

Mercado del alquiler

Diego Castillo

13/11/2020

```
library(dplyr)
```

```
##  
## 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
```

```
library(ggplot2)  
library(mice)
```

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

```
library(VIM)
```

```
## Loading required package: colorspace  
  
## Loading required package: grid  
  
## Loading required package: data.table  
  
##  
## Attaching package: 'data.table'  
  
## The following objects are masked from 'package:dplyr':  
##  
##   between, first, last
```

```
## VIM is ready to use.
## Since version 4.0.0 the GUI is in its own package VIMGUI.
##
## Please use the package to use the new (and old) GUI.

## Suggestions and bug-reports can be submitted at: https://github.com/alexkowa/VIM/issues

##
## Attaching package: 'VIM'

## The following object is masked from 'package:datasets':
##
## sleep

library(gridExtra)

##
## Attaching package: 'gridExtra'

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

Introducción

En el presente proyecto se realizará un análisis con el fin de encontrar las relaciones más importantes entre los atributos del mercado del alquiler de España y así resolver las preguntas de cuales son los patrones más importantes que tiene este mercado así como los sesgos del mismo. Para el desarrollo del proyecto de explotarán los datos de la Encuesta de Condiciones de Vida (INE, 2019), así como la información disponible del portal inmobiliario Idealista.

La Encuesta de Condiciones de Vida (ECV) es una operación estadística que se lleva a cabo de manera anual por el INE, cuyo objetivo es la producción sistemática de estadísticas comunitarias sobre la renta y las condiciones de vida de la población. Incluye datos transversales y longitudinales comparables y actualizados sobre la renta, el nivel y composición de la pobreza y la exclusión social, a escala nacional y europea.

Entre los diferentes aspectos a los cuales la ECV está orientada, es de especial relevancia para el presente proyecto los referentes a los ingresos y situación económica de los hogares, empleo, vivienda, movimientos migratorios y nivel de formación.

El portal inmobiliario Idealista es uno de los principales sitios web de anuncios de alquiler y venta de inmuebles, funciona en varios países europeo y tiene vasta presencia en España desde hace ya varios años. Idealista pone a disposición una API pública mediante la cual se puede realizar la consulta de la información del alquiler y venta según el lugar deseado.

En primer lugar, se realiza la extracción, tratamiento y exploración de los datos de la ECV.

Extracción de la información

Se realiza la extracción de la información correspondiente a la Encuesta de Condiciones de Vida (ECV) obtenida desde la web oficial del INE para el periodo 2019. El INE pone a disposición del público en general los datos de la ECV, los cuales están conformados por 4 ficheros en formato csv los cuales contienen información básica y detallada de los hogares y personas encuestadas. En el presente proyecto se realizará el análisis de los datos detallados de hogares y personas.

```
hogares <- read.csv('../data/data_collect/enc_condiciones_vida/esudb19h.csv', header= TRUE, sep= ',', s
personas <- read.csv('../data/data_collect/enc_condiciones_vida/esudb19p.csv', header= TRUE, sep= ',', s
hogaresBasic <- read.csv('../data/data_collect/enc_condiciones_vida/esudb19d.csv', header= TRUE, sep= '
```

Data cleaning

Selección de atributos

En un primer paso de limpieza de datos, se realiza la eliminación de aquellos atributos que no aportan o no son suficientemente relevantes para los análisis que se desean realizar en el presente proyecto.

```
# Eliminación de columnas del dataset de Hogares
AttrDelHog <- read.csv('../data/data_collect/enc_condiciones_vida/variables_eliminar_hogares.csv', head
hogaresClean <-hogares %>% select(-one_of(AttrDelHog$var))

# Eliminación de columnas del dataset de Personas
DelPer <- read.csv('../data/data_collect/enc_condiciones_vida/variables_eliminar_personas.csv', header=
personasClean <- personas %>% select(-one_of(DelPer$var))
```

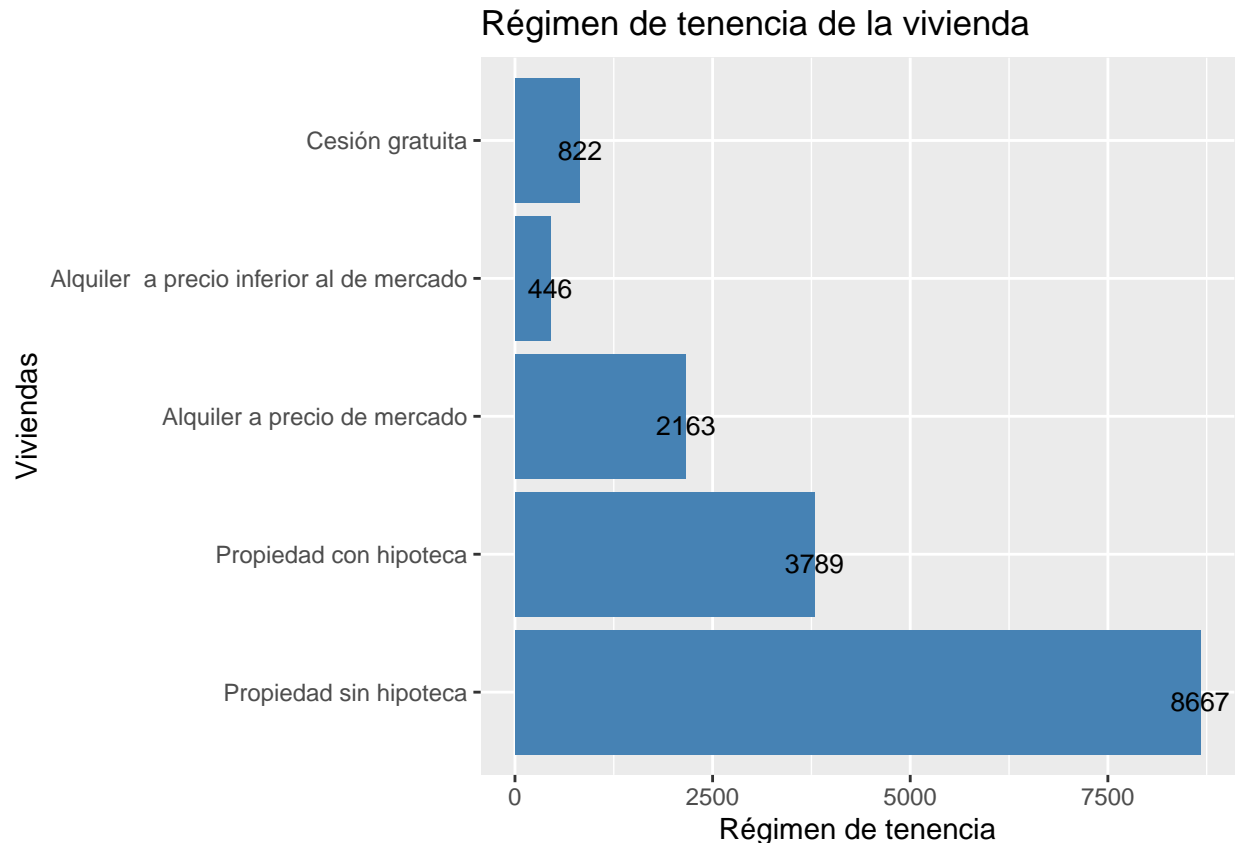
Análisis exploratorio de datos

Régimen de tenencia

Se verifica en primera instancia, el régimen de tenencia de las viviendas encuestadas, para determinar cual es el porcentaje de viviendas que tienen un régimen de alquiler.

