Proyecto de Minería de datos. Matrimonios de Guatemala

Ayleen Rubio 19003, Andrés Say 19705, Andreé Toledo 18439

25/3/2021

# Proyecto No. 1. Análisis de matrimonios en Guatemala

Actualmente en Guatemala es legal contraer matrimonio hasta cumplir la mayoría de edad, es decir los 18 años, tanto para los hombres como para las mujeres, sin embargo, esto no fue decretado en el Código Civil hasta agosto de 2017. Anteriormente en el 2015 ya se había hecho una reforma para elevar la edad mínima de matrimonio a los 18 años pero en estas reformas aún era posible contraer matrimonio a los 16 años con el permiso de un juez. Antes de todas estas reformas en los Códigos Civil y Penal, era posible contraer matrimonio desde que se cumplía los 14 años de edad y, en caso de agresiones sexuales, era permitido el matrimonio civil entre ofendida y ofensor en relaciones con menores de edad, lo cual cambió en el año 2005 al ser declarado Inconstitucional, posteriormente en el año 2009 La Ley contra la Violencia Sexual, Explotación y Trata de Personas agravó las sentencias en los casos de delitos sexuales con víctimas menores de 18 años. -unionestempranas.org

A pesar de que se haya prohibido contraer matrimonio siendo menores de edad, según una declaración del defensor de la Niñez de la PDH Luis de León en el año 2019, aún se encuentran registros de matrimonio entre dos personas menores, ocurriendo principalmente en departamentos occidentales indígenas de Huehuetenango, Quetzaltenango y San Marcos, además de la capital del país. Según se informó, algunas instituciones públicas y juristas no tenían conocimientos sobre tal prohibición, razón por la cual tuvieron lugar estos matrimonios sin algún tipo de dificultad. -Prensa Libre: redacción AFP, 2019.

En la presente investigación se busca estudiar los casos de matrimonio desde el año 2009 hasta el año 2019 para verificar si existe algún cambio en la ocurrencia de matrimonios a partir de los años de las reformas y para estudiar los casos de matrimonios con menores de edad y verificar si estos suelen ser con otras personas menores de edad o con mayores de edad, de ser el caso, se busca estudiar a su vez la diferencia de edad entre ambos cónyuges.

## Análisis exploratorio

Para poder hacer un análisis exploratorio de los datos de 2009 a 2019 ha sido necesario hacer una limpieza de estos en los siguientes aspectos: Principalmente se cuenta con las columnas que describen departamento de registro, municipio de registro, mes de registro, año de registro, clase de unión, número de nupcias contraídas por el hombre y por la mujer, edad del hombre y de la mujer, grupo étnico del hombre y de la mujer, nacionalidad del hombre y de la mujer, ocupación del hombre y de la mujer, escolaridad del hombre y de la mujer, área geográfica de ocurrencia y día de ocurrencia. Sin embargo, no todos estos datos se encontraban en todas las bases de datos o no contaban con el mismo nombre de variable o se encontraban con diferencias entre mayúsculas y minúsculas, de la misma manera se encontró que en algunos casos el año de ocurrencia se encontraba descrito únicamente como 9 o 10 en vez de contar con el año completo, por lo que fue necesario hacer cambios para generalizar todo. Como estándar se han utilizado los años 2015, 2016 y 2017, por lo que las bases de los años anteriores se han adaptado al formato de estas tres, agregando las columnas de número de nupcias del hombre y la mujer en los años 2010, 2011, 2012, 2013 y 2014, año de ocurrencia en 2012, 2013 y 2014, escolaridad del hombre y de la mujer y día de ocurrencia en el año 2009 y por último se ha agregamo la columna de área geográfica de ocurrencia en los años 2018 y 2019. Luego de agregar estar columnas se ha hecho las modificaciones necesarias en el formato de los datos o respectivos nombres de las columnas para que todos los años contaran con el mismo formato y pudieran ser unificados en una sola base de datos.

## Warning: package 'e1071' was built under R version 4.0.4

## Warning: package 'caret' was built under R version 4.0.4

## Loading required package: lattice

## Loading required package: ggplot2

##   
## 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

## ------------------------------------------------------------------------------

## You have loaded plyr after dplyr - this is likely to cause problems.  
## If you need functions from both plyr and dplyr, please load plyr first, then dplyr:  
## library(plyr); library(dplyr)

## ------------------------------------------------------------------------------

##   
## Attaching package: 'plyr'

## The following objects are masked from 'package:dplyr':  
##   
## arrange, count, desc, failwith, id, mutate, rename, summarise,  
## summarize

## Package 'mclust' version 5.4.7  
## 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 'randomForest' was built under R version 4.0.4

## randomForest 4.6-14

## Type rfNews() to see new features/changes/bug fixes.

##   
## Attaching package: 'randomForest'

## The following object is masked from 'package:dplyr':  
##   
## combine

## The following object is masked from 'package:ggplot2':  
##   
## margin

##   
## Attaching package: 'ggpubr'

## The following object is masked from 'package:plyr':  
##   
## mutate

## corrplot 0.84 loaded

## Warning: package 'caretEnsemble' was built under R version 4.0.4

##   
## Attaching package: 'caretEnsemble'

## The following object is masked from 'package:ggplot2':  
##   
## autoplot

## Warning: package 'psych' was built under R version 4.0.4

##   
## Attaching package: 'psych'

## The following object is masked from 'package:randomForest':  
##   
## outlier

## The following object is masked from 'package:mclust':  
##   
## sim

## The following objects are masked from 'package:ggplot2':  
##   
## %+%, alpha

## Warning: package 'mice' was built under R version 4.0.4

##   
## Attaching package: 'mice'

## The following object is masked from 'package:stats':  
##   
## filter

## The following objects are masked from 'package:base':  
##   
## cbind, rbind

## Warning: package 'GGally' was built under R version 4.0.4

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

## Warning: package 'clValid' was built under R version 4.0.4

## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated  
  
## Warning in `[<-.factor`(`\*tmp\*`, ri, value = c(2, 9, 2, 1, 1, 1, 1, 1, 2, :  
## invalid factor level, NA generated

## Informe de análisis exploratorio

En este informe usted podrá leer los hallazgos sobre los matrimonios, basado en los datos obtenidos del INE

## DEPREG MUPREG MESREG   
## Guatemala :172623 Guatemala : 87005 Diciembre: 83457   
## Huehuetenango : 63131 Villa Nueva : 15316 Enero : 76874   
## Alta Verapaz : 57971 Quetzaltenango : 12272 Marzo : 74823   
## San Marcos : 54984 Mixco : 11904 Mayo : 74794   
## Quiche : 53302 Cobán : 10906 Noviembre: 70123   
## Quetzaltenango: 50457 San Pedro Carchá: 10657 Abril : 69336   
## (Other) :377637 (Other) :682045 (Other) :380698   
## AÑOREG CLAUNI EDADHOM   
## Min. :2009 Matrimonio : 62104 Min. : 1.00   
## 1st Qu.:2012 Unión de hecho : 0 1st Qu.: 7.00   
## Median :2014 Comunidad absoluta : 44066 Median :11.00   
## Mean :2014 Separación absoluta : 15044 Mean :14.47   
## 3rd Qu.:2017 Comunidad de gananciales:672119 3rd Qu.:17.00   
## Max. :2020 No especificado : 36772 Max. :89.00   
##   
## EDADMUJ GETHOM GETMUJ   
## Min. : 1.00 Indigena :101309 Indigena : 98014   
## 1st Qu.: 7.00 No Indigena: 22152 No Indigena: 21926   
## Median :10.00 Ignorado : 69966 Ignorado : 77158   
## Mean :13.45 No indigena:104340 No indigena:100669   
## 3rd Qu.:16.00 NA's :532338 NA's :532338   
## Max. :92.00   
##   
## NACHOM NACMUJ   
## Guatemala :822684 Guatemala :821778   
## Estados Unidos de América: 1868 Estados Unidos de América: 1987   
## México : 1543 México : 1603   
## El Salvador : 999 El Salvador : 1591   
## Nicaragua : 487 Honduras : 797   
## Honduras : 359 Nicaragua : 653   
## (Other) : 2165 (Other) : 1696   
## CIUOHOM   
## Peones agropecuarios, pesqueros y forestales :121065   
## No especificado en otro grupo :116574   
## Agricultores y trabajadores calificados de cultivos extensivos :103076   
## Agricultores y trabajadores calificados de explotaciones agropecuarias con destino al mercado: 77926   
## Estudiante : 73917   
## Vendedores : 42023   
## (Other) :295524   
## CIUOMUJ NUNUHO NUNUMU   
## No especificado en otro grupo :410150 None :395909 None :395909   
## Oficios domésticos no remunerados:166759 Ignorado:279785 Ignorado:279548   
## Estudiante : 72134 1 :152743 1 :153671   
## Profesionales de la enseñanza : 28155 2 : 1635 2 : 954   
## Ignorado : 22439 3 : 24 3 : 15   
## Oficinistas : 13621 5 : 5 5 : 4   
## (Other) :116847 (Other) : 4 (Other) : 4   
## DEPOCU MUPOCU MESOCU   
## Guatemala :169944 Guatemala : 70183 Diciembre: 88368   
## Huehuetenango : 63128 Mixco : 19566 Noviembre: 82896   
## Alta Verapaz : 58011 Villa Nueva : 17955 Marzo : 76195   
## San Marcos : 54807 Quetzaltenango : 12705 Abril : 70995   
## Quiche : 53208 San Juan Sacatepéquez: 12109 Febrero : 69448   
## Quetzaltenango: 50267 Cobán : 10982 Octubre : 68300   
## (Other) :380740 (Other) :686605 (Other) :373903   
## AÑOOCU AREAGOCU ESCHOM ESCMUJ   
## Length:830105 Urbano :487340 Length:830105 Length:830105   
## Class :character Rural : 67815 Class :character Class :character   
## Mode :character Ignorado:123238 Mode :character Mode :character   
## 3 : 1   
## 11 : 3   
## None :151705   
## NA's : 3   
## DIAOCU   
## Length:830105   
## Class :character   
## Mode :character   
##   
##   
##   
##

