

Asignación 1: Introducción al lenguaje R

Dr. Marco Aurelio González Tagle

05 /08/ 2019

Índice

Objetivo	1
Indicaciones	1
Importar datos de trabajo	1
Selección de datos	3
Visualización de datos	4
Estadísticas básicas	4
Líneas de comando en R	4
Referencias	5

Objetivo

El objetivo de la siguiente asignación es la selección de un conjunto de datos en *R* mediante restricciones.

Indicaciones

Realice las actividades indicadas en los siguientes apartados de la actividad 1. Recuerde que el script debe estar 100 % funcional y compilado en formato *html* o *pdf*.

Importar datos de trabajo

Los datos de trabajo provienen del libro *Introductory probability & statistics, applications for forestry & natural sciences* de Kozak et al. (2008). El Cuadro 1 muestra un ejemplo de datos coletados de 50 árboles con siete variables. (i) Árbol se refiere al número de árbol, (ii) Fecha: mes de colecta en Marzo 2006; (iii) Especie: C: Cedro Rojo (Western red cedar); F: Douglasia verde (Douglas fir); H: Tsuga heterófila (western hemlock); (iv) Posición: clasificación de la copa: D: Dominante, C: codominante, I: Intermedio, S: suprimido; (v) Vecinos: número de vecinos en un radio de 5m; (vi) Diámetro: diámetro a la altura de pecho (1.3m); (vii) Altura total.

Cuadro 1: Conjunto de datos que contiene información colectada de 50 árboles.

Árbol	Fecha	Especie	Posición	Vecinos	Diámetro	Altura
1	12	F	C	4	15.3	14.78
2	12	F	D	3	17.8	17.07
3	9	C	D	5	18.2	18.28
4	9	H	S	4	9.7	8.79

Árbol	Fecha	Especie	Posición	Vecinos	Diámetro	Altura
5	7	H	I	6	10.8	10.18
6	10	C	I	3	14.1	14.90
7	10	C	C	2	17.1	15.34
8	12	C	D	2	20.6	17.22
9	16	F	C	4	18.2	15.15
10	14	F	I	5	16.1	14.66
11	8	H	D	3	14.2	17.43
12	5	H	D	6	14.8	17.45
13	12	F	I	2	19.1	14.18
14	5	C	I	2	16.7	13.40
15	12	C	S	4	18.9	10.40
16	20	H	S	3	12.4	11.52
17	15	H	C	0	17.3	14.61
18	20	F	D	1	22.7	21.46
19	15	C	C	4	15.1	17.82
20	14	C	I	3	17.7	11.38
21	14	C	S	5	13.4	8.50
22	13	C	I	4	16.2	12.8
23	14	F	D	1	18.5	18.71
24	20	F	I	4	15.0	14.48
25	21	F	C	2	18.8	14.81
26	5	H	I	4	15.8	12.01
27	2	H	I	3	16.1	11.70
28	22	C	C	3	15.4	16.03
29	22	C	I	0	17.8	14.46
30	18	C	S	1	18.5	8.47
31	16	C	I	3	14.1	11.22
32	16	C	C	5	14.8	12.34
33	17	F	C	4	15.5	16.79
34	17	F	I	6	13.8	16.06
35	18	F	S	4	13.0	13.20
36	20	H	C	2	18.2	14.30
37	22	H	C	0	22.3	16.84
38	20	H	I	3	17.8	13.84
39	17	C	I	4	13.1	11.31
40	17	C	I	6	12.8	13.20
41	16	C	C	3	13.3	13.75
42	23	F	C	3	15.6	14.60
43	23	H	C	4	16.6	12.56
44	22	C	I	5	13.0	10.88
45	24	C	I	4	10.2	13.93
46	23	F	I	3	14.4	12.68
47	24	C	S	6	7.7	10.00
48	25	C	S	5	9.9	8.69
49	25	H	D	1	20.4	16.73
50	24	H	D	3	20.9	16.25

- Ingresar los datos del inventario del cuadro 1 a Excel (omitir acentos en Excel).
- Importar la base de datos a excel a R en un objeto llamada *conjunto*.