```
# Se convierte la variable en Factor.
hogaresClean$HH021 <- factor(hogaresClean$HH021,levels = c(1,2,3,4,5),labels = c('Propiedad sin hipotec

# Grafico
g1.hh021 <- ggplot(hogaresClean, aes(HH021)) +
  geom_bar(fill="steelblue") +
  geom_text(aes(label=..count..), stat='count', vjust=1, color="black", size=3.5) +
  coord_flip() +
  ggtitle("Régimen de tenencia de la vivienda") +
  xlab("Viviendas") + ylab("Régimen de tenencia")
g1.hh021
```



El régimen de tenencia predominante es el de la propiedad con un total de 12.456 hogares encuestados que corresponde al 78.4% del total de la población de la encuesta. El motivo del presente análisis son aquellas viviendas que están en un régimen de alquiler, en donde se tiene que la tenencia de alquiler a precio de mercado se cuenta con un total de 2.163 registros que corresponde al 13,61% y para el régimen de tenencia de alquiler a precio inferior al mercado se tienen 446 registros que equivale al 2,8%.

El objetivo del presente proyecto es realizar el análisis de los hogares y las personas que residen en esos hogares en un régimen de tenencia de alquiler. No se toma en cuenta para el análisis los registros que corresponden a los hogares con un régimen de tenencia de alquiler a precio inferior al mercado, ya que dichos precios pueden afectar el valor real del mercado del alquiler.

```
hogaresAlquiler <- filter(hogaresClean, HH021 == 'Alquiler a precio de mercado')
```

Precio del alquiler

La variable HH060 es uno de los atributos más relevantes del dataset, ya que contiene el valor que las familias pagan por el alquiler de la vivienda en la que residen. Se verifica la existencia de valores nulos o perdidos

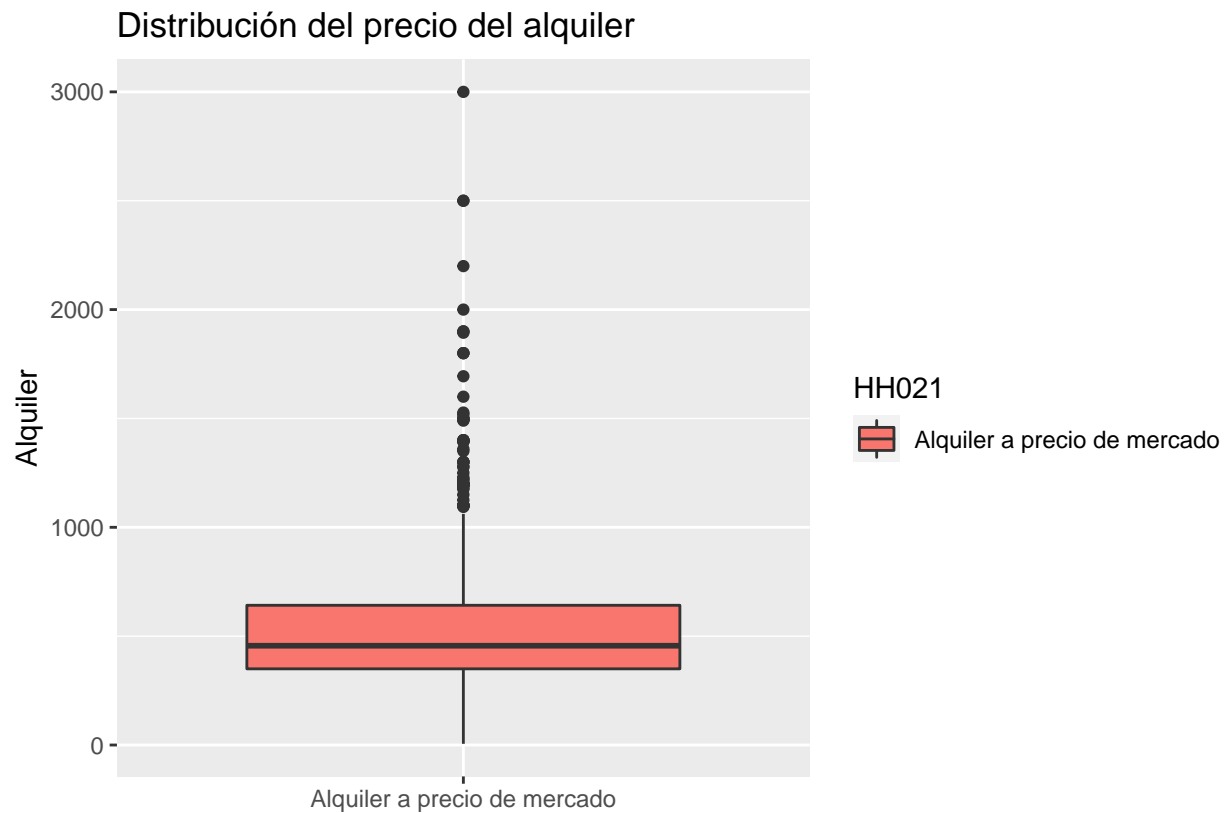
```
sum(is.na(hogaresAlquiler$HH060))
```

