Opérations sur les données

Nous allons travailler avec une Base de Données BDD dans un fichier unique avec SQLite

Les sources sur Internet

On va télécharger d'abord sur le site de l'INSEE https://www.insee.fr/fr/information/4316069 Le « Code officiel géographique au 1er janvier 2020 » sous forme de fichier csv .

Voici le fichier à télécharger : https://www.insee.fr/fr/statistiques/fichier/4316069/cog_ensemble_2020_csv.zip

Il faut décompresser ce fichier avec 7zip par exemple.

On travaillera uniquement avec 2 fichiers : region2020.csv departement2020.csv

Ces 2 fichiers peuvent être ouverts avec un tableur comme Calc de LibreOffice ou Excel.

Nous allons les insérer dans une base de données BDD au format SQLite.

2 solutions pour travailler avec SQLite:

1) Installer DB Browser for SQLite https://sqlitebrowser.org/dl/ Vous pouvez télécharger ce fichier qui peut être mis dans n'importe quel dossier sans droit administrateur : https://github.com/sqlitebrowser/sqlitebrowser/releases/download/v 3.12.1/DB.Browser.for.SQLite-3.12.1-win64.zip

2) En ligne: https://sqliteonline.com/

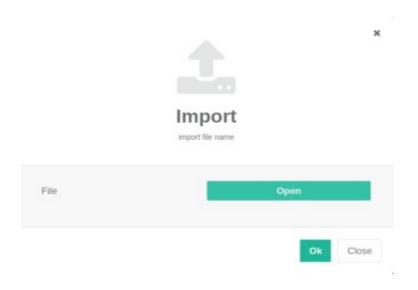
Solution utilisée dans ce diaporama

Cliquer à gauche sur SQLite

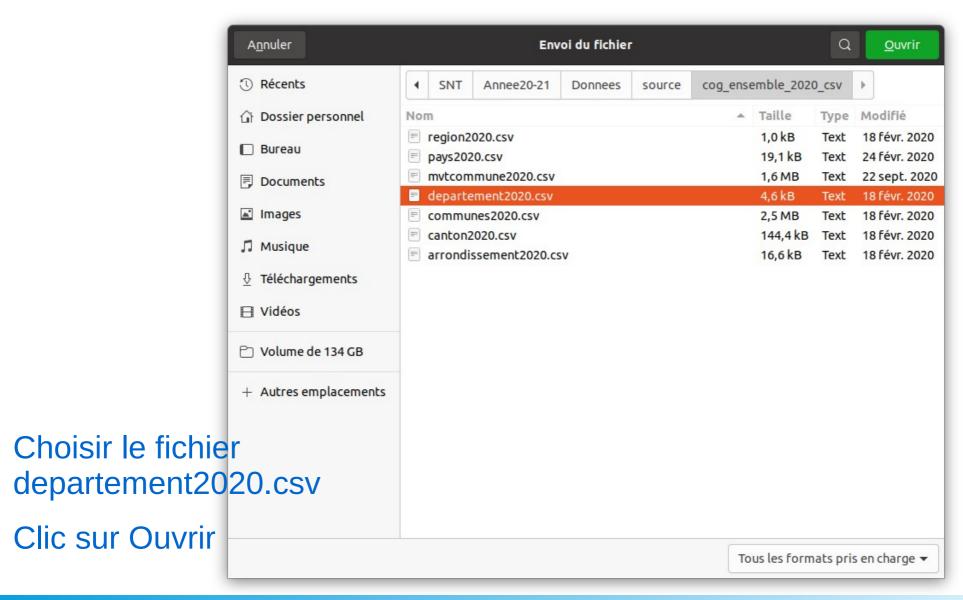
On va importer des tables avec Import

Dans les menus de la ligne du haut, cliquer sur Import

Dans les menus de la ligne du haut, cliquer sur Import



Clic sur Open



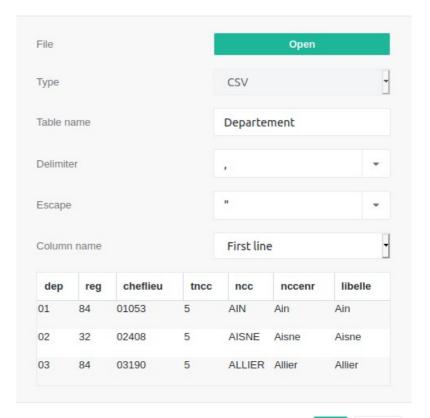
On vérifie que le fichier importer s'appelle departement2020.csv

