

Proyect

Cristopher Barrios, Elean Rivas, 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

## Warning: package 'ggplot2' was built under R version 4.2.3

## 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.

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 AÑOREG:

cuantitativa discreta DIAOCU: cuantitativa MESOCU: cualitativa ANOOCU: cuantitativa discreta DE-
 POCU: cualitativa MUPOCU: cualitativa EDADHOM: cuantitativa discreta EDADMUJ: cuantitativa disc-
 reta GETHOM: cualitativa GETMUJ: cualitativa NACHOM: cualitativa NACMUJ: cualitativa OCUHOM:
 cualitativa 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 numéri
for (i in 1:length(datasets)) {
  vars_numéricas <- sapply(datasets[[i]], is.numeric)
  print(names(vars_numéricas[vars_numéricas]))
}
```

```
## [1] "Depreg" "Mesreg" "Diaocu" "Mesocu" "Depocu" "Edadhom" "Edadmuj"
## [8] "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" "ocupamuj"
## [1] "depreg" "mesreg" "añoreg" "diaocu" "mesocu" "añooocu"
## [7] "depocu" "edadhom" "edadmuj" "grethom" "gretmuj" "nachom"
## [13] "nacmuj" "escohom" "escomuj" "ocupahom" "ocupamuj"
## [1] "DEPREG" "MESREG" "AÑOREG" "DIAOCU" "MESOCU" "DEPOCU" "EDADHOM"
## [8] "EDADMUJ" "GETHOM" "GETMUJ" "NACHOM" "NACMUJ" "ESCHOM" "ESCMUJ"
## [1] "DEPREG" "MESREG" "AÑOREG" "DIAOCU" "MESOCU" "DEPOCU" "EDADHOM"
## [8] "EDADMUJ" "PUEHOM" "PUEMUJ" "NACHOM" "NACMUJ" "ESCHOM" "ESCMUJ"
## [1] "DEPREG" "MESREG" "AÑOREG" "DIAOCU" "MESOCU" "DEPOCU" "EDADHOM"
## [8] "EDADMUJ" "PUEHOM" "PUEMUJ" "NACHOM" "NACMUJ" "ESCHOM" "ESCMUJ"
## [1] "DEPREG" "MESREG" "AÑOREG" "DIAOCU" "MESOCU" "AÑOOOCU" "DEPOCU"
## [8] "EDADHOM" "EDADMUJ" "PUEHOM" "PUEMUJ" "NACHOM" "NACMUJ" "ESCHOM"
## [15] "ESCMUJ" "CIUOHOM" "CIUOMUJ"
## [1] "DEPREG" "MESREG" "AÑOREG" "DIAOCU" "MESOCU" "AÑOOOCU" "DEPOCU"
## [8] "EDADHOM" "EDADMUJ" "PPERHOM" "PPERMUJ" "NACHOM" "NACMUJ" "ESCHOM"
## [15] "ESCMUJ" "CIUOHOM" "CIUOMUJ"
## [1] "DEPREG" "MESREG" "AÑOREG" "DIAOCU" "MESOCU" "AÑOOOCU" "DEPOCU"
## [8] "EDADHOM" "EDADMUJ" "PPERHOM" "PPERMUJ" "NACHOM" "NACMUJ" "ESCHOM"
## [15] "ESCMUJ" "CIUOHOM" "CIUOMUJ"
## [1] "DEPREG" "MESREG" "AÑOREG" "DIAOCU" "MESOCU" "AÑOOOCU" "DEPOCU"
## [8] "EDADHOM" "EDADMUJ" "PPERHOM" "PPERMUJ" "NACHOM" "NACMUJ" "ESCHOM"
## [15] "ESCMUJ" "CIUOHOM" "CIUOMUJ"
## [1] "DEPREG" "MESREG" "AÑOREG" "DIAOCU" "MESOCU" "AÑOOOCU" "DEPOCU"
## [8] "EDADHOM" "EDADMUJ" "PPERHOM" "PPERMUJ" "NACHOM" "NACMUJ" "ESCHOM"
## [15] "ESCMUJ" "CIUOHOM" "CIUOMUJ"
## [1] "DEPREG" "MESREG" "AÑOREG" "DIAOCU" "MESOCU" "AÑOOOCU" "DEPOCU"
## [8] "EDADHOM" "EDADMUJ" "PPERHOM" "PPERMUJ" "NACHOM" "NACMUJ" "ESCHOM"
## [15] "ESCMUJ" "CIUOHOM" "CIUOMUJ"
```

```
## [1] "DEPREG" "MESREG" "AÑOREG" "DIAOCU" "MESOCU" "AÑOOCU" "DEPOCU"
## [8] "EDADHOM" "EDADMUJ" "PPERHOM" "PPERMUJ" "NACHOM" "NACMUJ" "ESCHOM"
## [15] "ESCMUJ"
```

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

NUNUHO: “Número de nupcias del hombre” NUNUMU: “Número de nupcias de la mujer”

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

```
## # A tibble: 87,480 x 2
##   NUNUHO      NUNUMU
##   <dbl+lbl> <dbl+lbl>
## 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]
## # ... with 87,470 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: 9,621 x 2
##   EDADHOM      EDADMUJ
##   <dbl+lbl> <dbl+lbl>
## 1 29          25
## 2 36          37
## 3 34          31
## 4 48          33
## 5 44          28
## 6 39          27
## 7 45          40
## 8 999 [Ignorado] 999 [Ignorado]
## 9 36          31
## 10 999 [Ignorado] 999 [Ignorado]
## # ... with 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

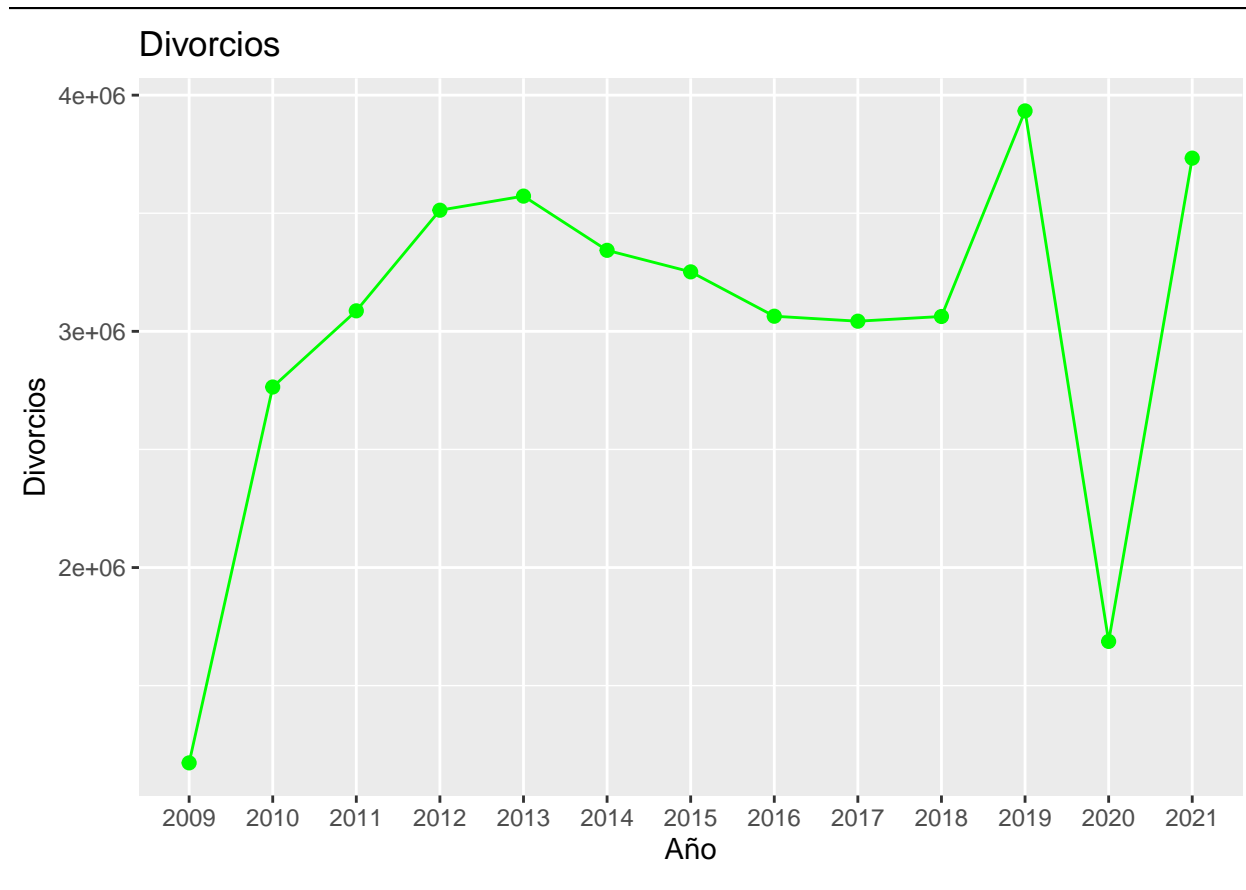
Genero: Es interesante ver que genero es más propenso a los divorcios, esto también puede estar relacionado con la cantidad de nupcias de una persona

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
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)
r dfD <- data.frame (año = c("2009", "2010", "2011", "2012", "2013", "2014", "2015",
"2016", "2017","2018","2019","2020","2021"), matrimonios = c(EdadMujD2009, EdadMujD2010,
EdadMujD2011, EdadMujD2012, EdadMujD2013, EdadMujD2014, EdadMujD2015, EdadMujD2016,
EdadMujD2017, EdadMujD2018, EdadMujD2019, EdadMujD2020, EdadMujD2021)) print(dfD)
##      año matrimonios ## 1  2009      1173016 ## 2  2010      2764309 ## 3  2011
3086834 ## 4  2012      3512781 ## 5  2013      3572498 ## 6  2014      3342987 ## 7  2015
3251831 ## 8  2016      3064039 ## 9  2017      3042894 ## 10 2018      3062850 ## 11 2019
3933259 ## 12 2020      1687235 ## 13 2021      3733294
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 =
"Divorcios")
```



Histogramas

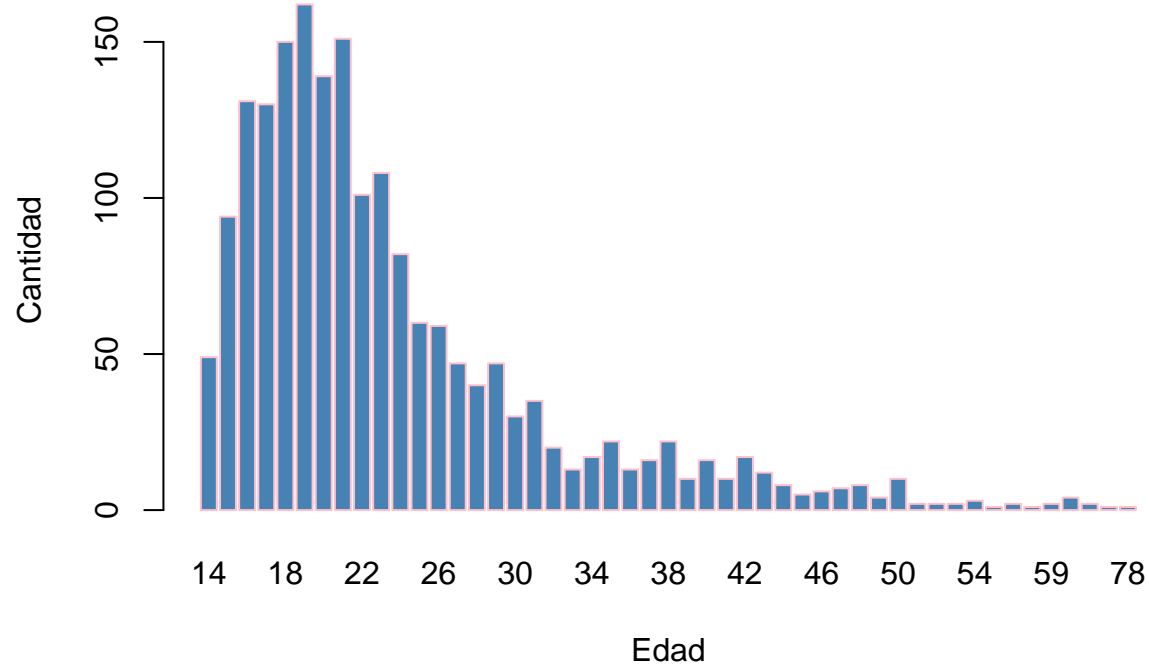
2009

