

Práctica2

Preparación de la realización del estudio

Primero cargamos la librería carData y knitr. Necesarias para la realización de la práctica. Por último guardaremos el dataset en una variable llamada data y investigamos el contenido usando head, tail y summary que nos dejen ver el principio, final y un resumen de los datos.

a) Solucionar los valores NA

R hace un gran esfuerzo por añadir funcionalidades para limpiar los datos antes de su análisis. Es por esto que contiene una función llamada na.omit que hará el trabajo por nosotros. Almacenamos los datos limpios en una variable s

```
s <- na.omit(data)
```

b y c) Encontrar el país con mayor y menor mortalidad infantil y mayor y menor renta

Para ello primero debemos de añadir una nueva columna que sean los diferentes países. Para ello añadiremos a p una nueva columna llamada pais que contendrá los diferentes nombres de los países que antes de esto estarían como nombre de las filas.

Una vez terminamos este tratamiento de datos será fácil calcular países con mejor y peor renta y mayor y menor mortalidad infantil. Simplemente filtramos los países por la renta máxima y mínima usando la función which.

```
row.names(s)
```

##	[1]	"Australia"	"Austria"
##	[3]	"Belgium"	"Canada"
##	[5]	"Denmark"	"Finland"
##	[7]	"France"	"West.Germany"
##	[9]	"Ireland"	"Italy"
##	[11]	"Japan"	"Netherlands"
##	[13]	"New.Zealand"	"Norway"
##	[15]	"Portugal"	"South.Africa"
##	[17]	"Sweden"	"Switzerland"
##	[19]	"Britain"	"United.States"
##	[21]	"Algeria"	"Ecuador"
##	[23]	"Indonesia"	"Iraq"
##	[25]	"Libya"	"Nigeria"
##	[27]	"Saudi.Arabia"	"Venezuela"
##	[29]	"Argentina"	"Brazil"
##	[31]	"Chile"	"Colombia"
##	[33]	"Costa.Rica"	"Dominican.Republic"
##	[35]	"Greece"	"Guatemala"
##	[37]	"Israel"	"Jamaica"

```
## [39] "Lebanon" "Malaysia"
## [41] "Mexico" "Nicaragua"
## [43] "Panama" "Peru"
## [45] "Singapore" "Spain"
## [47] "Taiwan" "Trinidad.and.Tobago"
## [49] "Tunisia" "Uruguay"
## [51] "Yugoslavia" "Zambia"
## [53] "Bolivia" "Cameroon"
## [55] "Congo" "Egypt"
## [57] "El.Salvador" "Ghana"
## [59] "Honduras" "Ivory.Coast"
## [61] "Jordan" "South.Korea"
## [63] "Liberia" "Moroco"
## [65] "Papua.New.Guinea" "Paraguay"
## [67] "Philippines" "Syria"
## [69] "Thailand" "Turkey"
## [71] "South.Vietnam" "Afganistan"
## [73] "Bangladesh" "Burma"
## [75] "Burundi" "Cambodia"
## [77] "Central.African.Republic" "Chad"
## [79] "Dahomey" "Ethiopia"
## [81] "Guinea" "India"
## [83] "Kenya" "Madagascar"
## [85] "Malawi" "Mali"
## [87] "Mauritania" "Niger"
## [89] "Pakistan" "Rwanda"
## [91] "Sierra.Leone" "Somalia"
## [93] "Sri.Lanka" "Sudan"
## [95] "Tanzania" "Togo"
## [97] "Uganda" "Upper.Volta"
## [99] "Southern.Yemen" "Yemen"
## [101] "Zaire"
```

```
p <- data.frame(s, pais = row.names(s))
```

```
names(p) <- c("renta", "mortalidad_infantil", "region", "productor_petroleo", "pais")
```

```
max(p$renta)
```

```
## [1] 5596
```

```
pais_mayor_renta <- p$pais[which.max(p$renta)];pais_mayor_renta
```

```
## [1] "Sweden"
```

```
pais_menor_renta <- p$pais[which.min(p$renta)];pais_menor_renta
```

```
## [1] "Mali"
```

```
pais_mayor_mi <- p$pais[which.max(p$mortalidad_infantil)];pais_mayor_mi
```

```
## [1] "Saudi.Arabia"
```