```
## [1] 32
```

Existen 32 valores vacíos o nulos para la variable HH060, se aplicará en este caso la imputación por ...

Se analiza a continuación la distribución de los valores del precio del alquiler de las viviendas mediante un diagrama de cajas.

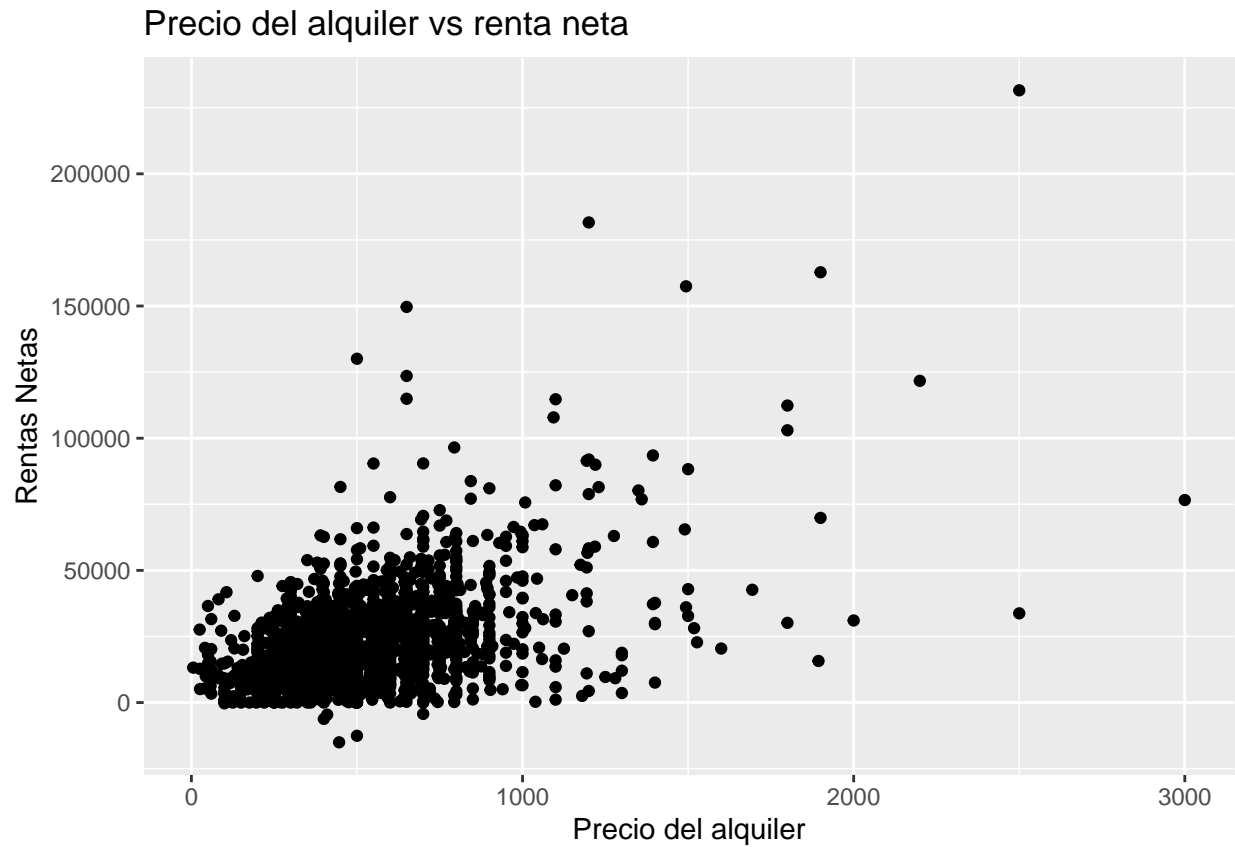
```
ggplot(data = subset(hogaresAlquiler, !is.na(HH060)), aes(x = HH021, y = HH060, fill = HH021)) + geom_boxplot()
ggtitle("Distribución del precio del alquiler") +
xlab("") + ylab("Alquiler")
```



Se observa que La mediana del valor a pagar por el alquiler de una vivienda corresponde a NA. Se puede observar que existen valores que superan el valor del tercer cuartil, por lo que se consideran como outliers, no se realizara la imputación de dichos valores ya que se entiende que existen ciertos casos en donde dependiendo de las características de las viviendas el valor del alquiler es muy elevado.

Precio del alquiler vs rentas netas de la vivienda

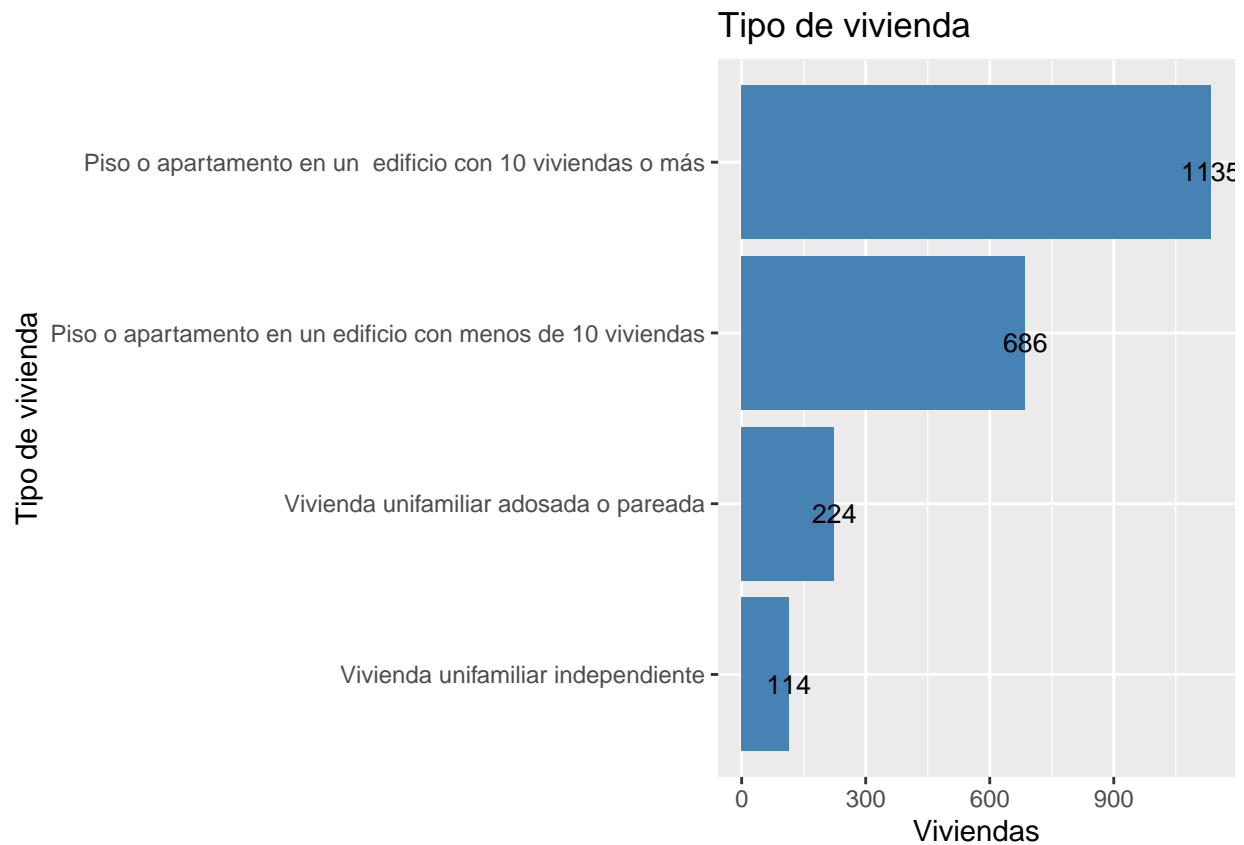
```
g1.rentasN <- ggplot(data = subset(hogaresAlquiler, !is.na(HH060))) +
  geom_point(aes(HH060, HY020)) +
  ggtitle("Precio del alquiler vs renta neta") +
  xlab("Precio del alquiler") + ylab("Rentas Netas")
g1.rentasN
```



Tipo de vivienda

