Ocupacion de Cama Hospitalaria

Alicia Perdices Guerra

3 de mayo, 2021

Contents

1.PROCESAMIENTO DE LOS DATOS.

- En primer lugar leemos el fichero:

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

- Realicemos una breve inspección de los datos

```
str(ocupacion_ch)
```

```
## 'data.frame':    310 obs. of  6 variables:
##  $ TIME              : int  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
##  $ GEO               : Factor w/ 31 levels "Austria","Belgium",..: 2 5 6 10 7 13 11 27 9 3 ...
##  $ ICHA_HC           : Factor w/ 1 level "Services of curative care": 1 1 1 1 1 1 1 1 1 ...
##  $ UNIT              : Factor w/ 1 level "Percentage": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Value             : Factor w/ 213 levels ":","45.60","47.50",..: 173 88 1 155 75 209 71 126 121 1
##  $ Flag.and.Footnotes: Factor w/ 5 levels "","b","bd","d",..: 1 1 1 1 1 1 4 1 1 1 ...
```

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

```
## [1] "TIME"               "GEO"                "ICHA_HC"
## [4] "UNIT"               "Value"              "Flag.and.Footnotes"
```

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

```
## [1] 310
```

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

```
## [1] 6
```

*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.Porcentaje.
- **ICHA__HC**: Variable cualitativa. Hace referencia a los Servicios de cuidados curativos
- **Value**: Variable cuantitativa. Indica el porcentaje de ocupación de camas hospitalarias por paises.
- **Fal.and.footnotes**. Notas sobre etiquetas. Eliminamos esta columna.

*Años de las mediciones:

```
unique(ocupacion_ch$TIME)
```

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

*Paises:

```
unique(ocupacion_ch$GEO)
```

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

*Unidad de las mediciones:

```
unique(ocupacion_ch$UNIT)
```

```
## [1] Percentage
## Levels: Percentage
```

- En relación a los Servicios de Cuidado Curativo

```
unique(ocupacion_ch$ICHA_HC)
```

```
## [1] Services of curative care
## Levels: Services of curative care
```

- Eliminamos la columna Fal.and.footnotes ya que no nos aporta información relevante.

```
ocupacion_ch<-ocupacion_ch[,-6]
```

- Tendríamos que resolver las posibles inconsistencias en relación al formato del valor numérico de la variable **Value** y convertirla a valor numérico.

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

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

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

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

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

```
##
## 91.4 91.9 92.6 93.3 93.8 <NA>
##    1    1    1    1    1   81
```

- Observamos que tenemos **81 valores perdidos**.Guardamos en la variable **idx** los índices de los registros con valores **NA** de la variable **Value**.

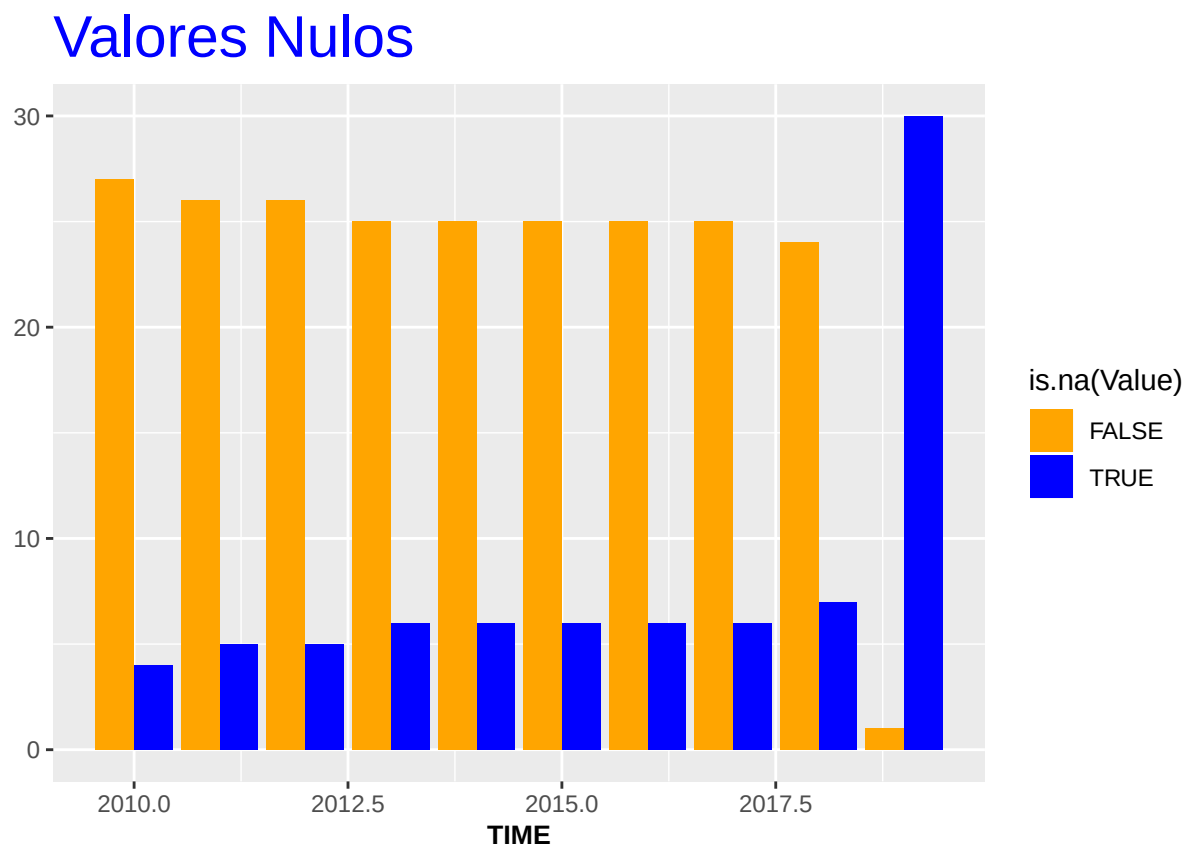
```
idx<-which(is.na(ocupacion_ch$Value))
length(idx)
```

```
## [1] 81
```

