

Tarea 12

Imanol

24/2/2021

Carga el conjunto de datos de la carpeta data llamado grades.txt en R y responde a las siguientes preguntas

```
data = read.table("../data/grades.txt")
head(data)
```

```
##      estudio      nota
## 1 Industrial 4.811598
## 2 Telematica 3.825744
## 3 Telematica 3.518318
## 4      Mates 0.000000
## 5 Telematica 5.879675
## 6 Informatica 3.554049
```

```
str(data)
```

```
## 'data.frame':    200 obs. of  2 variables:
## $ estudio: Factor w/ 4 levels "Industrial","Informatica",...: 1 4 4 3 4 2 3 2 4 4 ...
## $ nota   : num  4.81 3.83 3.52 0 5.88 ...
```

Preguntas de esta tarea

YO

1. Calcula la media, la mediana y la moda redondeada a dos cifras decimales de las notas numéricas de los exámenes tanto de forma conjunta como por estudio universitario. ¿En qué grupo observamos la nota más alta? ¿Qué grupo está por encima de la media de estudiantes?

```
round(mean(data$nota),2)
```

```
## [1] 3.87
```

```
round(median(data$nota),2)
```

```
## [1] 3.5
```

```
round(as.numeric(names(which(table(data$nota)==max(table(data$nota))))),2)
```

```
## [1] 0
```

```
round(by(data$nota, data$estudio, FUN = mean),2)
```

```
## data$estudio: Industrial
```

```
## [1] 3.58
```

```
## -----
```

```
## data$estudio: Informatica
```

```
## [1] 3.95
```

```
## -----
```

```
## data$estudio: Mates
```

```
## [1] 4.22
```

```
## -----
```

```
## data$estudio: Telematica
```

```
## [1] 3.7
```

```
round(by(data$nota, data$estudio, FUN = median),2)
```

```
## data$estudio: Industrial
```

```
## [1] 3.44
```

```
## -----
```

```
## data$estudio: Informatica
```

```
## [1] 3.48
```

```
## -----
```

```
## data$estudio: Mates
```

```
## [1] 4.1
```

```
## -----
```

```
## data$estudio: Telematica
```

```
## [1] 3.41
```

```
# Mates tiene la nota más alta
```

```
# Informática y Mates estan por encima de la media de estudiantes
```

2. ¿En qué grupo observamos más variación de notas? Justifica tu respuesta

```
by(data$nota, data$estudio, FUN = var)
```

```
## data$estudio: Industrial
```

```
## [1] 4.043913
```

```
## -----
```

```
## data$estudio: Informatica
```

```
## [1] 2.365377
```

```
## -----
```

```
## data$estudio: Mates
```

```
## [1] 4.819062
```

```
## -----
```

```
## data$estudio: Telematica
```

```
## [1] 3.646977
```

```
by(data$nota, data$estudio, FUN = sd)
```

```
## data$estudio: Industrial
## [1] 2.010948
## -----
## data$estudio: Informatica
## [1] 1.537978
## -----
## data$estudio: Mates
## [1] 2.195236
## -----
## data$estudio: Telematica
## [1] 1.909706
```

```
by(data$nota, data$estudio, FUN = IQR)
```

```
## data$estudio: Industrial
## [1] 2.152302
## -----
## data$estudio: Informatica
## [1] 2.030849
## -----
## data$estudio: Mates
## [1] 3.398879
## -----
## data$estudio: Telematica
## [1] 2.373383
```

El grupo con más variación de las notas es el de Mates ya que tiene una una varianza mayor y por lo t

