

Practica Final Parte 1

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

Contenidos:

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.	2
5. Boxplots y histogramas de las variables numericas.	5
6. Pie chart de variables no numéricas.	9
7. Valor medio de alguna de las variables numéricas según el barrio.	9
8. Boxplot de la variable precio, para los 5 barrios más caros y los 5 barrios más baratos.	12
9. Diagramas de dispersión de 4 variables numéricas 2 a 2, con un color para cada Vecindario.	14
10. Coeficientes de correlación 2 a 2 apartado anterior.	15
11. Con 2 variables numéricas, organizar sus valores en un máximo de 5 intervalos con la función cut.	15
12. A partir de los datos en intervalos obtenidos en el apartado anterior construid una tabla de contingencia de las dos variables y dibuja el diagrama de mosaico asociado a la mesa.	15

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.

```
#Renombramos todas las variables
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"
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.

	Lat.	Long.	Precio	Min.noches	N_reseñas	Reseñas_mes
datosEstadisticos	35.9368	14.4326	78.9676	3.3122	19.3657	2.0924
media	35.9368	14.4326	78.9676	3.3122	19.3657	2.0924
varianza	0.0037	0.0098	5661.4032	34.5577	1217.0306	18.3768
mínimo	35.8133	14.1954	7.0000	1.0000	0.0000	0.0000
Cuartil_1	35.8992	14.3722	35.0000	1.0000	0.0000	0.0000
Mediana	35.9156	14.4840	59.0000	2.0000	4.0000	0.0000
Cuartil_3	35.9530	14.4994	95.0000	3.0000	22.0000	2.0000
máximo	36.0802	14.5780	850.0000	90.0000	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	117
Gharghur	Ghasri	Ghaxaq	Gudja
17	80	11	18
Gzira	Hamrun	Iklin	Isla
360	27	8	79
Kalkara	Kercem	Kirkop	Lija
21	63	8	11
Luqa	Marsa	Marsascala	Marsaxlokk
12	6	295	59
Mdina	Mellieha	Mgarr	Mosta
13	462	65	83
Mqabba	Msida	Mtarfa	Munxar
3	340	5	145
Nadur	Naxxar	Paola	Pembroke
121	103	21	78
Pieta	Qala	Qormi	Qrendi
80	138	44	13
Rabat (Malta)	Rabat (Victoria)	Safi	San Giljan
113	116	7	831
San Gwann	San Lawrenz	San Pawl il-Bahar	Sannat
188	51	834	46
Santa Lucija	Santa Venera	Siggiewi	Sliema
1	23	25	922
Swieqi	Ta' Xbiex	Tarxien	Valletta
391	50	32	467
Xaghra	Xewkija	Xghajra	Zabbar
249	52	21	28
Zebbug (Ghawdex)	Zebbug (Malta)	Zejtun	Zurrieq
312	50	41	25

frecuenciaNombre

Joseph	Short Lets Malta
130	126
Maria	360 Estates
108	107
Victor	Baron
93	92
Robert	Mark
88	79
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
Stephanie	Alfred
22	21
Chiara	Jorge
21	21
Lorraine	Short Lets
21	21

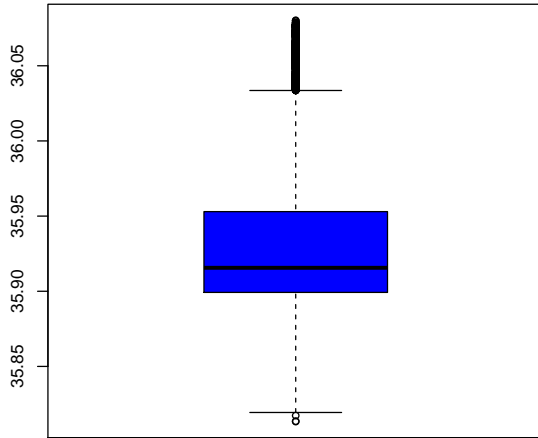
frecuenciaTipo

Entire home/apt	Hotel room	Private room	Shared room
5249	182	2737	318

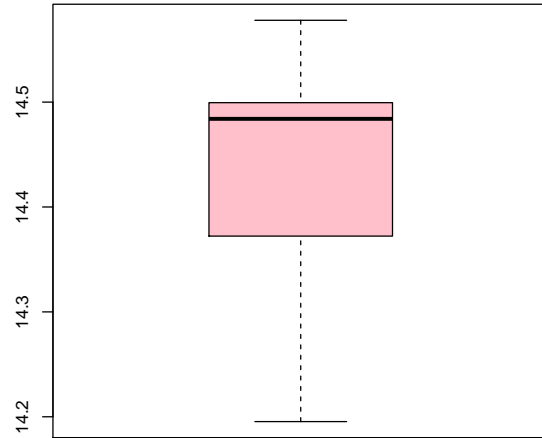
5. Boxplots y histogramas de las variables numericas.

Boxplots de las variables numericas.

