

Proyecto

Cristopher Barrios, Elean Rivas, Angel Higueros, Mariana David
16/2/2023

librerias

```
library(haven)
library(stats)
library(dplyr)
library(ggplot2)
library(cluster) #Para calcular la silueta
library(e1071)#para cmeans
library(mclust) #mixtures of gaussians
library(fpc) #Para hacer el plotcluster
library(NbClust) #Para determinar el número de clusters óptimo
library(factoextra) #Para hacer gráficos bonitos de clustering
library(hopkins) #Para revisar si vale la pena hacer agrupamiento
library(GGally) #Para hacer el conjunto de graficos
library(FeatureImpCluster) #Para revisar la importancia de las variables en los grupos. # nolint
library(pheatmap) #Para hacer mapa de calor
```

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
```

Nacimientos

```
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.

Las bases de datos de matrimonios cuentan con diferentes cantidades de variables, pero las 22 más comunes son: —Para divorcios— DEPREG: cualitativa MUPREG; cualitativa MESREG: cualitativa ANOREG: cuantitativa discreta DIAOCU: cuantitativa ANOOCU: cuantitativa discreta DEPOCU: cualitativa MUPOCU: cualitativa EDADHOM: cuantitativa discreta EDADMUJ: cuantitativa discreta GETHOM: cualitativa GETMUJ: cualitativa NACHOM: cualitativa NACMUJ: cualitativa OCUHOM: cuantitativa OCUMUJ: cualitativa MEVER: cualitativa ANOVER: cualitativa

Resumen de datos

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)
# Crear una lista con los conjuntos de datos
datasets <- list(D2009, D2010, D2011, D2012, D2013, D2014, D2015, D2016, D2017, D2018, D2019, D2020, D2021)

# Crear un bucle for para analizar cada conjunto de datos y obtener los nombres de las variables numéricas
for (i in 1:length(datasets)) {
  vars_numéricas <- sapply(datasets[[i]], is.numeric)
  print(names(vars_numéricas))
}
```

```

## [1] "Depreg"   "Mesreg"    "Diaocu"    "Mesocu"    "Depoci"    "Edadhom"   "Edadmuj"
## [2] "Gethom"   "Getmuj"    "Nachom"    "Nacmuj"    "Ocuhom"    "Ocumuj"    "Mever"
## [15] "Anover"
## [1] "depreg"   "mesreg"    "añoreg"    "diaocu"    "mesocu"    "añooocu"
## [7] "depocu"    "edadhom"   "edadmuj"   "grethom"   "gretmuj"   "nachom"
## [13] "nacmuj"   "escohom"   "escomuj"   "ocupahom"  "ocupamu"   "añooocu"
## [1] "depreg"   "mes reg"   "anoreg"   "diaocu"    "mesocu"    "añooocu"
## [7] "depocu"    "edadhom"   "edadmuj"   "grethom"   "gretmuj"   "nachom"
## [13] "nacmuj"   "escohom"   "escomuj"   "ocupahom"  "ocupamu"   "añooocu"
## [1] "DEPREG"   "MESREG"    "ANOREG"    "DIAOCU"   "MESOCU"    "DEPOCU"    "EDADHOM"
## [8] "EDADMUJ"  "GETHOM"    "GETMUY"    "NACHOM"   "NACMUJ"    "ESCHOM"    "ESCMUJ"
## [1] "DEPREG"   "MESREG"    "ANOREG"    "DIAOCU"   "MESOCU"    "DEPOCU"    "EDADHOM"
## [8] "EDADMUJ"  "PUEHOM"   "PUEMUJ"   "NACHOM"   "NACMUJ"    "ESCHOM"    "ESCMUJ"
## [1] "DEPREG"   "MESREG"    "ANOREG"    "DIAOCU"   "MESOCU"    "AÑOOOCU"   "DEPOCU"
## [8] "EDADHOM"  "EDADMUJ"   "PPERHOM"   "PPEMUJ"   "NACHOM"   "NACMUJ"    "ESCHOM"
## [15] "ESCMUJ"   "CIUOHOM"   "CIUOMUJ"   "CIUOHOM"  "CIUOMUJ"   "CIUOHOM"   "CIUOMUJ"
## [1] "DEPREG"   "MESREG"    "ANOREG"    "DIAOCU"   "MESOCU"    "AÑOOOCU"   "DEPOCU"
## [8] "EDADHOM"  "EDADMUJ"   "PPERHOM"   "PPEMUJ"   "NACHOM"   "NACMUJ"    "ESCHOM"
## [15] "ESCMUJ"   "CIUOHOM"   "CIUOMUJ"   "CIUOHOM"  "CIUOMUJ"   "CIUOHOM"   "CIUOMUJ"
## [1] "DEPREG"   "MESREG"    "ANOREG"    "DIAOCU"   "MESOCU"    "AÑOOOCU"   "DEPOCU"
## [8] "EDADHOM"  "EDADMUJ"   "PPERHOM"   "PPEMUJ"   "NACHOM"   "NACMUJ"    "ESCHOM"
## [15] "ESCMUJ"   "CIUOHOM"   "CIUOMUJ"   "CIUOHOM"  "CIUOMUJ"   "CIUOHOM"   "CIUOMUJ"
## [1] "DEPREG"   "MESREG"    "ANOREG"    "DIAOCU"   "MESOCU"    "AÑOOOCU"   "DEPOCU"
## [8] "EDADHOM"  "EDADMUJ"   "PPERHOM"   "PPEMUJ"   "NACHOM"   "NACMUJ"    "ESCHOM"
## [15] "ESCMUJ"   "CIUOHOM"   "CIUOMUJ"   "CIUOHOM"  "CIUOMUJ"   "CIUOHOM"   "CIUOMUJ"
## [1] "DEPREG"   "MESREG"    "ANOREG"    "DIAOCU"   "MESOCU"    "AÑOOOCU"   "DEPOCU"
## [8] "EDADHOM"  "EDADMUJ"   "PPERHOM"   "PPEMUJ"   "NACHOM"   "NACMUJ"    "ESCHOM"
## [15] "ESCMUJ"   "CIUOHOM"   "CIUOMUJ"   "CIUOHOM"  "CIUOMUJ"   "CIUOHOM"   "CIUOMUJ"
## [1] "DEPREG"   "MESREG"    "ANOREG"    "DIAOCU"   "MESOCU"    "AÑOOOCU"   "DEPOCU"
## [8] "EDADHOM"  "EDADMUJ"   "PPERHOM"   "PPEMUJ"   "NACHOM"   "NACMUJ"    "ESCHOM"
## [15] "ESCMUJ"   "CIUOHOM"   "CIUOMUJ"   "CIUOHOM"  "CIUOMUJ"   "CIUOHOM"   "CIUOMUJ"
## [1] "DEPREG"   "MESREG"    "ANOREG"    "DIAOCU"   "MESOCU"    "AÑOOOCU"   "DEPOCU"
## [8] "EDADHOM"  "EDADMUJ"   "PPERHOM"   "PPEMUJ"   "NACHOM"   "NACMUJ"    "ESCHOM"
## [15] "ESCMUJ"   "CIUOHOM"   "CIUOMUJ"   "CIUOHOM"  "CIUOMUJ"   "CIUOHOM"   "CIUOMUJ"
## [1] "DEPREG"   "MESREG"    "ANOREG"    "DIAOCU"   "MESOCU"    "AÑOOOCU"   "DEPOCU"
## [8] "EDADHOM"  "EDADMUJ"   "PPERHOM"   "PPEMUJ"   "NACHOM"   "NACMUJ"    "ESCHOM"
## [15] "ESCMUJ"   "CIUOHOM"   "CIUOMUJ"   "CIUOHOM"  "CIUOMUJ"   "CIUOHOM"   "CIUOMUJ"

```

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.

Tiempo: Es importante poder ver el cambio a través del tiempo, si ha habido un incremento o decremento, tanto en matrimonios como en divorcios

```
M2021[, c(7, 8)]
```

```
## # A tibble: 87,480 × 2
##   NUNUH0  NUNUMU
##   <dbl>|<dbl>  <dbl>|<dbl>
## 1  9 [Ignorado] 9 [Ignorado]
## 2  1          1
## 3  9 [Ignorado] 9 [Ignorado]
## 4  1          1
## 5  9 [Ignorado] 9 [Ignorado]
## 6  9 [Ignorado] 9 [Ignorado]
## 7  9 [Ignorado] 9 [Ignorado]
## 8  1          1
## 9  9 [Ignorado] 9 [Ignorado]
## 10 9 [Ignorado] 9 [Ignorado]
## # ... # i 87,479 more rows
```

Saber si una persona ha estado previamente casada y si esto influye en la posibilidad de divorcio, cómo hipótesis se espera que las personas que han tenido más de dos nupcias antes, son más propensas al divorcio

Edadhom: "Edad del hombre" Edadmuj: "Edad de la mujer"

```
D2021[, c(10, 11)]
```

```

## # A tibble: 9,611 × 2
##   EDADHOMM EDADMUJ
##   <dbl><dbl>
## 1  29      25
## 2  36      37
## 3  34      31
## 4  48      33
## 5  44      28
## 6  39      27
## 7  45      40
## # ... with 9,604 more rows
## # i 9,611 more rows

```

La edad puede ser un dato interesante a explorar, esto para saber si los jóvenes tienen más tendencia a casarse o divorciarse y si los matrimonios más duraderos tienen menos divorcios

Genero: Es interesante ver que genero es más propenso a los divorcios, esto también puede estar relacionado con la cantidad de nupcias de una persona

Efectos de la pandemia: Ver como los divorcios y matrimonios se comportaron a partir de marzo de 2020 que fue el momento en que la cuarentena empezó a hacerse efectiva

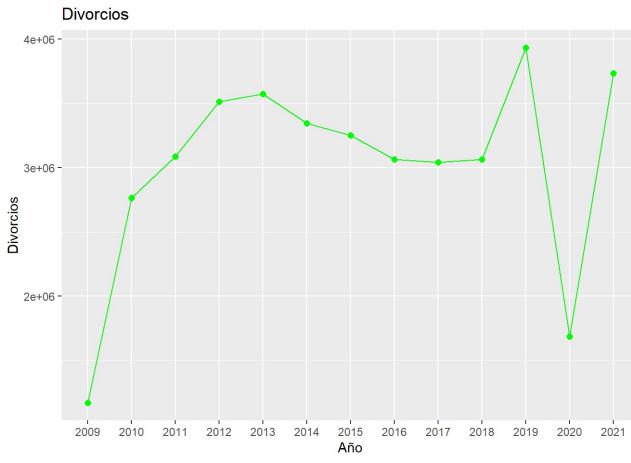
Graficos exploratorios

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

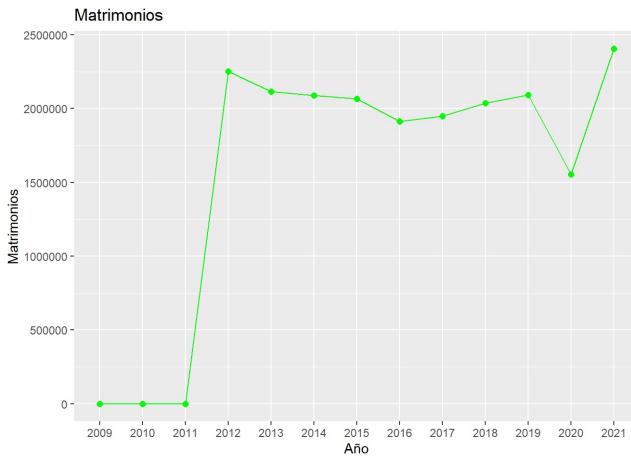
Analisis

Divorcios

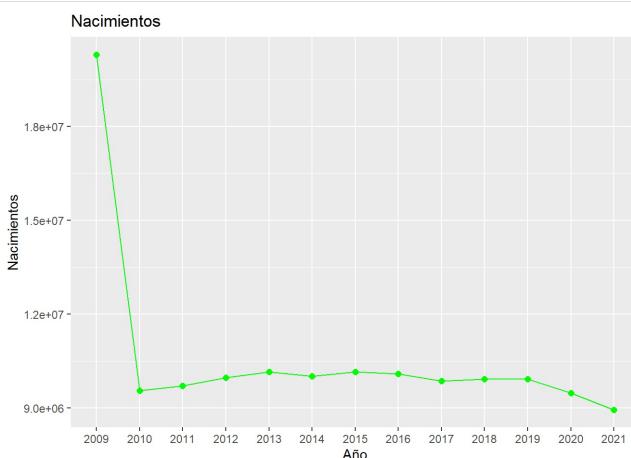
```
r EdadMujD2009 <- sum(D2009$edadmuj) EdadMujD2010 <- sum(D2010$edadmuj) EdadMujD2011 <- sum(D2011$edadmuj) EdadMujD2012 <- sum(D2012$EDADMUJ) EdadMujD2013 <- sum(D2013$edadmuj) EdadMujD2014 <- sum(D2014$edadmuj) EdadMujD2015 <- sum(D2015$edadmuj) EdadMujD2016 <- sum(D2016$edadmuj) EdadMujD2017 <- sum(D2017$edadmuj) EdadMujD2018 <- sum(D2018$edadmuj) EdadMujD2019 <- sum(D2019$edadmuj) EdadMujD2020 <- sum(D2020$edadmuj) EdadMujD2021 <- sum(D2021$edadmuj) EdadMujD2022 <- sum(D2022$edadmuj) EdadMujD2023 <- sum(D2023$edadmuj)
r dfD <- data.frame (año = c("2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", "2019", "2020", "2021"), matrimonios = c(EdadMujD2009
##     año matrimonios ## 1 2009      1173016 ## 2 2010      2764309 ## 3 2011      3086834 ## 4 2012      3512781 ## 5 2013      3572498 ## 6 2014      334298
r ggplot(dfD, aes(x=año, y=matrimonios, group = 1)) + geom point(size = 2, color = "green") + geom line(color="green") + labs(x = "Año", y = "Divorcios", title = "Matrimonios y Divorcios en Argentina 2009-2023")
```



