

Proyect

Cristopher Barrios, Elian, Angel Higueros, Mariana David

16/2/2023

librerias

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

## Package 'mclust' version 6.0.0
## Type 'citation("mclust")' for citing this R package in publications.

## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa

## Warning: package 'GGally' was built under R version 4.2.3

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg      ggplot2

## Warning: package 'FeatureImpCluster' was built under R version 4.2.3

## Loading required package: data.table

##
## Attaching package: 'data.table'

## The following objects are masked from 'package:dplyr':
##
##   between, first, last

## Warning: package 'pheatmap' was built under R version 4.2.3
```

datos

```

M2009 <- read_sav("Matrimonios/Matrimonio2009.sav") # nolint
M2010 <- read_sav("Matrimonios/Matrimonio2010.sav") # nolint
M2011 <- read_sav("Matrimonios/Matrimonio2011.sav") # nolint
M2012 <- read_sav("Matrimonios/Matrimonio2012.sav") # nolint
M2013 <- read_sav("Matrimonios/Matrimonio2013.sav") # nolint
M2014 <- read_sav("Matrimonios/Matrimonio2014.sav") # nolint
M2015 <- read_sav("Matrimonios/Matrimonio2015.sav") # nolint
M2016 <- read_sav("Matrimonios/Matrimonio2016.sav") # nolint
M2017 <- read_sav("Matrimonios/Matrimonio2017.sav") # nolint
M2018 <- read_sav("Matrimonios/Matrimonio2018.sav") # nolint
M2019 <- read_sav("Matrimonios/Matrimonio2019.sav") # nolint
M2020 <- read_sav("Matrimonios/Matrimonio2020.sav") # nolint
M2021 <- read_sav("Matrimonios/Matrimonio2021.sav") # nolint

```

Nacimiento

```

N2009 <- read_sav("Nacimientos/Nacimiento2009.sav") # nolint
N2010 <- read_sav("Nacimientos/Nacimiento2010.sav") # nolint
N2011 <- read_sav("Nacimientos/Nacimiento2011.sav") # nolint
N2012 <- read_sav("Nacimientos/Nacimiento2012.sav") # nolint
N2013 <- read_sav("Nacimientos/Nacimiento2013.sav") # nolint
N2014 <- read_sav("Nacimientos/Nacimiento2014.sav") # nolint
N2015 <- read_sav("Nacimientos/Nacimiento2015.sav") # nolint
N2016 <- read_sav("Nacimientos/Nacimiento2016.sav") # nolint
N2017 <- read_sav("Nacimientos/Nacimiento2017.sav") # nolint
N2018 <- read_sav("Nacimientos/Nacimiento2018.sav") # nolint
N2019 <- read_sav("Nacimientos/Nacimiento2019.sav") # nolint
N2020 <- read_sav("Nacimientos/Nacimiento2020.sav") # nolint
N2021 <- read_sav("Nacimientos/Nacimiento2021.sav") # nolint

```

Divorcios

```

D2009 <- read_sav("Divorcios/Divorcio2009.sav") # nolint
D2010 <- read_sav("Divorcios/Divorcio2010.sav") # nolint
D2011 <- read_sav("Divorcios/Divorcio2011.sav") # nolint
D2012 <- read_sav("Divorcios/Divorcio2012.sav") # nolint
D2013 <- read_sav("Divorcios/Divorcio2013.sav") # nolint
D2014 <- read_sav("Divorcios/Divorcio2014.sav") # nolint
D2015 <- read_sav("Divorcios/Divorcio2015.sav") # nolint
D2016 <- read_sav("Divorcios/Divorcio2016.sav") # nolint
D2017 <- read_sav("Divorcios/Divorcio2017.sav") # nolint
D2018 <- read_sav("Divorcios/Divorcio2018.sav") # nolint
D2019 <- read_sav("Divorcios/Divorcio2019.sav") # nolint
D2020 <- read_sav("Divorcios/Divorcio2020.sav") # nolint
D2021 <- read_sav("Divorcios/Divorcio2021.sav") # nolint

```

Descripción variables y observaciones

Comience describiendo cuantas variables y observaciones tiene disponibles, el tipo de cada una de las variables.

