



Schistosomiasis Control Initiative

Épuration et analyse des données avec Excel

Session 0 *Résumé des sessions 1 à 9*

0.1 *Session 1 – Introduction Excel*

1. Fondamentaux : types de données Excel, formatage et formatage conditionnel des cellules
Formats des nombres, fusion de cellules, création de lignes de séparation dans les cellules
2. Masquer des colonnes, filtrer et trier des lignes
3. Introduction aux formules et références de cellule relatives et plages de cellules
4. Comment utiliser les références de cellules relatives pour copier des formules
5. Fonctions importantes et fonction SI
6. Introduction à la fonction RECHERCHEV pour trouver des informations dans un tableau en fonction d'une variable clé

0.2 *Session 2 – Classeur de nettoyage des données et vérification des variables d'identité*

1. Créer un dossier sur lequel travailler et le renommer par exemple « session de nettoyage de données »
2. Créer des dossiers dans le dossier initial :
 - a. '1 Données brutes'
 - b. '2 Nettoyage des données'
 - c. '3 Données nettoyées'
 - d. '4 Analyse des données'
 - e. '5 Résultats'
3. Enregistrer les données brutes dans '1 Données brutes'.
4. Enregistrer le classeur de nettoyage des données dans '2 Nettoyage des données' sous un nom pertinent.
5. Enregistrer les données brutes dans '2 Nettoyage des données' sous 'données 0'. Désormais, utiliser ce fichier.
6. Prendre les données et le classeur de nettoyage des données.
7. Figurer les volets.
8. Créer des numéros de ligne.
9. Vérifier que le code école est unique à l'aide du tableau croisé dynamique. Effectuer les modifications requises éventuelles et enregistrer dans le classeur de nettoyage des données.
10. Exercice : vérifier que le code élève est unique à l'aide du tableau croisé dynamique. Effectuer les modifications requises éventuelles et enregistrer dans le classeur de nettoyage des données.
11. Enregistrer les données obtenues sous 'données 1'.
12. Créer un code école dans le fichier de l'élève. Démonstration de GAUCHE et exercice de STXT.
13. Vérifier que le code école est identique dans le groupe de données de l'école et de l'élève. Effectuer les modifications requises éventuelles et enregistrer dans le classeur de nettoyage des données.

14. Enregistrer les données obtenues sous 'données 2'. Enregistrer le classeur de nettoyage.

0.3 *Session 3 – Créer des synthèses de données brutes*

1. Utiliser 'données 2' et 'classeur de nettoyage de données'
2. Inclure la comparaison de feuilles de calcul
3. Enregistrer les données obtenues sous 'données 3'. Enregistrer le classeur de nettoyage.

0.4 *Session 4 – Vérifier le protocole*

1. Démonstration : vérifier le sexe avec un jeu de données fictives à l'aide des tableaux dynamiques croisés
2. Ouvrir 'données 3' et vérifier le sexe comme montré précédemment. Enregistrer les résultats dans le classeur de nettoyage des données.
3. Enregistrer le classeur de nettoyage des données et enregistrer le jeu de données sous 'données 4'.
4. Démonstration : vérifier l'âge avec un jeu de données fictives.
5. Vérifier l'âge dans le jeu de données 'données 4' et enregistrer dans le classeur de nettoyage des données.
6. Enregistrer 'données 4' et le classeur de nettoyage des données
7. Facultatif : inclure des tableaux dynamiques croisés

0.5 *Session 5 – Créer des variables parasitologiques*

1. Compléter le 'classeur de fonctions'. Mettre l'accent sur les fonctions 'si'.
2. À l'aide des 'données 4', créer les variables parasitologiques de :
 - a. S mansoni mean epg, inf et heavy inf (démonstration / exercice)
 - b. Ascaris mean epg, inf et heavy inf (exercice)
 - c. Ankylostomiose mean epg, inf et heavy inf (exercice)
 - d. Trichuris mean epg, inf et heavy inf (exercice)
 - e. S haem ep10ml, inf et heavy inf (exercice)
3. Enregistrer les données obtenues sous 'données 5'.

0.6 *Session 6 – Créer des synthèses de données nettoyées*

1. Utiliser 'données 5' et 'classeur de nettoyage de données'.
2. Enregistrer les données obtenues sous 'données 6'. Enregistrer le classeur de nettoyage.
3. Enregistrer les données nettoyées (données 6) dans '3 Données nettoyées' sous un nom pertinent. Saisir la date à la fin sous la forme 'aaaa-mm-jj'. Par exemple pour ces données 'ETH_2015_impact_clean_data_2016-10-15'

0.7 *Session 7 – Créer le tableau de résultats*

1. Enregistrer les données nettoyées dans '4 Analyses des données' sous 'création du tableau de résultats'.
2. Enregistrer 'modèle résultats' dans '5 Résultats' sous un nom pertinent, en incluant la date d'aujourd'hui. Par exemple 'ETH_2015_impact_2016-10-15'
3. Démonstration : résultats généraux SCH à l'aide des tableaux croisés dynamiques dans 'création du tableau de résultats' Copier les informations (à l'aide des valeurs copier) dans le tableau de résultats.
4. Utiliser CONCATENER x pour rassembler les âges minimum et maximum
5. Exercice : SCH par école, âge et variable parasitologique
6. Compléter le 'Classeur Index et Equiv' (index and match).
7. Démonstration : utiliser Index et Equiv pour importer les informations de l'école dans les tableaux de résultats classés par école.
8. Exercice : compléter le tableau de résultats pour STH.
9. Enregistrer le tableau de résultats.