```
## Matrimonios
r EdadMujM2009 <- sum(M2009$Edadmuj)
## Warning: Unknown or uninitialized column: `Edadmuj`.
r EdadMujM2010 <- sum(M2010$Edadmuj)
## Warning: Unknown or uninitialized column: `edadmuj`.
r EdadMujM2011 <- sum(M2011$Edadmuj)
## Warning: Unknown or uninitialized column: `edadmuj`.
r EdadMujM2012 <- sum(M2012$EDADMUJ) EdadMujM2013 <- sum(M2013$EDADMUJ) EdadMujM2014 <- sum(M2014$EDADMUJ) EdadMujM2015 <- sum(M2015$EDADMUJ) EdadMujM2016 <- sum(M2016$EDADMUJ) EdadMujM2017 <- sum(M2017$EDADMUJ) EdadMujM2018 <- sum(M2018$EDADMUJ) EdadMujM2019 <- sum(M2019$EDADMUJ) EdadMujM2020 <- sum(M2020$EDADMUJ)
r dfD <- data.frame(año = c("2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", "2019", "2020", "2021"), matrimonios = c(EdadMujM2009, EdadMujM2010, EdadMujM2011, EdadMujM2012, EdadMujM2013, EdadMujM2014, EdadMujM2015, EdadMujM2016, EdadMujM2017, EdadMujM2018, EdadMujM2019, EdadMujM2020, EdadMujM2021))
## año matrimonios ## 1 2009 0 ## 2 2010 0 ## 3 2011 0 ## 4 2012 2251360 ## 5 2013 2116792 ## 6 2014 208981
r ggplot(dfD, aes(x=año, y=matrimonios, group = 1)) + geom_point(size = 2, color = "green") + geom_line(color="green") + labs(x = "Año", y = "Matrimonios", title = "Matrimonios")
```



```
## Nacimientos
r EdadMujN2009 <- sum(N2009$Edadm) EdadMujN2010 <- sum(N2010$Edadm) EdadMujN2011 <- sum(N2011$Edadm) EdadMujN2012 <- sum(N2012$Edadm) EdadMujN2013 <- sum(N2013$Edadm) EdadMujN2014 <- sum(N2014$Edadm) EdadMujN2015 <- sum(N2015$Edadm) EdadMujN2016 <- sum(N2016$Edadm) EdadMujN2017 <- sum(N2017$Edadm) EdadMujN2018 <- sum(N2018$Edadm) EdadMujN2019 <- sum(N2019$Edadm) EdadMujN2020 <- sum(N2020$Edadm) EdadMujN2021 <- sum(N2021$Edadm)
r dNac <- data.frame(año = c("2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", "2019", "2020", "2021"), nacimientos = c(EdadMujN2009, EdadMujN2010, EdadMujN2011, EdadMujN2012, EdadMujN2013, EdadMujN2014, EdadMujN2015, EdadMujN2016, EdadMujN2017, EdadMujN2018, EdadMujN2019, EdadMujN2020, EdadMujN2021))
## año nacimientos ## 1 2009 20298288 ## 2 2010 9550498 ## 3 2011 9713100 ## 4 2012 9974089 ## 5 2013 10152809 ## 6 2014 1002153
r ggplot(dNac, aes(x=año, y=nacimientos, group = 1)) + geom_point(size = 2, color = "green") + geom_line(color="green") + labs(x = "Año", y = "Nacimientos", title = "Nacimientos")
```



Histogramas

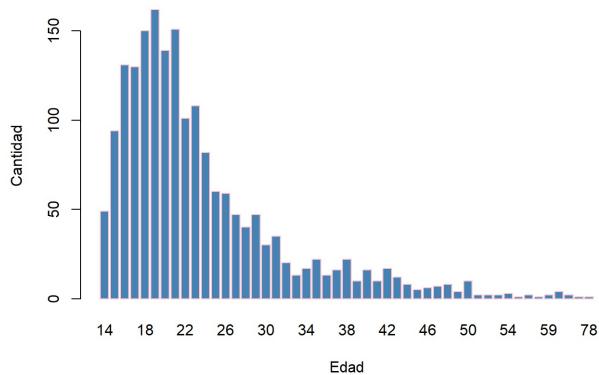
2009

```
library(dplyr)
library(ggplot2)

D2009 <- subset(D2009, Edadmuj < 999)

barplot(table(D2009$Edadmuj), main = "Edad de la mujer en divorcios 2009", xlab = "Edad", ylab = "Cantidad", col = "steelblue", border = "pink")
```

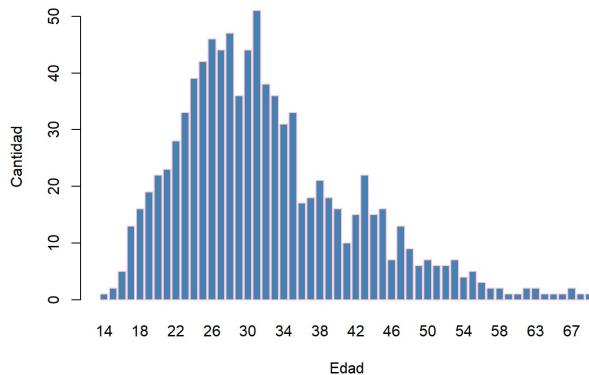
Edad de la mujer en divorcios 2009



2010

```
D2010 <- subset(D2010, edadmuj < 999)
barplot(table(D2010$edadmuj), main = "Edad de la mujer en divorcios 2010", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

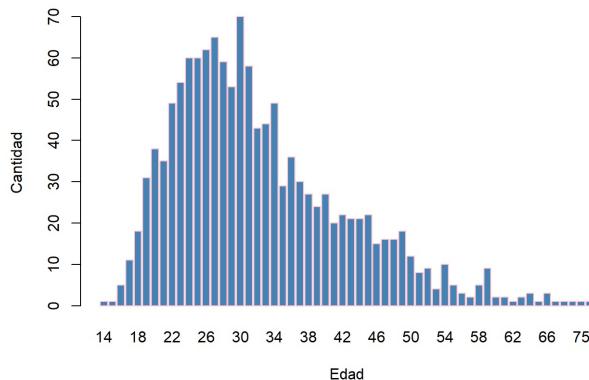
Edad de la mujer en divorcios 2010



2011

```
D2011 <- subset(D2011, edadmuj < 999)
barplot(table(D2011$edadmuj), main = "Edad de la mujer en divorcios 2011", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

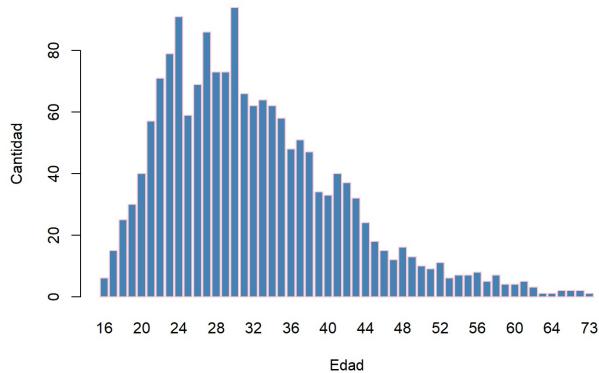
Edad de la mujer en divorcios 2011



2012

```
D2012 <- subset(D2012, EDADMUJ < 999)
barplot(table(D2012$EDADMUJ), main = "Edad de la mujer en divorcios 2012", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

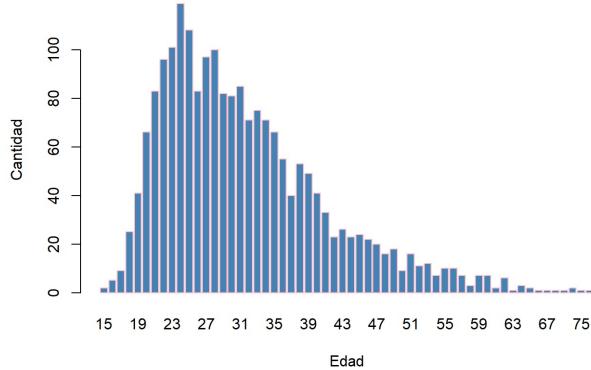
Edad de la mujer en divorcios 2012



2013

```
D2013 <- subset(D2013, EDADMUJ < 999)
barplot(table(D2013$EDADMUJ), main = "Edad de la mujer en divorcios 2013", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

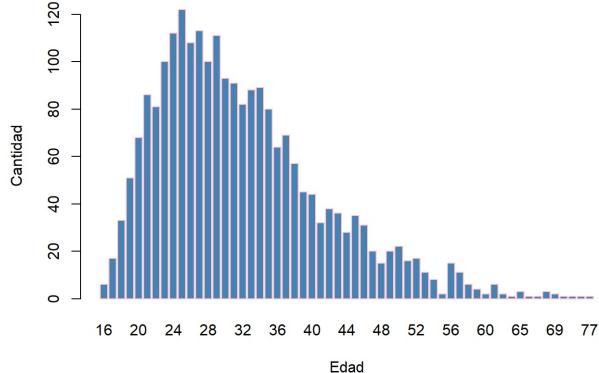
Edad de la mujer en divorcios 2013



2014

```
D2014 <- subset(D2014, EDADMUJ < 999)
barplot(table(D2014$EDADMUJ), main = "Edad de la mujer en divorcios 2014", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

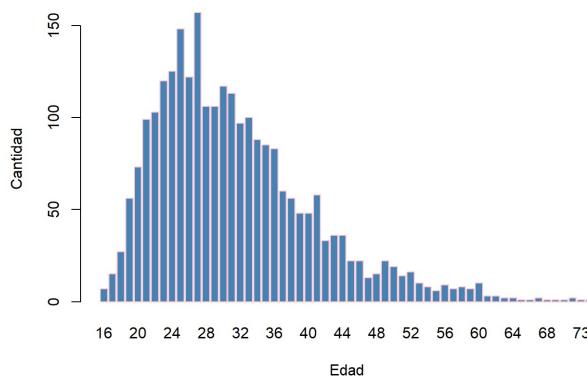
Edad de la mujer en divorcios 2014



2015

```
D2015 <- subset(D2015, EDADMUJ < 999)
barplot(table(D2015$EDADMUJ), main = "Edad de la mujer en divorcios 2015", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

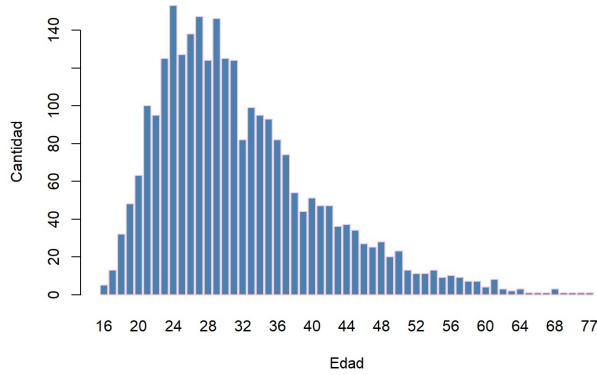
Edad de la mujer en divorcios 2015



2016

```
D2016 <- subset(D2016, EDADMUJ < 999)
barplot(table(D2016$EDADMUJ), main = "Edad de la mujer en divorcios 2016", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

