Compte rendu Travaux pratiques en langage R

Problème 1 :

**Pour charger les données :**

data = *read.csv*("psychology.csv", *TRUE*)  
*print*(data)

Le code permet de lire un fichier CSV nommé “psychology.csv” qui contient des données sur la psychologie des individus. Le paramètre TRUE (ou header = TRUE) indique que la première ligne du fichier contient les noms des variables. C’est-à-dire que dans le fichier csv à la 1er ligne on remarque : IBE, IP et qu’il est identifié comme un header. La fonction print (data) affiche le contenu du fichier dans la console.

**Question 1 :**

# ---- Question 1 ----  
# Statistiques descriptives pour l'IBE  
summary\_IBE = *summary*(data$IBE)  
mean\_IBE = *mean*(data$IBE)  
median\_IBE = *median*(data$IBE)  
sd\_IBE = *sd*(data$IBE)  
IQR\_IBE = *IQR*(data$IBE)  
  
# Affichage des résultats  
*cat*("Statistiques descriptives pour l'IBE:\n")  
*cat*("Sommaire:", summary\_IBE, "\n")  
*cat*("Moyenne:", mean\_IBE, "\n")  
*cat*("Médiane:", median\_IBE, "\n")  
*cat*("Écart-type:", sd\_IBE, "\n")  
*cat*("IQR:", IQR\_IBE, "\n")  
  
# Statistiques descriptives pour l'IP  
summary\_IP = *summary*(data$IP)  
mean\_IP = *mean*(data$IP)  
median\_IP = *median*(data$IP)  
sd\_IP = *sd*(data$IP)  
IQR\_IP = *IQR*(data$IP)  
  
# Affichage des résultats  
*cat*("\nStatistiques descriptives pour l'IP:\n")  
*cat*("Sommaire:", summary\_IP, "\n")  
*cat*("Moyenne:", mean\_IP, "\n")  
*cat*("Médiane:", median\_IP, "\n")  
*cat*("Écart-type:", sd\_IP, "\n")  
*cat*("IQR:", IQR\_IP, "\n")

Cette partie du code permet de calculer et d’afficher les statistiques descriptives pour la variable IBE et IP, qui représente l’indice de bien-être des individus (IBE) et de productivité (IP). La fonction summary(data$IBE) renvoie les valeurs minimale, maximale, moyenne, médiane et les quartiles de la variable. Les fonctions mean(data$IBE), median(data$IBE), sd(data$IBE) et IQR(data$IBE) renvoient respectivement la moyenne, la médiane, l’écart-type et l’intervalle interquartile de la variable. La fonction cat() permet d’afficher les résultats dans la console avec des sauts de ligne (\n) et des séparateurs (:) de la même manière qu’un print.Pour IP on utilise la même chose ce qui renvoie en console :

Statistiques descriptives pour l'IBE:

Sommaire: 1 24.25 50 48.29333 74.5 100

Moyenne: 48.29333

Médiane: 50

Écart-type: 29.28626

IQR: 50.25

Statistiques descriptives pour l'IP:

Sommaire: 0 4.09425 4.6505 4.475667 5.264 6.092

Moyenne: 4.475667

Médiane: 4.6505

Écart-type: 1.201006

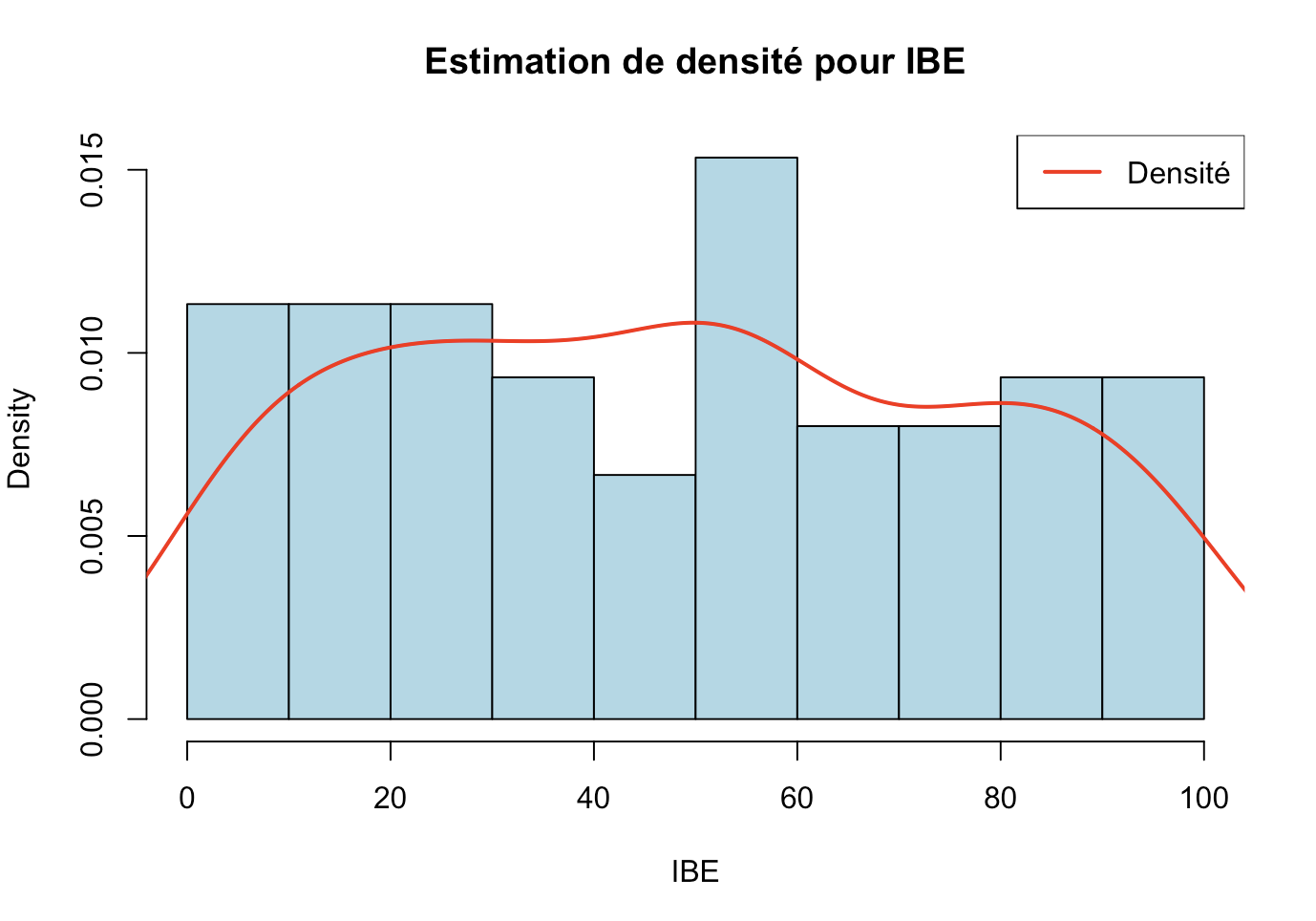
IQR: 1.16975

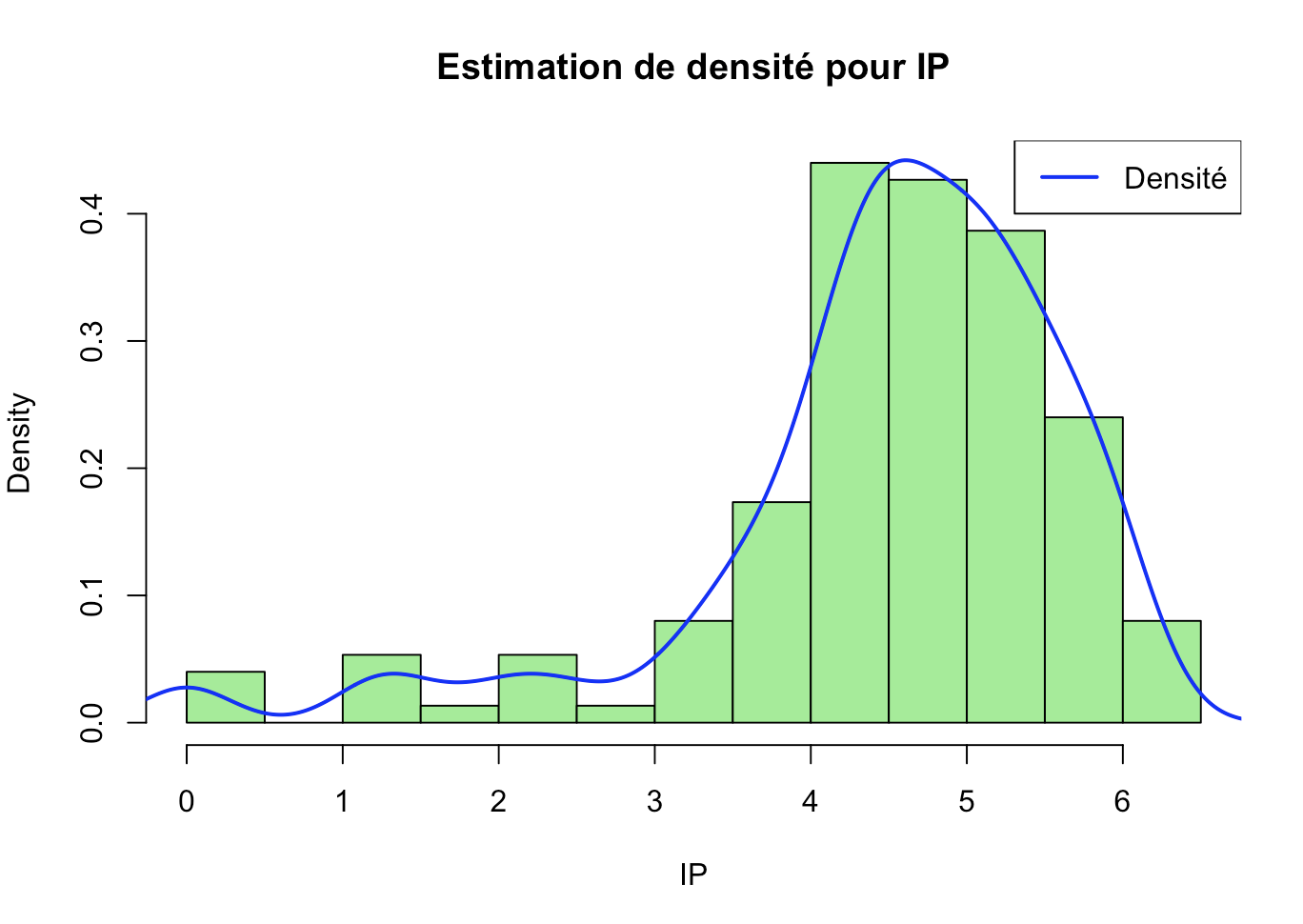
**Question 2 :**

# ---- Question 2 ----  
# Créer un histogramme pour la variable IBE  
*hist*(data$IBE, main="Estimation de densité pour IBE", xlab="IBE", col="lightblue", border="black", probability=*TRUE*)  
*lines*(*density*(data$IBE), col="red", lwd=2)  
  
# Ajouter une légende  
*legend*("topright", legend="Densité", col="red", lwd=2)  
  
# Créer un histogramme pour la variable IP  
*hist*(data$IP, main="Estimation de densité pour IP", xlab="IP", col="lightgreen", border="black", probability=*TRUE*)  
*lines*(*density*(data$IP), col="blue", lwd=2)  
  
