

Laboratorio 1  
Estadística Para Ingenieros  
Francisco Sucre 10-10717  
José Cipagauta 05-38040

## **PARTE I**

```
> #Primero, se necesitan los datos, attach para facil acceso después.  
> data = read.table("IndiceCrim.dat", header = T)  
> attach(data)  
> #1  
> #R = Índice de criminalidad, número de delitos conocidos por la policía por cada millón de habitantes.  
> summary(R) #Del summary obtenemos min, max, cuartiles, mediana, media  
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
 34.20  65.85   83.10   90.51  105.80  199.30  
> var(R) #Varianza  
[1] 1495.854  
> sd(R) #Desviación estándar  
[1] 38.67627  
> range(R) #Rango  
[1] 34.2 199.3  
> IQR(R) #Rango intercuartil  
[1] 39.9
```

**Análisis:** Con estos datos podemos ver en primer lugar que el índice de criminalidad es muy variante entre los estados, teniendo en cuenta que el máximo es 199, que la desviación estándar sea 38 (casi un cuarto del valor) indica que los valores están muy dispersos. Si vemos el resto de las medidas podemos ver que los datos están distribuidos de una manera relativamente uniforme, aunque si vemos el 3er cuartil podemos ver que el aumento es mayor después de 105 hasta llegar en 190, Así que podemos ver una diferencia notable en estos estados, una diferencia que si es causada por datos errados entonces afecta de manera grave al resto de las medidas.

```
> #LF = Proporción en participación en trabajos de fuerza por cada mil hombres con edad 14-24.  
> summary(LF)  
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   
 480.0  530.5   560.0   561.2   593.0   641.0  
> var(LF)  
[1] 1633.115  
> sd(LF)  
[1] 40.41181  
> range(LF)  
[1] 480 641  
> IQR(LF)  
[1] 62.5
```

>

**Análisis:** En este caso podemos ver que los datos no se encuentran tan dispersos, ya que la desviación estándar es más pequeña en comparación al máximo de los datos, los datos también tienen una distribución relativamente uniforme, en el 75% de los estados no se llega a una proporción de 600

> #X = Desigualdad en ingresos, el número de familias por mil que ganan por debajo de la mitad de la mediana de ingresos.

> summary(X)

```
Min. 1st Qu. Median Mean 3rd Qu. Max.
126.0 165.5 176.0 194.0 227.5 276.0
```

> var(X)

```
[1] 1591.696
```

> sd(X)

```
[1] 39.89606
```

> range(X)

```
[1] 126 276
```

> IQR(X)

```
[1] 62
```

**Análisis:** Podemos ver que los datos no andan muy dispersos ya que la desviación estándar no se encuentra en una proporción tan elevada comparada con el rango de los valores, también podemos ver que la mediana es algo distinta pero no es tanta la diferencia, podemos ver también que hay una distribución relativamente uniforme en los datos, estando el 25% de las muestras por encima de 230 y 25% por debajo de 165

## **PARTE II**

> #2

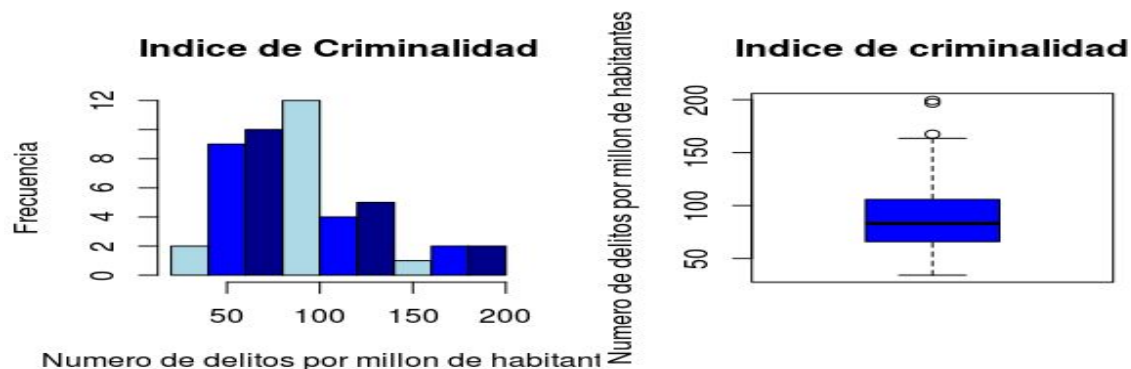
> #Mostrando dos gráficos, histograma y boxplot de cada variable

