

# Practica Final Parte 1

Xavier Vives, Josep Damià Ruiz, Antonio Marin y Martí Paredes

## Contents

1. Cargad en un dataframe los datos del fichero listings.csv y construid un nuevo data frame . . . . .	2
2. Renombrar las variables al castellano. . . . .	2
3. Calcular mínimo, máximo, media, varianza, cuartiles y mediana. . . . .	2
4. Tablas de frecuencias absolutas. . . . .	3
5. Boxplots y histogramas. . . . .	5
Pie chart de variables no numéricas . . . . .	18
9. Diagramas de dispersión de 4 variables numéricas 2 a 2, con un color para cada Vecindario. . . . .	19
10. Coeficientes de correlación 2 a 2 apartado anterior. . . . .	21

1. Cargad en un dataframe los datos del fichero listings.csv y construid un nuevo data frame

```
datos_raw = read.csv("listings.csv")
datos = datos_raw[-c(1,2,3,5,14,15)]
```

2. Renombrar las variables al castellano.

```
names(datos)[names(datos) == "host_name"] <- "Nombre_propietario"
names(datos)[names(datos) == "neighbourhood"] <- "Vecindario"
names(datos)[names(datos) == "latitude"] <- "Lat."
names(datos)[names(datos) == "longitude"] <- "Long."
names(datos)[names(datos) == "room_type"] <- "Tipo_habitación"
names(datos)[names(datos) == "price"] <- "Precio"
names(datos)[names(datos) == "minimum_nights"] <- "Min.noches"
names(datos)[names(datos) == "number_of_reviews"] <- "N_reseñas"
names(datos)[names(datos) == "number_of_reviews_ltm"] <- "Reseñas_mes"
names(datos)[names(datos) == "availability_365"] <- "Disponibilidad_año"

datos<- subset(datos, datos$Precio<1000)
datos<- subset(datos, datos$Min.noches<400)

names(datos)

## [1] "Nombre_propietario" "Vecindario"           "Lat."
## [4] "Long."                "Tipo_habitación"      "Precio"
## [7] "Min.noches"            "N_reseñas"            "Reseñas_mes"
## [10] "Disponibilidad_año"
```

3. Calcular mínimo, máximo, media, varianza, cuartiles y mediana.

```
media = unname(sapply(datos[c(3,4,6,7,8,9)],FUN=mean))
varianza = unname(sapply(datos[c(3,4,6,7,8,9)],FUN=var))
cuartiles = sapply(datos[c(3,4,6,7,8,9)],FUN=quantile)
rownames(cuartiles)<-c("mínimo","Cuartil_1","Mediana","Cuartil_3","máximo"))
datosEstadisticos=media
datosEstadisticos<-rbind(datosEstadisticos, media)
datosEstadisticos<-rbind(datosEstadisticos, varianza)
datosEstadisticos<-rbind(datosEstadisticos,cuartiles)
datosEstadisticos<-round(datosEstadisticos, digits=4)
datosEstadisticos

##                  Lat.    Long.     Precio Min.noches N_reseñas Reseñas_mes
## datosEstadisticos 35.9368 14.4326   78.9683     3.9576   19.3276    2.0872
## media             35.9368 14.4326   78.9683     3.9576   19.3276    2.0872
## varianza          0.0037  0.0098 5660.3554   224.0863 1215.0219   18.3423
## mínimo            35.8133 14.1954    7.0000     1.0000    0.0000    0.0000
## Cuartil_1         35.8992 14.3723   35.0000     1.0000    0.0000    0.0000
```

```

## Mediana      35.9156 14.4840   59.0000    2.0000   4.0000   0.0000
## Cuartil_3   35.9529 14.4994   95.0000    3.0000  22.0000   2.0000
## màximo       36.0802 14.5780  850.0000  365.0000 406.0000  69.0000

```

#### 4. Tablas de frecuencias absolutas.

```

frecuenciaVecinador = table(datos$Vecindario)
frecuenciaNombre = table(datos$Nombre_propietario)
frecuenciaNombre <- subset(frecuenciaNombre,frecuenciaNombre > 20)
frecuenciaNombre = sort(frecuenciaNombre, decreasing = TRUE)
frecuenciaTipo = table(datos$Tipo_habitación)
frecuenciaVecinador

```

	Attard	Balzan	Birgu	Birkirkara
##	17	19	82	134
##	Birzebbugia	Bormla	Dingli	Fgura
##	64	126	2	7
##	Floriana	Fontana	Ghajnsielem	Gharb
##	127	22	105	118
##	Għargħur	Għasri	Għaxaq	Gudja
##	17	80	11	18
##	Gżira	Hamrun	Iklin	Isla
##	360	27	9	79
##	Kalkara	Kercem	Kirkop	Lija
##	21	63	8	11
##	Luqa	Marsa	Marsascala	Marsaxlokk
##	12	6	296	59
##	Mdina	Mellieha	Mgarr	Mosta
##	13	462	65	83
##	Mqabba	Msida	Mtarfa	Munxar
##	3	341	5	145
##	Nadur	Naxxar	Paola	Pembroke
##	122	103	21	78
##	Pieta	Qala	Qormi	Qrendi
##	81	138	44	13
##	Rabat (Malta)	Rabat (Victoria)	Safi	San Giljan
##	113	116	7	833
##	San Gwann	San Lawrenz	San Pawl il-Bahar	Sannat
##	188	51	837	46
##	Santa Lucija	Santa Venera	Siggiewi	Sliema
##	1	23	25	925
##	Swieqi	Ta' Xbiex	Tarxien	Valletta
##	396	50	32	467
##	Xaghra	Xewkija	Xghajra	Zabbar
##	249	52	21	28
##	Zebbug (Għawdex)	Zebbug (Malta)	Zejtun	Zurrieq
##	313	51	41	25

```
frecuenciaNombre
```

```
##
```

