

1-Teoría

SETUP Y LIBRERIAS

Para la realización de la práctica cargaremos todos los datos usando un dataframe y a su vez importaremos la librería ggplot que se usara mas adelante

a) Obténgase una distribución de datos conjunta en intervalos de amplitud del 10% de la distribución, construir una tabla de frecuencias absolutas y relativas y definir cada representante de la clase.

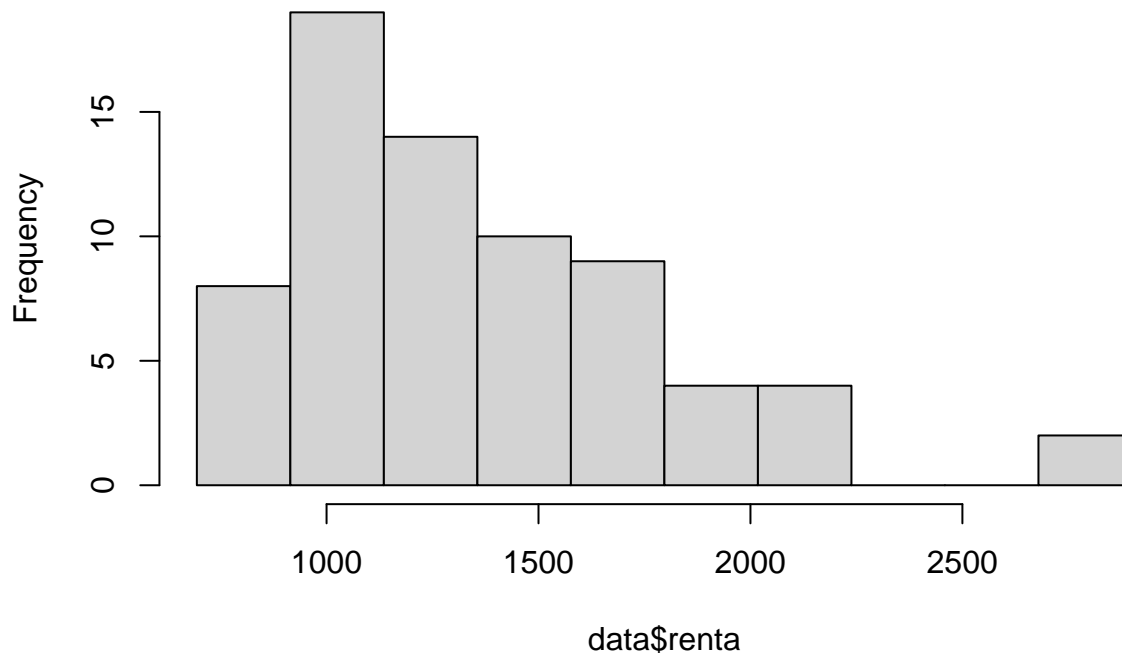
Para llevar a cabo esta tarea requeriremos el uso de varios comandos de R. Por una parte cut para dividir en intervalos los datos de los ingresos, por otra parte el comando hist que nos creara un histograma conteniendo los representantes de clase, por otra parte table que nos servirá para obtener las frecuencias absolutas y el resto son cálculos sencillos y operaciones con data frames.

```
intervalosRenta <- cut(data$ingresos, breaks=seq(694, 2900, 220.6), include.lowest=TRUE, right=TRUE)

data <- data.frame(interval = intervalosRenta, renta = data$ingresos, barrio = data$barrio)

#tabla de frecuencias absolutas
rep <- hist(data$renta, breaks=seq(694, 2900, 220.6), main="Histograma de las rentas")
```

Histograma de las rentas



```
tabla_frecuencias <- data.frame(table(data$interval), rep$mids)
names(tabla_frecuencias) <- c("Intervalos", "Freq. Absol", "Representantes de clase")
tabla_frecuencias <- data.frame(tabla_frecuencias, tabla_frecuencias$`Freq. Absol`/length(data$renta))
names(tabla_frecuencias) <- c("Intervalos", "Freq. Absol", "Representantes de clase", "Freq relativa")
tabla_frecuencias
```