> #Considerando mostrar los datos etiquetados y coloreados para fácil lectura

> par(mfrow = c(1,2))

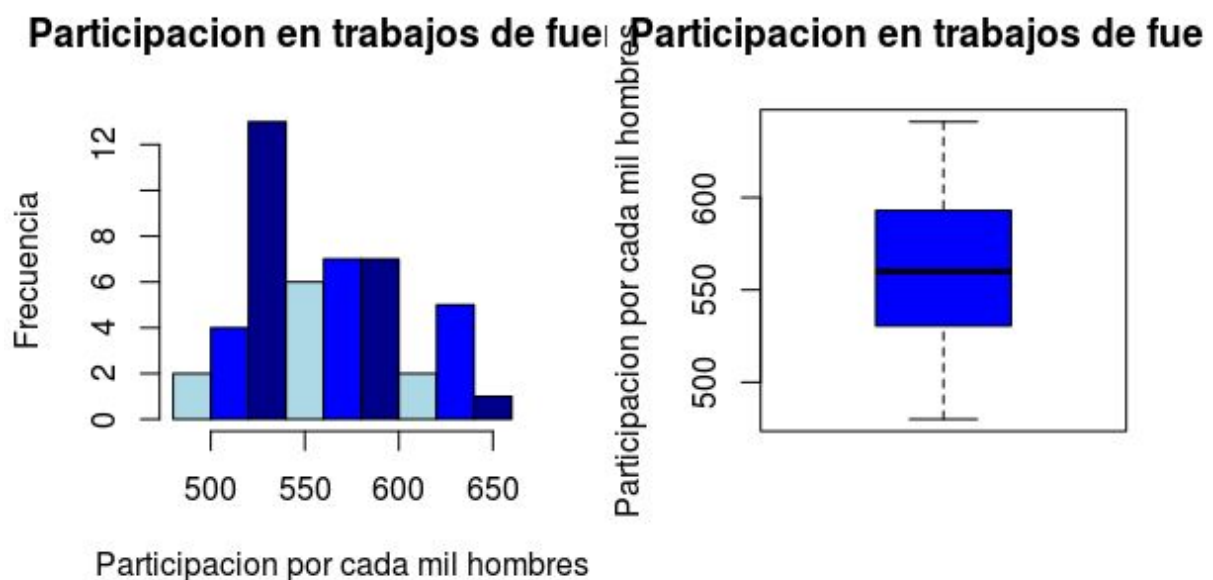
> hist(R, main = "Índice de Criminalidad", xlab = "Número de delitos por millón de habitantes", ylab = "Frecuencia", col = c("lightblue","blue","darkblue"))

> boxplot(R, main = "Indice de criminalidad", ylab = "Numero de delitos por millón de habitantes", col = c("blue"))



