

Corrigé

P1C3 - Produisez des graphiques propres avec Seaborn



Python

```
import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt


df = pd.read_csv("heart.csv")

sns.set_theme(style="whitegrid")


# Q1 : distribution de l'âge selon le diagnostic

sns.kdeplot(data=df, x="Age", hue="HeartDisease", fill=True,
palette="Set2")

plt.title("Les patients malades sont en moyenne plus âgés",
fontsize=13)

plt.xlabel("Âge")

plt.show()
```

```
# Q2 : relation MaxHR / âge selon le sexe

sns.lmplot(data=df,      x="Age",      y="MaxHR",      hue="Sex",
palette="Set1")

plt.title("La fréquence cardiaque maximale décline avec
l'âge", fontsize=13)

plt.show()
```

Pour Q1, le paramètre `hue="HeartDisease"` est la décision clé : sans lui, vous obtiendriez une seule courbe de densité, sans pouvoir comparer les deux groupes. Le paramètre `fill=True` colore la zone sous chaque courbe, ce qui facilite la comparaison visuelle. Sur ce dataset, la courbe des patients malades est clairement décalée vers la droite : leur âge moyen avoisine 56 ans, contre environ 50,5 ans pour les patients sains. La palette `"Set2"` a été choisie car `HeartDisease` est une variable catégorielle (0 ou 1) : une palette qualitative à deux teintes bien contrastées suffit.

Pour Q2, `lmplot` a été préféré à `regplot` parce que le sexe (`Sex`) doit être comparé : `lmplot` génère automatiquement une droite de régression par catégorie, alors que `regplot` ne propose pas nativement ce découpage. On observe une tendance baissière dans les deux groupes : la fréquence cardiaque maximale diminue globalement avec l'âge, qu'on soit un homme (725 patients dans le dataset) ou une femme (193 patients).

Variante à anticiper : que se passerait-il en remplaçant `kdeplot` par `histplot` ? Vous obtiendriez des barres au lieu d'une courbe lissée, ce qui reste valable, mais rend la comparaison entre les deux groupes plus difficile à lire dès que les effectifs diffèrent fortement, comme c'est le cas ici entre patients sains et malades. Le `kdeplot` reste préférable lorsque l'objectif est justement de comparer des distributions entre groupes.