0.8 *Session 8 – Créer des graphiques*

1. Enregistrer les données résultats dans '4 Analyses des données' sous 'création de graphiques'.
2. Démonstration : créer un graphique à barres de la prévalence de SCH.
3. Copier/coller le graphique définitif dans un Powerpoint sous forme d'image.
4. Enregistrer le Powerpoint dans '5 Résultats' sous un nom pertinent.
5. Exercice : créer le point de diffusion de *S. mansoni* vs *S. haematobium*.
6. Exercice : créer d'autres graphiques

0.9 *Session 9 – Saisie des données*

1. Qu'est-ce que la validation des données et aperçu des options de validation des données dans Excel
2. Utiliser la boîte de dialogue de validation des données
3. Démonstration : validation des données appliquées aux données CCA
4. Relier la liste des options à une valeur choisie pour une autre cellule

Session 1 Introduction Excel

1.1 Saisie des données, mise en forme des cellules et mise en forme conditionnelle

Les données saisies dans les cellules peuvent être de différents types :

1. Nombres : alignés à droite, peuvent aussi être une date, une heure, etc.
2. Texte : aligné à gauche. Pour saisir des nombres sous forme de texte, définir le format de cellule sur *texte* ou commencer la saisie par une apostrophe, par ex., '123.
3. Formule : toujours commencer par le signe égal, par ex., = A1 + A2

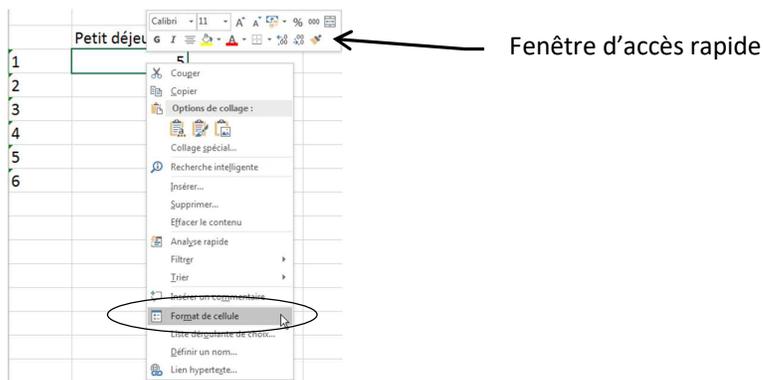


Poignée de recopie

Pour copier rapidement des données dans des cellules adjacentes, utiliser la « poignée de recopie ».

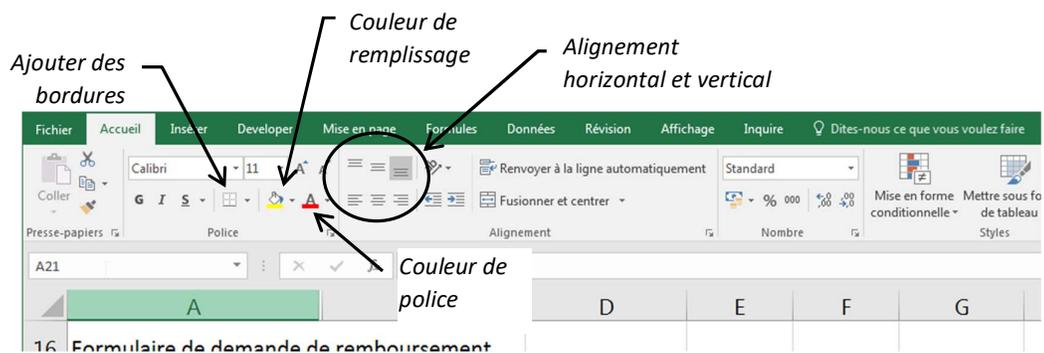
Pour accéder à *toutes* les options de mise en forme d'une cellule, utiliser le menu de contexte : clic droit sur la cellule à formater puis :

Cliquer sur « Format de cellule » ou utiliser les options d'accès rapide dans la fenêtre supérieure.

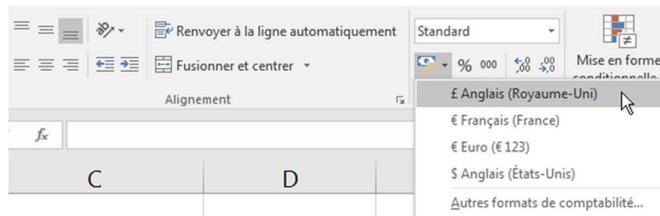


Démonstration

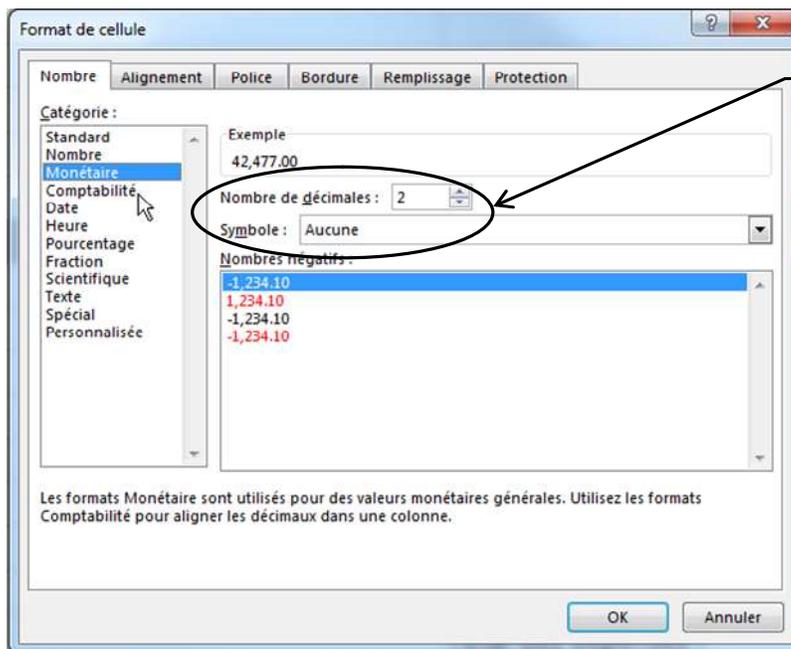
1. Ouvrir le fichier *Session_1.xlsx* sur la clé USB. Les onglets en bas de page indiquent le nom de la feuille de calcul. Aller sur la feuille *Démo 1*.
2. **Pour modifier la couleur d'une cellule ou d'un texte**, utiliser le menu de contexte ou le Ruban d'Accueil :



Les nombres peuvent être formatés en Devise, Date ou Pourcentage par exemple. Pour définir le format d'une cellule sur Devise, utiliser le bouton Devise du Ruban