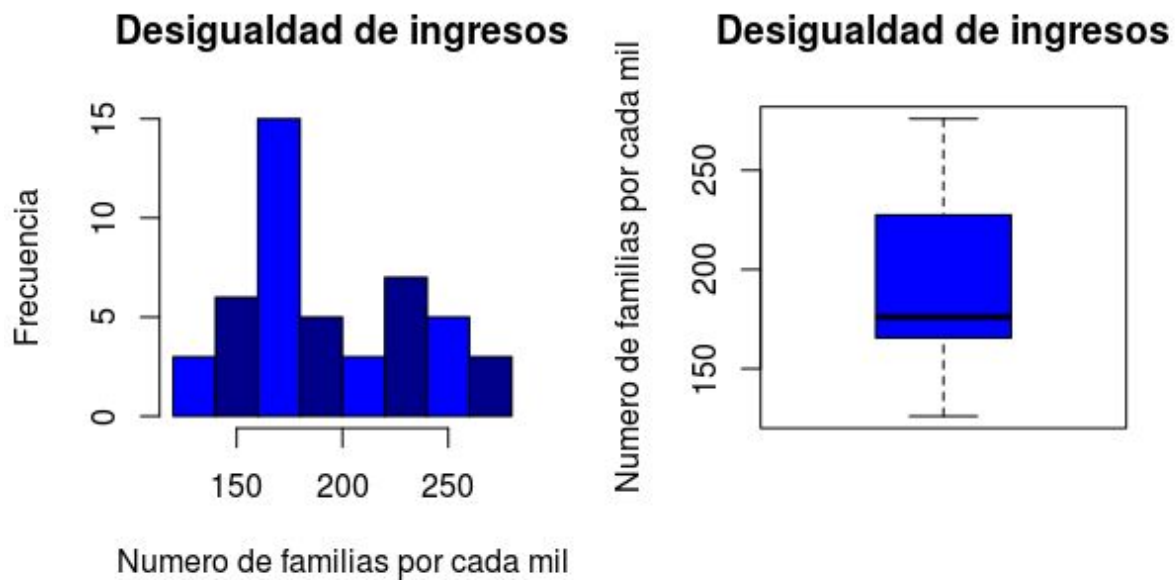
En estos gráficos podemos fácilmente observar que los índices de criminalidad más comunes están entre 50 y 100, teniendo un segundo grupo importante entre 100 y 150, y el resto se refieren a casos más aislados, y como podemos ver en el grafico de caja, las muestras de mayor valor parecen ser datos atípicos y podrían indicar algún dato erróneo, en general el índice de criminalidad es bastante bajo

```
> hist(LF, main = "Participación en trabajos de fuerza", xlab = "Participación por cada mil hombres", ylab = "Frecuencia", col = c("lightblue","blue","darkblue"))
> boxplot(LF, main = "Participación en trabajos de fuerza", ylab = "Participación por cada mil hombres", col = c("blue"))
```



En los gráficos podemos ver que en el caso de esta variable hay una distribución más uniforme de los valores, habiendo una concentración relativamente parecida entre 500 y 550, y la clase de 550 a 600, aunque el resto de las clas aún presentan un número considerable de muestras, en el grafico de caja podemos ver reflejado esto, y se ve perfectamente que entre esas dos clases hay una concentración similar y en esas 2 clases es donde se encuentran la mayoría de las muestras

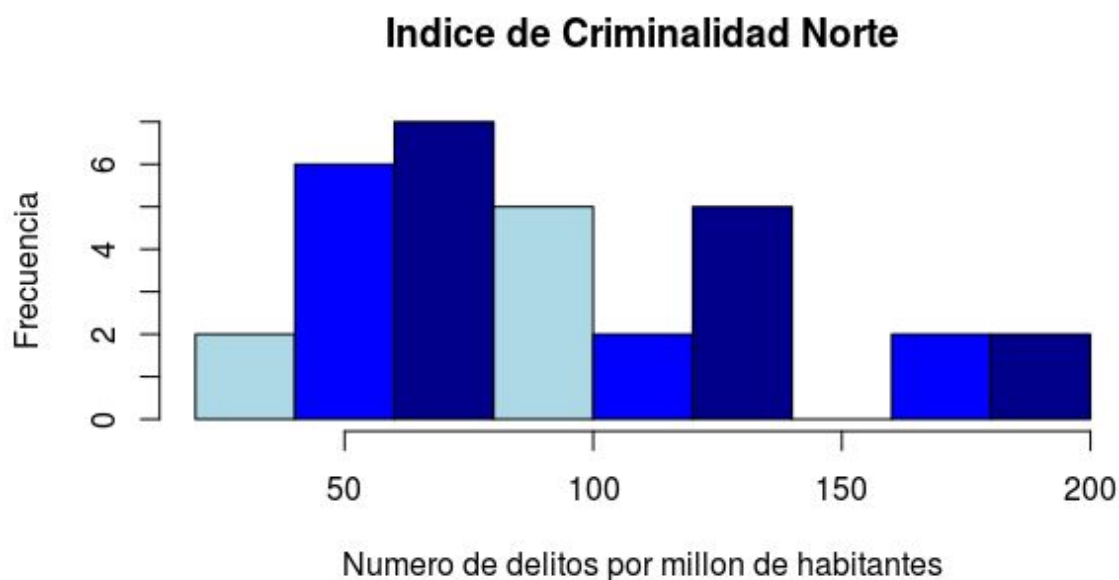
```
> hist(X, main = "Desigualdad de ingresos", xlab = "Número de familias por cada mil", ylab = "Frecuencia", col = c("blue","darkblue"))
> boxplot(X, main = "Desigualdad de ingresos", ylab = "Número de familias por cada mil", col = c("blue"))
```