# Ajouter une légende  
*legend*("topright", legend="Densité", col="blue", lwd=2)

Cette partie du code permet de créer des histogrammes pour les variables IBE et IP, afin de visualiser la distribution des données. La fonction hist() prend en paramètre la variable à représenter, le titre du graphique (main), le nom de l’axe des abscisses (xlab), la couleur des barres (col), la couleur des bordures (border) et le type de fréquence (probability=TRUE pour avoir des fréquences relatives). La fonction lines() permet de tracer la courbe de densité estimée par noyau (KDE) sur l’histogramme, en précisant la couleur (col) et l’épaisseur (lwd) de la ligne. La fonction legend() permet d’ajouter une légende au graphique, en indiquant la position (topright), le texte (legend), la couleur (col) et l’épaisseur (lwd) de la ligne.

Ainsi on a graphiquement pour IBE et IP avec les couleurs appliquées :

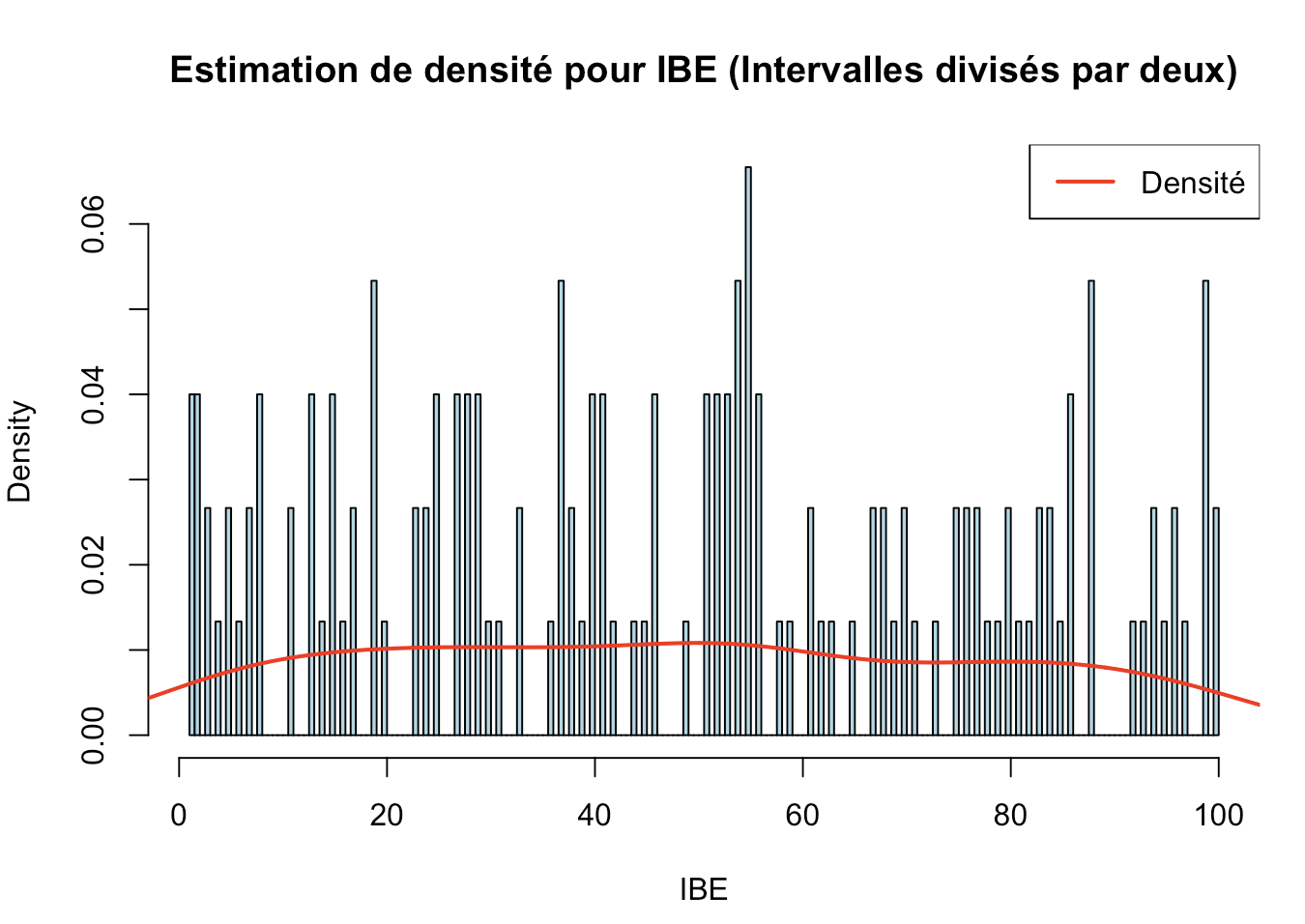