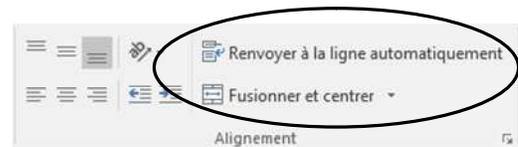


ou utiliser *Format de cellule...* dans le menu de contexte (clic droit sur la cellule) :

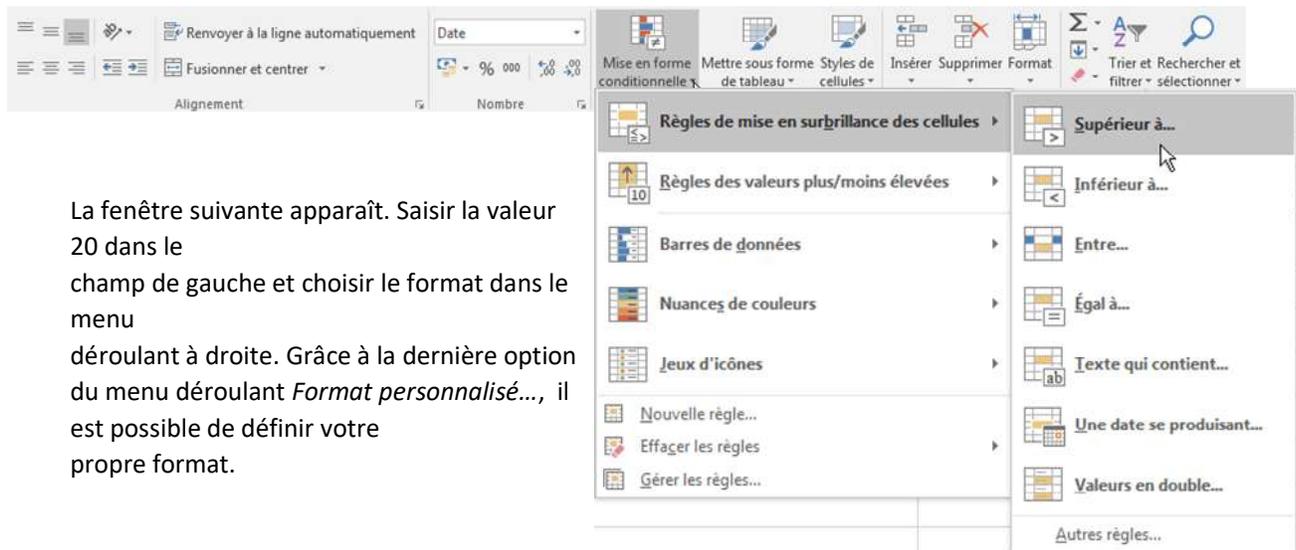


Définir le nombre de décimales et le symbole de la devise

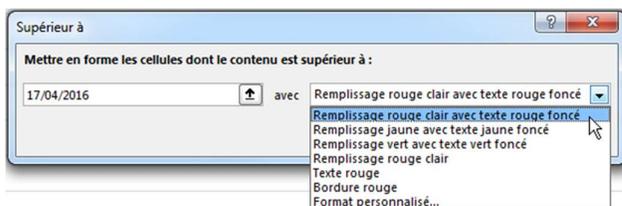
3. Pour **fusionner plusieurs cellules** en une, sélectionner les cellules à fusionner et cliquer sur le bouton *Fusionner et centrer* dans le Ruban d'Accueil. Pour **aller à la ligne** dans une cellule, cliquer sur le bouton *Renvoyer à la ligne automatiquement*. Les deux commandes apparaissent également dans le menu de contexte *Format de cellule* à l'onglet *Alignement*.



4. La **mise en forme conditionnelle** vous permet de définir le format d'une cellule, par ex., la couleur de la cellule, en fonction de son contenu. Par exemple, si toutes les valeurs supérieures à 20 doivent présenter un remplissage rouge clair et un texte rouge foncé, sélectionner la cellule à formater et cliquer sur *Mise en forme conditionnelle* dans le Ruban et choisir *Règles de mise en surbrillance des cellules* puis *Supérieur à...*



La fenêtre suivante apparaît. Saisir la valeur 20 dans le champ de gauche et choisir le format dans le menu déroulant à droite. Grâce à la dernière option du menu déroulant *Format personnalisé...*, il est possible de définir votre propre format.



Maintenant, à vous :

La feuille *Exercice 1* dans le deuxième onglet contient un formulaire de demande de remboursement similaire.

Mettre en forme le formulaire pour qu'il ressemble à celui-ci :

Formulaire de dépenses

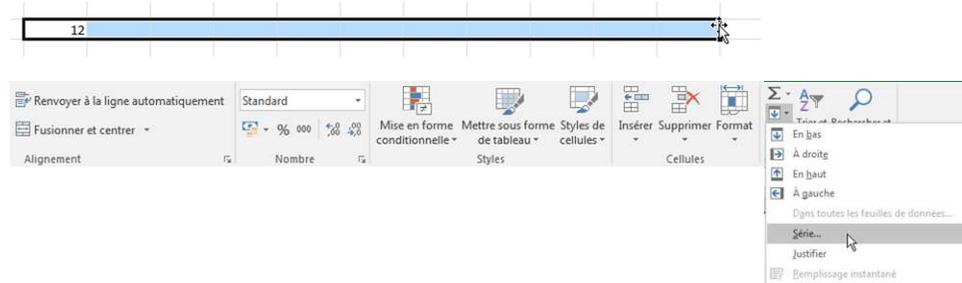
Nom	Udo Wittmann		Département	SCI		Motif du déplacement	5 jours de conférence sur les maladies tropicales négligées associés à une formation sur														
De	1/12/15																				
	Jour	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15					
	Date	1-déc.	2-déc.	3-déc.	4-déc.	5-déc.	6-déc.	7-déc.	8-déc.	9-déc.	10-déc.	11-déc.	12-déc.	13-déc.	14-déc.	15-déc.					
	Petit déjeuner et déjeuner	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00	Fr. 20.00					
	Billet entrée conférence	Fr. 110.00	Fr. 110.00	Fr. 110.00	Fr. 110.00	Fr. 110.00															
	Dîner	Fr. 12.00	Fr. 56.00	Fr. 56.00	Fr. 88.00	Fr. 100.00	Fr. 12.00	Fr. 15.00	Fr. 44.00	Fr. 30.00	Fr. 30.00	Fr. 65.00	Fr. 70.00	Fr. 110.00	Fr. 105.00	Fr. 12.00					
	Total des dépenses	€1,655.00																			

Utiliser la mise en forme conditionnelle pour marquer les dépenses de repas supérieures à 50 en jaune, et les dépenses de repas supérieures à 80 en rouge. Appliquer également une mise en forme conditionnelle similaire au total des dépenses.