```
library(dplyr)
library(ggplot2)

D2009 <- subset(D2009, Edadmuj < 999)

barplot(table(D2009$Edadmuj), main = "Edad de la mujer en divorcios 2009", xlab = "Edad", ylab = "Canti
```

Edad de la mujer en divorcios 2009

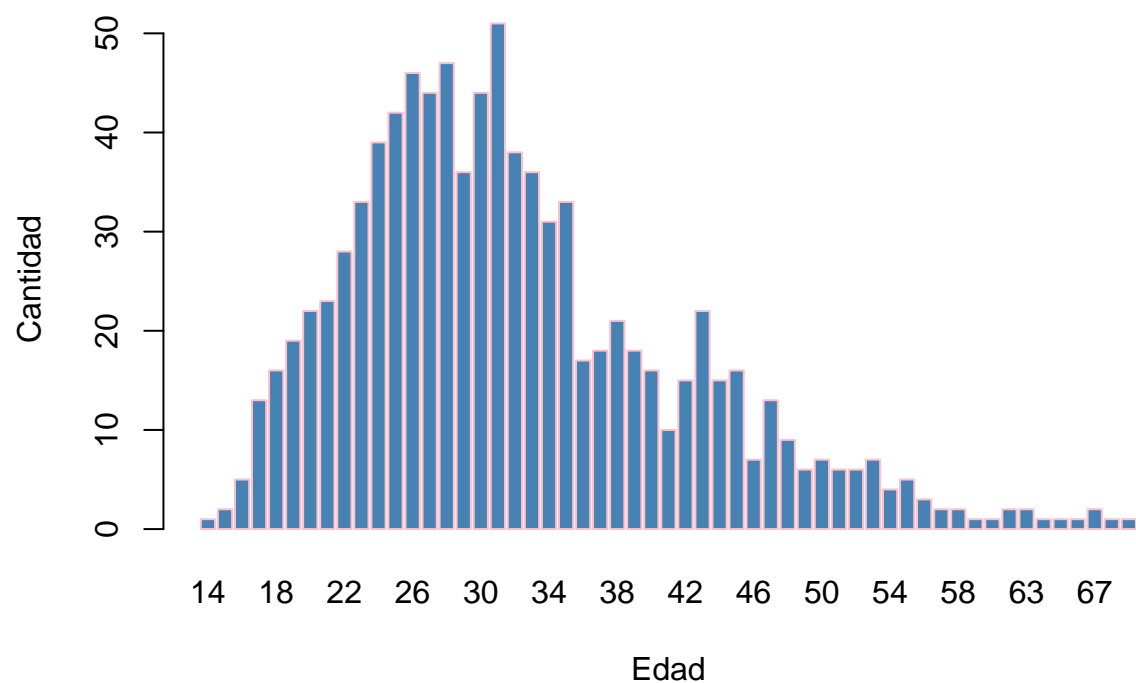


2010

```
D2010 <- subset(D2010, edadmuj < 999)
```

```
barplot(table(D2010$edadmuj), main = "Edad de la mujer en divorcios 2010", xlab = "Edad", ylab = "Cantidad")
```

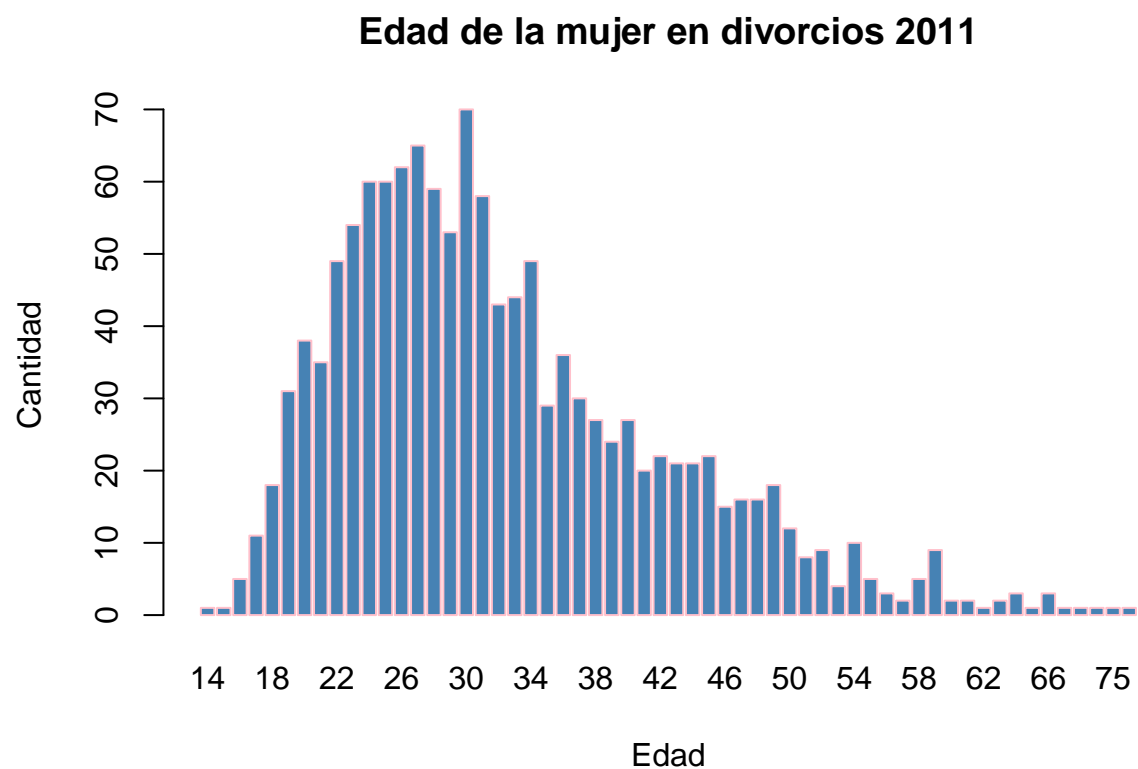
Edad de la mujer en divorcios 2010



2011

```
D2011 <- subset(D2011, edadmuj < 999)
```

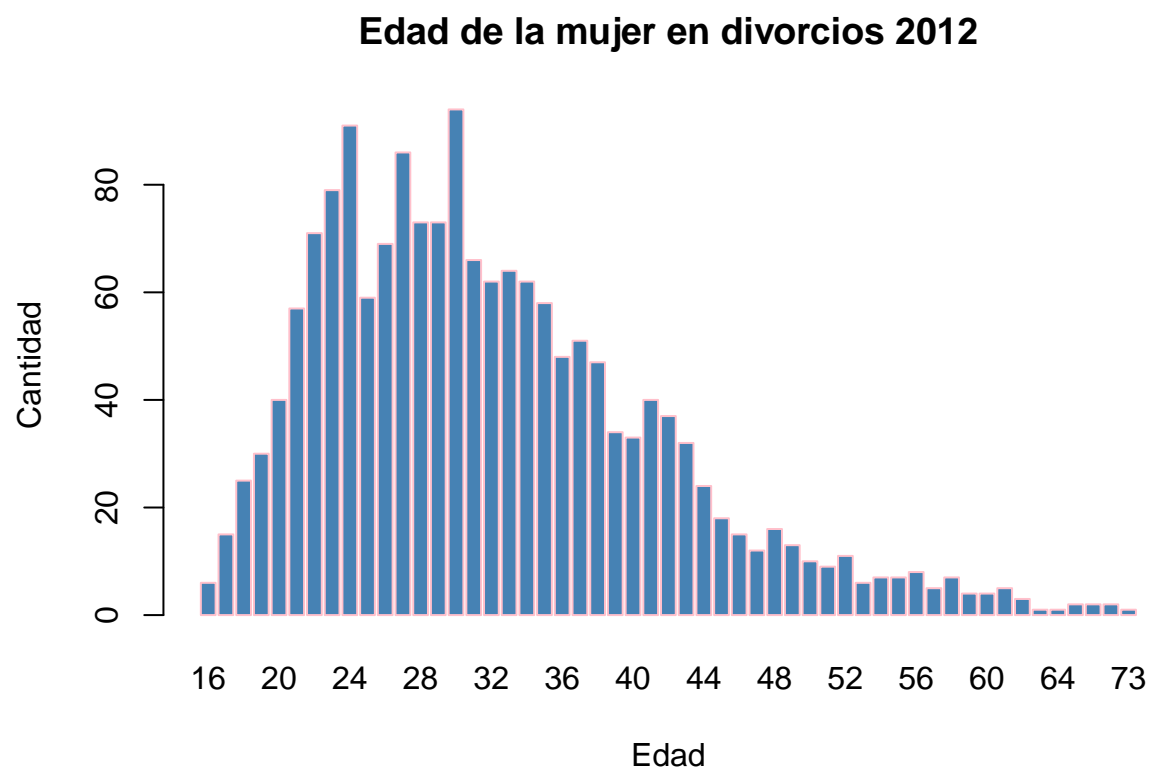
```
barplot(table(D2011$edadmuj), main = "Edad de la mujer en divorcios 2011", xlab = "Edad", ylab = "Cantidad")
```

2012

```
D2012 <- subset(D2012, EDADMUJ < 999)
```

```
barplot(table(D2012$EDADMUJ), main = "Edad de la mujer en divorcios 2012", xlab = "Edad", ylab = "Cantidad")
```

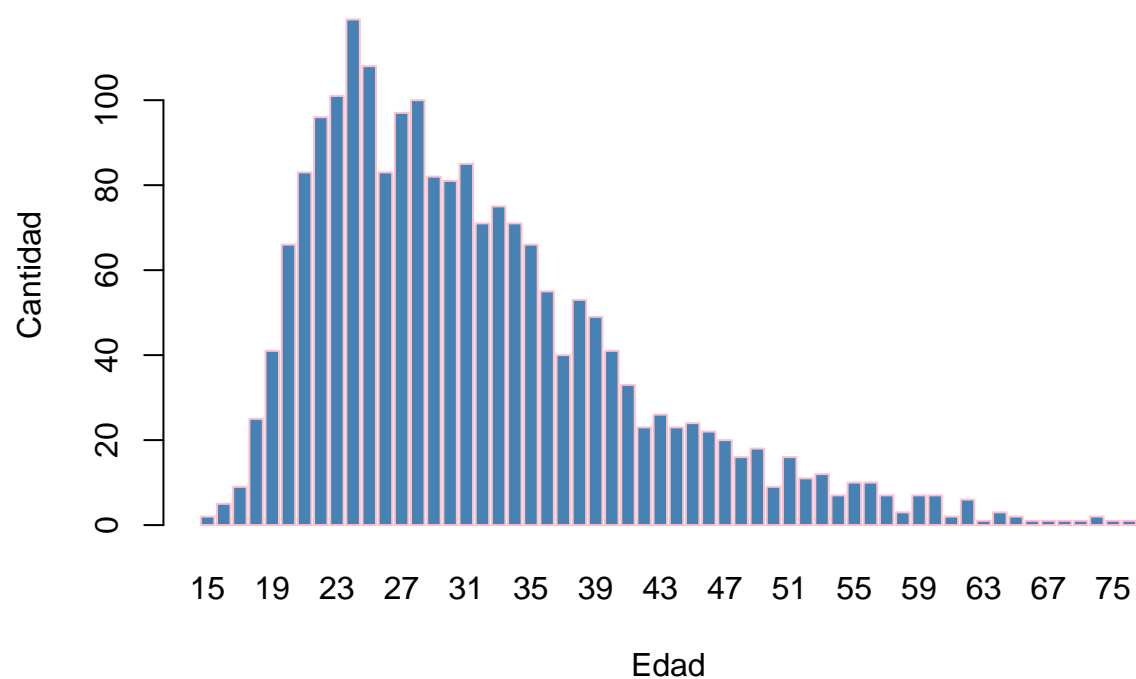


2013