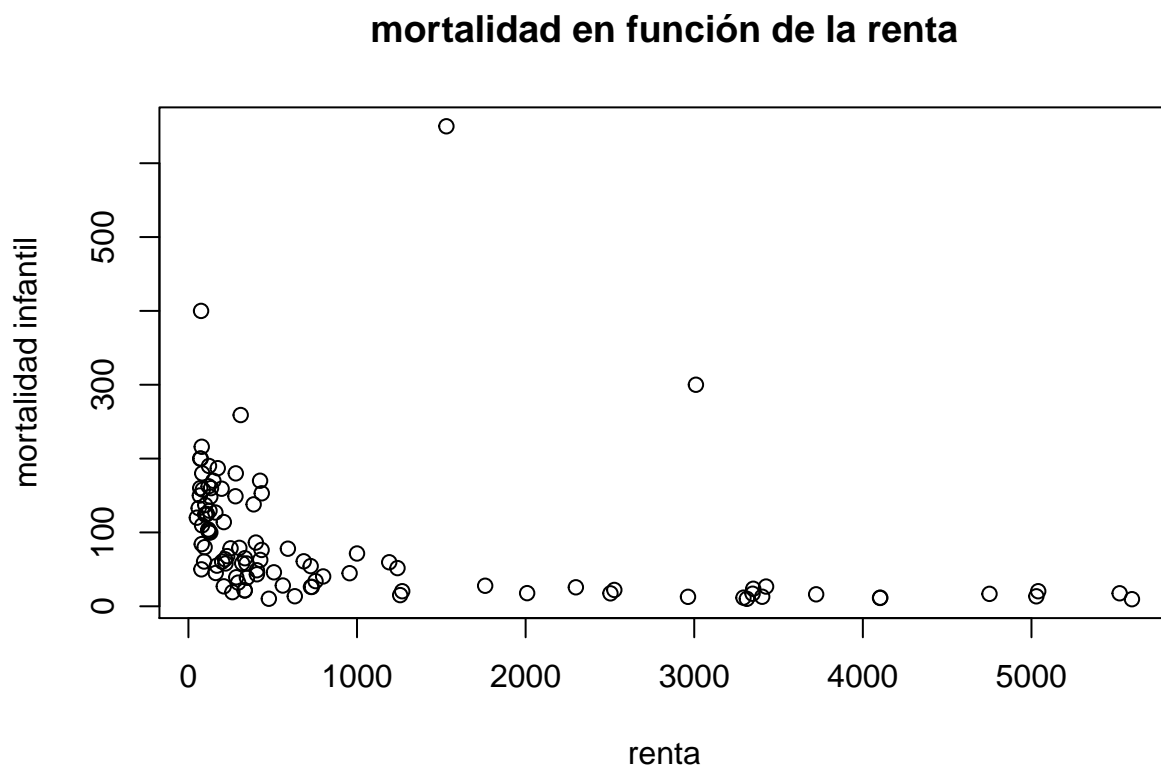
```
pais_menor_mi <- p$pais[which.min(p$mortalidad_infantil)];pais_menor_mi
```

```
## [1] "Sweden"
```

d) Analizar las rentas per cápita y la mortalidad infantil en función del continente. Calcular las medias, medianas y desviación estándar correspondientes.

