

Práctica7

Antes de empezar

Para realizar el setup correspondiente a esta práctica agregaremos multiples librerías que necesitaremos por una parte ggplot2 para realizar las gráficas, por otra parte knitr que nos agregara comandos tan útiles como kable, y por último tidyr que nos permitirá distribuir las tablas de datos y obtener estadísticas mas complejas.

```
setwd("~/Escritorio/R/Datasets")

educacion <- read.table("Datos_hsb.txt", sep = ",", header = TRUE, fileEncoding = "latin1")

library(ggplot2)
library(knitr)
library(tidyr)
```

a) Calcular medias de progreso por las distintas categorías razonar posibles conclusiones

Para calcular las diferentes medias haremos uso de las ya conocidas funciones mean, y aggregate de esa manera podremos crear una tabla con todas ellas de manera sencilla y visual.

```
p <- aggregate(cbind(P_Lectura, P_Escritura, P_Matematicas, P_Sociales, P_Ciencias)~Estatus_Socioeconomico, data = educacion, FUN = mean)
kable(p)
```

Estatus_Socioeconomico	P_Lectura	P_Escritura	P_Matematicas	P_Sociales	P_Ciencias
alto	56.50000	55.91379	56.17241	57.13793	55.44828
bajo	48.27660	50.61702	49.17021	47.31915	47.70213
medio	51.57895	51.92632	52.21053	52.03158	51.70526

```
s <- aggregate(cbind(P_Lectura, P_Escritura, P_Matematicas, P_Sociales, P_Ciencias)~Tipo_Programa, data = educacion, FUN = mean)
kable(s)
```

Tipo_Programa	P_Lectura	P_Escritura	P_Matematicas	P_Sociales	P_Ciencias
Academic preparatory	56.16190	56.25714	56.73333	56.69524	53.80000
General	49.75556	51.33333	50.02222	50.60000	52.44444
Vocational/technical	46.20000	46.76000	46.42000	45.02000	47.22000

```
p <- aggregate(cbind(P_Lectura, P_Escritura, P_Matematicas, P_Sociales, P_Ciencias)~Estatus_Socioeconomico+Raza+Sexo, data = educacion, FUN = mean)
kable(p)
```

Estatus_Socioeconomico	Raza	Sexo	P_Lectura	P_Escritura	P_Matematicas	P_Sociales	P_Ciencias
bajo	Asiática	Hombre	52.00000	44.00000	49.00000	41.00000	55.00000
medio	Asiática	Hombre	52.50000	61.50000	63.50000	51.00000	52.00000
alto	Blanca	Hombre	56.76000	54.44000	55.88000	58.80000	56.36000

Estatus_Socioeconomico	Raza	Sexo	P_Lectura	P_Escritura	P_Matematica	P_Sociales	P_Ciencias
bajo	Blanca	Hombre	49.33333	46.77778	48.55556	43.22222	51.88889
medio	Blanca	Hombre	54.23529	50.17647	54.44118	52.17647	55.58824
alto	Hispana	Hombre	48.66667	44.00000	51.66667	53.00000	43.00000
bajo	Hispana	Hombre	52.33333	45.33333	45.00000	41.00000	48.00000
medio	Hispana	Hombre	44.57143	44.14286	50.00000	44.57143	45.57143
alto	Negra	Hombre	50.00000	40.00000	39.00000	47.00000	49.00000
bajo	Negra	Hombre	43.50000	49.00000	46.50000	48.50000	40.50000
medio	Negra	Hombre	47.75000	47.75000	46.25000	49.75000	49.25000
alto	Asiática	Mujer	50.33333	61.66667	56.66667	52.66667	53.66667
bajo	Asiática	Mujer	61.00000	65.00000	68.00000	63.50000	56.00000
medio	Asiática	Mujer	47.00000	52.00000	49.33333	44.33333	44.66667
alto	Blanca	Mujer	57.82609	59.17391	57.86957	56.91304	57.43478
bajo	Blanca	Mujer	50.20000	56.00000	52.26667	49.33333	51.86667
medio	Blanca	Mujer	52.02564	55.10256	51.94872	53.89744	51.12821
alto	Hispana	Mujer	73.00000	61.00000	57.00000	66.00000	55.00000
bajo	Hispana	Mujer	42.00000	46.50000	45.33333	47.16667	42.66667
medio	Hispana	Mujer	45.00000	49.50000	42.25000	51.00000	46.50000
alto	Negra	Mujer	54.00000	51.50000	54.50000	52.50000	46.00000
bajo	Negra	Mujer	44.66667	47.88889	45.00000	47.11111	38.77778
medio	Negra	Mujer	49.00000	50.50000	52.00000	58.50000	44.00000

A simple vista con las medias ya podemos empezar a establecer varias relaciones entre los datos. Primero observamos que claramente el estatus económico de el alumno influye de manera drástica en las puntuaciones, siendo aquellos con estatus alto los que liderarán la lista seguidos por los de medio y por último los de estatus bajo.

Respecto al tipo de programa pasa algo bastante parecido aquellos que estan en preparatoria académica ostentan mas puntuación en todos los campos que los que estan en un programa general seguidos de el Vocacional/técnico.

No mencionaré los casos de sexo y raza ya que en el apartado D realizaremos un estudio mas exhaustivo de estos casos haciendo referencia a las medias citadas arriba.

b) Evaluar la influencia del centro en los resultados de aprendizaje de la época.

Para ello usaremos las herramientas brindadas por ggplot2 tanto el boxplot, como el barplot. A su vez también usaremos la librería tidyr para realizar gráficos mas detallados con todas las asignaturas.

