

Formation au Ministère de l'Agriculture

Les fondamentaux de la programmation

Sommaire

- 1 Algorithmique et Programmation
- 2 Variables et Types
- 3 Les Tests
- 4 Les Boucles
- 5 Les Fonctions
- 6 Principales opérations sur les données et/ou tableaux

0.1 Avant-propos

Ce diaporama de formation a été rédigé dans le but d'être le support visuel des formations dispensées au **SSM Agriculture**.

Cette formation s'adresse à des agents souhaitant découvrir l'écriture de syntaxe avec R sans toutefois avoir déjà écrit des programmes dans d'autres langages.

0.2 Avant-propos

La formation est donnée en distancielle sur une durée **d'une journée**.

Objectifs par cette formation

Ce support permet de :

- Connaître les principes fondamentaux de la programmation
- Avoir les pré-requis nécessaires à la formation « Initiation à la programmation R »

Pour information, les thèmes abordés sont:

- Algorithmes et programmes
- Les variables : nommage et typage
- Les différentes instructions : affectation, lecture, écriture
- Les tests de conditionnalité
- Les fonctions : utilité, arguments, résultats

~~Ils sont orientés pour être utile aux agents du SSM Agriculture et se concentrent sur une utilisation de R via [RStudio](#) qui est mise à disposition des agents sur la plateforme interne Cerise basée sur RStudio Workbench.~~



MINISTÈRE
DE L'AGRICULTURE
ET DE LA SOUVERAINETÉ
ALIMENTAIRE

*Liberté
Égalité
Fraternité*

1 Algorithmique et Programmation



1.1 Etapes de développement d'un logiciel ou programme

ANALYSER Comprendre les besoins du client.

- Dossier d'analyse
- Cahier des charges

CONCEVOIR Identifier tout ce qui va être utile pour répondre au besoin du client. => **Algorithmique**

- Documents de spécifications détaillées

REALISER Implémenter la solution choisie pour répondre au besoin du client. => **Programmation**

- Programmes prêts à exécuter



1.2 Définitions

ALGORITHME - Représentatif d'une logique -
Indépendant de tout langage de programmation -
Inexploitable en l'état par un ordinateur -
Compréhensible par tous - Suite d'instructions ou
d'étapes - Plusieurs représentations : schémas, texte
(pseudo-code)

PROGRAMME - Traduction d'un algorithme dans un
langage de programmation spécifique (lignes de code) -
Exécutable par le système

1.3 Analogie

Analogie avec la littérature dans le cadre d'un livre traduit en plusieurs langues

- L'écrivain réalise l'algorithme (imagination du récit, choix des personnages, découpage en chapitres....) et écrit le programme dans une langue donnée
- Les traducteurs transposent l'œuvre de l'écrivain dans d'autres langues en essayant d'être le plus fidèles possibles à l'histoire et aux émotions qu'il a voulu transmettre.



1.4 Structure d'un algorithme en pseudo-code

- Choix du nom de l'algorithme
- Déclaration des variables manipulées
- Initialisation des variables et liste des instructions

ALGORITHME < Nom >

VARIABLES

< Liste des variables >

DEBUT

< Liste des instructions >

FIN

1.5 Structure d'un algorithme en pseudo-code

Plusieurs variantes correctes du bloc d'instructions existent.

- Le choix des instructions tient compte de plusieurs facteurs :
 - optimisation,
 - lisibilité et compréhension,
 - possibilité de traduction dans plusieurs langages différents
- Le pool d'instructions disponibles peut être lié au langage de programmation mais aussi à la communauté de développeurs (bibliothèques mises à disposition....).

Exercices 1 à 3 (manipulation des instructions)

2 Variables et Types



2.1 Définitions

- Une **variable** est un symbole qui porte un nom et qui contient une valeur d'un type donné.

Si la valeur de la variable n'est pas modifiable dans le temps, le terme de **constante** peut être utilisé.

- **Déclaration** : indiquer l'existence d'une variable et préciser son type

`<nom_variable> : <nom_type>`

- **Initialisation** : donner une valeur initiale à une variable
- **Affectation/assignation** : donner une valeur à une variable à tout moment. La nouvelle valeur affectée à une variable écrase l'ancienne valeur.

`<nom_variable> <- <valeur>`

Exercices 4 (manipulation d'une variable)

2.2 Principaux types

Type	Utilité	Exemple
Booléen	Représente un état binaire, vrai ou faux	Vrai
Numérique	Représente un nombre quelconque	12345.6789
Caractère	Représente un caractère unique (comme une touche du clavier)	'c'
Chaîne de caractères	Représente un texte	"contenu de la chaîne"

Il est possible de ventiler la catégorie « **Numérique** » en 2 catégories distinctes :

- Les entiers
- Les réels

La famille des « **dates** » peut également être considérée comme un type de base pour les variables.

2.3 Opérations possibles

Pour les **numériques** : toutes les **opérations mathématiques** classiques

- + : addition
- - : soustraction
- * : multiplication
- / : division

Pour les **chaînes de caractères** :

- + : concaténation (assembler des caractères à la suite des autres)

Chaque langage de programmation décline ces types de base en plus ou moins de types de données propres. Le type choisi au moment de l'écriture du programme aura un impact sur l'allocation mémoire qui sera faite pour le stockage de la variable mais aussi sur la liste des fonctions utilisables.

Exercices 5 (manipulation de variables de types différents)

2.4 La valeur **NULL** (1/2)

Le mot clé **NULL** est utilisé pour indiquer que la valeur d'une variable est inconnue.

