Partie 2

Ex1:

5)- Oui, il existe une forte corrélation entre l'âge et la taille (correlation = 0.926317 Une **valeur de 0.926** signifie une **forte corrélation positive** entre l’âge et la taille).

EX3 :

3)-**Que pouvez-vous dire de ces données ?**

Boston est un dataset contenant des **données sur l’immobilier**(La variable cible est **medv** « valeur médiane des maisons »,Il y a **14 variables explicatives** (ex : crim = taux de criminalité, lstat = pourcentage de population pauvre, etc.),Certaines variables sont **corrélées** entre elles.

10)- Si medv n’est pas **normalement distribué**, un modèle linéaire simple pourrait ne pas être idéal.

14)- les **p-values** de lstat et age sont faibles (**< 0.05**), ces variables ont un effet significatif sur medv.

16)- log(lstat) améliore le **R²** du modèle et réduit la **p-value**, cela signifie qu'une transformation logarithmique est utile.

18)- une variable a une **corrélation élevée** avec medv (proche de 1 ou -1), elle est potentiellement importante pour la prédiction

20)- Un **modèle plus simple** avec seulement les variables **significatives** peut être plus performant.

23)- un **modèle de degré plus élevé** améliore le **R² ajusté** et réduit l’erreur, cela signifie que la relation entre medv et lstat **n’est pas strictement linéaire**.