##	Joseph	Short Lets Malta
##	131	126
##	Maria	360 Estates
##	108	107
##	Victor	Baron
##	94	92
##	Robert	Mark
##	88	80
##	Michael	Paul
##	70	68
##	GetawaysMalta	James
##	61	58
##	Andrea	Matthew
##	53	52
##	Simon	Chris
##	49	47
##	Karl	John
##	46	43
##	Mario	Ian
##	42	40
##	Jeffrey	Aaron
##	40	39
##	Mariella	Daniel
##	39	37
##	David	George
##	36	36
##	GetawaysMalta Neville	Carmen
##	36	35
##	Caroline	Rita
##	35	35
##	Barbara	Joe
##	34	33
##	Stefan	Alex
##	32	31
##	C'Est La Vie	Casa Rooms
##	31	31
##	Jonathan	Kevin
##	30	30
##	Vanesa	Med Malta
##	30	29
##	Mary	Thomas
##	28	28
##	Lisa	Peter
##	27	27
##	Adelbert	Amr
##	25	25
##	Andrew Paul - HolidayLetsMaltaGozo	
##	25	25
##	Ray	Alison
##	25	24
##	Anna	Josette
##	24	24
##	Ryan	Martin
##	24	23

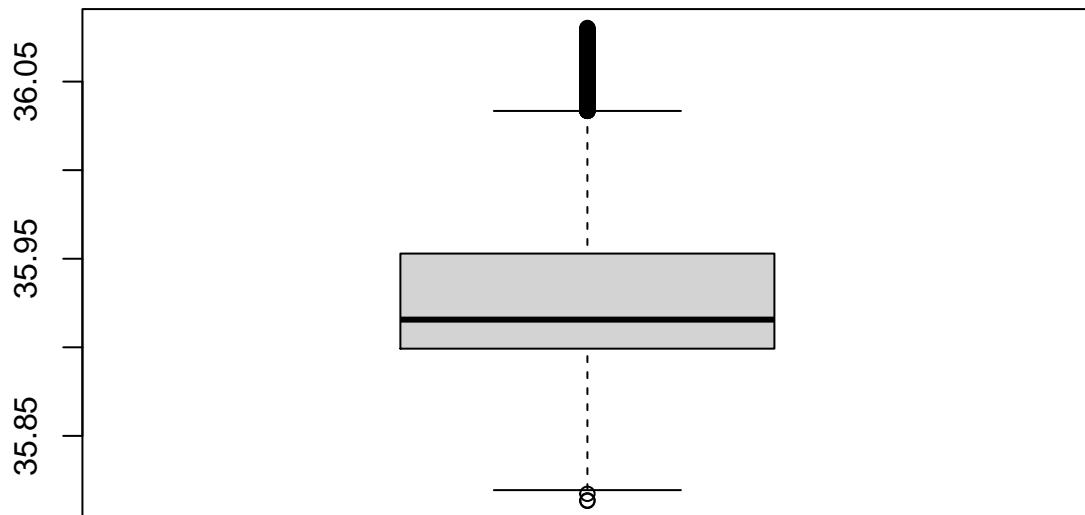
```
##           Richard          Francesca
##             23                  22
##           Frank          Gozo Village
##             22                  22
##           22
##           Stephanie        Alfred
##           22                  21
##           Chiara            Jorge
##             21                  21
##           Lorraine       Short Lets
##             21                  21
```

```
frecuenciaTipo
```

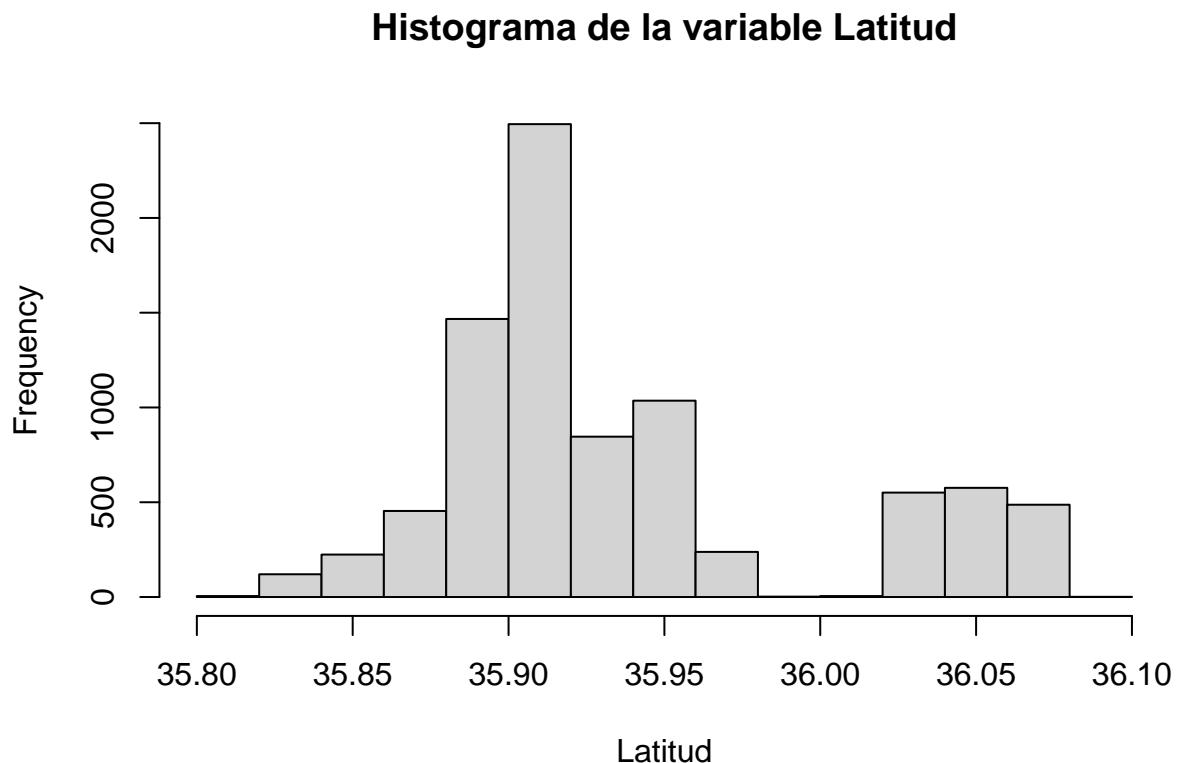
```
##
## Entire home/apt      Hotel room      Private room      Shared room
##      5268                 182              2739                318
```