Remarque :

- Pour des numériques : **NULL** est différent de 0
- Pour des chaînes de caractères : **NULL** est différent de « »
- Pour des booléens : **NULL** est différent de FAUX

Des opérations simples peuvent entraîner des erreurs (plantages) si des variables ne sont pas valorisées.

Exemple : addition de 2 numériques dont l'un a pour valeur **NULL**.

2.5 La valeur **NULL** (2/2)

Pour éviter ce type d'erreur :

- Initialiser le plus souvent possible les variables en début de programme
- Si le langage le permet, utiliser la notion de valeur par défaut lors de la déclaration des variables
- Si le langage le permet, utiliser une fonction permettant de remplacer toutes les valeurs NULL par une valeur passée en paramètre

Dans certains cas, une valeur vide peut être significative. Un mot clé spécifique au langage est utilisé.

Dans ce cas, la notion de « non applicable » ou « valeur indéfinie » est souvent employée.

2.6 Les tableaux

Un **tableau** est une variable qui peut contenir plusieurs valeurs simultanément.

Un tableau se présente comme une structure organisée sous forme de cellules ou cases qui peuvent être identifiées par un indice ou index.

Selon les langages de programmation, certaines règles peuvent différer :

- Un tableau peut ne contenir que des variables du même type
- La valeur de l'indice débute à 0 ou à 1

Un tableau peut avoir 1 ou plusieurs dimensions.

- 1 dimension : vecteur ou liste
- Plusieurs dimensions : matrice ou tableau multi-dimensionnel

2.7 Les tableaux

Les tableaux les plus fréquemment exploités au SSM Agriculture sont des **tables de données**.

Prénom	Sexe	Taille
Pierre	H	185
Anne	F	167
Marie	F	171
Charles	H	175

Caractères Caractères Numériques

Exemple d'un support de cours « Initiation à R »

Tableau TAB_IND

 = <nom_tableau>[num_ligne,num_colonne] = TAB_IND[3,2]

 = <nom_tableau>\$<nom_colonne>[num_ligne] = TAB_IND\$Sexe[3]

Une **colonne** ne peut contenir que des valeurs de même type.

Colonne = variable = vecteur

Toutes les colonnes possèdent le même nombre de valeurs. Un nom peut être attribué à chaque colonne. Ce nom de colonne peut être utilisé pour accéder à la valeur de la cellule.

Une **ligne** contient les valeurs observées pour chaque variable pour un élément donné. Ligne = observation = occurrence = élément

Les **indices ou index** de lignes et de colonnes débutent à 1.



MINISTÈRE
DE L'AGRICULTURE
ET DE LA SOUVERAINETÉ
ALIMENTAIRE

*Liberté
Égalité
Fraternité*

3 Les Tests



3.1 La condition - “SI”

```
SI < condition > = VRAI  
ALORS  
    < liste d'instructions >  
SINON  
    < liste d'instructions >  
FIN SI
```

- La condition est une expression booléenne qui ne peut avoir que 2 valeurs possibles : VRAI ou FAUX.
- Le bloc d'instructions SINON est facultatif.
- Plusieurs conditions SI peuvent être imbriquées.

Analogie avec Excel :

SI(test_logique;valeur_si_vrai;valeur_si_faux)
Vérifie si la condition est respectée et renvoie une valeur si le résultat d'une condition que vous avez spécifiée est VRAI, et une autre valeur si le résultat est FAUX.

3.2 La condition - “SI”

La condition exprimée dans un bloc SI peut correspondre à toute expression pouvant être valorisée par **VRAI** ou **FAUX**.

Cette expression peut être unitaire ou combinée par des opérateurs logiques (ET et OU). Les parenthèses sont utilisées pour encadrer les différentes conditions et gérer les priorités.

Exemples de conditions courantes :

- est égal à , différent de, supérieur, inférieur
- contient, commence par ...
- est vide, existe ...

Exemples de conditions combinées :

- Condition A ou Condition B
 - (Condition A et Condition B) ou Condition C
-

3.3 Exemples

```
ZoneOutreMer <- FAUX
SI CodeDepartement possède 3 caractères
ALORS
    ZoneOutreMer <- VRAI
FIN SI
```

```
SI MoyenneGenerale >= 10
ALORS
    ResultatExamen <- « Reçu »
SINON
    ResultatExamen <- « Ajourné »
FIN SI
```

```
SI MoyenneGenerale >= 10
ALORS
    ResultatExamen <- « Reçu »
    Mention <- « »
    SI MoyenneGenerale >= 16
    ALORS
        Mention <- « Très Bien »
    SINON
        SI MoyenneGenerale >= 14
        ALORS
            Mention <- « Bien »
        SINON
            SI MoyenneGenerale >= 12
            ALORS
                Mention <- « Assez Bien »
            FIN SI
        FIN SI
    FIN SI
SINON
    ResultatExamen <- « Ajourné »
    Mention <- « »
FIN SI
```

3.4 Les opérateurs logiques

La valeur d'une condition combinée dépend des valeurs de chaque condition unitaire utilisée, des opérateurs logiques et des priorités données par les parenthèses.



3.5 L'opérateur "OU"

« Pour que la condition combinée soit vraie, au moins une des 2 conditions doit être vraie ».

Condition A	Condition B	A OU B
VRAI	VRAI	VRAI
VRAI	FAUX	VRAI
FAUX	VRAI	VRAI
FAUX	FAUX	FAUX

3.6 L'opérateur "ET"

« Pour que la condition combinée soit vraie, les 2 conditions doivent être vraies »

Condition A	Condition B	A ET B
VRAI	VRAI	VRAI
VRAI	FAUX	FAUX
FAUX	VRAI	FAUX
FAUX	FAUX	FAUX

En l'absence de parenthèses, l'opérateur **ET** est prioritaire sur l'opérateur **OU**.

