

# rentHousesInfo

*Carlos A. García*

*November 19, 2019*

Llegim el document:

```
rentHousesInfo <- read.csv2("dataOutput.csv", header = TRUE, sep = ";", dec = ".")
summary(rentHousesInfo)
```

```
##   propertyCode
##   Min.      : 922769
##   1st Qu.:85678650
##   Median :87122480
##   Mean    :82582329
##   3rd Qu.:87567906
##   Max.    :87761745
##
##
##                                     thumbnail
##                                     : 10
## https://img3.idealista.com/blur/WEB_LISTING/0/id.pro.es.image.master/00/07/f4/729429492.jpg: 1
## https://img3.idealista.com/blur/WEB_LISTING/0/id.pro.es.image.master/00/26/6c/683326374.jpg: 1
## https://img3.idealista.com/blur/WEB_LISTING/0/id.pro.es.image.master/00/39/45/706380166.jpg: 1
## https://img3.idealista.com/blur/WEB_LISTING/0/id.pro.es.image.master/00/3c/77/729410670.jpg: 1
## https://img3.idealista.com/blur/WEB_LISTING/0/id.pro.es.image.master/00/5b/af/714633131.jpg: 1
## (Other) :3585
##      externalReference  numPhotos      floor      price
##                   : 611   Min.    : 0.00      :1337   Min.    : 350
## GC MASCARELL 1:    6   1st Qu.:11.00    1      : 600   1st Qu.: 1000
## 1082             :    3   Median :16.00    2      : 463   Median : 1450
## A-135            :    3   Mean    :19.01   bj      : 382   Mean    : 2085
## 059              :    2   3rd Qu.:24.00    3      : 333   3rd Qu.: 2300
## 1007             :    2   Max.    :92.00    4      : 191   Max.    :90000
## (Other)         :2973              (Other): 294
##      propertyType  operation      size      exterior
## chalet           : 885   rent:3600   Min.    : 22.0  False:1757
## countryHouse:    174              1st Qu.: 81.0  True :1843
## duplex           : 71              Median : 120.0
## flat             :2115              Mean    : 169.7
## penthouse        : 259              3rd Qu.: 183.0
## studio           : 96              Max.    :13700.0
##
##      rooms      bathrooms
##   Min.    : 0.000   Min.    : 1.000
##   1st Qu.: 2.000   1st Qu.: 1.000
##   Median : 3.000   Median : 2.000
##   Mean    : 2.806   Mean    : 2.125
##   3rd Qu.: 4.000   3rd Qu.: 2.000
##   Max.    :13.000   Max.    :100.000
##
##
##      address      province
## Santa Ponça      : 119   Balears (Illes):3600
## Portals Nous - Bendinat : 69
```

```

## barrio La Seu - Cort - Monti-Sion: 65
## barrio La Bonanova - Porto Pi : 64
## Marina Botafoc - Talamanca : 56
## barrio La Missio - Mercat : 47
## (Other) :3180
## municipality
## Palma de Mallorca :1380
## Calvià : 579
## Eivissa : 229
## Lluçmajor : 209
## Santa Eulalia del Río : 178
## Sant Josep de Sa Talaia: 175
## (Other) : 850
## district country latitude
## : 526 es:3600 Min. :38.67
## Ciutat Antigua : 313 1st Qu.:39.45
## Playa de Palma : 236 Median :39.54
## Genova - Bonanova - Sant Agustí : 219 Mean :39.44
## Sta Catalina - Son Armadans - Maritim: 205 3rd Qu.:39.57
## Santa Ponça : 149 Max. :40.06
## (Other) :1952
## longitude showAddress
## Min. :1.220 False:3064
## 1st Qu.:2.477 True : 536
## Median :2.632
## Mean :2.471
## 3rd Qu.:2.726
## Max. :4.292
##
## url hasVideo
## https://www.idealista.com/inmueble/1135092/ : 1 False:3486
## https://www.idealista.com/inmueble/1248620/ : 1 True : 114
## https://www.idealista.com/inmueble/1477283/ : 1
## https://www.idealista.com/inmueble/25344779/: 1
## https://www.idealista.com/inmueble/25543140/: 1
## https://www.idealista.com/inmueble/26348917/: 1
## (Other) :3594
## status newDevelopment hasLift
## : 674 False:3600 :1389
## good :2915 False: 724
## renew: 11 True :1487
##
##
##
##

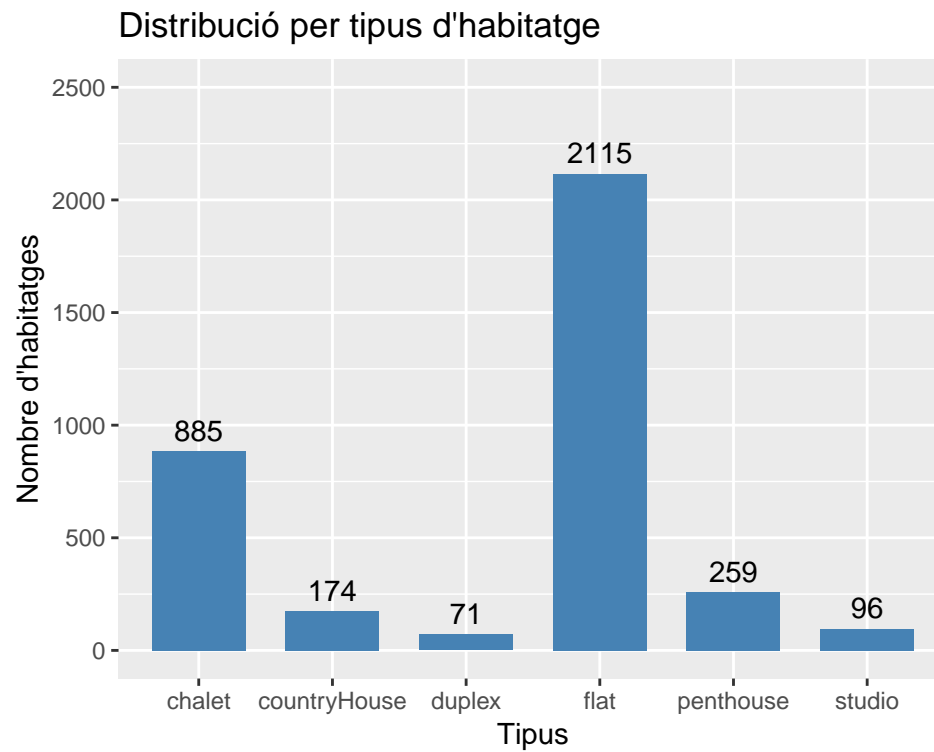
```