```
# Se convierte la variable en Factor.
hogaresAlquiler$HH010 <- factor(hogaresAlquiler$HH010, levels = c(1,2,3,4), labels = c('Vivienda unifamil.', 'Vivienda multifamil.', 'Vivienda colectiva', 'Vivienda institucional'))

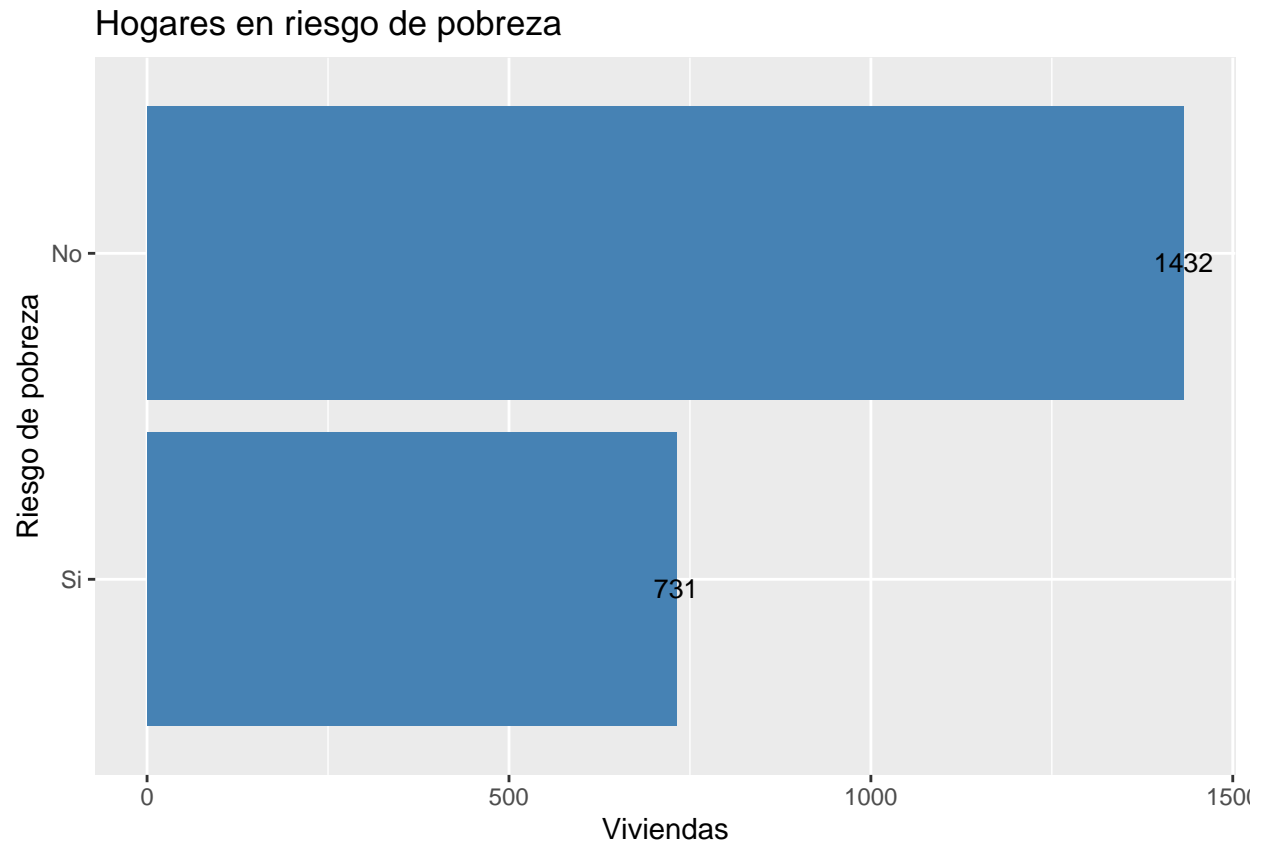
# Grafico
g.TipoViv <- ggplot(data = subset(hogaresAlquiler, !is.na(HH010)), aes(HH010)) +
  geom_bar(fill="steelblue") +
  geom_text(aes(label=..count..), stat='count', vjust=1, color="black", size=3.5) +
  coord_flip() +
  ggtitle("Tipo de vivienda") +
  xlab("Tipo de vivienda") + ylab("Viviendas")
g.TipoViv
```



Hogares en alquiler en riesgo de pobreza