3.7 L'opérateur "OU EXCLUSIF"

« Pour que la condition combinée soit vraie, au moins une des 2 conditions mais pas les 2 doit être vraie ».

Condition A	Condition B	A OU exclusif B
VRAI	VRAI	FAUX
VRAI	FAUX	VRAI
FAUX	VRAI	VRAI
FAUX	FAUX	FAUX

Illustration fonctionnelle :

Cas d'un menu proposant une formule « Plat + entrée ou dessert » à 20€

- Si le client prend un plat et une entrée, il peut bénéficier de la formule
- Si le client prend un plat et un dessert, il peut bénéficier de la formule
- Si le client prend un plat, une entrée et un dessert, il ne peut pas bénéficier de la formule.

3.8 Les Tests - “SELON”

SELON < valeur de la variable testée > **FAIRE**

CAS < liste 1 de valeurs > : < liste d'instructions >

CAS < liste 2 de valeurs > : < liste d'instructions >

CAS < liste N de valeurs > : < liste d'instructions >

AUTREMENT : < liste d'instructions >

FIN SELON

La liste de valeurs peut contenir une seule valeur, une suite de valeurs séparées par des virgules et/ou un intervalle de valeurs.

Exercices 7 et 8 (utilisation du test SELON)

3.9 Exemples

```

SI MoyenneGenerale >= 10
ALORS
    ResultatExamen <- « Reçu »
    Mention <- « »
    SI MoyenneGenerale >= 16
    ALORS
        Mention <- « Très Bien »
    SINON
        SI MoyenneGenerale >= 14
        ALORS
            Mention <- « Bien »
        SINON
            SI MoyenneGenerale >= 12
            ALORS
                Mention <- « Assez Bien »
            FIN SI
        FIN SI
    FIN SI
SINON
    ResultatExamen <- « Ajourné »
    Mention <- « »
FIN SI

```

```

Mention <- « »

SI MoyenneGenerale < 10
ALORS
    ResultatExamen <- « Ajourné »
SINON
    ResultatExamen <- « Reçu »
    SELON MoyenneGenerale FAIRE
        CAS [12-14[ : Mention <- « Assez Bien »
        CAS [14-16[ : Mention <- « Bien »
        CAS >= 16 : Mention <- « Très Bien »
    FIN SELON
FIN SI

```



MINISTÈRE
DE L'AGRICULTURE
ET DE LA SOUVERAINETÉ
ALIMENTAIRE

*Liberté
Égalité
Fraternité*

4 Les Boucles



4.1 Définition

Une **boucle** permet de répéter plusieurs fois un bloc d'instructions.

3 types de boucles coexistent :

- Boucle « **POUR** » : elle est utilisée lorsque le nombre de répétitions est connu à l'avance
- Boucle « **TANT QUE** » : elle est utilisée lorsque le nombre de répétitions n'est pas connu à l'avance et permet de répéter l'exécution du bloc d'instructions tant que la condition de bouclage est respectée
- Boucle « **REPETER** » ou « **FAIRE TANT QUE** » : variante de la boucle « TANT QUE » où le bloc d'instructions est exécuté avant le test de la bouclage.

Il est obligatoire de s'assurer que la condition de bouclage deviendra FAUSSE à un moment donné pour provoquer l'arrêt de l'exécution du bloc d'instructions. Dans le cas contraire, le programme ne s'arrêtera jamais (boucle infinie).

4.2 Les boucles POUR

POUR <variable> ALLANT DE <debut> A <fin>

FAIRE < liste d'instructions >

FIN POUR

Illustration :

Boucle permettant d'afficher le libellé des 12 mois de l'année.

```
1 VARIABLES :  
2     indice_boucle : ENTIER  
3     libelle_mois : CHAINE de CARACTERES  
4 DEBUT  
5     POUR indice_boucle ALLANT DE 1 A 12  
6     FAIRE  
7         libelle_mois <- donne_libelle_du_mois(indice_boucle)  
8         Afficher libelle_mois  
9     FIN POUR  
10 FIN
```

4.3 Les boucles TANT QUE

TANT QUE = VRAI

FAIRE < liste d'instructions >

FIN TANT QUE

Illustration :

Boucle permettant d'afficher le libellé des 12 mois de l'année.

```
VARIABLES :  
    indice_boucle : ENTIER  
    libelle_mois : CHAINE de CARACTERES  
DEBUT  
    Indice_boucle <- 1  
    TANT QUE indice_boucle <= 12  
    FAIRE  
        libelle_mois <- donne_libelle_du_mois(indice_boucle)  
        Afficher libelle_mois  
        indice_boucle <- indice_boucle + 1  
    FIN TANT QUE  
FIN
```


4.4 Les boucles REPETER

REPETER

< liste d'instructions >

JUSQU'À = VRAI

FIN REPETER

Illustration :

Boucle permettant d'afficher le libellé des 12 mois de l'année.

```
1 VARIABLES :
2     indice_boucle : ENTIER
3     libelle_mois : CHAINE de CARACTERES
4 DEBUT
5     Indice_boucle <- 1
6     REPETER
7         libelle_mois <- donne_libelle_du_mois(indice_boucle)
8         Afficher libelle_mois
9         indice_boucle <- indice_boucle + 1
10        JUSQU'À indice_boucle = 13
11    FIN REPETER
12 FIN
```



MINISTÈRE
DE L'AGRICULTURE
ET DE LA SOUVERAINETÉ
ALIMENTAIRE

*Liberté
Égalité
Fraternité*

5 Les Fonctions



5.1 Définition (1/2)

Une **fonction** correspond à un regroupement d'instructions dans le but de réaliser une action précise et qui a vocation à être réutilisé.