##	Intervalos	Freq. Absol	Representantes de clase	Freq relativa
## 1	[694,915]	8	804.3	0.11428571
## 2	(915,1.14e+03]	19	1024.9	0.27142857
## 3	(1.14e+03,1.36e+03]	14	1245.5	0.20000000
## 4	(1.36e+03,1.58e+03]	10	1466.1	0.14285714
## 5	(1.58e+03,1.8e+03]	9	1686.7	0.12857143
## 6	(1.8e+03,2.02e+03]	4	1907.3	0.05714286
## 7	(2.02e+03,2.24e+03]	4	2127.9	0.05714286
## 8	(2.24e+03,2.46e+03]	0	2348.5	0.00000000
## 9	(2.46e+03,2.68e+03]	0	2569.1	0.00000000
## 10	(2.68e+03,2.9e+03]	2	2789.7	0.02857143

b) Explicar el tipo de muestreo realizado y calcular las medias muestrales y las desviaciones estándar muestrales, del conjunto y por barrios

El muestreo realizado ha sido por agrupamientos. Se han buscado agrupaciones no homogéneas de residentes dentro de los barrios que serían los respectivos Clústers. De esta manera por ejemplo podemos comparar eficazmente el estado económico de cada grupo.

Para calcular las medias muestrales y las desviaciones estándar harémos uso de los comandos mean y sd, a su vez que aggregate para construir las tablas.

```
#MEDIAS Y SD GENERALES
```

```
mean(data$renta)
```

```
## [1] 1365.9
```

```
sd(data$renta)
```

```
## [1] 451.7381
```

```
#MEDIAS Y SD POR BARRIO
```

```
tabla_medias <- aggregate(renta~barrio, data=data, mean )
```

```
names(tabla_medias) <- c("Barrio", "Media renta")
```

```
tabla_sd <- aggregate(renta~barrio, data=data, sd )
```

```
names(tabla_sd) <- c("Barrio", "Desviacion estandar")
```

```
tabla_final <- merge(tabla_medias, tabla_sd)
```

```
tabla_final
```

```
##   Barrio Media renta Desviacion estandar
```

```
## 1    B1   1210.050         198.9640
```

```
## 2    B2   1642.158         511.1913
```

```
## 3    B3   1297.129         466.9886
```

c) Encontrar la mediana, los cuartiles y el rango intercuartílico.

Esto es muy fácil de realizar con r siempre y cuando se conozcan los comandos y conceptos adecuados. Por una parte usaremos median para calcular la mediana general y por otra parte quantile para calcular los cuartiles, a su vez usaremos quantile para calcular el cuartil 75% y el 25% y restarlos para así conseguir el recorrido intercuartílico.

```
renta <- sort(data$renta)
```

```
median(renta)
```

```
## [1] 1256
```

```
quantile(renta)
```

```
##      0%      25%      50%      75%     100%
```

```
## 694.00 1033.75 1256.00 1638.25 2900.00
```

```
recorrido_intercuartil <- quantile(renta, 0.75) - quantile(renta, 0.25)
```

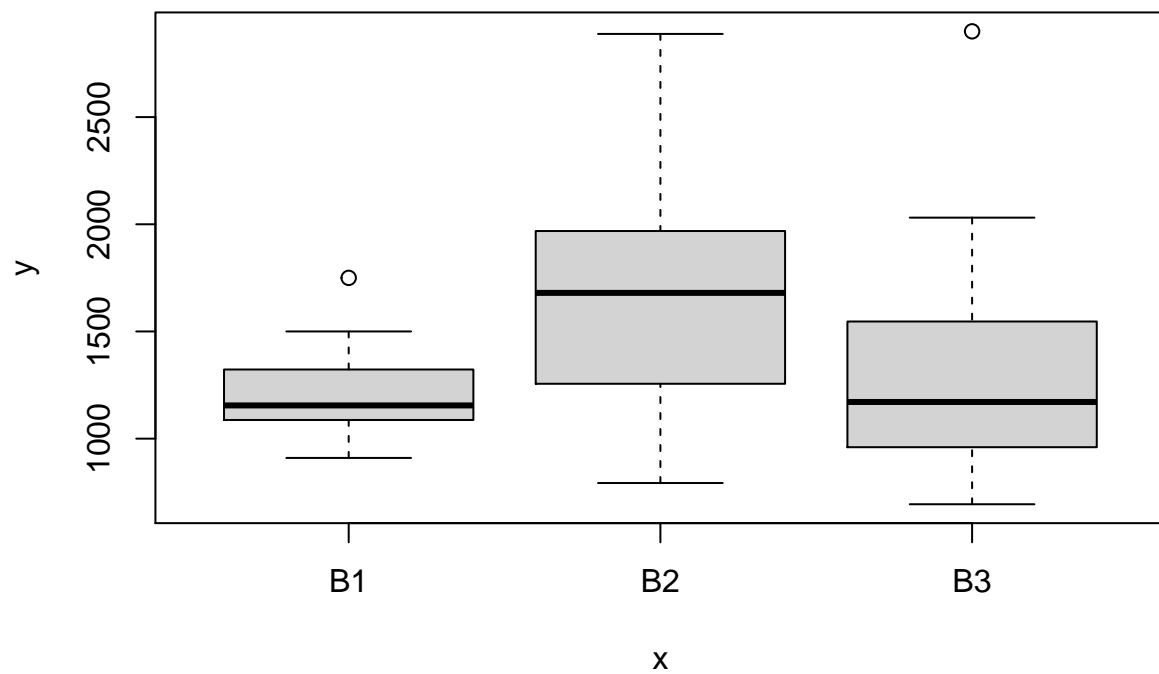
```
recorrido_intercuartil
```

```
##      75%
```

```
## 604.5
```