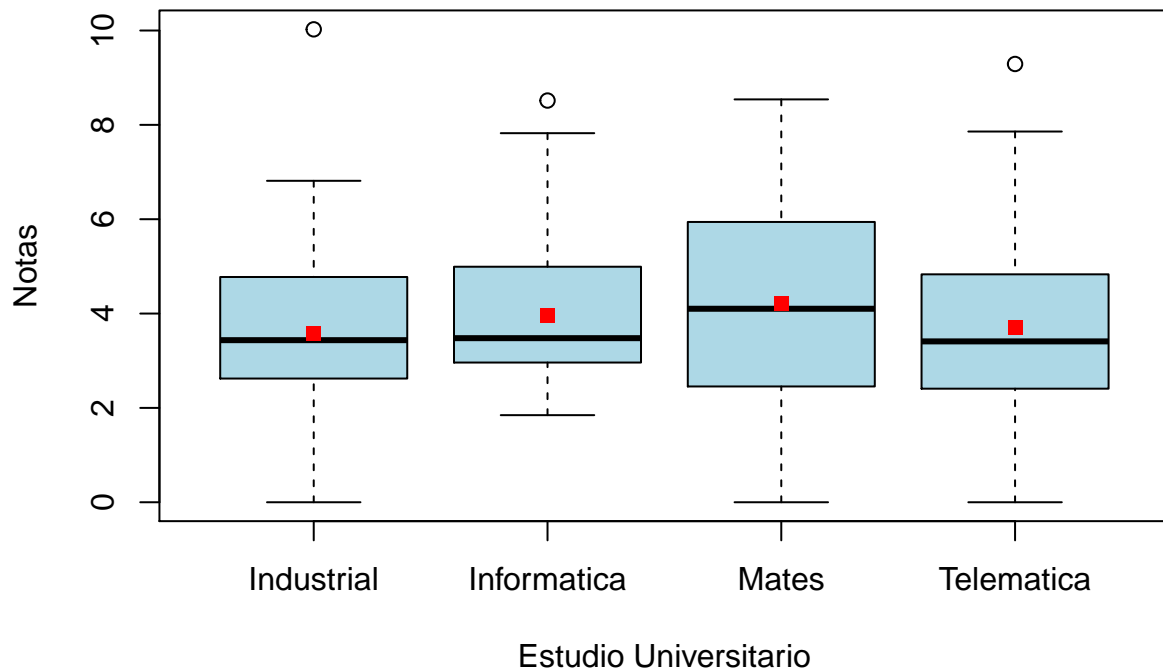
3. ¿Hay mucha diferencia entre el grupo que has respondido en el ejercicio 1 y en el ejercicio 2? Intenta dar una explicación objetiva del suceso

No hay mucha diferencia con el grupo de Mates pero en el grupo de Informatica escogido en la pregunta 1 si hay diferencia. Se puede deducir de los datos que Mates tiene una media ligeramente superior a Informática pero Informática varia mucho menos en las notas.

4. Dibuja en un único gráfico los cuatro diagramas de caja, uno para cada grupo. Añade así mismo la nota media, pon nombre, título y color al gráfico para documentar el gráfico correctamente.

```
boxplot(nota~estudio, data = data, col = "lightblue",
        main = "Diagrama de Cajas",
        xlab = "Estudio Universitario", ylab = "Notas")
medias = aggregate(nota~estudio, data = data, FUN = mean)
points(medias, col = "red", pch = 15)
```