```
D2013 <- subset(D2013, EDADMUJ < 999)
```

```
barplot(table(D2013$EDADMUJ), main = "Edad de la mujer en divorcios 2013", xlab = "Edad", ylab = "Cantidad")
```

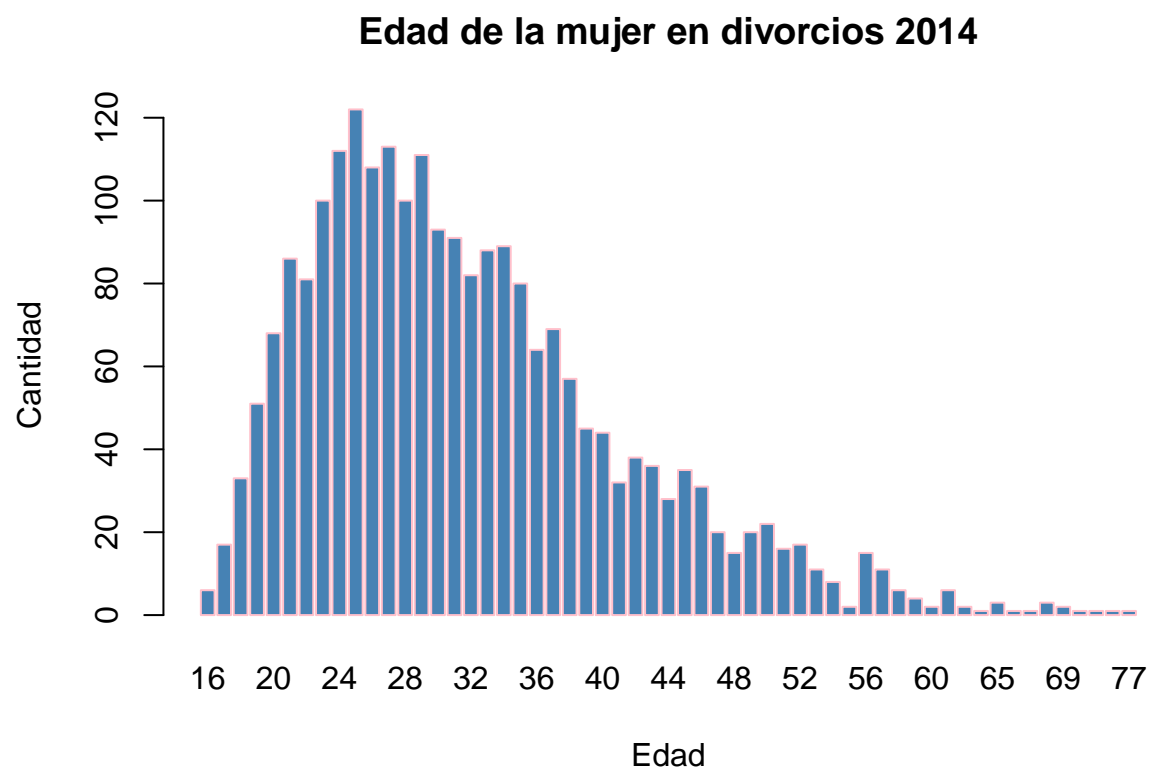
Edad de la mujer en divorcios 2013



2014

```
D2014 <- subset(D2014, EDADMUJ < 999)
```

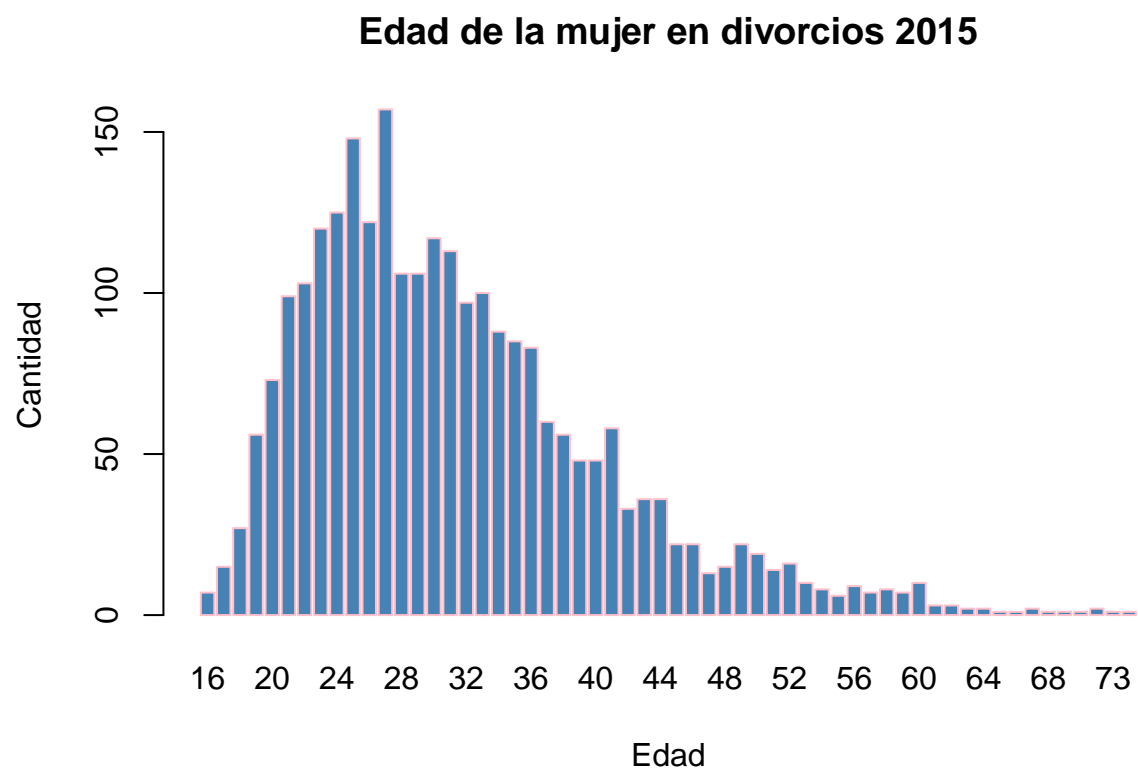
```
barplot(table(D2014$EDADMUJ), main = "Edad de la mujer en divorcios 2014", xlab = "Edad", ylab = "Cantidad")
```



2015

```
D2015 <- subset(D2015, EDADMUJ < 999)
```

```
barplot(table(D2015$EDADMUJ), main = "Edad de la mujer en divorcios 2015", xlab = "Edad", ylab = "Cantidad")
```

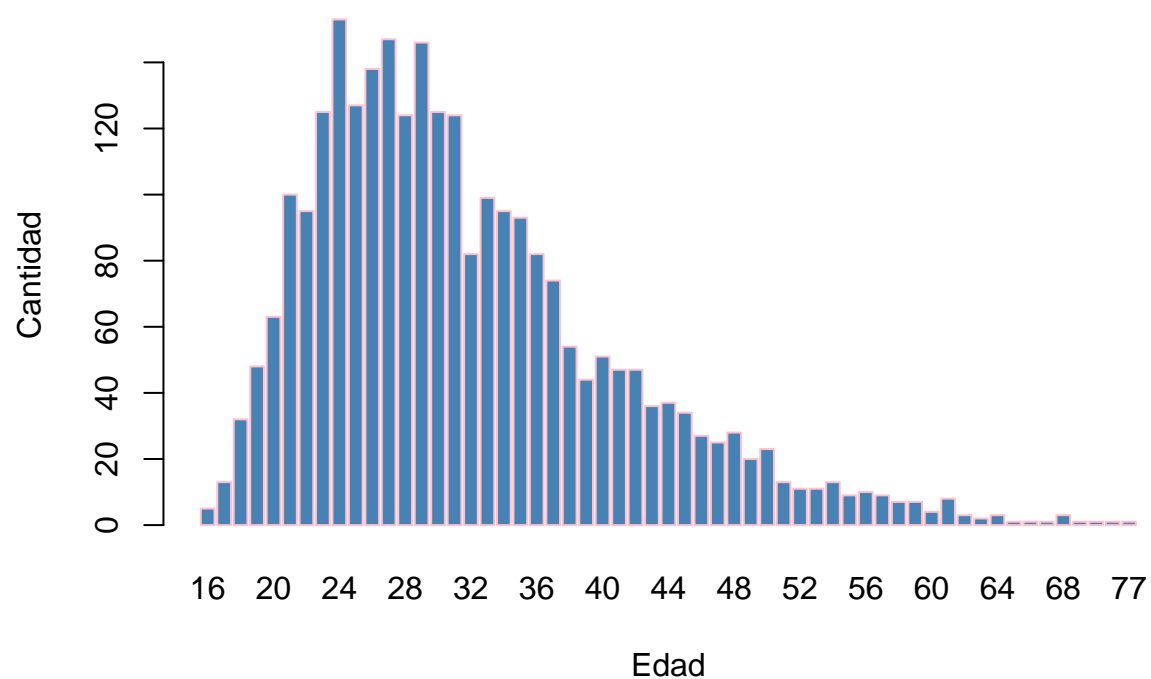


2016

```
D2016 <- subset(D2016, EDADMUJ < 999)
```

```
barplot(table(D2016$EDADMUJ), main = "Edad de la mujer en divorcios 2016", xlab = "Edad", ylab = "Cantidad")
```

Edad de la mujer en divorcios 2016

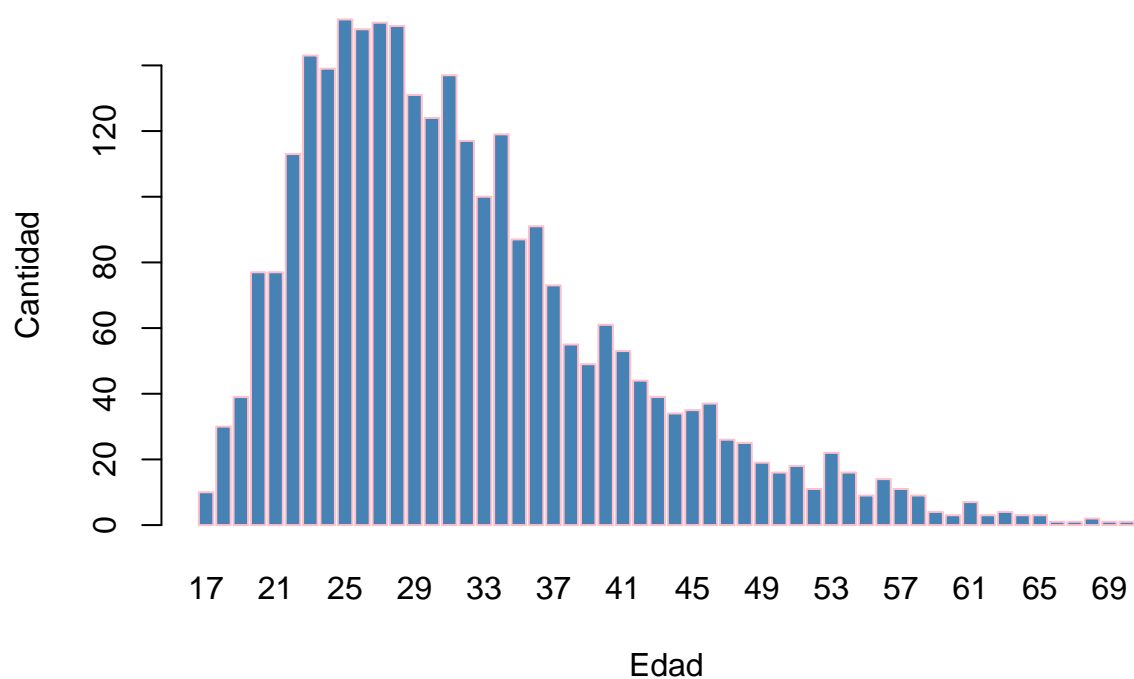


2017

