

# Chapitre 5 Manipuler plusieurs tables

Le package *dplyr* possède également plusieurs fonctions permettant de travailler sur deux tables. On va pouvoir regrouper ces fonctions en plusieurs catégories de manipulations :

- pour fusionner des informations de deux tables entre elles : jointures transformantes
- pour sélectionner des observations d'une table en fonction de celles présentes dans une autre table : jointures filtrantes
- pour traiter deux tables ayant les mêmes colonnes et sélectionner sur celles-ci des observations de l'une et l'autre : opérations ensemblistes
- Des manipulations visant à additionner deux tables ensembles : assemblages

**a**

x1	x2
A	1
B	2
C	3

**b**

x1	x3
A	T
B	F
D	T

+

=

**y**

x1	x2
A	1
B	2
C	3

**z**

x1	x2
B	2
C	3
D	4

+

=

### Jointures transformantes

**dplyr::left\_join(a, b, by = "x1")**  
Joindre à a les variables de b selon x1

x1	x2	x3
A	1	T
B	2	F
C	3	NA

**dplyr::right\_join(a, b, by = "x1")**  
Joindre à b les variables de a selon x1

x1	x3	x2
A	T	1
B	F	2
D	T	NA

**dplyr::inner\_join(a, b, by = "x1")**  
Joindre a et b en ne gardant que les observations des deux tableaux

x1	x2	x3
A	1	T
B	2	F

**dplyr::full\_join(a, b, by = "x1")**  
Joindre a et b en gardant toutes les observations

x1	x2	x3
A	1	T
B	2	F
C	3	NA
D	NA	T

### Jointures filtrantes

**dplyr::semi\_join(a, b, by = "x1")**  
Toutes les observations de a ayant des valeurs correspondantes dans b

x1	x2
A	1
B	2

**dplyr::anti\_join(a, b, by = "x1")**  
Toutes les observations de a n'ayant aucune correspondance dans b.

x1	x2
C	3

### Opérations ensemblistes

**dplyr::intersect(y, z)**  
Observations appartenant à y et z

x1	x2
B	2
C	3

**dplyr::union(y, z)**  
Observations appartenant à y et z ou l'un des 2

x1	x2
A	1
B	2
C	3
D	4

**dplyr::setdiff(y, z)**  
Observations appartenant à y et pas à z

x1	x2
A	1

### Assemblages

**dplyr::bind\_rows(y, z)**  
Ajoute z à y comme nouvelles lignes.

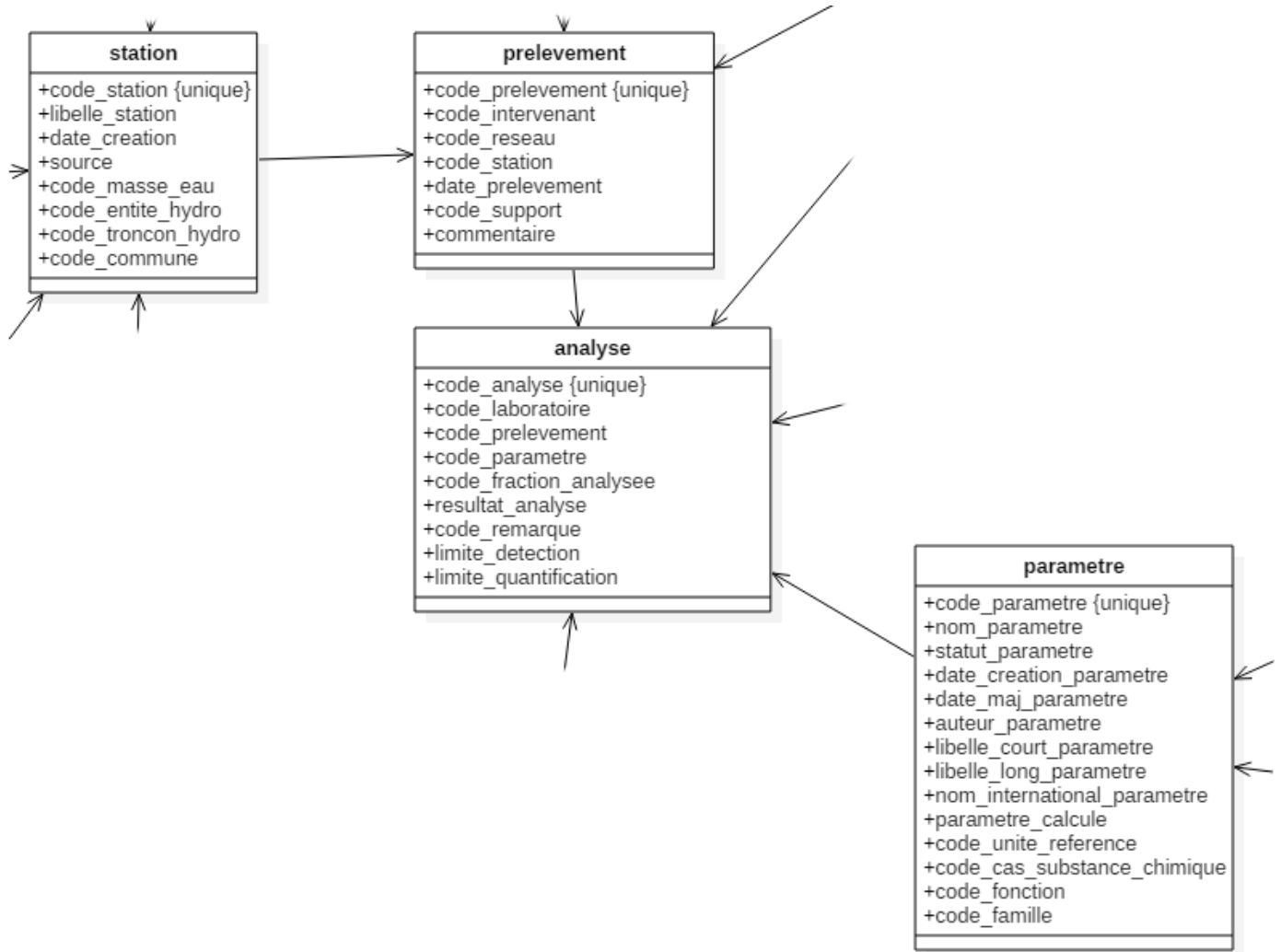
x1	x2
A	1
B	2
C	3
B	2
C	3
D	4

**dplyr::bind\_cols(y, z)**  
Ajoute z à y comme nouvelles colonnes.  
NB: matches rows by position.

x1	x2	x1	x2
A	1	B	2
B	2	C	3
C	3	D	4

## 5.1 Exercices

- reconstituer le dataframe "exercice" à partir des données contenues dans les tables "analyse", "prelevement" et "station" (jointures)



- calculer le nombre d'analyses réalisées sur des molécules (code\_parametre) présentes dans le référentiel
- produire une liste des **code\_parametre** associés à des analyses mais absents du référentiel
- produire une table des analyses "orphelines", c'est-à-dire qui ne correspondent pas à un prélèvement