Exercice facultatif :

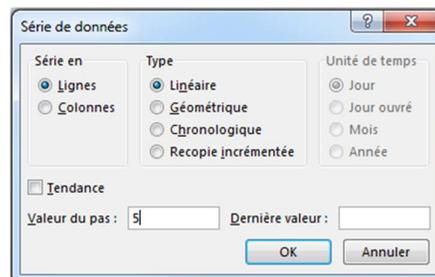
Modifier la dernière ligne des dépenses de repas : Nous dépensons 12 £ le premier jour puis chaque jour 5 £ de plus que la veille. Au lieu de saisir ces chiffres à la main, on peut utiliser l'outil de remplissage.

Tout d'abord, saisir 12 dans la première cellule puis faire glisser la souris sur les cellules que vous souhaitez remplir automatiquement :



Définir les options comme indiqué ci-dessous. La *Valeur du pas* désigne la valeur à ajoutée à chaque pas à la dernière valeur de la série. Dans cet exemple, ce sera 5.

Tester d'autres paramètres pour créer des séries différentes.



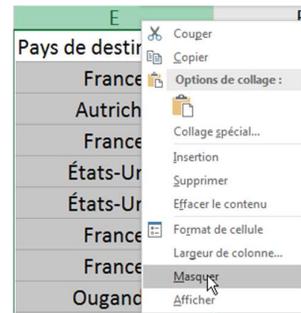
1.2 Masquer, trier, filtrer ; la barre d'état

Lorsqu'on travaille sur de grands tableaux, il est parfois difficile de suivre le travail effectué en différents endroits du tableau. Masquer des colonnes ou des lignes, filtrer et trier le tableau en fonction de la valeur d'une colonne peuvent faciliter la tâche.

Démonstration

1. Cliquer sur la feuille *Démo 2*.
2. Pour commencer, on applique la mise en forme définie dans *Démo 1* : faire pivoter les en-têtes de section, définir un format correct pour les dates et les devises, appliquer une mise en forme conditionnelle aux dépenses (toutes les dépenses de déjeuner supérieures à 5 £ doivent apparaître en rouge)
3. Pour mieux visualiser les dépenses, on **masque** les colonnes sans intérêt. Pour cela, mettre en surbrillance les colonnes

que l'on souhaite masquer et faire un clic droit sur l'en-tête de ces colonnes. Une fenêtre similaire à l'illustration de droite devrait apparaître. Cliquer sur Masquer pour masquer les colonnes sélectionnées.



- Mettre en surbrillance toutes les colonnes et faire un clic droit sur l'un des en-têtes de colonne. Pour appliquer un **filtre**, faire un clic droit sur le premier nom dans la colonne *Nom du voyageur*.

Sélectionner Filtrer puis *Filtrer par la valeur de la cellule sélectionnée*.

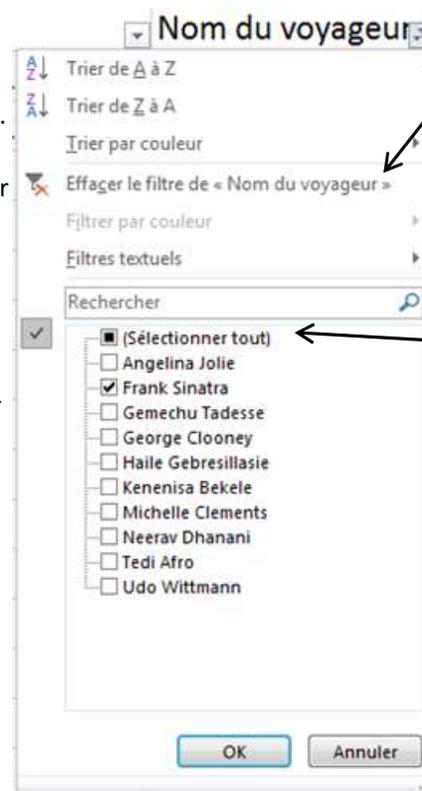
- Mettre en surbrillance la première entrée de la colonne Dépenses de dîner. Appuyer sur MAJ - Ctrl - ↓ pour sélectionner toutes les valeurs de cette colonne. La moyenne des cellules sélectionnées apparaît dans la **barre d'état** à droite. Pour modifier la barre d'état en fonction de vos besoins, faire un clic droit et cocher les valeurs nécessaires.

- Cliquer sur  pour visualiser la liste de toutes les valeurs de cette colonne. Sélectionner les noms dont vous souhaitez voir les colonnes. Cliquer sur *(Tout sélectionner)* pour sélectionner ou désélectionner toutes les valeurs.

Pour supprimer tous les filtres, cliquer sur Filtrer dans le Ruban des Données :



Que peut-on dire des dépenses d'Angelina Jolie ?
À quelle fréquence Frank Sinatra a-t-il voyagé ?