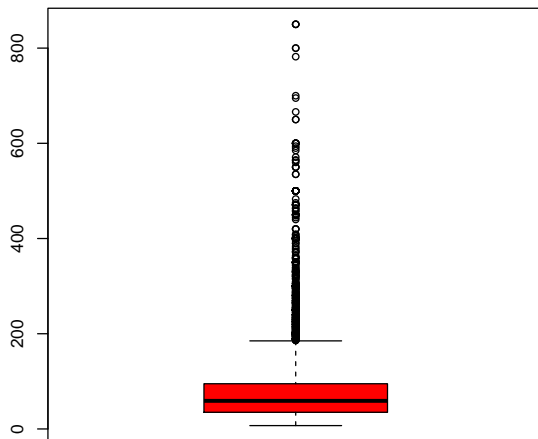
Boxplot de la variable Latitud



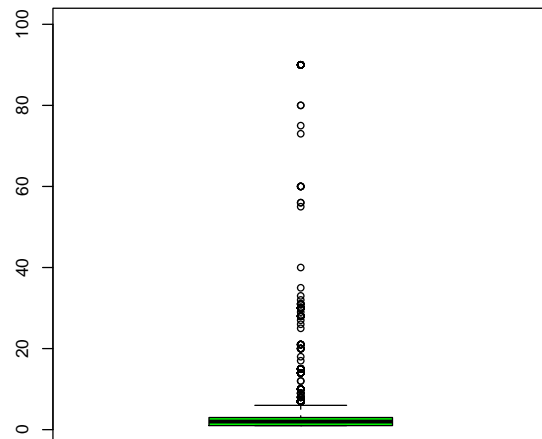
Boxplot de la variable Longitud



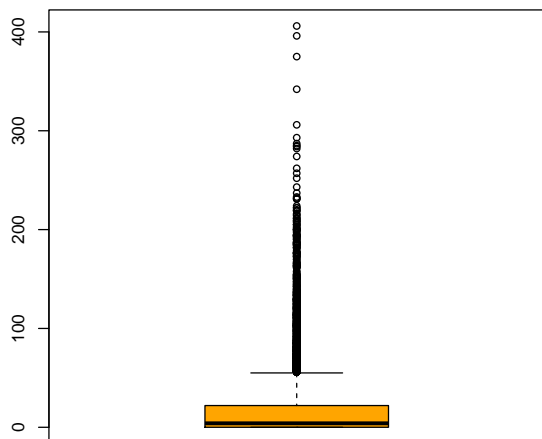
Boxplot de la variable Precio



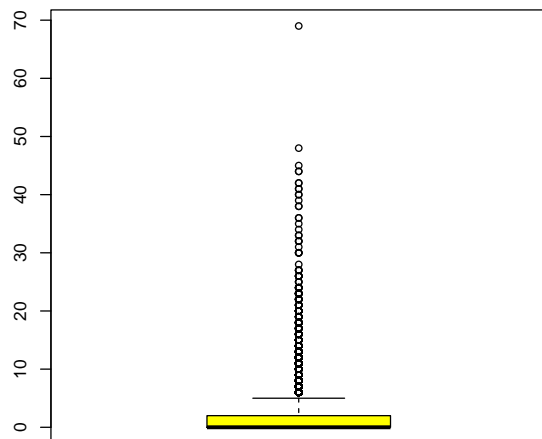
Boxplot de la variable Mínimo de Noches



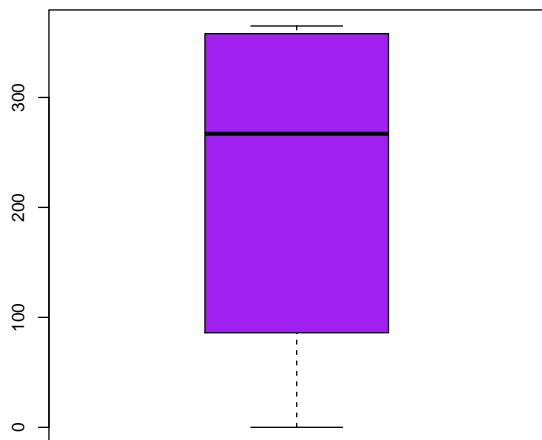
Boxplot de la variable Numeros de reseñas