En el siguiente cuadro se muestra el tipo de dato de cada variable:

## 'data.frame': 830105 obs. of 23 variables:  
## $ DEPREG : Factor w/ 24 levels "Guatemala","El Progreso",..: 12 11 13 14 14 13 14 14 4 17 ...  
## $ MUPREG : Factor w/ 337 levels "Guatemala","Santa Catarina Pinula",..: 167 156 193 233 233 198 237 237 50 277 ...  
## $ MESREG : Factor w/ 12 levels "Enero","Febrero",..: 4 5 1 8 10 6 10 12 8 8 ...  
## $ AÑOREG : num 2010 2009 2010 2009 2009 ...  
## $ CLAUNI : Factor w/ 6 levels "Matrimonio","Unión de hecho",..: 1 1 1 1 1 1 1 1 1 1 ...  
## $ EDADHOM : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ EDADMUJ : num 1 1 1 1 1 1 1 1 1 1 ...  
## $ GETHOM : Factor w/ 4 levels "Indigena","No Indigena",..: 1 3 3 1 1 3 1 1 1 1 ...  
## $ GETMUJ : Factor w/ 4 levels "Indigena","No Indigena",..: 1 3 3 1 1 3 1 1 1 2 ...  
## $ NACHOM : Factor w/ 123 levels "Canadá","Estados Unidos de América",..: 41 41 41 41 41 41 41 41 41 41 ...  
## $ NACMUJ : Factor w/ 115 levels "Canadá","Estados Unidos de América",..: 35 35 35 35 35 35 35 35 35 35 ...  
## $ CIUOHOM : Factor w/ 489 levels "Oficiales de las fuerzas armadas",..: 12 284 284 284 284 284 284 284 284 284 ...  
## $ CIUOMUJ : Factor w/ 487 levels "Oficiales de las fuerzas armadas",..: 33 440 440 440 440 440 440 440 440 440 ...  
## $ NUNUHO : Factor w/ 8 levels "1","2","3","5",..: 6 6 1 1 1 6 6 6 6 6 ...  
## $ NUNUMU : Factor w/ 8 levels "1","2","3","5",..: 6 6 1 1 1 6 6 6 6 6 ...  
## $ DEPOCU : Factor w/ 24 levels "Guatemala","El Progreso",..: 12 11 13 14 14 13 14 14 4 17 ...  
## $ MUPOCU : Factor w/ 348 levels "Guatemala","Santa Catarina Pinula",..: 167 156 193 233 233 198 237 237 50 277 ...  
## $ MESOCU : Factor w/ 12 levels "Enero","Febrero",..: 9 3 12 7 10 6 9 11 7 8 ...  
## $ AÑOOCU : chr "2009" "2009" "2009" "2009" ...  
## $ AREAGOCU: Factor w/ 6 levels "Urbano","Rural",..: 3 3 3 3 3 3 3 3 3 3 ...  
## $ ESCHOM : chr "None" "None" "None" "None" ...  
## $ ESCMUJ : chr "None" "None" "None" "None" ...  
## $ DIAOCU : chr "None" "None" "None" "None" ...

El dataset tiene 23 variables, el departamento, municipalidad, mes y año de registro, la casificación de la unión, las edades de los hombres y mujer, el grupo étnico de los hombres y mujeres, su nacionalidad, su oficio, la cantidad de nupcias de cada uno. Luego tambien se encuentran las variables de ocurrencia. el departamento, municipalidad, mes, año, area y dia de ocurrencia. Por último se encuentra la escolaridad de cada uno. Las variables más interesantes para el porpósito de la investigación son lass que proporcionan información de la fecha del matrimonio. Otras con gran potencial, son las edades de los cónyuges, ya que con ellas podemos analizar cual es la tendencia de matrimonios de menores, diferencias de edades etc.

## Análisis de variables cuantitativas

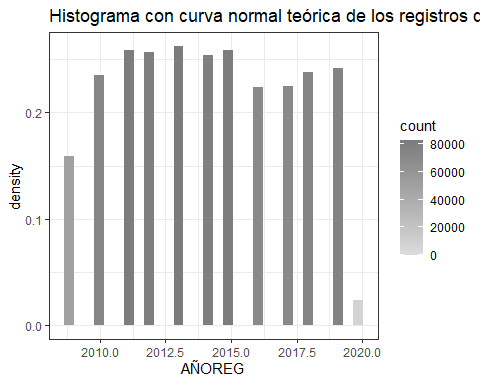
### Registros de matrimonio

#### Prueba de normalidad para registros de matrimonios:

## Warning in mean.default(datosMatrimonio$popularity): argument is not numeric or  
## logical: returning NA

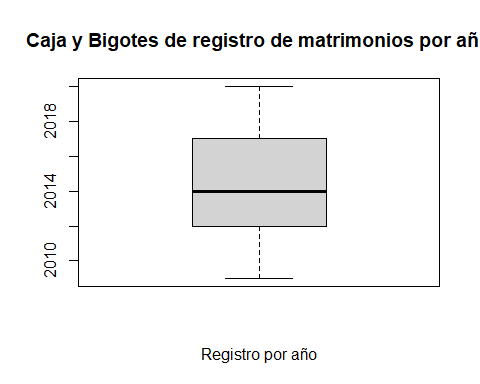
## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

## Warning: Removed 101 row(s) containing missing values (geom\_path).



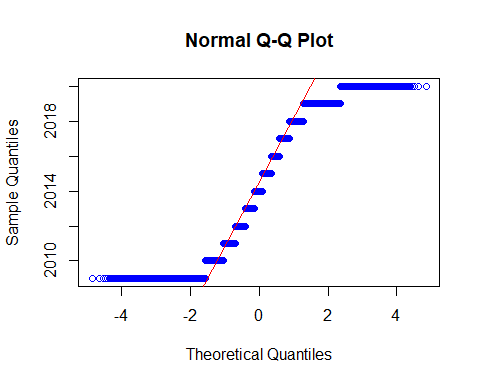
Como es de esperarse, no es posible observar que exista una relación en el histograma de registros de matrimonios con las reformas planteadas para la prohibición de matrimonios con menores de edad, ya que parece que la cantidad de matrimonios ha sido bastante constante durante los años. Según el histograma mostrado, la cantidad de matrimonios por año no sigue una distribución normal.

### Diagrama de caja y bigotes



Como dicho anteriormente, no hay muchos grupos atípicos ya que se concentran dentro de un intervalo.Según este diagrama, sí se cuenta con normalidad en la cantidad de matrimonios por año.

### Diagrama de qqnormal



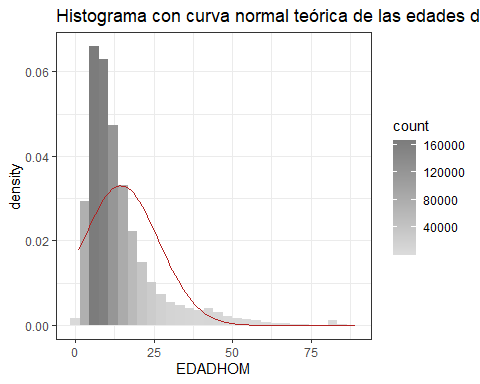
Aunque en el diagrama qqnormal parece que se sigue una distribución normal al encontrarse muchos de los datos alineados, hay bastantes datos que se encuentran fuera de la linea de tendencia, por lo que no se cumple con normalidad en la cantidad de matrimonios por año.

En conclusión, la cantidad de matrimonios por año no sigue una distribución normal.

