

# Hoja informativa: Relaciones entre datasets

## Práctica

```
# Crear un gráfico de dispersión

data.plot(x='column_x', y='column_y', kind='scatter');

# Crear gráficos de dispersión por pares para columnas DataFrames

pd.plotting.scatter_matrix(data)

# Construir un gráfico de celdas
# gridsize - número de celdas a lo largo del eje horizontal

data.plot(x='column_x', y='column_y', kind='hexbin', gridsize=20, sharex=False);

# Cálculo del coeficiente de correlación de Pearson
# El coeficiente no depende del procedimiento de cálculo

print(data['column_1'].corr(data['column_2']))
print(data['column_2'].corr(data['column_1']))

# Coeficiente de correlación de Pearson para cada par de columnas

data.corr()
```

## Teoría

**Gráfico de dispersión:** un gráfico en el que cada punto representa dos valores (sus valores X e Y)

**Gráfico binning hexagonal:** un gráfico dividido en celdas; su color indica el número de puntos que caen en esa celda

**Coeficiente de correlación de Pearson:** un número entre -1 y 1 que muestra cuánto cambia un valor cuando otro también cambia