## 5. Boxplots y histogramas.

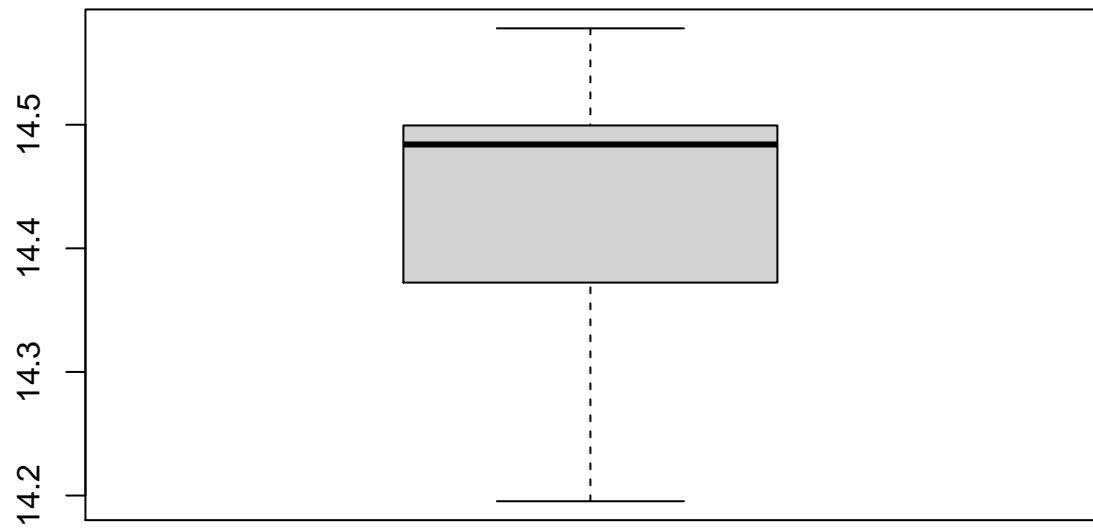
```
boxplot(datos$Lat.)
```



```
hist(datos$Lat., main= "Histograma de la variable Latitud", xlab = "Latitud")
```

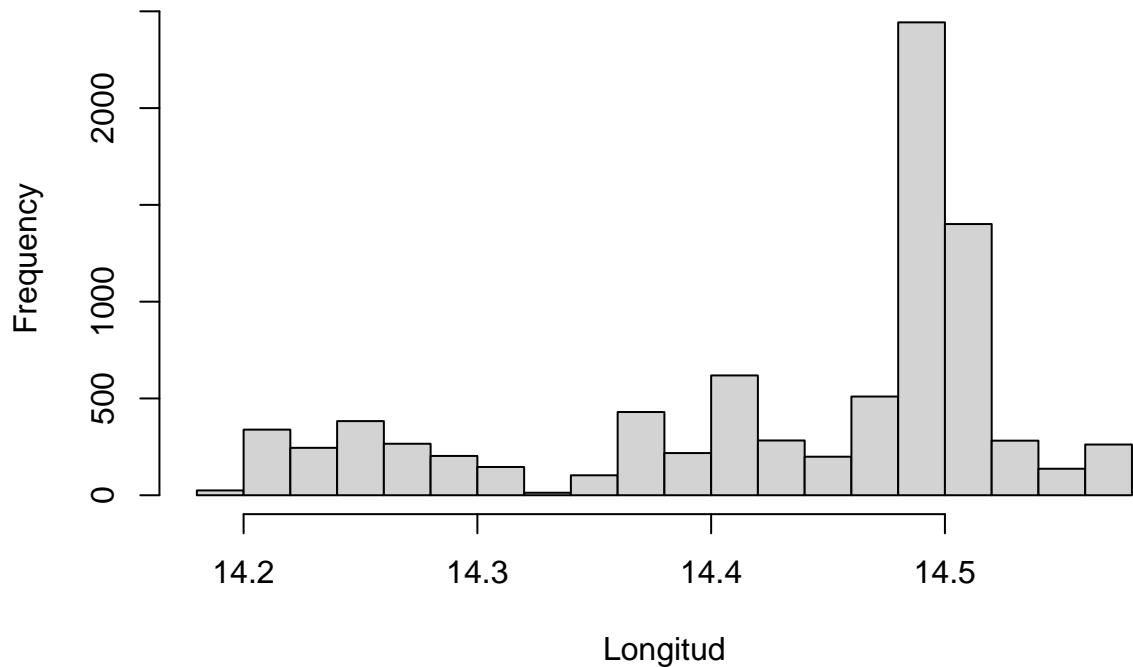


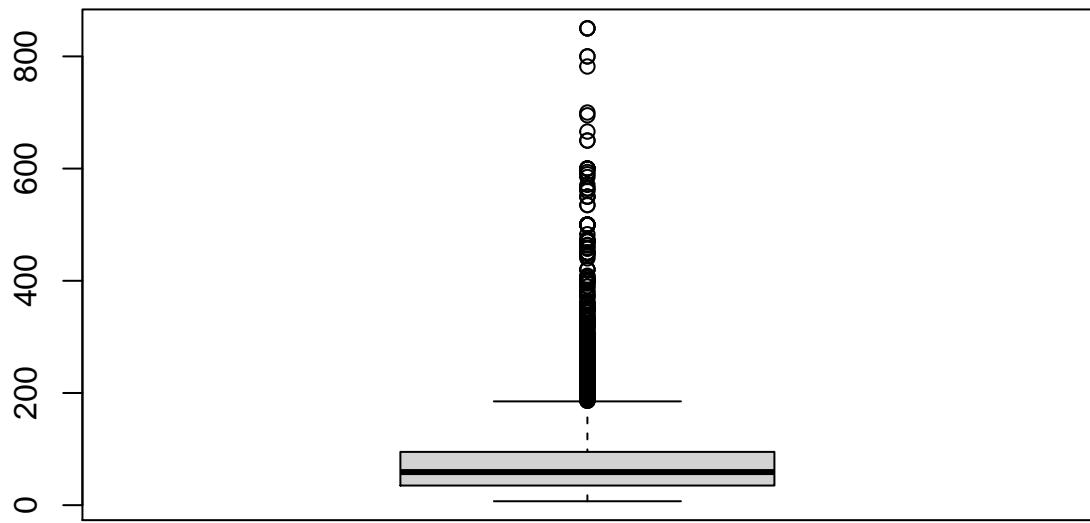
```
boxplot(datos$Long.)
```



```
hist(datos$Long., main= "Histograma de la variable Longitud", xlab = "Longitud")
```

