

Lab Gráficos en Lenguaje R

Objetivo

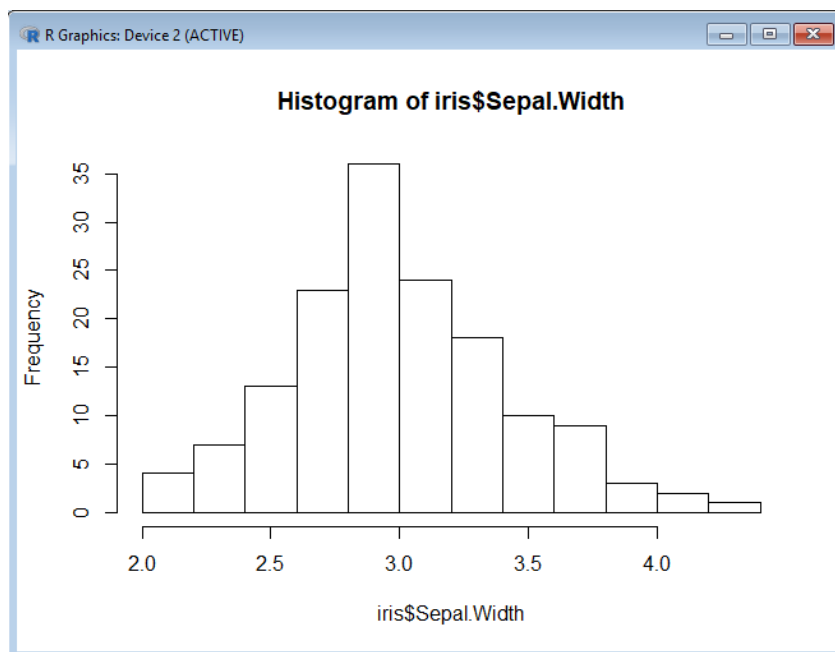
- Mostrar al participante las funciones más útiles para la creación de graficos en el entorno R

Contenido

Gráficos de Variables Continuas

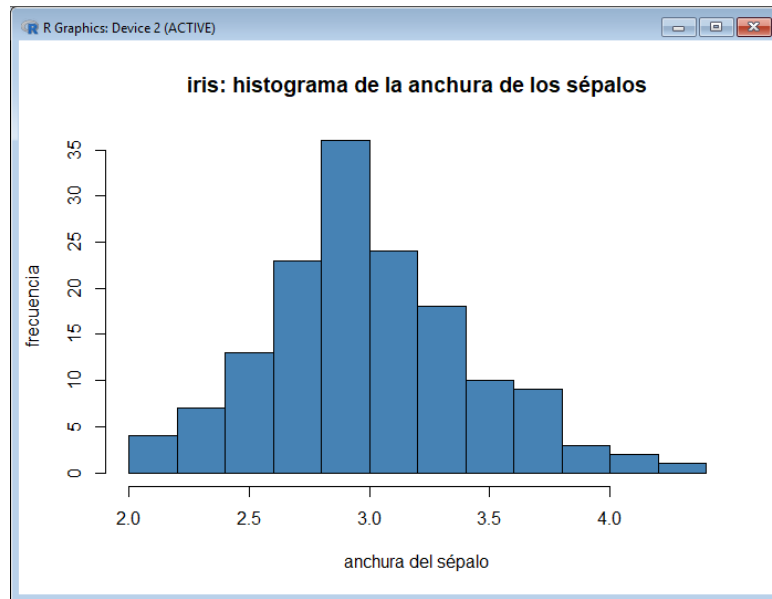
1. Inspecciona la columna Sepal.Width de la tabla iris. Crea un histograma con los dato de la columna Sepal.Width.

```
> hist(iris$Sepal.Width)
```



2. Personaliza el histograma del ítem anterior.

```
> hist(iris$Sepal.Width, main = "iris: histograma de la anchura de los sépalos",  
      xlab = "anchura del sépalo", ylab = "frecuencia",  
      col = "steelblue")
```