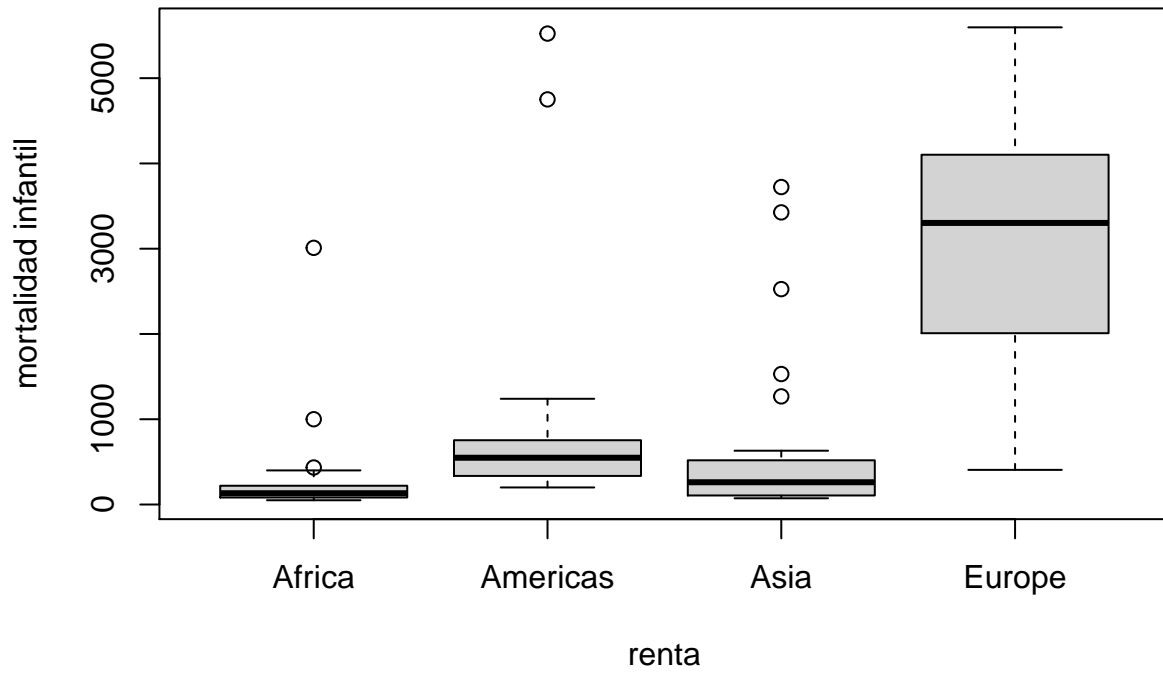
Para analizar la renta per capita y la mortalidad infantil en función del continente hemos creado unas graficas usando plot que muestran esta relación. A su vez usando la función mean (media), median (mediana), sd (standart desviation) y aggregate para crear las diferentes tablas hemos podido calcular con facilidad las medias, desviación estandar y medianas de todos los continentes.

```
plot(p$renta, p$mortalidad_infantil,  
     xlab="renta",  
     ylab="mortalidad infantil",  
     main="mortalidad en función de la renta")
```