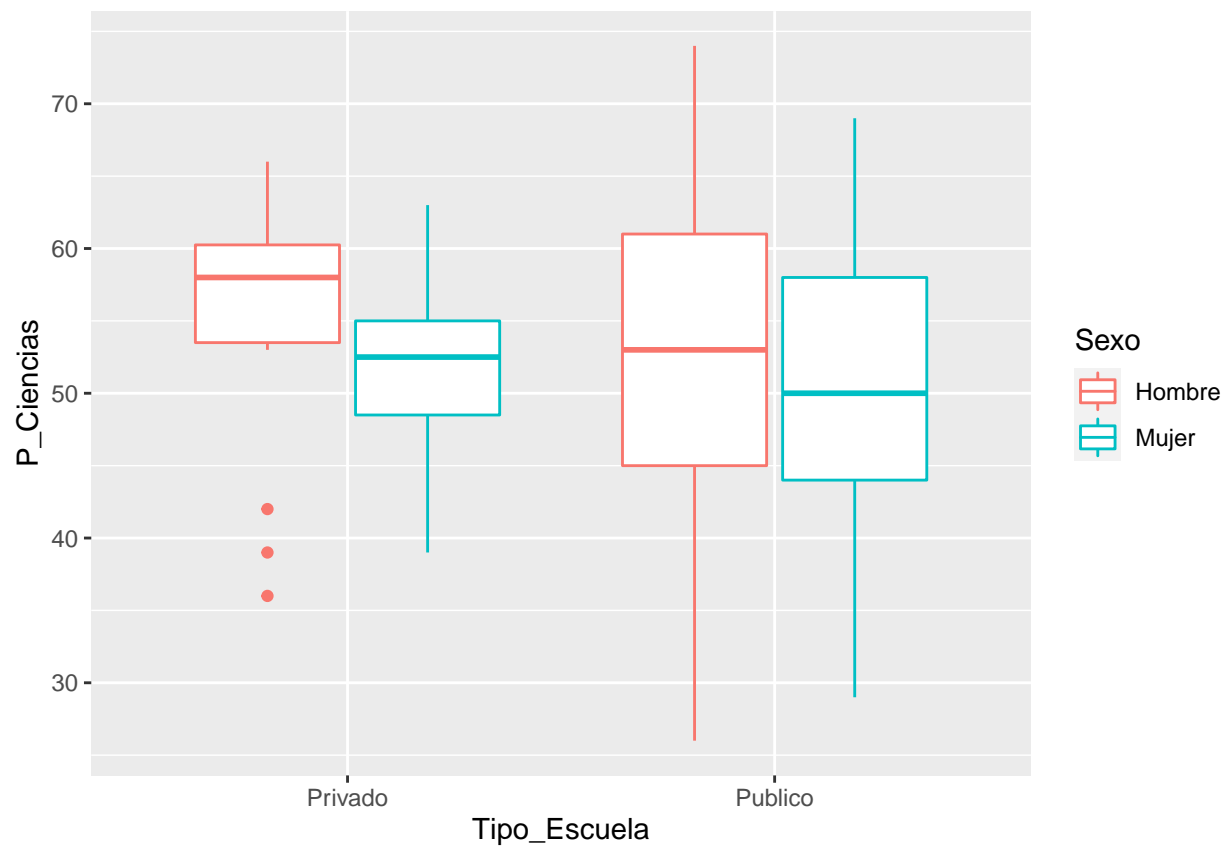
```
educacion_escuela <- aggregate(cbind(P_Lectura, P_Escritura, P_Matematicas, P_Sociales, P_Ciencias)~Tipo_Escuela,
                                data=educacion, FUN=mean)
library(ggplot2)
library(tidyr)
library(dplyr)
library(purrr)
```

Tipo_Escuela	P_Lectura	P_Escritura	P_Matematicas	P_Sociales	P_Ciencias
Privado	54.25000	55.53125	54.75000	54.78125	53.31250
Publico	51.84524	52.25000	52.24405	51.95238	51.57143

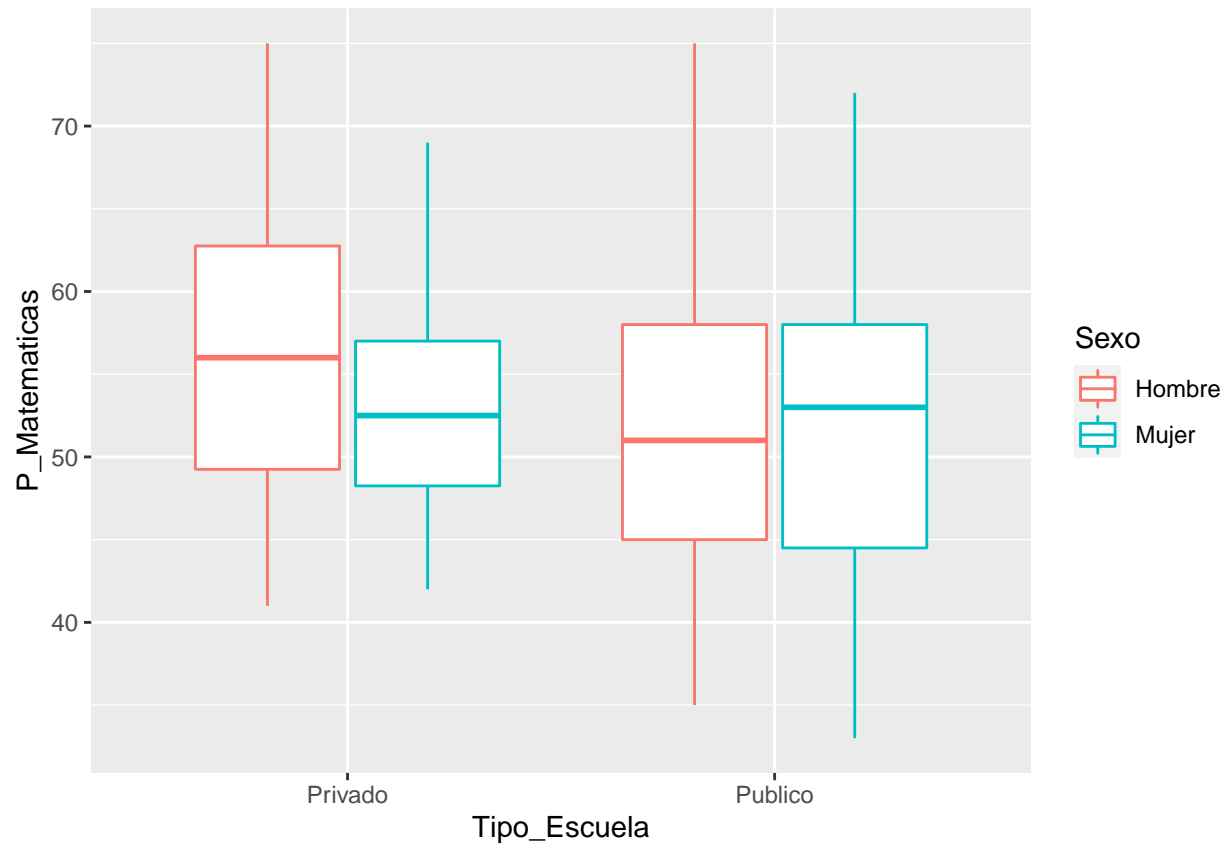
```
centros <- unique(educacion$Tipo_Escuela); centros
```

```
## [1] "Publico" "Privado"
```