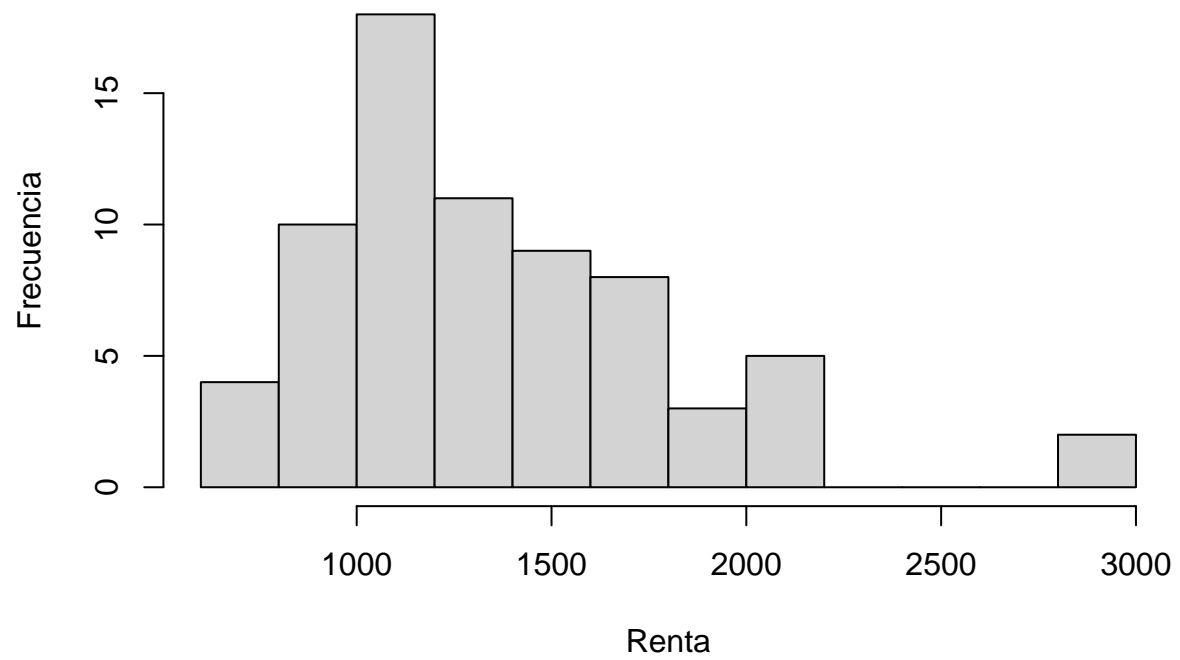
d) Elaborar un histograma con los datos, diferenciando el total y los distintos barrios.

```
plot(as.factor(data$barrio), data$renta )
```