```
hogaresAlquiler$vhPobreza <- factor(hogaresAlquiler$vhPobreza, levels = c(1,0), labels = c('Si', 'No'))

# Grafico
g.HogPob <- ggplot(data = subset(hogaresAlquiler, !is.na(vhPobreza)), aes(vhPobreza)) +
  geom_bar(fill="steelblue") +
  geom_text(aes(label=..count..), stat='count', vjust=1, color="black", size=3.5) +
  coord_flip() +
  ggtitle("Hogares en riesgo de pobreza") +
  xlab("Riesgo de pobreza") + ylab("Viviendas")
g.HogPob
```



Precio del alquiler vs Nacionalidad

Previo al análisis del precio del alquiler de viviendas por nacionalidad de los arrendatarios, se realiza la unión de los conjuntos de datos de hogares y personas, con el fin de lograr identificar a las personas que residen en una vivienda alquilada.

```
# Se genera la variable con el identificador de la vivienda
personasClean <- personasClean %>% mutate(HB030 = as.integer(substr(PB030,1,nchar(PB030)-2)))

# Se genera un nuevo dataset resultado del join entre personas y hogares y se filtran solo las personas
personasClean <- left_join(personasClean, hogaresAlquiler, by = 'HB030', copy = FALSE)
personasAlquiler <- filter(personasClean, HH021 == 'Alquiler a precio de mercado')

# Se genera el factor para verificar los labels de nacionalidad de las personas.
personasAlquiler$PB210 <- factor(personasAlquiler$PB210, levels = c(1,2,3), labels = c('España', 'Extranjero'))
```

En la ECV, las nacionalidad se agrupa en 3 categorías que son: personas nacidas en España, personas que nacieron en países miembros de la Unión Europea y personas nacidas en el resto de países del mundo. Se analiza la distribución del precio de las viviendas por nacionalidad así como el número de personas que viven bajo un régimen de tenencia de alquiler por nacionalidad.

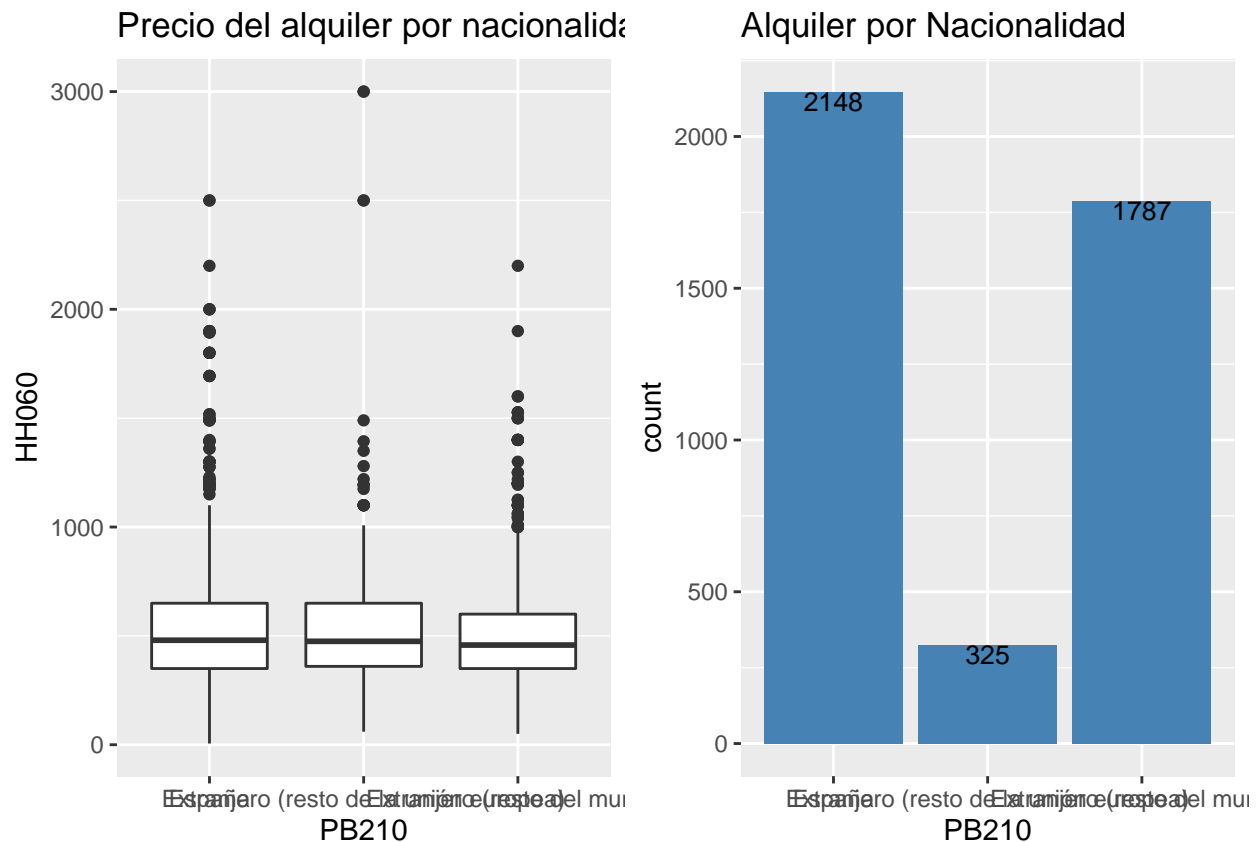
```
g.nacionalidad <- ggplot(data = subset(personasAlquiler, !is.na(PB210)), aes(PB210, HH060)) +
  geom_boxplot() +
  ggtitle('Precio del alquiler por nacionalidad')
```