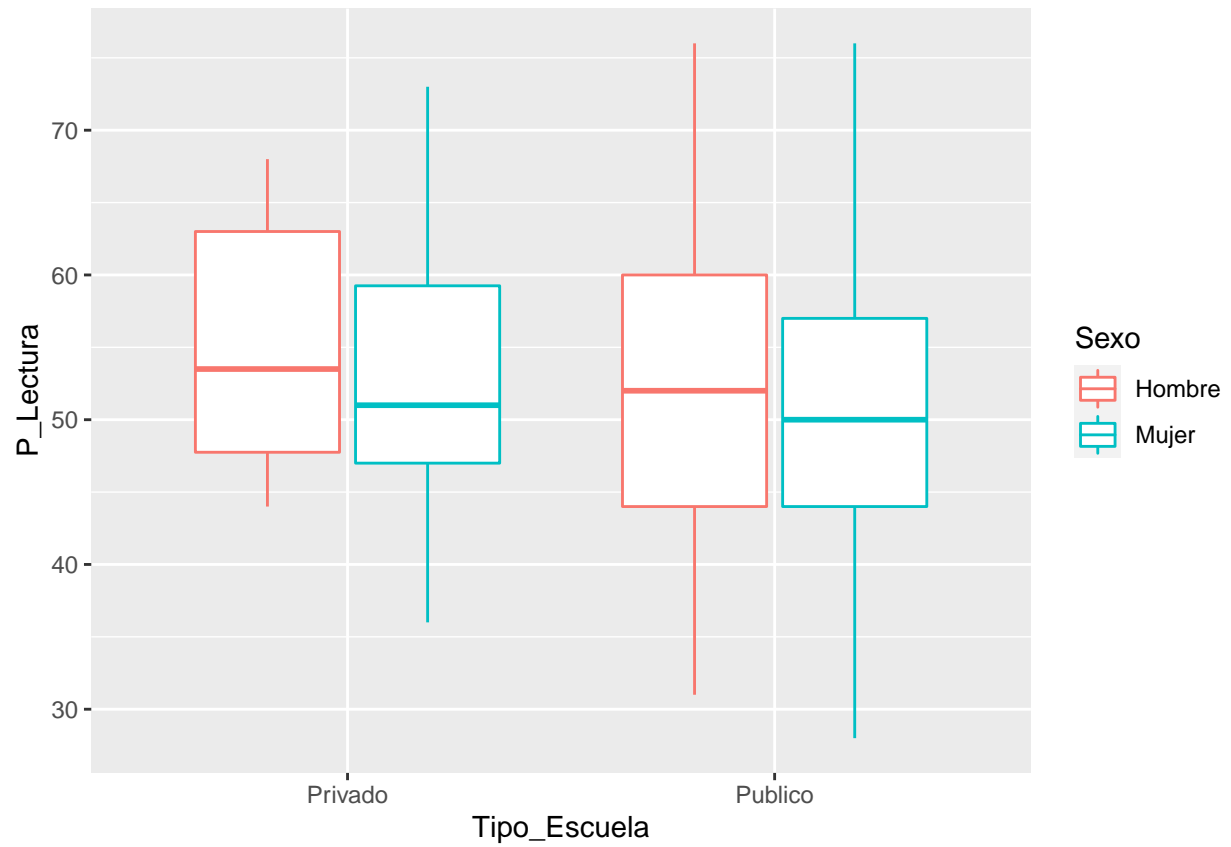
```
ggplot(data = educacion, mapping = aes(x = Tipo_Escuela, y = P_Ciencias)) + geom_boxplot() + aes(color = Sexo)
```



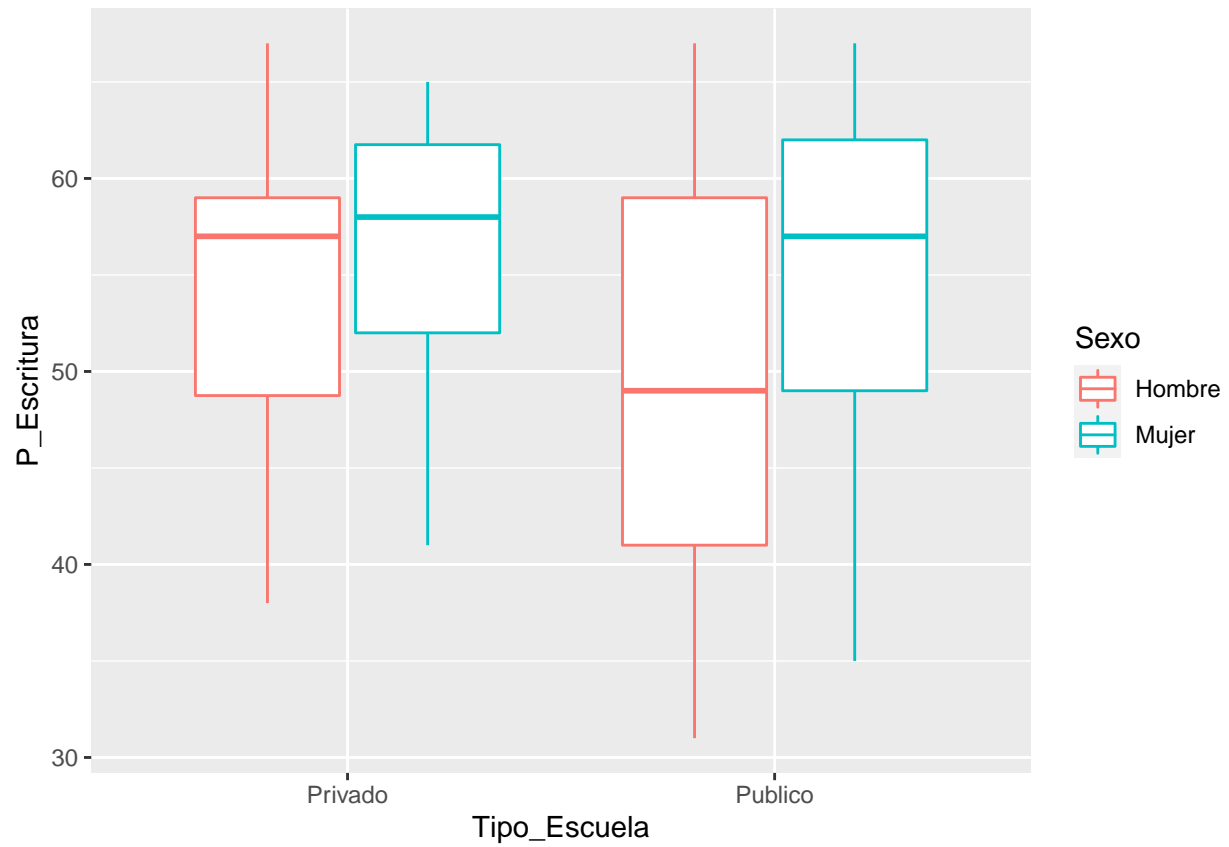
```
ggplot(data = educacion, mapping = aes(x = Tipo_Escuela, y = P_Matematicas)) + geom_boxplot() + aes(color = Sexo)
```



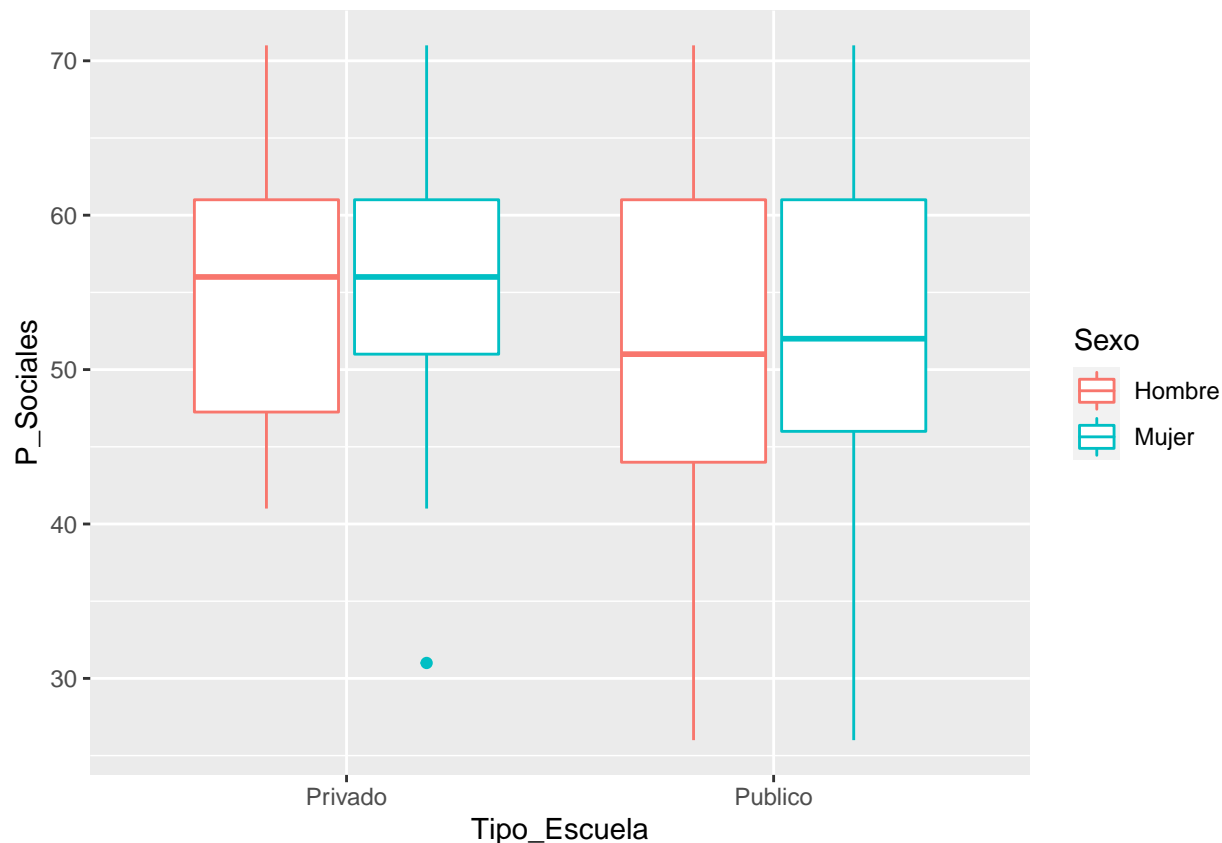
```
ggplot(data = educacion, mapping = aes(x = Tipo_Escuela, y = P_Lectura)) + geom_boxplot() + aes(colour = Sexo)
```



```
ggplot(data = educacion, mapping = aes(x = Tipo_Escuela, y = P_Escritura)) + geom_boxplot() + aes(col
```



```
ggplot(data = educacion, mapping = aes(x = Tipo_Escuela, y = P_Sociales)) + geom_boxplot() + aes(color = Sexo)
```