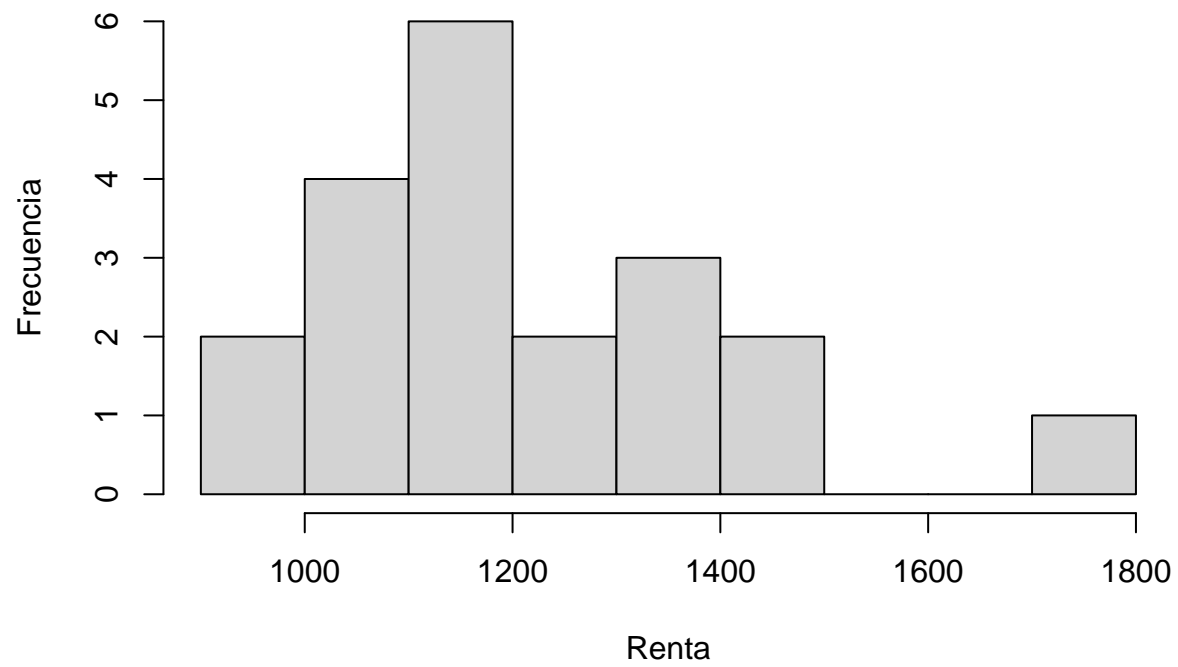
```
hist(data$renta, main="Histograma de la renta", xlab="Renta", ylab="Frecuencia")
```

Histograma de la renta



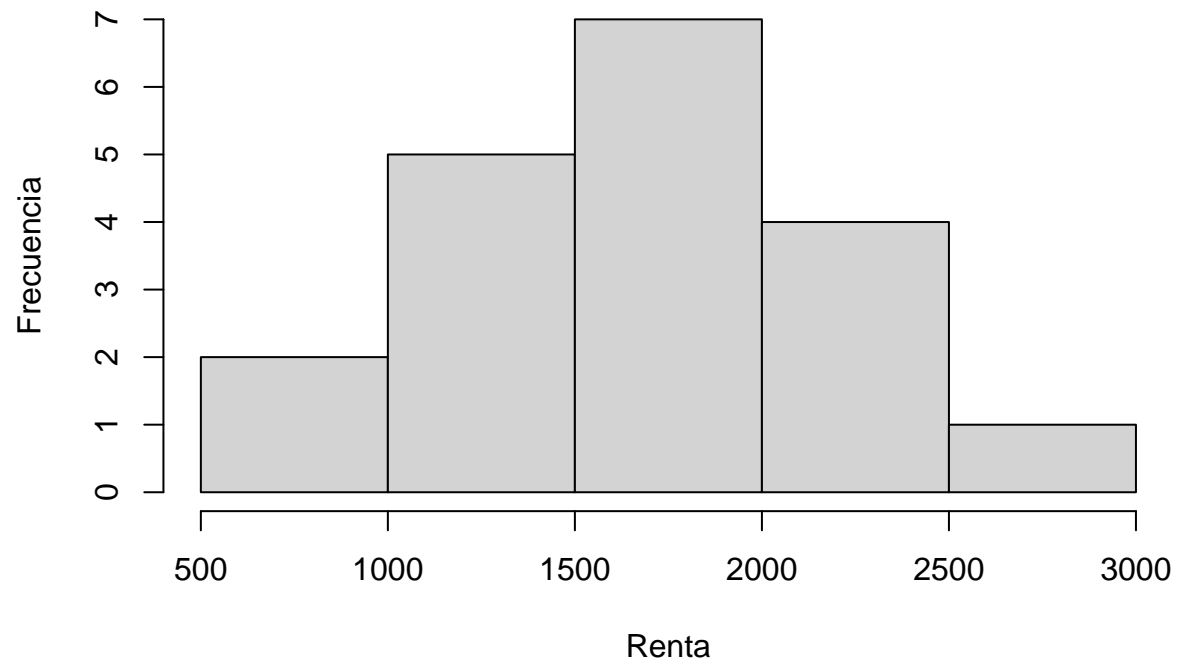
```
hist(data$renta[data$barrio=="B1"], main="Histograma renta en B1", xlab="Renta", ylab="Frecuencia")
```