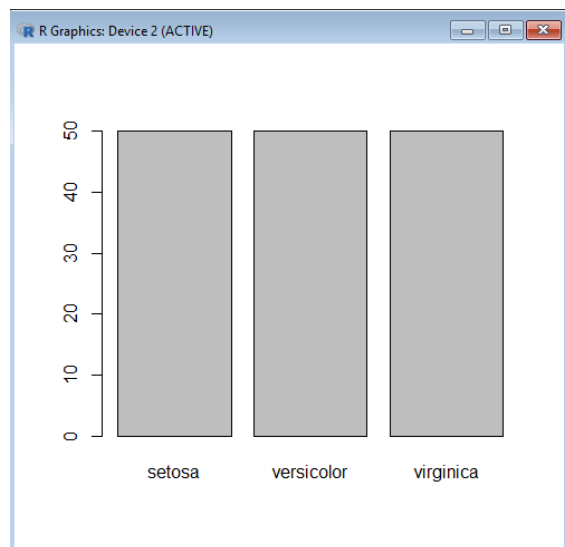


3. Estudia la distribución de las temperaturas en Nueva York (usa `airquality`).
4. Usa las funciones `png` y `jpeg` para guardar alguno de los gráficos anteriores en tu disco duro.

Gráficos de Variables Categóricas

5. Muestra cómo en `iris` existe el mismo número de observaciones de cada especie.

```
>barplot(table(iris$Species))
```



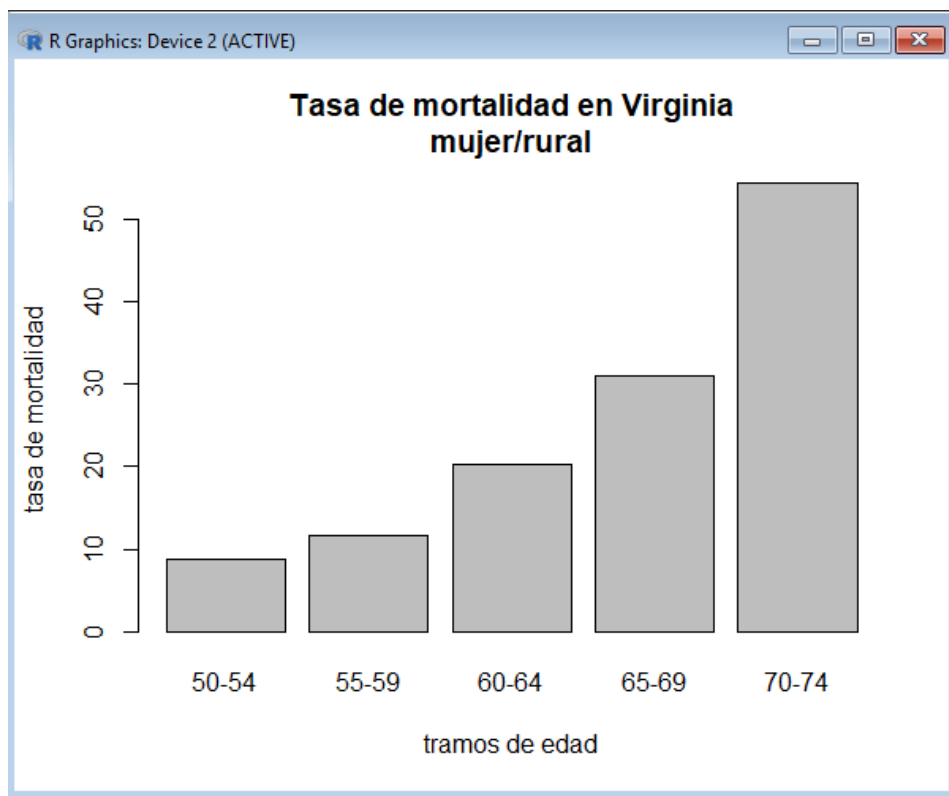
6. Usa los parámetros `main`, `xlab`, `ylob` y `col` mejora el aspecto del gráfico del ítem 5.