Ya observando las medias nos damos cuenta que en los centros privados las puntuaciones obtenidas son bastante mayores y si nos fijamos en los boxplot mostrados (donde no solo se muestra la diferencia de centros sino también de sexos) nos podemos fijar que no solo las medias son mejores sino que también existe mucha menos dispersión de los datos, por lo tanto las notas en las escuelas privadas la mayoría serán positivas al contrario que en la pública que hay gran diversidad.

Comencemos analizando las puntuaciones en ciencias: Primero podemos observar que las mujeres ostentan notas inferiores a los hombres tanto en la educación pública y privada. A su vez también observamos lo comentado anteriormente las notas en los centros privados son mas concentradas siendo la mayoría altas y en el público habiendo gran variedad.

En matemáticas en el privado es prácticamente igual a las ciencias los hombres tienen mejores puntuaciones que las mujeres y su dispersión es inferior. Sin embargo en el público es al contrario las mujeres poseen mejores notas que los hombres en promedio.

En lectura la gráfica nos muestra información muy parecida que en ciencias, hombres mejor puntuación media y centro privado teniendo gran ventaja respecto al público.

Sin embargo tanto en escritura y sociales parece que las diferencias entre sexos disminuyen llegando de nuevo en la escuela pública superar las mujeres a los hombres y en las privadas igualarlos, aunque sigue siendo notoria la diferencia entre los centros privados que se mantienen mas arriba de media y con menos dispersión y los públicos. En escritura pasa algo similar lo que esta vez las mujeres mantienen mejores calificaciones de media siendo no tan notoria la diferencia respecto a los hombres en la escuela privada sin embargo en la pública teniendo una diferencia considerable.

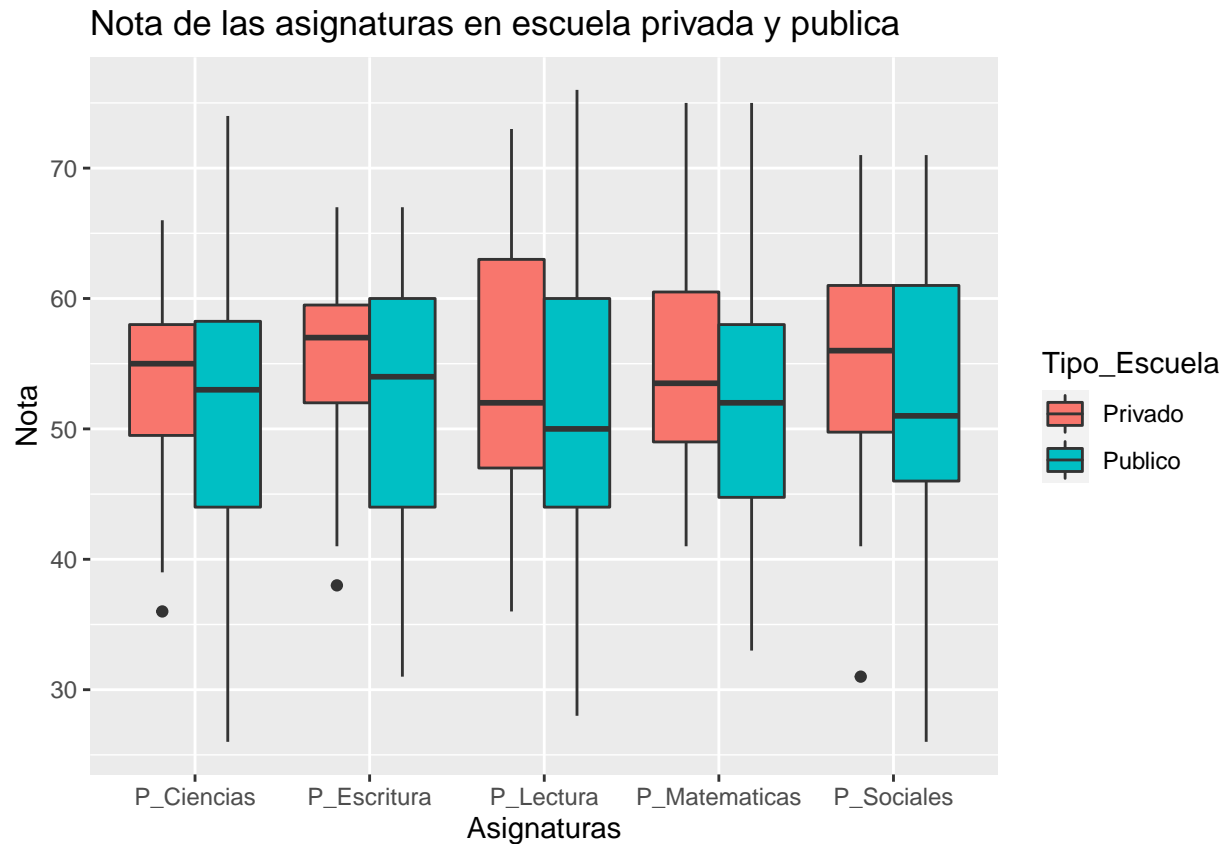
Dado lo analizado anteriormente podemos llegar a dos conclusiones: la primera que en los centros privados las notas son mejores que los públicos en promedio y ostentan menor dispersión entre las notas de los alumnos, y segundo que las mujeres tienen mejores calificaciones en escritura y asignaturas sociales, y a su vez los hombres le sacan ventaja en matemáticas, ciencias y lectura.