```
D2017 <- subset(D2017, EDADMUJ < 999)
```

```
barplot(table(D2017$EDADMUJ), main = "Edad de la mujer en divorcios 2017", xlab = "Edad", ylab = "Cantidad")
```

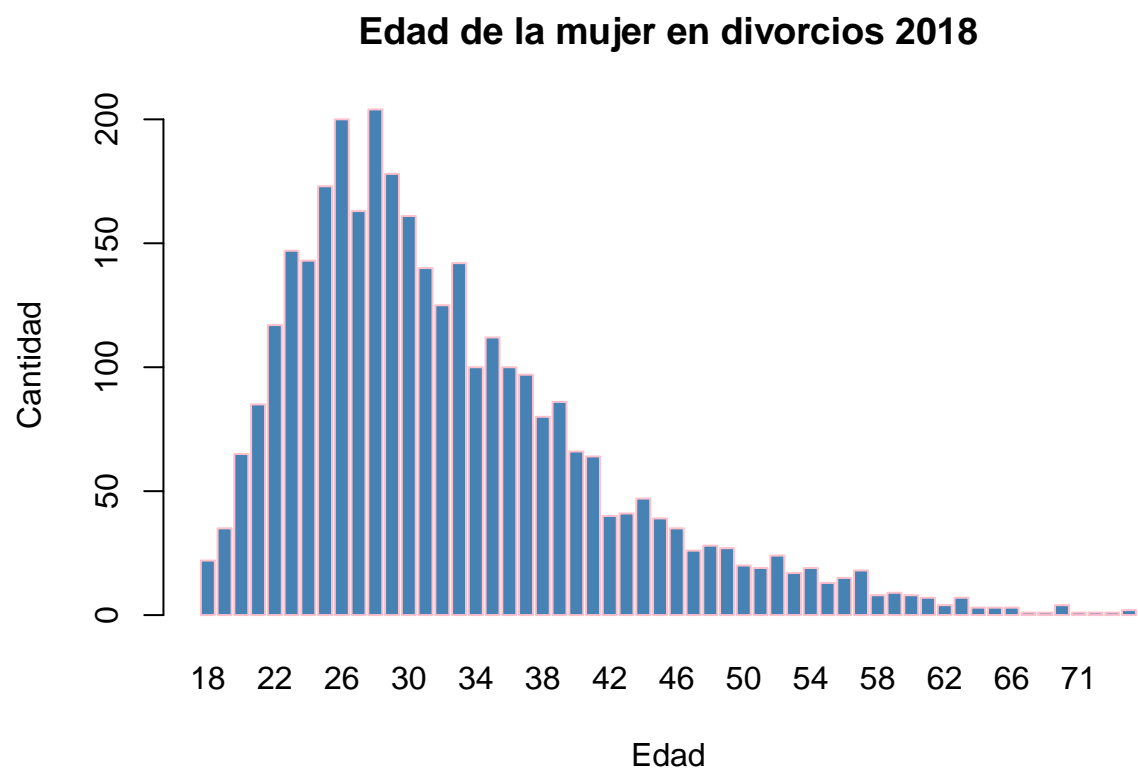
Edad de la mujer en divorcios 2017



2018

```
D2018 <- subset(D2018, EDADMUJ < 999)
```

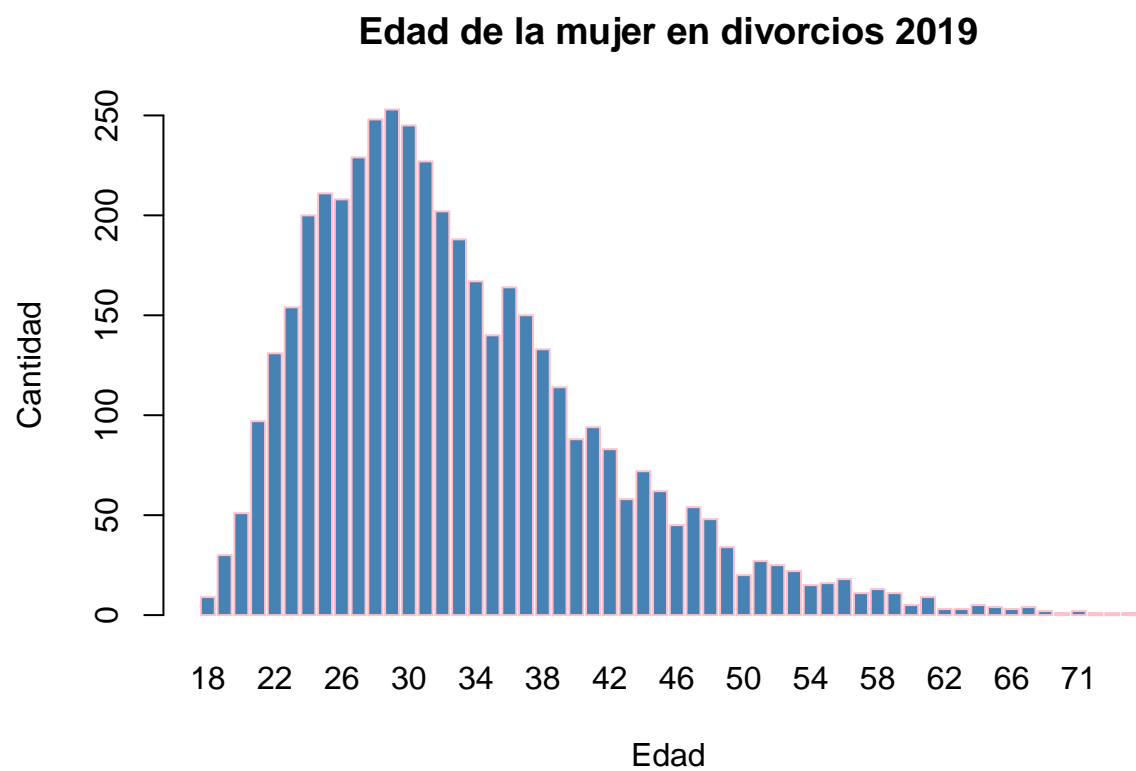
```
barplot(table(D2018$EDADMUJ), main = "Edad de la mujer en divorcios 2018", xlab = "Edad", ylab = "Cantidad")
```



2019

```
D2019 <- subset(D2019, EDADMUJ < 999)
```

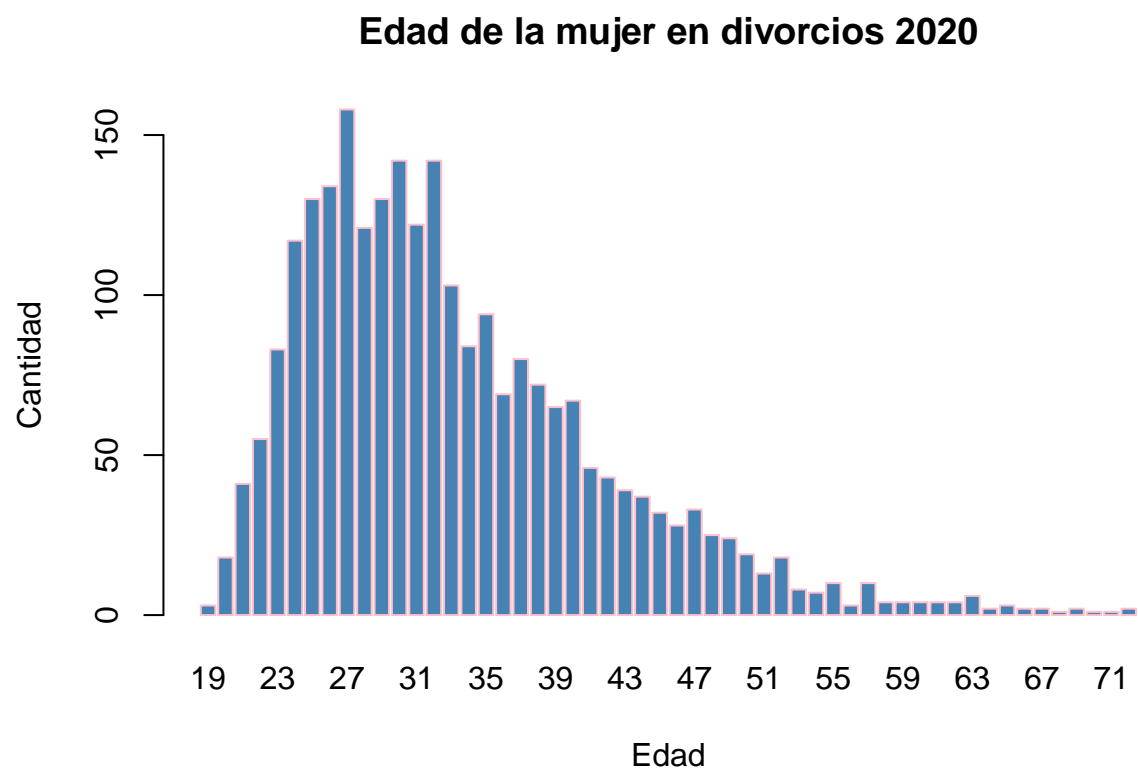
```
barplot(table(D2019$EDADMUJ), main = "Edad de la mujer en divorcios 2019", xlab = "Edad", ylab = "Cantidad")
```

2020

```
D2020 <- subset(D2020, EDADMUJ < 999)
```

```
barplot(table(D2020$EDADMUJ), main = "Edad de la mujer en divorcios 2020", xlab = "Edad", ylab = "Cantidad")
```

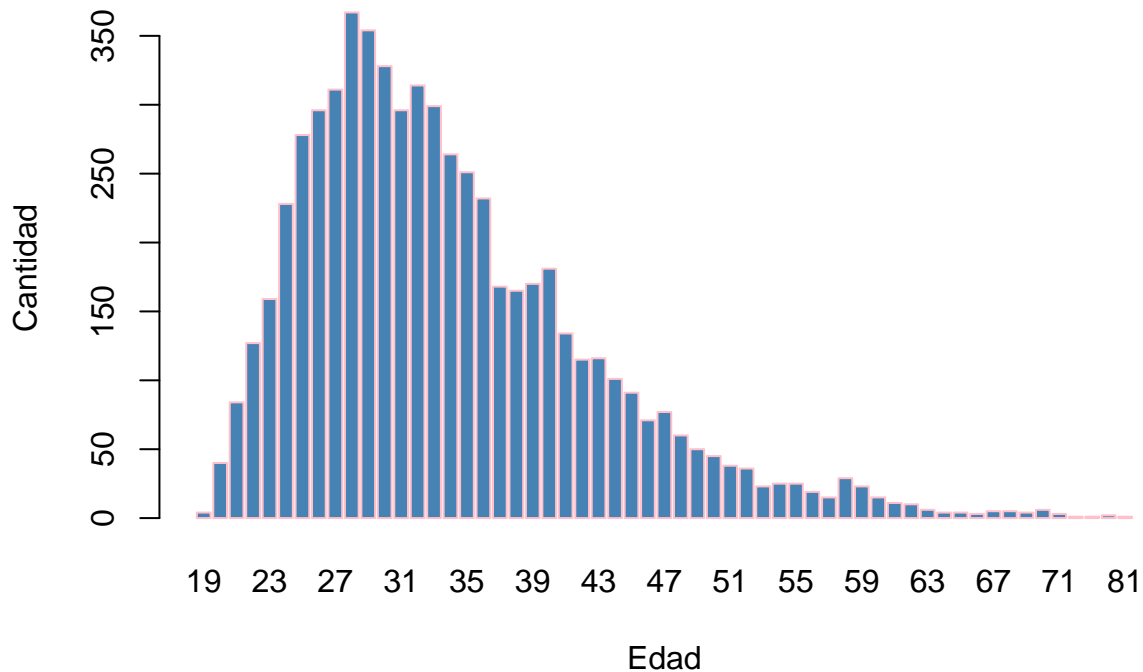


2021

```
D2021 <- subset(D2021, EDADMUJ < 999)
```

```
barplot(table(D2021$EDADMUJ), main = "Edad de la mujer en divorcios 2021", xlab = "Edad", ylab = "Cantidad")
```

Edad de la mujer en divorcios 2021



Convinando datos

```
colnames(D2009) <- c("DEPREG", "MUPREG", "MESREG", "AÑOREG", "DIAOCU", "MESOCU", "AÑOOCU", "DEPOCU", "MUPOCU")
colnames(D2010) <- c("DEPREG", "MUPREG", "MESREG", "AÑOREG", "DIAOCU", "MESOCU", "AÑOOCU", "DEPOCU", "MUPOCU")
colnames(D2011) <- c("DEPREG", "MUPREG", "MESREG", "AÑOREG", "DIAOCU", "MESOCU", "AÑOOCU", "DEPOCU", "MUPOCU")
colnames(D2012) <- c("DEPREG", "MUPREG", "MESREG", "AÑOREG", "DIAOCU", "MESOCU", "DEPOCU", "MUPOCU")
colnames(D2013) <- c("DEPREG", "MUPREG", "MESREG", "AÑOREG", "DIAOCU", "MESOCU", "DEPOCU", "MUPOCU")
colnames(D2014) <- c("DEPREG", "MUPREG", "MESREG", "AÑOREG", "DIAOCU", "MESOCU", "DEPOCU", "MUPOCU")
colnames(D2015) <- c("DEPREG", "MUPREG", "MESREG", "AÑOREG", "DIAOCU", "MESOCU", "AÑOOCU", "DEPOCU")