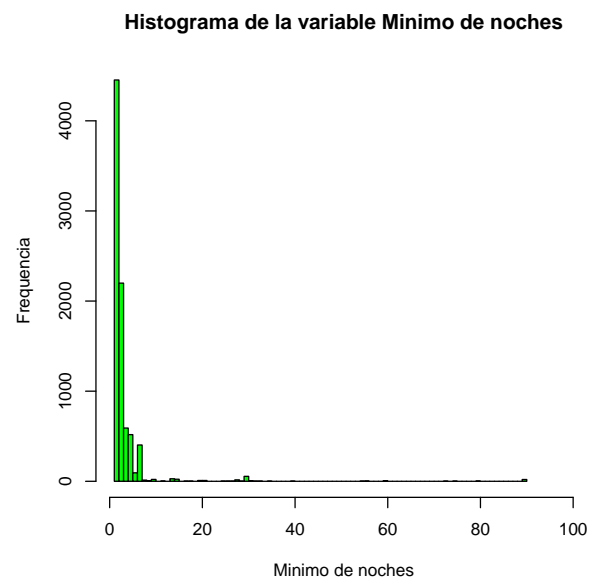
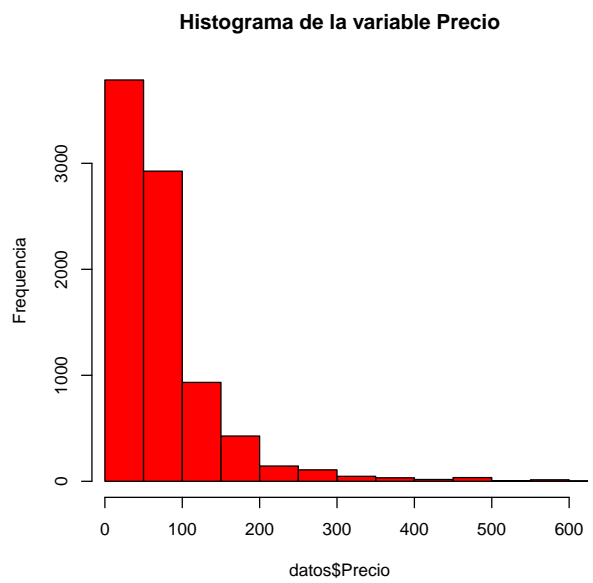
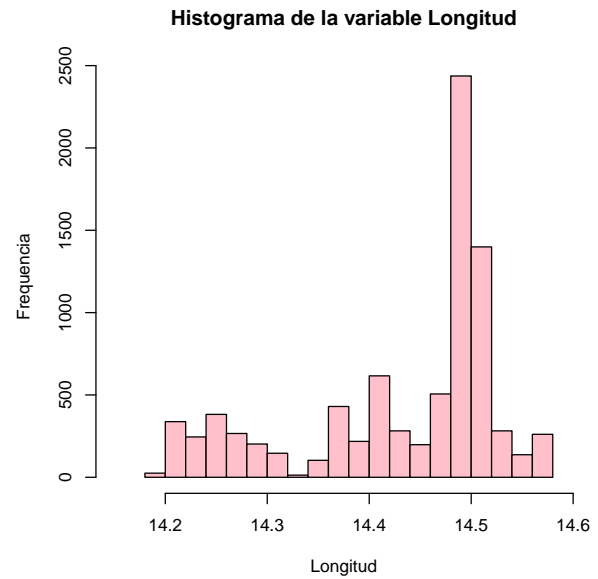
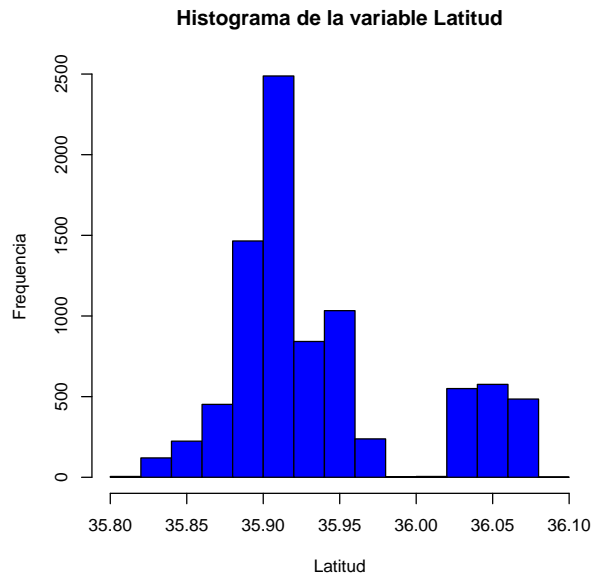
Boxplot de la variable Numeros de Reseñas por mes



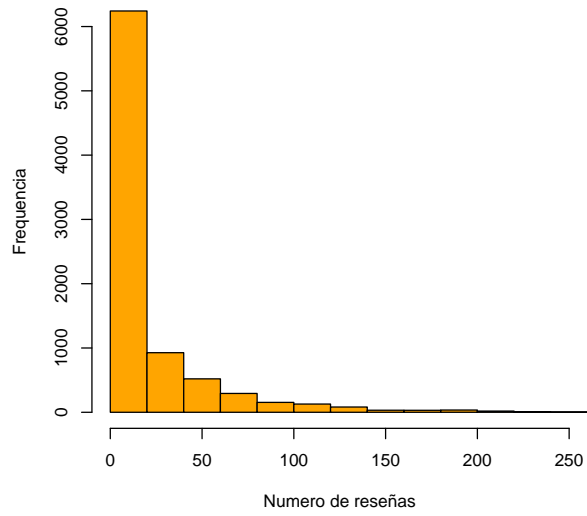
Boxplot de la variable Disponibilidad por año



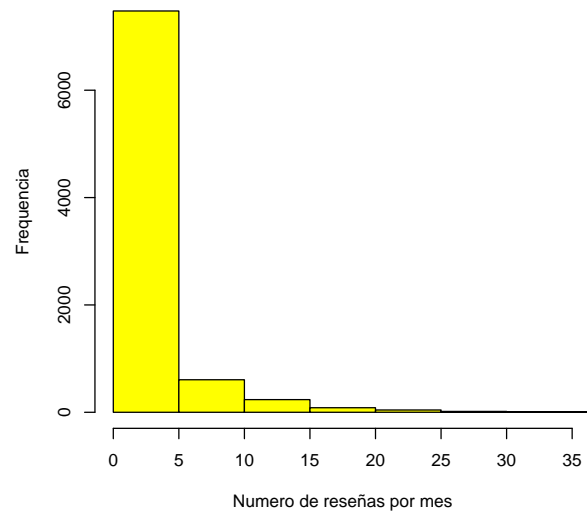
Histogramas de las variables numericas.