## Registros de Edad hombre

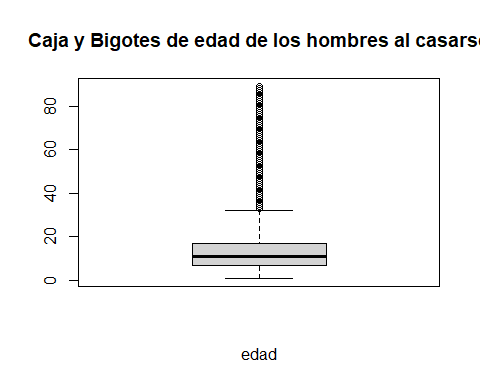
### Prueba de normalidad para la edad del hombre:

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



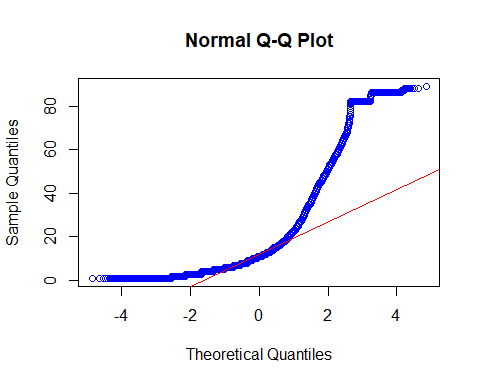
En este histograma se puede observar que los datos tienen relativamente forma de campana, sin embargo, es evidente que estos tienen un sesgo hacia la izquierda, esto quiere decir que usualmente la edad entre los hombres en un matrimonio se encuentra debajo de los 30. Sin embargo, es importante mencionar que como dato interesante se puede observar que hay casos de matrimonios en hombres de más de 75 años. Debido al sesgo, la edad de los hombres en matrimonios no sigue una distribución normal.

### Diagrama de caja y bigotes



Se evidencia la gran cantidad de datos atípicos en el diagrama, se encuentran muy dispersos arriba de la media y del extremo superior que se encuentra al rededor de los 30 años. Esto quiere decir que la edad de los hombres en el matrimonio no sigue una distribución normal.

### Diagrama qqnormal



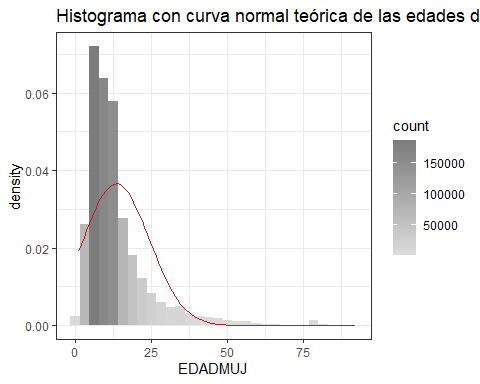
Como puede observarse, es muy notable en el diagrama qqnormal que los datos no siguen una distribución acorde a la linea de tendencia, esto quiere decir que la edad del hombre en el matrimonio no sigue una distribución normal.

En conclusión, la edad de los hombres en el matrimonio no sigue una distribución normal.

## Registros de edad de mujeres

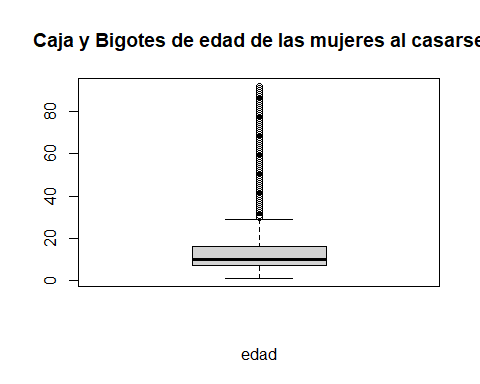
### Prueba de normalidad para la edad de la mujer:

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



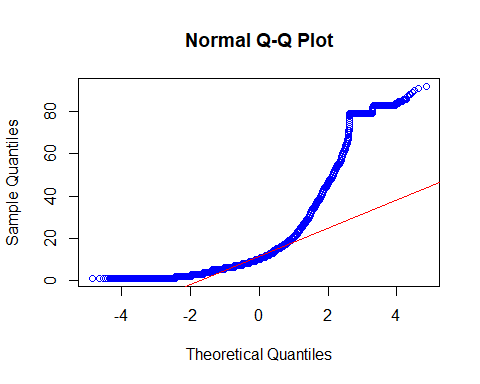
De la misma manera que en el histograma de hombres, hay una forma relativa de campana, sin embargo se nota un gran sesgo a la izquierda que demuestra que la edad de mujeres en el matrimonio suele ser debajo de los 25, aunque también se cuentan con algunos datos arriba de los 75 años. Debido al sesgo, la edad de mujeres en el matrimonio no sigue una distribución normal.

### Diagrama de caja y bigotes



En este diagrama de caja y bigotes es posible ver una gran cantidad de datos atípicos arriba del extremo superior que se encuentra al rededor de los 30 años, muy similar al diagrama de la edad de los hombres, por lo que de la misma manera no se cuenta con normalidad en la edad de las mujeres en el matrimonio.

### Diagrama qqnormal



Como puede observarse, los valores en este diagrama evidentemente no siguen con una distribución normal ya que no esán sobre la linea de tendencia, por lo que la edad de las mujeres en el matrimonio no sigue una distribución normal.

En conclusión, la edad de las mujeres en el matrimonio no sigue una distribución normal.

## Tabla de frecuencias para variables cualitativas

### Tabla de frecuencias para el departamento de ocurrencia

##   
## Guatemala El Progreso Sacatepequez Chimaltenango Escuintla   
## 169944 9553 23698 41224 40117   
## Santa Rosa Solola Totonicapán Quetzaltenango Suchitepequez   
## 20129 22668 2337 50267 37426   
## Retalhuleu San Marcos Huehuetenango Quiche Baja Verapaz   
## 21188 54807 63128 53208 15752   
## Alta Verapaz Petén Izabal Zacapa Chiquimula   
## 58011 1656 18851 13471 18968   
## Jalapa Jutiapa Totonicapan Peten   
## 16303 26970 27091 23338

Como puede observarse, el departamento en el que suele haber más matrimonios es en Guatemala, seguido por Huehuetenango y luego por Alta Verapaz.