Important, il faut remplir:

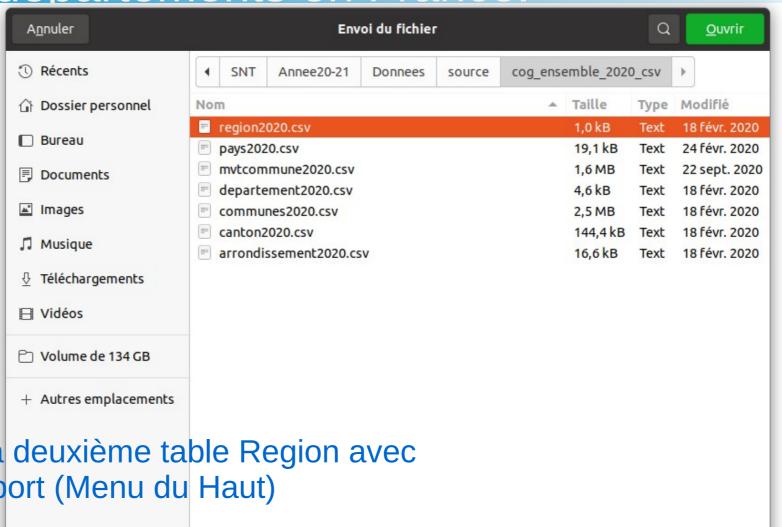
Table name avec **Departement** (sans accent)

Il faut remplacer **New-auto** dans **Column name** par **First line**









On va ajouter la deuxième table Region avec de nouveau Import (Menu du Haut)

Ouvrir le fichier region2020.csv

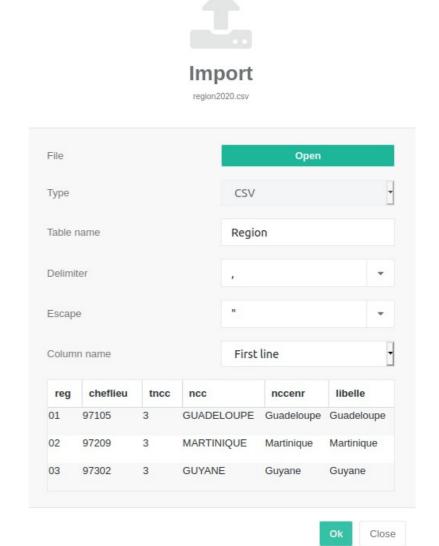
Tous les formats pris en charge ▼

On vérifie que le fichier importer s'appelle region2020.csv

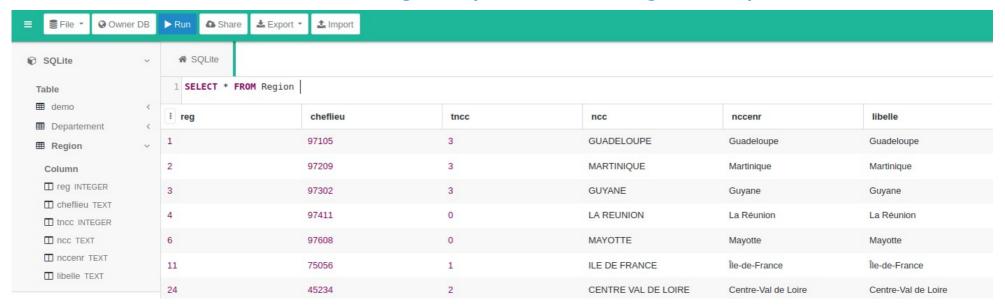
Important, il faut remplir:

Table name avec **Region** (sans accent)

Il faut remplacer **New-auto** dans **Column name** par **First line**



Double-clic sur la table Region (colonne de gauche)



Tous les champs sont affichés, la vue affichée correspond à la requête SQL :

SELECT * FROM Region;

Sélectionner tous les champs (*) depuis la table Région

Si on ne veut afficher que 2 champs, voici un exemple :

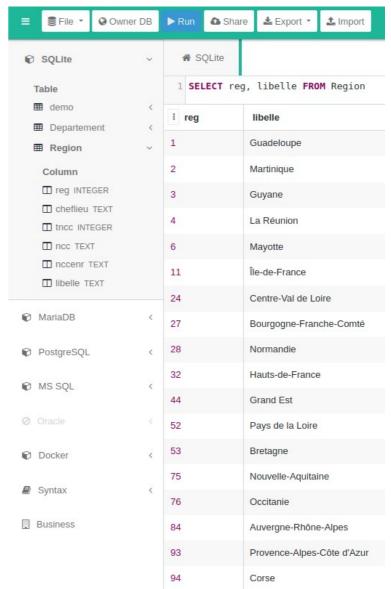
SELECT reg, libelle FROM Region;

SELECT permet donc de choisir les champs que l'on souhaite voir.

FROM indique la table utilisée, ici Region

N'oubliez pas d'exécuter la requête avec ► RUN

Il faut mettre un «; » à la fin de la requête pour être compatible avec d'autres logiciels.



Requêtes SQL

Nous venons de faire notre première requête, mais c'est quoi une requête ?

Juste à titre informatif, voici une requête SELECT qui possède presque toutes les commandes possibles :

SELECT*

FROM table

WHERE condition

GROUP BY expression

HAVING condition

{ UNION | INTERSECT | EXCEPT }

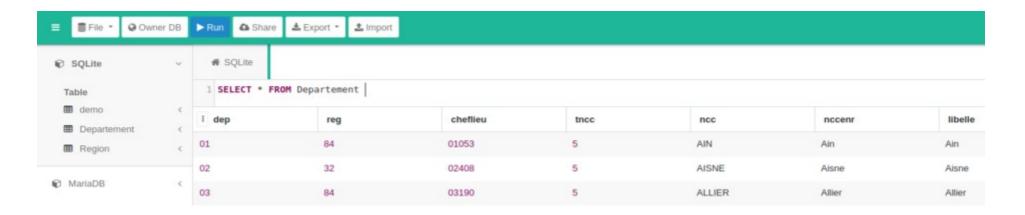
ORDER BY expression

A noter : cette requête imaginaire sert principalement d'aide-mémoire pour retenir l'ordre d'apparition de chacune des commandes au sein d'une requête.

Changeons de table en double-cliquant sur la table Departement à gauche.

Tous les champs sont affichés, la vue affichée correspond à la commande SQL :

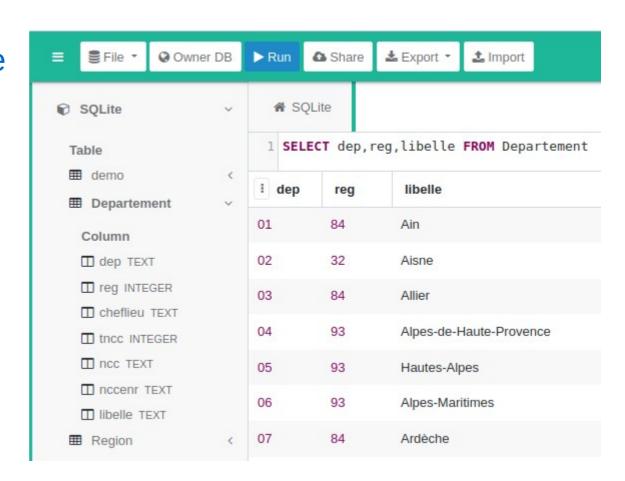
SELECT * FROM Departement;