- Investiga el argumento `horiz` de `barplot` para crear un gráfico de barras horizontales.
- Usa la ayuda para ver el tipo de datos que almacena `VADeaths`.

```
>?VADeaths
```

- Usando `VADeaths` grafica la tasa de mortalidad por edad.

```
>barplot(VADeaths[, 2], xlab = "tramos de edad", ylab = "tasa de mortalidad",  
main = "Tasa de mortalidad en Virginia\nmujer/rural")
```



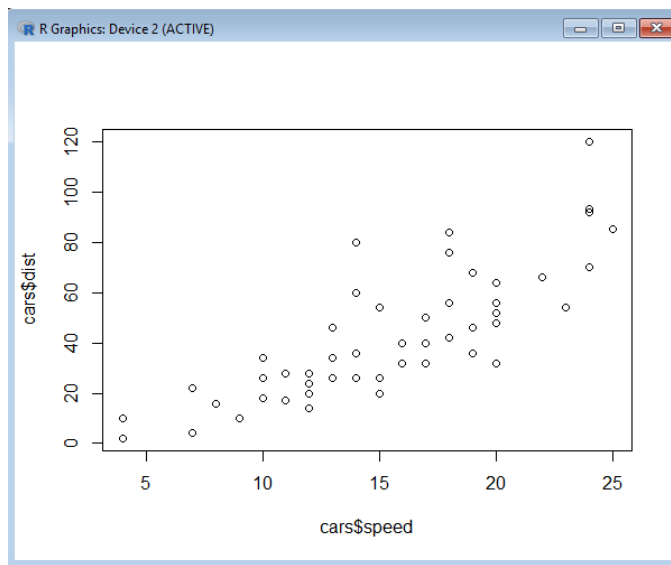
Nota: También puedes usar `pie()`

```
>pie(VADeaths[, 2], main="Mortalidad por Edad")
```

Gráficos de Dispersión

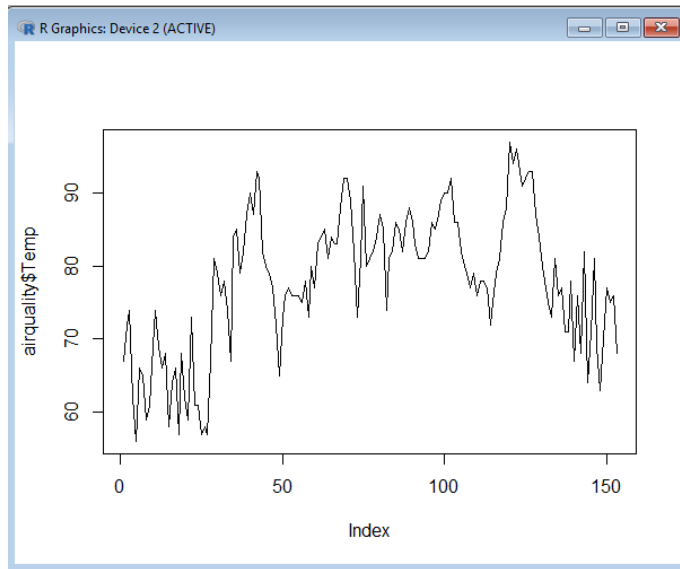
- Los gráficos de dispersión muestran la relación entre dos variables numéricas. Grafica la distancia de frenado versus la velocidad del auto, para un conjunto de datos en tabla `cars`.

```
>plot(cars$speed, cars$dist)
```