### Tabla de frecuencias para el municipio de ocurrencia

##   
## Guatemala Santa Catarina Pinula   
## 70183 3797   
## San José Pinula San José del Golfo   
## 4751 1117   
## Palencia Chinautla   
## 3234 4758   
## San Pedro Ayampuc Mixco   
## 2629 19566   
## San Pedro Sacatepéquez San Juan Sacatepéquez   
## 8561 12109   
## San Raymundo Chuarrancho   
## 2295 1034   
## Fraijanes Amatitlán   
## 2492 7371   
## Villa Nueva Villa Canales   
## 17955 7992   
## Petapa Guastatoya   
## 5455 1977   
## Morazán San Agustín Acasaguastlán   
## 655 1963   
## San Cristóbal Acasaguastlán El Jícaro   
## 339 699   
## Sansare Sanarate   
## 695 2382   
## San Antonio la Paz Antigua Guatemala   
## 843 4687   
## Jocotenango Pastores   
## 1096 922   
## Sumpango Santo Domingo Xenacoj   
## 2524 635   
## Santiago Sacatepéquez San Bartolomé Milpas Altas   
## 2231 482   
## San Lucas Sacatepéquez Santa Lucía Milpas Altas   
## 2131 1321   
## Magdalena Milpas Altas Santa María de Jesús   
## 721 1413   
## Ciudad Vieja San Miguel Dueñas   
## 2122 764   
## Alotenango San Antonio Aguas Calientes   
## 1695 727   
## Santa Catarina Barahona Chimaltenango   
## 226 7018   
## San José Poaquil San Martín Jilotepeque   
## 1783 4624   
## Comalapa Santa Apolonia   
## 3106 1367   
## Tecpán Guatemala Patzún   
## 6010 4379   
## Pochuta Patzicía   
## 513 2521   
## Santa Cruz Balanyá Acatenango   
## 676 1398   
## Yepocapa San Andrés Itzapa   
## 1672 2038   
## Parramos Zaragoza   
## 1039 1687   
## El Tejar Escuintla   
## 1393 7576   
## Santa Lucía Cotzumalguapa La Democracia   
## 6924 5123   
## Siquinalá Masagua   
## 1356 2068   
## Tiquisate La Gomera   
## 3785 3545   
## Guanagazapa San José   
## 879 3538   
## Iztapa Palín   
## 867 3638   
## San Vicente Pacaya Nueva Concepción   
## 954 3627   
## Cuilapa Barberena   
## 2472 3539   
## Santa Rosa de Lima Casillas   
## 850 954   
## San Rafael las Flores Oratorio   
## 534 1081   
## San Juan Tecuaco Chiquimulilla   
## 455 2754   
## Taxisco Santa María Ixhuatán   
## 1604 990   
## Guazacapán Santa Cruz Naranjo   
## 1072 816   
## Pueblo Nuevo Viñas Nueva Santa Rosa   
## 1254 1753   
## Sololá San José Chacayá   
## 5687 270   
## Santa María Visitación Santa Lucía Utatlán   
## 156 1522   
## Nahualá Santa Catarina Ixtahuacán   
## 3066 1862   
## Santa Clara la Laguna Concepción   
## 839 189   
## San Andrés Semetabaj Panajachel   
## 675 1229   
## Santa Catarina Palopó San Antonio Palopó   
## 162 425   
## San Lucas Tolimán Santa Cruz la Laguna   
## 1867 276   
## San Pablo la Laguna San Marcos la Laguna   
## 248 100   
## San Juan la Laguna San Pedro la Laguna   
## 607 655   
## Santiago Atitlán Totonicapán   
## 2832 9184   
## San Cristóbal Totonicapán San Francisco el Alto   
## 2751 3714   
## San Andrés Xecul Momostenango   
## 2076 8260   
## Santa María Chiquimula Santa Lucía la Reforma   
## 2362 486   
## San Bartolo Quetzaltenango   
## 594 12705   
## Salcajá Olintepeque   
## 1368 2031   
## San Carlos Sija Sibilia   
## 1567 566   
## Cabricán Cajolá   
## 1138 299   
## San Miguel Siguilá Ostuncalco   
## 346 3113   
## San Mateo Concepción Chiquirichapa   
## 366 889   
## San Martín Sacatepéquez Almolonga   
## 1322 1317   
## Cantel Huitán   
## 2877 747   
## Zunil Colomba   
## 1177 3361   
## San Francisco la Unión El Palmar   
## 367 1614   
## Coatepeque Génova   
## 7782 1583   
## Flores Costa Cuca La Esperanza   
## 1397 1408   
## Palestina de los Altos Mazatenango   
## 924 5927   
## Cuyotenango San Francisco Zapotitlán   
## 3359 1695   
## San Bernardino San José el Idolo   
## 891 726   
## Santo Domingo Suchitepéquez San Lorenzo   
## 1775 1793   
## Samayac San Pablo Jocopilas   
## 1611 2701   
## San Antonio Suchitepéquez San Miguel Panán   
## 3461 550   
## San Gabriel Chicacao   
## 360 3489   
## Patulul Santa Bárbara   
## 2179 2280   
## San Juan Bautista Santo Tomás la Unión   
## 401 1615   
## Zunilito Pueblo Nuevo   
## 991 653   
## Río Bravo Retalhuleu   
## 1801 6716   
## San Sebastián Santa Cruz Muluá   
## 2144 1062   
## San Martín Zapotitlán San Felipe   
## 879 1474   
## San Andrés Villa Seca Champerico   
## 2288 1874   
## Nuevo San Carlos El Asintal   
## 2353 2397   
## San Marcos San Antonio Sacatepéquez   
## 3625 1246   
## Comitancillo San Miguel Ixtahuacán   
## 2450 1855   
## Concepción Tutuapa Tacaná   
## 1884 3693   
## Sibinal Tajumulco   
## 680 1202   
## Tejutla San Rafael Pié de la Cuesta   
## 2082 1068   
## Nuevo Progreso El Tumbador   
## 1595 2256   
## El Rodeo Malacatán   
## 740 6153   
## Catarina Ayutla   
## 1861 2220   
## Ocós San Pablo   
## 1426 2083   
## El Quetzal La Reforma   
## 1917 1000   
## Pajapita Ixchiguán   
## 1590 1103   
## San José Ojetenán San Cristóbal Cucho   
## 986 1084   
## Sipacapa Esquipulas Palo Gordo   
## 918 692   
## Río Blanco Huehuetenango   
## 318 9416   
## Chiantla Malacatancito   
## 4679 1163   
## Cuilco Nentón   
## 2815 2164   
## San Pedro Necta Jacaltenango   
## 2307 2737   
## Soloma Ixtahuacán   
## 2802 1418   
## La Libertad San Miguel Acatán   
## 4697 1125   
## San Rafael la Independencia Todos Santos Cuchumatán   
## 719 1506   
## San Juan Atitán Santa Eulalia   
## 919 2222   
## San Mateo Ixtatán Colotenango   
## 1355 930   
## San Sebastián Huehuetenango Tectitán   
## 1558 417   
## Concepción Huista San Juan Ixcoy   
## 1174 889   
## San Antonio Huista San Sebastián Coatán   
## 1030 1488   
## Barillas Aguacatán   
## 6432 3152   
## San Rafael Petzal San Gaspar Ixchil   
## 294 272   
## Santiago Chimaltenango Santa Ana Huista   
## 462 466   
## Unión Cantinil Santa Cruz del Quiché   
## 746 6820   
## Chiché Chinique   
## 1737 920   
## Zacualpa Chajul   
## 1775 2548   
## Chichicastenango Patzité   
## 7921 488   
## San Antonio Ilotenango San Pedro Jocopilas   
## 989 1755   
## Cunén San Juan Cotzal   
## 2062 1380   
## Joyabaj Nebaj   
## 4189 4152   
## San Andrés Sajcabajá Uspantán   
## 1270 3728   
## Sacapulas San Bartolomé Jocotenango   
## 2916 625   
## Canillá Chicamán   
## 682 1929   
## Ixcán Pachalum   
## 4831 491   
## Salamá San Miguel Chicaj   
## 3765 1686   
## Rabinal Cubulco   
## 2724 2886   
## Granados El Chol   
## 626 793   
## San Jerónimo Purulhá   
## 1377 1895   
## Cobán Santa Cruz Verapaz   
## 10982 1797   
## San Cristóbal Verapaz Tactic   
## 4214 2541   
## Tamahú Tucurú   
## 884 1539   
## Panzós Senahú   
## 2123 2188   
## San Pedro Carchá San Juan Chamelco   
## 10909 3366   
## Lanquín Cahabón   
## 1334 2598   
## Chisec Chahal   
## 4442 1413   
## Fray Bartolomé de las Casas Santa Catalina la Tinta   
## 4532 1477   
## Raxruhá Flores   
## 1672 2386   
## San Benito San Andrés   
## 2678 1462   
## San Francisco Santa Ana   
## 780 504   
## Dolores San Luis   
## 1708 3616   
## Sayaxché Melchor de Mencos   
## 3437 1119   
## Poptún Las Cruces   
## 2902 1065   
## Puerto Barrios Livingston   
## 5714 2879   
## El Estor Morales   
## 2771 4697   
## Los Amates Zacapa   
## 2790 3891   
## Estanzuela Río Hondo   
## 944 1431   
## Gualán Teculután   
## 2626 1004   
## Usumatlán Cabañas   
## 559 601   
## San Diego La Unión   
## 434 1404   
## Huité Chiquimula   
## 360 6234   
## San José La Arada San Juan Ermita   
## 389 749   
## Jocotán Camotán   
## 2179 1523   
## Olopa Esquipulas   
## 1217 2442   
## Concepción Las Minas Quetzaltepeque   
## 597 1577   
## San Jacinto Ipala   
## 529 1532   
## Jalapa San Pedro Pinula   
## 7932 2431   
## San Luis Jilotepeque San Manuel Chaparrón   
## 1313 435   
## San Carlos Alzatate Monjas   
## 844 1464   
## Mataquescuintla Jutiapa   
## 1884 7829   
## El Progreso Santa Catarina Mita   
## 2005 1459   
## Agua Blanca Asunción Mita   
## 1061 2795   
## Yupiltepeque Atescatempa   
## 765 1058   
## Jerez El Adelanto   
## 191 434   
## Zapotitlán Comapa   
## 414 1796   
## Jalpatagua Conguaco   
## 1845 838   
## Moyuta Pasaco   
## 1585 483   
## San José Acatempa Quesada   
## 1084 1328   
## Extranjero Ignorado   
## 0 0   
## San José La Maquina La Blanca   
## 77 736   
## El Chal San Jorge   
## 394 217   
## 01-1 08-1   
## 5 1   
## 09-1 12-1   
## 1 2   
## Sipacate San José la Maquina   
## 91 831   
## Petatán 03-1   
## 103 1   
## 05-1 06-1   
## 1 1   
## 07-1 0928   
## 1 2   
## 11-1 13-1   
## 1 1

De la misma manera, es Guatemala el municipio que cuenta con la mayor cantidad de matrimonios.

### Tabla de frecuencias para la escolaridad del hombre

##   
## Básico Diversificado DIversificado Ignorado Ninguno   
## 120498 150209 39441 53223 119964   
## None Post Grado Postgrado Primaria Universitario   
## 62104 46 135 255480 29005

Como puede observarse, el nivel de escolaridad más común entre los hombres es primaria, con 255,480 casos, mientras que el más bajo es el de post grado con 181 casos.

### Tabla de frecuencias para la escolaridad de la Mujer

##   
## Básico Diversificado DIversificado Ignorado Ninguno   
## 98827 146175 37165 68360 170223   
## None Post Grado Postgrado Primaria Universitario   
## 62104 33 93 222076 25049

Como puede observarse, el nivel de escolaridad más común entre las mujeres es primaria, con 222076 casos, mientras que el más bajo es el de post grado con 126 casos.

