Tecnología Médica

Alicia Perdices Guerra

3 de mayo, 2021

Contents

1.PROCESAMIENTO DE LOS DATOS.

- En primer lugar leemos el fichero:

```
tec<-read.csv("C:/temp/TecnologiaMedica.csv",sep= ",")
```

- Realicemos una breve inspección de los datos

```
str(tec)
```

```
## 'data.frame':    15750 obs. of  7 variables:
##  $ TIME              : int  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
##  $ GEO               : Factor w/ 35 levels "Albania","Austria",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ UNIT              : Factor w/ 3 levels "Inhabitants per ...",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ FACILITY          : Factor w/ 5 levels "Angiography units",..: 2 2 2 5 5 5 3 3 3 1 ...
##  $ ICHA_HP           : Factor w/ 3 levels "Hospitals","Hospitals and providers of ambulatory health
##  $ Value             : Factor w/ 3061 levels ":","0.00","0.01",..: 740 695 2755 490 490 2 1 1623 1 1
##  $ Flag.and.Footnotes: Factor w/ 5 levels "","b","d","e",..: 3 3 1 1 1 1 1 1 1 1 ...
```

```
colnames(tec) #Nombre de las variables
```

```
## [1] "TIME"               "GEO"                "UNIT"
## [4] "FACILITY"           "ICHA_HP"            "Value"
## [7] "Flag.and.Footnotes"
```

```
nrow(tec) #Número de registros
```

```
## [1] 15750
```

```
ncol(tec) #Número de variables
```

```
## [1] 7
```

*Observamos las siguientes variables:

- **TIME**: variable cuantitativa. Indica el año en el que se ha realizado la medida, en este caso el valor de la variable "Value". Se ha cargado bien como número entero.
- **GEO**: variable cualitativa. Indica el país o región en el que se ha realizado la medida.Se ha cargado bien como factor.
- **UNIT**: variable cualitativa. Indica la medida de la variable valor.Se ha cargado bien como factor.
- **FACILITY**: variable cualitativa. Indica el tipo de tecnología médica.
- **ICHA__HP**: variable cualitativa. Indica donde se usa el recurso tecnológico.
- **Value**: Variable cuantitativa. Indica el número de recursos recursos tecnológicos médicos por países.Se ha cargado mal como factor.
- **Fal.and.footnotes**. Notas sobre etiquetas. Eliminamos esta columna.

*Años de las mediciones:

```
unique(tec$TIME)
```

```
##  [1] 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
```

*Paises:

```
unique(tec$GEO)
```

```
##  [1] Belgium
##  [2] Bulgaria
##  [3] Czechia
##  [4] Denmark
##  [5] Germany (until 1990 former territory of the FRG)
##  [6] Estonia
##  [7] Ireland
##  [8] Greece
##  [9] Spain
## [10] France
## [11] Croatia
## [12] Italy
## [13] Cyprus
## [14] Latvia
## [15] Lithuania
## [16] Luxembourg
## [17] Hungary
## [18] Malta
## [19] Netherlands
## [20] Austria
## [21] Poland
## [22] Portugal
## [23] Romania
## [24] Slovenia
## [25] Slovakia
## [26] Finland
## [27] Sweden
## [28] Iceland
## [29] Liechtenstein
## [30] Switzerland
## [31] United Kingdom
## [32] North Macedonia
## [33] Albania
## [34] Serbia
## [35] Turkey
## 35 Levels: Albania Austria Belgium Bulgaria Croatia Cyprus Czechia ... United Kingdom
```

*Unidad de las mediciones:

```
unique(tec$UNIT)
```

```
## [1] Number                          Inhabitants per ...
## [3] Per hundred thousand inhabitants
## Levels: Inhabitants per ... Number Per hundred thousand inhabitants
```

- Tipo de recursos tecnológicos:

```
unique(tec$FACILITY)
```

```
## [1] Computed Tomography Scanners     Magnetic Resonance Imaging Units
## [3] Gamma cameras                    Angiography units
## [5] Lithotriptors
## 5 Levels: Angiography units Computed Tomography Scanners ... Magnetic Resonance Imaging Units
```