Chaque langage de programmation propose un pool de fonctions de base plus ou moins riche qui peuvent être regroupées ou non dans des librairies.

Tout développeur peut utiliser des fonctions déjà écrites ou créer des fonctions personnalisées pour des traitements répondant à ses besoins.

Une fonction :

- Possède un nom
- Peut nécessiter des données en entrée (paramètres)
- Peut fournir une valeur en retour (paramètre de sortie ou résultat)

5.2 Définition (2/2)

Toutes ces informations sont mentionnées dans la 1er ligne du script qui se nomme la **signature**.

Plusieurs fonctions portant le même nom peuvent coexister. Dans la majorité des cas, elles possèdent des signatures différentes.

Exemple :

La fonction `isNull(<variable>)` qui retourne un booléen. Cette fonction a été implémentée dans de nombreux langages pour des paramètres d'entrée de plusieurs types (entier, chaînes de caractères)



5.3 Syntaxe

FONCTION <type du résultat> <nom de la fonction> (<type du param1> <nom du param1, ...>)

< liste d'instructions >

RETOURNER <valeur du résultat>

FIN

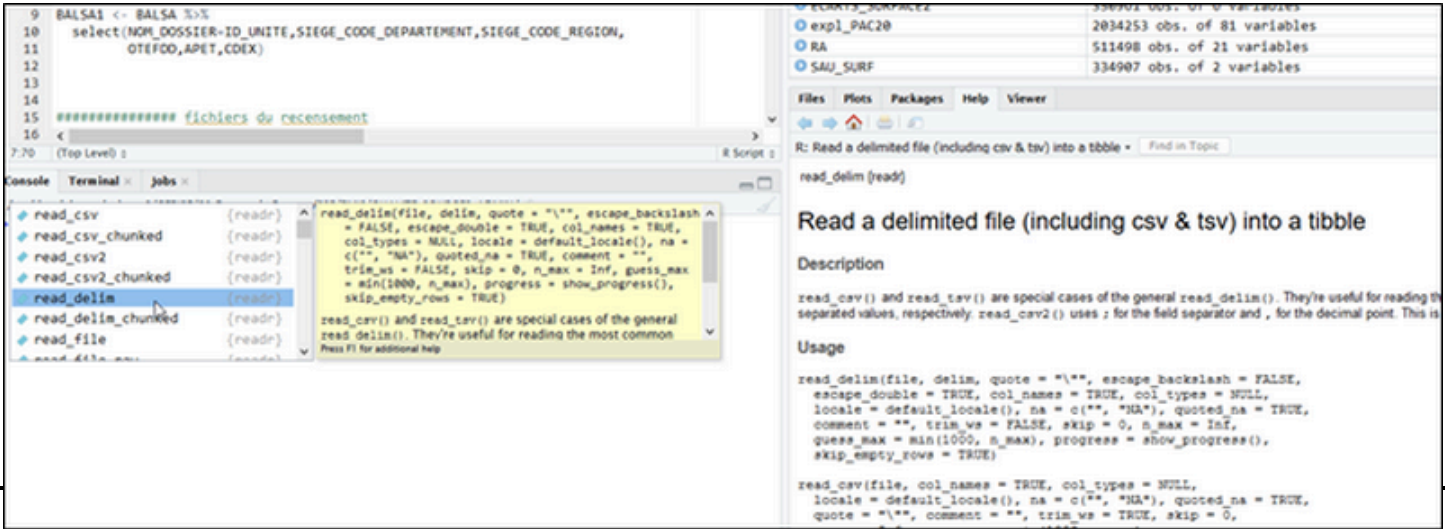
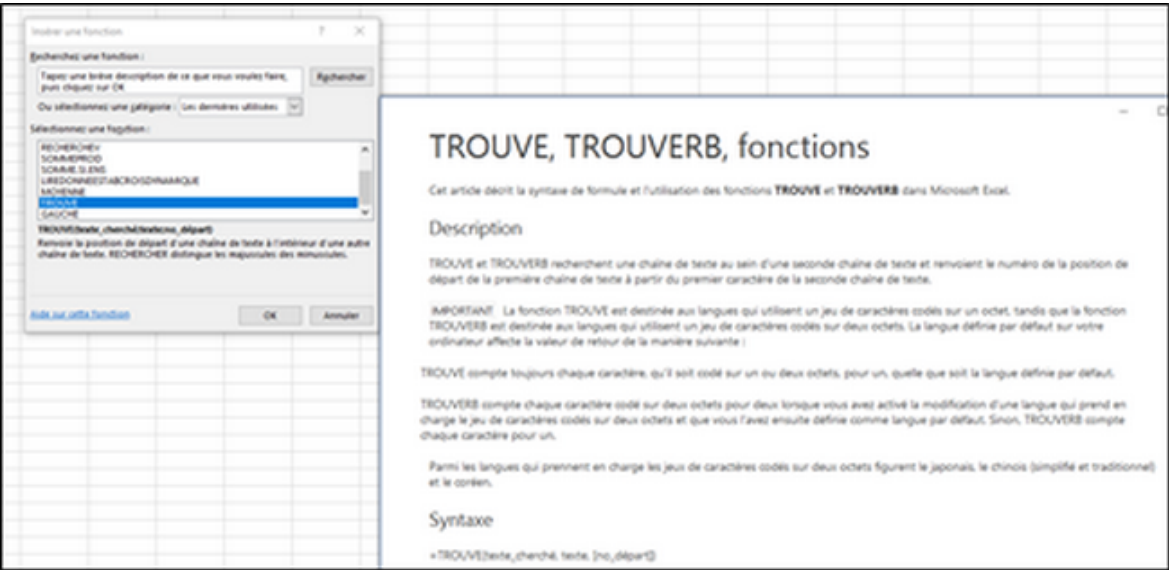
Une fonction est dite **récursive** lorsqu'elle s'appelle elle-même. Dans ce cas, son code doit forcément comprendre une condition d'arrêt.