### Tabla de frecuencias para el grupo étnico del hombre

##   
## Indigena No Indigena Ignorado No indigena   
## 101309 22152 69966 104340

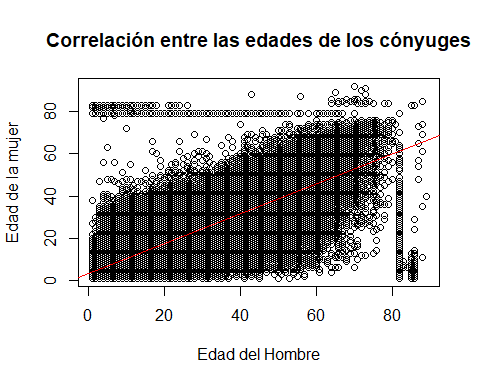
Coomo puede observarse, el grupo étnico más común al que pertenecen los hombres en la base de datos es no indígena.

### Tabla de frecuencias para el grupo étnico de la mujer

##   
## Indigena No Indigena Ignorado No indigena   
## 98014 21926 77158 100669

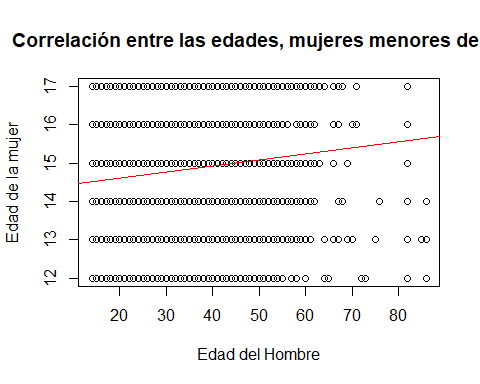
De la misma manera, puede observarse que el grupo étnico más común al que pertenecen las mujeres en la base de datos es no indígena.

## Correlacion entre edades de los cónyuges



Evidentementeno se puede observar correlación, sin embargo, es posible observar que hay hombres con edades muy altas con mujeres debajo de los 18 años,cabe mencionar que es importante filtrar los datos atipicos o norepresentativos del dataset

Como se puede ver, existen muchos datos extraños dentro del data set, estudiándolo se ha encontrado una gran cantidad de 1’s y otros datos en donde la edad de los “cónyuges” es incluso menor a 5 años, por lo que la presencia de estos datos nos ha llevado a buscar la naturaleza de las variables, por lo que revisando la lista de variables del INE de cada año, pudimos darnos cuenta que la edad más pequeña para mujeres durante los años fue 12 y de los hombres 14, por lo tanto se hizo un nuevo dataset para estudiar con precisión lo que deseamos estudiar. Se ha filtrado de la base de datos de matrimonios los valores en donde la edad de la mujer era menor a 18 (por el objetivo) y mayor a 12 años (para eliminar los datos no significativos o no lógicos), asimismo se filtró los datos en donde el hombre es menor a 14 años (de igual manera para eliminar esos datos confusos).



Filtrando los datos en donde las mujeres son menores a 18 años, es posible observar que una gran cantidad de mujeres se encuentran comprometidas con hombres mucho mayores a ellas, existen datos atípicos muy extensos y preocupantes, cabe mencionar que los datos que se encuentran en el fondo (muy pocos pero si existen) se encuentran así debido a que los datos no son indicados en el diagrama.

## Departamentos en donde se llevan a cabo los matrimonios de mujeres menores de edad

##   
## Guatemala El Progreso Sacatepequez Chimaltenango Escuintla   
## 9408 555 1150 1314 2018   
## Santa Rosa Solola Totonicapán Quetzaltenango Suchitepequez   
## 1221 865 82 1936 1321   
## Retalhuleu San Marcos Huehuetenango Quiche Baja Verapaz   
## 820 2194 2532 1891 512   
## Alta Verapaz Petén Izabal Zacapa Chiquimula   
## 2531 106 1304 880 1038   
## Jalapa Jutiapa Totonicapan Peten   
## 966 1718 803 1492

Puede observarse que muchos de estos matrimonios se llevan a cabo en el departamento de Guatemala, donde una gran población son ladinos.