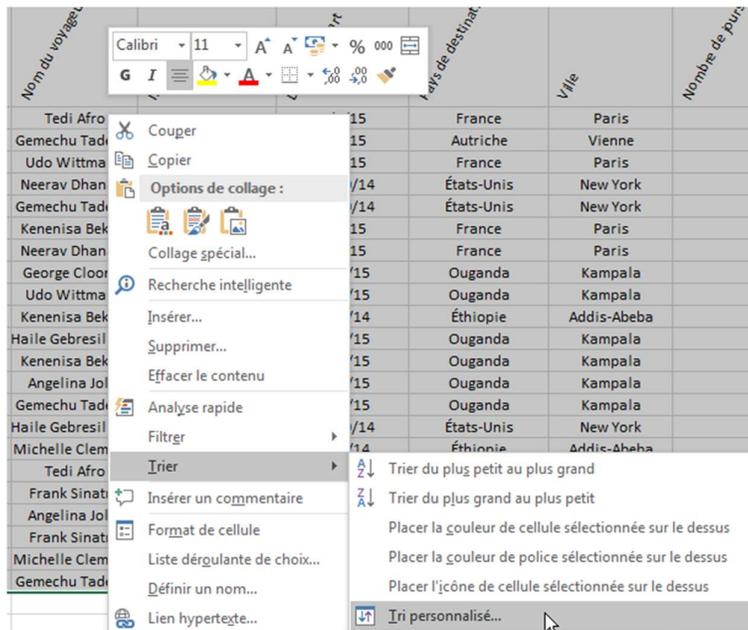


Cliquer ici pour supprimer le filtre

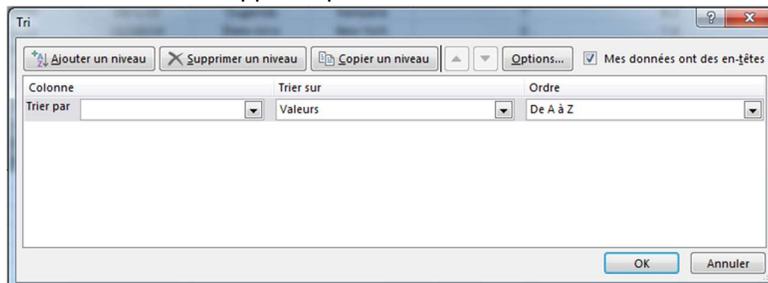
Cliquer sur (Tout sélectionner) pour désélectionner tous les éléments puis cocher les noms que vous souhaitez voir apparaître

- Pour **trier** tout le tableau en fonction des noms des voyageurs, sélectionner tout le tableau (sans les en-têtes). Faire un clic droit sur la colonne *Nom du*

voyageur et choisir *Trier*. Le sous-menu vous propose à présent différents accès rapides pour trier le tableau. Choisir *Tri personnalisé...* pour cette démonstration.



La fenêtre suivante apparaît pour définir le tri.



Trier par : Choisir la colonne à utiliser comme clé de tri, ici *Nom du voyageur*

Trier sur : Choisir *Valeurs*. Cela signifie que les valeurs des cellules de la colonne *Nom du voyageur* doivent être utilisées comme clé de tri.

Ordre : De A à Z, ce sera l'ordre croissant, de Z à A l'ordre décroissant.

Cliquer sur OK pour trier.

Quelle personne a voyagé le plus ?

Maintenant, à vous

8. Aller sur la feuille *Exercice 2*.
9. Créer une nouvelle colonne vide avant la première colonne.
Numéroter chaque cellule de cette colonne dans l'ordre pour chaque observation.
10. Masquer la colonne qui contient le numéro dépense.

11. Utiliser le filtre pour déterminer qui a tendance à dépenser plus (moins) que la moyenne lors des conférences.
12. Trier le tableau par conférence et indiquer celui qui a dépensé le plus pour chaque conférence.
Dans l'ensemble, qui a dépensé le plus ?

Exercice facultatif

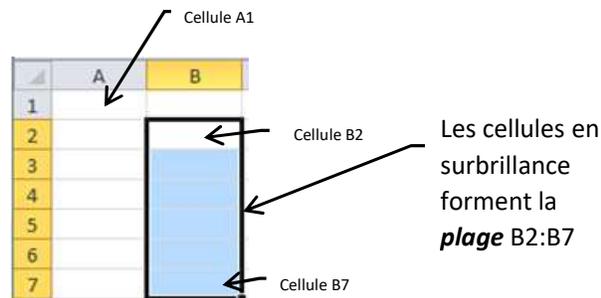
Voici un petit aperçu de la section suivante. Déterminer le coût total de chaque ligne en saisissant la formule = H2 + I2 + J2 dans la cellule K2 comme illustré ci-dessous.

G	H	I	J	K
Nom de la conférence	Tarif conférence	Repas	Déplacement	Total en \$
ASTMH	590	90	45	=H2+I2+J2

Utiliser la poignée de recopie pour copier la formule dans les autres lignes. Vérifier les nouvelles formules créées à l'aide de la poignée de recopie.

1.3 Références de cellules, plages de cellules et formules

Chaque cellule peut être désignée par sa colonne et sa ligne.



L'expression A1 est la **référence de cellule** désignant la cellule en haut à gauche.

L'expression B2:B7 désigne la place de cellules débutant par la cellule B2 et terminant par la cellule B7. Les B2, B3, B4, B5, B6, B7 font toutes partie de la plage B2:B7.

Démonstration

13. Créer la situation décrite dans la capture d'écran à droite :
remplir la plage de cellules B2:B16 des nombres 1 à 15
(utiliser la poignée de recopie). Placer le curseur dans la
cellule C3 et saisir le signe égal. Ainsi, on indique à Excel
qu'une formule va être saisie dans cette cellule. Saisir la
formule
= B3 + B2
On peut également cliquer sur les cellules respectives pour
créer les références de cellule. Appuyer sur Entrée pour
afficher le résultat.

À présent, utiliser la poignée de recopie pour copier la
formule dans les cellules sous C3.

Les références de cellule s'adaptent à la nouvelle position de
la formule, c'est-à-dire que ces références ne sont pas des
positions absolues de cellules mais des positions relatives à
la cellule où se situe la formule. Les références de cellules
dont nous parlons sont donc appelées des **références de
cellule relatives**.