D2009 <- D2009 %>% mutate(AÑOREG = as.numeric(AÑOREG))
D2010 <- D2010 %>% mutate(AÑOREG = as.numeric(AÑOREG))
D2011 <- D2011 %>% mutate(AÑOREG = as.numeric(AÑOREG))
D2012 <- D2012 %>% mutate(AÑOREG = as.numeric(AÑOREG))
D2013 <- D2013 %>% mutate(AÑOREG = as.numeric(AÑOREG))
D2014 <- D2014 %>% mutate(AÑOREG = as.numeric(AÑOREG))
D2015 <- D2015 %>% mutate(AÑOREG = as.numeric(AÑOREG))
D2016 <- D2016 %>% mutate(AÑOREG = as.numeric(AÑOREG))
```

```
D2017 <- D2017 %>% mutate(AÑOREG = as.numeric(AÑOREG))
D2018 <- D2018 %>% mutate(AÑOREG = as.numeric(AÑOREG))
D2019 <- D2019 %>% mutate(AÑOREG = as.numeric(AÑOREG))
D2020 <- D2020 %>% mutate(AÑOREG = as.numeric(AÑOREG))
D2021 <- D2021 %>% mutate(AÑOREG = as.numeric(AÑOREG))
```

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

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

```
#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, 9, 22, 3, ...
## ..@ 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"
```

```

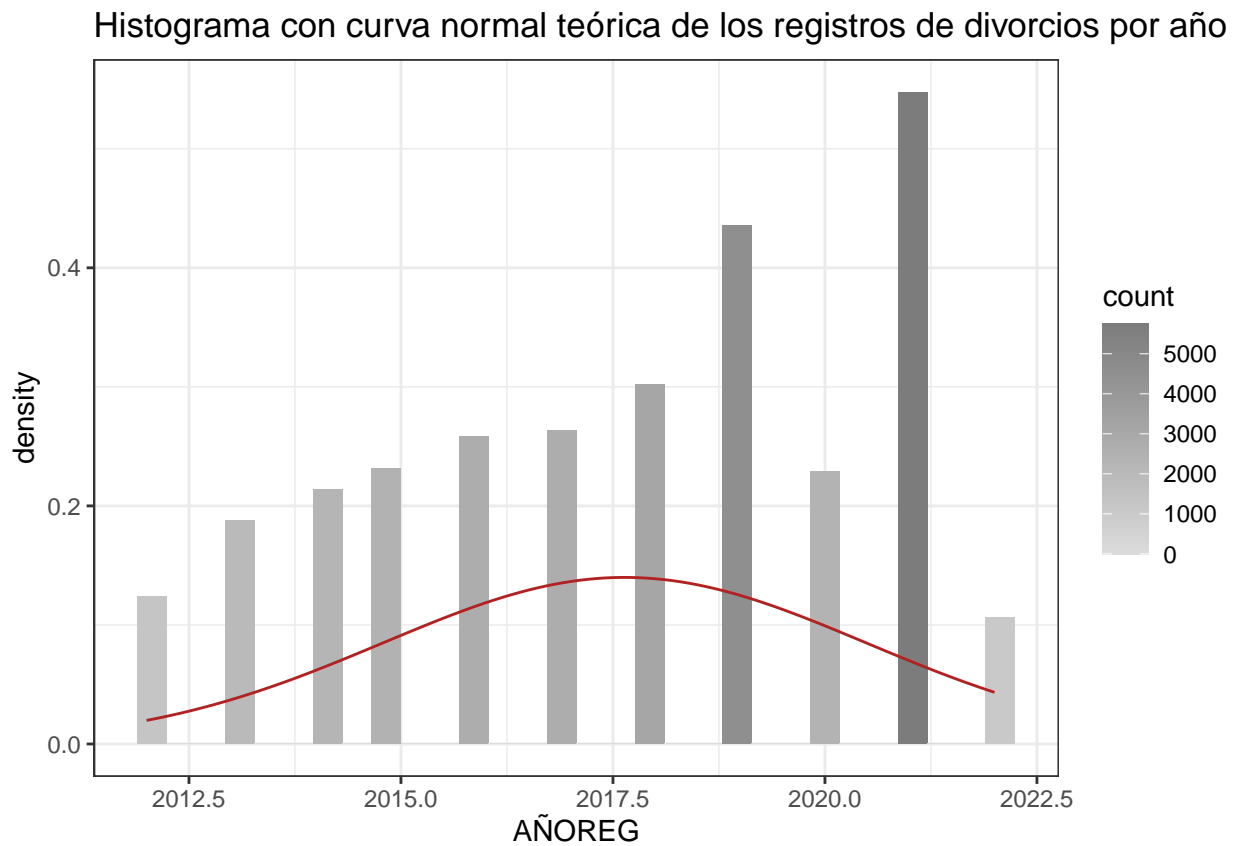
## $ MUPREG : chr+lbl [1:30378] 1708, 1213, 0101, 1416, 0101, 0101, 0101, 0101, 0101...
## ..@ labels: Named chr [1:342] "1010" "2010" "0110" "1210" ...
## ..- attr(*, "names")= chr [1:342] "San Antonio Suchitepéquez" "San Jacinto" "San Juan Sacatepé"
## ..@ label : chr "Municipio de registro"
## $ MESREG : 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"
## $ AÑOREG : 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, 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" "Sacatepequez" "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 Sacatepé"
## ..@ 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"
## $ EDADMUJ: 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"
## $ PPERMUJ: dbl+lbl [1:30378] 2, 2, 2, 2, 2, 9, 9, 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"
## $ NACMUJ : dbl+lbl [1:30378] 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 [1:30378] 9, 5, 5, 1, 4, 9, 5, 5, 5, 9, 4, 9, 3, 5, 9, 9, 3, 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 del hombre"
## $ ESCMUJ : 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" ...
## $ CIUOMUJ: chr [1:30378] "110" "1120" "1120" "1120" ...
## $ AÑOOCU : num [1:30378] NA NA NA NA NA NA NA NA NA NA ...
```

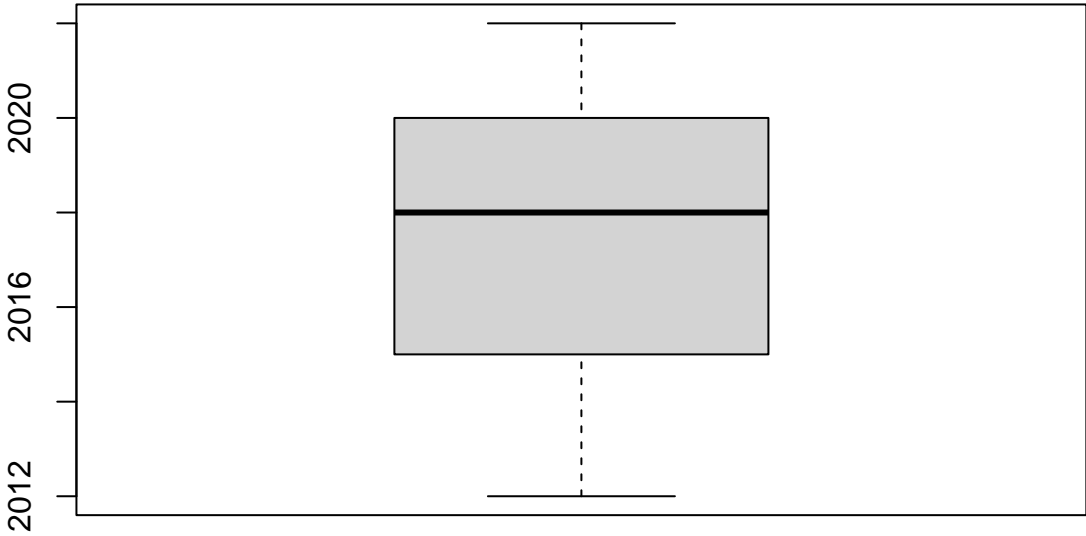
```
## Warning: The dot-dot notation ('..density..') was deprecated in ggplot2 3.4.0.
## i Please use 'after_stat(density)' instead.
```

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



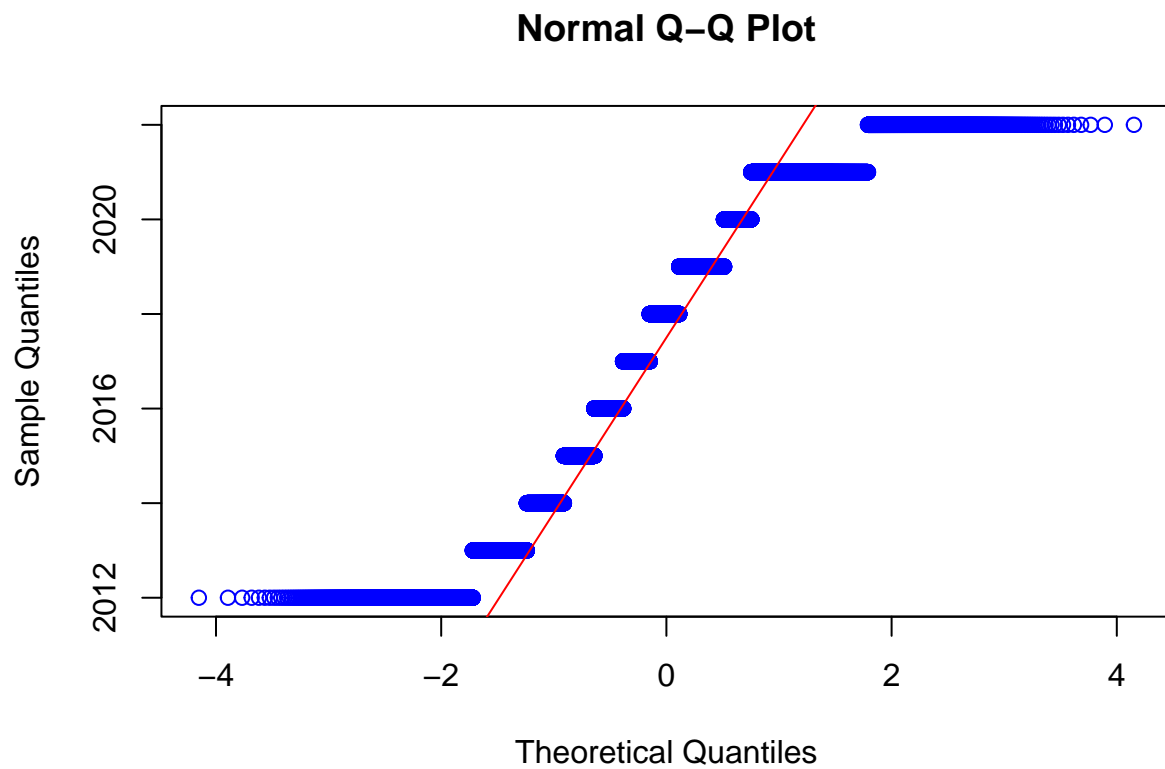
```
boxplot(divorcios$AÑOREG, main = "Caja y Bigotes de registro de divorcios por año (2012 - 2021)", xlab = "AÑOREG")
```

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:

```
## Don't know how to automatically pick scale for object of type  
## <haven_labelled/vctrs_vctr/double>. Defaulting to continuous.  
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```


Histograma con curva normal teórica de las edades de los hombres (2012)