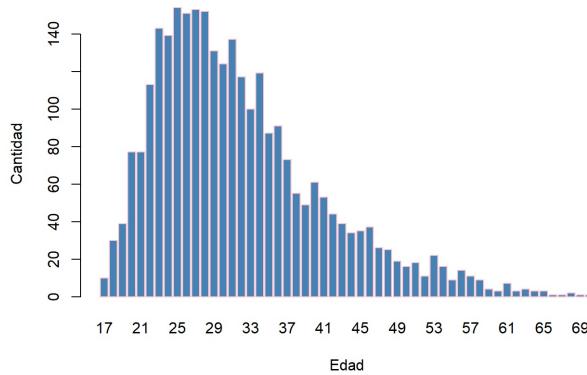
Edad de la mujer en divorcios 2016



2017

```
D2017 <- subset(D2017, EDADMUJ < 999)
barplot(table(D2017$EDADMUJ), main = "Edad de la mujer en divorcios 2017", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

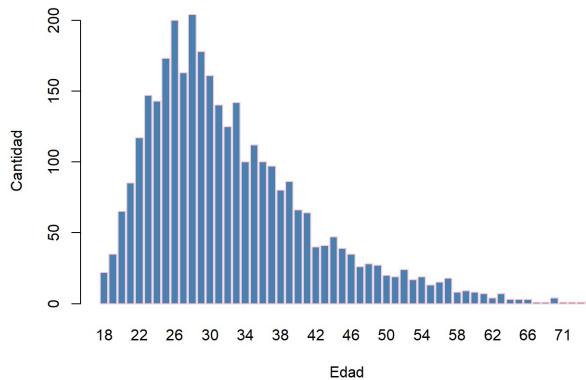
Edad de la mujer en divorcios 2017



2018

```
D2018 <- subset(D2018, EDADMUJ < 999)
barplot(table(D2018$EDADMUJ), main = "Edad de la mujer en divorcios 2018", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

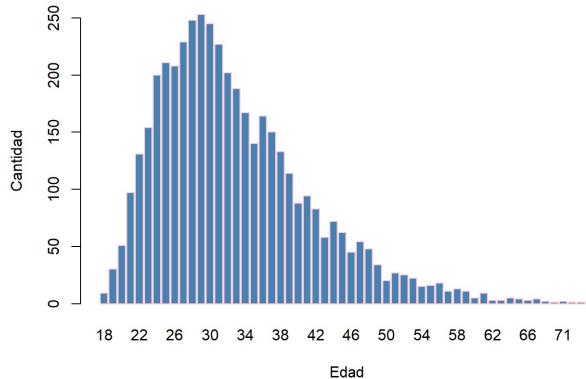
Edad de la mujer en divorcios 2018



2019

```
D2019 <- subset(D2019, EDADMUJ < 999)
barplot(table(D2019$EDADMUJ), main = "Edad de la mujer en divorcios 2019", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

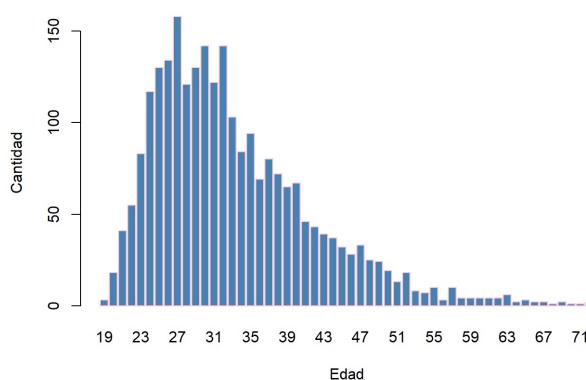
Edad de la mujer en divorcios 2019



2020

```
D2020 <- subset(D2020, EDADMUJ < 999)
barplot(table(D2020$EDADMUJ), main = "Edad de la mujer en divorcios 2020", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

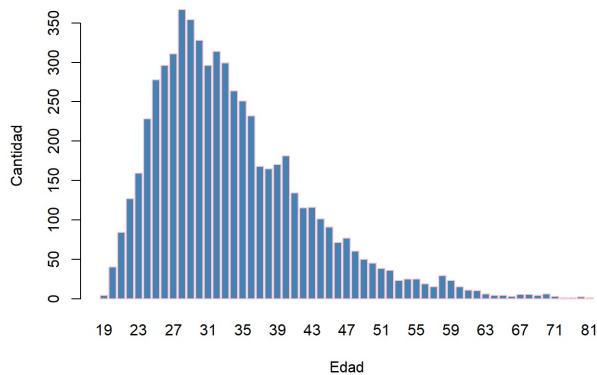
Edad de la mujer en divorcios 2020



2021

```
D2021 <- subset(D2021, EDADMUJ < 999)
barplot(table(D2021$EDADMUJ), main = "Edad de la mujer en divorcios 2021", xlab = "Edad", ylab = "Cantidad", col = "steelblue" , border = "pink")
```

Edad de la mujer en divorcios 2021



Combinando datos

```

colnames(D2009) <- c("DEPREG", "MUPREG", "MESREG", "ANOREG", "DIAOCU", "MESOCU", "ANOCU", "DEPOCU", "MUPOCU", "EDADHOM", "EDADMUJ", "PPERHOM", "PPERMUJ", "NACHOM", "NACMUJ", "CIUOHOM", "CIUOMUJ", "Mever", "Anover")
colnames(D2010) <- c("DEPREG", "MUPREG", "MESREG", "ANOREG", "DIAOCU", "MESOCU", "ANOCU", "DEPOCU", "MUPOCU", "EDADHOM", "EDADMUJ", "PPERHOM", "PPERMUJ", "NACHOM", "NACMUJ", "ESCHOM", "ESCMUJ", "CIUOHOM", "CIUOMUJ")
colnames(D2011) <- c("DEPREG", "MUPREG", "MESREG", "ANOREG", "DIAOCU", "MESOCU", "ANOCU", "DEPOCU", "MUPOCU", "EDADHOM", "EDADMUJ", "PPERHOM", "PPERMUJ", "NACHOM", "NACMUJ", "ESCHOM", "ESCMUJ", "CIUOHOM", "CIUOMUJ")
colnames(D2012) <- c("DEPREG", "MUPREG", "MESREG", "ANOREG", "DIAOCU", "MESOCU", "ANOCU", "DEPOCU", "MUPOCU", "EDADHOM", "EDADMUJ", "PPERHOM", "PPERMUJ", "NACHOM", "NACMUJ", "ESCHOM", "ESCMUJ", "CIUOHOM", "CIUOMUJ")
colnames(D2013) <- c("DEPREG", "MUPREG", "MESREG", "ANOREG", "DIAOCU", "MESOCU", "DEPOCU", "MUPOCU", "EDADHOM", "EDADMUJ", "PPERHOM", "PPERMUJ", "NACHOM", "NACMUJ", "ESCHOM", "ESCMUJ", "CIUOHOM", "CIUOMUJ")
colnames(D2014) <- c("DEPREG", "MUPREG", "MESREG", "ANOREG", "DIAOCU", "MESOCU", "DEPOCU", "MUPOCU", "EDADHOM", "EDADMUJ", "PPERHOM", "PPERMUJ", "NACHOM", "NACMUJ", "ESCHOM", "ESCMUJ", "CIUOHOM", "CIUOMUJ")
colnames(D2015) <- c("DEPREG", "MUPREG", "MESREG", "ANOREG", "DIAOCU", "MESOCU", "ANOCU", "DEPOCU", "MUPOCU", "EDADHOM", "EDADMUJ", "PPERHOM", "PPERMUJ", "NACHOM", "NACMUJ", "ESCHOM", "ESCMUJ", "CIUOHOM", "CIUOMUJ")

D2009 <- D2009 %>% mutate(ANOREG = as.numeric(ANOREG))
D2010 <- D2010 %>% mutate(ANOREG = as.numeric(ANOREG))
D2011 <- D2011 %>% mutate(ANOREG = as.numeric(ANOREG))
D2012 <- D2012 %>% mutate(ANOREG = as.numeric(ANOREG))
D2013 <- D2013 %>% mutate(ANOREG = as.numeric(ANOREG))
D2014 <- D2014 %>% mutate(ANOREG = as.numeric(ANOREG))
D2015 <- D2015 %>% mutate(ANOREG = as.numeric(ANOREG))
D2016 <- D2016 %>% mutate(ANOREG = as.numeric(ANOREG))
D2017 <- D2017 %>% mutate(ANOREG = as.numeric(ANOREG))
D2018 <- D2018 %>% mutate(ANOREG = as.numeric(ANOREG))
D2019 <- D2019 %>% mutate(ANOREG = as.numeric(ANOREG))
D2020 <- D2020 %>% mutate(ANOREG = as.numeric(ANOREG))
D2021 <- D2021 %>% mutate(ANOREG = as.numeric(ANOREG))

D2009 <- D2009 %>% mutate(ANOCU = as.numeric(ANOCU))
D2010 <- D2010 %>% mutate(ANOCU = as.numeric(ANOCU))
D2011 <- D2011 %>% mutate(ANOCU = as.numeric(ANOCU))

D2009 <- D2009 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2010 <- D2010 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2011 <- D2011 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2012 <- D2012 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2013 <- D2013 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2014 <- D2014 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2015 <- D2015 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2016 <- D2016 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2017 <- D2017 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2018 <- D2018 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2019 <- D2019 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2020 <- D2020 %>% mutate(CIUOHOM = as.character(CIUOHOM))
D2021 <- D2021 %>% mutate(CIUOHOM = as.character(CIUOHOM))

D2009 <- D2009 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2010 <- D2010 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2011 <- D2011 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2012 <- D2012 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2013 <- D2013 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2014 <- D2014 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2015 <- D2015 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2016 <- D2016 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2017 <- D2017 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2018 <- D2018 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2019 <- D2019 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2020 <- D2020 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))
D2021 <- D2021 %>% mutate(CIUOMUJ = as.character(CIUOMUJ))

#divorcios <- bind_rows(D2009, D2010, D2011, D2012, D2013, D2014, D2015, D2016, D2017, D2018, D2019, D2020, D2021)
#divorcios <- bind_rows( D2012, D2013, D2014, D2015, D2016, D2017, D2018, D2019, D2020, D2021)
#l = list(D2012, D2013, D2014, D2015, D2016, D2017, D2018, D2019, D2020, D2021)
#divorcios <- do.call("rbind", l)

str(divorcios)

```

```

### tibble [30,378 x 19] (S3:tbl_df/tbl/data.frame)
### $ DEPREG : dbl+lbl [1:30378] 17, 12, 1, 14, 1, 1, 1, 1, 1, 1, 9, 22, 3, ...
### ..@ labels: Named num [1:22] 1 2 3 4 5 6 7 8 9 10 ...
### ..@ label : chr "Departamento de registro"
### $ MUPREG : chr+lbl [1:30378] 1708, 1213, 0101, 1416, 0101, 0101, 0101, 0101, 0101...
### ..@ labels: Named chr [1:342] "1010" "2010" "0110" "1210" ...
### ..@ attr(*, "names")= chr [1:342] "San Antonio Suchitepéquez" "San Jacinto" "San Juan Sacatepéquez" "Tejutla" ...
### ..@ label : chr "Municipio de registro"
### $ MESREG : dbl+lbl [1:30378] 3, 5, 4, 6, 10, 2, 8, 11, 9, 8, 4, 1, 4, ...
### ..@ labels: Named num [1:12] 1 2 3 4 5 6 7 8 9 10 ...
### ..@ attr(*, "names")= chr [1:12] "Enero" "Febrero" "Marzo" "Abril" ...
### ..@ label : chr "Mes de registro"
### $ ANOREG : num [1:30378] 2012 2012 2012 2012 2012 ...
### $ DIAOCU : num [1:30378] 16 3 27 28 12 9 11 3 29 16 ...
### $ MESOCU : dbl+lbl [1:30378] 2, 2, 3, 5, 3, 1, 6, 8, 2, 5, 5, 2, 4, ...
### ..@ label : chr "Mes de ocurrencia"
### ..@ format.spss: chr "F2.0"
### ..@ labels : Named num [1:12] 1 2 3 4 5 6 7 8 9 10 ...
### ..@ attr(*, "names")= chr [1:12] "Enero" "Febrero" "Marzo" "Abril" ...
### $ DEPOCU : dbl+lbl [1:30378] 17, 12, 1, 14, 1, 1, 1, 1, 1, 1, 9, 22, 1, ...
### ..@ labels: Named num [1:22] 1 2 3 4 5 6 7 8 9 10 ...
### ..@ attr(*, "names")= chr [1:22] "Guatemala" "El Progreso" "Sacatepéquez" "Chimaltenango" ...
### ..@ label : chr "Departamento de ocurrencia"
### $ MUPOCU : chr+lbl [1:30378] 1703, 1215, 0101, 1401, 0101, 0101, 0101, 0101, 0101...
### ..@ labels: Named chr [1:342] "1010" "2010" "0110" "1210" ...
### ..@ attr(*, "names")= chr [1:342] "San Antonio Suchitepéquez" "San Jacinto" "San Juan Sacatepéquez" "Tejutla" ...
### ..@ label : chr "Municipio de ocurrencia"
### $ EDADHOM: dbl+lbl [1:30378] 999, 35, 33, 31, 27, 37, 41, 36, 46, 38, 3...
### ..@ label : chr "Edad del hombre"
### ..@ format.spss: chr "F4.0"
### ..@ labels : Named num 999
### ..@ attr(*, "names")= chr "Ignorado"
### $ EDADMU: dbl+lbl [1:30378] 33, 30, 32, 28, 29, 30, 42, 30, 37, 35, 35, 33, 38, ...
### ..@ label : chr "Edad de la mujer"
### ..@ format.spss: chr "F4.0"
### ..@ labels : Named num 999
### ..@ attr(*, "names")= chr "Ignorado"
### $ PPERHOM: dbl+lbl [1:30378] 9, 2, 9, 9, 2, 9, 2, 2, 2, 2, 1, 9, 2, 2, 1, 2, 2, 9...
### ..@ labels: Named num [1:6] 1 2 9 3 4 5
### ..@ attr(*, "names")= chr [1:6] "Indigena" "No indigena" "Ignorado" "Xinca" ...
### ..@ label : chr "Grupo Etnico del hombre"
### $ PPERMU: dbl+lbl [1:30378] 2, 2, 2, 2, 9, 9, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 9...
### ..@ labels: Named num [1:6] 1 2 9 3 4 5
### ..@ attr(*, "names")= chr [1:6] "Indigena" "No indigena" "Ignorado" "Xinca" ...
### ..@ label : chr "Grupo Etnico de la mujer"
### $ NACHOM : dbl+lbl [1:30378] 320, 320, 320, 320, 320, 320, 320, 320, 320, 320, 32...
### ..@ labels: Named num [1:105] 32 56 68 84 124 156 170 188 192 222 ...
### ..@ attr(*, "names")= chr [1:105] "Argentina" "Bélgica" "Bolivia" "Belice" ...
### ..@ label : chr "Nacionalidad del hombre"
### $ NACMU : dbl+lbl [1:30378] 320, 320, 320, 320, 320, 320, 320, 320, 320, 320, 32...
### ..@ labels: Named num [1:103] 78 84 172 188 192 218 222 276 320 340 ...
### ..@ attr(*, "names")= chr [1:103] "Brasil" "Belice" "Colombia" "Costa Rica" ...
### ..@ label : chr "Nacionalidad de la mujer"
### $ ESCHOM : dbl+lbl [1:30378] 9, 5, 5, 1, 4, 9, 5, 5, 9, 4, 9, 3, 5, 9, 9, 3, ...
### ..@ labels: Named num [1:8] 1 2 3 4 5 9 6 0
### ..@ attr(*, "names")= chr [1:8] "Ninguna" "Primaria" "Básico" "Diversificado" ...
### ..@ label : chr "Escolaridad del hombre"
### $ ESCMU : dbl+lbl [1:30378] 4, 9, 5, 4, 5, 9, 5, 5, 5, 5, 4, 4, 4, 5, 4, 5, 5, 9...
### ..@ labels: Named num [1:8] 1 2 3 4 5 9 6 0
### ..@ attr(*, "names")= chr [1:8] "Ninguna" "Primaria" "Básico" "Diversificado" ...
### ..@ label : chr "Escolaridad de la mujer"
### $ CIUOHOM: chr [1:30378] "9712" "110" "2142" "8189" ...
### $ CIUMUJ: chr [1:30378] "10" "1120" "1120" "1120" ...
### $ AÑOOCU : num [1:30378] NA NA NA NA NA NA NA NA NA ...

