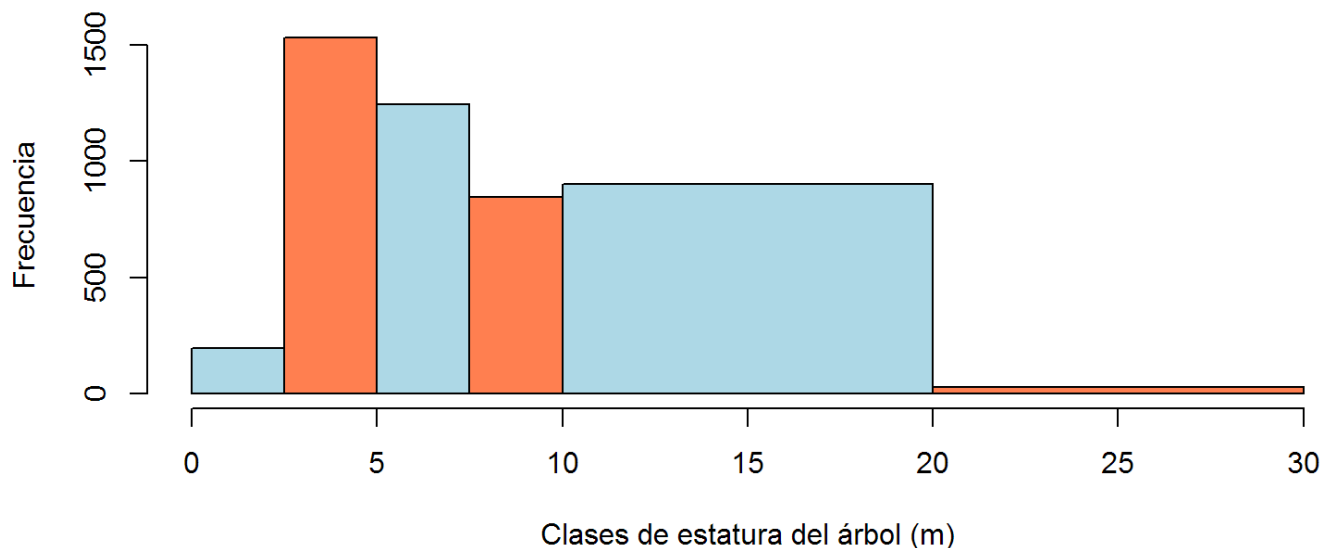


Las graficas pueden ser controlados por medio de una gran cantidad de parámetros.

```
hist(arbolado$Alt,  
     col = "lightgray",  
     xlab = "Clases de estatura del árbol (m)", ylab = "Frecuencia",  
     main = "Histograma de estaturas")
```

**Histograma de estaturas**

```
H0 <- hist(arbolado$Alt,  
           col = c("lightblue","coral"),  
           xlab = "Clases de estatura del árbol (m)", ylab = "Frecuencia",  
           main = "Histograma de estaturas",  
           breaks = c(0,2.5,5,7.5,10,20,30), prob = FALSE)
```

**Histograma de estaturas**

```
str(H0)
```

```
## List of 6
## $ breaks : num [1:7] 0 2.5 5 7.5 10 20 30
## $ counts : int [1:6] 196 1529 1242 844 900 28
## $ density : num [1:6] 0.0165 0.1291 0.1048 0.0712 0.019 ...
## $ mids : num [1:6] 1.25 3.75 6.25 8.75 15 25
## $ xname : chr "arbolado$Alt"
## $ equidist: logi FALSE
## - attr(*, "class")= chr "histogram"
```

---

A continuar el día Jueves 20 de octubre del 2016