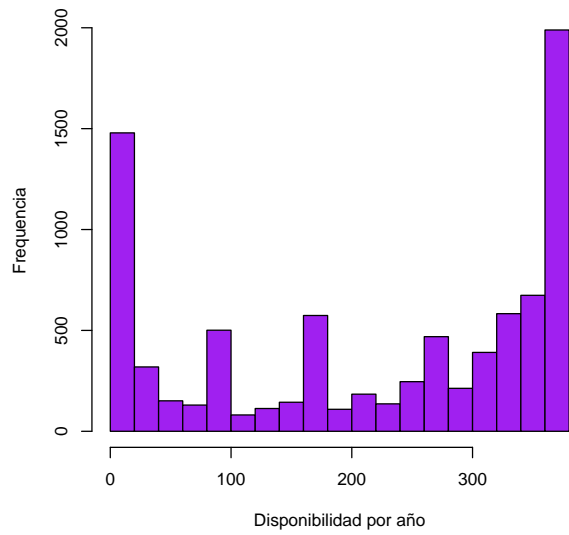
Histograma de la variable Numero de reseñas



Histograma de la variable Numero de reseñas por mes

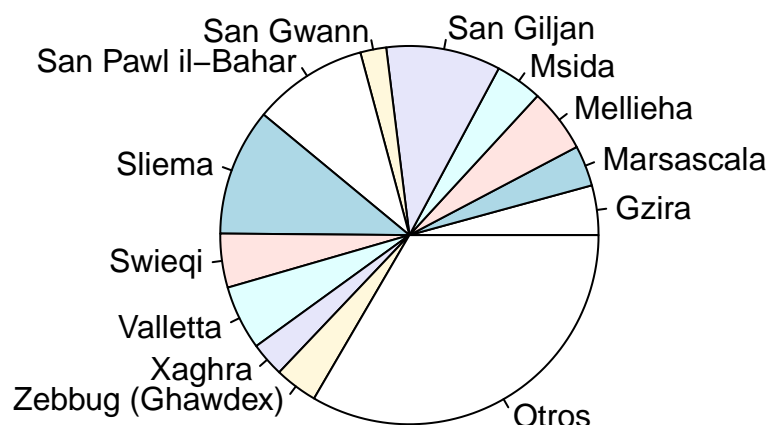


Histograma de la variable disponibilidad por año



6. Pie chart de variables no numéricas.

Diagrama de tarta de la cantidad de habitaciones por cada barrio