Histograma renta en B1



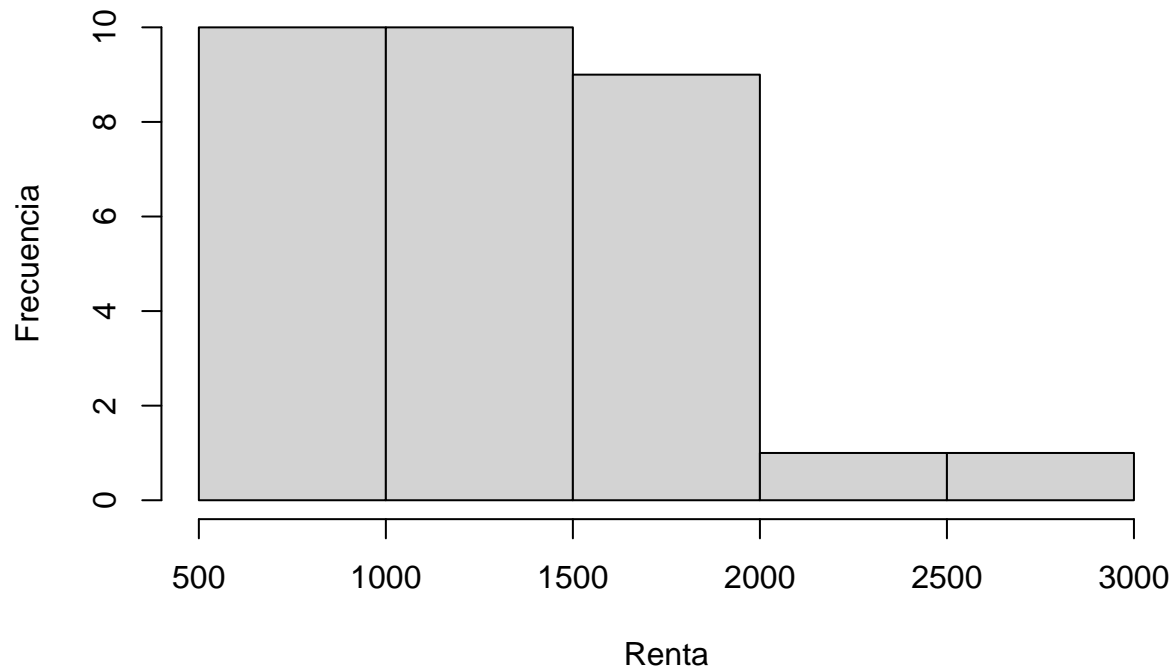
```
hist(data$renta[data$barrio=="B2"], main="Histograma renta en B2", xlab="Renta", ylab="Frecuencia")
```

Histograma renta en B2



```
hist(data$renta[data$barrio=="B3"], main="Histograma renta en B3", xlab="Renta", ylab="Frecuencia")
```