L'utilisation de fonctions permet souvent de décomposer un traitement complexe en plusieurs traitements plus simples. La compréhension et la maintenance des programmes sont alors facilitées.

Toute fonction doit être **documentée** pour permettre sa réutilisation ultérieure. En général, l'espace de développement donne la possibilité d'accéder à la description de la fonction et éventuellement à des exemples d'utilisation. Les sites web de communautés de développeurs sont aussi très riches en informations.

5.4 Usage

Avant de se lancer dans l’écriture d’une fonction, il faut au préalable rechercher si elle n’existe pas déjà.



6 Principales opérations sur les données et/ou tableaux

6.1 Importer des données à partir d'un fichier

Objectif : charger le contenu d'un fichier de données dans un tableau manipulable dans l'environnement de traitement des données.

Exemples :

- charger dans CERISE le contenu d'un fichier CSV
- ouvrir un fichier Excel

Paramètres possibles :

- emplacement et nom du fichier
- encodage du fichier (UTF8, ASCII...)
- présence d'entêtes avec le nom des variables/colonnes
- type de certaines variables/colonnes
- nom de la variable de type tableau initialisée avec les données chargées

6.2 Sélectionner des variables/colonnes

Objectif : conserver uniquement certaines variables/colonnes d'un tableau de données afin de se concentrer sur les informations utiles ou pertinentes.

Paramètres nécessaires :

- nom du tableau concerné
- nom ou numéro d'ordre des variables/colonnes sélectionnées

Remarques :

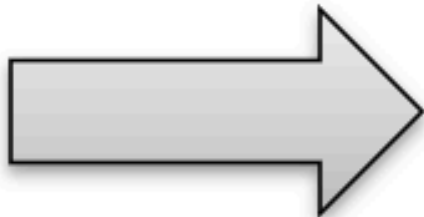
- l'opération de sélection de variables/colonnes n'influence pas le nombre de lignes du tableau
- la réduction du volume des données induite par cette opération améliore la performance des traitements
- certains langages, dont R, permettent de préciser les colonnes à enlever au lieu de celles à conserver (« sélectionner toutes les colonnes sauf »)

6.3 Sélectionner des variables/colonnes

Sélectionner des variables/colonnes

ASSOLEMENT

Exploitant	Département	Parcelle	Surface (ha)	Culture
EXPL1	81	1	13,33	BTH
EXPL2	81	1	5,27	ORH
EXPL1	11	1	11,50	PPT
EXPL2	2A	2	2,70	BTH
EXPL3	29	2	15,20	MIS



Exploitant	Culture
EXPL1	BTH
EXPL2	ORH
EXPL1	PPT
EXPL2	BTH
EXPL3	MIS

Sélection des variables « Exploitant » et « Culture » du tableau ASSOLEMENT

6.4 Renommer des variables/colonnes

Objectif : modifier les noms des variables/colonnes d'un tableau afin de les rendre plus compréhensibles, clairs, homogènes ou adaptés à l'analyse.

Paramètres nécessaires :

- nom du tableau concerné
- nom ou numéro d'ordre de la variable/colonne à modifier
- nouveau nom de la variable/colonne

Remarques :

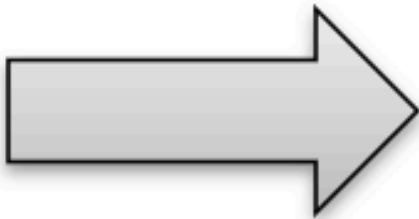
- certains environnements de traitement des données ou langages peuvent être restrictifs pour les intitulés de variables/colonnes (présence de caractères spéciaux ou d'accents, nombre de caractères, mots-clés interdits ...)
- certaines structures adoptent des règles de nommage communes pour faciliter la compréhension et la maintenance de leurs projets ou traitements. Si elles existent, elles doivent être respectées.

6.5 Renommer des variables/colonnes

Renommer des variables/colonnes

REGIONS

Code	Libellé
91	Languedoc-Roussillon
73	Midi-Pyrénées
76	Occitanie
53	Bretagne
24	Centre Val de Loire
94	Corse



REGIONS

Code Région	Libellé
91	Languedoc-Roussillon
73	Midi-Pyrénées
76	Occitanie
53	Bretagne
24	Centre Val de Loire
94	Corse

Renommage de la variable « Code » du tableau REGION en « Code Région »

6.6 Filtrer des données

Objectif : conserver uniquement certaines lignes d'un tableau de données en fonction d'un critère défini afin de se concentrer sur les informations utiles ou pertinentes.

Paramètres nécessaires :

- nom du tableau concerné
- critère de filtre reposant sur une ou plusieurs variables/colonnes du tableau

Remarques :

- l'opération de filtrage des données n'influence pas le nombre de variables/colonnes du tableau
- la réduction du volume des données induite par cette opération améliore la performance des traitements

6.7 Filtrer des données

Filtrer des données

DEPARTEMENTS

Code	Libellé	Région	Début de validité	Fin de validité
11	Aude	91	01/01/1956	31/12/2015
11	Aude	76	01/01/2016	
2A	Corse du Sud	94	01/01/1956	
29	Finistère	53	01/01/1956	
56	Morbihan	53	01/01/1956	
81	Tarn	73	01/01/1956	31/12/2015
81	Tarn	76	01/01/2016	