	A	B	C	D
1				
2		1		
3		2	3	
4		3	=B4+B3	
5		4	7	
6		5	9	
7		6	11	
8		7	13	
9		8	15	
10		9	17	
11		10	19	
12		11	21	
13		12	23	
14		13	25	
15		14	27	
16		15	29	

14. À présent, aller sur la feuille « Démo 3 » dans le fichier Session_1.xlsx.
Placer le curseur dans la cellule G62 sous le nombre d'œufs d'Ascaris.
Saisir la formule
=MOYENNE(G2:G61)
La plage de cellules peut être saisie en cliquant sur la cellule G61 et en remontant jusqu'à la
cellule G2. Relâcher le bouton de la souris.
La cellule affiche désormais la moyenne du nombre d'œufs d'Ascaris.
Recopier la formule vers la droite avec la poignée de recopie permet de créer la même
formule mais pour l'ankylostomiase et Trichuris.
MOYENNE () est une fonction. Il existe d'autres fonctions importantes
- | | |
|------------------|---|
| MIN(G2:G61) | trouve la valeur minimale dans une plage de cellules |
| MAX(G2:G61) | trouve la valeur maximale dans une plage de cellules |
| SOMME(G2:G61) | indique la somme des valeurs de la plage de cellules. |
| NB(G2:G61) | indique le nombre de cellules dans la plage de cellules. |
| VAR.S(G2:G61) | indique l'estimation de la variance de
la population en fonction des données de la plage G2:G61.
(en langage courant : c'est la variance que l'on utiliserait
normalement) |
| ECARTYPE(G2:G61) | indique l'écart type, c'est-à-dire la racine carrée de la variance
située au-dessus. |
- Il existe beaucoup d'autres fonctions dont nous parlerons plus tard.

15. La fonction SI :

La fonction SI indique une valeur dépendant d'une condition : sa forme générale est :

=SI(**condition**, valeur à indiquer si la condition est **vraie**, valeur à indiquer si la condition est **fausse**)

Placer le curseur dans la cellule J2 et saisir la formule

=SI(G2 > 0, « infecté par l'ASC », « pas infecté par l'ASC »)

La fonction SI indiquera le texte « infecté par l'ASC » dans la cellule J2 si la valeur de G2 est supérieure à 0, sinon elle indiquera « pas infecté par l'ASC » dans J2. Appuyer sur Entrée.

G	H	I	J	K
Œufs d'Ascaris par gramme	Œufs d'ankylostomiase par g	Œufs de trichuris par gramme	Infecté par Ascaris	infection majeure
0	0	24	=SI(G2 > 0, "infecté par l'ASC", "pas infecté par l'Asc")	
0	72	0	SI(test_logique, [valeur_si_vrai], [valeur_si_faux])	

Maintenant, à vous

16. Cliquer sur la feuille Exercice 3.
17. Comme pour la démonstration, cette feuille contient les données d'infection pour la schistosomiase. Trouver
 - a) le nombre moyen d'œufs pour chaque infection
 - b) le nombre maximal d'œufs pour chaque infection
 - c) la variance du nombre d'œufs pour chaque infection.
 - c) insérer une colonne et la remplir avec 1 si l'infection par SCH haematobium de cette personne est massive et 0 dans le cas contraire. Utiliser une fonction SI. Faire de même pour l'autre infection.

Exercice facultatif

Créer une colonne qui indique le statut d'infection pour SCH mansoni. Utiliser la fonction SI.

Nombre d'œufs = 0 : pas infecté

Nombre d'œufs compris entre 1 et 100 : infection légère

Nombre d'œufs compris entre 101 et 400 : infection modérée

Nombre d'œufs supérieurs à 400 : infection massive

Créer un tableau croisé dynamique pour le statut d'infection.

1.4 La fonction RECHERCHEV

Supposons que nous disposons d'un tableau des données parasitologiques dans lequel chaque école est identifiée par un code. Dans un autre fichier, nous avons les mêmes codes d'école associés aux noms des écoles.

Si l'on souhaite ajouter le nom de l'école à chaque observation dans le premier tableau, il nous faut rechercher le code de l'école dans le second tableau et copier le nom correspondant dans la ligne concernée du premier fichier. Cette recherche peut être faite automatiquement avec la fonction RECHERCHEV (le « v » signifie verticale).

Nous prenons pour exemple la feuille DÉMO 4 contenant les données parasitologiques et les codes des écoles associés à leur nom dans la feuille « Nom des écoles ».

18. Ouvrir le tableau des données et insérer une colonne Nom de l'école

	A	B	C	D	E	F	G	H
1	Code école	Nom d'école	ID élève	Asc	Ankylostomiase			
2	100	=RECHERCHEV(
3	100	RECHERCHEV(valeur_cherchée, table_matrice, no_index_col, [valeur_proche])						
4	100		100.03	1	1			
5	101		101.01	0	1			
6	101		101.02	1	0			
7	101		101.03	0	0			
8	102		102.01	0	0			

19. À la place du nom de l'école, saisir la formule
=recherchev(A2, 'nom de l'école'!A1:B4, 2)

Le code d'école que nous recherchons, 100 dans ce cas

La zone de la feuille "Noms des écoles" qui contient le code de l'école en première colonne

Les noms des écoles doivent provenir de la deuxième colonne « Nom des écoles »

	A	B
1		
2	Code école	Nom école
3	100	École A
4	101	École B
5	102	École C
6	103	École D
7		

20. Avant de copier cette fonction dans les cellules inférieures, on doit définir les références dans le tableau des Noms des écoles comme des **références de cellules absolues** :
recherchev(A2, 'nom de l'école'!\$A\$1:\$B\$4, 2)

Le symbole \$ permet de conserver la partie située juste après le \$ pour qu'elle ne soit pas ajustée à la nouvelle position de la formule. Comme on souhaite conserver totalement la référence à la zone, on insère les symboles \$ avant chaque référence de ligne et colonne.

On utilise ensuite la poignée de recopie pour copier la fonction recherchev dans les cellules d'en dessous.