7. Valor medio de alguna de las variables numéricas según el barrio.

Santa Lucija	Mtarfa	Qormi	San Gwann
24.00000	27.20000	36.31818	37.90426
Safi	Paola	Kirkop	Birkirkara
38.28571	39.66667	40.87500	44.91791
Pembroke	Mosta	Msida	Gudja
46.32051	49.38554	55.35588	55.44444
Swieqi	Birzebbugia	Munxar	Marsa
56.46036	58.26562	58.84138	59.50000
Pieta	Attard	San Pawl il-Bahar	Naxxar
59.50000	61.64706	64.32614	67.46602
Balzan	Gzira	Zebbug (Ghawdex)	Rabat (Malta)
68.00000	68.37500	68.68910	69.32743
Ghaxaq	Luqa	Marsascala	Xghajra
70.45455	70.50000	72.21695	72.85714
Birgu	Hamrun	Fontana	San Giljan
73.06098	75.11111	76.59091	77.41155
Mgarr	Qrendi	Sliema	Isla
78.36923	78.61538	79.73102	79.93671
Rabat (Victoria)	Kalkara	Floriana	Tarxien
80.10345	81.09524	82.57480	85.00000
Marsaxlokk	Bormla	Mqabba	Santa Venera

85.54237	86.53175	86.66667	86.73913
Ghajnsielem	Valletta	Siggiewi	Lija
88.69524	88.80300	90.12000	90.54545
Nadur	Iklin	Qala	Zabbar
90.76860	93.62500	97.31884	97.82143
Zejtun	Gharghur	Xewkija	Zebbug (Malta)
103.36585	105.41176	111.30769	113.10000
Mellieha	Mdina	Gharb	Xaghra
114.67965	116.30769	118.09402	121.40161
Kercem	San Lawrenz	Dingli	Sannat
127.76190	127.96078	132.50000	132.84783
Ta' Xbiex	Zurrieq	Ghasri	Fgura
138.50000	148.56000	155.82500	163.57143

Valor medio de la variable Número de reseñas en base a los distintos barrios.

Marsa	Dingli	Attard	Zurrieq
1.333333	2.000000	5.764706	8.080000
Swieqi	Ta' Xbiex	Zejtun	San Gwann
8.219949	9.000000	9.000000	9.196809
Pembroke	Ghasri	Naxxar	Gharghur
9.538462	9.712500	10.184466	10.823529