DEPARTEMENTS

Code	Libellé	Région	Début de validité	Fin de validité
11	Aude	76	01/01/2016	
2A	Corse du Sud	94	01/01/1956	
29	Finistère	53	01/01/1956	
56	Morbihan	53	01/01/1956	
81	Tarn	76	01/01/2016	

Suppression des lignes n'étant plus valides (date de fin de validité renseignée) dans le tableau DEPARTEMENTS

6.8 Filtrer des données

ASSOLEMENT

Exploitant	Département	Parcelle	Surface (ha)	Culture
EXPL1	81	1	13,33	BTH
EXPL2	81	1	5,27	ORH
EXPL1	11	1	11,50	PPT
EXPL2	2A	2	2,70	BTH
EXPL3	29	2	15,20	MIS



ASSOLEMENT

Exploitant	Département	Parcelle	Surface (ha)	Culture
EXPL2	81	1	5,27	ORH

- Maintien des lignes du tableau ASSOLEMENT :
- Associées à l'exploitant n°2
 - Parcelles dont la superficie est supérieure à 5 ha

6.9 Trier des données

Objectif : classer les données d'un tableau selon les valeurs d'une ou plusieurs variables/colonnes.

Paramètres nécessaires :

- nom du tableau concerné
- nom des variables/colonnes utilisées pour le classement
- ordre de tri

Remarques : - si plusieurs variables/colonnes sont impliquées dans un tri, l'ordre de tri peut être différent pour chacune d'elles

- tous les langages ne traitent pas de la même manière les valeurs vides.

6.10 Trier des données

CULTURES

Code	Libellé	Saison de semis
BTH	Blé tendre d'hiver	Automne
ORH	Orge d'hiver	Automne
CZH	Colza	Eté
PPT	Prairie temporaire	Indifférent
MIS	Maïs	Printemps
TRN	Tournesol	Printemps
AIL	Ail	Hiver



CULTURES

Code	Libellé	Saison de semis
AIL	Ail	Hiver
BTH	Blé tendre d'hiver	Automne
CZH	Colza	Eté
MIS	Maïs	Printemps
ORH	Orge d'hiver	Automne
PPT	Prairie temporaire	Indifférent
TRN	Tournesol	Printemps

Classement des données du tableau CULTURES par ordre alphabétique croissant du « Code »

6.11 Trier des données

ASSOLEMENT

Exploitant	Département	Parcelle	Surface (ha)	Culture
EXPL1	81	1	13,33	BTH
EXPL2	81	1	5,27	ORH
EXPL1	11	1	11,50	PPT
EXPL2	2A	2	2,70	BTH
EXPL3	29	2	15,20	MIS



ASSOLEMENT

Exploitant	Département	Parcelle	Surface (ha)	Culture
EXPL1	81	1	13,33	BTH
EXPL1	11	1	11,50	PPT
EXPL2	81	1	5,27	ORH
EXPL2	2A	2	2,70	BTH
EXPL3	29	2	15,20	MIS

Classement des données du tableau ASSOLEMENT selon l'ordre de tri :

- N° exploitant croissant
- Surface des parcelles décroissante

6.12 Enrichir les données par ajout d'une variable/colonne

Objectif : compléter les informations d'un tableau en ajoutant une variable/colonne nécessaire à l'analyse.

Paramètres nécessaires :

- nom du tableau concerné
- nom de la nouvelle variable/colonne
- type de la nouvelle variable/colonne
- règle d'initialisation des valeurs de la nouvelle variable/colonne

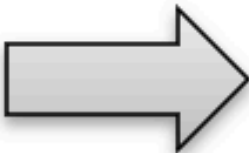
Remarque :

- les règles d'initialisation des valeurs peuvent être simples (une seule valeur pour toutes les lignes) ou plus complexes (calcul à partir des données de plusieurs variables/colonnes, valeurs différentes selon des conditions définies ...)

6.13 Enrichir les données par ajout d'une variable/colonne

ASSOLEMENT

Exploitant	Département	Parcelle	Surface (ha)	Culture
EXPL1	81	1	13,33	BTH
EXPL2	81	1	5,27	ORH
EXPL1	11	1	11,50	PPT
EXPL2	2A	2	2,70	BTH
EXPL3	29	2	15,20	MIS



ASSOLEMENT

Année	Exploitant	Département	Parcelle	Surface (ha)	Culture
2025	EXPL1	81	1	13,33	BTH
2025	EXPL1	11	1	11,50	PPT
2025	EXPL2	81	1	5,27	ORH
2025	EXPL2	2A	2	2,70	BTH
2025	EXPL3	29	2	15,20	MIS

Ajout de la colonne « Année » dont la valeur est initialisée à 2025 dans le tableau ASSOLEMENT

6.14 Enrichir les données par jointure/appariement

Objectif : compléter les informations d'un tableau en l'associant à un autre par rapprochement des valeurs présentes dans des variables/colonnes communes.

Paramètres nécessaires :

- nom des 2 tableaux concernés
- nom des variables/colonnes utilisées pour le rapprochement
- type de jointure souhaitée

Remarques :

- les types de jointure possibles définissent le devenir des lignes du tableau qui sont présentes dans un tableau mais pas dans l'autre.