Si je ne veux afficher que le numéro des départements et le numéro des régions ainsi que le nom des départements

On pourra faire la requête SQL :

SELECT dep,reg,libelle FROM Departement;

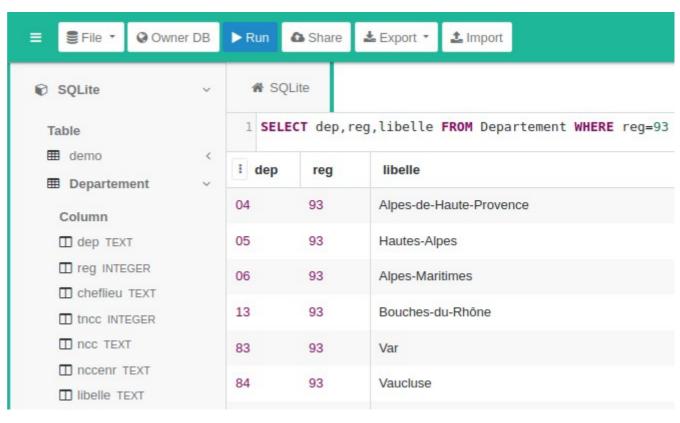


Si je regarde les premiers départements dans la diapositive précédente, je vois que 3 départements sont dans la même région dont le numéro est 93.

Je peux avoir la liste des départements de la région dont le numéro est 93 avec la requête SQL :

SELECT dep,reg,libelle FROM Departement WHERE reg=93

Cette instruction
WHERE + condition
permet de n'afficher que
les lignes qui respectent
cette condition.



Malheureusement, le numéro 93 n'est pas très pratique, quelle est cette région ?

Si je remonte à la requête SQL : SELECT reg, libelle FROM Region ;

L'avant dernière ligne m'indique :

reg 93 libelle Provence-Alpes-Côte d'Azur

Je peux aussi obtenir cette ligne (row) avec la requête SQL:

SELECT reg, libelle FROM Region WHERE reg=93;



Remarque: on peut ajouter des requêtes SQL en cliquant sur le plus à droite, on voit ici que nous avons la requête 2 avec sqlite2.

Comment mettre cette information dans ma requête?

Il faut déjà mettre les 2 tables Departement et Region

On peut mettre les deux requêtes en une avec UNION mais il faut mettre le même nombre de colonnes.

Voici cette requête SQL:

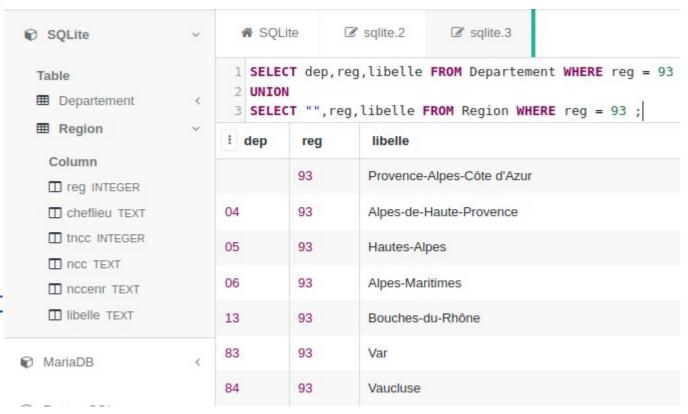
SELECT dep,reg,libelle FROM Departement WHERE reg=93 UNION

SELECT "",reg,libelle FROM Region WHERE reg=93;

Problème les noms des champs sont les mêmes dans les deux tables mais les champs libellé n'ont pas la même signification.

Dans notre cas Le libellé indique tantôt le nom de la région, tantôt le nom du département.

Cette requête est donc inutilisable.