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

```
library(ggplot2)
library(scales)
g = ggplot(ocupacion_ch, 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)
```

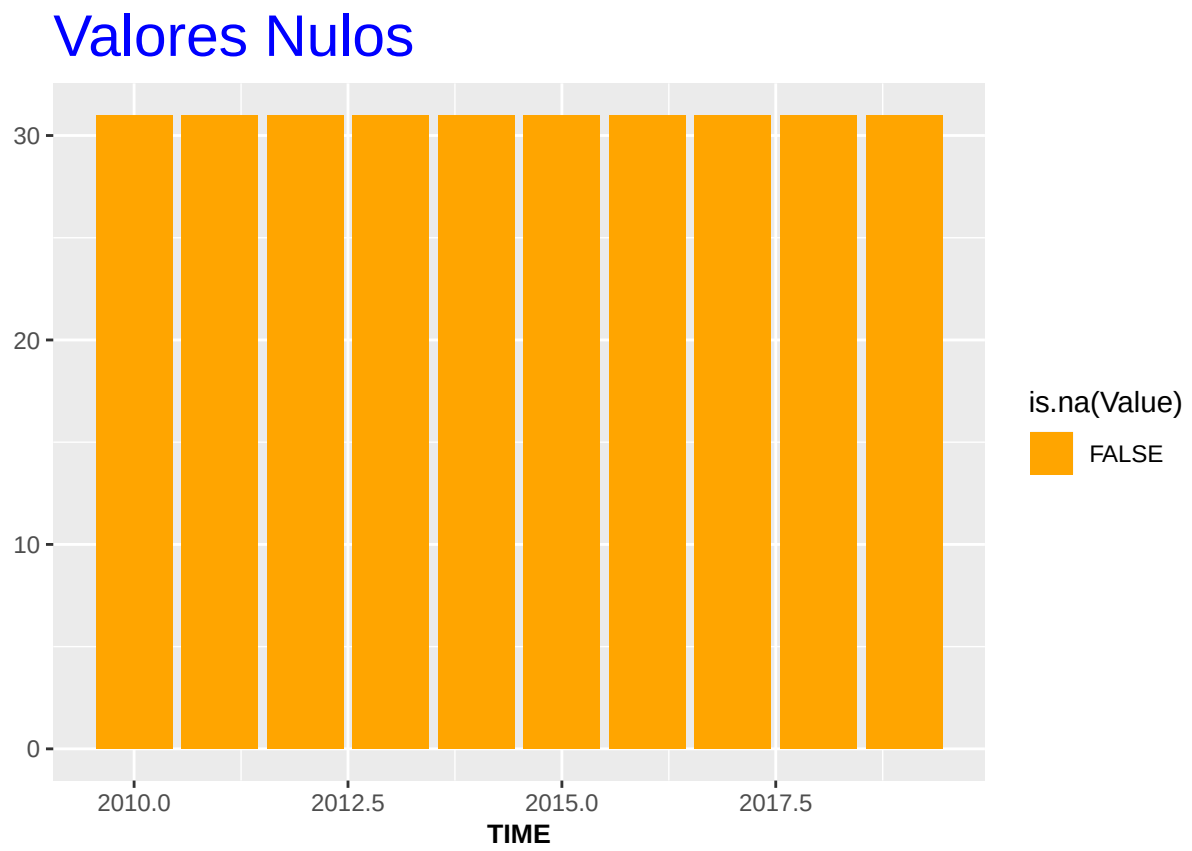
```
## Loading required package: colorspace
## Loading required package: grid
## VIM is ready to use.
## Suggestions and bug-reports can be submitted at: https://github.com/statistikat/VIM/issues
##
## Attaching package: 'VIM'
## The following object is masked from 'package:datasets':
##
##     sleep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

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

- Comprobamos que no tenemos valores nulos después de la imputación

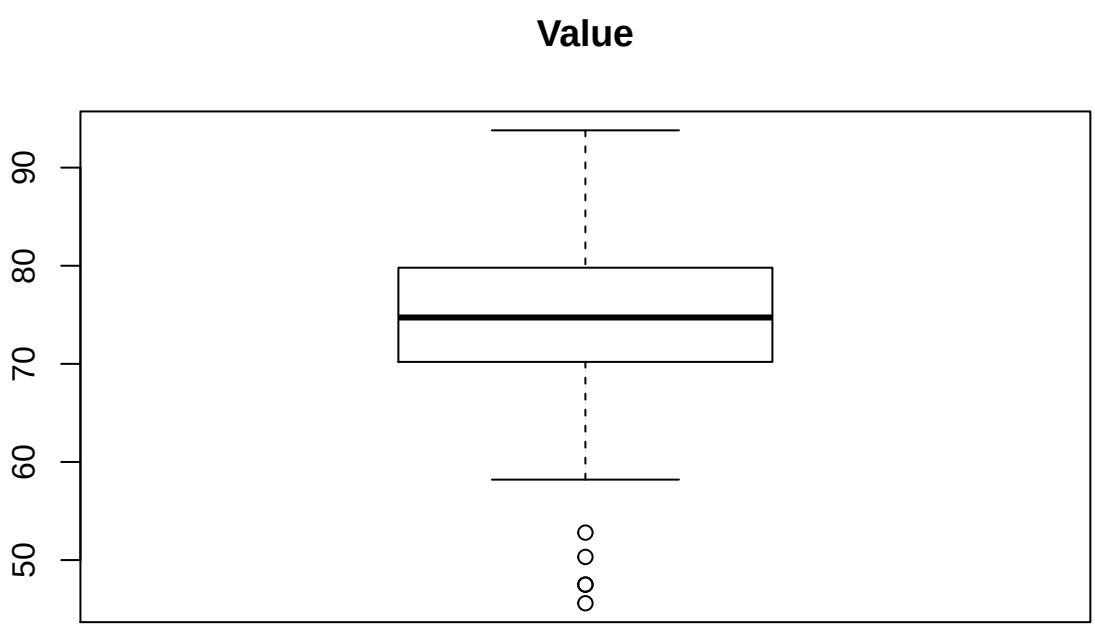
```
g = ggplot(ocupacion_ch, aes(TIME, fill=is.na(Value)) ) +
labs(title = "Valores Nulos")+ylab("") +
theme(plot.title = element_text(size = rel(2), colour = "blue"))
```