Mostrem la distribució de les variables categòriques:

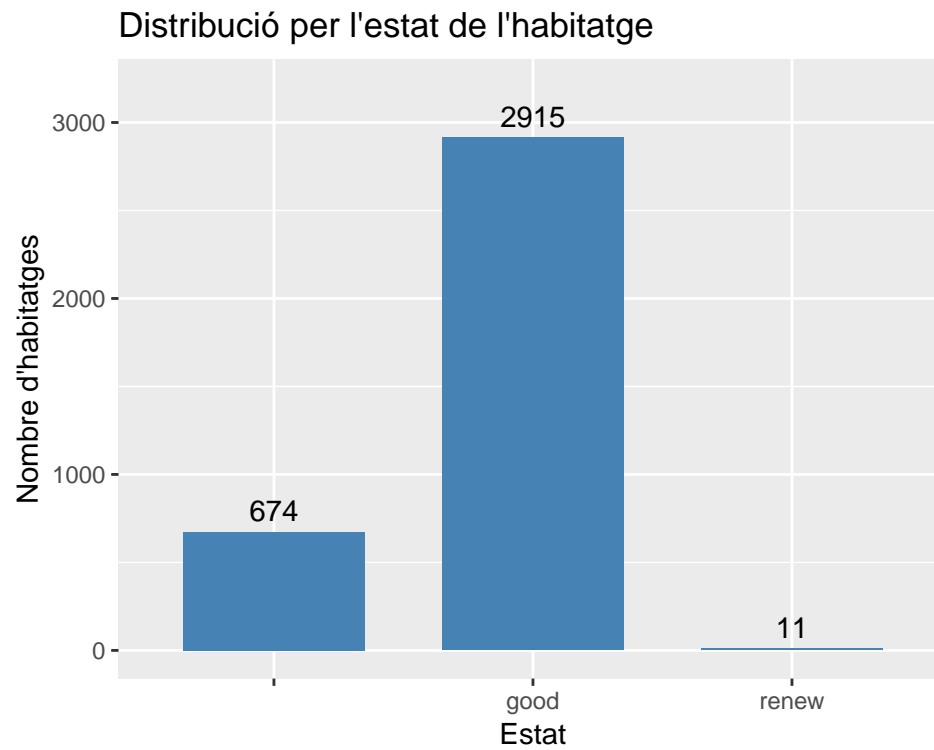
```

ggplot(rentHousesInfo, aes(x=factor(propertyType))) +
  labs(title="Distribució per tipus d'habitatge", x="Tipus", y = "Nombre d'habitatges")+
  geom_bar(stat="count", width=0.7, fill="steelblue") +
  ylim(0, 2500) +
  geom_text(stat="count", aes(label=..count..), vjust=-0.5)