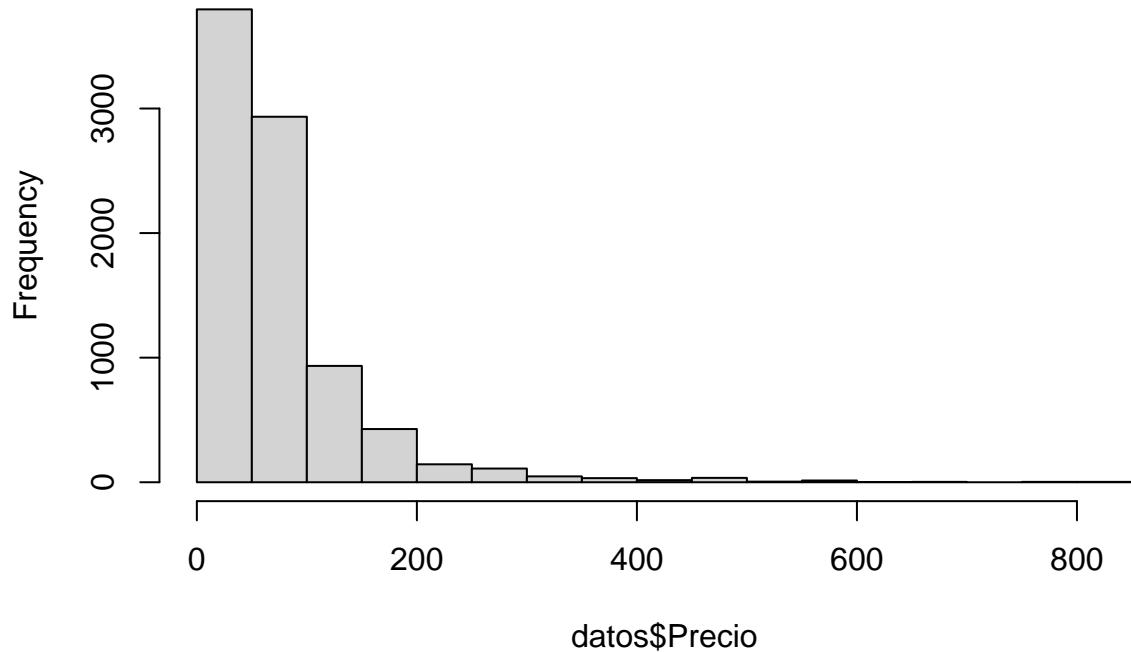
## Histograma de la variable Longitud



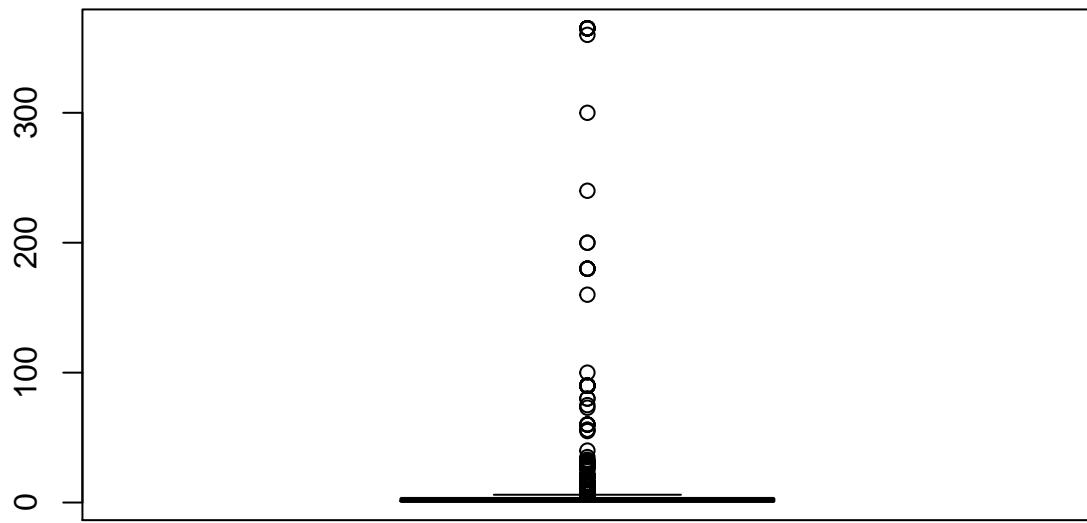


```
L1 = c(0,100,250,1000,2500,5000,10000)
hist(datos$Precio, main= "Histograma de la variable Precio")
```