San Lawrenz	Birkirkara	Kercem	Msida
11.019608	11.059701	11.142857	11.155882
Fontana	Xaghra	Zebbug (Malta)	Sannat
11.863636	12.485944	12.860000	12.978261
Gharb	Mosta	Zabbar	Tarxien
13.000000	13.457831	13.821429	14.781250
Lija	Rabat (Victoria)	Mellieha	Santa Venera
15.454545	15.577586	15.755411	15.913043
Ghaxaq	Marsaxlokk	San Giljan	Siggiewi
16.000000	16.338983	17.108303	17.440000
San Pawl il-Bahar	Xghajra	Ghajnsielem	Marsascala
17.995204	18.095238	18.171429	18.254237
Zebbug (Ghawdex)	Qala	Hamrun	Qrendi
18.746795	18.913043	19.333333	19.692308
Gzira	Qormi	Sliema	Nadur
21.219444	21.590909	21.817787	21.892562
Birzebbugia	Paola	Bormla	Mtarfa
23.031250	23.619048	24.095238	25.000000
Xewkija	Balzan	Rabat (Malta)	Munxar
25.057692	25.473684	25.929204	26.041379
Iklin	Kalkara	Mgarr	Pieta
26.750000	26.904762	28.646154	28.912500
Birgu	Valletta	Safi	Isla
32.731707	36.901499	42.714286	43.708861
Floriana	Luqa	Gudja	Fgura
47.771654	51.000000	54.555556	54.571429
Mqabba	Mdina	Kirkop	Santa Lucija
61.666667	73.230769	78.875000	117.000000

Valor medio de la variable Disponibilidad por año en base a los distintos barrios.

Zurrieq	Ghaxaq	Luqa	Santa Lucija
---------	--------	------	--------------

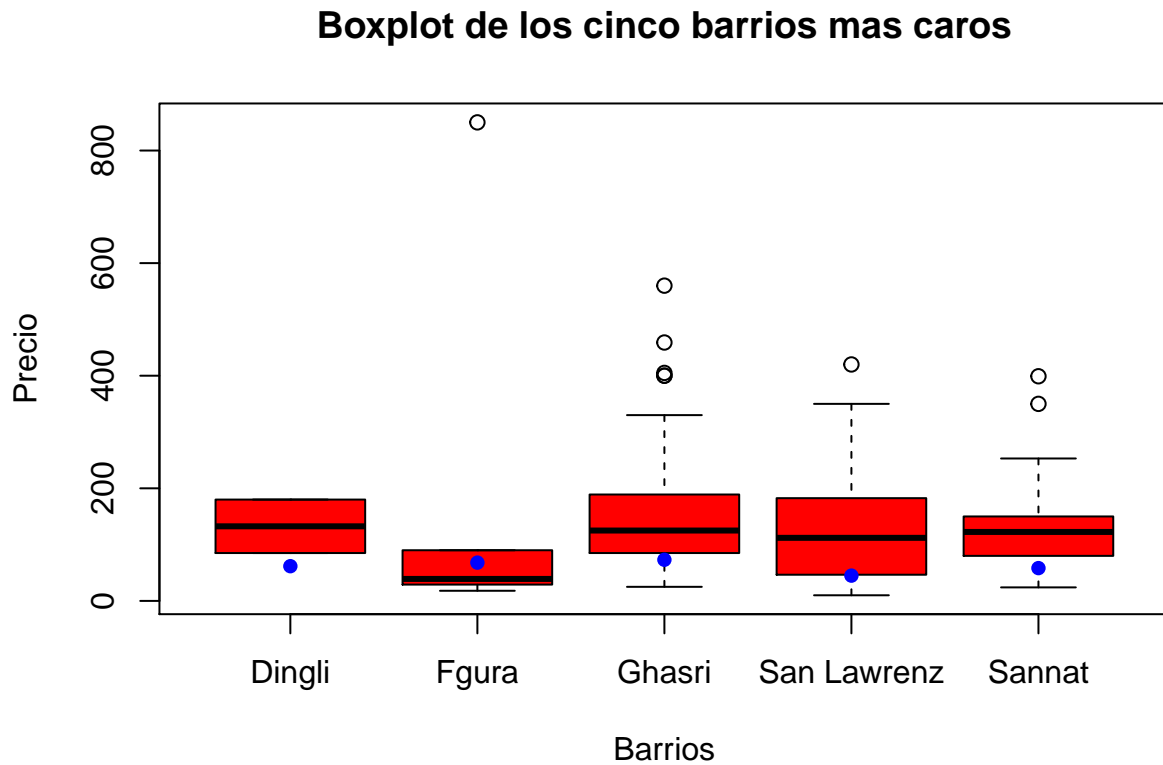
166.0800	167.8182	175.1667	179.0000
Balzan	Paola	Qala	Dingli
180.7895	181.1429	182.0072	182.5000
San Gwann	Fgura	Qormi	Pieta
183.8085	185.5714	186.5682	190.5500
Sliema	Pembroke	Birkirkara	Fontana
191.1638	191.6026	192.0597	192.4545
Santa Venera	Marsa	Mgarr	Gzira
193.6522	195.8333	198.4000	199.1000