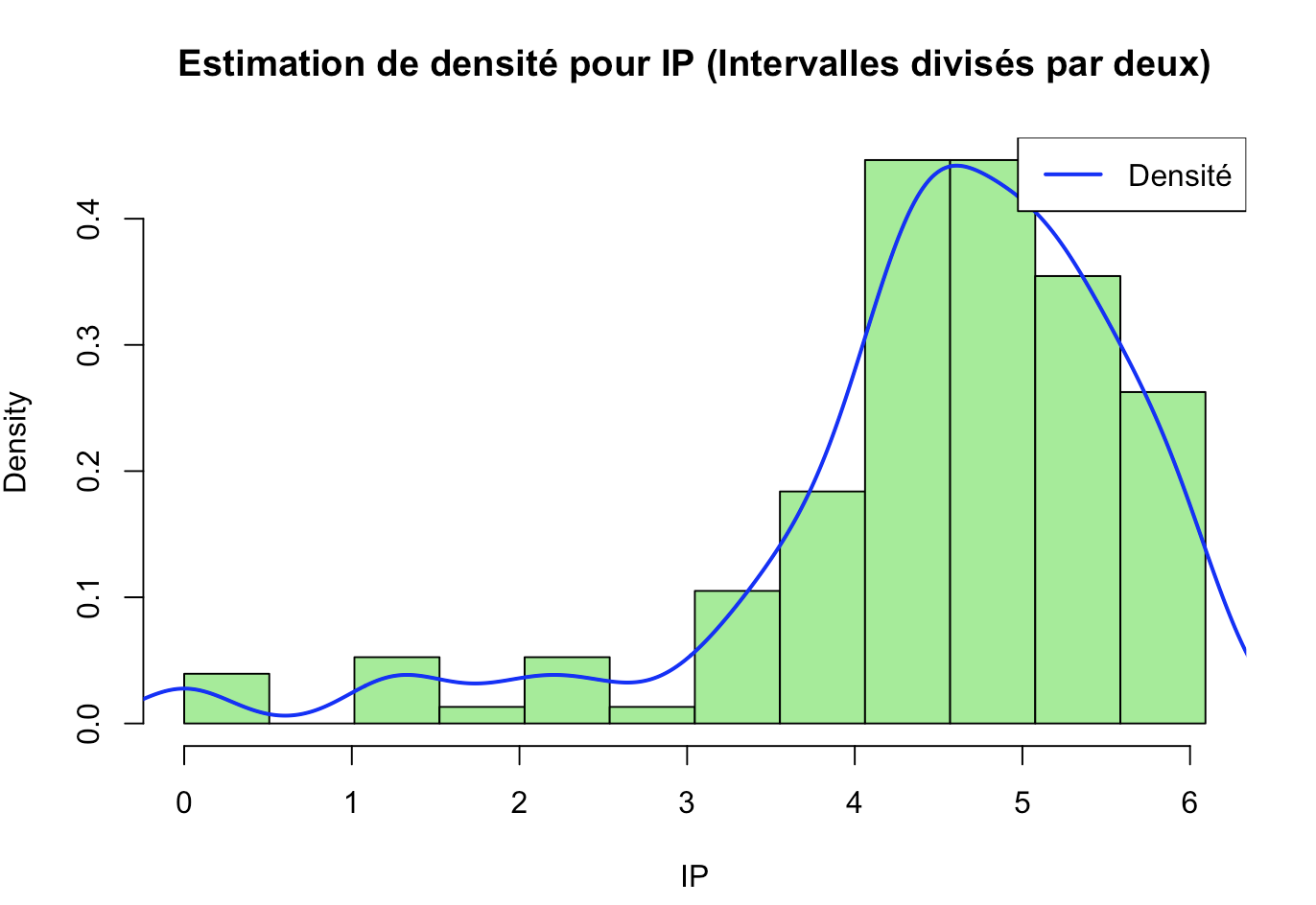


**Question 3 :**

# ---- Question 3 ----  
  
# Diviser par deux la largeur des intervalles pour IBE  
new\_breaks\_ibe = *seq*(*min*(data$IBE), *max*(data$IBE), length.out = (*length*(*seq*(*min*(data$IBE), *max*(data$IBE))) - 1) \* 2 + 1)  
  
# Créer un histogramme pour la variable IBE avec des intervalles divisés par deux  
*hist*(data$IBE, breaks = new\_breaks\_ibe, main="Estimation de densité pour IBE (Intervalles divisés par deux)", xlab="IBE", col="lightblue", border="black", probability=*TRUE*)  
*lines*(*density*(data$IBE), col="red", lwd=2)  
  
# Ajouter une légende  
*legend*("topright", legend="Densité", col="red", lwd=2)  
  
# Diviser par deux la largeur des intervalles pour IP  
new\_breaks\_ip = *seq*(*min*(data$IP), *max*(data$IP), length.out = (*length*(*seq*(*min*(data$IP), *max*(data$IP))) - 1) \* 2 + 1)  
  
# Créer un histogramme pour la variable IP avec des intervalles divisés par deux  
*hist*(data$IP, breaks = new\_breaks\_ip, main="Estimation de densité pour IP (Intervalles divisés par deux)", xlab="IP", col="lightgreen", border="black", probability=*TRUE*)  
*lines*(*density*(data$IP), col="blue", lwd=2)  
  
# Ajouter une légende  
*legend*("topright", legend="Densité", col="blue", lwd=2)

Ce code permet de créer des histogrammes avec des intervalles plus fins pour les variables IBE et IP, afin de mieux apprécier la forme de la distribution. La fonction seq() permet de créer une séquence de valeurs régulièrement espacées entre le minimum et le maximum de la variable, en précisant le nombre de valeurs (length.out). Ce nombre est calculé en multipliant par deux le nombre d’intervalles de l’histogramme précédent, moins un. La fonction hist() prend en paramètre le vecteur des intervalles (breaks) à utiliser pour créer l’histogramme. Les autres fonctions sont les mêmes que pour la question 2. Ce qui donne graphiquement :

