

# Recursos Técnicos Hospitalarios

Alicia Perdices Guerra

3 de mayo, 2021

## Contents

### 1.PROCESAMIENTO DE LOS DATOS.

- En primer lugar leemos el fichero:

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

- Realicemos una breve inspección de los datos

```
str(recursos)
```

```
## 'data.frame': 1860 obs. of 6 variables:
## $ TIME : int 2010 2010 2010 2010 2010 2010 2010 2010 2010 2010 ...
## $ GEO : Factor w/ 31 levels "Austria","Belgium",...: 2 2 2 2 2 2 3 3 3 3 ...
## $ FACILITY : Factor w/ 6 levels "Day care places altogether",...: 4 1 6 3 5 2 4 1 6 3 ...
## $ UNIT : Factor w/ 1 level "Number": 1 1 1 1 1 1 1 1 1 1 ...
## $ Value : Factor w/ 616 levels ":", "0.00", "1 014.00",...: 29 451 184 46 197 475 1 1 1 1
## $ Flag.and.Footnotes: Factor w/ 5 levels "","b","d","e",...: 1 1 1 1 1 1 1 1 1 1 ...
```

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

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

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

```
## [1] 1860
```

```
ncol(recursos) #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.
- **FACILITY**: variable cualitativa. Indica el tipo de recurso hospitalario.
- **Value**: Variable cuantitativa. Indica el número de recursos hospitalarios por países. Se ha cargado mal como factor.
- **Fal.and.footnotes**. Notas sobre etiquetas. Eliminamos esta columna.

\*Años de las mediciones:

```
unique(recursos$TIME)
```

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

\*Países:

```
unique(recursos$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] Liechtenstein
## [27] Switzerland
## [28] United Kingdom
## [29] North Macedonia
## [30] Serbia
## [31] Turkey
## 31 Levels: Austria Belgium Bulgaria Croatia Cyprus Czechia Denmark ... United Kingdom
```

\*Unidad de las mediciones:

```
unique(recursos$UNIT)
```

```
## [1] Number
## Levels: Number
```

- Tipo de recursos:

```
unique(recursos$FACILITY)
```

```
## [1] Operation theatres in hospital Day care places altogether
## [3] Surgical day care places Oncological day care place
## [5] Psychiatric day care place Geriatric day care places
## 6 Levels: Day care places altogether ... Surgical day care places
```

- Eliminamos la columna Fal.and.footnotes.

```
recursos<-recursos[,-6]
```

- 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.

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

## Warning: NAs introducidos por coerción

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

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

```
##
## 72536 72789 74646 75452 77297 <NA>
##      1      1      1      1      1 1030
```

- Observamos que tenemos **1030 valores perdidos**. Guardamos en la variable **idx** los índices de los registros con valores **NA** de la variable **Value**.

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

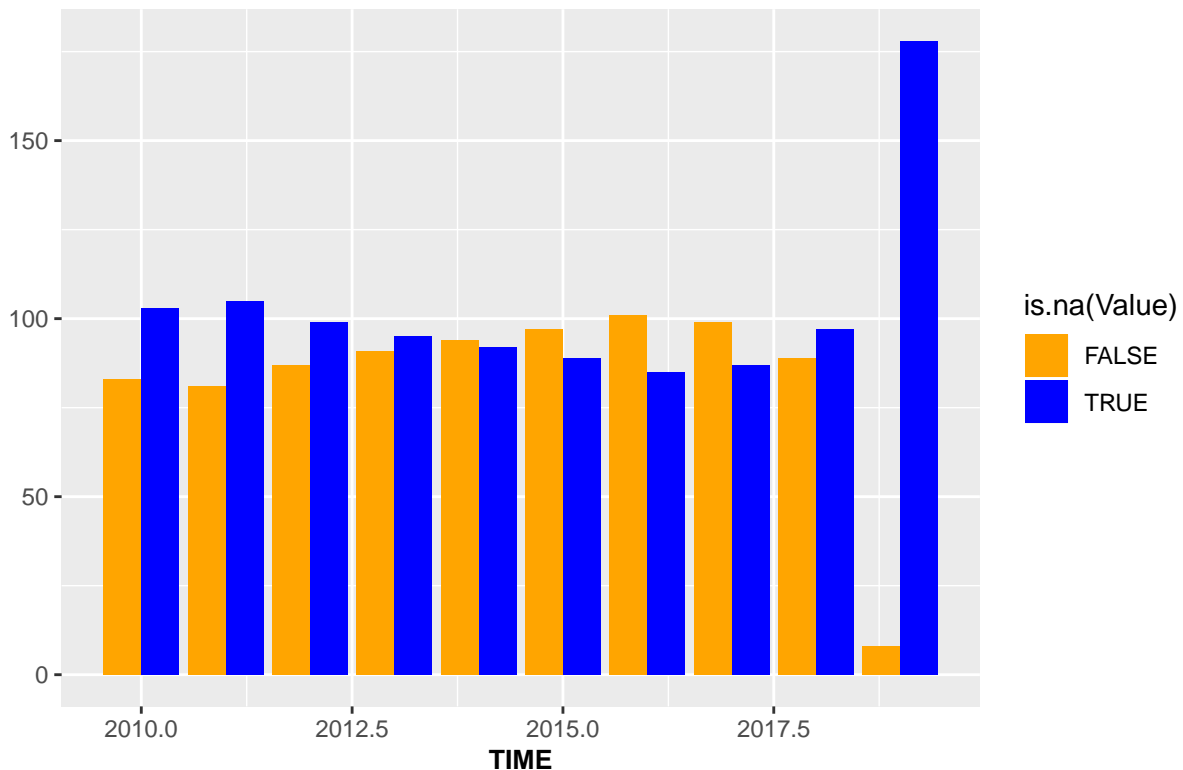
```
## [1] 1030
```

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

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

## Valores Nulos



- 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 robusta 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(recursos, variable=c("Value"),k=3)
```