Comme les deux tables ont un champ commun reg nous allons utiliser SQL INNER JOIN https://sql.sh/cours/jointures/inner-join

Intersection de 2 ensembles

Pour utiliser ce type de jointure il convient

d'utiliser une requête SQL avec cette syntaxe :

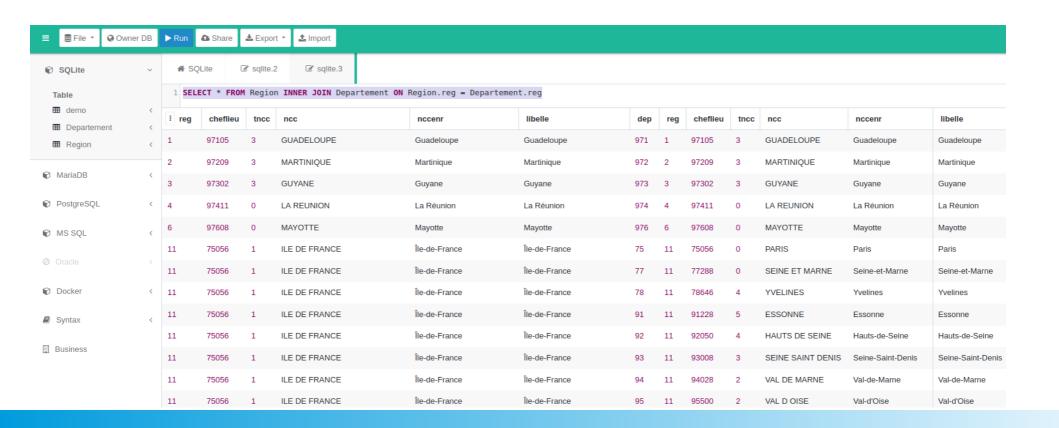


On peut aussi faire cette requête :

SELECT * FROM table1,table2 WHERE table1.id = table2.fk_id;

Voici la requête pour nos 2 tables :

SELECT * FROM Region INNER JOIN Departement ON Region.reg = Departement.reg;



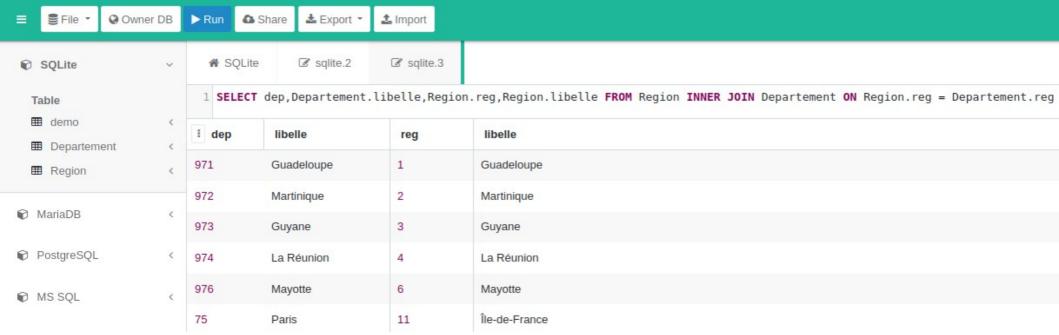
Toujours trop de colonnes et des colonnes avec le même nom, nous allons nous limiter à 4 champs :

le numéro de département le nom du département, le numéro de région et le nom de la région.

Pour afficher le nom du département : champ libelle de département nous écrirons Departement.libelle et pour le nom de la région : champ libelle de région nous écrirons Region.libelle

Voici notre requête :

SELECT dep,Departement.libelle,Region.reg,Region.libelle FROM Region INNER JOIN Departement ON Region.reg = Departement.reg

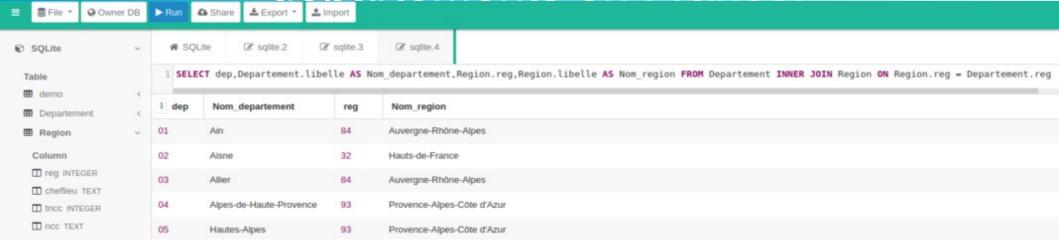


Voici notre requête :

SELECT dep,Departement.libelle,Region.reg,Region.libelle FROM Region INNER JOIN Departement ON Region.reg = Departement.reg;

Il reste un problème, les champs restent les mêmes pour l'affichage.

