

# Devoir 1

La base de données `visaprem` du paquet `hecmulti` contiennent les profils de 1294 clients d'une institution bancaire française avant la zone euro. Les données ont été collectées lors d'une enquête mensuelle.

1. Transformez la variable catégorielle `sexe` en variable binaire entière avec `homme=0`, `femme=1`, et `NA` pour les valeurs manquantes.
2. Ré-étiquetez les situations familiales (`famiq`) selon que la personne est seule (`seu`) ou en couple (`cou`). Transformez les valeurs manquantes (`inc` pour inconnu) en `NA` (voir `dplyr::na_if`).
3. Éliminez les observations pour lesquelles la variable `age` est manquante.
4. Créez une colonne, `nbsd`, qui représente le nombre total de jours à débit des trois derniers mois. Éliminez les variables utilisées lors de la création de `nbsd`.
5. Considérez le nombre total de cartes `ntcas`. Y a-t-il des incohérences en lien avec les autres variables?
6. Que représentent les variables manquantes résiduelles de `zocnb`? *Indice: voir la question précédente.*
  - Expliquez pourquoi il serait logique de remplacer ces valeurs manquantes par des valeurs numériques (laquelle).
  - Effectuez la modification.
7. Produisez un histogramme de la variable ancienneté du compte (`relat`) avec `ggplot`. Que remarquez-vous?
8. Produisez un nuage de point de `relat` et `age` et commentez (quel est le lien entre `relat` et `age`)?

La plupart des manipulations sont à effectuer directement; votre code et la base de données feront foi de votre travail.

**Indication:** Si vous modifiez une variable, assurez-vous d'écraser la colonne existante (par exemple, la base de données devrait contenir une colonne `sexe` encodée 0L/1L/NA).

Vous devez remettre trois fichiers sur ZoneCours,

- un rapport au format PDF

- votre code **R** ou un fichier Rmarkdown
- la base de données créée à la suite des manipulations, au format CSV (valeurs séparées par des virgules), à l'aide de la commande `write.csv(db, file = "d1_matricule.csv")` où `db` est le nom de votre base de données.

Tous vos fichiers seront nommés selon la convention `d1_matricule.extension` en remplaçant `matricule` par votre numéro d'étudiant(e) et `.extension` par `.pdf`, `.R`, `.Rmd` et `.csv`, selon le type du fichier.

Assurez-vous que vous n'avez pas supprimé la variable `matric`.

**Astuces:** vous pouvez utiliser la fonction `knitr::purl()` pour extraire le code **R** d'un fichier Rmarkdown (extension `.Rmd`).