```
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(ocupacion_ch$Value, main="Value")
```



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

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

```
##
## 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
##   31   31   31   31   31   31   31   31   31   31
```

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

```
##
##                                    Austria
##                                         10
##                                    Belgium
##                                         10
##                                    Croatia
##                                         10
##                                     Cyprus
##                                         10
##                                    Czechia
##                                         10
##                                    Denmark
##                                         10
##                                    Estonia
##                                         10
##                                    Finland
##                                         10
##                                     France
```

```
##                                                 10
## Germany (until 1990 former territory of the FRG)
##                                                 10
##                                             Greece
##                                                 10
##                                            Hungary
##                                                 10
##                                            Ireland
##                                                 10
##                                              Italy
##                                                 10
##                                             Latvia
##                                                 10
##                                       Liechtenstein
##                                                 10
##                                          Lithuania
##                                                 10
##                                         Luxembourg
##                                                 10
##                                              Malta
##                                                 10
##                                         Montenegro
##                                                 10
##                                        Netherlands
##                                                 10
##                                             Norway
##                                                 10
##                                           Portugal
##                                                 10
##                                             Serbia
##                                                 10
##                                           Slovakia
##                                                 10
##                                           Slovenia
##                                                 10
##                                              Spain
##                                                 10
##                                             Sweden
##                                                 10
##                                        Switzerland
##                                                 10
##                                             Turkey
##                                                 10
##                                     United Kingdom
##                                                 10
```

```r
table(ocupacion_ch$UNIT, useNA = "ifany")
```

```
##
## Percentage
##        310
```

```r
table(ocupacion_ch$ICHA_HC, useNA = "ifany")
```

```
##
```

```
## Services of curative care
##                       310
```

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:

```
str(ocupacion_ch)
```

```
## 'data.frame':    310 obs. of  6 variables:
##  $ TIME     : int  2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
##  $ GEO      : Factor w/ 31 levels "Austria","Belgium",..: 2 5 6 10 7 13 11 27 9 3 ...
##  $ ICHA_HC  : Factor w/ 1 level "Services of curative care": 1 1 1 1 1 1 1 1 1 1 ...
##  $ UNIT     : Factor w/ 1 level "Percentage": 1 1 1 1 1 1 1 1 1 1 ...
##  $ Value    : num  80.7 71.3 79 79 70.8 ...
##  $ Value_imp: logi  FALSE FALSE TRUE FALSE FALSE FALSE ...
```

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

```
write.csv(ocupacion_ch, file="OcupacionCamaHospitalaria_clean.csv", row.names = FALSE)
```