En la parte inferior podemos encontrar gráficos mas generales que muestran también lo citado anteriormente.

```
educacion2 <- educacion %>%
  gather(Campo, Value, P_Ciencias, P_Escritura, P_Lectura, P_Matematicas, P_Sociales, -Tipo_Escuela)

stat1 <- ggplot(educacion2, aes(x = Campo, y = Value, fill = Tipo_Escuela)) + geom_boxplot(position = "dodge")

stat1 + labs(title = "Nota de las asignaturas en escuela privada y publica", x = "Asignaturas", y = "Nota")
```

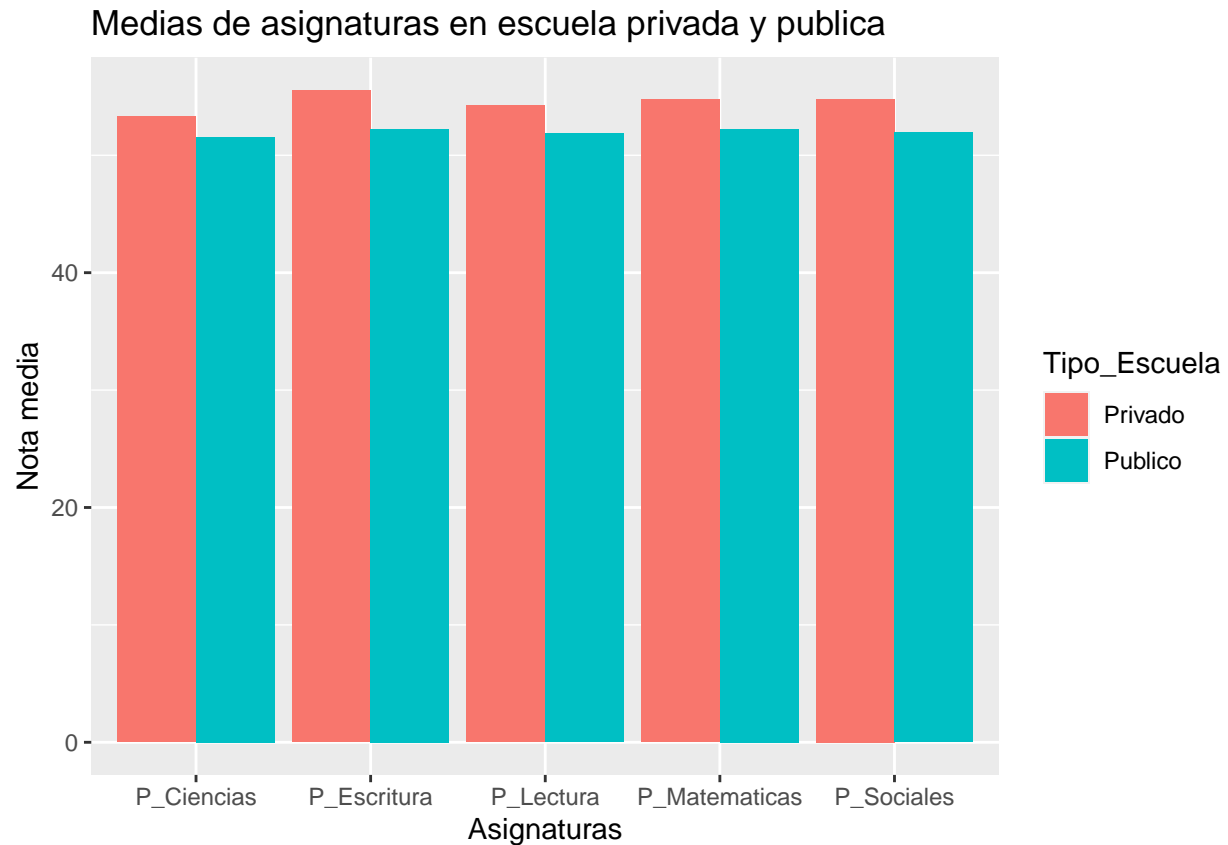


```
#using plot$out for outliers #str analisis tipos de datos

educacion2 <- educacion_escuela %>%
  gather(Campo, Value, -Tipo_Escuela)

stat2 <- ggplot(educacion2, aes(x = Campo, y = Value, fill = Tipo_Escuela)) + geom_col(position = "dodge")

stat2 + labs(title = "Medias de asignaturas en escuela privada y publica", x = "Asignaturas", y = "Nota")
```

c) Estudiar posibles relaciones entre la capacidad de lectura y el aprendizaje de otras disciplinas

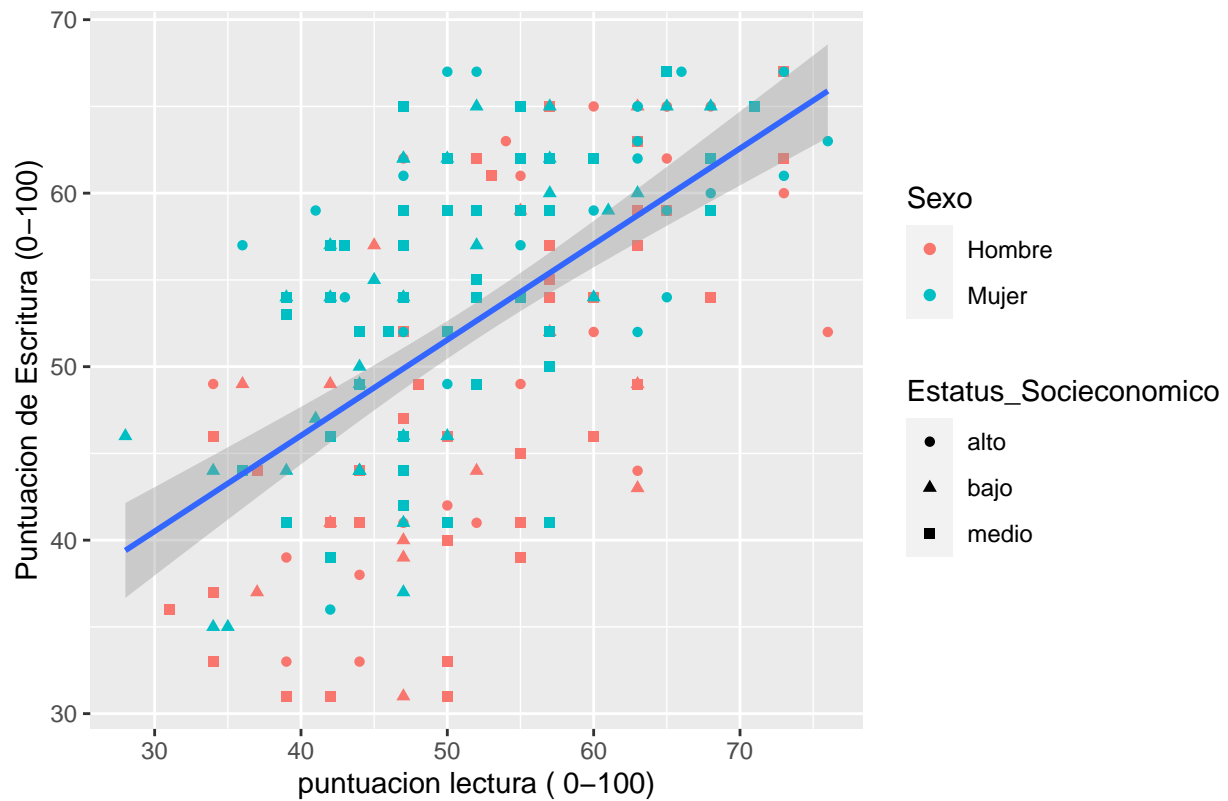
Para ello compararemos la puntuación de lectura y la puntuación en otras asignaturas para comprobar si existe algún tipo de relación lineal entre la lectura y otros campos.

```
g0 <- ggplot(data = educacion, aes(x = P_Lectura)) +
  xlab("puntuacion lectura ( 0-100)") +
  ggtitle("Relacion entre la capacidad de lectura y aprendizaje en otras materias")

g0+aes(y=P_Escritura)+
  geom_point(aes(shape=Estatus_Socieconomico, colour=Sexo))+
  geom_smooth(method="lm")+
  ylab("Puntuacion de Escritura (0-100)")

## `geom_smooth()` using formula 'y ~ x'
```

