On souhaite étudier les résultats d'une enquête sur 178 clients d'une banque. Pour chaque client, on a noté l'âge du client `AGE`, le degré d'éducation `ED`, le nombre de mois consécutifs de travail avec le même employeur `EMP`, le nombre d'années d'habitation à l'adresse actuelle `ADRE`, le revenu annuel du foyer en milliers de dollars `REVENU`, le rapport des crédits en cours sur le revenu en pourcentage `CREDIT` et une variable indiquant si le client a déjà été à découvert ou non `DECOUVERT`.

1. Récupérer le fichier `banque.txt` et charger les données dans `R` :
   `donnees = read.table(file.choose(),header=T)`
2. Il y a 7 variables mesurées sur l'échantillon. Quel est le type de chacune de ces variables ?
3. On veut tout d'abord présenter les résultats concernant l'âge des clients. C'est une variable quantitative continue.
   (a) On affiche les données sur l'âge avec `donnees$AGE` ou `donnees[,1]` (première colonne).
   (b) On peut calculer les paramètres statistiques de cette variable : `mean(donnees$AGE)`, `var(donnees$AGE)`, `sd(donnees$AGE)`, `median(donnees$AGE)`, `quantile(donnees$AGE)`, `min(donnees$AGE)`, `max(donnees$AGE)`.
   (c) On peut tracer le diagramme en boîte de l'âge : `boxplot(donnees$AGE)`.
   (d) Pour tracer l'histogramme de l'âge : `hist(donnees$AGE)`.
4. On veut maintenant représenter le degré d'éducation des clients. C'est une variable qualitative ordinale.
   (a) Afficher les données.
   (b) Les paramètres statistiques de cette variable sont uniquement les effectifs de chaque modalité. On les obtient avec la fonction `summary` :
   `effectifsED=summary(donnees$ED)`
   (attention : l'ordre utilisé est l'ordre alphabétique).
   (c) Calculer les fréquences : `effectifsED/sum(effectifsED)*100`
   (d) On peut tracer le diagramme en barres : `barplot(effectifsED)`.
5. On veut enfin représenter le découvert des clients. C'est une variable qualitative nominale (binaire).
   (a) Afficher les données.
   (b) Les paramètres statistiques de cette variable sont là-aussi les effectifs de chaque modalité. Calculer ces effectifs et les fréquences.
   (c) On peut tracer le diagramme en secteurs : `pie(effectifsDECOUVERT)`.
6. On veut maintenant étudier le revenu suivant le degré d'éducation. Il faut pour cela séparer les données de revenu en quatre groupes (suivant le degré d'éducation).

(a) On peut utiliser la fonction `split` pour séparer une variable en fonction d'un facteur : `RevenuED=split(donnees$REVENU,donnees$ED)`

(b) Pour visualiser les quatre groupes simultanément, on peut tracer les diagramme en boîtes `boxplot(donnees$REVENU~donnees$ED)`, ou encore `boxplot(RevenuED)`.

(c) On peut calculer les paramètres statistiques pour chacune des catégories : on utilise la fonction `sapply`

```
sapply(RevenuED,mean)
sapply(RevenuED,var)
sapply(RevenuED,sd)
```

7. On veut ensuite représenter les effectifs de degré d'éducation suivant le groupe `DECOUVERT`. Il faut pour cela séparer les données de degré d'éducation en deux groupes (`DECOUVERT oui ou non`).

   (a) Créer deux vecteurs `EDoui` et `EDnon` :

   ```
   EDsplitDECOUVERT=split(donnees$ED,donnees$DECOUVERT)
   EDoui = EDsplitDECOUVERT$"oui"
   EDnon = EDsplitDECOUVERT$"non"
   ```

   (b) Créer deux vecteurs qui contiennent les effectifs des deux vecteurs précédents.

   (c) Associer ces deux tableaux en une seule matrice `X=rbind(vecteur1,vecteur2)` (chaque tableau d'effectifs est en ligne).

   (d) On peut alors tracer le diagramme en barres :

   ```
   barplot(X,beside=TRUE,legend.text=c("Oui","Non")).
   ```

8. On veut découper la variable `AGE` en quatre catégories d'âges : on utilise pour cela la fonction `cut`

   ```
   catAge = cut(donnees$AGE,4)
   ```

   représenter alors le `CREDIT` en fonction de ces quatre catégories. Calculer les paramètres statistiques pour chacune des catégories.

9. Représenter la répartition d'effectifs de découvert oui/non pour les quatre catégories d'âge.
10. Représenter le revenu en fonction du découvert (boxplot).
11. En découpant le revenu en cinq catégories, représenter la variable `CREDIT` en fonction de ces catégories.
12. Représenter la répartition d'effectifs de découvert oui/non pour les cinq catégories de revenu.
13. Construire un tableau `dataOui` qui ne contient que les données du tableau initial pour les clients ayant déjà eu un découvert. Trouver le nombre de clients dans ce cas. Calculer l'âge moyen et l'écart-type de l'âge de ces clients.
14. Reprendre la question précédente pour les clients n'ayant pas eu de découvert.
15. Tracer deux histogrammes représentant l'âge des clients (ayant eu un découvert ou non).
16. Créer un graphique représentant l'âge moyen des clients avec ou sans découvert et ajouter une "barre d'erreur" $\pm$ l'écart-type.