```

```

## Warning: The dot-dot notation `..density..` was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(density)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

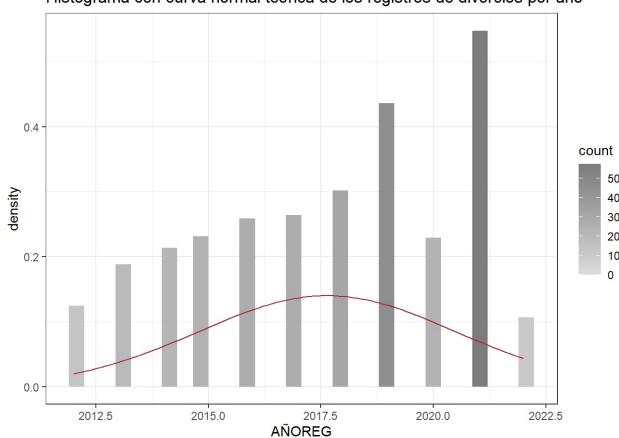
```

```

## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.

```

Histograma con curva normal teórica de los registros de divorcios por año

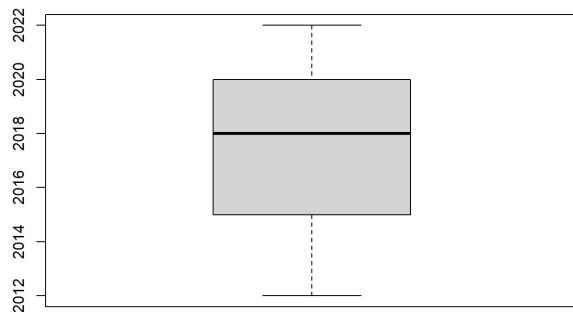


```

boxplot(divorcios$AÑOREG, main = "Caja y Bigotes de registro de divorcios por año (2012 - 2021)", xlab = "Registro por año")

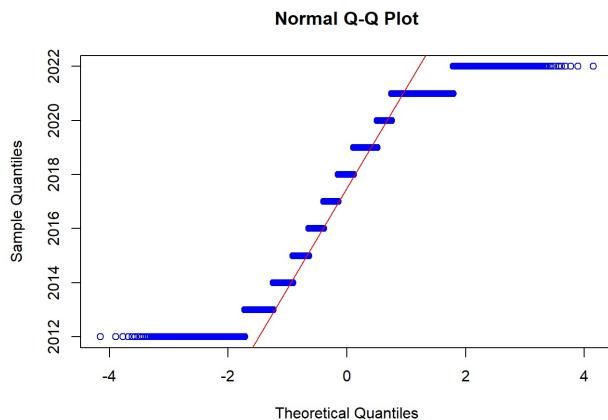
```

Caja y Bigotes de registro de divorcios por año (2012 - 2021)



Registro por año

Diagrama de qqnormal



Registros de Edad hombre

Prueba de normalidad para la edad del hombre:

```
## `stat_bin()` using `bins = 30` . Pick better value with `binwidth` .
```

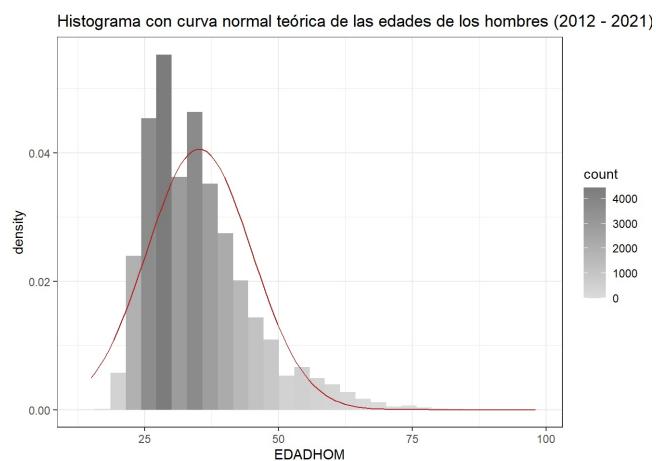
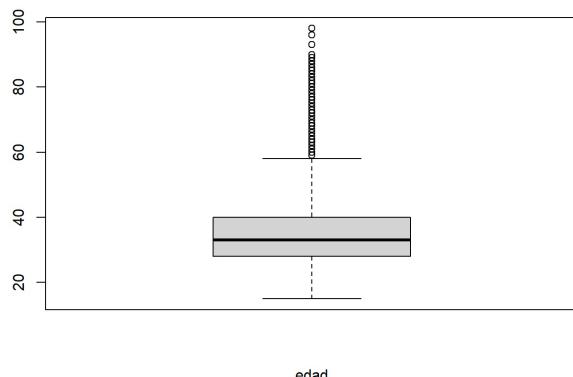


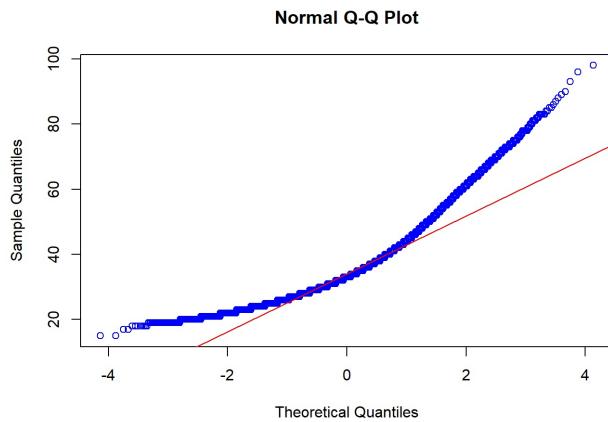
Diagrama de caja y bigotes

Caja y Bigotes de edad de los hombres al divorciarse (2012 - 2021)



edad

Diagrama qqnormal



Registros de edad de mujeres

Prueba de normalidad para la edad de la mujer:

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

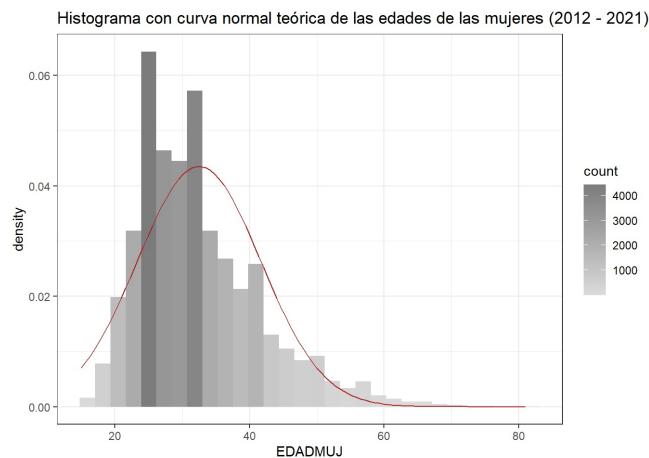


Diagrama de caja y bigotes

Caja y Bigotes de edad de las mujeres al divorciarse (2012 - 2021)

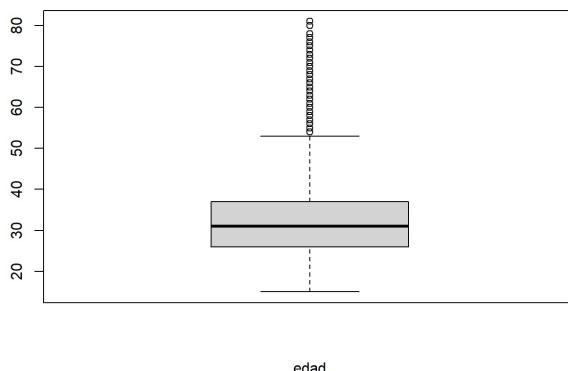


Diagrama qqnormal

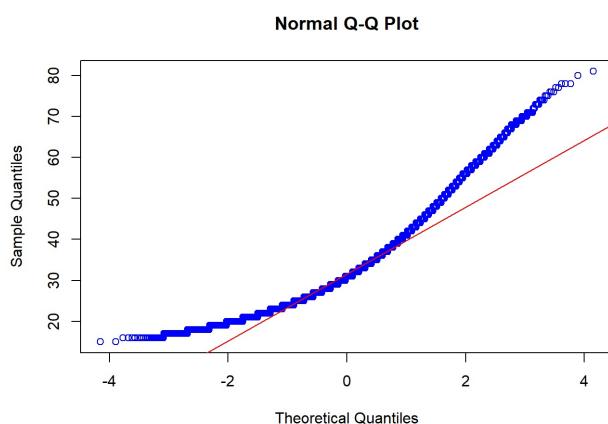


Tabla de frecuencias para variables cualitativas

Tabla de frecuencias para el departamento de ocurrencia y representación gráfica

```

D2009 <- D2009 %>%
  mutate_if(is.labelled,list(as_factor))
D2010 <- D2010 %>%
  mutate_if(is.labelled,list(as_factor))
D2011 <- D2011 %>%
  mutate_if(is.labelled,list(as_factor))
D2012 <- D2012 %>%
  mutate_if(is.labelled,list(as_factor))
D2013 <- D2013 %>%
  mutate_if(is.labelled,list(as_factor))
D2014 <- D2014 %>%
  mutate_if(is.labelled,list(as_factor))
D2015 <- D2015 %>%
  mutate_if(is.labelled,list(as_factor))
D2016 <- D2016 %>%
  mutate_if(is.labelled,list(as_factor))
D2017 <- D2017 %>%
  mutate_if(is.labelled,list(as_factor))
D2018 <- D2018 %>%
  mutate_if(is.labelled,list(as_factor))
D2019 <- D2019 %>%
  mutate_if(is.labelled,list(as_factor))
D2020 <- D2020 %>%
  mutate_if(is.labelled,list(as_factor))
D2021 <- D2021 %>%
  mutate_if(is.labelled,list(as_factor))

divorciosLabels <- bind_rows( D2012, D2013, D2014, D2015, D2016, D2017, D2018, D2019, D2020, D2021)

##      Guatemala   El Progreso Sacatepéquez Chimaltenango Escuintla
## 10008          554           482        815       1418
## Santa Rosa      Solola    Totonicapan Quetzaltenango Suchitepéquez
## 898            312           404        2751       705
## Retalhuleu     San Marcos Huehuetenango Quiché    Baja Verapaz
## 912          1455          1107        635       645
## Alta Verapaz     Petén         Izabal    Zacapa    Chiquimula
## 892          579           978        762       798
## Jalapa          Jutiapa    Sacatepéquez Sololá    Totonicapán
## 604          1474          162         91       171
## Suchitepéquez   Quiché        Petén
## 267            274          225

##      Guatemala   El Progreso Sacatepéquez Chimaltenango Escuintla
## 10008          554           482        815       1418
## Santa Rosa      Solola    Totonicapan Quetzaltenango Suchitepéquez
## 898            312           404        2751       705
## Retalhuleu     San Marcos Huehuetenango Quiché    Baja Verapaz
## 912          1455          1107        635       645
## Alta Verapaz     Petén         Izabal    Zacapa    Chiquimula
## 892          579           978        762       798
## Jalapa          Jutiapa    Sacatepéquez Sololá    Totonicapán
## 604          1474          162         91       171
## Suchitepéquez   Quiché        Petén
## 267            274          225