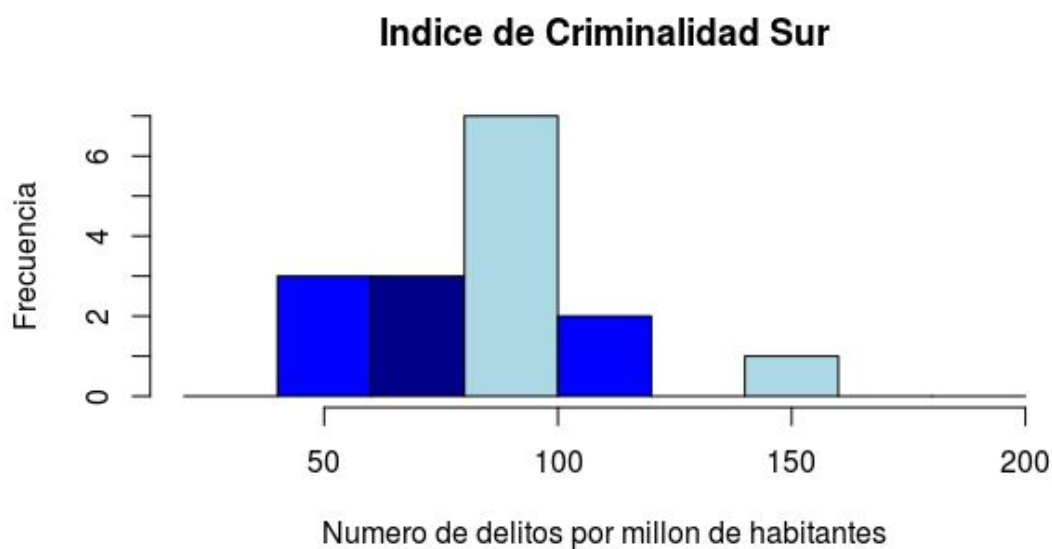
En los gráficos podemos ver que en el caso de esta variable tenemos una distribución diferente, a los gráficos anteriores, aunque no hay datos lo suficientemente dispersos para considerarse atípicos, si podemos ver que el segundo cuartil los datos son muy parecidos mientras que en el 3ro los datos son mucho más dispersos

### PARTE III

```
> #3
> #Mostrando ambos graficos lado a lado para facil comparacion
> #Creando los breakpoints para el histograma, se necesitan los mismos para poder ser comparables
> breakpoints <- c(20,40,60,80,100,120,140,160,180,200)
> hist(R[S==0], breaks = breakpoints, main = "Indice de Criminalidad Norte", xlab = "Numero de delitos por millon de habitantes", ylab = "Frecuencia", col = c("lightblue","blue","darkblue"))
```

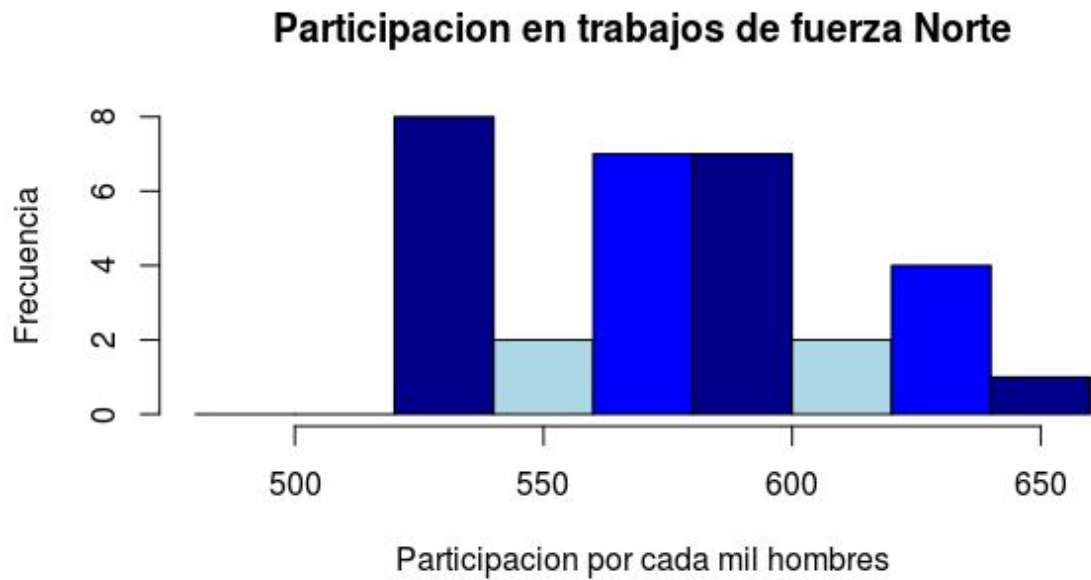


```
> hist(R[S==1], breaks = breakpoints, main = "Indice de Criminalidad Sur", xlab = "Numero de delitos por millon de habitantes", ylab = "Frecuencia", col = c("lightblue","blue","darkblue"))
```