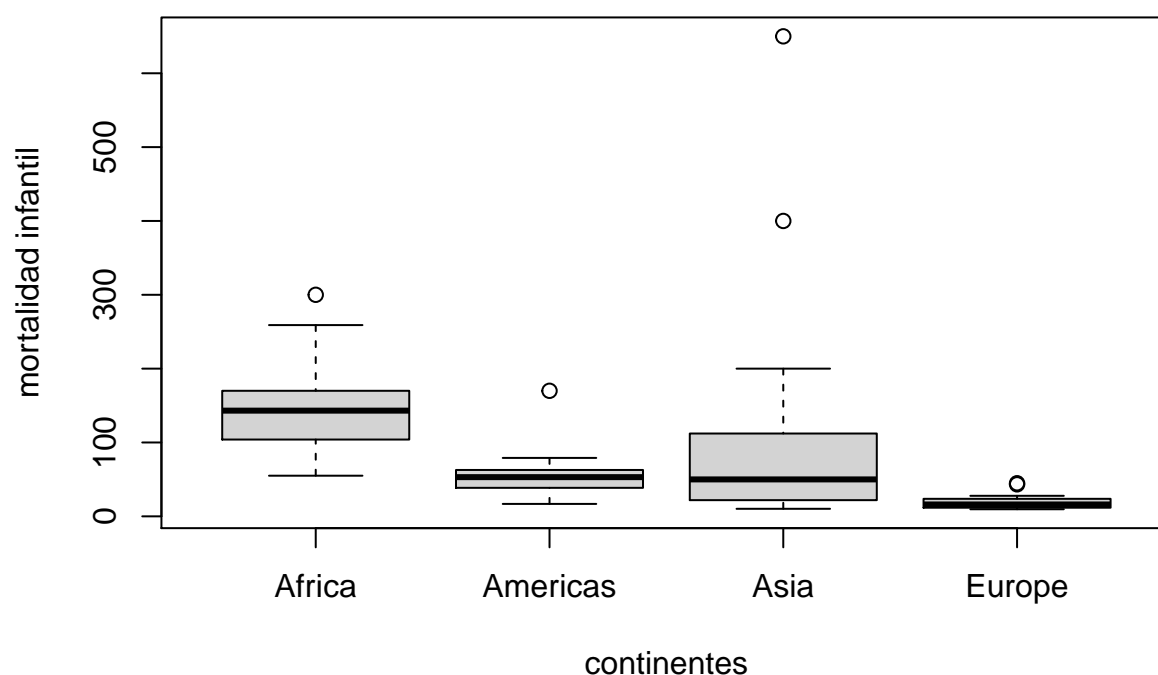
```
plot(p$region, p$renta,  
     xlab="renta",  
     ylab="mortalidad infantil",  
     main="renta en función del pais")
```

renta en función del pais



```
plot(p$region, p$mortalidad_infantil,  
     xlab="continentes",  
     ylab="mortalidad infantil",  
     main="mortalidad infantil en función de los continentes")
```

mortalidad infantil en función de los continentes



#MEDIA Y MEDIANA GLOBALES DE RENTA PER CAPITA

```
mean(p$renta)
```

```
## [1] 1022.337
```

```
median(p$renta)
```

```
## [1] 334
```

```
sd(p$renta)
```

```
## [1] 1435.642
```

```
tabla_medias <- aggregate(cbind(income, infant)~region, data=s, mean)
kable(tabla_medias)
```

region	income	infant
Africa	273.2353	142.29118
Americas	978.0455	55.12273
Asia	656.4815	96.17037
Europe	3040.2222	19.25556

```
tabla_medianas <- aggregate(cbind(income, infant)~region, data=s, median)
kable(tabla_medianas)
```

region	income	infant
Africa	132.0	143.15
Americas	548.5	53.05
Asia	261.0	50.00
Europe	3302.0	16.05

```
tabla_sd <- aggregate(cbind(income, infant)~region, data=s, sd)
kable(tabla_sd)
```

region	income	infant
Africa	514.3173	54.94327
Americas	1380.0034	31.59109
Asia	1002.6715	137.46338
Europe	1439.2799	10.47418