*Lugar de uso de los recursos tecnológicos.

```
unique(tec$ICHA_HP)
```

```
## [1] Hospitals and providers of ambulatory health care
## [2] Hospitals
## [3] Providers of ambulatory health care
## 3 Levels: Hospitals ... Providers of ambulatory health care
```

- Eliminamos la columna Fal.and.footnotes.

```
tec<-tec[,-7]
```

- Tendríamos que convertir la columna Value a numérico porque se ha cargado como factor y es erróneo. El resto de variables tienen el tipo correcto.

```
tec$Value<-as.character(tec$Value)
tec$Value<-(gsub(',','.',tec$Value) )
tec$Value<-(gsub(' ','',tec$Value) )
tec$Value<-as.numeric(tec$Value)
```

```
## Warning: NAs introducidos por coerción
```

- Comprobamos que valores tenemos en la columna **Value**:

```
tail(table(tec$Value, useNA = "ifany"))
```

```
##
## 23371348.25    46444832    46480882  46620044.5  46773054.5        <NA>
##           1           1           1           1           1        7005
```

- Observamos que tenemos **7005 valores perdidos**.Guardamos en la variable **idx** los índices de los registros con valores **NA** de la variable **Value**.

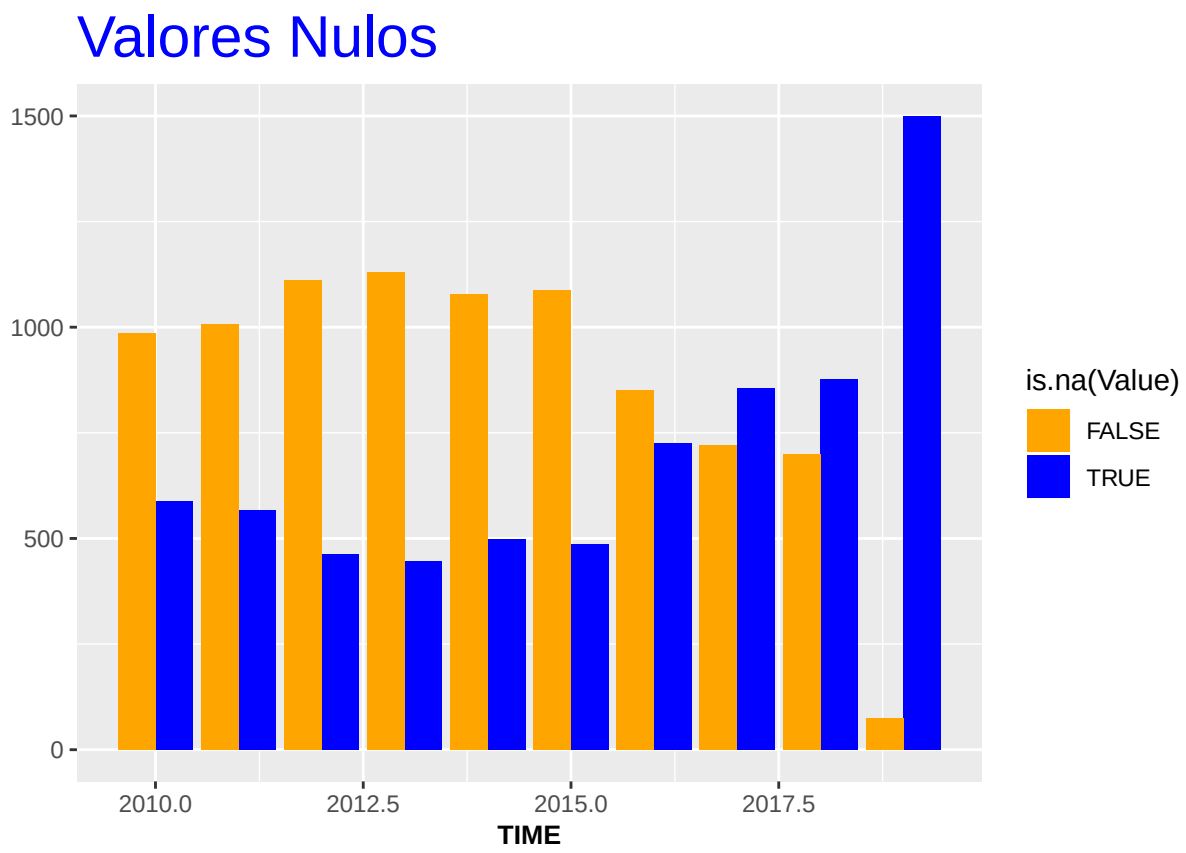
```
idx<-which(is.na(tec$Value))
length(idx)
```

```
## [1] 7005
```