Diagrama de Cajas

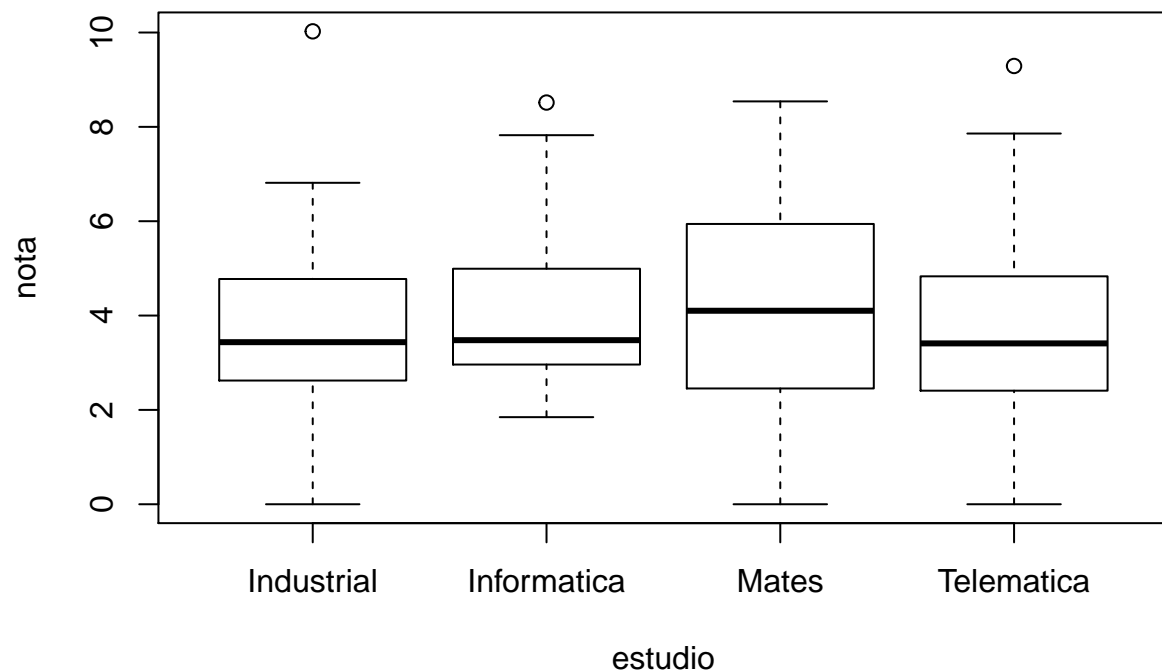


5. ¿Observas algunos valores atípicos en el boxplot anterior? ¿A qué grupo pertenece?

```
boxplot(nota~estudio, data = data)$out
```

```
## [1] 10.024687 8.515653 9.290465
```

```
boxplot(nota~estudio, data = data)$group
```



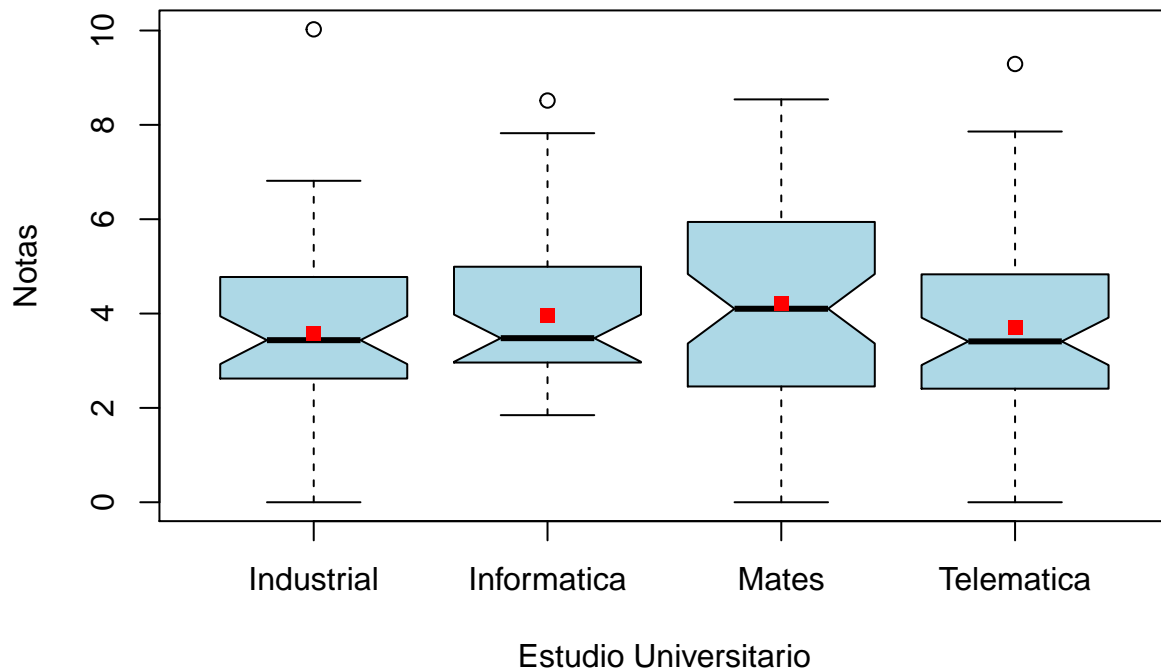
```
## [1] 1 2 4
```

Se observa un valor atipico de 10.02 para Industrial, de 8.52 para Informatica y de 9.29 para Telemat

6. Indica a partir del boxplot anterior en qué grupo observamos más variedad de notas.

```
boxplot(nota~estudio, data = data, col = "lightblue",
        main = "Diagrama de Cajas",
        xlab = "Estudio Universitario", ylab = "Notas", notch = TRUE)
medias = aggregate(nota~estudio, data = data, FUN = mean)
points(medias, col = "red", pch = 15)
```

Diagrama de Cajas



Se observa mas variedad de notas en Mates ya que su caja es de mayor tamaño que las demas y los bigot

7. En base a todo lo anterior, ¿puedes afirmar de forma estadística que el examen de carrera ha ido mejor en algún estudio que en los demás?

Yo diría que le ha ido mejor a Informatica ya que tiene la segunda mejor media con la variacion de datos más pequeña. Además en el diagrama de cajas se puede ver que el bigote inferior es el mas corto.