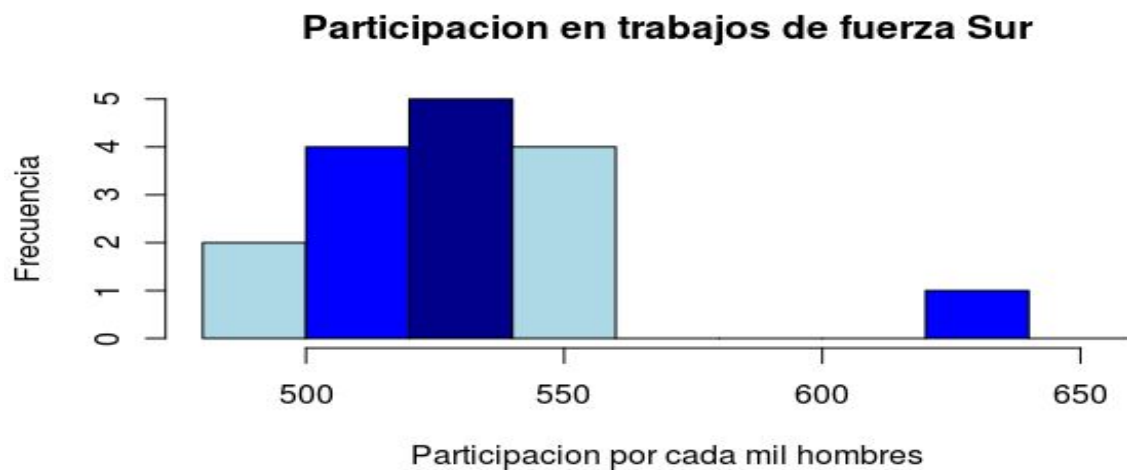


**En los estados del norte hay una mayor concentración de muestras, de lo cual podemos concluir que en el norte hay mayor índice de criminalidad**

```
> #Nuevos breakpoints para nuevos datos
> breakpoints <- c(480,500,520,540,560,580,600,620,640,660)
> hist(LF[S==0], breaks = breakpoints, main = "Participación en trabajos de fuerza Norte", xlab = "Participación por cada mil hombres", ylab = "Frecuencia", col = c("lightblue","blue","darkblue"))
```