Nous allons utiliser un alias des champs avec AS pour l'affichage.



Voici la requête :

SELECT dep,Departement.libelle AS Nom_departement, Region.reg, Region.libelle AS Nom_region FROM Departement INNER JOIN Region ON Region.reg = Departement.reg;

Si on veut trier par nom de région on ajoutera : ORDER BY Nom_region

Table demo	<	<pre>SELECT dep,Departement.libelle AS Nom_departement,Region.reg,Region.libelle AS Nom_region PROM Departement INNER JOIN Region ON Region.reg = Departement.reg ORDER BY Nom_region</pre>				
■ Departement ■ Region	·	i dep	Nom_departement	reg	Nom_region	
Column		01	Ain	84	Auvergne-Rhône-Alpes	
☐ reg INTEGER		03	Allier	84	Auvergne-Rhône-Alpes	
☐ cheflieu TEXT ☐ tncc INTEGER		07	Ardèche	84	Auvergne-Rhône-Alpes	
□ ncc TEXT		15	Cantal	84	Auvergne-Rhône-Alpes	
nccenr TEXT		26	Drôme	84	Auvergne-Rhône-Alpes	
☐ libelle TEXT		38	Isère	84	Auvergne-Rhône-Alpes	
	<	42	Loire	84	Auvergne-Rhône-Alpes	
	<	43	Haute-Loire	84	Auvergne-Rhône-Alpes	
S MC COL	<	63	Puy-de-Dôme	84	Auvergne-Rhône-Alpes	
		69	Rhône	84	Auvergne-Rhône-Alpes	
⊘ Oracle		73	Savoie	84	Auvergne-Rhône-Alpes	
	<	74	Haute-Savoie	84	Auvergne-Rhône-Alpes	
		21	Côte-d'Or	27	Bourgogne-Franche-Comté	
Svntax	(

Comme on le voit le tri se fait par région, on commence par les départements de la région Auvergne-Rhône-Alpes, ainsi de suite ...

Si maintenant, je ne veux afficher que les départements d'une région, je pourrais utiliser WHERE pour choisir la Bretagne par exemple. SELECT dep,Departement.libelle AS Nom_departement, Region.reg, Region.libelle as Nom region

FROM Departement INNER JOIN Region

ON Region.reg = Departement.reg

WHERE
Nom_region =
'Bretagne'
ORDER BY
Nom_region;

2 FROM I	<pre>dep,Departement.libelle AS Nom_departement,Region.reg,Region.libelle AS Nom_region epartement INNER JOIN Region ON Region.reg = Departement.reg Nom_region = 'Bretagne' BY Nom_region</pre>			
dep	Nom_departement	reg	Nom_region	
22	Côtes-d'Armor	53	Bretagne	
29	Finistère	53	Bretagne	
35	Ille-et-Vilaine	53	Bretagne	
56	Morbihan	53	Bretagne	

Voici comment à partir de deux tables distinctes qui ont un champ commun reg (numéro de région) nous avons pu retrouver les départements qui font partie d'une région.

Que faut-il faire si on veut afficher les départements de la région Occitanie ?