## Histograma de la variable Precio

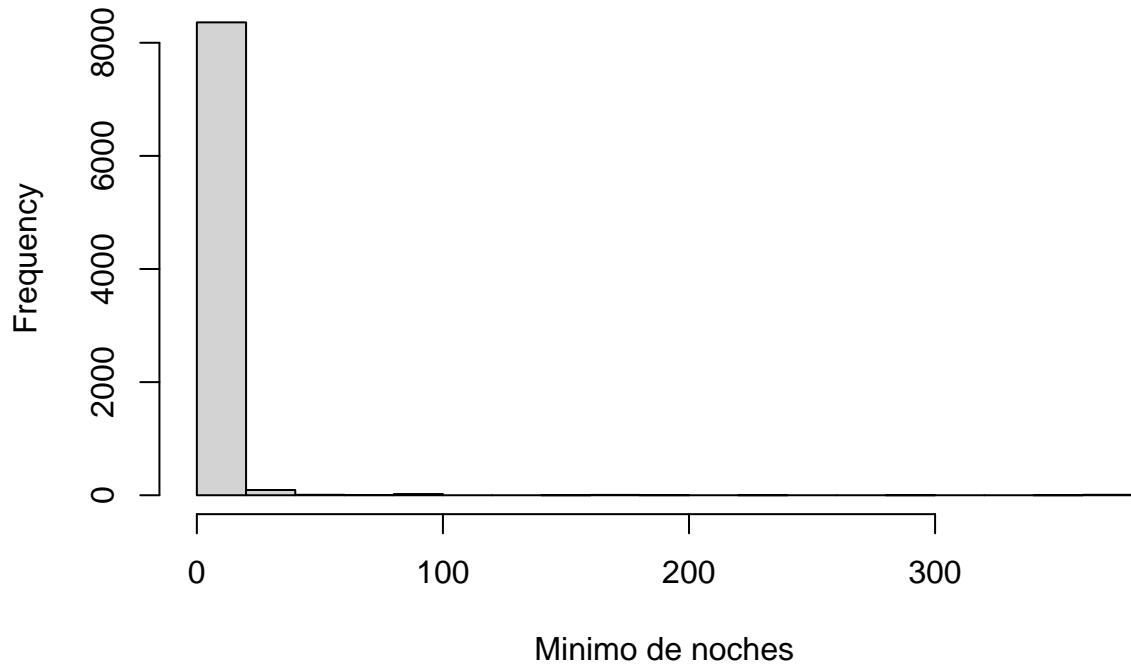


```
boxplot(datos$Min.noches)
```

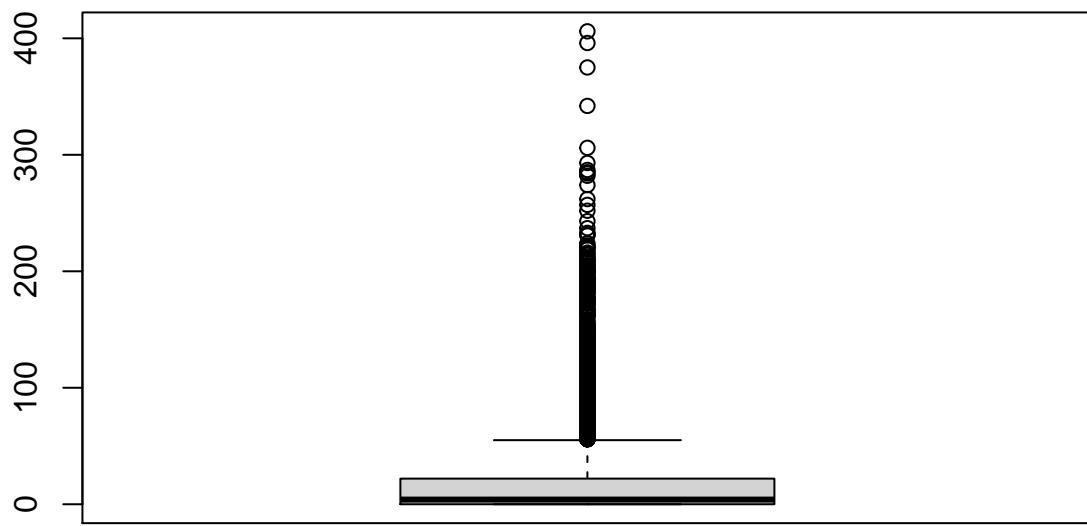


```
hist(datos$Min.noches, main= "Histograma de la variable Minimo de noches", xlab = "Minimo de noches")
```