```

Distribución de departamento de ocurrencia (2012 - 2021)

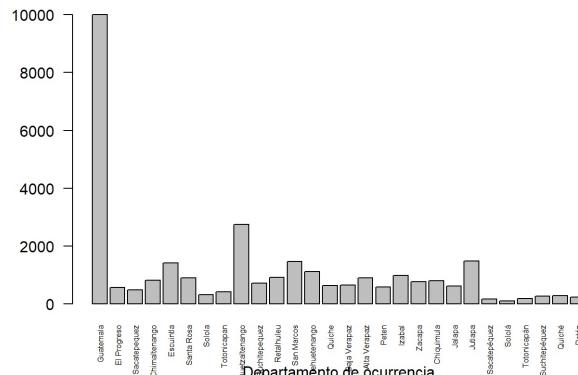


Tabla de frecuencias para el municipio de ocurrencia

```

##      Guatemala      Santa Catarina Pinula
## 6367          239
## San José Pinula  San José del Golfo
## 176            34
## Palencia        Chinautla
## 79             199
## San Pedro Ayampuc Mixco
## 95             779
## San Pedro Sacatepéquez San Juan Sacatepéquez
## 185            131
## San Raymundo    Chuarrancho
## 91             19
## Fraijanes        Amatitlán
## 82             356
## Villa Nueva     Villa Canales
## 749            314
## Petapa          Guastatoya
## 231            206
## Moraízán        San Agustín Acasaguastlán
## 29             58
## San Cristóbal Acasaguastlán El Jícaro
## 16             53
## Sansare         Sanarate
## 33             107
## San Antonio la Paz Antigua Guatemala
## 52             262
## Jocotenango     Pastores
## 62             19
## Sumpango         Santo Domingo Xenacoj
## 18              7
## Santiago Sacatepéquez San Bartolomé Milpas Altas
## 30              12
## San Lucas Sacatepéquez Santa Lucía Milpas Altas
## 81              41
## Magdalena Milpas Altas Santa María de Jesús
## 19              5
## Ciudad Vieja    San Miguel Dueñas
## 35              15
## Alotenango San Antonio Aguas Calientes
## 12              21
## Santa Catarina Barahona Chimaltenango
## 5               296
## San José Poaquil  San Martín Jilotepeque
## 18              47
## Comalapa         Santa Apolonia
## 40              8
## Tecpán Guatemala Patzún
## 90              53

```

##	Pochuta	Patzicia
##	4	45
##	Santa Cruz Balanyá	Acatenango
##	14	19
##	Yepocapa	San Andrés Itzapa
##	10	40
##	Parramos	Zaragoza
##	25	69
##	El Tejar	Escuintla
##	37	524
##	Santa Lucía Cotzumalguapa	La Democracia
##	170	105
##	Siquinalá	Masagua
##	32	28
##	Tiquisate	La Gomera
##	156	72
##	Guanagazapa	San José
##	16	174
##	Iztapa	Palín
##	46	84
##	San Vicente Pacaya	Nueva Concepción
##	22	89
##	Cuillapa	Barberena
##	182	156
##	Santa Rosa de Lima	Casillas
##	47	40
##	San Rafael las Flores	Oratorio
##	15	37
##	San Juan Tecuaco	Chiquimulilla
##	11	134
##	Taxisco	Santa María Ixhuatán
##	49	33
##	Guazacapán	Santa Cruz Naranjo
##	61	33
##	Pueblo Nuevo Viñas	Nueva Santa Rosa
##	35	65
##	Sololá	San José Chacayá
##	158	1
##	Santa María Visitación	Santa Lucía Utatlán
##	0	30
##	Nahualá	Santa Catarina Ixtahuacán
##	34	19
##	Santa Clara la Laguna	Concepción
##	13	3
##	San Andrés Semetabaj	Panajachel
##	5	52
##	Santa Catarina Palopó	San Antonio Palopó
##	4	7
##	San Lucas Tolimán	Santa Cruz la Laguna
##	31	5
##	San Pablo la Laguna	San Marcos la Laguna
##	1	0
##	San Juan la Laguna	San Pedro la Laguna
##	14	11
##	Santiago Atitlán	Totonicapán
##	15	290
##	San Cristóbal Totonicapán	San Francisco el Alto
##	39	53
##	San Andrés Xecul	Momostenango
##	29	111
##	Santa María Chiquimula	Santa Lucía la Reforma
##	25	12
##	San Bartolo	Quetzaltenango
##	16	1166
##	Salcajá	Olintepeque
##	131	67
##	San Carlos Sija	Sibilia
##	57	16
##	Cabricán	Cajolá
##	17	6
##	San Miguel Sigüilá	Ostuncalco
##	9	77
##	San Mateo	Concepción Chiquirichapa
##	20	30
##	San Martín Sacatepéquez	Almolonga
##	36	29
##	Cantel	Huitán
##	121	14
##	Zunil	Colombia
##	51	87
##	San Francisco la Unión	El Palmar
##	10	38
##	Coatepeque	Génova
##	501	113
##	Flores Costa Cuca	La Esperanza
##	58	79
##	Palestina de los Altos	Mazatenango
##	18	382
##	Cuyotenango	San Francisco Zapotitlán
##	92	67
##	San Bernardino	San José el Idolo
##	57	11
##	Santo Domingo Suchitepéquez	San Lorenzo
##	15	28
##	Samayac	San Pablo Jocopilas
##	44	23
##	San Antonio Suchitepéquez	San Miguel Panán
##	60	4
##	San Gabriel	Chicacao
##	12	38
##	Patulul	Santa Bárbara
##	28	24
##	San Juan Bautista	Santo Tomás la Unión
##	6	22
##	Zunilito	Pueblo Nuevo
##	17	29
##	Río Bravo	Retalhuleu
##	37	484
##	San Sebastián	Santa Cruz Muluá
##	79	27
##	San Martín Zapotitlán	San Felipe
##	29	46
##	San Andrés Villa Seca	Champerico
##	75	78
##	Nuevo San Carlos	El Asintal
##	50	44
##	San Marcos	San Antonio Sacatepéquez
##	167	20
##	Comitancillo	San Miguel Ixtahuacán
##	12	9
##	Concepción Tutuapa	Tacaná
##	13	21
##	Sibinal	Tajumulco
##	3	8
##	Tejutla San Rafael Pié de la Cuesta	44
##	57	

##	Nuevo Progreso	El Tumbador
##	43	42
##	El Rodeo	Malacatán
##	11	232
##	Catarina	Ayutla
##	35	151
##	Ocós	San Pablo
##	105	58
##	El Quetzal	La Reforma
##	71	29
##	Pajapita	Ixchiguán
##	49	36
##	San José Ojetenán	San Cristóbal Cuchío
##	5	47
##	Sipacapa	Esquipulas Palo Gordo
##	9	8
##	Río Blanco	Huehuetenango
##	26	516
##	Chiantla	Malacatancito
##	46	36
##	Cuilco	Nentón
##	24	12
##	San Pedro Necta	Jacaltenango
##	16	23
##	Soloma	Ixtahuacán
##	44	23
##	La Libertad	San Miguel Acatán
##	108	7
##	San Rafael la Independencia	Todos Santos Cuchumatán
##	4	37
##	San Juan Atitán	Santa Eulalia
##	14	15
##	San Mateo Ixtatán	Colotenango
##	2	14
##	San Sebastián Huehuetenango	Tectitán
##	7	4
##	Concepción Huista	San Juan Ixcoy
##	7	9
##	San Antonio Huista	San Sebastián Coatán
##	16	37
##	Bárillas	Aguacatán
##	33	21
##	San Rafael Petzal	San Gaspar Ixchil
##	5	2
##	Santiago Chimaltenango	Santa Ana Huista
##	5	4
##	Unión Cantinil	Santa Cruz del Quiché
##	2	280
##	Chiché	Chinique
##	16	23
##	Zacualpa	Chajul
##	33	8
##	Chichicastenango	Patzité
##	79	5
##	San Antonio Ilotenango	San Pedro Jocopilas
##	22	12
##	Cunén	San Juan Cotzal
##	33	10
##	Joyabaj	Nebaj
##	105	73
##	San Andrés Sajcabajá	Uspantán
##	12	37
##	Sacapulas	San Bartolomé Jocotenango
##	41	5
##	Canillá	Chicamán
##	23	18
##	Ixcán	Pachalum
##	44	30
##	Salamá	San Miguel Chicaj
##	304	23
##	Rabinal	Cubulco
##	84	73
##	Granados	El Chol
##	28	29
##	San Jerónimo	Purulhá
##	80	24
##	Cobán	Santa Cruz Verapaz
##	453	32
##	San Cristóbal Verapaz	Tactic
##	68	48
##	Tamahú	Tucurú
##	6	6
##	Panzós	Senahú
##	7	13
##	San Pedro Carchá	San Juan Chamelco
##	97	51
##	Lanquín	Cahabón
##	6	5
##	Chisec	Chahal
##	32	5
##	Fray Bartolomé de las Casas	Santa Catalina la Tinta
##	48	11
##	Raxruhá	Flores
##	4	108
##	San Benito	San Andrés
##	208	41
##	San Francisco	Santa Ana
##	37	18
##	Dolores	San Luis
##	49	27
##	Sayaxché	Melchor de Mencos
##	36	37
##	Poptún	Las Cruces
##	129	12
##	Puerto Barrios	Livingston
##	617	58
##	El Estor	Morales
##	22	197
##	Los Amates	Zacapa
##	84	395
##	Estanzuela	Río Hondo
##	56	85
##	Gualán	Teculután
##	107	32
##	Usumatlán	Cabañas
##	13	18
##	San Diego	La Unión
##	27	15
##	Huité	Chiquimula
##	9	308
##	San José La Arada	San Juan Ermita
##	27	21
##	Jocotán	Camotán
##	31	15
##	Olopa	Esquipulas
##	24	116

Concepción	Las Minas	Quetzaltepeque
##	78	53
##	San Jacinto	Ipala
##	10	115
##	Jalapa	San Pedro Pinula
##	387	29
##	San Luis Jilotepeque	San Manuel Chaparrón
##	31	19
##	San Carlos Alzataate	Monjas
##	10	99
##	Mataquescuintla	Jutiapa
##	29	500
##	El Progreso	Santa Catarina Mita
##	152	91
##	Agua Blanca	Asunción Mita
##	78	160
##	Yupiltepeque	Atescatempa
##	23	49
##	Jerez	El Adelanto
##	13	9
##	Zapotitlán	Comapa
##	13	39
##	Jalpatagua	Conguaco
##	114	9
##	Moyuta	Pasaco
##	91	5
##	San José Acatempa	Quesada
##	81	47
##	San José La Maquina	La Blanca
##	0	8
##	El Chal	San Jorge
##	6	5
##	Sipacate	San José la Maquina
##	0	4
##	Petatán	Extranjero
##	0	0
##	Ignorado	
##	0	

Distribución de municipio de ocurrencia (2012 - 2021)

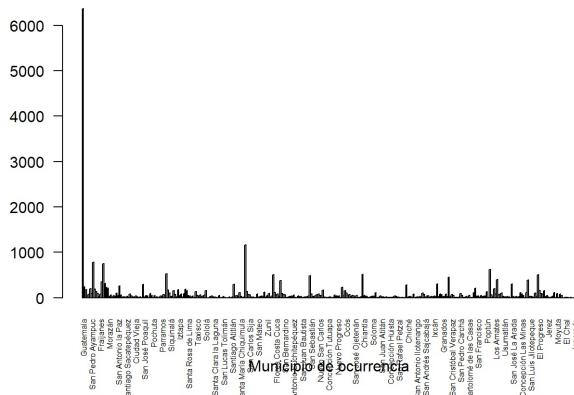


Tabla de frecuencias para la escolaridad del hombre

```

##          Ninguna      Primaria      Básico Diversificado Universitario
##            318           4846         3138       10566          3149
##        Ignorado      Ninguno     Postgrado      Básica    Post Grado
##            5275           2752           7          280           47
##        Doctorado          0
##
```