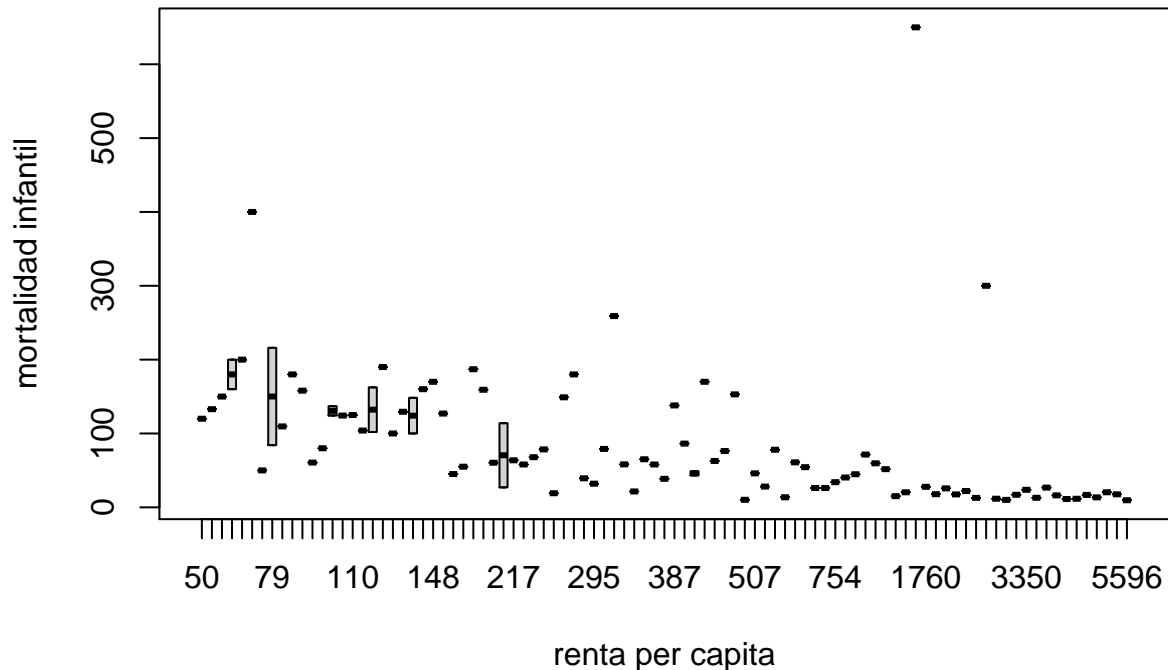
e) Visualizar gráficamente, utilizando `boxplot()`, una comparativa de los datos de renta por un lado y de mortalidad infantil por otro. Establecer justificadamente las conclusiones.

Como observamos en el gráfico posterior como norma general conforme aumenta la renta per capita disminuye drásticamente los casos de mortalidad infantil. Es lógico pensar que el motivo es debido a que en los países con mayores recursos y capacidades monetarias los jóvenes se dedican en gran parte al estudio y la violencia en las calles decrece ya que las necesidades principales en general estan cubiertas (comida, agua...) aunque esto no puede saberse con los datos que tenemos.

A su vez, podemos observar a su vez la presencia de outliers en las zonas mas desarrolladas, esto puede deberse a países con gran diferencia entre ricos y pobres, puntos concretos de violencia, etc... un ejemplo de esto es Arabia Saudi uno de los países con mayor renta per capita y a su vez el país con mayor mortalidad infantil como podemos observar debajo.

```
boxplot(mortalidad_infantil~renta, data = p, main="mortalidad infantil por rpc", xlab="renta per capita")
```

mortalidad infantil por rpc



```
p$pais[which.max(p$mortalidad_infantil)]
```

```
## [1] "Saudi.Arabia"
```

```
p$renta[p$pais == "Saudi.Arabia"]
```

```
## [1] 1530
```

f) Encontrar si hay influencia significativa entre la renta per cápita y la mortalidad infantil en países no productores