```
vars1_numéricas <- sapply(D2009, is.numeric) #Análisis de divorcios 2009 # nolint
vars2_numéricas <- sapply(D2010, is.numeric) #Análisis de divorcios 2010 # nolint
print(vars1_numéricas)
```

```
## Depreg Mupreg Mesreg Anoreg Diaocu Mesocu Anoocu Depocu Mupocu Edadhom
## TRUE FALSE TRUE FALSE TRUE TRUE FALSE TRUE FALSE TRUE
## Edadmuj Gethom Getmuj Nachom Nacmuj Ocuhom Ocumuj Mever Anover
## TRUE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
```

```
print(vars2_numéricas)
```

```
## depreg mupreg mesreg añoreg diaocu mesocu añoocu depocu
## TRUE FALSE TRUE TRUE TRUE TRUE TRUE TRUE
## mupocu edadhom edadmuj grethom gretmuj nachom nacmuj escohom
## FALSE TRUE TRUE TRUE TRUE TRUE TRUE TRUE
## escomuj ocupahom ocupamuj
## TRUE TRUE TRUE
```

Resumen de datos para verificar si son numéricas

Haga un resumen de las variables numéricas e investigue si siguen una distribución normal y tablas de frecuencia para las variables categóricas, escriba lo que vaya encontrando.

```
summary(M2009)
```

```
##          DEPREG          MUPREG          MESREG          AÑOREG
## Min.   : 1.000   Length:62104   Min.   : 1.000   Min.   : 9.000
## 1st Qu.: 4.000   Class :character   1st Qu.: 3.000   1st Qu.: 9.000
## Median :10.000   Mode  :character   Median : 6.000   Median : 9.000
## Mean   : 9.728                      Mean   : 6.259   Mean   : 9.196
## 3rd Qu.:14.000                      3rd Qu.: 9.000   3rd Qu.: 9.000
## Max.   :22.000                      Max.   :12.000   Max.   :10.000
##          CLAUNI          EDADHOM          EDADMUJ          GETHOM          GETMUJ
## Min.   :1   Min.   : 16.00   Min.   : 14.00   Min.   :1.000   Min.   :1.000
## 1st Qu.:1   1st Qu.: 21.00   1st Qu.: 19.00   1st Qu.:1.000   1st Qu.:1.000
## Median :1   Median : 25.00   Median : 22.00   Median :2.000   Median :2.000
## Mean   :1   Mean   : 63.16   Mean   : 65.16   Mean   :3.881   Mean   :3.941
## 3rd Qu.:1   3rd Qu.: 32.00   3rd Qu.: 29.00   3rd Qu.:9.000   3rd Qu.:9.000
## Max.   :1   Max.   :999.00   Max.   :999.00   Max.   :9.000   Max.   :9.000
##          NACHOM          NACMUJ          OCUHOM          OCUMUJ          NUPHON
## Min.   : 30.0   Min.   : 30   Min.   : 110   Min.   : 110   Min.   : 1.00
## 1st Qu.:320.0   1st Qu.:320   1st Qu.:6111   1st Qu.:9711   1st Qu.: 1.00
## Median :320.0   Median :320   Median :7112   Median :9711   Median :99.00
## Mean   :317.6   Mean   :317   Mean   :7447   Mean   :9047   Mean   :54.58
## 3rd Qu.:320.0   3rd Qu.:320   3rd Qu.:9712   3rd Qu.:9712   3rd Qu.:99.00
## Max.   :320.0   Max.   :320   Max.   :9999   Max.   :9999   Max.   :99.00
##          NUPMUJ          DEPOCU          MUPOCU          MESOCU
## Min.   : 1.00   Min.   : 1.000   Length:62104   Min.   : 1.000
## 1st Qu.: 1.00   1st Qu.: 4.000   Class :character   1st Qu.: 3.000
## Median :99.00   Median :10.000   Mode  :character   Median : 6.000
## Mean   :54.65   Mean   : 9.724                      Mean   : 6.553
```

```
## 3rd Qu.:99.00 3rd Qu.:14.000 3rd Qu.:10.000
## Max. :99.00 Max. :22.000 Max. :12.000
## AÑOOCU AREAG
## Min. :9 Min. :9
## 1st Qu.:9 1st Qu.:9
## Median :9 Median :9
## Mean :9 Mean :9
## 3rd Qu.:9 3rd Qu.:9
## Max. :9 Max. :9
```

Resumen de datos para verificar si son numéricas

Variables importantes

Cruce las variables que considere que son las más importantes para hallar los elementos clave que lo pueden llevar a comprender lo que está causando el problema encontrado.

Graficos exploratorios

Haga gráficos exploratorios que le de ideas del estado de los datos.

Clustering

Haga un agrupamiento (clustering) e interprete los resultados.

```
#D2010 <- na.omit(D2010)

#D2009_norm <- scale(D2009)

#cluster9 <- hclust(dist(D2009_norm), method="ward.D2")
```