Distribución de escolaridad de los hombres (2012 - 2021)

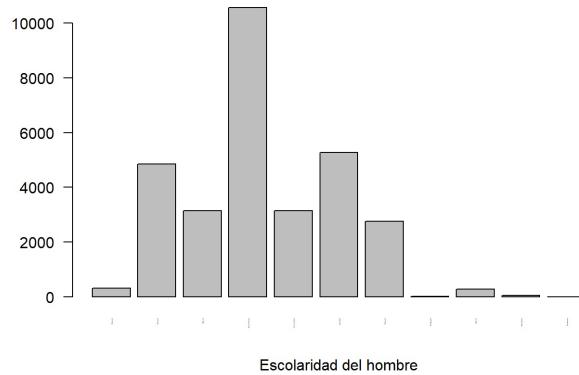


Tabla de frecuencias para la escolaridad de la Mujer

```

##          Ninguna      Primaria      Básico Diversificado Universitario
##          347           5118        3439       11462        3082
## Ignorado Ninguno      Postgrado      Básica Post Grado
##          3534           3074         4          289          29
## Doctorado
##          0

```

Distribución de escolaridad de las mujeres (2012 - 2021)

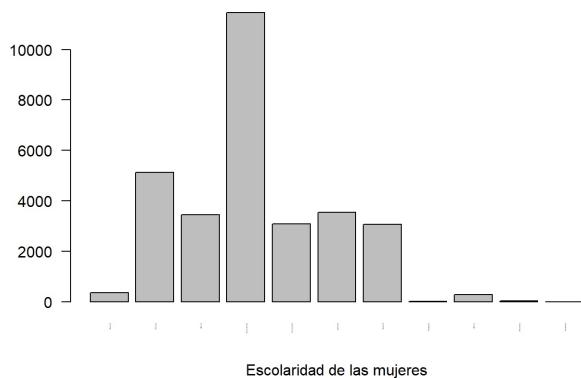


Tabla de frecuencias para el grupo étnico del hombre

			Indigena	Ladino / Mestizo
##	Garifuna	Ignorado		
##	18	6141	182	19355
##	Maya	No indigena	Otro	Xinka
##	3564	820	293	5

Distribución de etnia de los hombres (2012 - 2021)

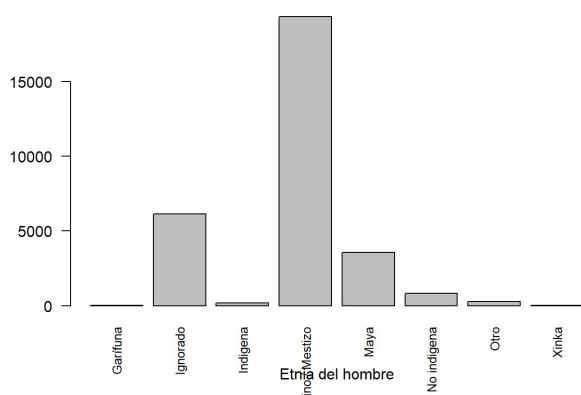
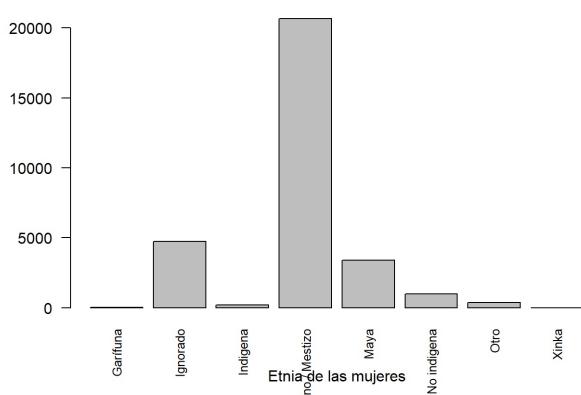


Tabla de frecuencias para el grupo étnico de la mujer

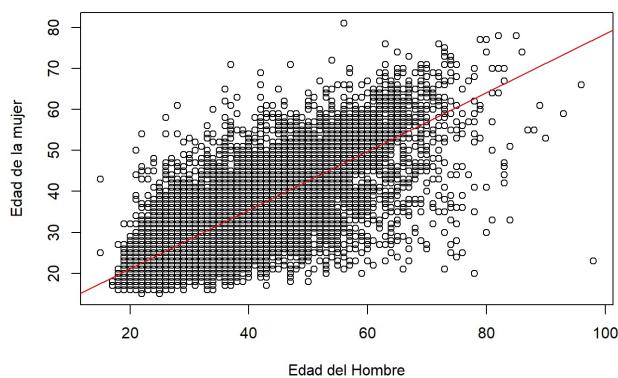
##	Garífuna	Ignorado	Indigena	Ladino / Mestizo
##	21	4740	192	20671
##	Maya	No indigena	Otro	Xinka
##	3401	987	363	3

Distribución de etnia de las mujeres (2012 - 2021)



Correlacion entre edades de los cónyuges

Correlación entre las edades de los cónyuges (2012 - 2021)



Ocupacion en Hombre 2009-2021

ocupacion en Mujeres 2009-2021

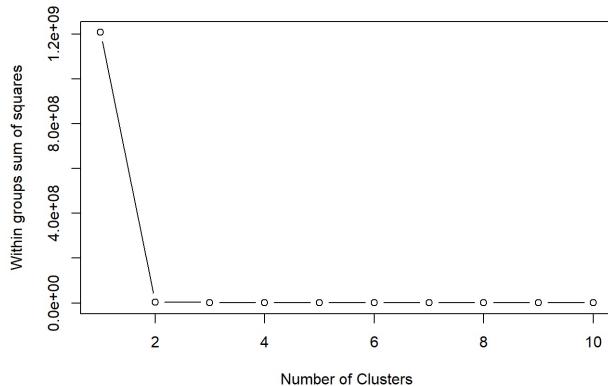
Clustering

Haga un agrupamiento (clustering) e interprete los resultados.

```
divorcios$AÑOOCU <- as.numeric(factor(divorcios$AÑOOCU))
datosClustering <- divorcios[,c("AÑOREG","EDADHOM","EDADMUJ","AÑOOCU")]
data_omit <- na.omit(datosClustering)
summary(datosClustering)
```

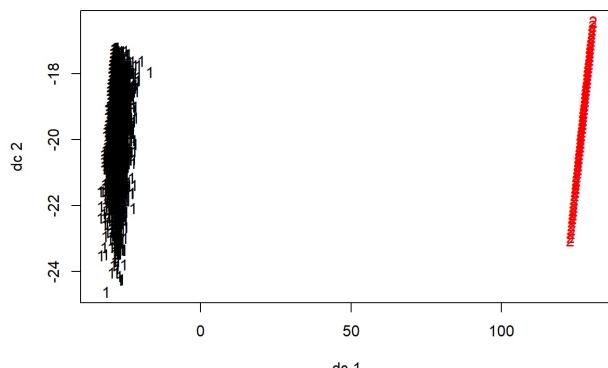
```
##      AÑOREG       EDADHOM       EDMAMUJ      AÑOOCU
## Min. :2012   Min. : 15.0   Min. :15.00   Min. :1.000
## 1st Qu.:2015  1st Qu.: 29.0   1st Qu.:26.00  1st Qu.:3.000
## Median :2018   Median : 34.0   Median :31.00  Median :5.000
## Mean   :2018   Mean   :106.8   Mean  :32.49   Mean  :4.482
## 3rd Qu.:2020  3rd Qu.: 42.0   3rd Qu.:37.00  3rd Qu.:7.000
## Max.  :2022   Max.  :999.0   Max. :81.00   Max. :7.000
##                         NA's :6627
```

```
## Warning: did not converge in 10 iterations
```

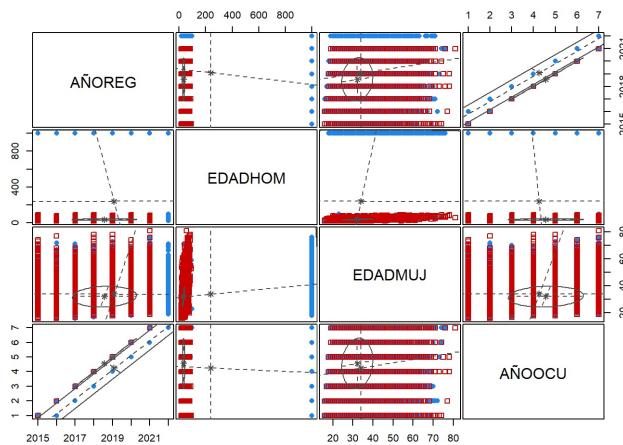


Edad mujeres y hombres en divorcio entre 2009-2021 por medio de k-means

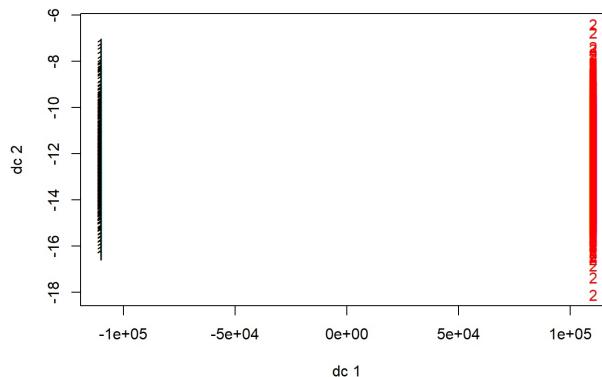
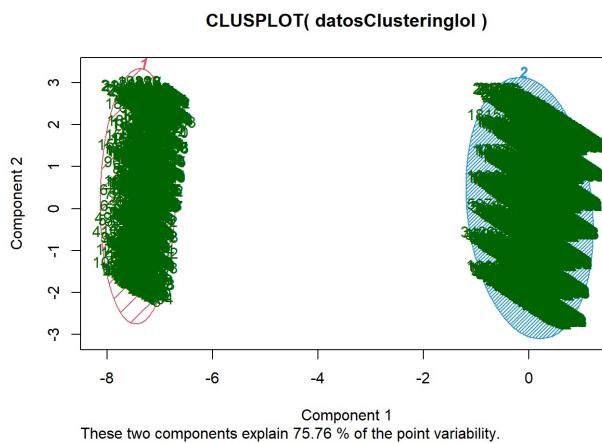
Agrupamiento por medio de k-means



Mixture of gaussians



Fuzzy C-means



```
##   AÑOREG     EDADHOM     EDADMUJ     AÑOOCU      grupo
## Min. :2015  Min. :18.00  Min. :16.00  Min. :1.000  Min. :1
## 1st Qu.:2017 1st Qu.:28.00 1st Qu.:26.00 1st Qu.:3.000 1st Qu.:1
## Median :2019 Median :33.00  Median :30.00  Median :5.000  Median :1
## Mean   :2019 Mean  :35.34  Mean  :32.17  Mean  :4.516  Mean  :1
## 3rd Qu.:2021 3rd Qu.:40.00 3rd Qu.:37.00 3rd Qu.:7.000 3rd Qu.:1
## Max.  :2022 Max. :98.00  Max. :81.00  Max. :7.000  Max. :1
```

```
##   AÑOREG     EDADHOM     EDADMUJ     AÑOOCU      grupo
## Min. :2015  Min. :999   Min. :18.00  Min. :1.000  Min. :2
## 1st Qu.:2016 1st Qu.:999   1st Qu.:34.00 1st Qu.:2.000 1st Qu.:2
## Median :2018 Median :999   Median :40.00  Median :4.000  Median :2
## Mean   :2018 Mean  :999   Mean  :40.99  Mean  :3.921  Mean  :2
## 3rd Qu.:2020 3rd Qu.:999   3rd Qu.:46.75 3rd Qu.:6.000 3rd Qu.:2
## Max.  :2022 Max. :999   Max. :76.00  Max. :7.000  Max. :2
```

```
##   AÑOREG     EDADHOM     EDADMUJ     AÑOOCU      grupo
## Min. :2015  Min. :18.00  Min. :16.00  Min. :1.000  Min. :1
## 1st Qu.:2017 1st Qu.:30.0  1st Qu.:27.00 1st Qu.:3.000 1st Qu.:1
## Median :2019 Median :36.0  Median :32.00  Median :4.000  Median :1
## Mean   :2019 Mean  :239.6  Mean  :34.01  Mean  :4.252  Mean  :1.212
## 3rd Qu.:2021 3rd Qu.:56.0  3rd Qu.:39.00 3rd Qu.:6.000 3rd Qu.:1.000
## Max.  :2022 Max. :999.0  Max. :76.00  Max. :7.000  Max. :2.000
## mxGau
## Min. :1
## 1st Qu.:1
## Median :1
## Mean   :1
## 3rd Qu.:1
## Max.  :1
```

```

##   AÑOREG     EDADHOM     EDADMUJ     AÑOOCU      grupo
## Min. :2015 Min. :18.00 Min. :16.00 Min. :1.000 Min. :1
## 1st Qu.:2017 1st Qu.:28.00 1st Qu.:26.00 1st Qu.:3.000 1st Qu.:1
## Median :2019 Median :33.00 Median :30.00 Median :5.000 Median :1
## Mean   :2019 Mean   :35.32 Mean   :32.18 Mean   :4.566 Mean   :1
## 3rd Qu.:2021 3rd Qu.:40.00 3rd Qu.:37.00 3rd Qu.:7.000 3rd Qu.:1
## Max.  :2021 Max.  :90.00 Max.  :81.00 Max.  :7.000 Max.  :1
##       mxGau
##       Min. :2
## 1st Qu.:2
## Median :2
## Mean   :2
## 3rd Qu.:2
## Max.  :2