• Requête pour obtenir les départements de la région Occitanie

Nous pouvons aussi compter les départements d'une région avec Count(*) qui compte le nombre de lignes par région avec le GROUP BY Nom_region

Voici la requête :

SELECT Region.libelle as Nom_region,COUNT(*) AS nombre_departement FROM Departement INNER JOIN Region ON Region.reg = Departement.reg GROUP BY Nom_region

Table ⊞ demo ⊞ Departement	< <	<pre>SELECT Region.libelle AS Nom_region,COUNT * AS nombre_departement FROM Departement INNER JOIN Region ON Region.reg = Departement.reg GROUP BY Nom_regionORDER BY Nom_region</pre>		
■ Region	~	Nom_region	nombre_departement	
Column		Auvergne-Rhône-Alpes	12	
☐ reg INTEGER ☐ cheflieu TEXT		Bourgogne-Franche-Comté	8	
☐ tncc INTEGER		Bretagne	4 6	
ncc TEXT		Centre-Val de Loire		
□ libelle TEXT		Corse	2	
	<	Grand Est	10	
Manabb		Guadeloupe	1	
	<	Guyane	1	
	<	Hauts-de-France	5	
⊘ Oracle		La Réunion	1	
O Gradio		Martinique	1	
Docker Docker	<	Mayotte	1	
Syntax	<	Normandie	5	
Business		Nouvelle-Aquitaine	12	
pg Caomicoo		Occitanie	13	
		Pays de la Loire	5	
		Provence-Alpes-Côte d'Azur	6	
		Île-de-France	8	

On peut savoir combien de régions ont moins de 6 départements avec HAVING condition

Voici la requête commentée :

SELECT Region.libelle,Count(Departement.libelle) as 'nb_dep' FROM Departement

INNER JOIN Region ON Departement.reg = Region.reg --intersection des 2 tables Departement et Region si et seulement si les numéros de la région sont égaux

Group By Region.reg -- regroupe tous les lignes dont le n° de région est identique -- l'instruction Count(Departement.libelle) compte le nombre des départements pour chacun des numéros de régions

HAVING Count(Departement.libelle) < 6 -- les régions qui ont moins de 6 départements

ORDER BY Region.libelle;-- on trie de manière croissante avec le nom de la région

Bretagne	4
Corse	2
Guadeloupe	1
Guyane	1
Hauts-de-France	5
La Réunion	1
Martinique	1
Mayotte	1
Normandie	5
Pays de la Loire	5

Pour finir voici une requête plus complexe avec 2 SELECT :

SELECT Departement.libelle,Departement.dep, Departement.reg, Region.libelle, nb dep

FROM (SELECT COUNT(Departement.dep) AS 'nb_dep', Region.libelle AS 'lib' FROM Departement INNER JOIN Region ON Departement.reg = Region.reg GROUP BY Region.libelle) AS 'count_dep'

INNER Join (Departement INNER JOIN Region ON Departement.reg = Region.reg)

ON Region.libelle = count_dep.lib

ORDER BY Region.libelle;

```
SELECT Departement.libelle ,Departement.dep,Departement.reg,Region.libelle, nb_dep

FROM

(SELECT COUNT(Departement.dep) AS 'nb_dep', Region.libelle AS 'lib' FROM Departement INNER JOIN Region ON Departement.reg = Region.reg GROUP BY Region.libelle) AS 'count_dep'

INNER JOIN (Departement INNER JOIN Region ON Departement.reg = Region.reg)

ON Region.libelle = count_dep.lib

ORDER BY Region.libelle;
```

i libelle	dep	reg	libelle:1	nb_dep
Ain	01	84	Auvergne-Rhône-Alpes	12
Allier	03	84	Auvergne-Rhône-Alpes	12
Ardèche	07	84	Auvergne-Rhône-Alpes	12
Cantal	15	84	Auvergne-Rhône-Alpes	12
Drôme	26	84	Auvergne-Rhône-Alpes	12
Isère	38	84	Auvergne-Rhône-Alpes	12
Loire	42	84	Auvergne-Rhône-Alpes	12
Haute-Loire	43	84	Auvergne-Rhône-Alpes	12
Puy-de-Dôme	63	84	Auvergne-Rhône-Alpes	12
Rhône	69	84	Auvergne-Rhône-Alpes	12
Savoie	73	84	Auvergne-Rhône-Alpes	12
Haute-Savoie	74	84	Auvergne-Rhône-Alpes	12
Côte-d'Or	21	27	Bourgogne-Franche-Comté	8