- Grafiquemos la información que contiene la variable **Value**

```
library(ggplot2)
library(scales)
g = ggplot(tec, aes(TIME, fill=is.na(Value)) ) +
labs(title = "Valores Nulos")+ylab("") +
theme(plot.title = element_text(size = rel(2), colour = "blue"))

g+geom_bar(position="dodge") + scale_fill_manual(values = alpha(c("orange", "blue"), 1)) +
theme(axis.title.x = element_text(face="bold", size=10))
```



- En caso de detectar algún valor anómalo (en nuestro caso los NAS) en las variables tendríamos que realizar una imputación de esos valores o bien sustituyéndolos por la media o usando el algoritmo KNN (k-Nearest Neighbour) con los 3 vecinos más cercanos usando la distancia que consideremos, en este caso usaremos Gower(Mediana), por ser una medida más robusa frente a extremos.

```
library(VIM)
```

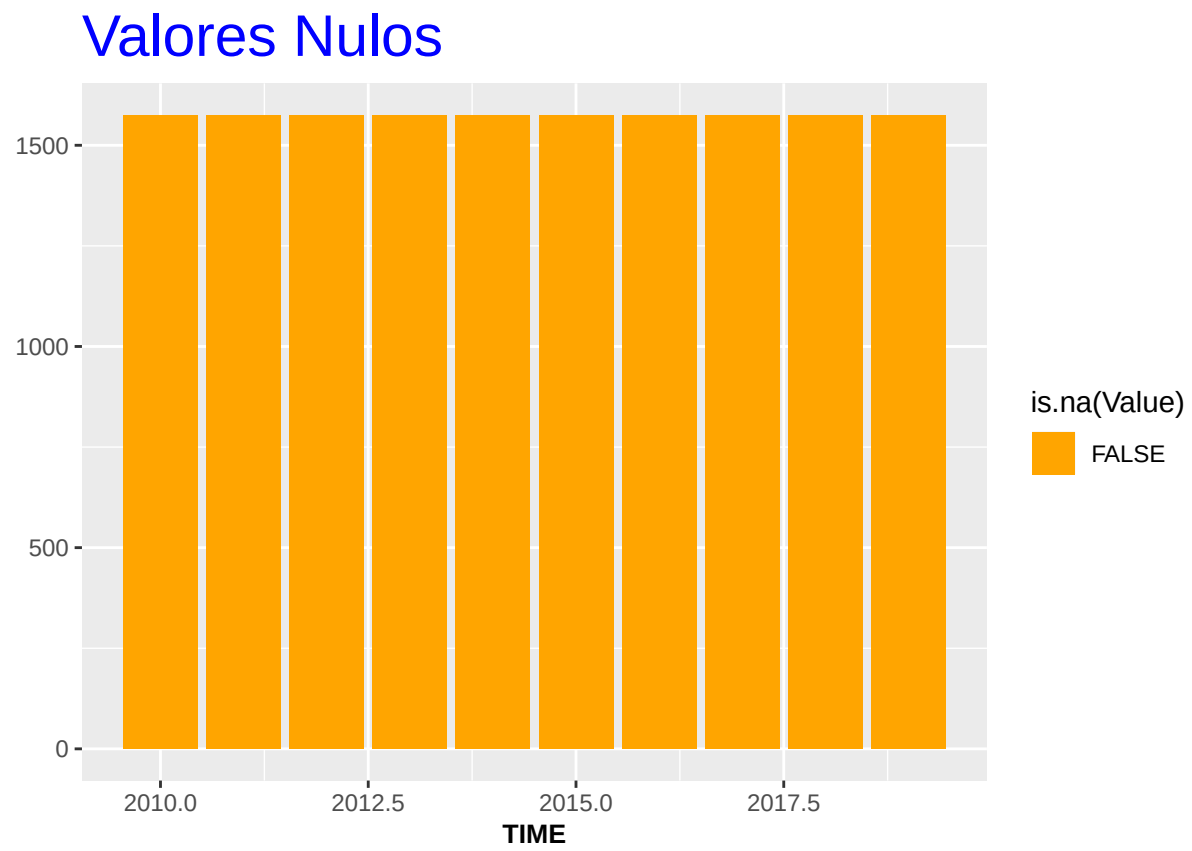
```
## Loading required package: colorspace
## Loading required package: grid
## VIM is ready to use.
## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues
##
## Attaching package: 'VIM'
## The following object is masked from 'package:datasets':
##
##     sleep
```

```
output<-kNN(tec, variable=c("Value"),k=3)
tec<-output
```