```

```

##   AÑOREG     EDADHOM     EDADMUJ     AÑOOCU      grupo
## Min. :2015 Min. :999 Min. :18.00 Min. :1.000 Min. :2
## 1st Qu.:2016 1st Qu.:999 1st Qu.:34.00 1st Qu.:2.000 1st Qu.:2
## Median :2018 Median :999 Median :40.00 Median :4.000 Median :2
## Mean   :2018 Mean   :999 Mean   :40.99 Mean   :3.921 Mean   :2
## 3rd Qu.:2020 3rd Qu.:999 3rd Qu.:46.75 3rd Qu.:6.000 3rd Qu.:2
## Max.  :2022 Max.  :999 Max.  :76.00 Max.  :7.000 Max.  :2
##       mxGau
##       FCGrupos
##       Min. :1 Min. :1
## 1st Qu.:1 1st Qu.:1
## Median :1 Median :1
## Mean   :1 Mean   :1
## 3rd Qu.:1 3rd Qu.:1
## Max.  :1 Max.  :1

```

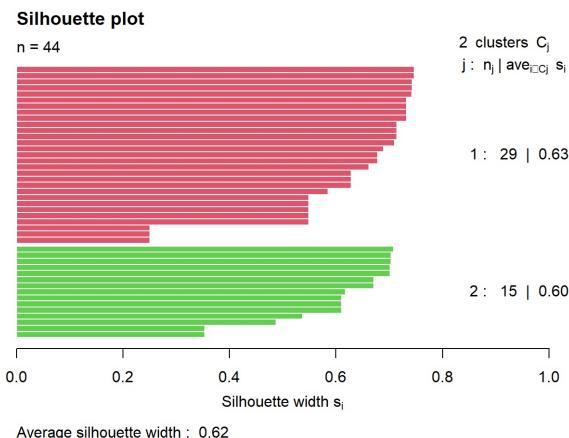
```

##   AÑOREG     EDADHOM     EDADMUJ     AÑOOCU      grupo
## Min. :2015 Min. :18.00 Min. :16.00 Min. :1.000 Min. :1
## 1st Qu.:2017 1st Qu.:28.00 1st Qu.:26.00 1st Qu.:3.000 1st Qu.:1
## Median :2019 Median :33.00 Median :30.00 Median :5.000 Median :1
## Mean   :2019 Mean   :35.34 Mean   :32.17 Mean   :4.516 Mean   :1
## 3rd Qu.:2021 3rd Qu.:40.00 3rd Qu.:37.00 3rd Qu.:7.000 3rd Qu.:1
## Max.  :2022 Max.  :98.00 Max.  :81.00 Max.  :7.000 Max.  :1
##       mxGau
##       FCGrupos
##       Min. :1.000 Min. :2
## 1st Qu.:2.000 1st Qu.:2
## Median :2.000 Median :2
## Mean   :1.778 Mean   :2
## 3rd Qu.:2.000 3rd Qu.:2
## Max.  :2.000 Max.  :2

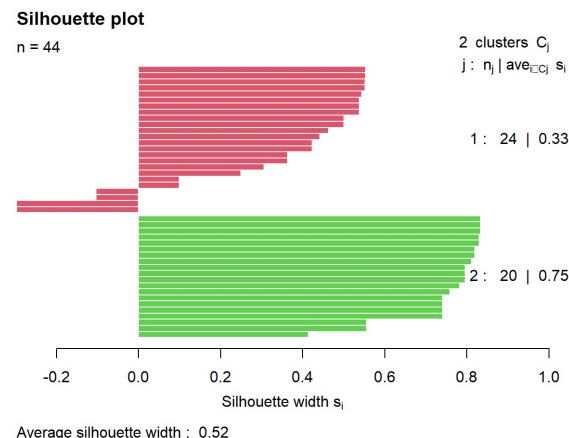
```

Evaluación de la tendencia de agrupamiento:

En hombres



En mujeres



Predicciones

divorciosLabels

```

# str(divorciosLabels)

# divorciosLabels$EDADMUJ <- as.numeric(as.character(divorciosLabels$EDADMUJ)) # Convierte a numérico
# divorciosLabels$clasification <- ifelse(divorciosLabels$EDADMUJ <= 17, "Menor", ifelse(divorciosLabels$EDADMUJ
>= 18, "Mayor", NA))
#
# print(unique(divorciosLabels$clasification)) # Imprime los valores únicos de la columna "clasification"
# divorciosLabels$y <- factor(divorciosLabels$clasification)
#
# str(divorciosLabels) # Verifica la estructura de los datos nuevamente

```

```

# divorciosLabels$EDADHOM <- as.numeric(as.character(divorciosLabels$EDADHOM))
# divorciosLabels$EDADMUJ <- as.numeric(as.character(divorciosLabels$EDADMUJ))
#
# datosMDD <- divorciosLabels %>% filter(EDADMUJ >= 12)
# datosMDD <- datosMDD %>% filter(EDADHOM >= 14)
# datosMDD <- datosMDD[,-20]
# datosMDD <- datosMDD[,-10]
# datosMDD <- datosMDD[complete.cases(datosMDD), ]
#
# str(datosMDD)
#
# datosMDD$DEPREG <- as.factor(datosMDD$DEPREG)
# datosMDD$MUPREG <- as.factor(datosMDD$MUPREG)
# datosMDD$MESREG <- as.factor(datosMDD$MESREG)
# datosMDD$ANOREG <- as.factor(datosMDD$ANOREG)
# datosMDD$DIAOCU <- as.factor(datosMDD$DIAOCU)
# datosMDD$MESOCU <- as.factor(datosMDD$MESOCU)
# datosMDD$DEPOCU <- as.factor(datosMDD$DEPOCU)
# datosMDD$MUPOCU <- as.factor(datosMDD$MUPOCU)
# datosMDD$EDADHOM <- as.factor(datosMDD$EDADHOM)
# datosMDD$PERHOM <- as.factor(datosMDD$PERHOM)
# datosMDD$PERMUJ <- as.factor(datosMDD$PERMUJ)
# datosMDD$NACHOM <- as.factor(datosMDD$NACHOM)
# datosMDD$NACMUJ <- as.factor(datosMDD$NACMUJ)
# datosMDD$ESCHOM <- as.factor(datosMDD$ESCHOM)
# datosMDD$ESCMUJ <- as.factor(datosMDD$ESCMUJ)
# datosMDD$CIUOHOM <- as.factor(datosMDD$CIUOHOM)
# datosMDD$CIUOMUJ <- as.factor(datosMDD$CIUOMUJ)
# datosMDD$ANOCU <- as.factor(datosMDD$ANOCU)
#
# str(datosMDD)

```

```

# datosMDD <- datosMDD[,c(5,6,7,8,9,10,15,16,17,18,19)]
# porciento <- 70/100
# set.seed(123)
# trainRowsNumber<-sample(nrow(datosMDD),porciento*nrow(datosMDD))
# train<-datosMDD[trainRowsNumber,]
# test<-datosMDD[-trainRowsNumber,]
#
# str(datosMDD)

```

```
divorcios <- subset(divorcios, EDAD == 18)
divorcios <- subset(divorcios, EDAD == 19)
str(divorcios)
```

```

## tibble [28,120 x 19] (S3:tbl_df/tbl/data.frame)
## $ DEPREG : dbl+lbl [1:28120] 12, 1, 14, 1, 1, 1, 1, 1, 1, 9, 3, 5, 16, ...
##   ..@ labels: Named num [1:22] 1 2 3 4 5 6 7 8 9 10 ...
##   ..@ attr(*, "names")= chr [1:22] "Guatemala" "El Progreso" "Sacatepequez" "Chimaltenango" ...
##   ..@ label : chr "Departamento de registro"
## $ MUPREG : chr+lbl [1:28120] 1213, 0101, 1416, 0101, 0101, 0101, 0101, 0101, 0101...
##   ..@ labels: Named chr [1:342] "1010" "2010" "0110" "1210" ...
##   ..@ attr(*, "names")= chr [1:342] "San Antonio Suchitepéquez" "San Jacinto" "San Juan Sacatépéquez" "Tejutla" ...
##   ..@ label : chr "Municipio de registro"
## $ MESREG : dbl+lbl [1:28120] 5, 4, 6, 10, 2, 8, 11, 9, 8, 4, 4, 10, 4, ...
##   ..@ labels: Named num [1:12] 1 2 3 4 5 6 7 8 9 10 ...
##   ..@ attr(*, "names")= chr [1:12] "Enero" "Febrero" "Marzo" "Abril" ...
##   ..@ label : chr "Mes de registro"
## $ ANOREG : num [1:28120] 2012 2012 2012 2012 2012 ...
## $ DIAOCU : num [1:28120] 3 27 28 12 9 11 3 29 16 30 ...
## $ MESOCU : dbl+lbl [1:28120] 2, 3, 5, 3, 1, 6, 8, 2, 5, 5, 4, 7, 9, ...
##   ..@ label : chr "Mes de ocurrencia"
##   ..@ format.spss: chr "F2.0"
##   ..@ labels : Named num [1:12] 1 2 3 4 5 6 7 8 9 10 ...
##   ..@ attr(*, "names")= chr [1:12] "Enero" "Febrero" "Marzo" "Abril" ...
## $ DEPOCU : dbl+lbl [1:28120] 12, 1, 14, 1, 1, 1, 1, 1, 1, 9, 1, 5, 16, ...
##   ..@ labels: Named num [1:22] 1 2 3 4 5 6 7 8 9 10 ...
##   ..@ attr(*, "names")= chr [1:22] "Guatemala" "El Progreso" "Sacatepequez" "Chimaltenango" ...
##   ..@ label : chr "Departamento de ocurrencia"
## $ MUPOCU : chr+lbl [1:28120] 1215, 0101, 1401, 0101, 0101, 0101, 0101, 0101, 0101...
##   ..@ labels: Named chr [1:342] "1010" "2010" "0110" "1210" ...
##   ..@ attr(*, "names")= chr [1:342] "San Antonio Suchitepéquez" "San Jacinto" "San Juan Sacatépéquez" "Tejutla" ...
##   ..@ label : chr "Municipio de ocurrencia"
## $ EDADHOM: dbl+lbl [1:28120] 35, 33, 31, 27, 37, 41, 36, 46, 38, 35, 36, 27, 29, ...
##   ..@ label : chr "Edad del hombre"
##   ..@ format.spss: chr "F4.0"
##   ..@ labels : Named num 999
##   ..@ attr(*, "names")= chr "Ignorado"
## $ EDADMUJ: dbl+lbl [1:28120] 30, 32, 28, 29, 30, 42, 30, 37, 35, 35, 38, 27, 26, ...
##   ..@ label : chr "Edad de la mujer"
##   ..@ format.spss: chr "F4.0"
##   ..@ labels : Named num 999
##   ..@ attr(*, "names")= chr "Ignorado"
## $ PPERHOM: dbl+lbl [1:28120] 2, 9, 2, 9, 2, 2, 2, 1, 2, 2, 1, 2, 2, 9, 2, 9, ...
##   ..@ labels: Named num [1:6] 1 2 9 3 4 5
##   ..@ attr(*, "names")= chr [1:6] "Indigena" "No indigena" "Ignorado" "Xinca" ...
##   ..@ label : chr "Grupo Etnico del hombre"
## $ PPERMUJ: dbl+lbl [1:28120] 2, 2, 2, 9, 9, 2, 2, 2, 2, 2, 2, 2, 2, 9, 2, 2...
##   ..@ labels: Named num [1:6] 1 2 9 3 4 5
##   ..@ attr(*, "names")= chr [1:6] "Indigena" "No indigena" "Ignorado" "Xinca" ...
##   ..@ label : chr "Grupo Etnico de la mujer"
## $ NACHOM: dbl+lbl [1:28120] 320, 320, 320, 320, 320, 320, 320, 320, 320, 320, 32...
##   ..@ labels: Named num [1:105] 32 56 68 84 124 156 170 188 192 222 ...
##   ..@ attr(*, "names")= chr [1:105] "Argentina" "Bélgica" "Bolivia" "Belice" ...
##   ..@ label : chr "Nacionalidad del hombre"
## $ NACMUJ: dbl+lbl [1:28120] 320, 320, 320, 320, 320, 320, 320, 320, 320, 320, 32...
##   ..@ labels: Named num [1:103] 76 84 170 188 192 218 222 276 320 340 ...
##   ..@ attr(*, "names")= chr [1:103] "Brasil" "Belice" "Colombia" "Costa Rica" ...
##   ..@ label : chr "Nacionalidad de la mujer"
## $ ESCHOM: dbl+lbl [1:28120] 5, 1, 4, 9, 5, 5, 5, 9, 4, 3, 5, 9, 9, 3, 9, 4, 1...
##   ..@ labels: Named num [1:8] 1 2 3 4 5 9 6 0
##   ..@ attr(*, "names")= chr [1:8] "Ninguna" "Primaria" "Básico" "Diversificado" ...
##   ..@ label : chr "Escolaridad del hombre"
## $ ESCMUJ: dbl+lbl [1:28120] 9, 5, 4, 5, 9, 5, 5, 5, 4, 4, 5, 4, 5, 5, 9, 4, 4...
##   ..@ labels: Named num [1:8] 1 2 3 4 5 9 6 0
##   ..@ attr(*, "names")= chr [1:8] "Ninguna" "Primaria" "Básico" "Diversificado" ...
##   ..@ label : chr "Escolaridad de la mujer"
## $ CIUOHOM: chr [1:28120] "110" "2142" "8189" "1346" ...
## $ CIUOMUJ: chr [1:28120] "1120" "1120" "1120" "1212" ...
## $ AÑOOCU : num [1:28120] NA ...
## 
```

```

divorcios$EDADHOM <- as.numeric(as.character(divorcios$EDADHOM))
divorcios$EDADMUJ <- as.numeric(as.character(divorcios$EDADMUJ))

datosMD <- divorcios %>% filter(EDADMUJ >= 12)
datosMD <- datosMD %>% filter(EDADHOM >= 14)
datosMD <- datosMD[,20]
datosMD <- datosMD[,10]
datosMD <- datosMD[complete.cases(datosMD), ]

datosMD$DEPREG <- as.factor(datosMD$DEPREG)
datosMD$MUPREG <- as.factor(datosMD$MUPREG)
datosMD$MESREG <- as.factor(datosMD$MESREG)
datosMD$ANOREG <- as.factor(datosMD$ANOREG)
datosMD$DIAOCU <- as.factor(datosMD$DIAOCU)
datosMD$MESOCU <- as.factor(datosMD$MESOCU)
datosMD$DEPOCU <- as.factor(datosMD$DEPOCU)
datosMD$MUPOCU <- as.factor(datosMD$MUPOCU)
datosMD$EDADHOM <- as.factor(datosMD$EDADHOM)
datosMD$PPERHOM <- as.factor(datosMD$PPERHOM)
datosMD$PERMUJ <- as.factor(datosMD$PERMUJ)
datosMD$NACHOM <- as.factor(datosMD$NACHOM)
datosMD$NACMUJ <- as.factor(datosMD$NACMUJ)
datosMD$ESCHOM <- as.factor(datosMD$ESCHOM)
datosMD$ESCMUJ <- as.factor(datosMD$ESCMUJ)
datosMD$CIUOHOM <- as.factor(datosMD$CIUOHOM)
datosMD$CIUOMUJ <- as.factor(datosMD$CIUOMUJ)
datosMD$AÑOOCU <- as.factor(datosMD$AÑOOCU)

datosMD <- datosMD[,c(5,6,7,8,9,10,15,16,17,18,19)]
porciento <- 70/100
set.seed(123)
trainRowsNumber<-sample(nrow(datosMD),porciento*nrow(datosMD))
train<-datosMD[trainRowsNumber,]
test<-datosMD[-trainRowsNumber,]

str(datosMD)

## # tibble [22,977 x 11] (S3:tbl_df/tbl/data.frame)
## $ DIAOCU : Factor w/ 31 levels "1","2","3","4",...: 8 24 23 4 6 24 25 3 17 18 ...
## $ MESOCU : Factor w/ 12 levels "1","2","3","4",...: 7 6 6 2 4 7 6 8 7 2 ...
## $ DEPOCU : Factor w/ 22 levels "1","2","3","4",...: 19 1 12 15 11 9 1 1 16 15 ...
## $ MUPOCU : Factor w/ 338 levels "0101","0102",...: 291 1 180 250 156 110 11 1 263 247 ...
## $ EDADHOM: Factor w/ 74 levels "18","19","20",...: 29 23 33 11 4 8 34 20 35 24 ...
## $ PPERHOM: Factor w/ 6 levels "1","2","3","4",...: 4 6 6 6 6 4 6 6 1 4 ...
## $ ESCMUJ : Factor w/ 8 levels "0","1","2","3",...: 8 6 8 5 5 5 5 4 4 ...
## $ CIUOHOM: Factor w/ 45 levels "01","03","1",...: 45 45 45 45 45 45 45 45 45 45 ...
## $ CIUOMUJ: Factor w/ 42 levels "01","03","1",...: 41 8 42 41 41 41 41 41 42 ...
## $ AÑOOCU : Factor w/ 7 levels "1","2","3","4",...: 1 1 1 1 1 1 1 ...
## $ y : Factor w/ 2 levels "Mayor","Menor": 1 1 1 1 1 1 1 1 1 1 ...