Nadur	Lija	Rabat (Malta)	Siggiewi
199.5950	200.0000	200.5752	200.9600
Gharghur	San Pawl il-Bahar	San Giljan	Zejtun
204.2353	204.8237	205.3105	206.7805
Mqabba	Swieqi	Naxxar	Msida
207.3333	207.9974	209.2816	209.4500
Hamrun	Xaghra	Floriana	Kercem
212.3704	216.5020	217.7638	217.9683
Rabat (Victoria)	Munxar	San Lawrenz	Ghasri
218.2328	220.9379	223.5686	226.3750
Ghajnsielem	Zebbug (Malta)	Kirkop	Marsascala
229.6190	229.7400	229.8750	231.5390
Mosta	Safi	Sannat	Bormla
231.5422	233.0000	233.7174	234.7222
Zebbug (Ghawdex)	Ta' Xbiex	Kalkara	Xghajra
237.3269	241.1200	244.6667	247.0476
Birgu	Birzebbugia	Xewkija	Mellieha
248.9878	249.0312	250.4423	251.3074
Qrendi	Gharb	Isla	Zabbar
251.3846	252.5897	253.2658	254.2143
Mdina	Valletta	Marsaxlokk	Tarxien
260.4615	262.8201	274.2542	277.9375
Iklin	Gudja	Attard	Mtarfa
298.5000	309.1667	310.8235	316.0000

Valor medio de la variable Mínimo de noches en base a los distintos barrios.

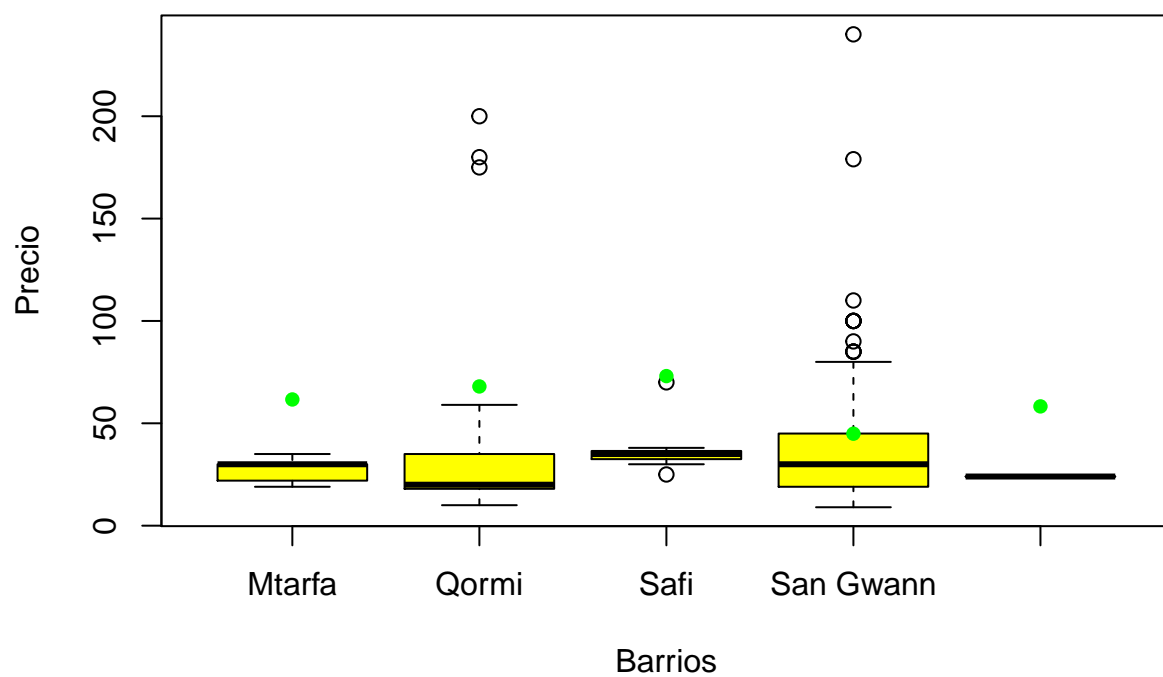
Kirkop	Santa Lucija	Luqa	Mtarfa
1.000000	1.000000	1.583333	1.600000
Gudja	Fontana	Paola	Safi
1.666667	1.772727	1.904762	2.000000
Hamrun	Qrendi	Xewkija	Rabat (Victoria)
2.037037	2.153846	2.307692	2.318966
Floriana	Kercem	Gharghur	Zabbar
2.330709	2.349206	2.352941	2.357143
Munxar	Tarxien	Zejtun	Valletta
2.434483	2.437500	2.439024	2.464668
Isla	Bormla	Lija	Marsaxlokk
2.468354	2.515873	2.545455	2.559322
San Lawrenz	Mgarr	Zurrieq	Mosta
2.568627	2.584615	2.600000	2.614458
Xaghra	Gharb	Mqabba	Xghajra
2.658635	2.666667	2.666667	2.666667
Pembroke	Ghasri	Ghajnsielem	Siggiewi