## Municipios en donde se llevan a cabo los matrimonios de mujeres menores de edad

## depMun  
## Guatemala Santa Catarina Pinula   
## 4565 226   
## San José Pinula San José del Golfo   
## 233 67   
## Palencia Chinautla   
## 145 200   
## San Pedro Ayampuc Mixco   
## 113 1116   
## San Pedro Sacatepéquez San Juan Sacatepéquez   
## 288 452   
## San Raymundo Chuarrancho   
## 76 44   
## Fraijanes Amatitlán   
## 145 338   
## Villa Nueva Villa Canales   
## 938 357   
## Petapa Guastatoya   
## 286 115   
## Morazán San Agustín Acasaguastlán   
## 42 95   
## San Cristóbal Acasaguastlán El Jícaro   
## 21 42   
## Sansare Sanarate   
## 47 139   
## San Antonio la Paz Antigua Guatemala   
## 54 363   
## Jocotenango Pastores   
## 54 36   
## Sumpango Santo Domingo Xenacoj   
## 94 28   
## Santiago Sacatepéquez San Bartolomé Milpas Altas   
## 79 29   
## San Lucas Sacatepéquez Santa Lucía Milpas Altas   
## 144 79   
## Magdalena Milpas Altas Santa María de Jesús   
## 25 45   
## Ciudad Vieja San Miguel Dueñas   
## 65 27   
## Alotenango San Antonio Aguas Calientes   
## 40 29   
## Santa Catarina Barahona Chimaltenango   
## 13 275   
## San José Poaquil San Martín Jilotepeque   
## 69 169   
## Comalapa Santa Apolonia   
## 84 31   
## Tecpán Guatemala Patzún   
## 142 125   
## Pochuta Patzicía   
## 13 60   
## Santa Cruz Balanyá Acatenango   
## 25 43   
## Yepocapa San Andrés Itzapa   
## 73 48   
## Parramos Zaragoza   
## 42 49   
## El Tejar Escuintla   
## 66 349   
## Santa Lucía Cotzumalguapa La Democracia   
## 285 249   
## Siquinalá Masagua   
## 60 105   
## Tiquisate La Gomera   
## 231 197   
## Guanagazapa San José   
## 37 203   
## Iztapa Palín   
## 55 164   
## San Vicente Pacaya Nueva Concepción   
## 42 211   
## Cuilapa Barberena   
## 157 134   
## Santa Rosa de Lima Casillas   
## 47 85   
## San Rafael las Flores Oratorio   
## 34 67   
## San Juan Tecuaco Chiquimulilla   
## 40 169   
## Taxisco Santa María Ixhuatán   
## 107 64   
## Guazacapán Santa Cruz Naranjo   
## 55 55   
## Pueblo Nuevo Viñas Nueva Santa Rosa   
## 59 148   
## Sololá San José Chacayá   
## 198 12   
## Santa María Visitación Santa Lucía Utatlán   
## 4 43   
## Nahualá Santa Catarina Ixtahuacán   
## 94 42   
## Santa Clara la Laguna Concepción   
## 19 6   
## San Andrés Semetabaj Panajachel   
## 23 61   
## Santa Catarina Palopó San Antonio Palopó   
## 9 19   
## San Lucas Tolimán Santa Cruz la Laguna   
## 85 11   
## San Pablo la Laguna San Marcos la Laguna   
## 13 13   
## San Juan la Laguna San Pedro la Laguna   
## 30 43   
## Santiago Atitlán Totonicapán   
## 140 295   
## San Cristóbal Totonicapán San Francisco el Alto   
## 92 108   
## San Andrés Xecul Momostenango   
## 64 196   
## Santa María Chiquimula Santa Lucía la Reforma   
## 92 17   
## San Bartolo Quetzaltenango   
## 21 625   
## Salcajá Olintepeque   
## 72 48   
## San Carlos Sija Sibilia   
## 46 20   
## Cabricán Cajolá   
## 40 15   
## San Miguel Siguilá Ostuncalco   
## 5 109   
## San Mateo Concepción Chiquirichapa   
## 18 30   
## San Martín Sacatepéquez Almolonga   
## 30 48   
## Cantel Huitán   
## 78 23   
## Zunil Colomba   
## 33 99   
## San Francisco la Unión El Palmar   
## 12 78   
## Coatepeque Génova   
## 316 45   
## Flores Costa Cuca La Esperanza   
## 44 71   
## Palestina de los Altos Mazatenango   
## 31 260   
## Cuyotenango San Francisco Zapotitlán   
## 130 60   
## San Bernardino San José el Idolo   
## 29 26   
## Santo Domingo Suchitepéquez San Lorenzo   
## 85 53   
## Samayac San Pablo Jocopilas   
## 36 66   
## San Antonio Suchitepéquez San Miguel Panán   
## 106 13   
## San Gabriel Chicacao   
## 10 97   
## Patulul Santa Bárbara   
## 96 69   
## San Juan Bautista Santo Tomás la Unión   
## 11 55   
## Zunilito Pueblo Nuevo   
## 32 13   
## Río Bravo Retalhuleu   
## 87 294   
## San Sebastián Santa Cruz Muluá   
## 67 45   
## San Martín Zapotitlán San Felipe   
## 42 59   
## San Andrés Villa Seca Champerico   
## 82 81   
## Nuevo San Carlos El Asintal   
## 82 68   
## San Marcos San Antonio Sacatepéquez   
## 172 38   
## Comitancillo San Miguel Ixtahuacán   
## 90 79   
## Concepción Tutuapa Tacaná   
## 86 190   
## Sibinal Tajumulco   
## 31 54   
## Tejutla San Rafael Pié de la Cuesta   
## 56 52   
## Nuevo Progreso El Tumbador   
## 83 71   
## El Rodeo Malacatán   
## 24 236   
## Catarina Ayutla   
## 83 113   
## Ocós San Pablo   
## 47 80   
## El Quetzal La Reforma   
## 58 25   
## Pajapita Ixchiguán   
## 68 38   
## San José Ojetenán San Cristóbal Cucho   
## 49 16   
## Sipacapa Esquipulas Palo Gordo   
## 53 28   
## Río Blanco Huehuetenango   
## 17 399   
## Chiantla Malacatancito   
## 136 60   
## Cuilco Nentón   
## 132 84   
## San Pedro Necta Jacaltenango   
## 89 102   
## Soloma Ixtahuacán   
## 107 45   
## La Libertad San Miguel Acatán   
## 254 58   
## San Rafael la Independencia Todos Santos Cuchumatán   
## 46 67   
## San Juan Atitán Santa Eulalia   
## 44 72   
## San Mateo Ixtatán Colotenango   
## 77 31   
## San Sebastián Huehuetenango Tectitán   
## 75 27   
## Concepción Huista San Juan Ixcoy   
## 33 43   
## San Antonio Huista San Sebastián Coatán   
## 37 47   
## Barillas Aguacatán   
## 251 106   
## San Rafael Petzal San Gaspar Ixchil   
## 6 20   
## Santiago Chimaltenango Santa Ana Huista   
## 15 29   
## Unión Cantinil Santa Cruz del Quiché   
## 33 235   
## Chiché Chinique   
## 45 39   
## Zacualpa Chajul   
## 53 71   
## Chichicastenango Patzité   
## 269 9   
## San Antonio Ilotenango San Pedro Jocopilas   
## 25 37   
## Cunén San Juan Cotzal   
## 64 58   
## Joyabaj Nebaj   
## 172 148   
## San Andrés Sajcabajá Uspantán   
## 62 136   
## Sacapulas San Bartolomé Jocotenango   
## 62 15   
## Canillá Chicamán   
## 28 76   
## Ixcán Pachalum   
## 255 32   
## Salamá San Miguel Chicaj   
## 140 41   
## Rabinal Cubulco   
## 67 72   
## Granados El Chol   
## 34 34   
## San Jerónimo Purulhá   
## 66 58   
## Cobán Santa Cruz Verapaz   
## 498 79   
## San Cristóbal Verapaz Tactic   
## 132 93   
## Tamahú Tucurú   
## 25 57   
## Panzós Senahú   
## 128 109   
## San Pedro Carchá San Juan Chamelco   
## 463 127   
## Lanquín Cahabón   
## 78 133   
## Chisec Chahal   
## 168 85   
## Fray Bartolomé de las Casas Santa Catalina la Tinta   
## 199 83   
## Raxruhá Flores   
## 74 163   
## San Benito San Andrés   
## 164 88   
## San Francisco Santa Ana   
## 49 21   
## Dolores San Luis   
## 142 199   
## Sayaxché Melchor de Mencos   
## 194 73   
## Poptún Las Cruces   
## 213 65   
## Puerto Barrios Livingston   
## 391 235   
## El Estor Morales   
## 160 325   
## Los Amates Zacapa   
## 193 247   
## Estanzuela Río Hondo   
## 63 104   
## Gualán Teculután   
## 193 67   
## Usumatlán Cabañas   
## 35 34   
## San Diego La Unión   
## 33 71   
## Huité Chiquimula   
## 20 344   
## San José La Arada San Juan Ermita   
## 29 35   
## Jocotán Camotán   
## 91 70   
## Olopa Esquipulas   
## 36 160   
## Concepción Las Minas Quetzaltepeque   
## 54 91   
## San Jacinto Ipala   
## 30 98   
## Jalapa San Pedro Pinula   
## 474 96   
## San Luis Jilotepeque San Manuel Chaparrón   
## 80 29   
## San Carlos Alzatate Monjas   
## 52 96   
## Mataquescuintla Jutiapa   
## 139 448   
## El Progreso Santa Catarina Mita   
## 180 111   
## Agua Blanca Asunción Mita   
## 79 224   
## Yupiltepeque Atescatempa   
## 36 53   
## Jerez El Adelanto   
## 14 25   
## Zapotitlán Comapa   
## 25 82   
## Jalpatagua Conguaco   
## 124 53   
## Moyuta Pasaco   
## 117 33   
## San José Acatempa Quesada   
## 33 81   
## Extranjero Ignorado   
## 0 0   
## San José La Maquina La Blanca   
## 3 43   
## El Chal San Jorge   
## 29 13   
## 01-1 08-1   
## 0 0   
## 09-1 12-1   
## 0 0   
## Sipacate San José la Maquina   
## 8 43   
## Petatán 03-1   
## 1 0   
## 05-1 06-1   
## 0 0   
## 07-1 0928   
## 0 0   
## 11-1 13-1   
## 0 0

De la misma manera, puede observarse que la mayor cantidad de ocurrencias se encuentra en Guatemala.

## La etnia de la mujer en donde es más común este evento

## etniaMujer  
## Indigena No Indigena Ignorado No indigena   
## 3229 1053 3114 4898

Se puede observar que no hay una diferencia significativa en estos grupos, lo cual es importante para saber si el hecho de que ocurran los matrimonios con menores de edad, es algo que se ve influido por la cultura.

## La etnia del hombre en donde es más común este evento

## etniaHombre  
## Indigena No Indigena Ignorado No indigena   
## 3127 1058 3251 4858

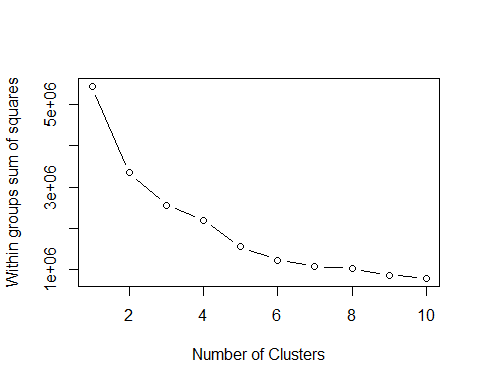
Se puede observar que no hay una diferencia significativa en estos grupos, lo cual es importante para saber si el hecho de que ocurran los matrimonios con menores de edad, es algo que se ve influido por la cultura.

# ———————————–Clustering——————————–

## AÑOREG EDADHOM EDADMUJ AÑOOCU   
## Min. :2009 Min. :14.00 Min. :12.00 Min. :1.000   
## 1st Qu.:2012 1st Qu.:15.00 1st Qu.:13.00 1st Qu.:4.000   
## Median :2015 Median :17.00 Median :15.00 Median :6.000   
## Mean :2014 Mean :18.81 Mean :14.59 Mean :5.976   
## 3rd Qu.:2017 3rd Qu.:20.00 3rd Qu.:16.00 3rd Qu.:9.000   
## Max. :2020 Max. :86.00 Max. :17.00 Max. :9.000

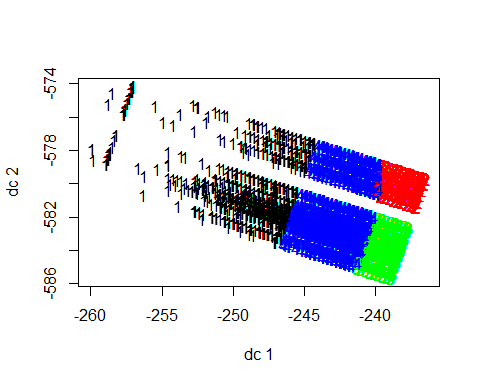
Se trabajará con las siguientes variables que tienen formato int para formas grupos: AÑOREG, ALOOCU, EDADHOM,EDADMUJ. Esto debido a que son variables que se pueden clasificar por grupos y los resultados pueden ser interpretados.