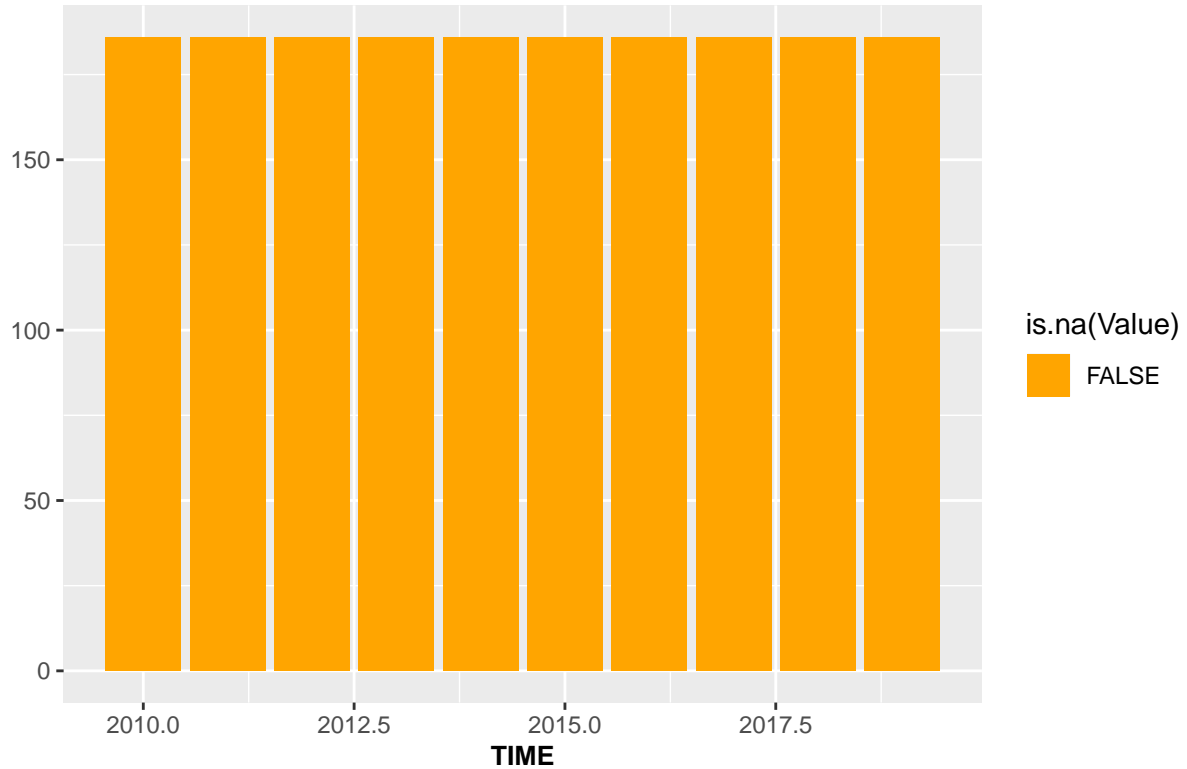
```
recursos<-output
```

- Comprobamos que no tenemos valores nulos después de la imputación

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

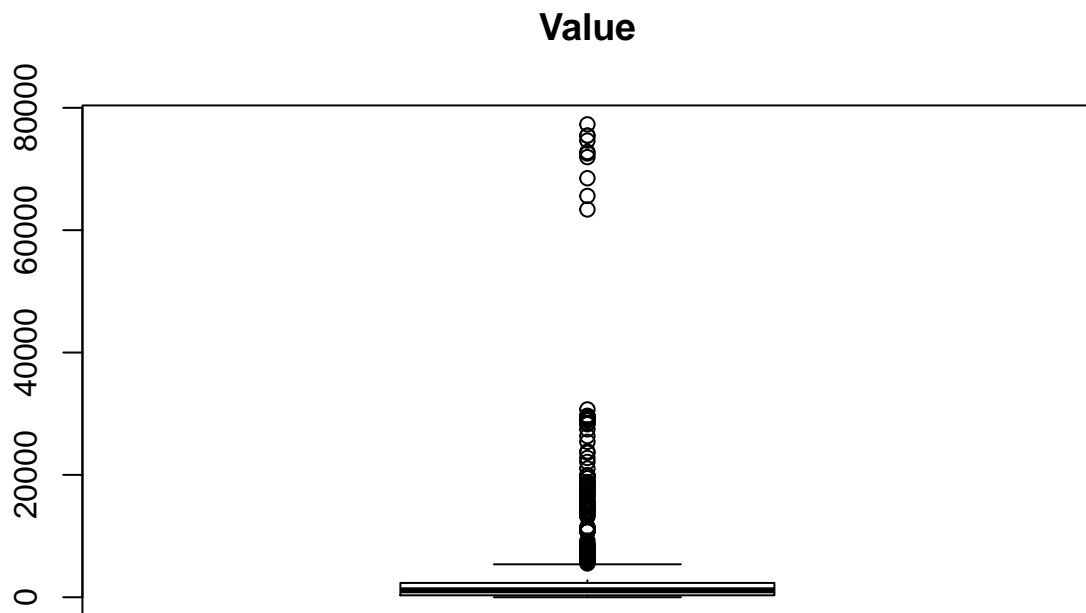
```
g+geom_bar(position="dodge") + scale_fill_manual(values = alpha(c("orange", "blue"), 1)) +  
theme(axis.title.x = element_text(face="bold", size=10))
```

## Valores Nulos



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

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



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

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

```
##
## 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
## 186 186 186 186 186 186 186 186 186 186
```

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

```
##
## Austria
## 60
## Belgium
## 60
## Bulgaria
## 60
## Croatia
## 60
## Cyprus
## 60
## Czechia
## 60
## Denmark
## 60
## Estonia
## 60
## France
```

```

##                                     60
## Germany (until 1990 former territory of the FRG)
##                                     60
##                                     Greece
##                                     60
##                                     Hungary
##                                     60
##                                     Ireland
##                                     60
##                                     Italy
##                                     60
##                                     Latvia
##                                     60
##                                     Liechtenstein
##                                     60
##                                     Lithuania
##                                     60
##                                     Luxembourg
##                                     60
##                                     Malta
##                                     60
##                                     Netherlands
##                                     60
##                                     North Macedonia
##                                     60
##                                     Poland
##                                     60
##                                     Portugal
##                                     60
##                                     Romania
##                                     60
##                                     Serbia
##                                     60
##                                     Slovakia
##                                     60
##                                     Slovenia
##                                     60
##                                     Spain
##                                     60
##                                     Switzerland
##                                     60
##                                     Turkey
##                                     60
##                                     United Kingdom
##                                     60

```

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

```

##
## Number
## 1860

```

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

```
##
```

```
##      Day care places altogether      Geriatric day care places
##                                310                                310
##      Oncological day care place Operation theatres in hospital
##                                310                                310
##      Psychiatric day care place      Surgical day care places
##                                310                                310
```

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

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

```
write.csv(recursos, file="RecursosTecnicos_hospitalarios_clean.csv", row.names = FALSE)
```