OTROS 1

1. Calcula la media, la mediana y la moda redondeada a dos cifras decimales de las notas numéricas de los exámenes tanto de forma conjunta como por estudio universitario. ¿En qué grupo observamos la nota más alta? ¿Qué grupo está por encima de la media de estudiantes?

```
round(summary(data[,2]),2) # media, mediana forma conjunta
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.00   2.60   3.50   3.87   5.12   10.02
```

```
as.numeric(names(which(table(data$nota) == max(table(data$nota))))) # moda de forma global
```

```
## [1] 0
```

```
by(data$nota, data$estudio, FUN = summary) # agrupados por carrera universitaria
```

```
## data$estudio: Industrial
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000  2.622   3.435   3.577   4.775   10.025
## -----
## data$estudio: Informatica
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.846  2.962   3.478   3.954   4.992   8.516
## -----
## data$estudio: Mates
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000  2.534   4.101   4.215   5.932   8.540
## -----
## data$estudio: Telematica
##   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000  2.443   3.410   3.704   4.816   9.290
```

La nota mas alta se encuentra en la carrera de Industrial (10.025).. aunque es un valor extraño si es

Los estudiantes de Matemáticas tienen una media superior (4.215) sobre los otros grupos.

2. ¿En qué grupo observamos más variación de notas? Justifica tu respuesta

```
aggregate(nota~estudio, data = data, FUN = sd)
```

```
##      estudio      nota
## 1 Industrial 2.010948
## 2 Informatica 1.537978
## 3      Mates 2.195236
## 4 Telematica 1.909706
```

```
aggregate(nota~estudio, data = data, FUN = IQR) # Rango intercuartilico
```

```
##      estudio      nota
## 1 Industrial 2.152302
## 2 Informatica 2.030849
## 3      Mates 3.398879
## 4 Telematica 2.373383
```

El grupo con más variación de las notas es el de Mates ya que tiene una varianza mayor y por lo tanto

3. ¿Hay mucha diferencia entre el grupo que has respondido en el ejercicio 1 y en el ejercicio 2? Intenta dar una explicación objetiva del suceso

```
table(data$estudio)
```

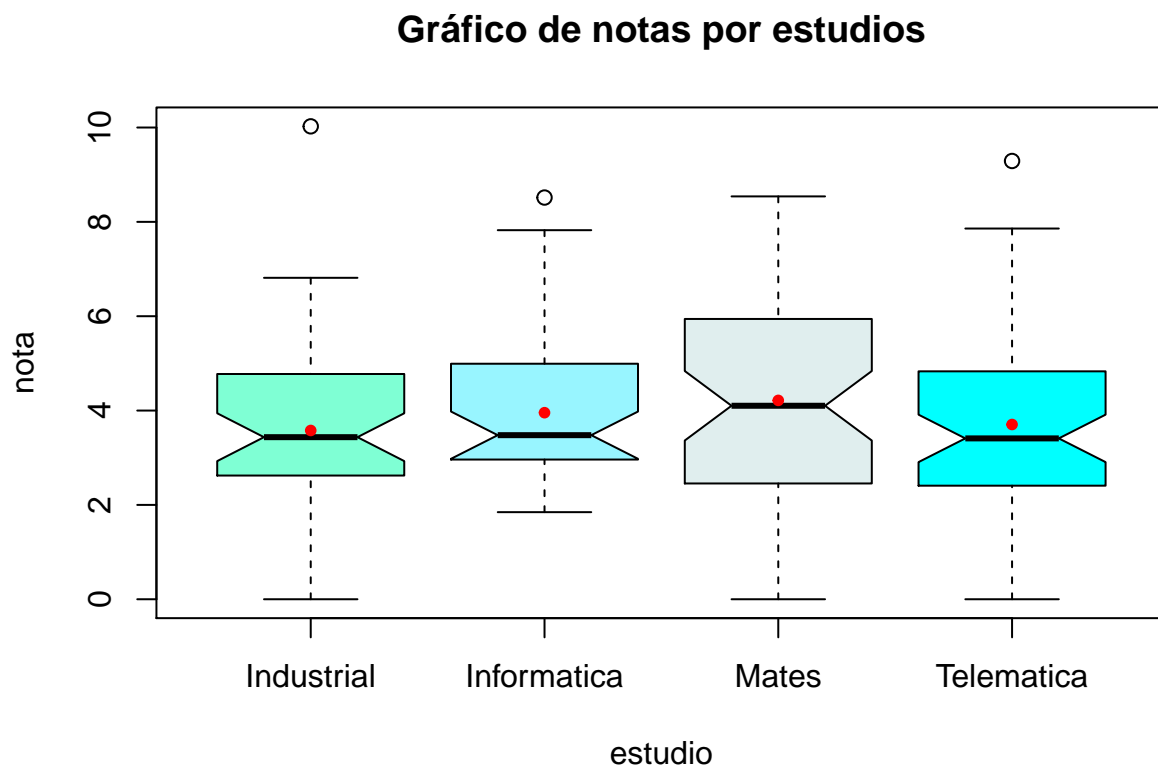
```
##
##  Industrial Informatica      Mates Telematica
##           45           41           56           58
```