- Comprobamos que no tenemos valores nulos después de la imputación

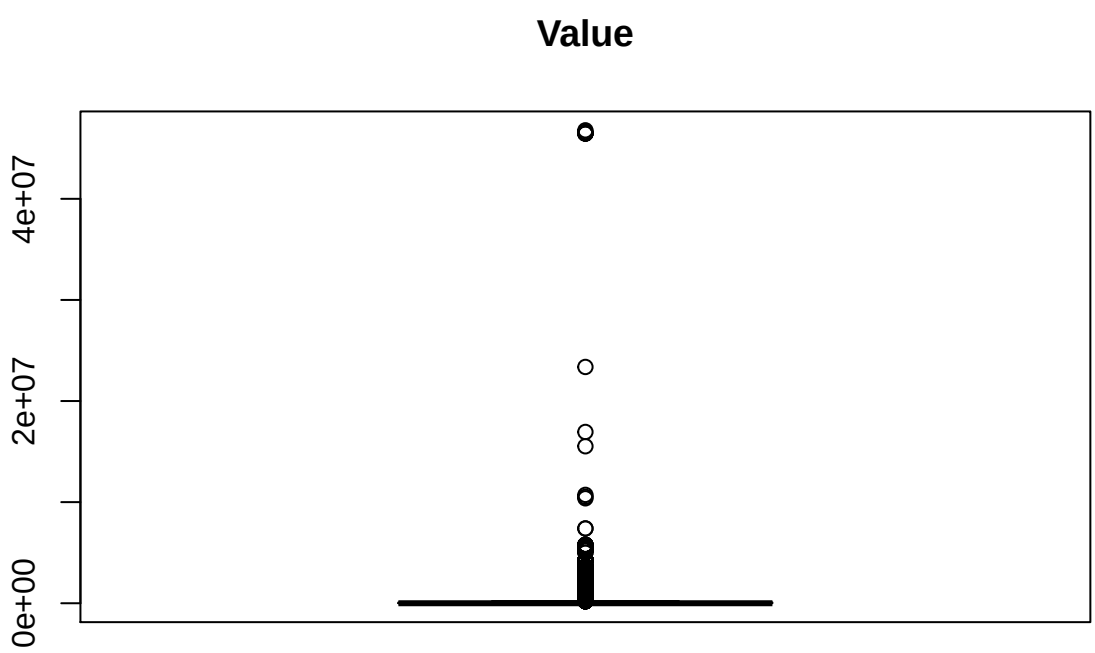
```
g = ggplot(tec, aes(TIME, fill=is.na(Value)) ) +
labs(title = "Valores Nulos")+ylab("") +
theme(plot.title = element_text(size = rel(2), colour = "blue"))
```

```
g+geom_bar(position="dodge") + scale_fill_manual(values = alpha(c("orange", "blue"), 1)) +
theme(axis.title.x = element_text(face="bold", size=10))
```