## Histograma de la variable Minimo de noches

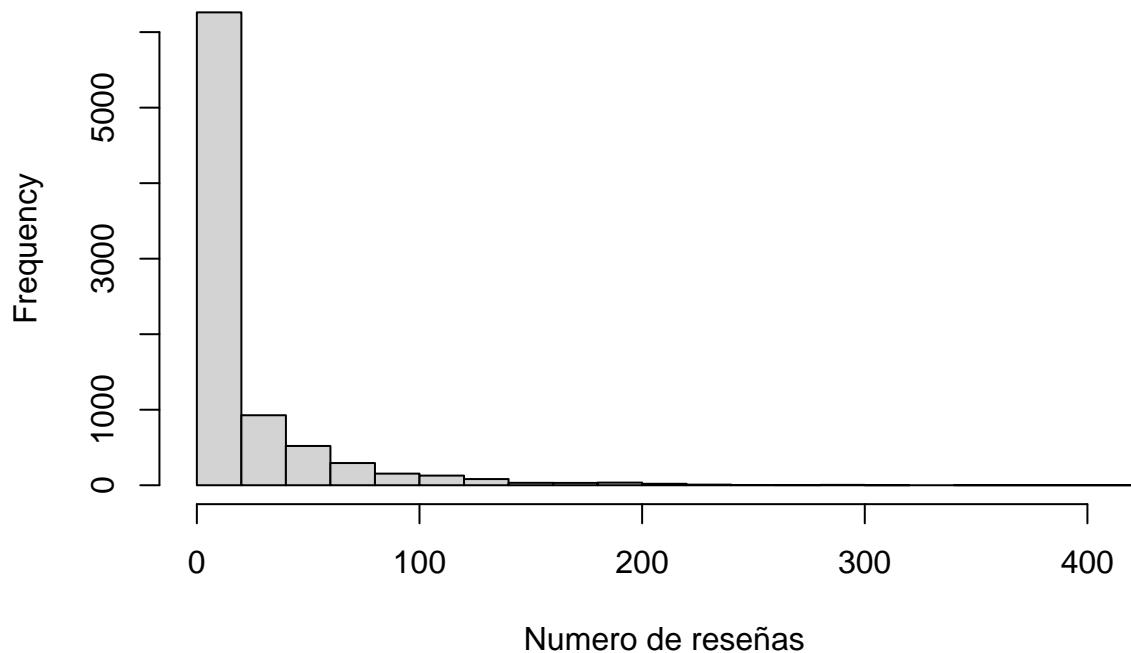


```
boxplot(datos$N_reseñas)
```

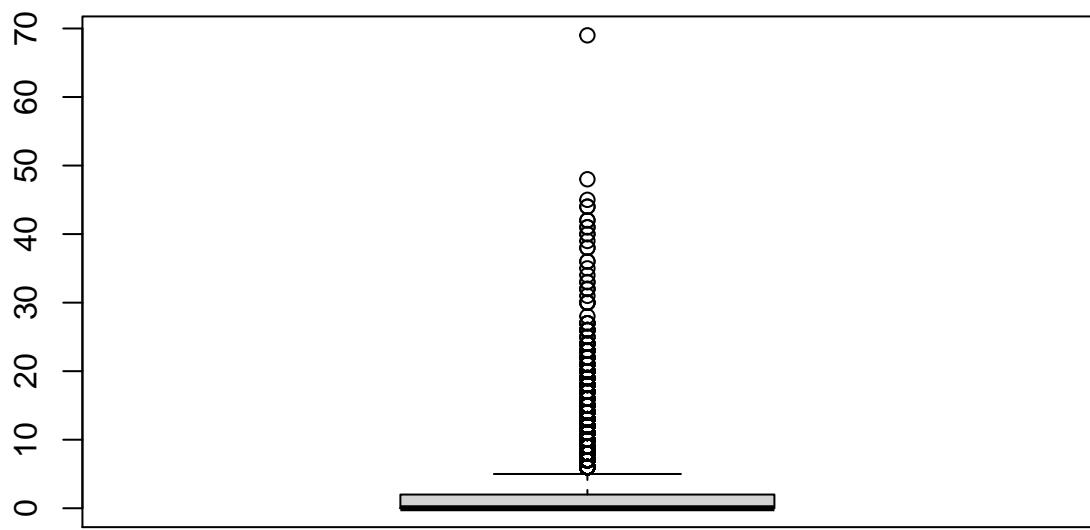


```
hist(datos$N_reseñas, main= "Histograma de la variable Numero de reseñas", xlab = "Numero de reseñas")
```

## Histograma de la variable Número de reseñas

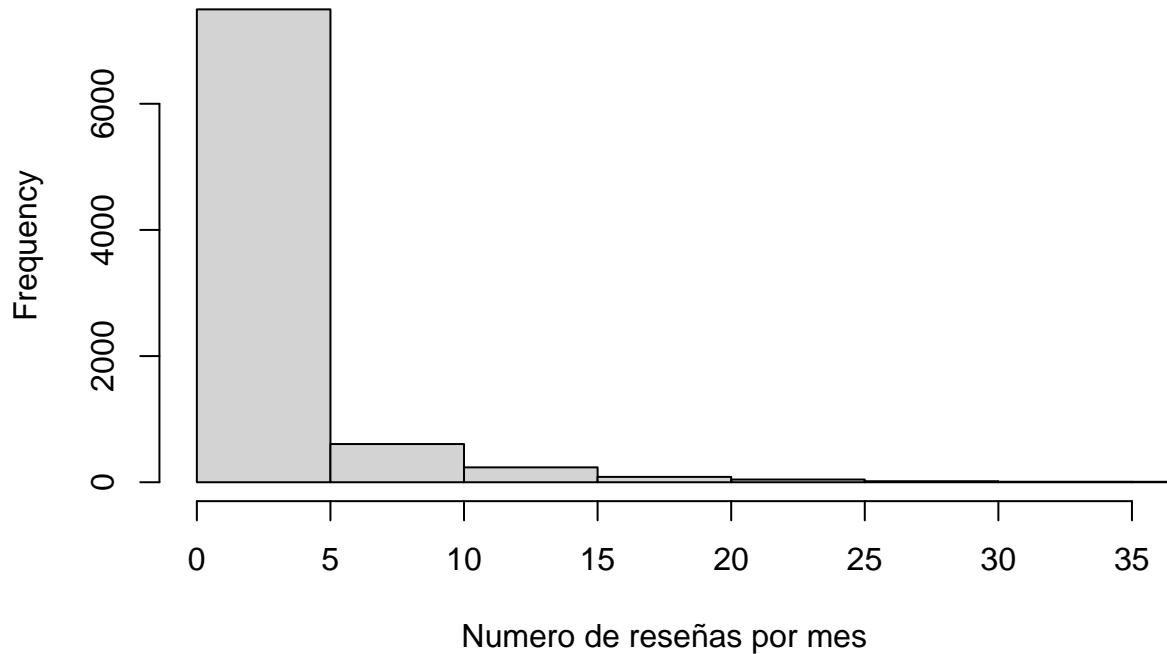


```
boxplot(datos$Reseñas_mes)
```

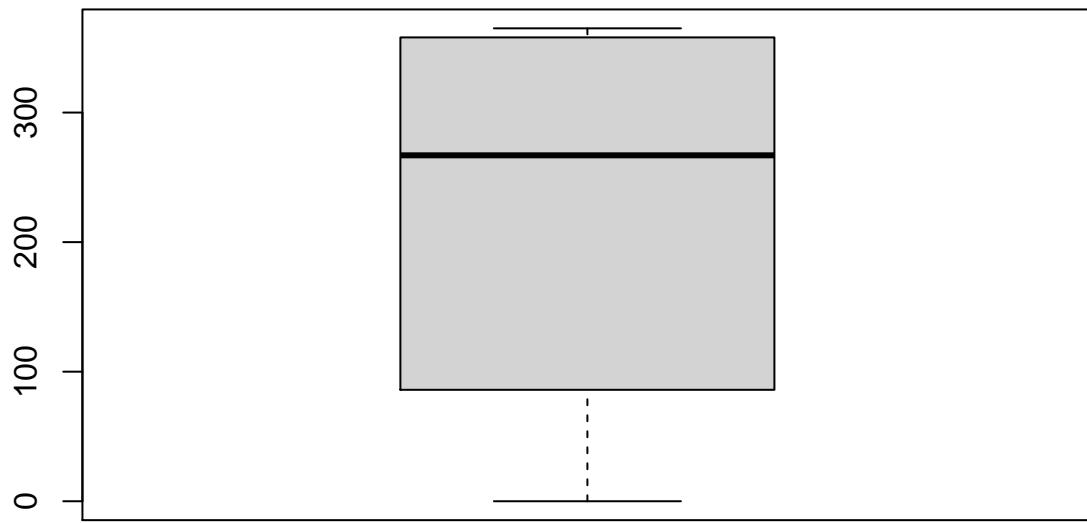


```
hist(datos$Reseñas_mes, main= "Histograma de la variable Numero de reseñas por mes", xlab = "Numero de :")
```