De los ejercicios anteriores observamos que la media mayor está en la carrera de Matemáticas pero a su vez tiene la mayor desviación.. esto puede ser causado por tener un mayor número de observaciones.

Si buscamos la carrera con menor desviación y datos mas específicos, nos iríamos Industriales, que a su vez tiene un IQR menor.

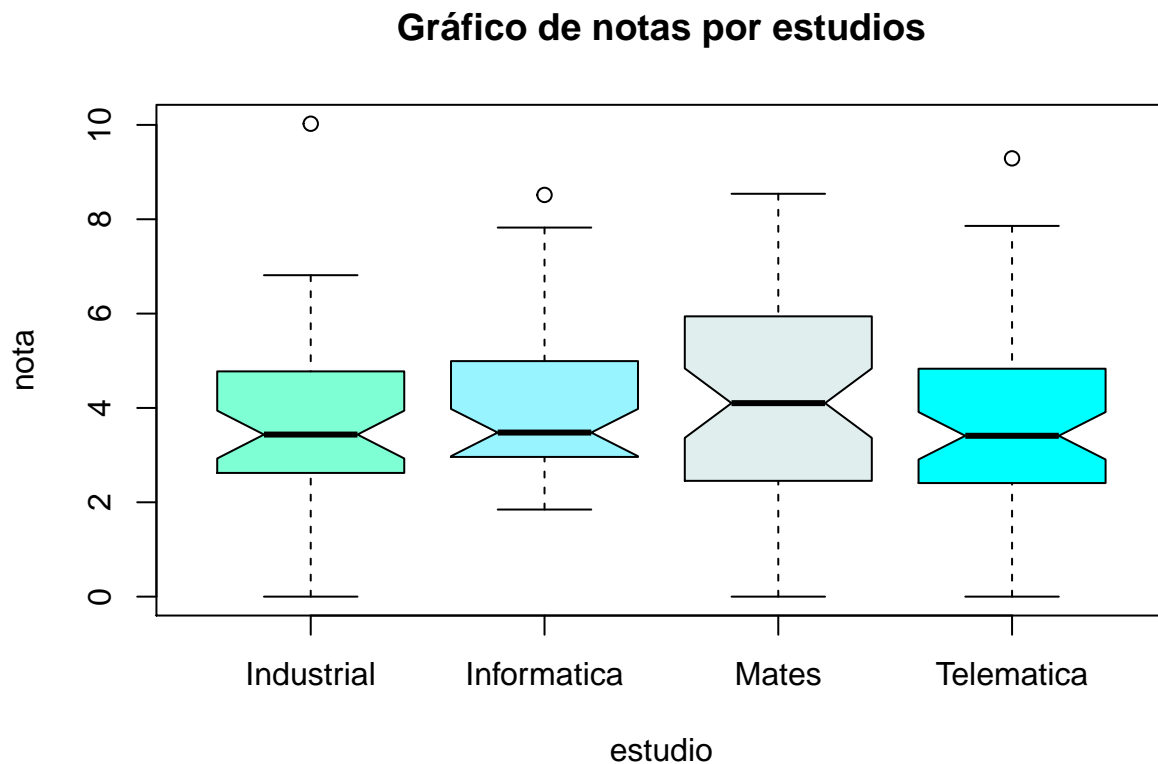
4. Dibuja en un único gráfico los cuatro diagramas de caja, uno para cada grupo. Añade así mismo la nota media, pon nombre, título y color al gráfico para documentar el gráfico correctamente.

```
boxplot(nota~estudio, data = data, main = "Gráfico de notas por estudios", col =c("aquamarine","cadetblue",
      notch = TRUE)
media = aggregate(nota~estudio, data = data, FUN = mean)
points(media, col = "red", pch = 20)
```



5. ¿Observas algunos valores atípicos en el boxplot anterior? ¿A qué grupo pertenece?

```
str(boxplot(nota~estudio, data = data, main = "Gráfico de notas por estudios", col =c("aquamarine","cad",
notch = TRUE))
```