	A	B
1	Code école	Nom d'école
2	100	=RECHERCHEV(A2, 'nom des écoles'!\$A\$2:\$B\$6, 2)
3	100	=RECHERCHEV(A3, 'nom des écoles'!\$A\$2:\$B\$6, 2)
4	100	=RECHERCHEV(A4, 'nom des écoles'!\$A\$2:\$B\$6, 2)
5	101	=RECHERCHEV(A5, 'nom des écoles'!\$A\$2:\$B\$6, 2)
6	101	=RECHERCHEV(A6, 'nom des écoles'!\$A\$2:\$B\$6, 2)
7	101	=RECHERCHEV(A7, 'nom des écoles'!\$A\$2:\$B\$6, 2)
8	102	=RECHERCHEV(A8, 'nom des écoles'!\$A\$2:\$B\$6, 2)
9	102	=RECHERCHEV(A9, 'nom des écoles'!\$A\$2:\$B\$6, 2)
10	102	=RECHERCHEV(A10, 'nom des écoles'!\$A\$2:\$B\$6, 2)
11	103	=RECHERCHEV(A11, 'nom des écoles'!\$A\$2:\$B\$6, 2)
12	103	=RECHERCHEV(A12, 'nom des écoles'!\$A\$2:\$B\$6, 2)
13	103	=RECHERCHEV(A13, 'nom des écoles'!\$A\$2:\$B\$6, 2)

La référence relative s'est adaptée à la nouvelle position

Les références absolues n'ont pas changé

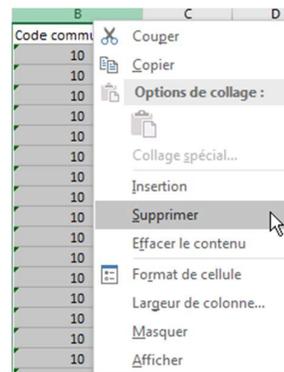
21. Pour remplacer la formule recherchev par le nom de l'école, sélectionner les noms des écoles, cliquer sur copie puis faire un clic droit sur la zone sélectionnée et choisir Collage spécial... - Coller valeurs – Valeurs (V).

Maintenant, à vous

1. Dans la feuille Exercice 4 du fichier Session 1, vous trouverez les mêmes données parasitologiques que pour l'exercice 3. Insérer une colonne « Nom de la communauté » entre le Code de communauté et le Code de district.
2. Dans la feuille suivante « Noms de communauté », vous trouverez une liste des codes de communautés et leurs noms. Utiliser la fonction recherchev pour copier le nom de communauté approprié de la Liste des communautés dans la nouvelle colonne de la feuille Exercice 4.

Exercice facultatif

1. Vous devez tout d'abord procéder à l'exercice 4. Veiller à ce que la feuille Exercice 4 contienne désormais le Nom de communauté pour chaque observation.
2. Convertir les résultats de recherchev en texte. Pour cela, sélectionner la colonne Nom de communauté dans la feuille Exercice 4 et appuyer sur Ctrl-C pour copier. Faire un clic droit dans la zone sélectionnée et cliquer sur Collage spécial – Coller valeurs ou cliquer sur  .
3. Supprimer la colonne Code de communautés en sélectionnant la colonne puis en faisant un clic droit sur son en-tête et Supprimer.
4. Créer une colonne ID élève et utiliser RECHERCHEV et CONCATÉNER pour créer un ID élève composé du CodeDistrict.CodeCommunauté.Codeélève. Utiliser RECHERCHEV pour rechercher le code de communauté correspondant à un nom de communauté donné dans la liste de l'Exercice 4. Utiliser une seule formule pour créer l'ID élève.



Session 2 Classeur de nettoyage des données et vérifications des données initiales

2.1 Qu'est-ce que le nettoyage des données ?

Les données brutes sont les données telles qu'elles ont été saisies dans le système ou la feuille de calcul.

Les données nettoyées sont les données après vérification et modification éventuelle. Les données doivent toujours être nettoyées avant analyse, et chaque modification effectuée doit être notée.

2.2 Le classeur de nettoyage des données

Chez SCI, nous utilisons un classeur de nettoyage des données pour nettoyer les données. Pour quoi faire ?

- Pour veiller à toujours effectuer les mêmes vérifications sur les données
- Pour vérifier si le protocole a été suivi
- Pour garder une trace de toutes les modifications apportées aux données

Le classeur de nettoyage des données est un classeur Excel contenant des feuilles pré-crées qui permet de faciliter le nettoyage des données.

Pour nettoyer les données, les étapes à suivre sont les suivantes :

1. Créer des dossiers de travail
2. Enregistrer les données brutes
3. Vérifier que les variables ID sont correctes
4. Compléter les synthèses de données brutes
5. Vérifier le respect du protocole
6. Créer des variables parasitologiques
7. Faire la synthèse finale des données
8. Enregistrer les données nettoyées
9. Analyser les données nettoyées

Pendant cette session, nous travaillerons sur les étapes 1 à 3.

2.3 Créer des dossiers de travail

Il est recommandé de toujours utiliser la même structure de dossier pour nettoyer des données. Ainsi, vous trouverez toujours les informations dont vous avez besoin même si vous n'avez pas beaucoup consulté les fichiers.

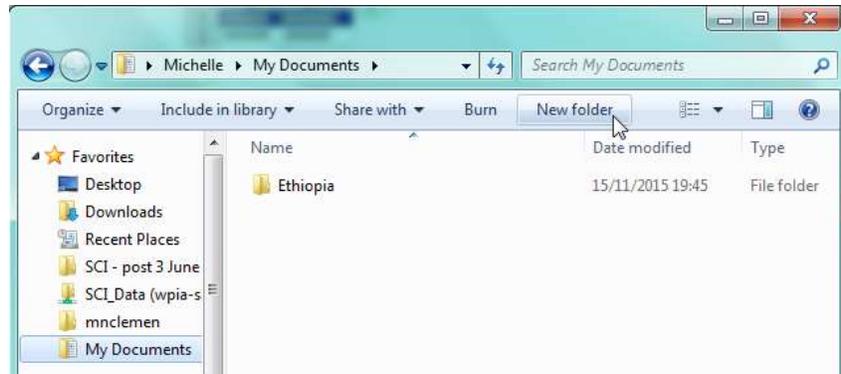
Pour commencer, nous allons créer les dossiers de travail :