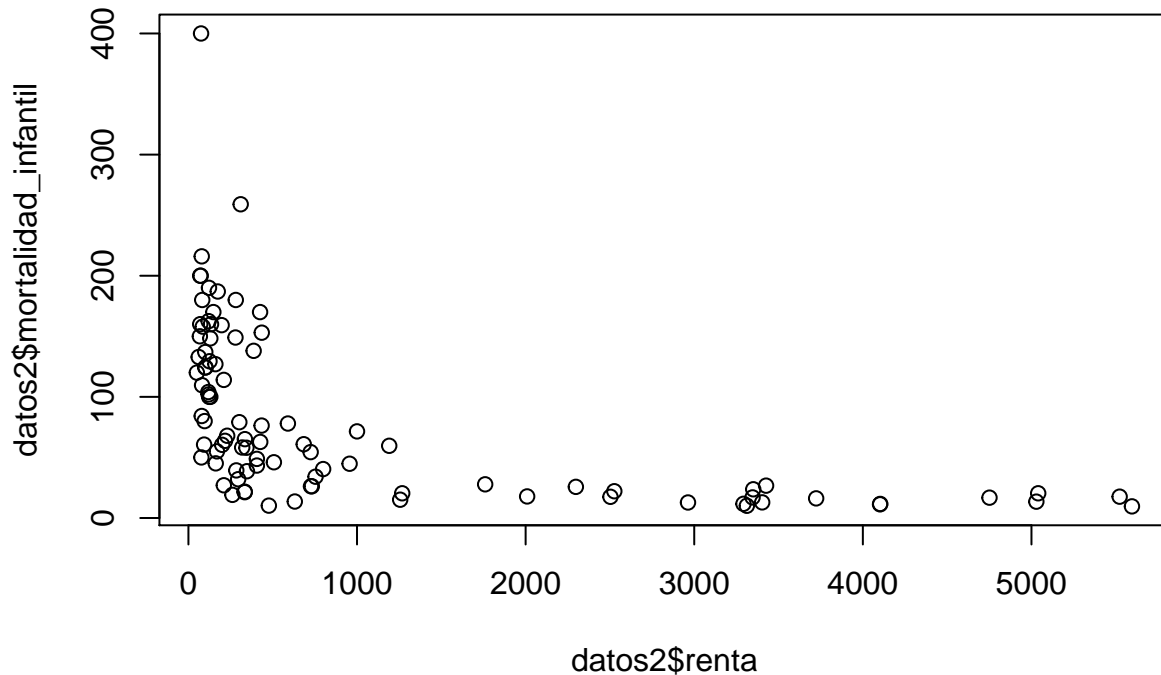
Para ello creamos un nuevo data frame con la renta y mortalidad infantil de los países no productores de petróleo. A su vez mostraremos un plot de estos y la media de la mortalidad infantil en países no productores y si productores.

Como observamos sigue habiendo una fuerte correlación entre las rentas y la mortalidad infantil a si que este suceso no se da únicamente en los países productores de petróleo.

Sin embargo es curioso observar las medias de la mortalidad infantil en países no productores y si productores. Los no productores tienen una cifra casi de la mitad que los que producen petróleo mostrando que aunque mientras menor renta seguirá siendo mayor la mortalidad infantil, la producción de petróleo influye drásticamente en el aumento de la mortalidad infantil.

También podría analizarse de la siguiente manera, los países que producen petróleo son los que menor renta tienen y los que no los que mayor, sin embargo debajo se han calculado las medias de las rentas de ambos grupos y aunque la diferencia es palpable, no es abrumadora, mostrando que definitivamente la producción de petróleo influye de forma negativa en la mortalidad infantil.

```
datos2 <- data.frame(renta = p$renta[p$productor_petroleo == "no"], mortalidad_infantil = p$mortalidad_infantil)
plot(datos2$renta, datos2$mortalidad_infantil)
```



```
mean(datos2$mortalidad_infantil)
```

```
## [1] 81.89462
```

```
mean(p$mortalidad_infantil[p$productor_petroleo=="yes"])
```

```
## [1] 172.2
```

```
mean(datos2$renta)
```

```
## [1] 1031.57
```

```
mean(p$renta[p$productor_petroleo=="yes"])
```

```
## [1] 915
```

