Semana 2

Juan Carlos GirÃƒÂ³n

June 5, 2018

## Objetivos

1. Cargar Archivos de Texto en url a R.
2. Crear conexiones a url para cargar archivos Rdata.
3. Cargar archivos formato “dat”.
4. Obtener registros con base al nombre de fila.

## DAT file

Time-Series Analyses of Beaver Body Temperatures Penny S. Reynolds. Longitudinal data Model selection Regression diagnostics Sampling Time series

df <- read.table("http://www.stats.ox.ac.uk/pub/datasets/csb/ch11b.dat",header = FALSE)  
head(df)

## V1 V2 V3 V4 V5  
## 1 1 307 930 36.58 0  
## 2 2 307 940 36.73 0  
## 3 3 307 950 36.93 0  
## 4 4 307 1000 37.15 0  
## 5 5 307 1010 37.23 0  
## 6 6 307 1020 37.24 0

## Txt file from http:

Decathlon results

data<-read.table(file="http://www.sthda.com/upload/decathlon.txt")   
head(data)

## V1 V2 V3 V4 V5 V6 V7 V8  
## 1 name 100m Long.jump Shot.put High.jump 400m 110m.hurdle Discus  
## 2 SEBRLE 11.04 7.58 14.83 2.07 49.81 14.69 43.75  
## 3 CLAY 10.76 7.4 14.26 1.86 49.37 14.05 50.72  
## 4 KARPOV 11.02 7.3 14.77 2.04 48.37 14.09 48.95  
## 5 BERNARD 11.02 7.23 14.25 1.92 48.93 14.99 40.87  
## 6 YURKOV 11.34 7.09 15.19 2.1 50.42 15.31 46.26  
## V9 V10 V11 V12 V13 V14  
## 1 Pole.vault Javeline 1500m Rank Points Competition  
## 2 5.02 63.19 291.7 1 8217 1  
## 3 4.92 60.15 301.5 2 8122 1  
## 4 4.92 50.31 300.2 3 8099 1  
## 5 5.32 62.77 280.1 4 8067 1  
## 6 4.72 63.44 276.4 5 8036 1

str(data)

## 'data.frame': 42 obs. of 14 variables:  
## $ V1 : Factor w/ 42 levels "Averyanov","Barras",..: 22 32 9 16 5 40 39 42 21 20 ...  
## $ V2 : Factor w/ 34 levels "10.44","10.5",..: 20 22 7 21 21 31 26 27 9 34 ...  
## $ V3 : Factor w/ 34 levels "6.61","6.68",..: 34 28 21 17 14 11 29 17 18 4 ...  
## $ V4 : Factor w/ 39 levels "12.68","13.07",..: 39 25 15 22 14 31 16 6 10 20 ...  
## $ V5 : Factor w/ 20 levels "1.85","1.86",..: 20 14 2 12 5 16 9 10 18 7 ...  
## $ V6 : Factor w/ 41 levels "400m","46.81",..: 1 29 21 5 12 33 9 8 30 31 ...  
## $ V7 : Factor w/ 37 levels "110m.hurdle",..: 1 19 4 5 28 34 9 7 14 24 ...  
## $ V8 : Factor w/ 40 levels "37.92","39.83",..: 40 19 38 35 8 30 10 28 20 31 ...  
## $ V9 : Factor w/ 19 levels "4.2","4.4","4.42",..: 19 14 12 12 17 8 12 3 3 12 ...  
## $ V10: Factor w/ 42 levels "50.31","50.62",..: 42 35 27 1 34 36 4 12 17 5 ...  
## $ V11: Factor w/ 39 levels "1500m","262.1",..: 1 35 38 37 28 20 23 9 31 2 ...  
## $ V12: Factor w/ 29 levels "1","10","11",..: 29 1 12 22 23 24 25 26 27 28 ...  
## $ V13: Factor w/ 40 levels "7313","7404",..: 40 30 29 27 24 23 22 18 17 10 ...  
## $ V14: Factor w/ 3 levels "1","2","Competition": 3 1 1 1 1 1 1 1 1 1 ...