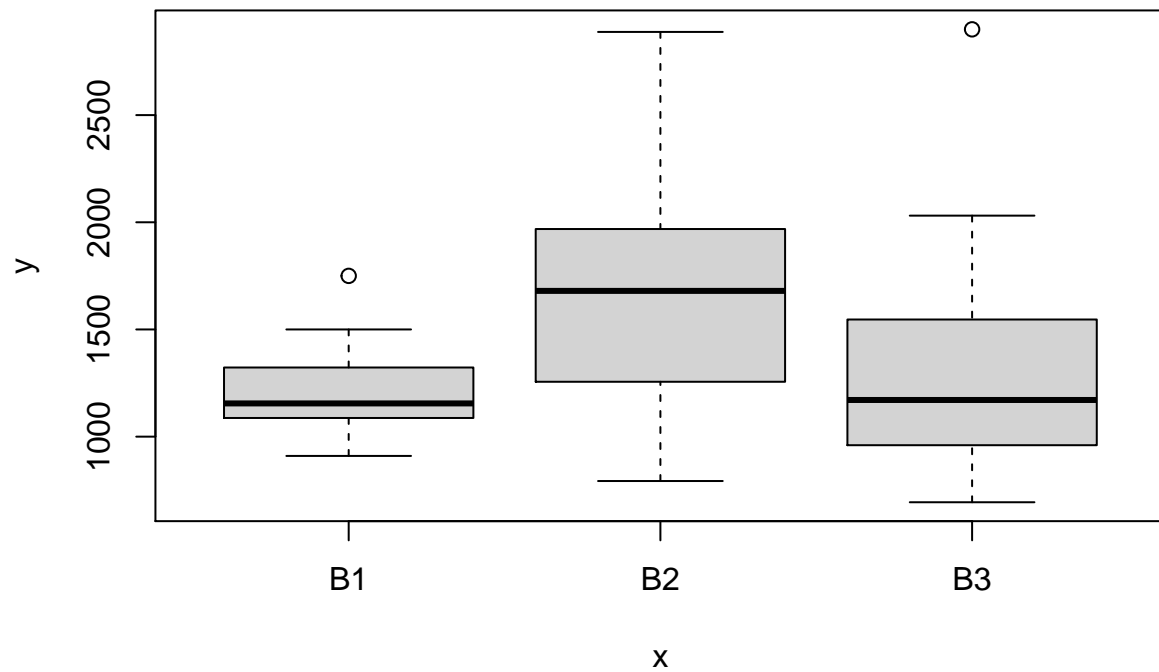
Histograma renta en B3



e y f) Construir un diagrama de caja y mostrar los casos atípicos y otros elementos relevantes del mismo.

Para analizar los outliers crearemos un gráfico y usaremos su propiedad out. Como podemos observar estos valores son los 1750 y 2900 pertenecientes al B1 y B3 respectivamente. Por otra parte en el histograma podemos observar los valores fuera de rango son aquellos mas cercanos a los 3000 se pueden observar de forma sencilla debido la gran separación que existe entre estos y los valores cercanos a los 2200.

```
graph <- plot(as.factor(data$barrio),data$renta )
```

```
graph$out
```

```
## [1] 1750 2900
```

```
graph
```

```
## $stats
```

```
##      [,1]  [,2]  [,3]
```

```
## [1,]  910.0  793.0  694.0
```

```
## [2,] 1087.0 1256.0  960.0
```

```
## [3,] 1155.0 1680.0 1171.0
```

```
## [4,] 1322.5 1968.5 1546.5
```

```
## [5,] 1500.0 2888.0 2031.0
```

```
##
```

```
## $n
```

```
## [1] 20 19 31
```

```
##
```

```
## $conf
```

```
##      [,1]      [,2]      [,3]
```

```
## [1,] 1071.798 1421.735 1004.565
```

```
## [2,] 1238.202 1938.265 1337.435
```

```
##
```

```
## $out
```

```
## [1] 1750 2900
```

```
##
```