Para saber la mejor cantidad de clusters a utilizar, se hará lo siguiente:



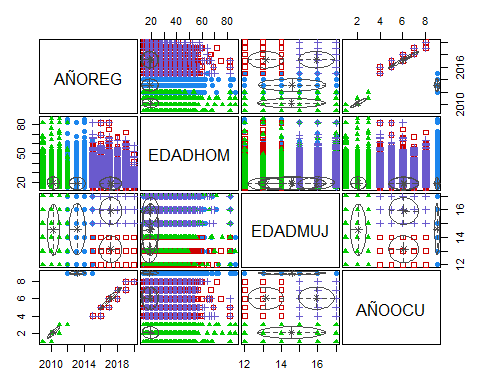
Puede observarse en el gráfico que el codo se encuentra en el valor 4, por lo que esta es la cantidad adecuada de clusters que se deben utilizar.

## Agrupamiento por medio de k-means



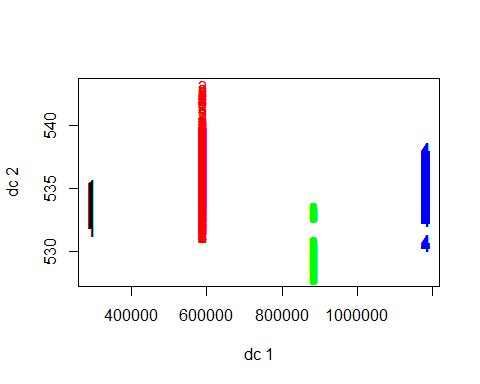
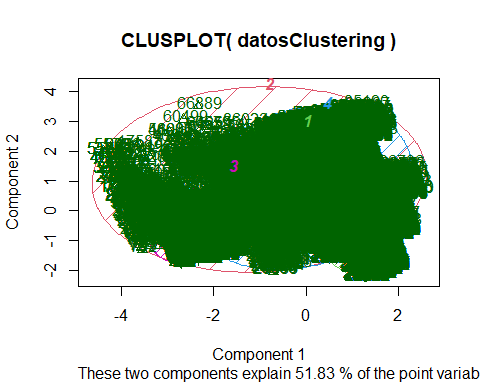
Se puede observar cómo los datos se encuentran divididos en 4 grupos, verde, rojo, azul y negr, en donde todos los grupos, excepto el azul, parecen estar bastante concentrados en un punto.

## Mixture of gaussians



En este gráfico se puede observar como en cada combinación de gráficos sucede la combinación de grupos y se observan los clusters bastante separados y bien identificados.

## Fuzzy C-means



En el primer gráfico, clusplot, se puede ver detrás las regiones y como se encuentran los datos, además se indica que se explica casi 64% del punto de variabilidad.

En el segundo gráfico se encuentran ya separados los clusters y es posible observar la variabilidad entre ellos.

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2009 Min. :40.00 Min. :12.00 Min. :1.00 Min. :1   
## 1st Qu.:2011 1st Qu.:42.00 1st Qu.:13.00 1st Qu.:3.00 1st Qu.:1   
## Median :2014 Median :47.00 Median :15.00 Median :6.00 Median :1   
## Mean :2014 Mean :52.38 Mean :14.67 Mean :5.47 Mean :1   
## 3rd Qu.:2016 3rd Qu.:55.00 3rd Qu.:16.00 3rd Qu.:9.00 3rd Qu.:1   
## Max. :2020 Max. :86.00 Max. :17.00 Max. :9.00 Max. :1

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2009 Min. :14.00 Min. :12.00 Min. :1.00 Min. :2   
## 1st Qu.:2010 1st Qu.:15.00 1st Qu.:13.00 1st Qu.:1.00 1st Qu.:2   
## Median :2010 Median :16.00 Median :15.00 Median :2.00 Median :2   
## Mean :2010 Mean :16.92 Mean :14.52 Mean :2.09 Mean :2   
## 3rd Qu.:2011 3rd Qu.:19.00 3rd Qu.:16.00 3rd Qu.:3.00 3rd Qu.:2   
## Max. :2012 Max. :24.00 Max. :17.00 Max. :3.00 Max. :2

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2012 Min. :14.00 Min. :12.00 Min. :4.00 Min. :3   
## 1st Qu.:2014 1st Qu.:15.00 1st Qu.:13.00 1st Qu.:6.00 1st Qu.:3   
## Median :2016 Median :16.00 Median :15.00 Median :7.00 Median :3   
## Mean :2016 Mean :16.56 Mean :14.54 Mean :7.13 Mean :3   
## 3rd Qu.:2018 3rd Qu.:18.00 3rd Qu.:16.00 3rd Qu.:9.00 3rd Qu.:3   
## Max. :2020 Max. :22.00 Max. :17.00 Max. :9.00 Max. :3

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2009 Min. :22.00 Min. :12.00 Min. :1.000 Min. :4   
## 1st Qu.:2013 1st Qu.:23.00 1st Qu.:13.00 1st Qu.:4.000 1st Qu.:4   
## Median :2015 Median :25.00 Median :15.00 Median :7.000 Median :4   
## Mean :2015 Mean :26.61 Mean :14.84 Mean :6.329 Mean :4   
## 3rd Qu.:2017 3rd Qu.:29.00 3rd Qu.:16.00 3rd Qu.:9.000 3rd Qu.:4   
## Max. :2020 Max. :39.00 Max. :17.00 Max. :9.000 Max. :4

Observando los resúmenes de cada cluster podemos darnos cuenta que en el clustering por kmeans podemos observar que una de las divisiones más evidentes se realizó por la edad del hombre, en donde el grupo 1 tiene los que se encuentran entre 14-22, el 2 entre 22-39, el 3 entre 40-86 y el 4 parece ser un complemento del 1 en donde se encuentran los de 14-24. Las edades de las mujeres de los grupos no parece ser un factor determinante al igual que el año de registro.

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2012 Min. :14.00 Min. :12.00 Min. :9 Min. :1.000   
## 1st Qu.:2012 1st Qu.:15.00 1st Qu.:13.00 1st Qu.:9 1st Qu.:3.000   
## Median :2013 Median :17.00 Median :15.00 Median :9 Median :3.000   
## Mean :2013 Mean :18.74 Mean :14.59 Mean :9 Mean :3.138   
## 3rd Qu.:2014 3rd Qu.:20.00 3rd Qu.:16.00 3rd Qu.:9 3rd Qu.:3.000   
## Max. :2015 Max. :86.00 Max. :17.00 Max. :9 Max. :4.000   
## mxGau   
## Min. :1   
## 1st Qu.:1   
## Median :1   
## Mean :1   
## 3rd Qu.:1   
## Max. :1

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2015 Min. :14.00 Min. :12.00 Min. :4.000 Min. :1.000   
## 1st Qu.:2016 1st Qu.:15.00 1st Qu.:12.00 1st Qu.:5.000 1st Qu.:3.000   
## Median :2017 Median :17.00 Median :13.00 Median :6.000 Median :3.000   
## Mean :2017 Mean :18.27 Mean :13.05 Mean :6.072 Mean :3.121   
## 3rd Qu.:2018 3rd Qu.:20.00 3rd Qu.:14.00 3rd Qu.:7.000 3rd Qu.:3.000   
## Max. :2020 Max. :82.00 Max. :14.00 Max. :8.000 Max. :4.000   
## mxGau   
## Min. :2   
## 1st Qu.:2   
## Median :2   
## Mean :2   
## 3rd Qu.:2   
## Max. :2

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2009 Min. :14.00 Min. :12.00 Min. :1.00 Min. :1.000   
## 1st Qu.:2010 1st Qu.:15.00 1st Qu.:13.00 1st Qu.:1.00 1st Qu.:2.000   
## Median :2010 Median :17.00 Median :15.00 Median :2.00 Median :2.000   
## Mean :2010 Mean :19.18 Mean :14.55 Mean :2.09 Mean :2.192   
## 3rd Qu.:2011 3rd Qu.:20.00 3rd Qu.:16.00 3rd Qu.:3.00 3rd Qu.:2.000   
## Max. :2012 Max. :86.00 Max. :17.00 Max. :3.00 Max. :4.000   
## mxGau   
## Min. :3   
## 1st Qu.:3   
## Median :3   
## Mean :3   
## 3rd Qu.:3   
## Max. :3

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2015 Min. :14.00 Min. :15.00 Min. :4.000 Min. :1.000   
## 1st Qu.:2016 1st Qu.:15.00 1st Qu.:15.00 1st Qu.:5.000 1st Qu.:3.000   
## Median :2017 Median :17.00 Median :16.00 Median :6.000 Median :3.000   
## Mean :2017 Mean :19.04 Mean :15.99 Mean :6.075 Mean :3.165   
## 3rd Qu.:2018 3rd Qu.:21.00 3rd Qu.:17.00 3rd Qu.:7.000 3rd Qu.:3.000   
## Max. :2020 Max. :82.00 Max. :17.00 Max. :8.000 Max. :4.000   
## mxGau   
## Min. :4   
## 1st Qu.:4   
## Median :4   
## Mean :4   
## 3rd Qu.:4   
## Max. :4

En este resúmen podemos identificar que se agrupo principalmente en relación al año, el grupo 1 se encuentran los años 2015-2020, 2 2012-2015, 3 2009,2012 y el 4to parece ser un complemento del 1ro. las edades parecen ser similares a través de los grupos.