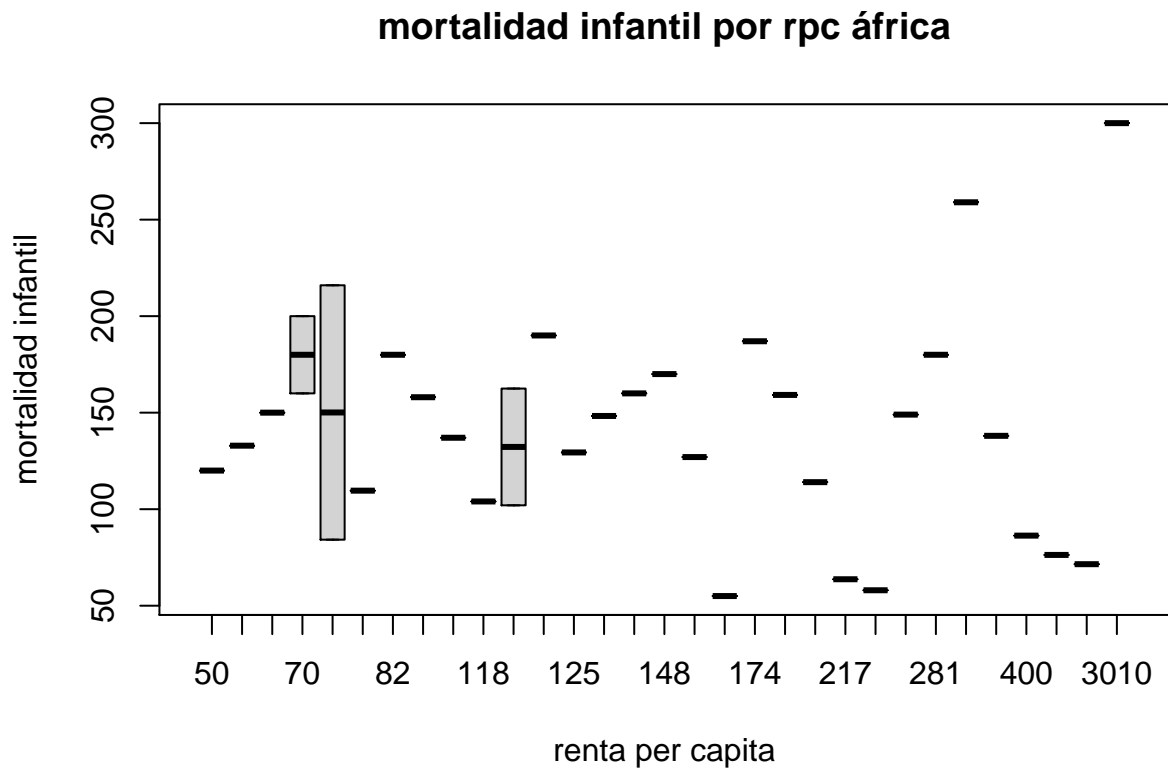
g) realizar el apartado e en el continente Áfricano

En África la mortalidad infantil independientemente de la renta no baja excepto en dos casos concretos muy curiosos cuya renta excede o iguala 1000 euros, sudafrica y libya. Entre sudafrica y libya encontramos una gran dicotomía digna de estudio por una parte ambos son países con una renta que excede o iguala los 1000 euros pero la mortalidad infantil en sudafrica es la mayor en todo el continente y en libya tenemos una de las menores. No tenemos datos con los que poder analizar la causa de estos acontecimientos y aunque intentemos estudiar la situación social de ambos (puesto que este podría ser el motivo) ambos son países con

gran separación entre ricos y pobres, constantes guerras y en situación desfavorecida, aunque en Sudafrica denota mayores problemáticas como el ya derogado apartheid, los diamantes de sangre, la gran explotación esclava, la criminalidad y su inmensa corrupción.