```

Eliminando para liberar memoria

```

rm(clarax)
rm(clorax)
rm(D2009)
rm(D2010)
rm(D2011)
rm(D2012)
rm(D2013)
rm(D2014)
rm(D2015)
rm(D2016)
rm(D2017)
rm(D2018)
rm(D2019)
rm(D2020)
rm(D2021)
rm(data.omit)
rm(datasets)
rm(datosclustering)
rm(datosClusteringlol)
rm(df0)
rm(dnac)
rm(fcmf)
rm(g1FCM)
rm(g1PMC)
rm(g2FCM)
rm(g2PMC)
rm(gkm1)
rm(gkm2)
rm(km)
rm(M2009)
rm(M2010)
rm(M2011)
rm(M2012)
rm(M2013)
rm(M2014)
rm(M2015)
rm(M2016)
rm(M2017)
rm(M2018)
rm(M2019)
rm(M2020)
rm(M2021)
rm(mc)
rm(N2009)
rm(N2010)
rm(N2011)
rm(N2012)
rm(N2013)
rm(N2014)
rm(N2015)
rm(N2016)
rm(N2017)
rm(N2018)
rm(N2019)
rm(N2020)
rm(N2021)

```

Algoritmos

Predicción por medio de Naive Bayes

```

withCallingHandlers({
  suppressPackageStartupMessages(library(caret))
  suppressWarnings({
    library(RWeka)
    modelo <- naiveBayes(train$y ~ ., data = train)
    predBayes <- predict(modelo, newdata = train[, 1:11])
  })
}, warning = function(w) {
  if (grepl("package 'RWeka' was built under R version", w$message))
    return(NULL)
  else
    invokeRestart("muffleWarning")
})

cm<- caret::confusionMatrix(predBayes,train$y)
cm

## Confusion Matrix and Statistics
##
##             Reference
## Prediction Mayor Menor
##     Mayor 15894    7
##     Menor   150   32
##
##                   Accuracy : 0.9902
##                   95% CI : (0.9886, 0.9917)
## No Information Rate : 0.9976
## P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.2867
##
## McNemar's Test P-Value : <2e-16
##
##                   Sensitivity : 0.9907
##                   Specificity : 0.8285
## Pos Pred Value : 0.9996
## Neg Pred Value : 0.1758
## Prevalence : 0.9976
## Detection Rate : 0.9882
## Detection Prevalence : 0.9887
## Balanced Accuracy : 0.9056
##
## 'Positive' Class : Mayor
##
```

```

predBayesM <- predict(modelo, newdata = test[,1:11])
test$PredBayes <- predBayesM
cm <- caret::confusionMatrix(test$PredBayes, test$y)
cm

## Confusion Matrix and Statistics
##
##             Reference
## Prediction Mayor Menor
##     Mayor 6818    6
##     Menor   65    5
##
##                   Accuracy : 0.9897
##                   95% CI : (0.987, 0.9919)
## No Information Rate : 0.9984
## P-Value [Acc > NIR] : 1
##
##                   Kappa : 0.121
##
## McNemar's Test P-Value : 5.847e-12
##
##                   Sensitivity : 0.99056
##                   Specificity : 0.45455
## Pos Pred Value : 0.99912
## Neg Pred Value : 0.07143
## Prevalence : 0.99840
## Detection Rate : 0.98898
## Detection Prevalence : 0.98985
## Balanced Accuracy : 0.72255
##
## 'Positive' Class : Mayor
##
```

Prediccion por redes neuronales: RWEKA

```

#datosML <- lapply(datosMD, as.factor)
#install.packages("RWeka")
library (caret)
library(RWeka)
NB <- make_Weka_classifier("weka/classifiers/functions/MultilayerPerceptron")
NB
WOW(NB)
nnodos='11'
modelo.bp<-NB(datosMD$y~., data=datosMD,subset = trainRowsNumber, control=Weka_control(H=nnodos, N=1, G=TRUE), options=NULL)
test$prediccionWeka<-predict(modelo.bp, newdata = test[,1:11])

cfmWeka<-confusionMatrix(test$prediccionWeka,test$y)
cfmWeka

```

```

## Confusion Matrix and Statistics
##
##             Reference
## Prediction Mayor Menor
##     Mayor      0      0
##     Menor   6883    11
##
##           Accuracy : 0.0016
##                 95% CI : (8e-04, 0.0029)
## No Information Rate : 0.9984
## P-Value [Acc > NIR] : 1
##
##           Kappa : 0
##
## McNemar's Test P-Value : <2e-16
##
##           Sensitivity : 0.000000
##           Specificity : 1.000000
## Pos Pred Value :      NaN
## Neg Pred Value : 0.001596
##          Prevalence : 0.998404
## Detection Rate : 0.000000
## Detection Prevalence : 0.000000
## Balanced Accuracy : 0.500000
##
## 'Positive' Class : Mayor
##

```

Prediccion por medio de árboles de decisión

```

library(caret)
trainData <- train[, 1:11]
trainLabels <- train$y
ct <- trainControl(method = "cv", number = 10, verboseIter = TRUE)
modelorf <- suppressWarnings(train(x = trainData, y = trainLabels, method = "rpart", trControl = ct))
prediccionADVC <- suppressWarnings(predict(modelorf, newdata = test[, 1:11]))
test$predADVC <- prediccionADVC

```

```
cfmCaret <- confusionMatrix(test$predADVC, test$y)
cfmCaret
```

```

## Confusion Matrix and Statistics
##
##             Reference
## Prediction Mayor Menor
##     Mayor   6883      0
##     Menor      0    11
##
##           Accuracy : 1
##                 95% CI : (0.9995, 1)
## No Information Rate : 0.9984
## P-Value [Acc > NIR] : 1.656e-05
##
##           Kappa : 1
##
## McNemar's Test P-Value : NA
##
##           Sensitivity : 1.0000
##           Specificity : 1.0000
## Pos Pred Value : 1.0000
## Neg Pred Value : 1.0000
##          Prevalence : 0.9984
## Detection Rate : 0.9984
## Detection Prevalence : 0.9984
## Balanced Accuracy : 1.0000
##
## 'Positive' Class : Mayor
##

```

```

library(e1071)

# Crear modelo SVM
model_svm <- svm(y ~ ., data = train)
predictions_svm <- predict(model_svm, newdata = test)
confusionMatrix(predictions_svm, test$y)
```

```

## Confusion Matrix and Statistics
##
##             Reference
## Prediction Mayor Menor
##     Mayor   6883    11
##     Menor      0      0
##
##           Accuracy : 0.9984
##                 95% CI : (0.9971, 0.9992)
## No Information Rate : 0.9984
## P-Value [Acc > NIR] : 0.579267
##
##           Kappa : 0
##
## McNemar's Test P-Value : 0.002569
##
##           Sensitivity : 1.0000
##           Specificity : 0.0000
## Pos Pred Value : 0.9984
## Neg Pred Value :      NaN
##          Prevalence : 0.9984
## Detection Rate : 0.9984
## Detection Prevalence : 1.0000
## Balanced Accuracy : 0.5000
##
## 'Positive' Class : Mayor
##
```

```

library(rpart)

# Crear modelo Árbol de Decisión
model_tree <- rpart(y ~ ., data = train)
predictions_tree <- predict(model_tree, newdata = test, type = "class")
confusionMatrix(predictions_tree, test$y)
```

```
## Confusion Matrix and Statistics
##
##             Reference
## Prediction Mayor Menor
##      Mayor     6878    10
##      Menor      5     1
##
##                  Accuracy : 0.9978
##                  95% CI : (0.9964, 0.9988)
##      No Information Rate : 0.9984
##      P-Value [Acc > NIR] : 0.9876
##
##                  Kappa : 0.1167
##
## McNemar's Test P-Value : 0.3017
##
##                  Sensitivity : 0.99927
##                  Specificity : 0.09091
##      Pos Pred Value : 0.99855
##      Neg Pred Value : 0.16667
##                  Prevalence : 0.99840
##      Detection Rate : 0.99768
## Detection Prevalence : 0.99913
##      Balanced Accuracy : 0.54509
##
##      'Positive' Class : Mayor
##
```