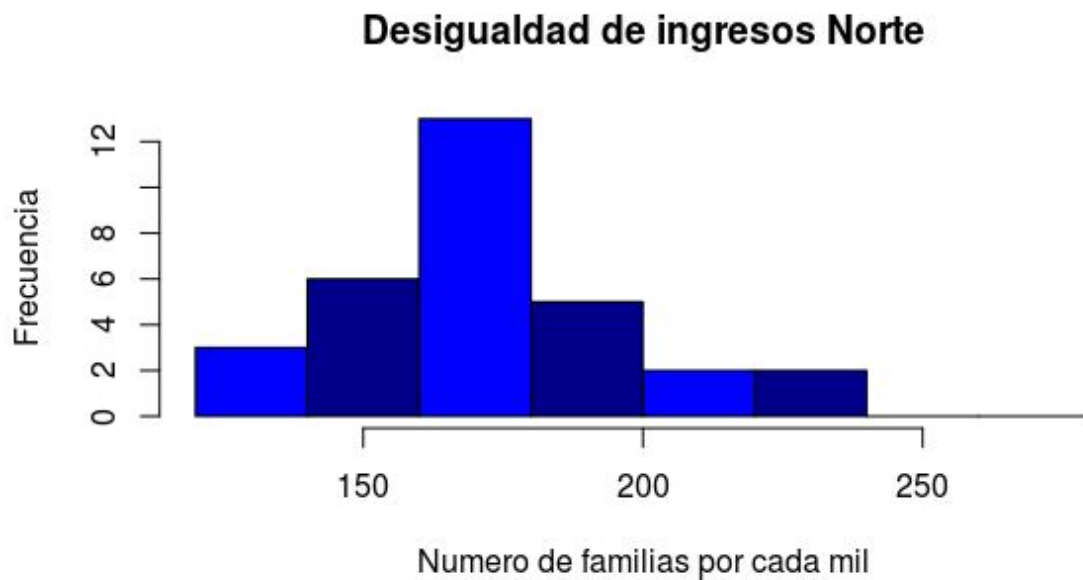


```
> hist(LF[S==1], breaks = breakpoints, main = "Participación en trabajos de fuerza Sur", xlab = "Participación por cada mil hombres", ylab = "Frecuencia", col = c("lightblue","blue","darkblue"))
```

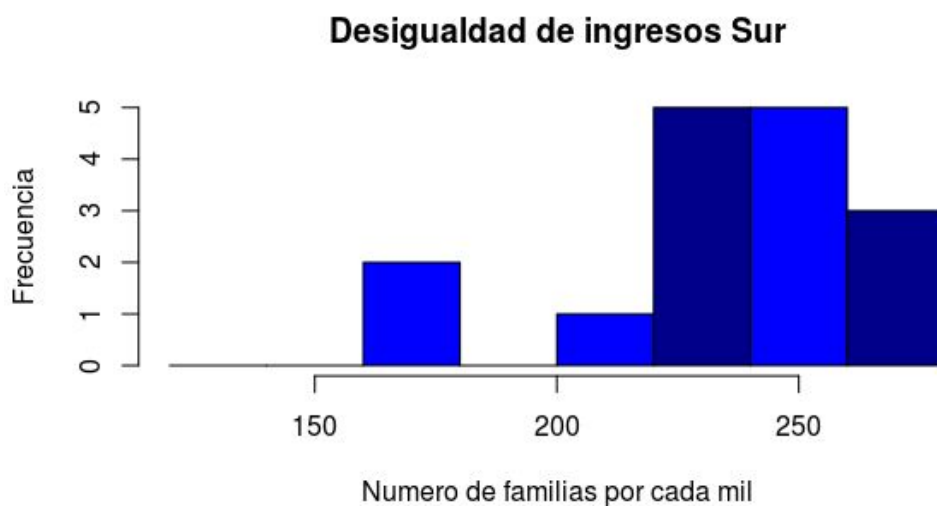


En los estados del norte hay una mayor concentración de muestras y además las muestras en el norte son de mayor valor, con lo cual podemos concluir que hay mayor participación en trabajos de fuerza en los estados del norte

```
> #De nuevo, nuevos datos, nuevos breakpoints
> breakpoints <- c(120,140,160,180,200,220,240,260,280)
> hist(X[S==0], breaks = breakpoints, main = "Desigualdad de ingresos Norte", xlab = "Número de familias por cada mil", ylab = "Frecuencia", col = c("blue","darkblue"))
```