## Histograma de la variable Número de reseñas por mes

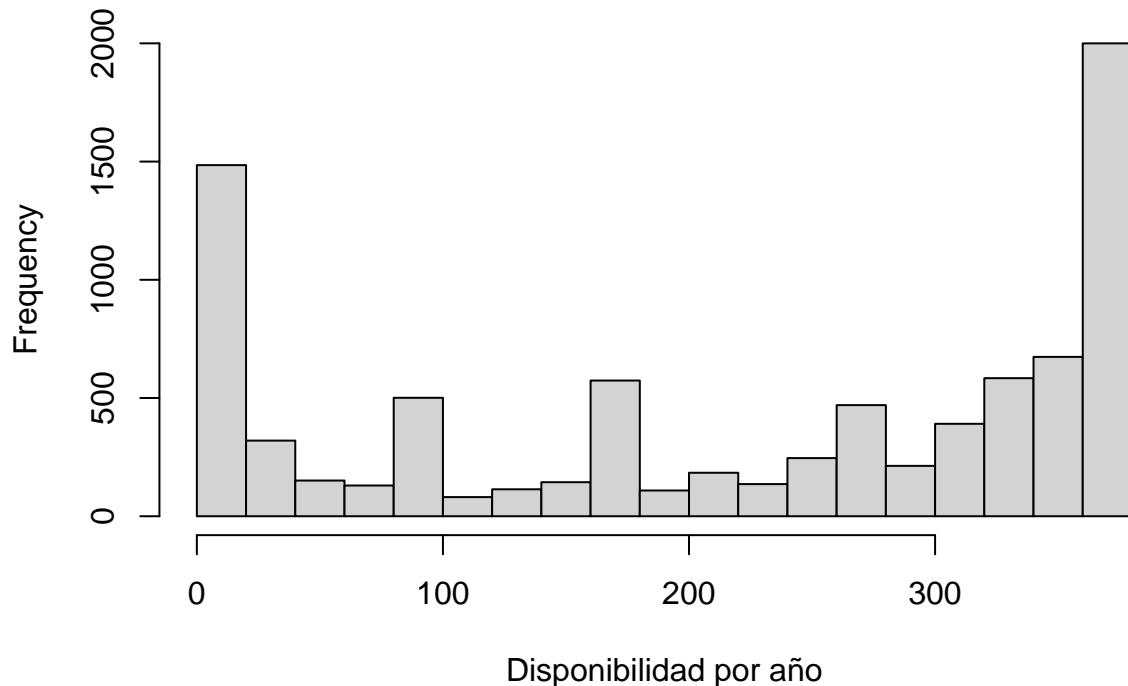


```
boxplot(datos$Disponibilidad_año)
```



```
hist(datos$Disponibilidad_año, main= "Histograma de la variable disponibilidad por año", xlab = "Disponibili
```

## Histograma de la variable disponibilidad por año



## Pie chart de variables no numéricas

```
frecuenciaVecindario = table(datos$Vecindario) #Hacemos una tabla de la frecuencias de la cantidad de habitacionesTotal=max(cumsum(frecuenciaVecindario)) #Número total de habitaciones en Malta

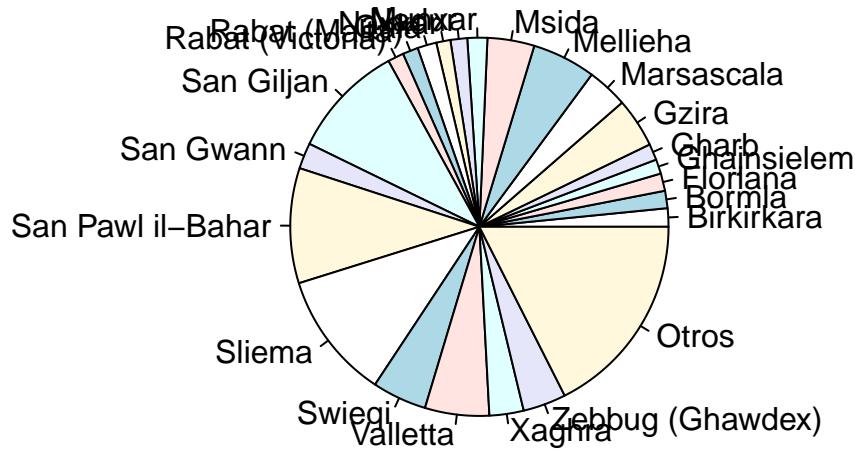
frecuenciaVecindario <- frecuenciaVecindario[frecuenciaVecindario>0.01*habitacionesTotal] #Excluimos las casas que tienen menos de 1% del total

numeroDeOtros= habitacionesTotal-max(cumsum(frecuenciaVecindario)) #Conseguir el número de "Otros" que quedan

names(numeroDeOtros)="Otros"
frecuenciaVecindario=append(frecuenciaVecindario, numeroDeOtros) #Añadir Otros al final de la tabla

pie(frecuenciaVecindario, main="Diagrama de tarta de la cantidad de habitaciones por cada barrio") #Graficar el pie
```