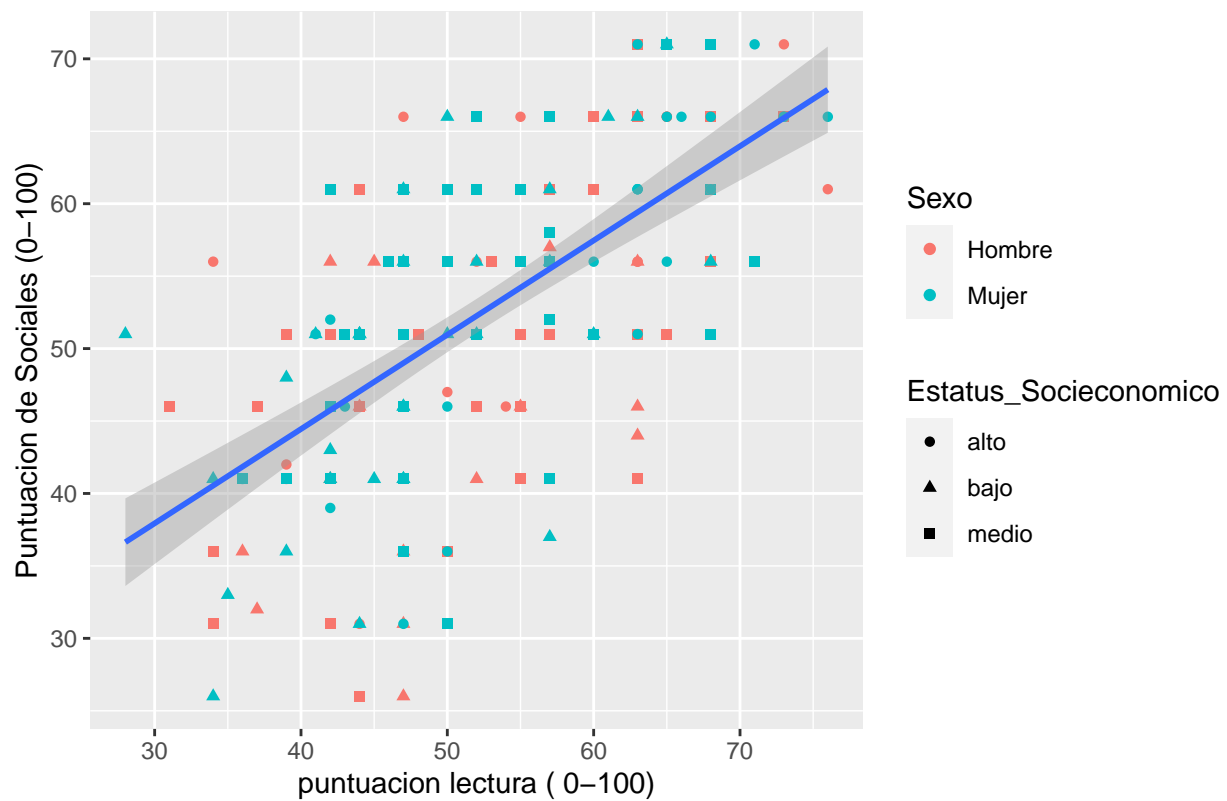
Relacion entre la capacidad de lectura y aprendizaje en otras materias



```
g0+aes(y=P_Sociales)+
geom_point(aes(shape=Estatus_Socioeconomico, colour=Sexo))+
geom_smooth(method="lm")+
ylab("Puntuacion de Sociales (0-100)")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

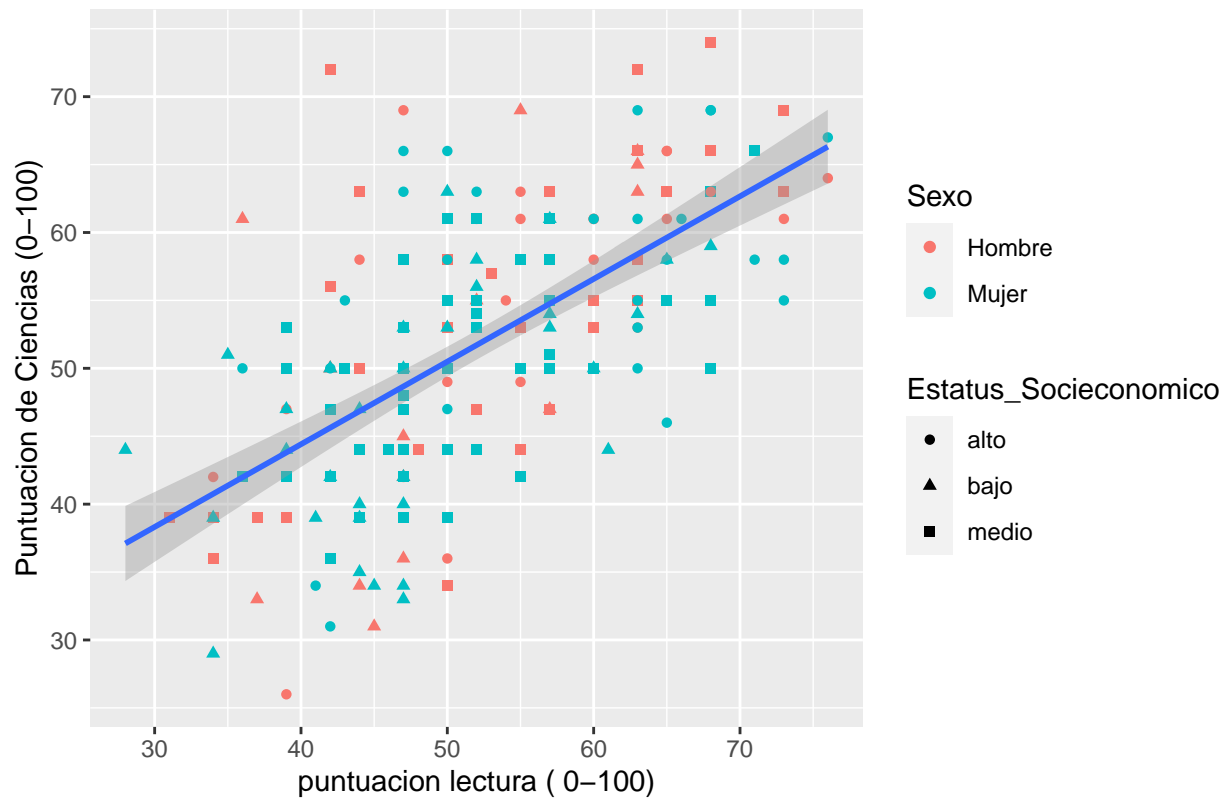
Relacion entre la capacidad de lectura y aprendizaje en otras materias