2.756410	2.775000	2.790476	2.800000

Birgu	Zebbug (Malta)	Mdina	Fgura
2.817073	2.840000	2.846154	2.857143
Qala	Qormi	Zebbug (Ghawdex)	Ta' Xbiex
2.876812	2.886364	2.907051	3.180000
San Giljan	Sliema	Kalkara	Rabat (Malta)
3.225030	3.247289	3.285714	3.318584
Marsa	Nadur	San Pawl il-Bahar	Sannat
3.333333	3.446281	3.484412	3.521739
Pieta	Birzebbugia	San Gwann	Swieqi
3.775000	3.812500	3.829787	3.867008
Mellieha	Ghaxaq	Msida	Marsascala
3.904762	4.090909	4.200000	4.216949
Birkirkara	Dingli	Gzira	Naxxar
4.425373	4.500000	4.505556	4.912621
Balzan	Iklin	Santa Venera	Attard
5.315789	6.625000	7.608696	7.823529

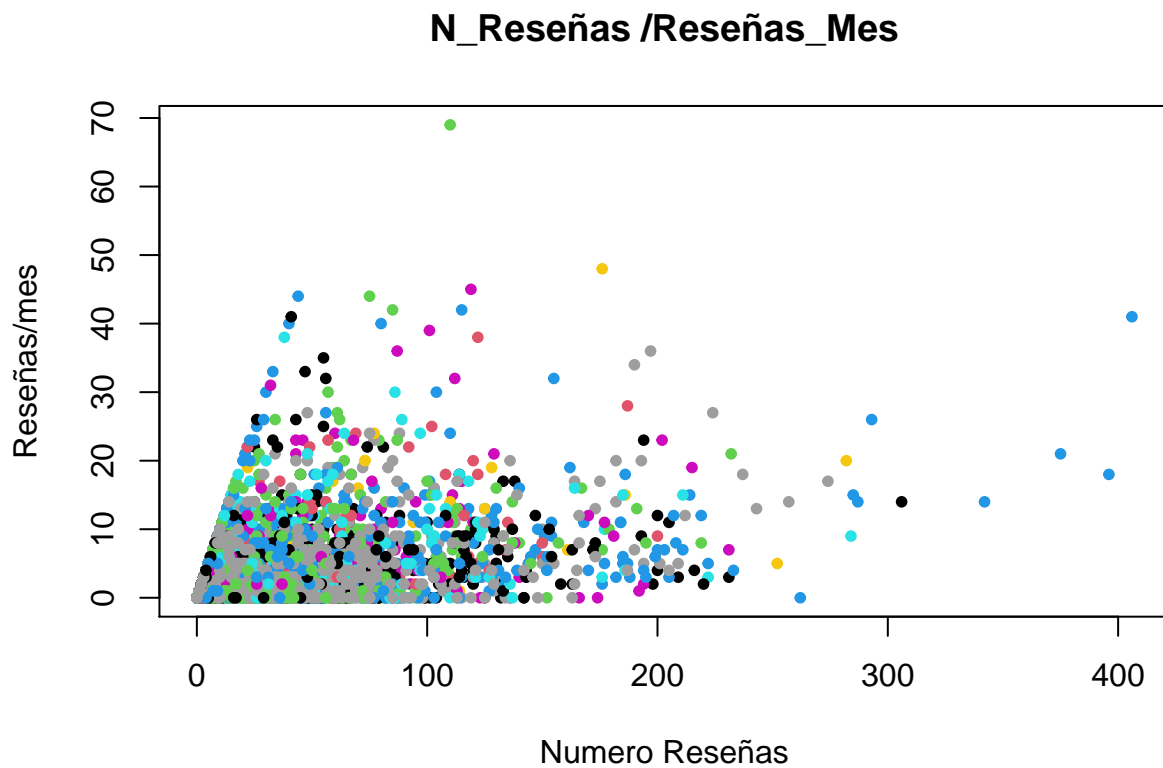
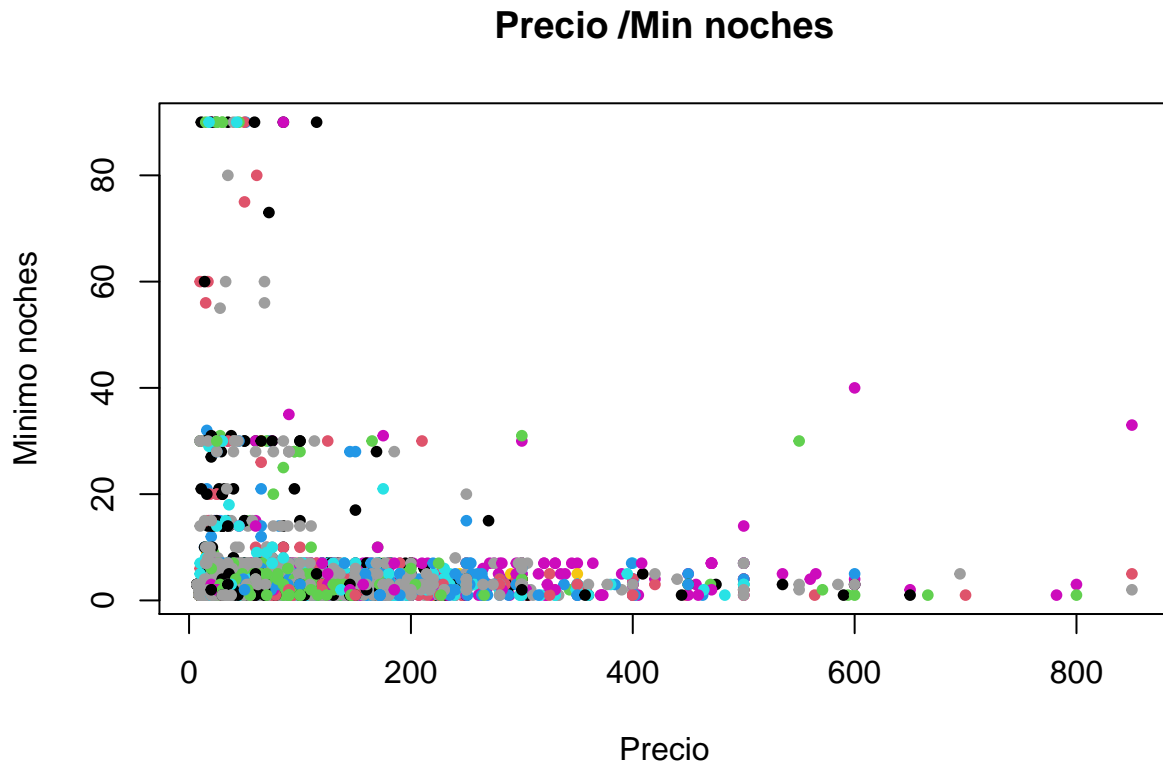
8. Boxplot de la variable precio, para los 5 barrios más caros y los 5 barrios más baratos.



Boxplot de los cinco barrios mas baratos



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



10. Coeficientes de correlación 2 a 2 apartado anterior.

```
[1] 0.005161337
```

```
[1] 0.5037265
```

11. Con 2 variables numéricas, organizar sus valores en un máximo de 5 intervalos con la función cut.

```
Cut1 <- cut(x = datos$Precio, breaks = 5)
Cut2 <- cut(x = datos$N_reseñas, breaks = 5)
```

12. A partir de los datos en intervalos obtenidos en el apartado anterior construid una tabla de contingencia de las dos variables y dibuja el diagrama de mosaico asociado a la mesa.

	Cut2				
Cut1	(-0.406,81.2]	(81.2,162]	(162,244]	(244,325]	(325,406]
(6.16,176]	7361	380	89	10	4
(176,344]	502	6	1	0	0
(344,513]	102	1	0	0	0
(513,681]	21	0	1	0	0
(681,851]	8	0	0	0	0