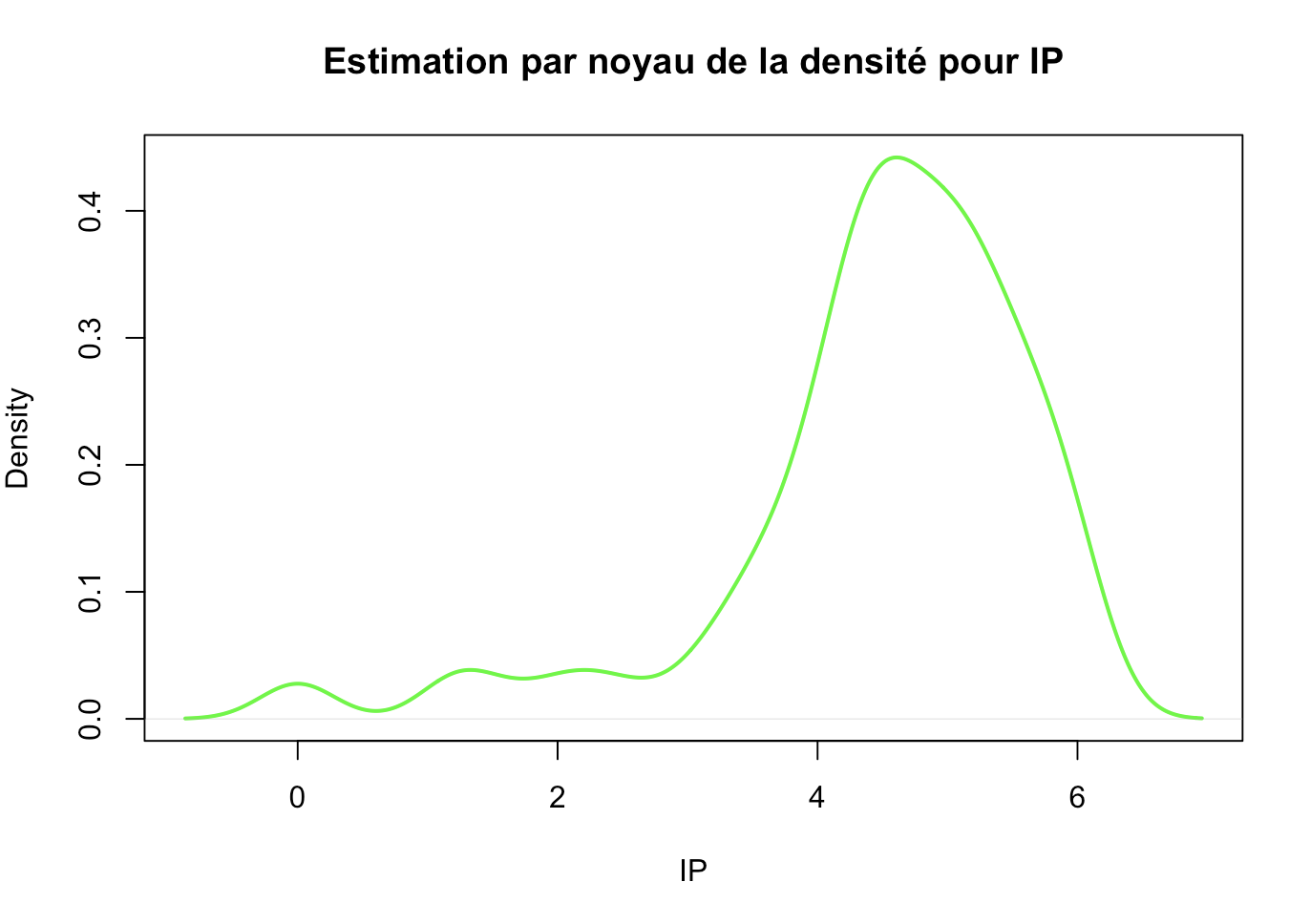
****

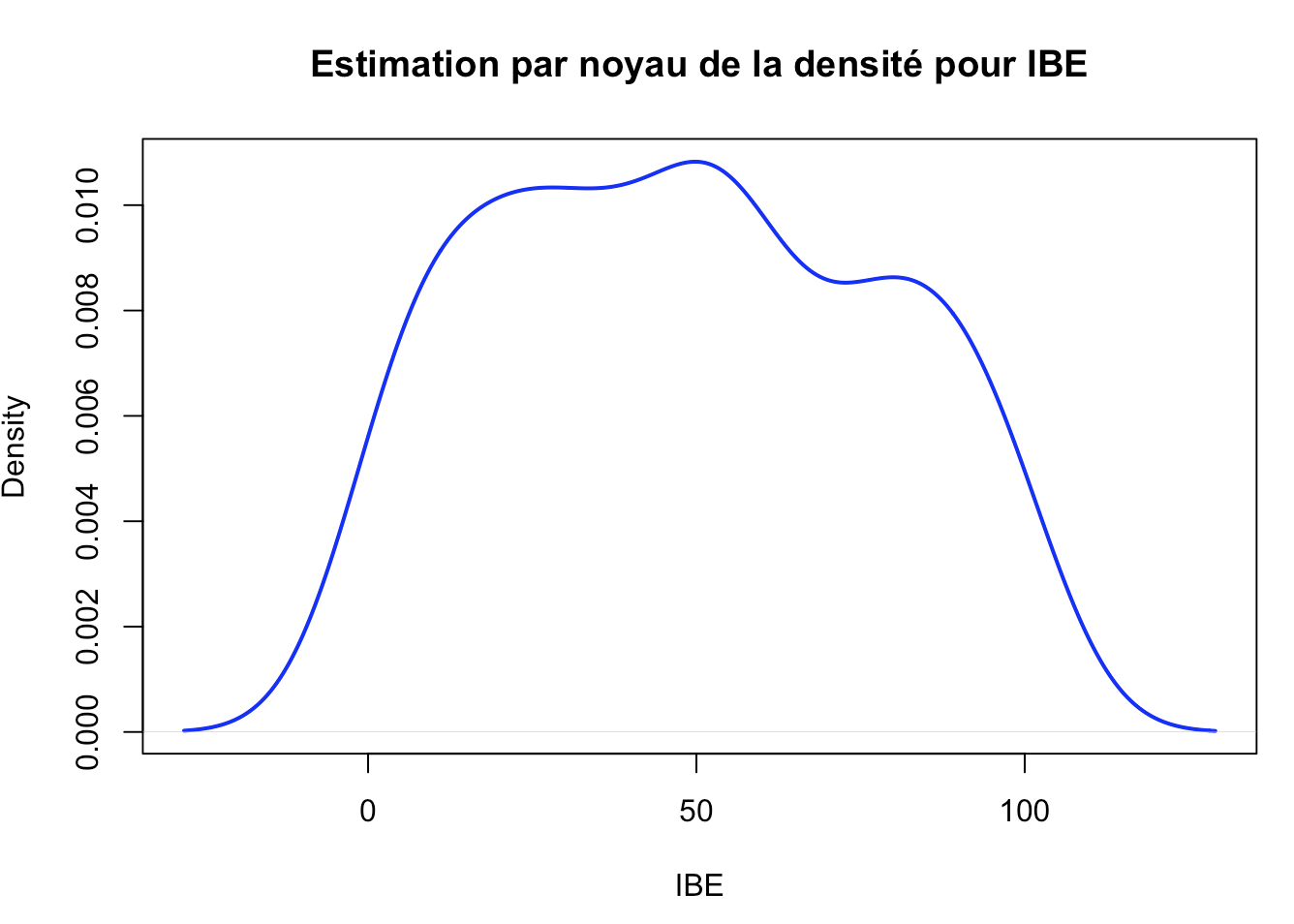
****

**Question 4 :**

# ---- Question 4 ----  
# Estimation par noyau de la densité pour IBE  
density\_ibe = *density*(data$IBE)  
  
# Tracer le KDE pour IBE  
*plot*(density\_ibe, main="Estimation par noyau de la densité pour IBE", xlab="IBE", col="blue", lwd=2)  
  
# Estimation par noyau de la densité pour IP  
density\_ip = *density*(data$IP)  
  
# Tracer le KDE pour IP  
*plot*(density\_ip, main="Estimation par noyau de la densité pour IP", xlab="IP", col="green", lwd=2)

Cette partie du code permet de tracer les courbes de densité estimées par noyau (KDE) pour les variables IBE et IP, sans utiliser d’histogramme. La fonction density() permet de calculer la densité par noyau pour une variable. La fonction plot() permet de tracer la courbe de densité, en précisant le titre du graphique (main), le nom de l’axe des abscisses (xlab), la couleur (col) et l’épaisseur (lwd) de la ligne.

****

****

**Question 5 :**