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2012 Min. :14.00 Min. :12.00 Min. :4.000 Min. :3   
## 1st Qu.:2014 1st Qu.:14.00 1st Qu.:13.00 1st Qu.:6.000 1st Qu.:3   
## Median :2016 Median :16.00 Median :14.00 Median :7.000 Median :3   
## Mean :2016 Mean :15.69 Mean :14.45 Mean :7.095 Mean :3   
## 3rd Qu.:2018 3rd Qu.:17.00 3rd Qu.:16.00 3rd Qu.:9.000 3rd Qu.:3   
## Max. :2020 Max. :19.00 Max. :17.00 Max. :9.000 Max. :3   
## mxGau FCGrupos  
## Min. :1.000 Min. :1   
## 1st Qu.:1.000 1st Qu.:1   
## Median :2.000 Median :1   
## Mean :2.317 Mean :1   
## 3rd Qu.:4.000 3rd Qu.:1   
## Max. :4.000 Max. :1

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2009 Min. :30.00 Min. :12.00 Min. :1.000 Min. :1.000   
## 1st Qu.:2012 1st Qu.:32.00 1st Qu.:13.00 1st Qu.:3.000 1st Qu.:1.000   
## Median :2014 Median :35.00 Median :15.00 Median :6.000 Median :4.000   
## Mean :2014 Mean :39.52 Mean :14.73 Mean :5.816 Mean :3.033   
## 3rd Qu.:2017 3rd Qu.:42.00 3rd Qu.:16.00 3rd Qu.:9.000 3rd Qu.:4.000   
## Max. :2020 Max. :86.00 Max. :17.00 Max. :9.000 Max. :4.000   
## mxGau FCGrupos  
## Min. :1.000 Min. :2   
## 1st Qu.:1.000 1st Qu.:2   
## Median :3.000 Median :2   
## Mean :2.512 Mean :2   
## 3rd Qu.:4.000 3rd Qu.:2   
## Max. :4.000 Max. :2

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2009 Min. :14 Min. :12.00 Min. :1.000 Min. :2.00   
## 1st Qu.:2010 1st Qu.:15 1st Qu.:13.00 1st Qu.:1.000 1st Qu.:2.00   
## Median :2010 Median :16 Median :15.00 Median :2.000 Median :2.00   
## Mean :2010 Mean :17 Mean :14.53 Mean :2.073 Mean :2.03   
## 3rd Qu.:2011 3rd Qu.:19 3rd Qu.:16.00 3rd Qu.:3.000 3rd Qu.:2.00   
## Max. :2012 Max. :27 Max. :17.00 Max. :3.000 Max. :4.00   
## mxGau FCGrupos  
## Min. :3 Min. :3   
## 1st Qu.:3 1st Qu.:3   
## Median :3 Median :3   
## Mean :3 Mean :3   
## 3rd Qu.:3 3rd Qu.:3   
## Max. :3 Max. :3

## AÑOREG EDADHOM EDADMUJ AÑOOCU grupo   
## Min. :2009 Min. :18.00 Min. :12.00 Min. :1.00 Min. :2.00   
## 1st Qu.:2013 1st Qu.:20.00 1st Qu.:13.00 1st Qu.:5.00 1st Qu.:3.00   
## Median :2015 Median :21.00 Median :15.00 Median :7.00 Median :3.00   
## Mean :2015 Mean :22.03 Mean :14.87 Mean :6.91 Mean :3.46   
## 3rd Qu.:2017 3rd Qu.:24.00 3rd Qu.:16.00 3rd Qu.:9.00 3rd Qu.:4.00   
## Max. :2020 Max. :29.00 Max. :17.00 Max. :9.00 Max. :4.00   
## mxGau FCGrupos  
## Min. :1.000 Min. :4   
## 1st Qu.:1.000 1st Qu.:4   
## Median :2.000 Median :4   
## Mean :2.349 Mean :4   
## 3rd Qu.:4.000 3rd Qu.:4   
## Max. :4.000 Max. :4

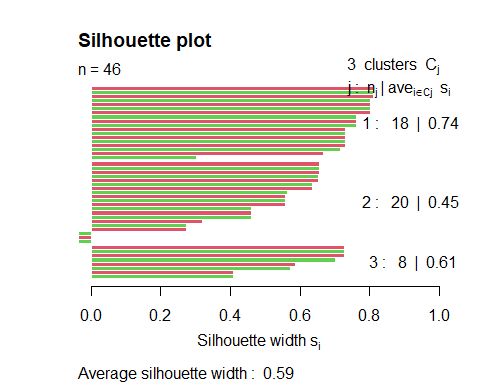
En el resúmer por fuzzy se puede identificar que la media de la edad de las mujeres parece ser constante, los años un poco variados pero en cierta manera se observa una relación entre año y edad del hombre.

## Evaluación de la tendencia de agrupamiento:

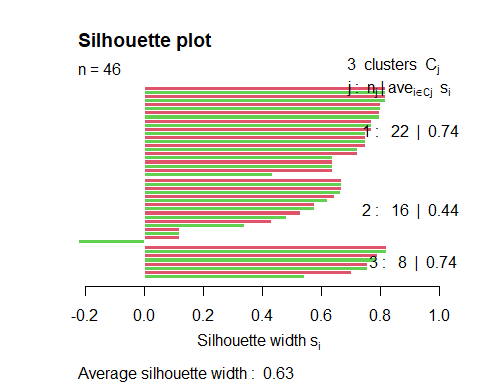
En esté análisis de datos consiste en examinar si los datos son agrupables, es decir, si los datos contienen cualquier estructura de agrupación inherente. Esto debe comprobarse antes de aplicar el análisis de agrupación.

Hopkins estadística: si el valor de Hopkins estadística está cerca de cero muy por debajo de 0,5, entonces podemos concluir que el conjunto de datos es significativamente clusterable. VAT evaluación visual de la tendencia de agrupamiento: la cubeta detecta la tendencia de agrupamiento en una forma visual contando el número de bloques oscuros o coloreados de forma cuadrada a lo largo de la diagonal en una imagen de rectangulos.

### En hombres



### En mujeres



## Conclusiones:

La ley guatemalteca, establece la mayoría de edad, como determinante de la libre aptitud para contraer matrimonio, no obstante, por medio de la simple autorización de los padres, tutores o padres adoptivos en su caso, o un Juez a través de la dispensa judicial, los menores de edad pueden contraer matrimonio civil, a los 16 años los varones, y a los 14 las mujeres. A través de los datos analizados, se han podido observar la gran cantida de matrimonios con ambos o un individuo menores de edad por ende es una realidad que se utiliza de manera recurrente la autorización de los padres para poder contraer matrimonio cuando aún no se ha cumplido la mayoría de edad. Esto no significa que la autorización de los padres sea cien por ciento veridica o no forzada.

Durante la experiencia en el análsis, ha dejado ver muchos casos en los que aunque uno, o los dos contrayentes son menores de edad, han logrado realizar los fines del matrimonio, especialmente en cuanto a permanencia, fidelidad y auxilio mutuo, por lo que no debe negárseles indiscriminadamente su derecho, pero también es cierto, que las épocas han cambiado, y en la actualidad, los matrimonios entre adolescentes son más propensos a la separación, o al divorcio al enfrentar situaciones adversas. Esto conlleva a analizar los posibles matrimonios infantiles y es que este se considera una forma de matrimonio forzoso, ya que no se cuenta con el consentimiento pleno, libre e informado de una de las partes o de ninguna de ellas.

Dentro de la información generada a partir de los datos del INE, sobre los años de las mujeres que han contraido matrimonio de 12 a 25 años de edad se observa que, en la categoría de casadas, los porcentajes más altos se encuentran en las mujeres del rango de 21 a 25 años, de igual manera en la categoría de alguna vez unida o casada, las mujeres en este rango de edad, representan más del 70%. Un 10% de quienes responden que están casadas son niñas y adolescentes del rango de 15 a 17 años, mientras que el 25.4% en esta categoría de casadas pertenece al rango de mujeres de 18 a 20 años. Se destaca que las más jóvenes, adolescentes de 12 a 14 años, representan un 3.5% de la categoría haber estado alguna vez casadas. Estos datos han variado a partir de las reformas impuestas en Reforma el artículo 81 del Código Civil, Decreto Ley 106, artículos: 81, relativo a la aptitud para contraer matrimonio y establece que la edad mínima para contraer matrimonio es de 18 años; 82, relativo a la excepción a la edad mínima y la establece en 16 años, cuando concurran razones fundadas; 84, relativo a la autorización judicial para que los menores de edad puedan contraer matrimonio; 177 en el que prohíbe la unión de menores de edad.Publicada el Lunes, 23 de noviembre del 2015.