```
g0+aes(y=P_Ciencias)+
geom_point(aes(shape=Estatus_Socioeconomico, colour=Sexo))+
geom_smooth(method="lm")+
ylab("Puntuacion de Ciencias (0-100)")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

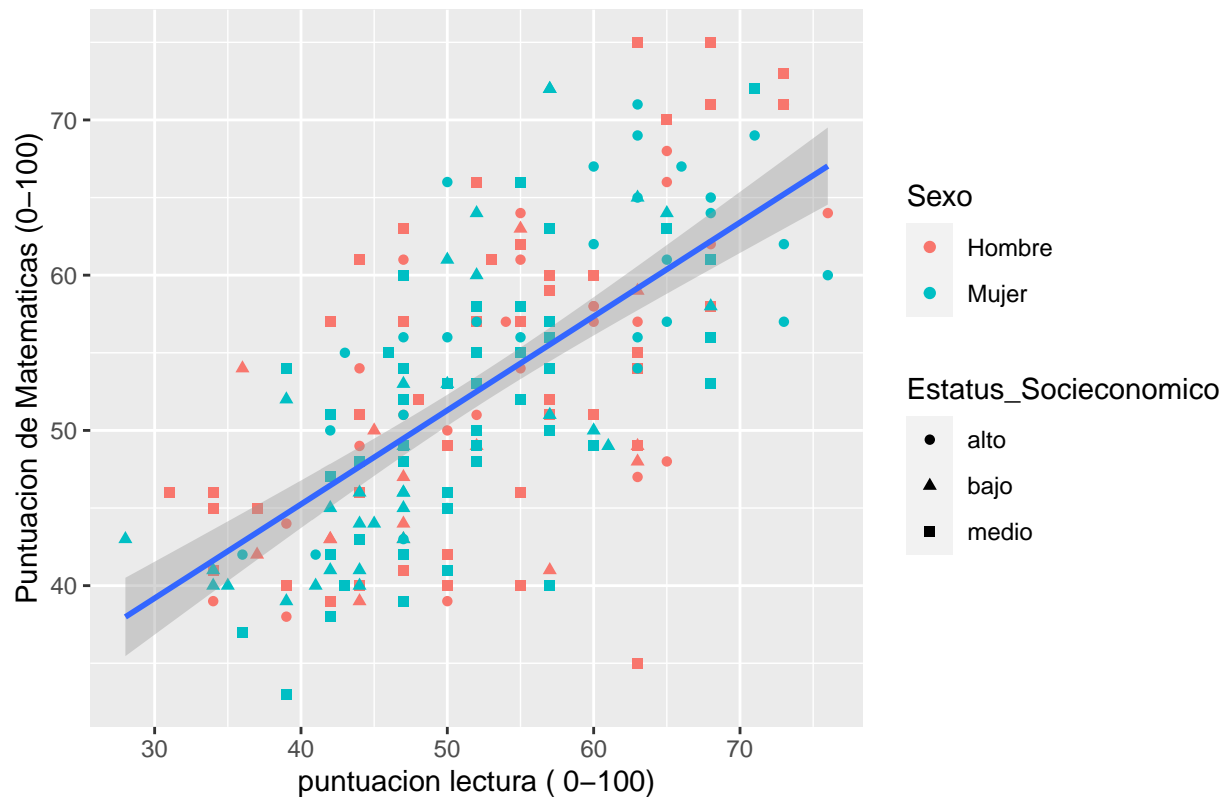
Relacion entre la capacidad de lectura y aprendizaje en otras materias



```
g0+aes(y=P_Matematicas)+
geom_point(aes(shape=Estatus_Socioeconomico, colour=Sexo))+
geom_smooth(method="lm")+
ylab("Puntuacion de Matematicas (0-100)")
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Relacion entre la capacidad de lectura y aprendizaje en otras materias



Los resultados con bastante reveladores, si bien esperaríamos que la lectura estuviera bastante relacionada con la escritura cosa que se ve en el gráfico que relaciona ambos, a su vez el resto de materias también tienen una fuerte relación incluidas las mas numéricas como matemáticas o ciencias.

Observamos esto en todos los sexos y estatus económicos mostrados en la leyenda. Podemos concluir que aquellos alumnos que tienen buena puntuación en lectura también la tendrán en el resto de asignaturas o al menos le ayudará a tenerla.

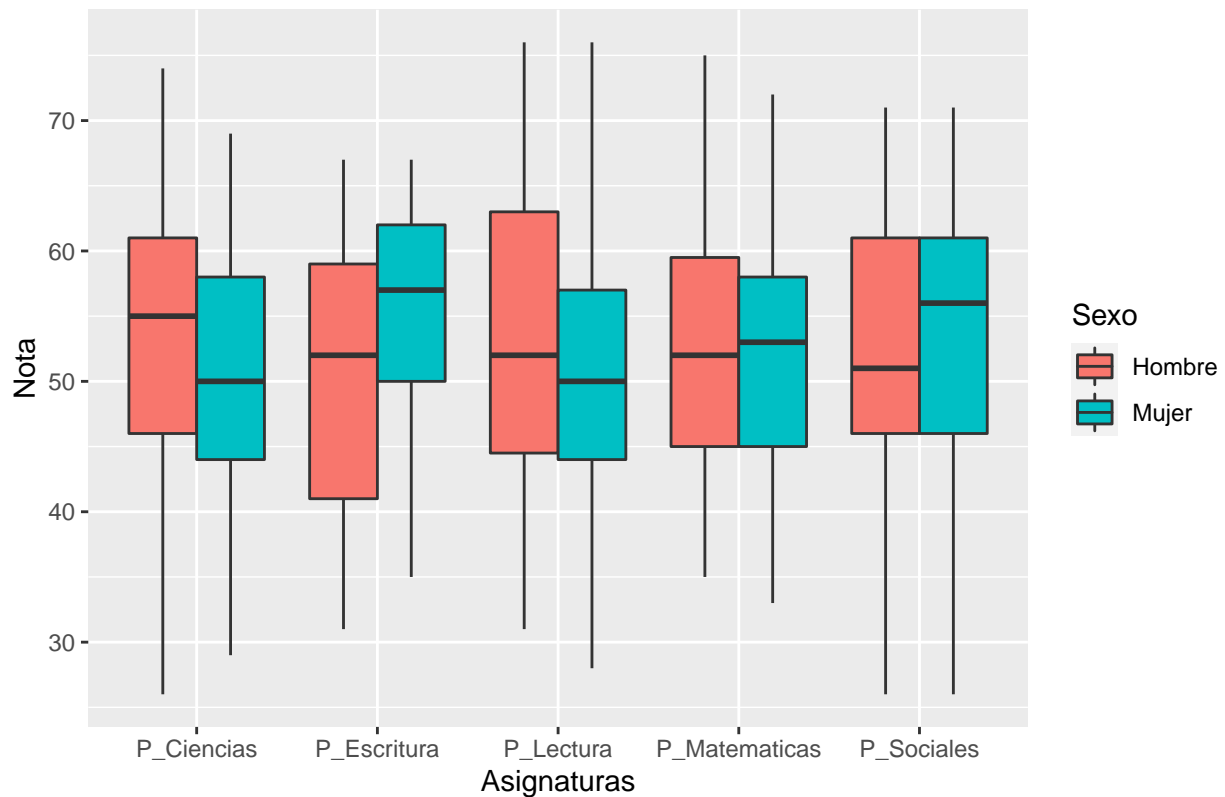
d) Visualizar si el sexo o la raza de origen en 1980 tenían influencias significativas en el nivel de aprendizaje

Para ello igual que en el apartado b realizaremos gráficas usando ggplot y tydir de forma mas general y visual usando tanto boxplot como barplots. A su vez también calcularemos tablas de medias usando aggregate. De esa forma podremos llegar a mejores conclusiones.