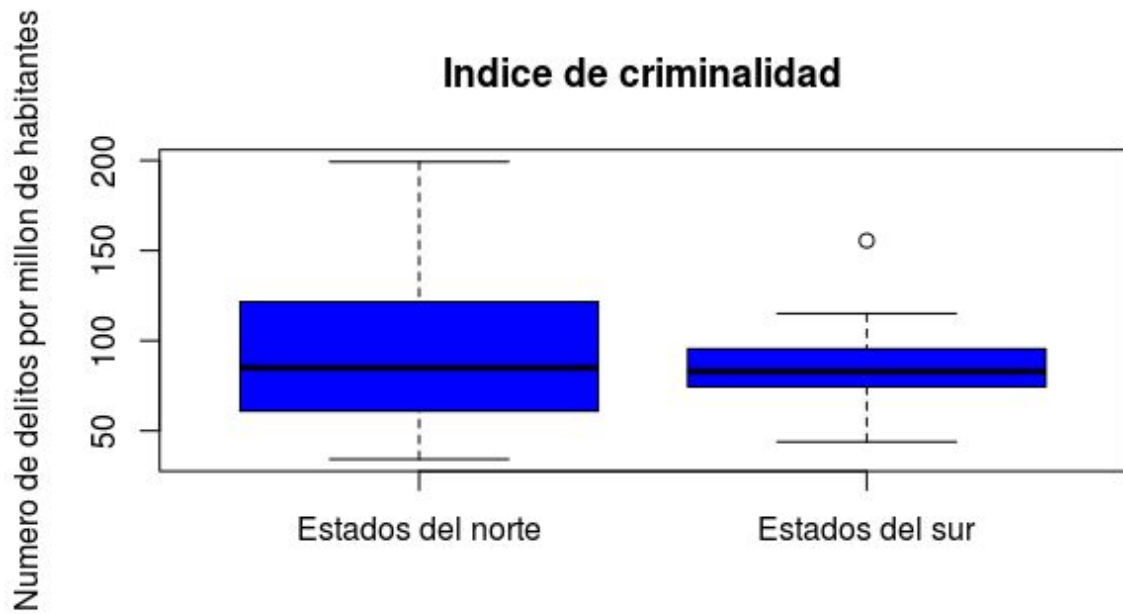


```
> hist(X[S==1], breaks = breakpoints, main = "Desigualdad de ingresos Sur", xlab =
"Número de familias por cada mil", ylab = "Frecuencia", col = c("blue","darkblue"))
```

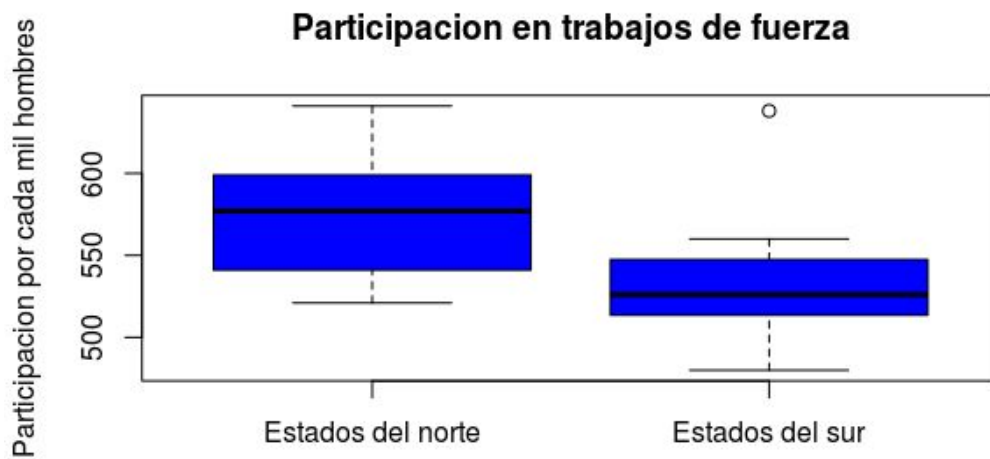


En ambos estados parecen haber concentraciones relativamente similares de datos, pero en los estados del sur las muestras están concentradas en los valores mas altos, lo cual demuestra que en los estados del sur hay mayor desigualdad de ingresos

```
> #Ahora mostrando los boxplots en una misma gráfica para facil comparacion
> par(mfrow = c(1,1))
> boxplot(R[S==0], R[S==1], main = "Indice de criminalidad", ylab = "Numero de delitos por
millon de habitantes", names = c("Estados del norte", "Estados del sur"), col = c("blue"))
```



```
> boxplot(LF[S==0], LF[S==1], main = "Participación en trabajos de fuerza", ylab =
"Participación por cada mil hombres", names = c("Estados del norte", "Estados del sur"), col
= c("blue"))
```



```
> boxplot(X[S==0], X[S==1], main = "Desigualdad de ingresos", ylab = "Número de familias
por cada mil", names = c("Estados del norte", "Estados del sur"), col = c("blue"))
```



## Desigualdad de ingresos