6.15 Enrichir les données par jointure/appariement

ASSOLEMENT

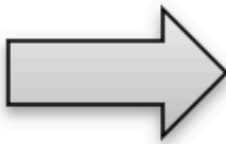
Exploitant	Département	Parcelle	Surface (ha)	Culture
EXPL1	81	1	13,33	BTH
EXPL2	81	1	5,27	ORH
EXPL1	11	1	11,50	PPT
EXPL2	2A	2	2,70	BTH
EXPL3	29	2	15,20	MIS

CULTURES

Code	Libellé	Saison de semis
BTH	Blé tendre d'hiver	Automne
ORH	Orge d'hiver	Automne
CZH	Colza	Eté
PPT	Prairie temporaire	Indifférent
MIS	Maïs	Printemps
TRN	Tournesol	Printemps
AIL	Ail	Hiver

ASSOLEMENT

Exploitant	Département	Parcelle	Surface (ha)	Culture	Libellé culture
EXPL1	81	1	13,33	BTH	Blé tendre d'hiver
EXPL2	81	1	5,27	ORH	Orge d'hiver
EXPL1	11	1	11,50	PPT	Prairie temporaire
EXPL2	2A	2	2,70	BTH	Blé tendre d'hiver
EXPL3	29	2	15,20	MIS	Maïs



Ajout de la colonne « Libellé culture » dans le tableau ASSOLEMENT par jointure avec le tableau CULTURES avec appariement sur la variable du « code culture ».

6.16 Enrichir les données par jointure/appariement

ASSOLEMENT					DEPARTEMENTS					REGIONS	
Exploitant	Département	Parcelle	Surface (ha)	Culture	Code	Libellé	Région	Début de validité	Fin de validité	Code	Libellé
EXPL1	81	1	13,33	BTH	11	Aude	91	01/01/1956	31/12/2015	91	Languedoc-Roussillon
EXPL2	81	1	5,27	ORH	11	Aude	76	01/01/2016		73	Midi-Pyrénées
EXPL1	11	1	11,50	PPT	2A	Corse du Sud	94	01/01/1956		76	Occitanie
EXPL2	2A	2	2,70	BTH	29	Finistère	53	01/01/1956		53	Bretagne
EXPL3	29	2	15,20	MIS	56	Morbihan	53	01/01/1956		24	Centre Val de Loire
					81	Tarn	73	01/01/1956	31/12/2015	94	Corse
					81	Tarn	76	01/01/2016			

→

Ajout de la colonne « Libellé région » dans le tableau ASSOLEMENT par jointure avec les tableaux DEPARTEMENTS et REGIONS avec appariement sur les variables « code département » et « code région ».

Exploitant	Département	Parcelle	Surface (ha)	Culture	Libellé région
EXPL1	81	1	13,33	BTH	Midi-Pyrénées
EXPL1	81	1	13,33	BTH	Occitanie
EXPL2	81	1	5,27	ORH	Midi-Pyrénées
EXPL2	81	1	5,27	ORH	Occitanie
EXPL1	11	1	11,50	PPT	Languedoc-Roussillon
EXPL1	11	1	11,50	PPT	Occitanie
EXPL2	2A	2	2,70	BTH	Corse
EXPL3	29	2	15,20	MIS	Bretagne

6.17 Enrichir les données par jointure/appariement

CULTURES

Code	Libellé
BTH	Blé tendre d'hiver
ORH	Orge d'hiver

SAISONS DE SEMIS

Libellé
Automne
Hiver
Printemps
Eté
Indifférent

CULTURES

Code	Libellé culture	Libellé période de semis
BTH	Blé tendre d'hiver	Automne
BTH	Blé tendre d'hiver	Hiver
BTH	Blé tendre d'hiver	Printemps
BTH	Blé tendre d'hiver	Eté
BTH	Blé tendre d'hiver	Indifférent
ORH	Orge d'hiver	Automne
ORH	Orge d'hiver	Hiver
ORH	Orge d'hiver	Printemps
ORH	Orge d'hiver	Eté
ORH	Orge d'hiver	Indifférent

Ajout de la colonne « Libellé période de semis » dans le tableau CULTURES par jointure avec le tableau SAISONS DE SEMIS.

Sans colonne commune entre les 2 tableaux, toutes les combinaisons possibles sont créées (produit cartésien)

6.18 Agréger les données

Objectif : regrouper des lignes d'un tableau selon une ou plusieurs variables/colonnes afin de calculer des valeurs agrégées pour chacun des groupes.

Paramètres nécessaires :

- nom du tableau concerné
- nom des variables/colonnes utilisées pour le regroupement
- type d'agrégat souhaité (comptage, somme, moyenne, pourcentage, minimum, maximum ...)

Exemples :

- à partir d'un tableau de parcelles (département, commune, référence cadastrale, superficie), calculer la superficie moyenne d'une parcelle pour chaque commune
- à partir d'un tableau de personnes (nom, prénom, date de naissance), calculer le nombre de personnes selon leur tranche d'âge

6.19 Agréger des données

ASSOLEMENT

Exploitant	Département	Parcelle	Surface (ha)	Culture
EXPL1	81	1	13,33	BTH
EXPL2	81	1	5,27	ORH
EXPL1	11	1	11,50	PPT
EXPL2	2A	2	2,70	BTH
EXPL3	29	2	15,20	MIS

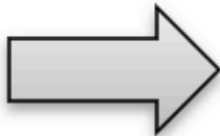


TABLEAU 1

Exploitant	Nombre de parcelles
EXPL1	2
EXPL2	2
EXPL3	1

Calcul du nombre de parcelles (« comptage ») par exploitant à partir des données du tableau ASSOLEMENT

TABLEAU 2



Culture	Cumul des surfaces
BTH	16,03
ORH	5,27
PPT	11,50
MIS	15,20

Cumul de la superficie des parcelles (« somme ») par culture à partir des données du tableau ASSOLEMENT

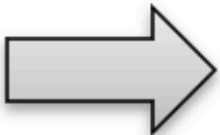


TABLEAU 3

Département	Superficie moyenne des parcelles
81	9,30
11	11,50
2A	2,70
29	15,20

Calcul de la superficie moyenne des parcelles (« moyenne ») par département à partir des données du tableau ASSOLEMENT

6.20 Gérer des doublons

Objectif : identifier les lignes répétées dans un tableau de données et choisir le traitement adéquat pour l'analyse.

