

3.1 La base de données `biere.sas7bdat` contient de l'information sur 35 types de bières

- `biere` : type de bière (variable de type chaîne de caractères)
  - `cout` : coût pour 12 onces
  - `calories` : nombre de calories
  - `sodium` : sodium en mg.
  - `alcool` : pourcentage d'alcool
  - `classement` : évaluation subjective de la qualité de la bière, soit très bon (1), soit bon (2), soit correct (3).
- (a) Procédez à une analyse descriptive des variables. D'après vous, quelles variables pourraient être utiles pour le regroupement? Déterminez s'il est nécessaire de standardiser les variables.
- (b) Procédez à un regroupement hiérarchique à l'aide de la méthode de Ward et d'une mesure de dissemblance euclidienne. Combien de groupes choisiriez-vous sur la base des critères couverts en classe?
- (c) Répétez l'analyse, en utilisant cette fois la méthode des  $k$ -moyennes. Essayez de varier les germes servant à l'initialisation de l'algorithme (en prenant le barycentre de la solution hiérarchique pour  $k$  groupe, par exemple).
- (d) Variez le nombre de groupe  $k$  et représentez graphiquement le regroupement en projetant les classes sur les deux premières composantes principales. Quel nombre de regroupements serait adéquat sur la base d'une inspection visuelle?
- (e) Interprétez les groupes obtenus à l'aide des méthodes hiérarchiques et non-hiérarchiques et décrivez dans vos mots les différents profils.