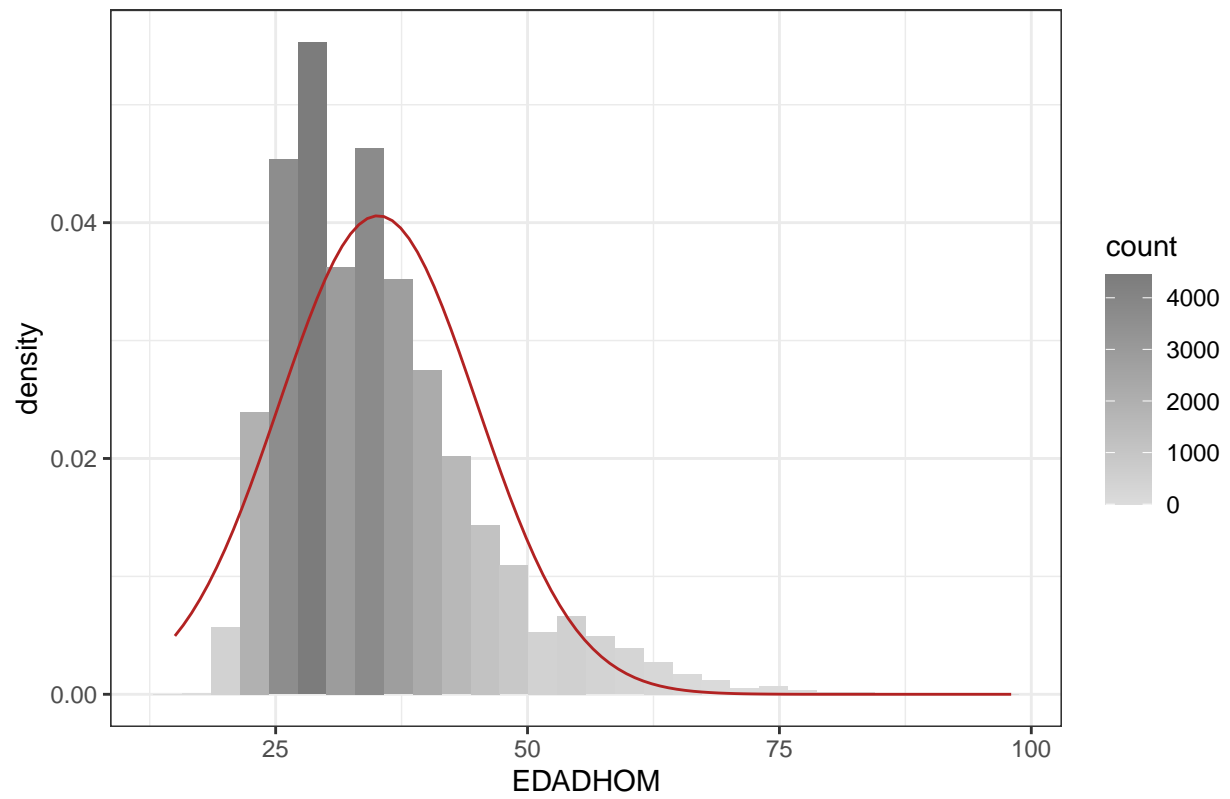


Diagrama de caja y bigotes

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

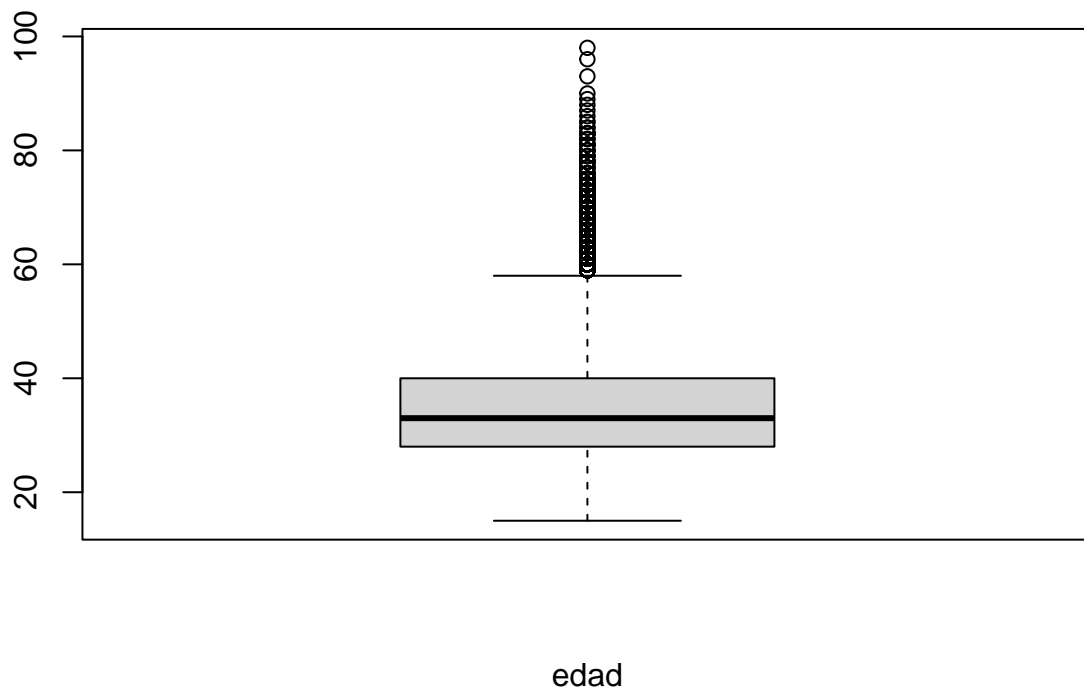
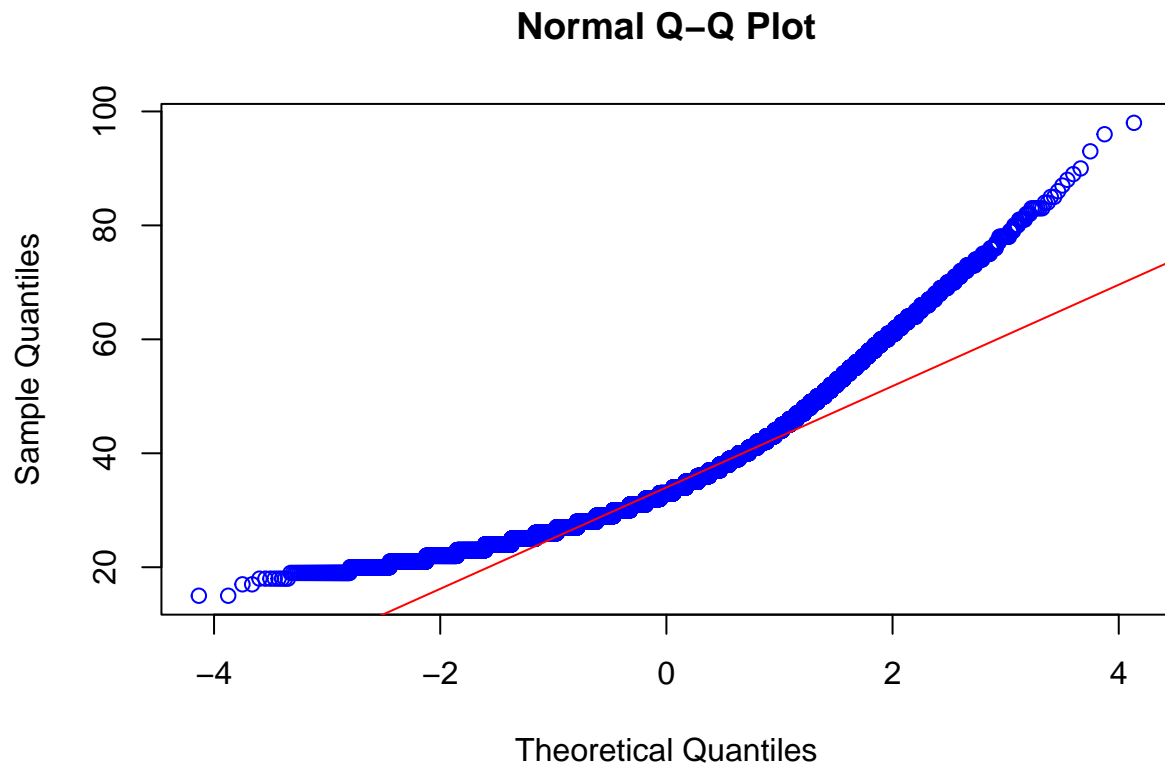


