

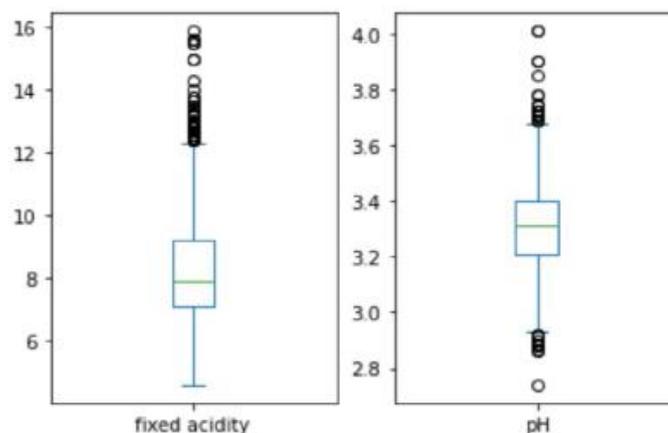
### 3.4. Outliers y diagramas de cajas y bigotes.

Resumen la distribución de cada variable cuantitativa, dibujándose la mediana y una caja alrededor del primer (Q1) y tercer (Q3) cuartiles. Permite visualizar cuáles son los datos centrales, los datos adyacentes y los datos atípicos, si los hubiera.

- El rango intercuartílico (RI) viene representado por la altura de la caja ( $Q3-Q1$ ).
- Un punto que esté 1.5 veces el RI por encima del tercer cuartil o 1.5 veces el RI por debajo del primer cuartil, es un posible valor atípico u outlier (representados como círculos en el ejemplo 43).
- El bigote superior se representa como una línea desde Q3 hasta el máximo punto excluyendo posibles outliers. De la misma forma, el bigote inferior se representa como una línea desde el Q1 hasta el mínimo punto que no cae en el rango de posibles outliers.
- El conocimiento del dominio del problema es fundamental a la hora de descartar posibles outliers.

#### Ejercicio 94\_01. Diagrama de cajas y bigotes en Python

```
import pandas as pd
import matplotlib.pyplot as plt
nomFichero = "winequality-red.csv"
# No es necesario añadir nombres de columnas, ya que el csv los trae
# Se usa ; como separador, y no la coma por defecto
datos = pd.read_csv(nomFichero, sep=";")
# Seleccionamos algunas variables y representamos sus diagramas
datos2 = datos.loc[:, ["fixed acidity", "pH"]]
# Layout, 1 fila y 2 columnas
datos2.plot(kind="box", subplots=True, layout=(1,2), sharex=False, sharey=False)
plt.show()
```



**Autores:** María Inmaculada Rodríguez García , María Gema Carrasco García, Javier González Enrique, Juan Jesús Ruiz Aguilar, Ignacio J. Turias Domínguez. [Universidad de Cádiz](https://www.usc.es/)

### Ejercicio 94\_02. Diagrama de cajas y bigotes en MATLAB

```
% Importar datos desde un archivo CSV en MATLAB
nomFichero = 'winequality-red.csv';
datos = readtable(nomFichero, 'Delimiter', ',');
% Seleccionar algunas variables
columnas_seleccionadas = {'fixedAcidity', 'pH'};
datos2 = datos(:, columnas_seleccionadas);
% Configurar el tamaño del gráfico
figure('Position', [100, 100, 800, 400]);
% Mostrar diagramas de caja para las variables seleccionadas
subplot(1, 2, 1);
boxplot(datos2.('fixedAcidity'));
title('Diagrama de caja para "fixedAcidity"');
xlabel('fixed acidity');
ylabel('Valores');
subplot(1, 2, 2);
boxplot(datos2.('pH'));
title('Diagrama de caja para "pH"');
xlabel('pH');
ylabel('Valores');
```