La primera fila del data set contiene el nombre de cada una de las columnas. La estructura del data set muestra un error ya que cada columna se encuentra en formato de factor,

#Agregar header TRUE para tener nombre de columnas.  
  
data<-read.table(file="http://www.sthda.com/upload/decathlon.txt",header = T)   
head(data)

## name X100m Long.jump Shot.put High.jump X400m X110m.hurdle Discus  
## 1 SEBRLE 11.04 7.58 14.83 2.07 49.81 14.69 43.75  
## 2 CLAY 10.76 7.40 14.26 1.86 49.37 14.05 50.72  
## 3 KARPOV 11.02 7.30 14.77 2.04 48.37 14.09 48.95  
## 4 BERNARD 11.02 7.23 14.25 1.92 48.93 14.99 40.87  
## 5 YURKOV 11.34 7.09 15.19 2.10 50.42 15.31 46.26  
## 6 WARNERS 11.11 7.60 14.31 1.98 48.68 14.23 41.10  
## Pole.vault Javeline X1500m Rank Points Competition  
## 1 5.02 63.19 291.7 1 8217 1  
## 2 4.92 60.15 301.5 2 8122 1  
## 3 4.92 50.31 300.2 3 8099 1  
## 4 5.32 62.77 280.1 4 8067 1  
## 5 4.72 63.44 276.4 5 8036 1  
## 6 4.92 51.77 278.1 6 8030 1

str(data)

## 'data.frame': 41 obs. of 14 variables:  
## $ name : Factor w/ 41 levels "Averyanov","Barras",..: 31 9 16 5 39 38 41 21 20 13 ...  
## $ X100m : num 11 10.8 11 11 11.3 ...  
## $ Long.jump : num 7.58 7.4 7.3 7.23 7.09 7.6 7.3 7.31 6.81 7.56 ...  
## $ Shot.put : num 14.8 14.3 14.8 14.2 15.2 ...  
## $ High.jump : num 2.07 1.86 2.04 1.92 2.1 1.98 2.01 2.13 1.95 1.86 ...  
## $ X400m : num 49.8 49.4 48.4 48.9 50.4 ...  
## $ X110m.hurdle: num 14.7 14.1 14.1 15 15.3 ...  
## $ Discus : num 43.8 50.7 49 40.9 46.3 ...  
## $ Pole.vault : num 5.02 4.92 4.92 5.32 4.72 4.92 4.42 4.42 4.92 4.82 ...  
## $ Javeline : num 63.2 60.1 50.3 62.8 63.4 ...  
## $ X1500m : num 292 302 300 280 276 ...  
## $ Rank : int 1 2 3 4 5 6 7 8 9 10 ...  
## $ Points : int 8217 8122 8099 8067 8036 8030 8004 7995 7802 7733 ...  
## $ Competition : int 1 1 1 1 1 1 1 1 1 1 ...

Esto soluciona el error de estructura. Podemos ver que la variable name se exporta como factor y las restantes como número.

### Accesar a fila por nombre de competidor.

data<-read.table(file="http://www.sthda.com/upload/decathlon.txt",header = T, row.names = 1,sep="\t")   
  
data["CLAY",]

## X100m Long.jump Shot.put High.jump X400m X110m.hurdle Discus  
## CLAY 10.76 7.4 14.26 1.86 49.37 14.05 50.72  
## Pole.vault Javeline X1500m Rank Points Competition  
## CLAY 4.92 60.15 301.5 2 8122 1

### Accesar a resultado específico.

data["CLAY","Shot.put"]

## [1] 14.26

## Ejercicio:

El data set presenta nombres duplicados en mayúscula y minúscula. Crear un formato de nombres en el cual cada registro se encuentre en mayúsculas y no se presenten duplicados.

Nota: No pueden existir nombres de fila duplicados.