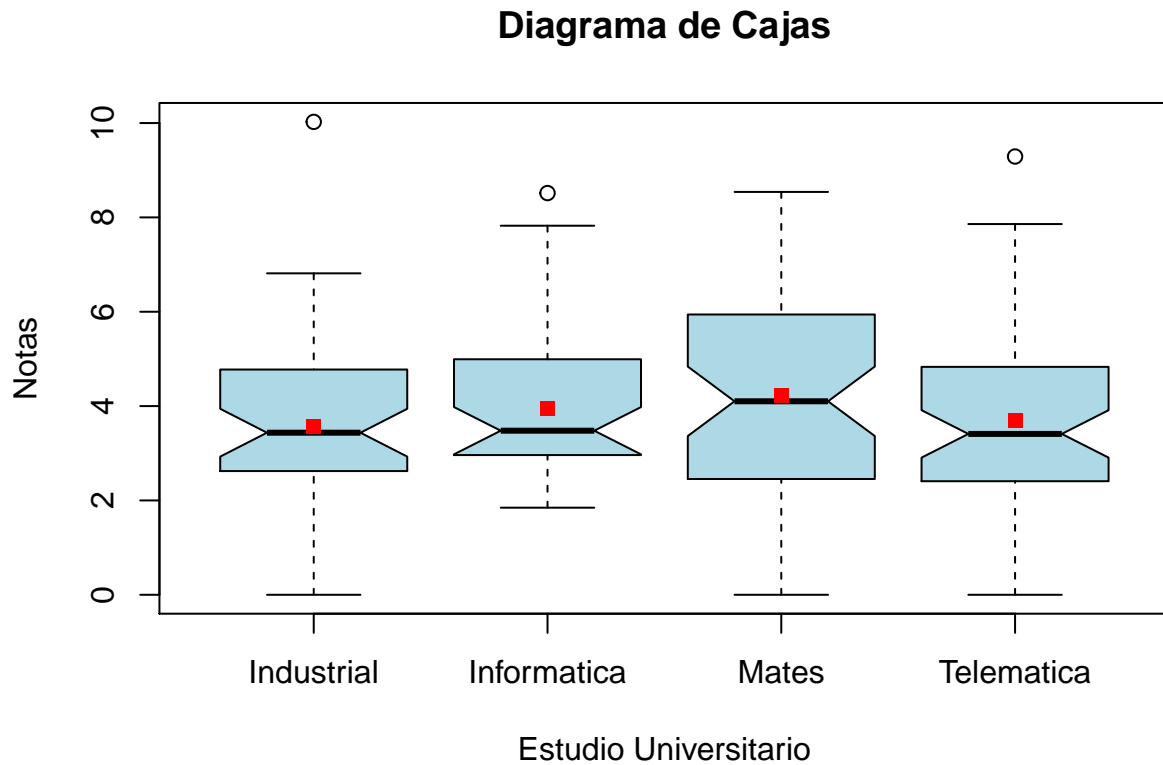
```
## List of 6
## $ stats: num [1:5, 1:4] 0 2.62 3.44 4.77 6.81 ...
## $ n : num [1:4] 45 41 56 58
## $ conf : num [1:2, 1:4] 2.93 3.94 2.98 3.98 3.37 ...
## $ out : num [1:3] 10.02 8.52 9.29
## $ group: num [1:3] 1 2 4
## $ names: chr [1:4] "Industrial" "Informatica" "Mates" "Telematica"
```

Se observan tres valores atípicos (10.02 8.52 9.29) en los grupos de Industrial, Informática y Telemática

6. Indica a partir del boxplot anterior en qué grupo observamos más variedad de notas.

```
boxplot(nota~estudio, data = data, col = "lightblue",
main = "Diagrama de Cajas",
```

```
xlab = "Estudio Universitario", ylab = "Notas", notch = TRUE)
medias = aggregate(nota~estudio, data = data, FUN = mean)
points(medias, col = "red", pch = 15)
```



Se observa mas variedad de notas en Mates ya que su caja es de mayor tamaño que las demas y los bigots

7. En base a todo lo anterior, ¿puedes afirmar de forma estadística que el examen de carrera ha ido mejor en algún estudio que en los demás?

Con el gráfico vemos que los estudiantes de Informática y Telemática han obtenido un mayor número de notas por encima de la media.

Y en mayor detalle la carrera de Informática presenta una desviación menor.