Diagrama qqnormal



Registros de edad de mujeres

Prueba de normalidad para la edad de la mujer:

```
## Don't know how to automatically pick scale for object of type  
## <haven_labelled/vctrs_vctr/double>. Defaulting to continuous.  
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```

Histograma con curva normal teórica de las edades de las mujeres (2012)

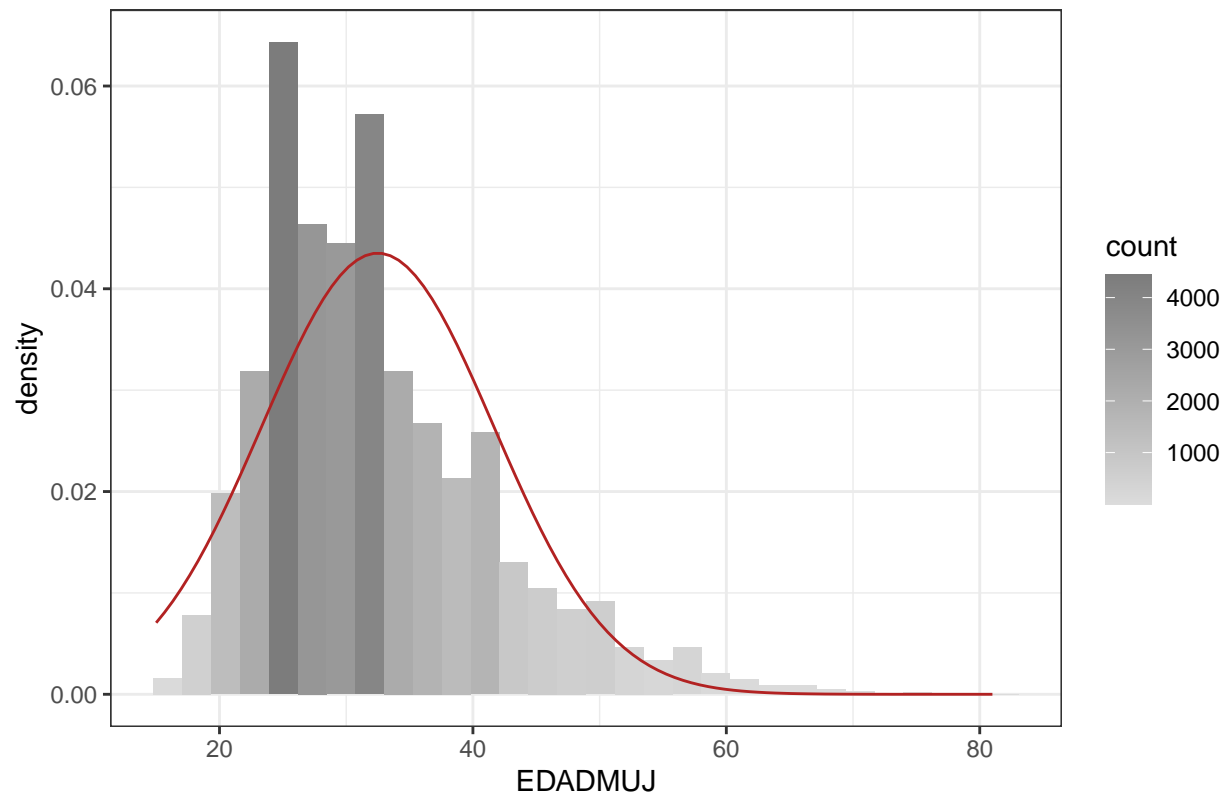


Diagrama de caja y bigotes

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

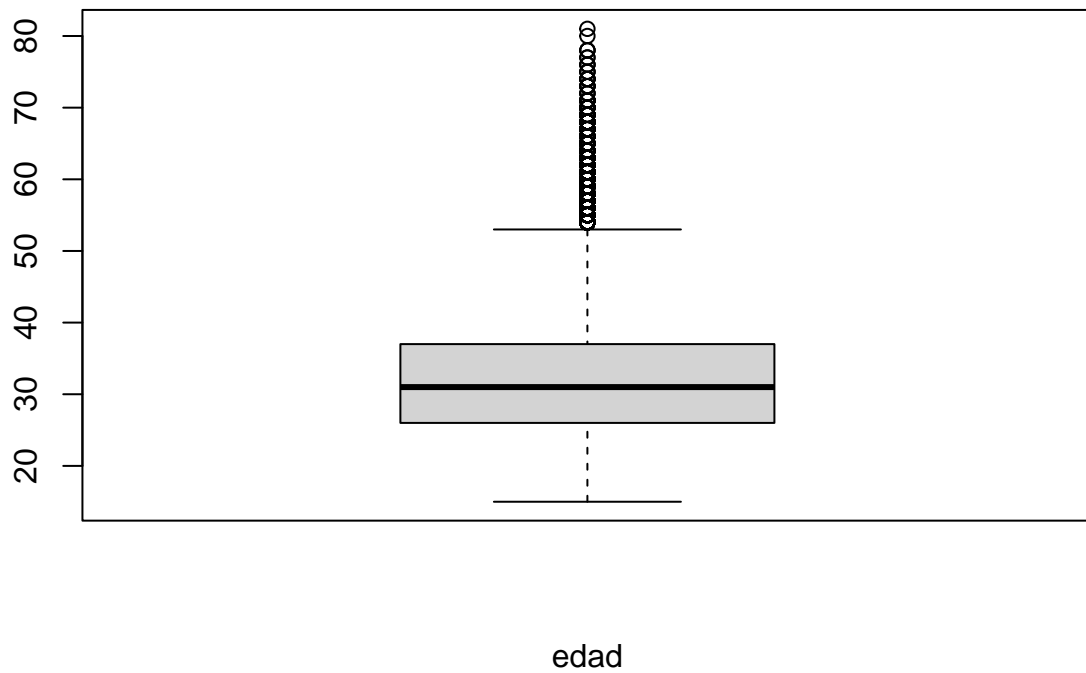


Diagrama qqnormal

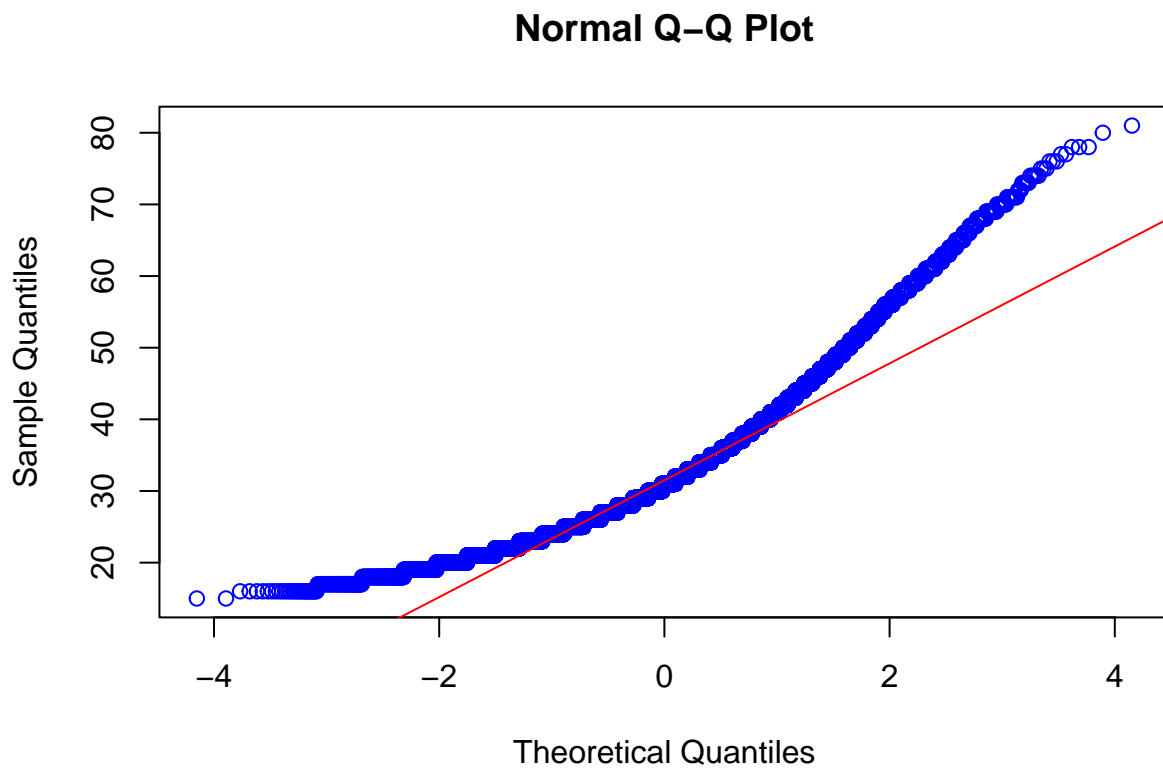


Tabla de frecuencias para variables cualitativas

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      Sacatepequez      Chimaltenango      Escuintla
##      10008           554           482           815           1418
##      Santa Rosa      Solola      Totonicapan      Quetzaltenango      Suchitepequez
##      898           312           404           2751           705
##      Retalhuleu      San Marcos      Huehuetenango      Quiche      Baja Verapaz
##      912           1455           1107           635           645
##      Alta Verapaz      Peten      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)

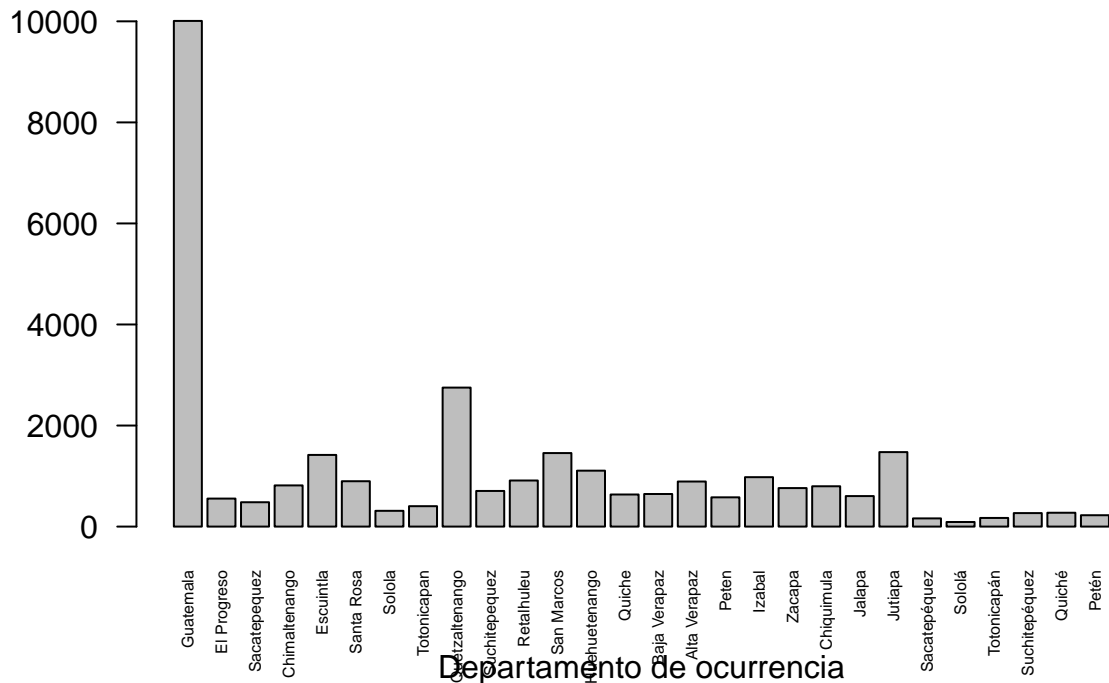


Tabla de frecuencias para el municipio de ocurrencia

##		
##	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	Chuarrrancho
##	91	19
##	Fraijanes	Amatitlán
##	82	356
##	Villa Nueva	Villa Canales
##	749	314
##	Petapa	Guastatoya
##	231	206
##	Morazán	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	Patzicía
##	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
##	Cuilapa	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 Siguilá	Ostuncalco
##	9	77
##	San Mateo	Concepción Chiquirichapa
##	20	30
##	San Martín Sacatepéquez	Almolonga
##	36	29
##	Cantel	Huitán
##	121	14
##	Zunil	Colomba
##	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
##	57	44
##	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 Cucho
##	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
##	Barillas	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 Alzatate	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)

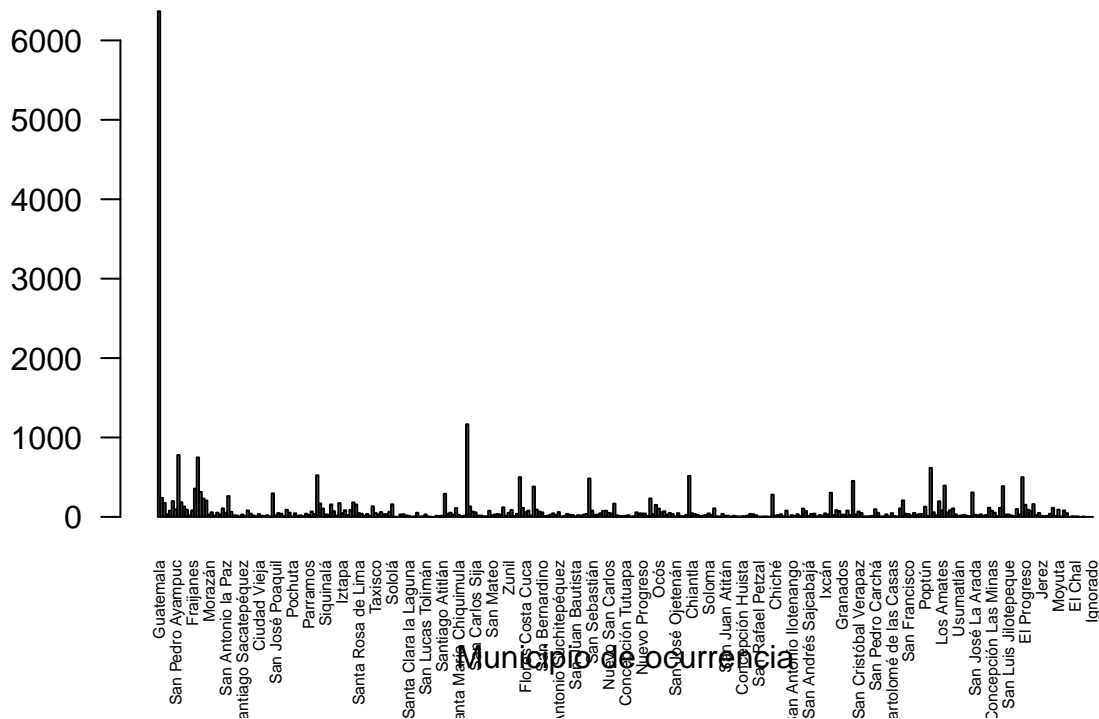


Tabla de frecuencias para la escolaridad del hombre

##					
##	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)

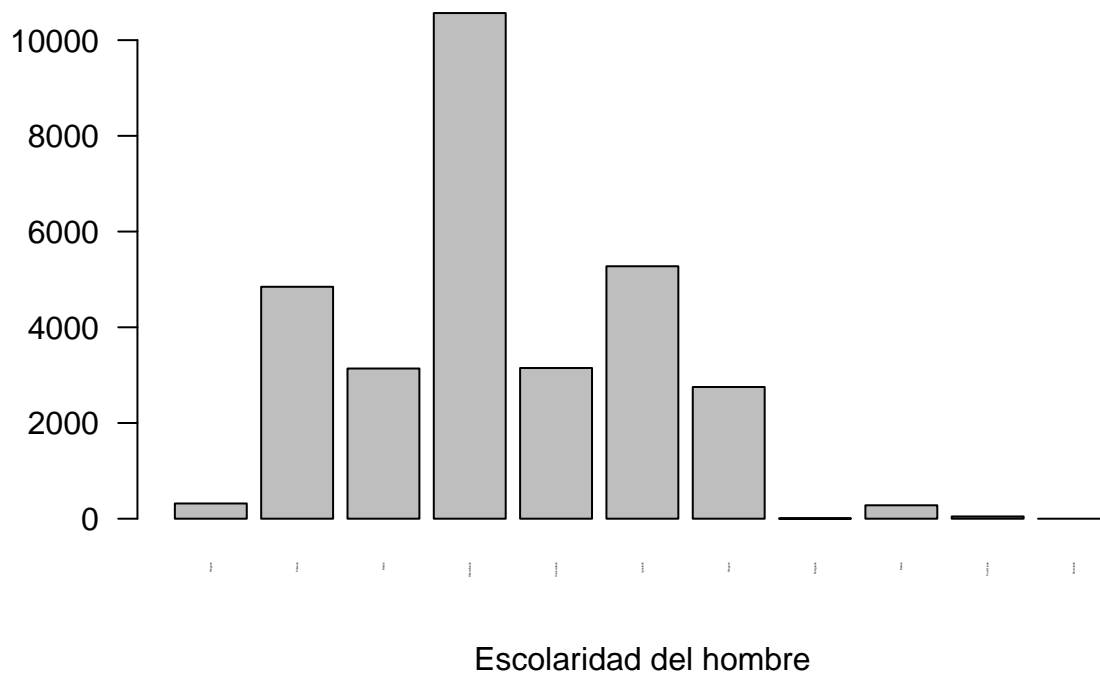


Tabla de frecuencias para la escolaridad de la Mujer