```

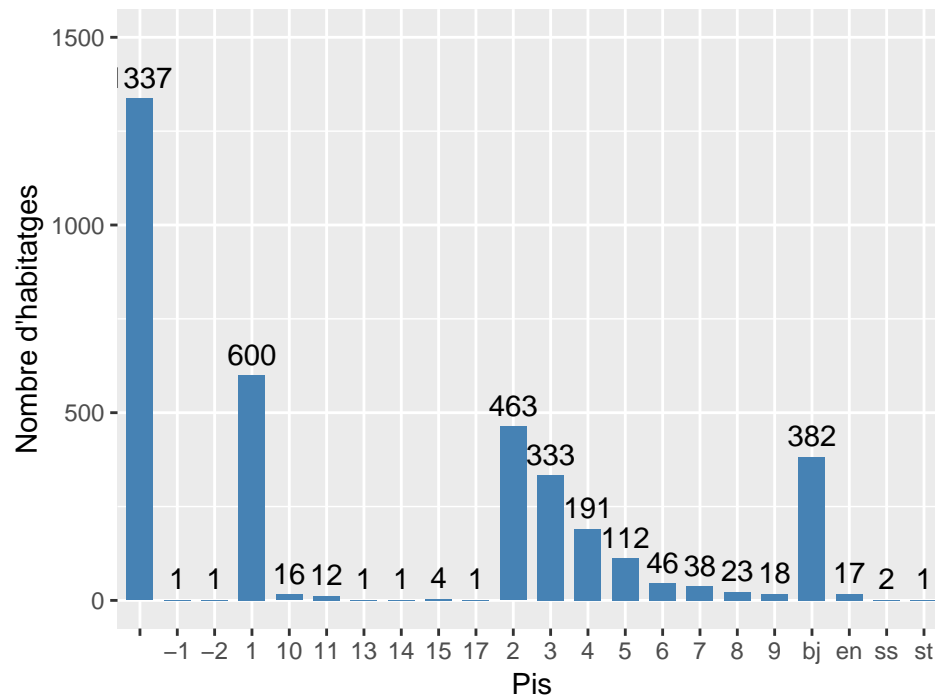


```
ggplot(rentHousesInfo, aes(x=factor(status))) +  
  labs(title="Distribució per l'estat de l'habitatge", x="Estat", y = "Nombre d'habitatges") +  
  geom_bar(stat="count", width=0.7, fill="steelblue") +  
  ylim(0, 3200) +  
  geom_text(stat="count", aes(label=..count..), vjust=-0.5)
```