- Con el siguiente gráfico, observaremos que la variable **Value** tiene outliers o valores extremos

```
boxplot(tec$Value, main="Value")
```



- Por otro lado, revisamos para el resto de columnas si tenemos valores NA.(desconocidos o perdidos)

```
table(tec$TIME, useNA = "ifany")
```

```
##
## 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
## 1575 1575 1575 1575 1575 1575 1575 1575 1575 1575
```

```
table(tec$GEO, useNA = "ifany")
```

```
##
##                                          Albania
##                                              450
##                                          Austria
##                                              450
##                                          Belgium
##                                              450
##                                         Bulgaria
##                                              450
##                                          Croatia
##                                              450
##                                           Cyprus
##                                              450
##                                          Czechia
##                                              450
##                                          Denmark
##                                              450
##                                          Estonia
```

```
##                                             450
##                                          Finland
##                                             450
##                                           France
##                                             450
## Germany (until 1990 former territory of the FRG)
##                                             450
##                                           Greece
##                                             450
##                                          Hungary
##                                             450
##                                          Iceland
##                                             450
##                                          Ireland
##                                             450
##                                            Italy
##                                             450
##                                           Latvia
##                                             450
##                                    Liechtenstein
##                                             450
##                                        Lithuania
##                                             450
##                                       Luxembourg
##                                             450
##                                            Malta
##                                             450
##                                      Netherlands
##                                             450
##                                   North Macedonia
##                                             450
##                                           Poland
##                                             450
##                                         Portugal
##                                             450
##                                          Romania
##                                             450
##                                           Serbia
##                                             450
##                                         Slovakia
##                                             450
##                                         Slovenia
##                                             450
##                                            Spain
##                                             450
##                                           Sweden
##                                             450
##                                      Switzerland
##                                             450
##                                           Turkey
##                                             450
##                                   United Kingdom
##                                             450
```

```
table(tec$UNIT, useNA = "ifany")
```

```
##
##              Inhabitants per ...                             Number
##                             5250                               5250
## Per hundred thousand inhabitants
##                             5250
```

```
table(tec$FACILITY, useNA = "ifany")
```

```
##
##                Angiography units     Computed Tomography Scanners
##                             3150                               3150
##                    Gamma cameras                     Lithotriptors
##                             3150                               3150
## Magnetic Resonance Imaging Units
##                             3150
```

```
table(tec$ICHA_HP, useNA = "ifany")
```

```
##
##                                         Hospitals
##                                              5250
## Hospitals and providers of ambulatory health care
##                                              5250
##               Providers of ambulatory health care
##                                              5250
```

Observamos que no existen ahora valores perdidos después de la imputación.La suma de las cantidades de cada variable, suman el total.

La estructura de los datos quedaría de la siguiente forma:

```
str(tec)
```

```
## 'data.frame':    15750 obs. of  7 variables:
##  $ TIME     : int  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
##  $ GEO      : Factor w/ 35 levels "Albania","Austria",..: 3 3 3 3 3 3 3 3 3 3 ...
##  $ UNIT     : Factor w/ 3 levels "Inhabitants per ...",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ FACILITY : Factor w/ 5 levels "Angiography units",..: 2 2 2 5 5 5 3 3 3 1 ...
##  $ ICHA_HP  : Factor w/ 3 levels "Hospitals","Hospitals and providers of ambulatory health care",..:
##  $ Value    : num  152 144 8 116 116 0 152 322 8 133 ...
##  $ Value_imp: logi  FALSE FALSE FALSE FALSE FALSE FALSE ...
```

- Finalmente, creamos un fichero con toda la información corregida.

```
write.csv(tec, file="TecnologiaMedica_clean.csv", row.names = FALSE)
```