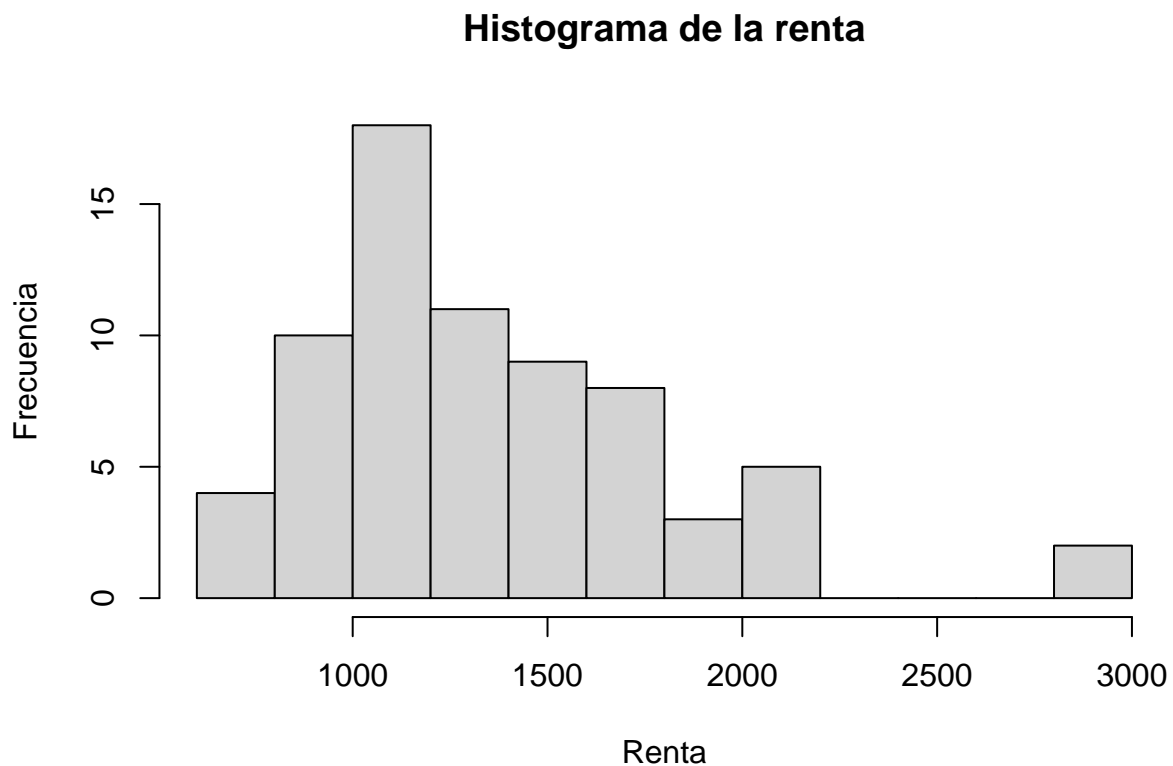
```
## $group
```

```
## [1] 1 3
```

```
##
## $names
## [1] "B1" "B2" "B3"
data$barrio[data$renta == 1750]

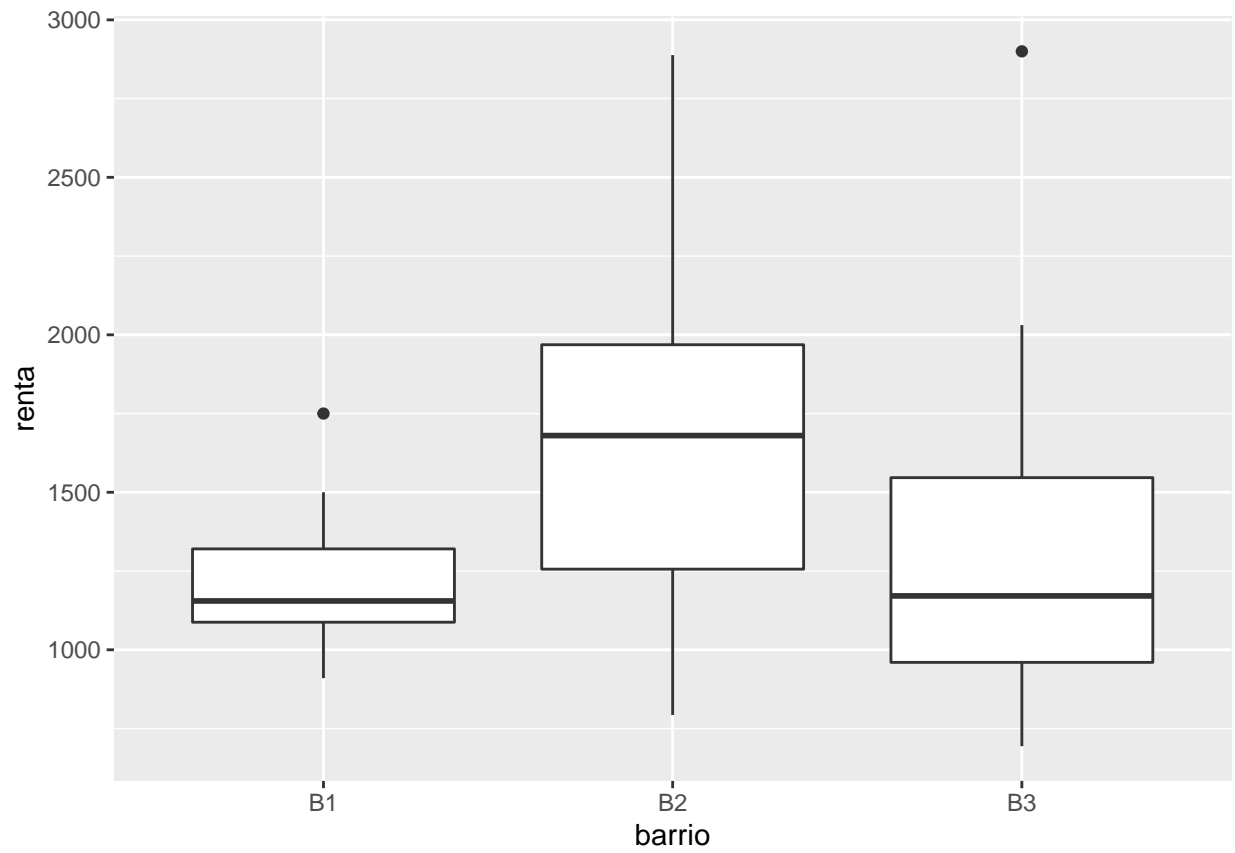
## [1] "B1"
data$barrio[data$renta == 2900]

## [1] "B3"
hist(data$renta, main="Histograma de la renta", xlab="Renta", ylab="Frecuencia")
```