Paramètres nécessaires :

- nom du tableau concerné
- nom des variables/colonnes utilisées pour la recherche des doublons

Remarque :

- plusieurs lignes d'un tableau peuvent être qualifiées de doublons même si la totalité des variables/données ne sont pas identiques

Ex : dans une liste de personnes (n°INSEE, nom, prénom, lieu de naissance), plusieurs lignes référençant le même numéro INSEE peuvent être considérées comme doublons car cette variable/colonne est identifiante (1 numéro est attribué à un seul individu) et unique pour une personne (un individu ne possède qu'un seul numéro).

6.21 Gérer des doublons

EXPLOITANTS INDIVIDUELS

Exploitant	N°Pacage	Nom	Prénom	Date de naissance
EXPL1	081123123	DUPOND	Marie	01/01/1975
EXPL2	081999999	DURAND	Jean	01/01/1992
EXPL1	081123123	DUPOND	Marie	01/01/1975
EXPL3	081999999	MARTINEZ	Jean	01/01/1962
EXPL4	029123123	MARTIN	Louis	01/01/1970



« vrai » doublon
Les données de la totalité des colonnes sont identiques (même ligne présente plusieurs fois)



Doublon fonctionnel
Le n°Pacage représente 2 exploitants individuels alors qu'il est supposé être unique. Problème de cohérence fonctionnelle des données.

6.22 Concaténer des tables

Objectif : assembler plusieurs tableaux de données

- par ajout de lignes (si les tableaux possèdent les même variables/colonnes)
- par ajout de colonnes sans rapprochement des données (si les tableaux possèdent les mêmes objets en ligne)

Paramètres nécessaires :

- nom des tableaux concernés
- type de concaténation (ajout de lignes ou ajout de colonnes)

Remarque :

- la concaténation de tables par ajout de colonnes ne doit pas être confondue avec un appariement. Les données des tableaux ne font pas l'objet de rapprochement selon des variables/colonnes communes.

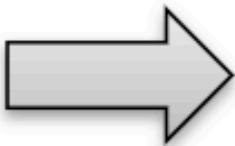
6.23 Concaténer des tables

ASSOLEMENT PETITES PARCELLES

Exploitant	Département	Parcelle	Surface (ha)	Culture
EXPL2	81	1	5,27	ORH
EXPL2	2A	2	2,70	BTH

ASSOLEMENT GRANDES PARCELLES

Exploitant	Département	Parcelle	Surface (ha)	Culture
EXPL1	81	1	13,33	BTH
EXPL1	11	1	11,50	PPT
EXPL3	29	2	15,20	MIS



Création du tableau ASSOLEMENT par concaténation des tableaux ASSOLEMENT PETITES PARCELLES et ASSOLEMENT GRANDES PARCELLES (ajout de lignes)

ASSOLEMENT

Exploitant	Département	Parcelle	Surface (ha)	Culture
EXPL2	81	1	5,27	ORH
EXPL2	2A	2	2,70	BTH
EXPL1	81	1	13,33	BTH
EXPL1	11	1	11,50	PPT
EXPL3	29	2	15,20	MIS

6.24 Concaténer des tables

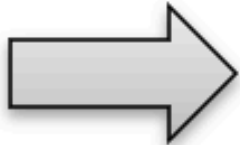
EXPLOITANTS INDIVIDUELS

Exploitant	N°Pacage	Nom	Prénom	Date de naissance
EXPL1	081123123	DUPOND	Marie	01/01/1975
EXPL2	081999999	DURAND	Jean	01/01/1992
EXPL4	029123123	MARTIN	Louis	01/01/1970

ADRESSES EXPLOITANTS

Ligne adresse	Code postal	Commune
La Métairie Rouge	81100	CASTRES
La Borie Neuve	81100	CASTRES
La Ferme aux Libellules	29510	BRIEC

EXPLOITANTS INDIVIDUELS AVEC ADRESSE



Exploitant	N°Pacage	Nom	Prénom	Date de naissance	Ligne adresse	Code postal	Commune
EXPL1	081123123	DUPOND	Marie	01/01/1975	La Métairie Rouge	81100	CASTRES
EXPL2	081999999	DURAND	Jean	01/01/1992	La Borie Neuve	81100	CASTRES
EXPL4	029123123	MARTIN	Louis	01/01/1970	La Ferme aux Libellules	29510	BRIEC

Concaténation des tables EXPLOITANTS INDIVIDUELS et ADRESSES EXPLOITANTS par ajout des colonnes.
A manipuler avec précaution !

6.25 Transposer une table

Objectif : modifier la structure d'un tableau de façon à présenter :

- en lignes des données qui étaient présentes en variables/colonnes
- en variables/colonnes des données qui étaient présentes en lignes

Exemple issu de la formation R du SSM Agriculture :

Exemple issu de la formation R du MASA:

Annee	Mois	Mesure	Temperature
2020	Janvier	Min	-10
2020	Janvier	Max	12
2020	Janvier	Moy	3
2020	Février	Min	-6
2020	Février	Max	14
2020	Février	Moy	5

Annee	Mois	Max	Min	Moy
2020	Janvier	12	-10	3
2020	Février	14	-6	5

tab2



Les valeurs de la variable/colonne « Mesure » entraînent la création de nouvelles variables/colonnes qui sont initialisées avec les valeurs de la variable/colonne « Temperature »



Les 3 variables Max, Min et Moy sont basculées en lignes et les valeurs associées alimentent une nouvelle variable/colonne nommée « Temperature ».