```
##
##      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)

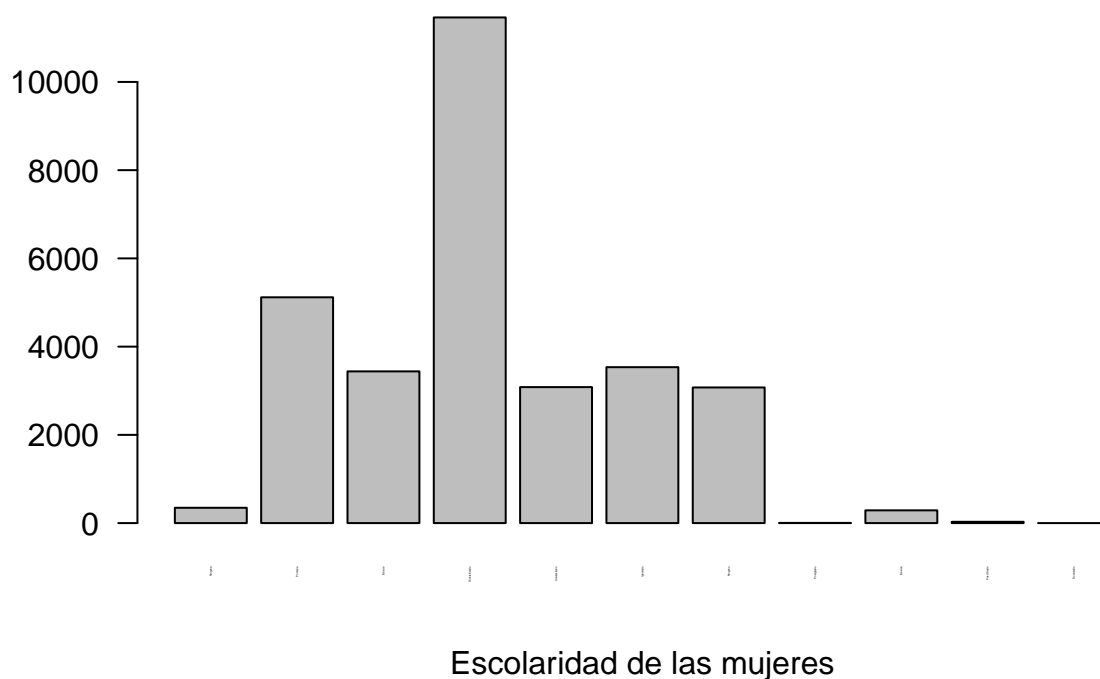


Tabla de frecuencias para el grupo étnico del hombre

##	Indigena	No indigena	Ignorado	Maya
##	182	820	6141	3564
##	Garífuna	Xinka Mestizo / ladino	Otro	
##	17	1	4247	293
##	Xinka	Garifuna Mestizo / Ladino	Ladino / Mestizo	
##	4	1	1773	13335

Distribución de etnia de los hombres (2012 – 2021)

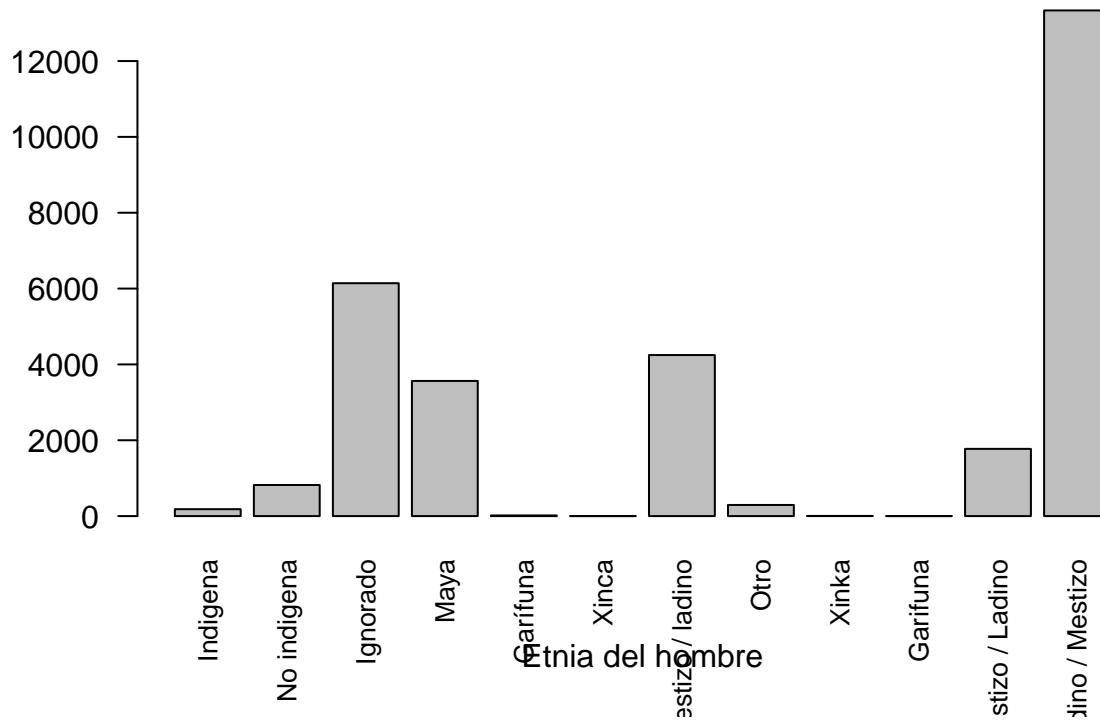
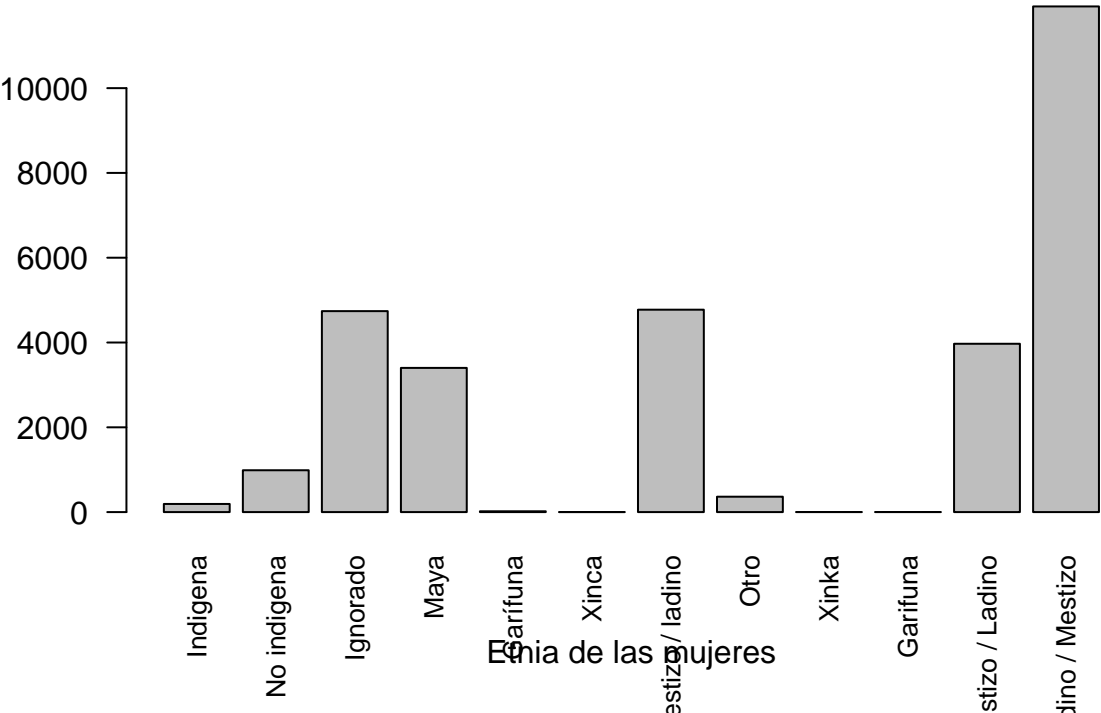


Tabla de frecuencias para el grupo étnico de la mujer

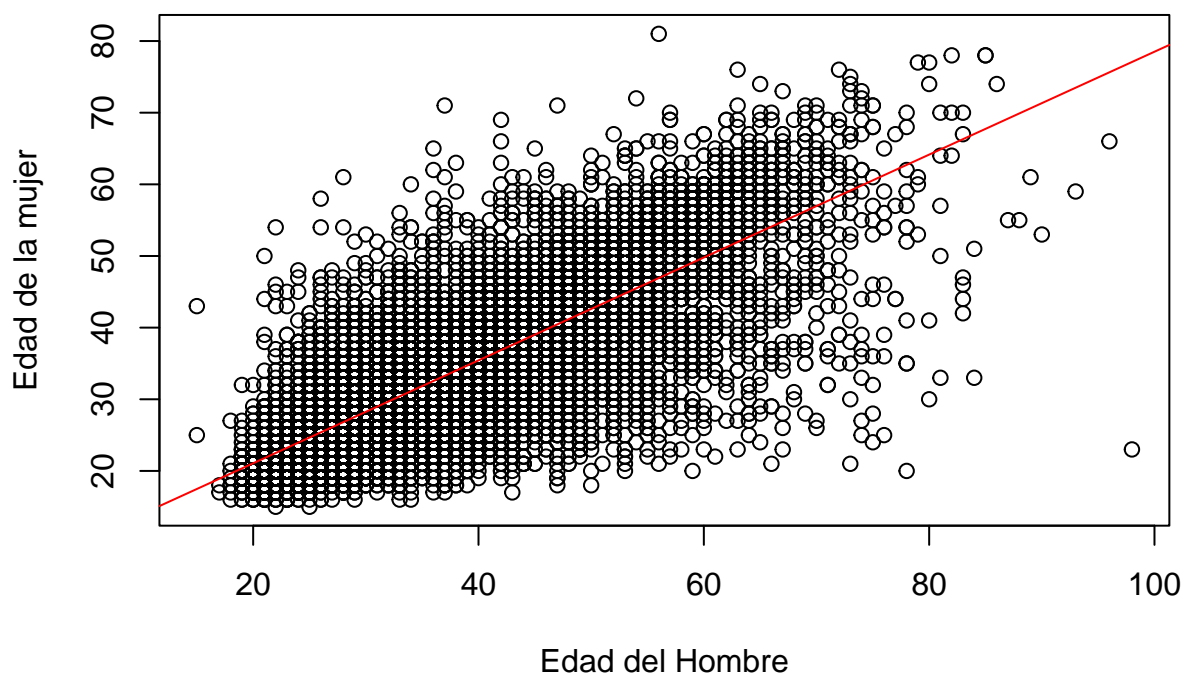
##				
##	Indígena	No indígena	Ignorado	Maya
##	192	987	4740	3401
##	Garífuna	Xinca	Mestizo / ladino	Otro
##	19	1	4774	363
##	Xinca	Garífuna	Mestizo / Ladino	Ladino / Mestizo
##	2	2	3970	11927

Distribución de etnia de las mujeres (2012 – 2021)



Correlacion entre edades de los cónyuges

Correlación entre las edades de los cónyuges (2012 – 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	EDADMUJ	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 :6027