```
g.nacionalidadCount <- ggplot(data = subset(personasAlquiler, !is.na(PB210)), aes(PB210)) +
  geom_bar(fill="steelblue") +
  geom_text(aes(label=..count..), stat='count', vjust=1, color="black", size=3.5) +
  ggtitle('Alquiler por Nacionalidad')

grid.arrange(g.nacionalidad, g.nacionalidadCount, ncol=2)
```

```
## Warning: Removed 53 rows containing non-finite values (stat_boxplot).
```



Se puede observar que para los 3 grupos de tipos de personas residentes se tiene una mediana similar, lo que indica que las personas que viven pagando un alquiler pagan casi lo mismo independientemente de su status migratorio. En el caso del número de personas, se tiene que las personas españolas cuentan con mas presencia en el mercado del alquiler con un total del 50,4%, en cambio las personas extranjeras que no pertenecen a países miembros de la Unión Europea cuentan con una proporción del 41,9% y por último, las personas extranjeras que vienen de países de la Unión Europea representan el 7,62%.

Precio del alquiler por el grupo de edad

En la ECV, encontramos el año de nacimiento de las personas encuestadas, a partir de dicho valor se calcula la edad a la fecha de la encuesta y se genera una nueva variable que contiene la edad de las personas agrupadas por las categorías: Menor de 30 años o mayor o igual a 30 años.

```
personasAlquiler <- personasAlquiler %>% mutate(edad = 2019 - PB140)
personasAlquiler <- personasAlquiler %>% mutate(grupos.edad = case_when(edad < 30 ~ 'Menor de 30',
                                                                           edad >= 30 ~ 'Mayor o igual a 30'))
```

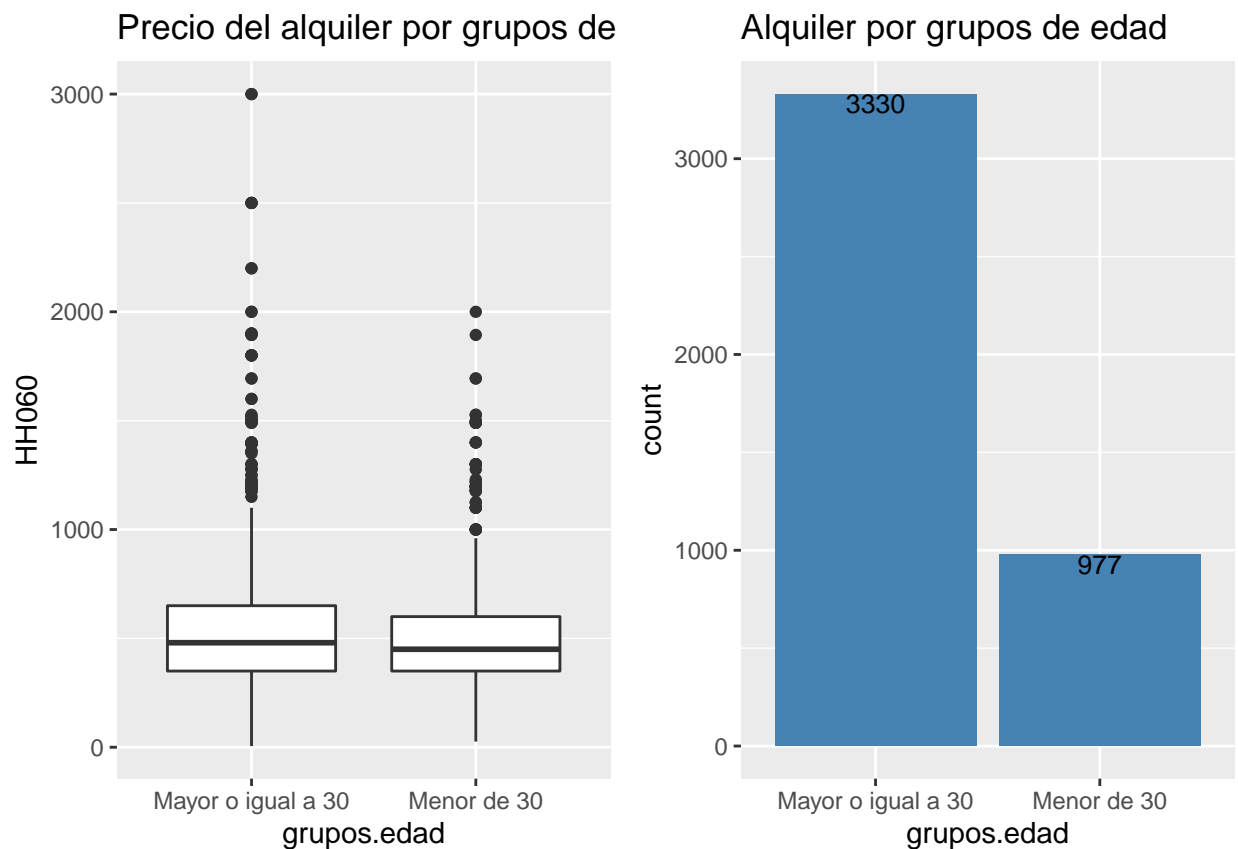
Se analiza mediante gráficos, cuál es la distribución del precio del alquiler de viviendas para las personas según su grupo de edad. Además, se analiza el número de personas por cada uno de estos grupos.