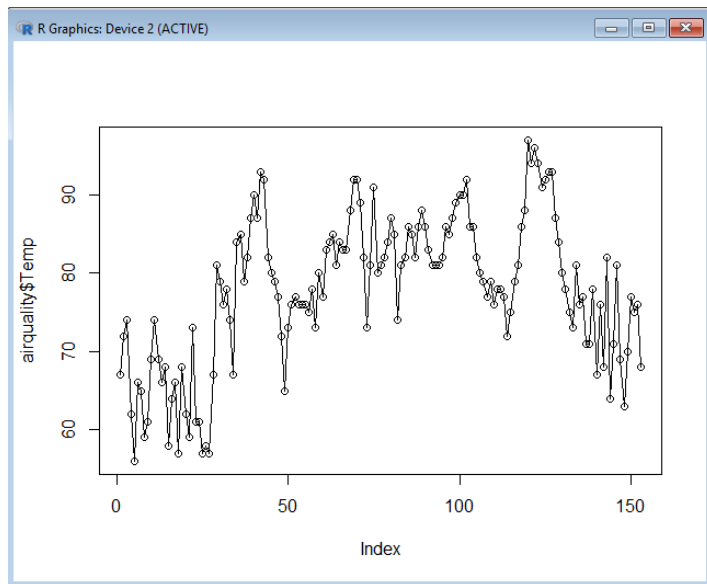
11. Usa los parámetros main, xlab, ylab y col para mejorar el aspecto del grafico del ítem anterior.
12. Representa gráficamente la anchura del sépalo contra su longitud (usando iris). Interpreta el gráfico.
13. Cuando una de las variables tiene un orden determinado (por ejemplo, es una variable temporal) pueden utilizarse líneas para unir los puntos de un diagrama de dispersión. Utilizando el hecho de que las observaciones de airquality están ordenadas temporalmente, crea un gráfico de dispersión usando la columna temp en airquality uniendo los puntos (medidas).

```
>plot(airquality$Temp, type = "l")
```



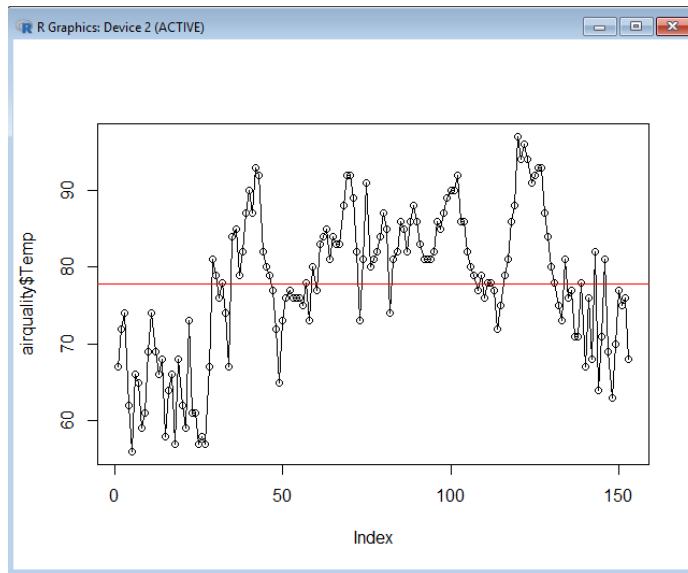
14. En los gráficos con R se pueden añadir capas adicionales a los gráficos, ejecuta las siguientes instrucciones.

```
>plot(airquality$Temp)
>lines(airquality$Temp)
```