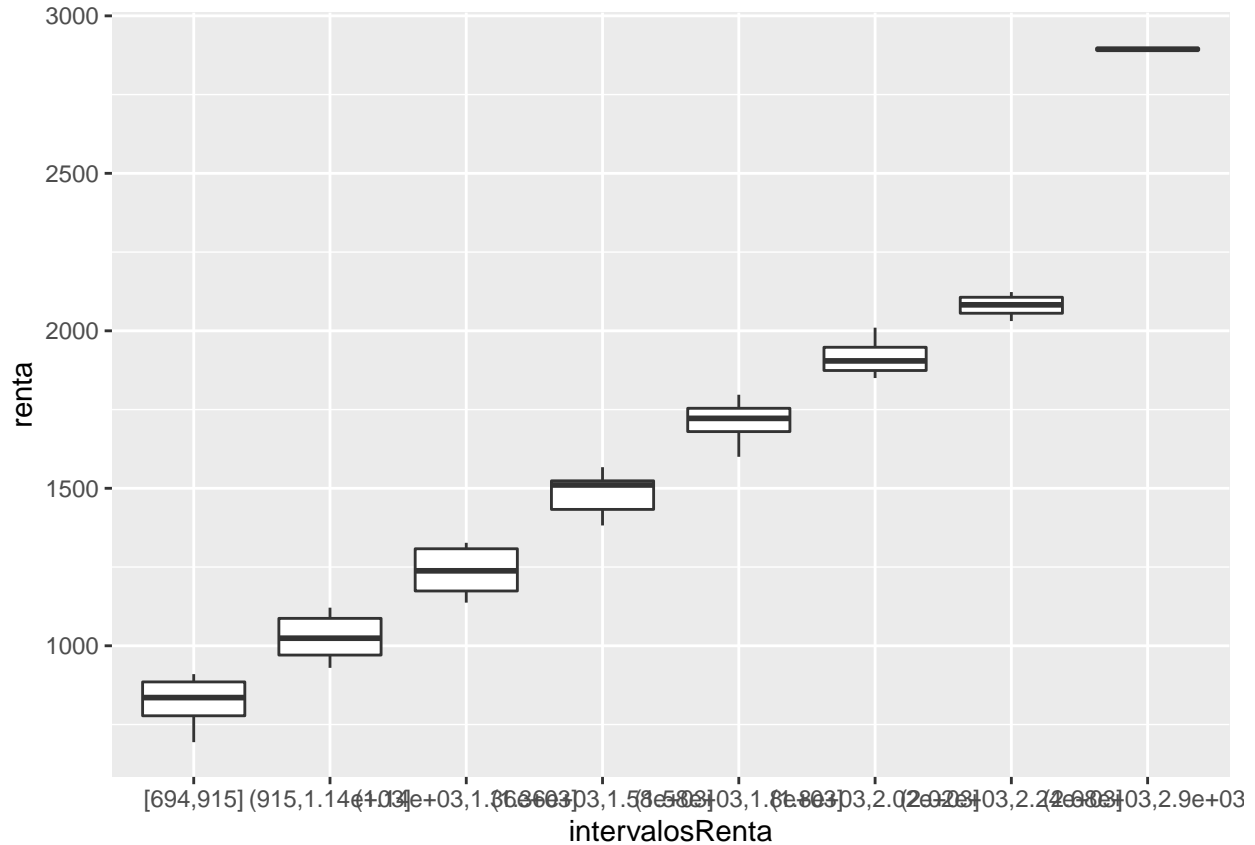


h) Utilizar la librería ggplot2 para visualizar los apartados d) y e) cuando se hace el análisis según barrios.

```
ggplot(data = data, aes(barrio, renta)) + geom_boxplot()
```



```
ggplot(data = data, aes(intervalosRenta, renta)) + geom_boxplot()
```



i) Establecer justificadamente las conclusiones a partir de las medidas utilizadas.

Por una parte podemos observar claramente que el barrio con mejor estado económico es el B2 siendo aquel con tanto las rentas mas altas asi como las mejores rentas en promedio. Por otra parte tenemos b2 cuya media es superior en b3, sin embargo son dos casos curiosos por una parte la desviación es muy pequeña en B1 por lo tanto existe poca dispersión en sus datos y es por ello que su media excede a b3 pero por otra parte en B3 existen casos con mejores rentas que en b1 aunque teniendo mayor dispersión.