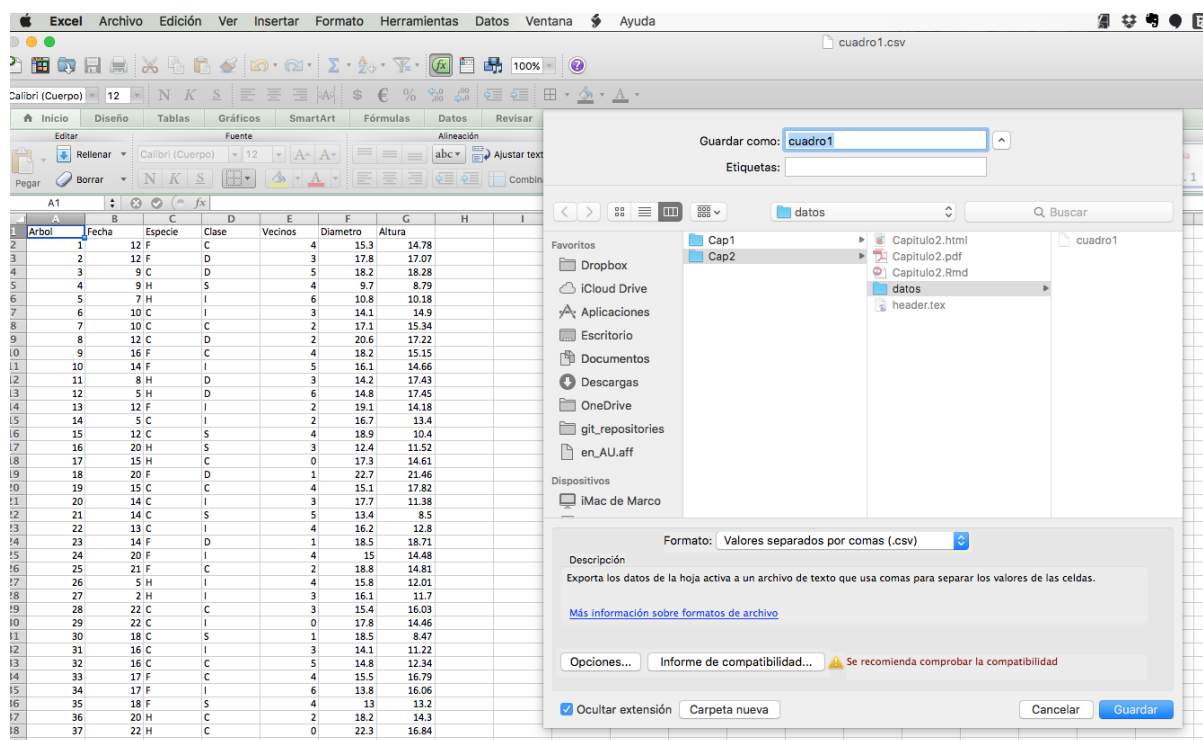


Figura 1: Guardar los datos de Excel en formato .csv para importar los datos a la consola de R.

- El objeto conjunto debe contener 7 variables y 50 observaciones para realizar sus actividades.

Selección de datos

- Aplicar la función `subset` para la variable **Altura** de acuerdo a las siguientes indicaciones:
 - Incluir los datos iguales o menores a la media (objeto en R se llame: **H.media**)
 - Incluir los datos menores a 16.5 m (objeto en R se llame: **H.16**)
- Aplicar la función `subset` para la variable **Vecinos**
 - Incluir los árboles que tengan un número de vecinos iguales o menores a 3 (Objeto en R: **Vecinos-3**)
 - Incluir los árboles que tengan un número de vecinos mayores a 4 (Objeto en R: **Vecinos-4**)
- Aplicar la función `subset` para la variable **Diametro**
 - Incluir los diámetros menores a la media (objeto en R: **DBH-media**)
 - Incluir los diámetros mayores a 16 (Objeto en R **DBH-16**)
- Aplicar la función `subset` para la variable **Especie**
 - Incluir la especie **Cedro Rojo**
 - Incluir la especie **Tsuga heterófila** y **Douglasia verde**
- Determinar cuantas observaciones son menores o iguales a 16.9 cm de **Diametro**
- Determinar cuantas observaciones son mayores a 18.5 metros de **Altura**

Visualización de datos

Con la función `hist` generar los histogramas para los objetos creados en el apartado anterior

- Altura, H.media y H.16
- Vecinos, Vecinos-3, Vecinos-4
- Diametro, DBH-media, DBH-16

Estadísticas básicas

Determinar la media (`mean`) de los objetos (variable y respectivos subsets), así como su desviación estándar (`sd`).

- Altura, H.media y H.16
- Vecinos, Vecinos-3, Vecinos-4
- Diametro, DBH-media, DBH-16

Líneas de comando en R

Los datos pueden descargarse del servidor de *dropbox* utilizando la paquetería *repmis* utilizando el siguiente código

```
library(repmis)

## Registered S3 method overwritten by 'R.oo':
##   method      from
##   throw.default R.methodsS3

conjunto <- source_data("https://www.dropbox.com/s/hmsf07bbayxv6m3/cuadro1.csv?dl=1")

## Downloading data from: https://www.dropbox.com/s/hmsf07bbayxv6m3/cuadro1.csv?dl=1
## SHA-1 hash of the downloaded data file is:
## 2bdde4663f51aa4198b04a248715d0d93498e7ba

head(conjunto)

##   Arbol Fecha Especie Clase Vecinos Diametro Altura
## 1     1    12      F     C      4     15.3   14.78
## 2     2    12      F     D      3     17.8   17.07
## 3     3     9      C     D      5     18.2   18.28
## 4     4     9      H     S      4      9.7    8.79
## 5     5     7      H     I      6     10.8   10.18
## 6     6    10      C     I      3     14.1   14.90
```

O se pueden importar guardando los datos en Excel y despues importarlos

```
conjunto <- read.csv("cuadro1.csv", header=TRUE)
```



Referencias

Kozak, A, RA Kozak, CL Staudhammer, and SB Watte. 2008. *Introductory Probability & Statistics. Applications for Forestry & the Natural Sciences*. Cambridge: Cambridge University Press.