```
g1.Edad <- ggplot(data = subset(personasAlquiler, !is.na(grupos.edad)), aes(grupos.edad, HH060)) +
  geom_boxplot() +
  ggtitle('Precio del alquiler por grupos de edad')

g.edadCount <- ggplot(personasAlquiler, aes(grupos.edad)) +
  geom_bar(fill="steelblue") +
  geom_text(aes(label=..count..), stat='count', vjust=1, color="black", size=3.5) +
  ggtitle('Alquiler por grupos de edad')

grid.arrange(g1.Edad, g.edadCount, ncol=2)
```

Warning: Removed 57 rows containing non-finite values (stat_boxplot).



Se observa que independientemente de los grupos de edad, el precio del alquiler es similar, aunque en los hogares encuestados se observa una mayor presencia de personas con una edad igual o mayor de 30 años.

Alquiler por nivel de instrucción de las personas

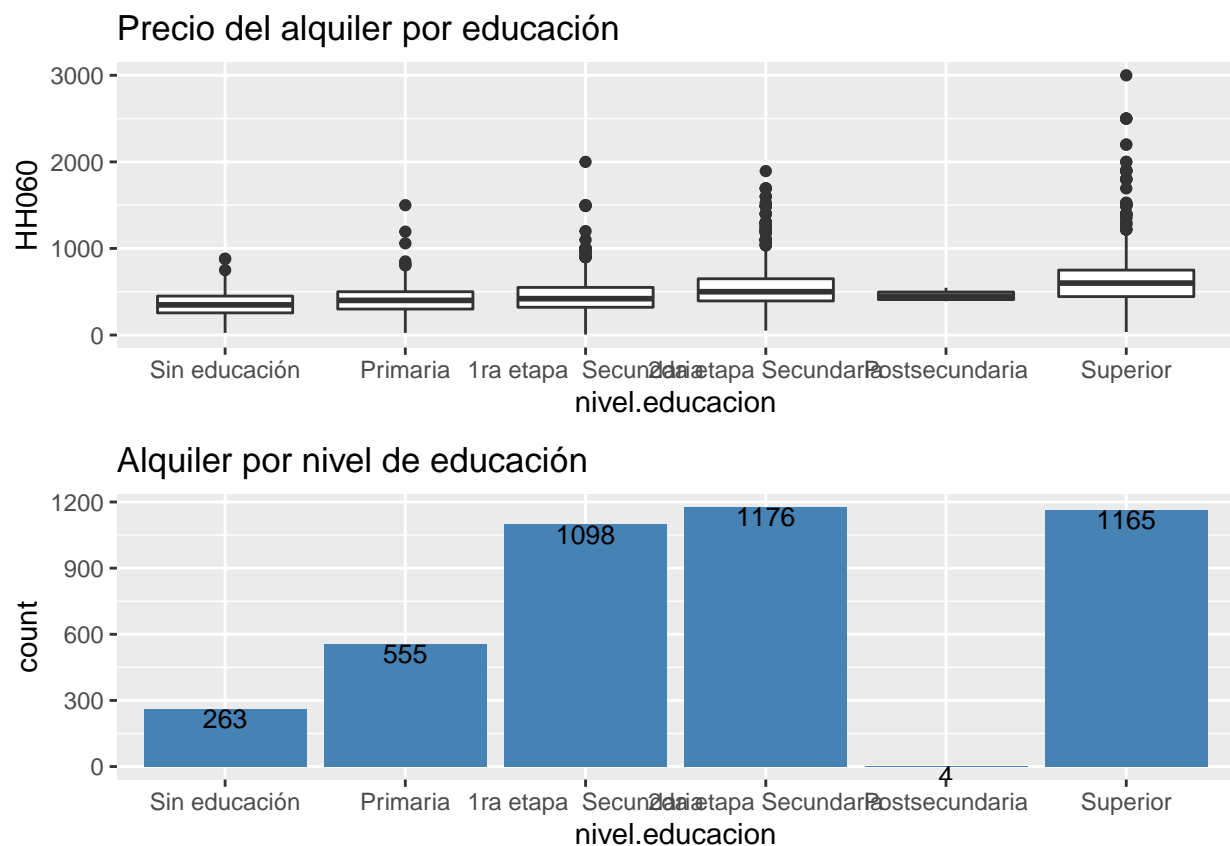
```
# Se genera el factor para verificar los labels de los niveles de instrucción de las personas
personasAlquiler <- personasAlquiler %>% mutate(nivel.educacion = substr(PE040,1,1))
personasAlquiler$nivel.educacion <- factor(personasAlquiler$nivel.educacion, levels = c('0','1','2','3'))
```

```
g.educacion <- ggplot(data = subset(personasAlquiler, !is.na(nivel.educacion)), aes(nivel.educacion, HH060)) +
  geom_boxplot() +
  ggtitle('Precio del alquiler por educación')

g.educacionCount <- ggplot(data = subset(personasAlquiler, !is.na(nivel.educacion)), aes(nivel.educacion, count)) +
  geom_bar(fill="steelblue") +
  geom_text(aes(label=..count..), stat='count', vjust=1, color="black", size=3.5) +
  ggtitle('Alquiler por nivel de educación')

grid.arrange(g.educacion, g.educacionCount, nrow=2)
```

Warning: Removed 53 rows containing non-finite values (stat_boxplot).



Se puede observar que el valor que pagan las personas por un alquiler, varía de acuerdo al nivel de educación que posean. Entre más alto el nivel de educación más pagan por el alquiler de una vivienda. Aunque se puede entender que, al poseer una persona un nivel de educación más alto puede también tener más ingresos que una persona que no cuenta con un nivel de educación y por lo tanto puede pagar más por un alquiler.

Número de habitantes por hogar

Se visualiza el número de personas mayores de 16 años que habitan en un hogar alquilado.