```
#sexo
educacion2 <- educacion %>%
gather(Campo, Value, P_Ciencias, P_Escritura, P_Lectura, P_Matematicas, P_Sociales, -Sexo)

stat1 <- ggplot(educacion2, aes(x = Campo, y = Value, fill=Sexo)) + geom_boxplot(position="dodge")
stat1 + labs(title="Nota de las asignaturas en escuela privada y publica", x = "Asignaturas", y = "Nota")
```

Nota de las asignaturas en escuela privada y publica



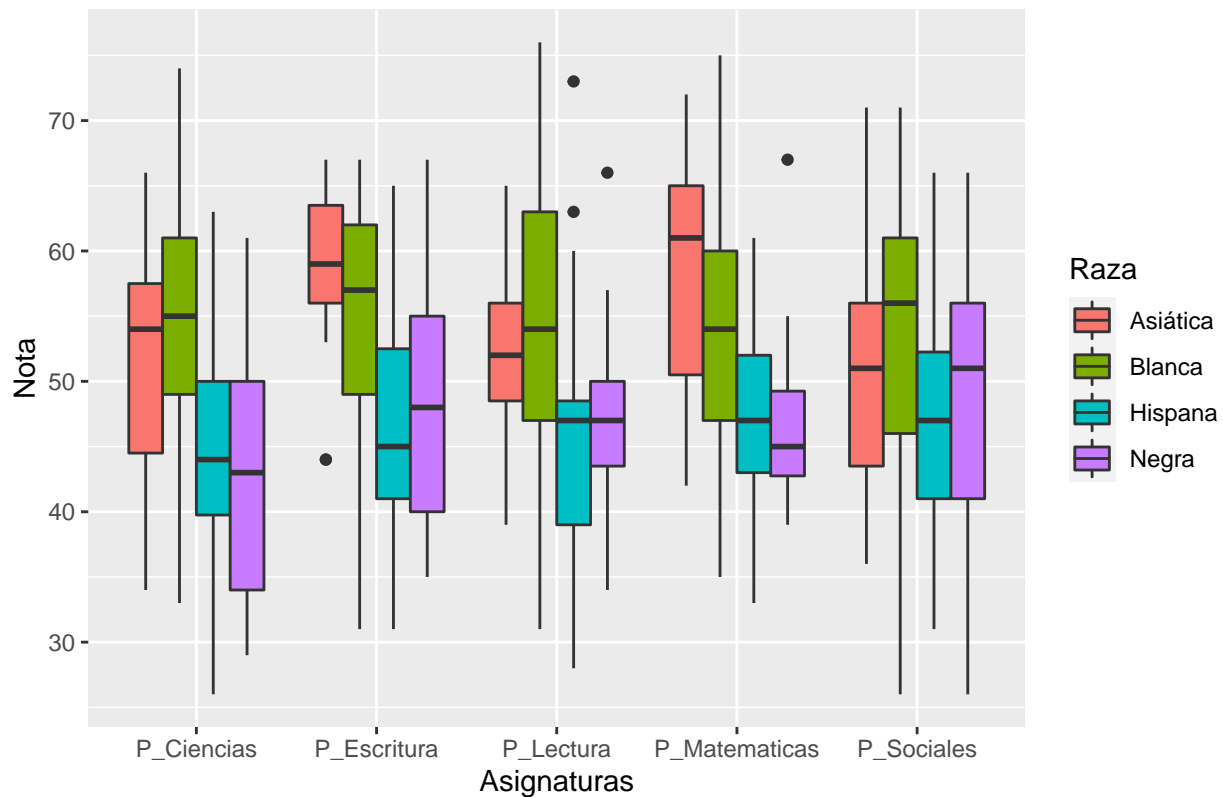
```
p <- aggregate(cbind(P_Lectura, P_Escritura, P_Matematicas, P_Sociales, P_Ciencias)~Sexo, data = educacion,
kable(p)
```

Sexo	P_Lectura	P_Escritura	P_Matematicas	P_Sociales	P_Ciencias
Hombre	52.82418	50.12088	52.94505	51.79121	53.23077
Mujer	51.73394	54.99083	52.39450	52.91743	50.69725

```
#raza
educacion2 <- educacion %>%
gather(Campo, Value, P_Ciencias, P_Escritura, P_Lectura, P_Matematicas, P_Sociales, -Raza)

stat1 <- ggplot(educacion2, aes(x = Campo, y = Value, fill=Raza)) + geom_boxplot(position="dodge")
stat1 + labs(title="Nota de las asignaturas en diferentes razas", x = "Asignaturas", y = "Nota")
```

Nota de las asignaturas en diferentes razas



```
p <- aggregate(cbind(P_Lectura, P_Escritura, P_Matematicas, P_Sociales, P_Ciencias)~Raza, data = educac,
kable(p)
```

Raza	P_Lectura	P_Escritura	P_Matematicas	P_Sociales	P_Ciencias
Asiática	51.90909	58.00000	57.27273	51.00000	51.45455
Blanca	53.92414	54.05517	53.97241	53.68276	54.20000
Hispana	46.66667	46.45833	47.41667	47.79167	45.37500
Negra	46.80000	48.20000	46.75000	49.45000	42.80000

En primer lugar observamos la gráfica general que relaciona asignaturas con su correspondiente nota en hombres y mujeres.

En este gráfico no distingue entre centros privados y públicos a si que veremos una perspectiva diferente a la vista en el apartado b. Primero observamos como los hombres sacan mejores puntuaciones en Ciencias con gran diferencia tanto en promedio como en las mejores notas.

A su vez en lectura también ocurre lo mismo pero de forma menos remarcada sin embargo se ve como aquellos que obtienen las notas mas altas son hombres, aunque en promedio no se diferencian tanto un grupo de otro.

Igual para en matemáticas, si bien parece que las mujeres en promedio tienen mejores notas aquellos que mas destacan son hombres. En sociales y escritura pasa todo lo contrario las mujeres no solo destacan mas sino también se ven superiores a los hombres en promedio, siendo sociales aquella asignatura que mas diferencia hay entre mujeres y hombres, mostrandose estas primeras superiores de forma considerable en promedio.

En cuanto a las razas no hay tanta variedad aquellos con mejores puntuaciones son o bien de raza asiática o bien de raza blanca, mostrando una amplia brecha entre la raza hispana y negra. Siendo esto así en ciencias,

escritura, lectura y matemáticas. En algunas asignaturas los asiáticos tienen mejores puntuaciones y en otra los blancos aunque con poca diferencia.

Aquella asignatura que mantiene mas igualada la situación es sociales donde a pesar de que las medias para los blancos y asiáticos siguen siendo superiores no lo son por mucho a los hispanos y negros.