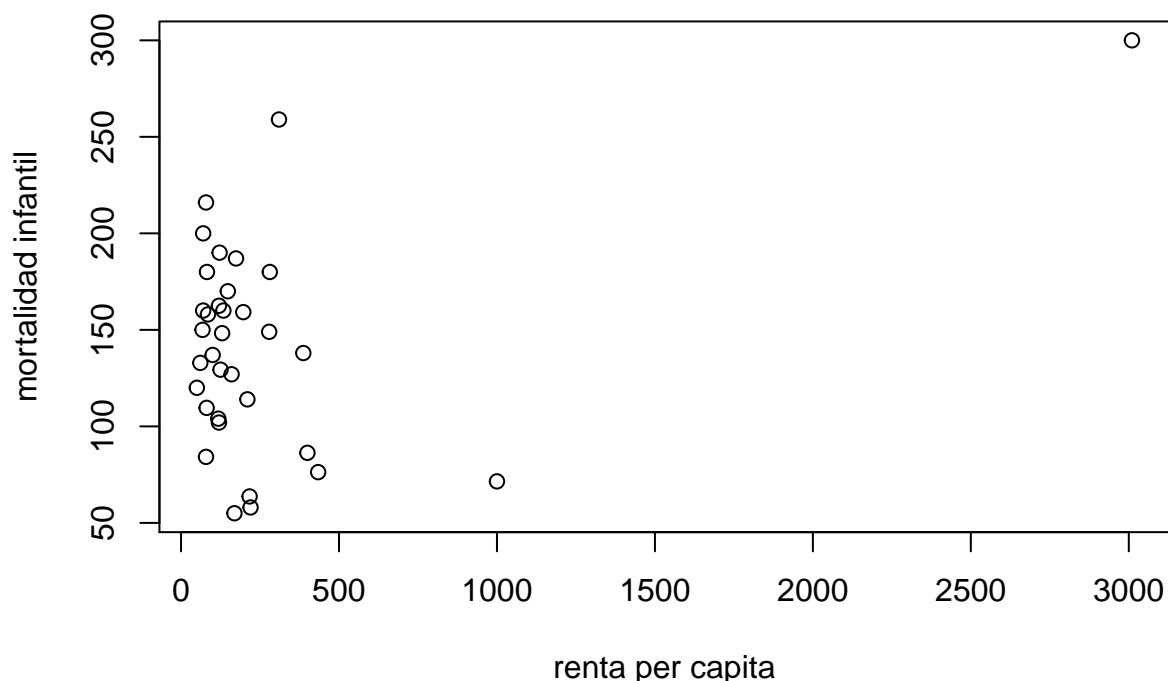
```
datos_africa <- data.frame(pais = p$pais[p$region=="Africa"],renta = p$renta[p$region=="Africa"], mortalidad_infantil = p$mortalidad_infantil)

boxplot(mortalidad_infantil~renta, data = datos_africa, main="mortalidad infantil por rpc áfrica", xlab="renta per capita", ylab="mortalidad infantil")
```



```
plot( datos_africa$renta, datos_africa$mortalidad_infantil,main="mortalidad infantil por rpc áfrica", xlab="renta per capita", ylab="mortalidad infantil")
```

mortalidad infantil por rpc África



```
datos_africa$pais[datos_africa$renta > 800]
```

```
## [1] "South.Africa" "Libya"
```

```
datos_africa$mortalidad_infantil[datos_africa$pais=="Libya"]
```

```
## [1] 300
```

```
datos_africa$mortalidad_infantil[datos_africa$pais=="South.Africa"]
```

```
## [1] 71.5
```

h) Visualizar los datos porcentuales por continente utilizando el comando pie()

Como podemos observar la mayor parte de la gráfica la ocupa África uno de los continentes con mayor violencia y menor renta per capita, siguiendole asia, américa y por último europa, el que menos porción tiene.

```
mortalidad_infantil_sum = na.omit(aggregate( s$infant~s$region , data = s, sum))
colnames(mortalidad_infantil_sum) <- c("region", "infant")
mortalidad_infantil_sum$infant
```

```
## [1] 4837.9 1212.7 2596.6 346.6
```

```
pie( mortalidad_infantil_sum$infant, labels=mortalidad_infantil_sum$region)
```