```
## [1] 3.120269
```

Se tiene una media de 3.1202693 personas residentes en viviendas con un régimen de tenencia de alquiler.

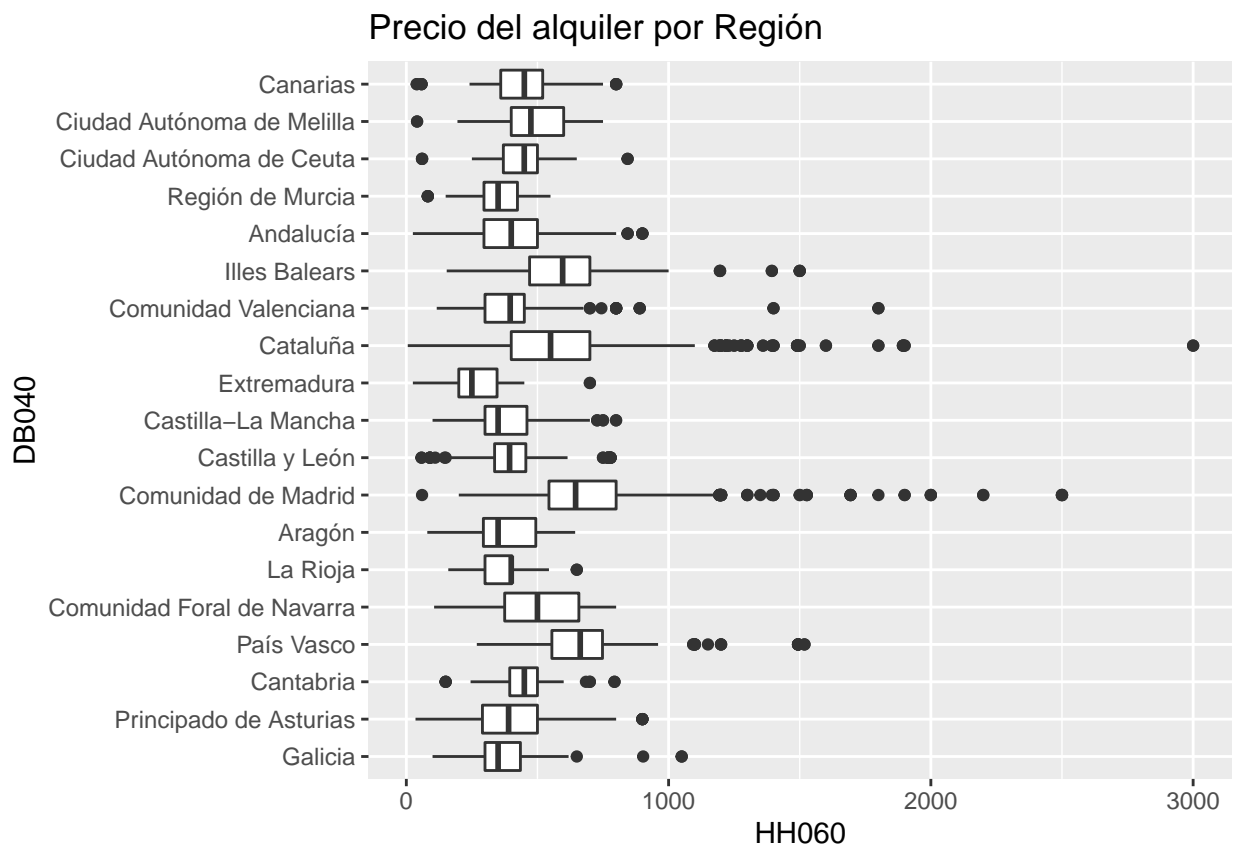
Alquiler por Región

```
# Se realiza un Join con el dataframe que contiene los datos de las regiones
hogaresBasic <- hogaresBasic %>% rename(HB030 = DB030)
personasAlquiler <- left_join(personasAlquiler, hogaresBasic, by = 'HB030', copy = FALSE)

# Se genera los nombres de las regiones para la presentación
personasAlquiler$DB040 <- factor(personasAlquiler$DB040, levels = c('ES11', 'ES12', 'ES13', 'ES21', 'ES22', 'ES23', 'ES24', 'ES25', 'ES26', 'ES27', 'ES28', 'ES29', 'ES30', 'ES31', 'ES32', 'ES33', 'ES34', 'ES35', 'ES36', 'ES37', 'ES38', 'ES39', 'ES40', 'ES41', 'ES42', 'ES43', 'ES44', 'ES45', 'ES46', 'ES47', 'ES48', 'ES49', 'ES50', 'ES51', 'ES52', 'ES53', 'ES54', 'ES55', 'ES56', 'ES57', 'ES58', 'ES59', 'ES60', 'ES61', 'ES62', 'ES63', 'ES64', 'ES65', 'ES66', 'ES67', 'ES68', 'ES69', 'ES70', 'ES71', 'ES72', 'ES73', 'ES74', 'ES75', 'ES76', 'ES77', 'ES78', 'ES79', 'ES80', 'ES81', 'ES82', 'ES83', 'ES84', 'ES85', 'ES86', 'ES87', 'ES88', 'ES89', 'ES90', 'ES91', 'ES92', 'ES93', 'ES94', 'ES95', 'ES96', 'ES97', 'ES98', 'ES99', 'ES00'))

g.Region <- ggplot(personasAlquiler, aes(DB040, HH060)) +
  geom_boxplot() +
  ggtitle('Precio del alquiler por Región') +
  coord_flip()
g.Region
```

```
## Warning: Removed 57 rows containing non-finite values (stat_boxplot).
```



Se observa que en algunas regiones como Comunidad de Madrid, Cataluña, Illes Balears la medianada del precio del alquiler es mayor que en el resto de Regiones de España.