## Diagrama de tarta de la cantidad de habitaciones por cada barrio

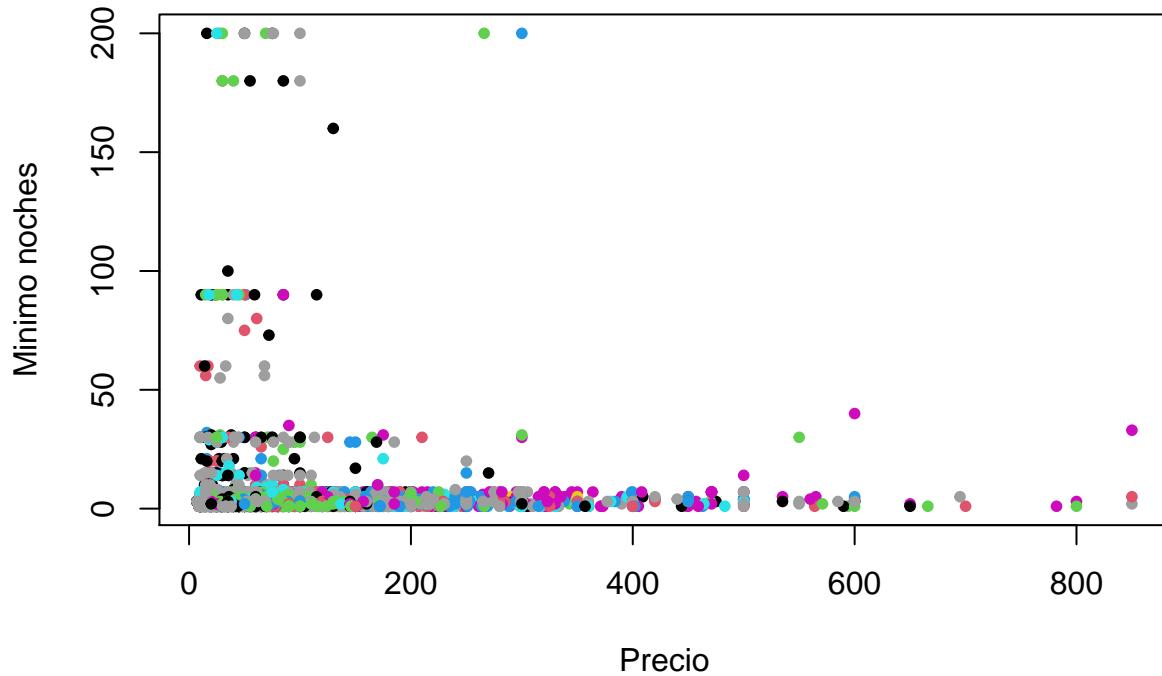


### 9. Diagramas de dispersión de 4 variables numéricas 2 a 2, con un color para cada Vecindario.

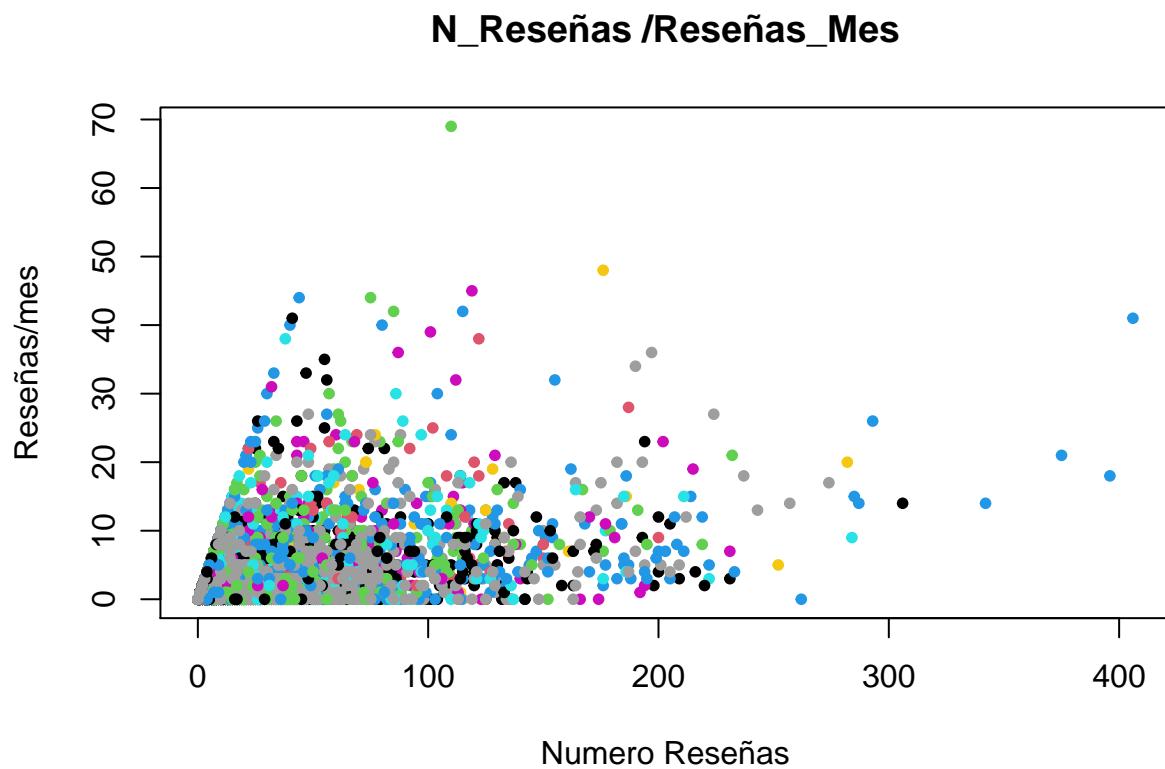
```
VecindarioNum <-as.factor(datos$Vecindario) # Pasamos el vector vecindario a un vector de factores

datos$Min.noches <- ifelse(datos$Min.noches > 200, 200, datos$Min.noches) #creamos el primer plot
plot(
  x = datos$Precio,
  y = datos$Min.noches,
  xlab = "Precio",
  ylab = "Minimo noches",
  pch = 20, # solid dots increase the readability of this data plot
  col = VecindarioNum, # El color se asigna dependiendo del vecindario
  main = "Precio /Min noches"
)
```

## Precio /Min noches



```
plot(  
  x = datos$N_reseñas,  
  y = datos$Reseñas_mes,  
  xlab = "Numero Reseñas",  
  ylab = "Reseñas/mes",  
  pch = 20, # solid dots increase the readability of this data plot  
  col = VecindarioNum, # El color se asigna dependiendo del vecindario  
  main = "N_Reseñas /Reseñas_Mes"  
)
```



#### 10. Coeficientes de correlación 2 a 2 apartado anterior.

```
cor(x = datos$Precio, y = datos$Min.noches)

## [1] 0.003649116

cor(x = datos$N_reseñas, y = datos$Reseñas_mes)

## [1] 0.503904
```