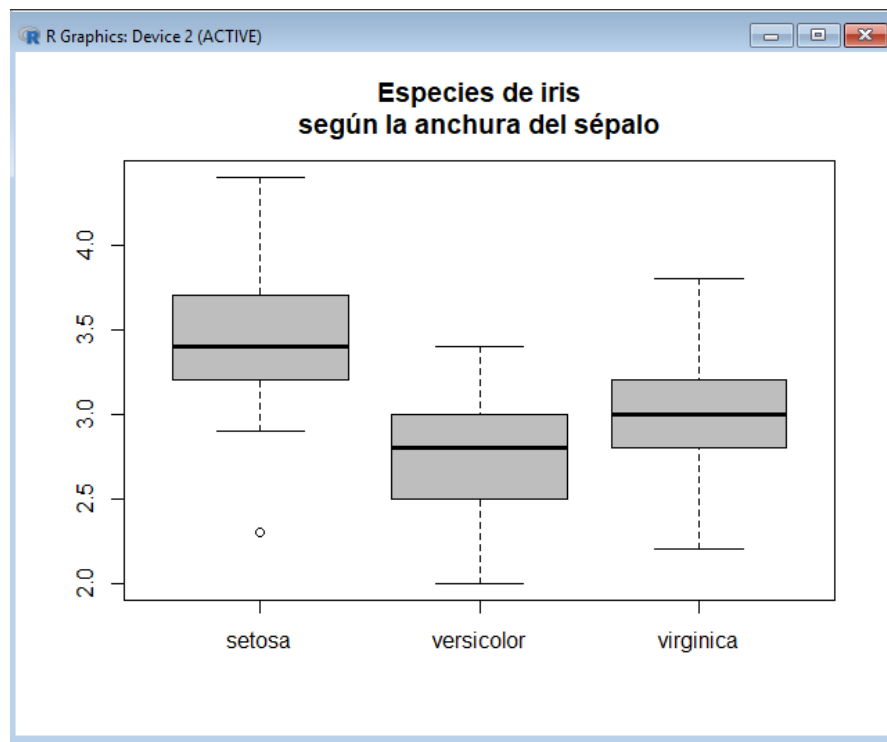
15. Ejecuta las siguientes instrucciones e interpreta el grafico resultante.

```
>plot(airquality$Temp)
>lines(airquality$Temp)
>abline(h = mean(airquality$Temp), col = "red")
```



16. Consulta la ayuda de la función `abline` y úsala para añadir líneas (no solo horizontales) a alguno de los gráficos anteriores.
17. Consulta `?par`, una página de ayuda en R que muestra gran cantidad de parámetros modificables en un gráfico. Investiga y usa `col`, `lty` y `lwd`.
18. Los diagramas de cajas (`boxplot`) estudian la distribución de una variable continua en función de una variable categórica. Estudiar la distribución de la anchura del sépalo en iris en función de la especie usando diagramas de cajas.

```
>boxplot(iris$Sepal.Width ~ iris$Species, col = "gray",  
         main = "Especies de iris\nsegún la anchura del sépalo")
```



Nota: La notación $y \sim x$ es muy común en R y significa que vas a hacer algo con y en función de x

19. Muestra la distribución de las temperaturas en Nueva York en función del mes.
20. Crea una tabla adicional seleccionando todas las columnas menos mes y día; luego haz un plot de ella y trata de encontrar relaciones (cualitativas) entre la temperatura y el viento, o el ozono,...
21. Usando el conjunto de datos mtcars (consulta ?mtcars), averigua:
 - cuál es el modelo que menos consume
 - cuál es el consumo medio de los modelos de 4 cilindros
 - Grafica los datos del consumo por vehículo y dibuja una línea con el dato de la media del consumo
22. Crea un histograma de la temperatura en Nueva York (usando airquality) y después usa abline para dibujar una línea vertical roja en la media de la distribución. Puedes obtener la media con summary o bien aplicando la función mean a la columna de interés.
23. Carga el fichero pisasci2006.csv que contiene los promedios de los resultados en ciencia notas por país en las pruebas PISA de 2006.
 - Identifica y muestra la línea correspondiente a Peru (Country vs Edu)

- Haz una gráfica en que se muestre la nota (Edu) en función de los ingresos (income)
- Construye un histograma con la nota y añade una recta vertical roja en la correspondiente a Peru