

Annexe : SQL et R

<http://revue.sesamath.net/spip.php?article982>

patrick.raffinat@univ-pau.fr

Je reprends ici un extrait de TP proposé à mes étudiants d'IUT.

	A	B	C
1	nom	sexe	age
2	Jim	h	21
3	Eve	f	19
4	Max	h	
5			
6			

	A	B
1	nom	math
2	Jim	12
3	Max	15

- 1) Entrez ces informations dans le tableur de LibreOffice, puis enregistrez le classeur.
- 2) Exportez les deux feuilles de calcul au format « csv » (commande Fichier / Enregistrer Sous, puis choisir le type csv avant d'enregistrer une feuille) :

Nom du fichier :

Type :

Avant de sauvegarder, veillez à choisir une tabulation comme séparateur de champ :

Export de fichiers texte

Options de champ

Jeu de caractères : Europe occidentale (Windows-1252/WinLatin 1)

Séparateur de champ : {Tabulation}

Séparateur de texte : "

- 3) Ouvrez le fichier pers.csv avec l'éditeur de texte Notepad++, ce qui vous permettra de vérifier que c'est un fichier texte et même d'en visualiser les caractères spéciaux (tabulation et fin de ligne)

```

1 nom>sexe ->ageCRLF
2 Jim>h ->21CRLF
3 Eve>f ->19CRLF
4 Max>h ->CRLF
  
```

4) Dans votre environnement R local, exécutez les commandes suivantes, puis visualisez les tableaux de données :

```
setwd("C:/Users/raffinat/Documents/pages/sesamath/article16") # à adapter
etud = read.table("etud.csv", header=TRUE, sep="\t", na.strings="NA", dec=",", strip.white=TRUE)
notes = read.table("notes.csv", header=TRUE, sep="\t", na.strings="NA", dec=",", strip.white=TRUE)
```

5) Exécutez les commandes suivantes :

```
library(sqldf) # pour charger cette librairie
hommes = sqldf("select nom,age from etud where sexe='h' ")
print(hommes)
```

6) Exécutez les requêtes suivantes à l'aide de sqldf :

a) nom, sexe et age des étudiants dont on connaît l'âge

```
> sqldf('select nom, sexe, age from etud where age is not null')
  nom sexe age
1 Jim   h   21
2 Eve   f   19
```

b) âge moyen par sexe

```
> sqldf("select sexe, avg(age) as age_moyen from etud group by sexe")
  sexe age_moyen
1    f         19
2    h         21
```

d) jointure des tables etud et notes , puis nuage de points avec age et note de math

	nom	sexe	age	math
1	Jim	h	21	12
2	Eve	f	19	NA
3	Max	h	NA	15

```
> library(sqldf)
Le chargement a nécessité le package : gsubfn
Le chargement a nécessité le package : proto
Le chargement a nécessité le package : RSQLite
> jointure = sqldf("select etud.nom,sexe,age,math
+                 from etud LEFT OUTER JOIN notes ON etud.nom=notes.nom")
> view(jointure)
> plot(jointure$age, jointure$math)
```