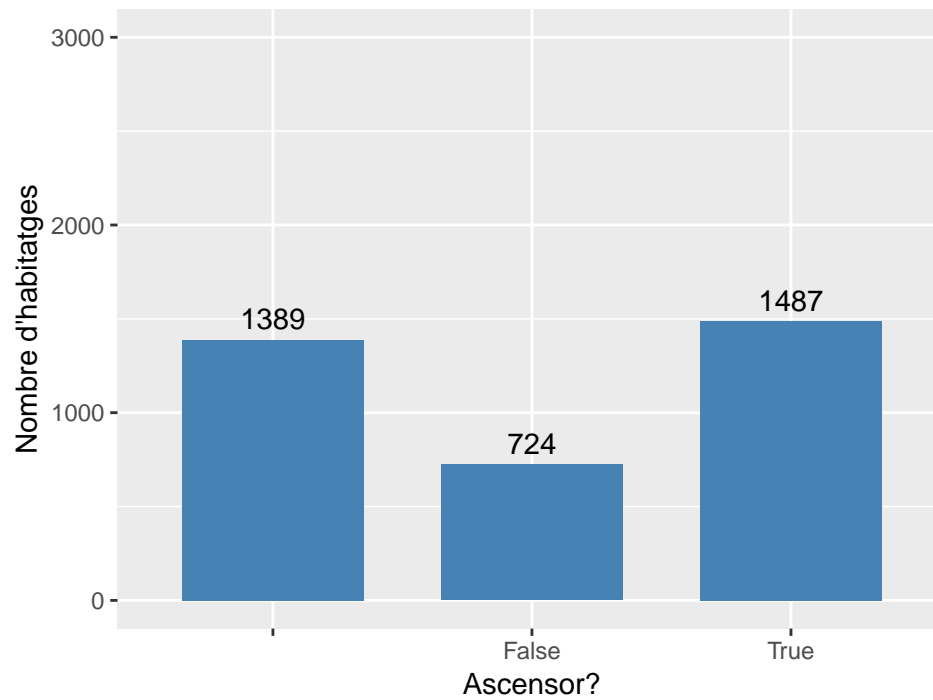
```
ggplot(rentHousesInfo, aes(x=factor(floor))) +  
  labs(title="Distribució pel pis de l'habitatge", x="Pis", y = "Nombre d'habitatges")+  
  geom_bar(stat="count", width=0.7, fill="steelblue") +  
  ylim(0, 1500) +  
  geom_text(stat="count", aes(label=..count..), vjust=-0.5)
```

Distribució pel pis de l'habitatge



```
ggplot(rentHousesInfo, aes(x=factor(hasLift))) +
  labs(title="Distribució depenent de si té ascensor", x="Ascensor?", y = "Nombre d'habitatges")+
  geom_bar(stat="count", width=0.7, fill="steelblue") +
  ylim(0, 3000) +
  geom_text(stat="count", aes(label=..count..), vjust=-0.5)
```

Distribució dependent de si té ascensor



```
ggplot(rentHousesInfo, aes(x=factor(municipality))) +  
  labs(title="Distribució d'habitatges per municipi", x="Municipi", y = "Nombre d'habitatges")+  
  geom_bar(stat="count", width=0.1, fill="steelblue") +  
  ylim(0, 1500) +  
  geom_text(stat="count", aes(label=..count..), vjust=0.5, hjust=-0.5, size=2) +  
  theme(axis.text.y = element_text(angle = 0, hjust = 0, size=8))+  
  coord_flip()
```

## Distribució d'habitatges per municipi



Si mostrem el nombre d'habitatges per preu, la gràfica no mostra res ja que hi ha massa dispersió. Llevant els outliers (preus superiors a la mitja + dues vegades la desviació estàndard) podem veure la distribució de la majoria dels lloguers.

```
rentHousesInfoPriceSD <- sd(rentHousesInfo$price)
rentHousesInfoPriceMean <- mean(rentHousesInfo$price)

rentHousesInfoWithoutOutliers <- filter(rentHousesInfo, price <= rentHousesInfoPriceMean + 2 * rentHousesInfoPriceSD)

ggplot() +
  xlab("Preu del lloguer") + ylab("Nombre de lloguers") + ggtitle("Nombre de lloguers per preu") +
  geom_density(data=rentHousesInfoWithoutOutliers, aes(x = price), fill="steelblue")
```