1. Ouvrir l'Explorateur Windows en faisant un clic droit sur le bouton Windows en bas à gauche

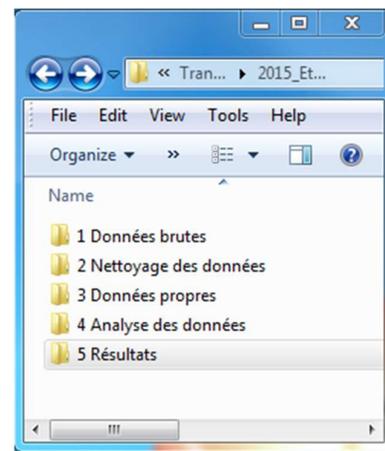
de l'écran  puis cliquer sur « Ouvrir l'Explorateur Windows »



2. Naviguer jusqu'à un emplacement où les données de cet exercice peuvent être enregistrées – par exemple « Documents » est un dossier qui convient mais tout autre emplacement où l'on peut facilement retrouver les données convient également.
3. Créer un nouveau dossier en appuyant sur le bouton « Nouveau dossier »



4. Nommer ce nouveau dossier « Atelier de nettoyage des données » ou autre nom pertinent
5. Dans « Atelier de nettoyage de données », créer des dossiers nommés :
 - 1 Données brutes
 - 2 Nettoyage des données
 - 3 Données propres
 - 4 Analyse des données
 - 5 Résultats
6. Vos dossiers doivent ressembler à ceci :

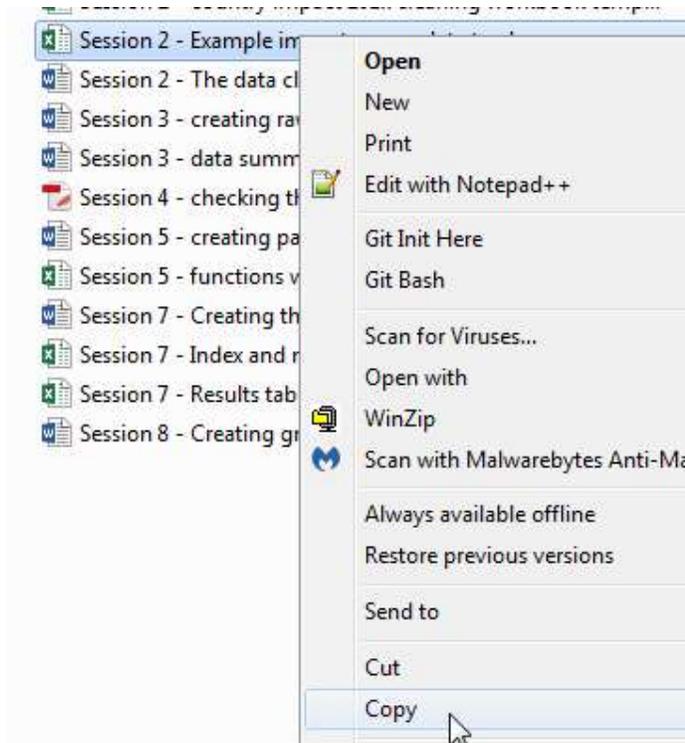


2.4 *Enregistrer des données brutes*

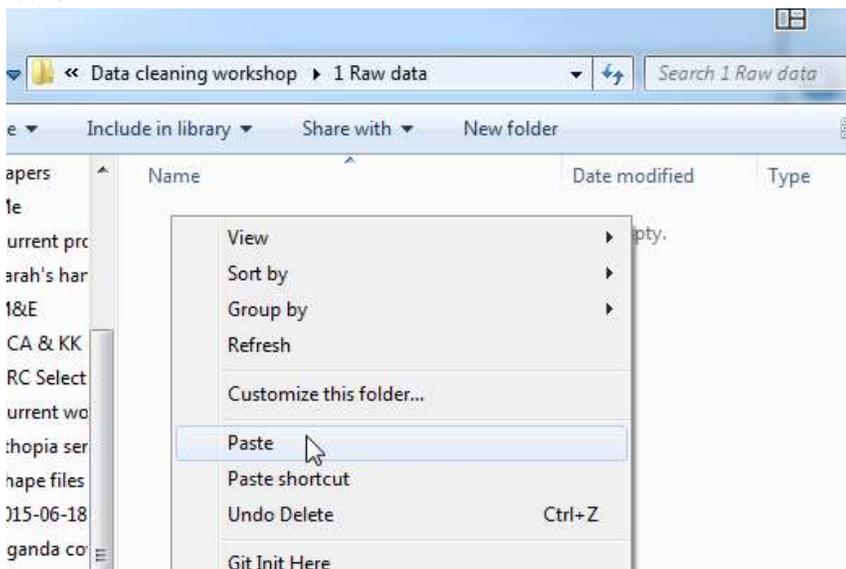
Chez SCI, nous veillons toujours à enregistrer une copie des données brutes reçues pour pouvoir les consulter ultérieurement si nécessaire. Après enregistrement, on ne touche plus à ces données.

Enregistrer les données brutes et le classeur de nettoyage des données :

1. Ouvrir une nouvelle fenêtre de l'explorateur (étape 1 précédente)
2. Naviguer vers le dossier de l'atelier sur votre disque dur ou la clé USB
3. Placer les deux fenêtres de l'explorateur l'une à côté de l'autre
4. Faire un clic droit sur les données appelées 'Session 2 – exemple données enquête impact à nettoyer' sur la clé et sélectionner 'copier'



5. Ouvrir '1 Données brutes' et coller le fichier dans le dossier en faisant un clic droit puis 'coller'

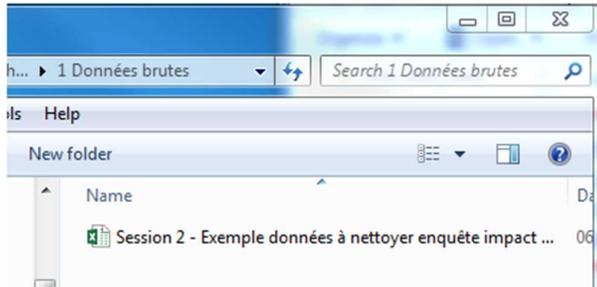


6. Vérifier que votre dossier '1 Données brutes' ressemble à ceci :

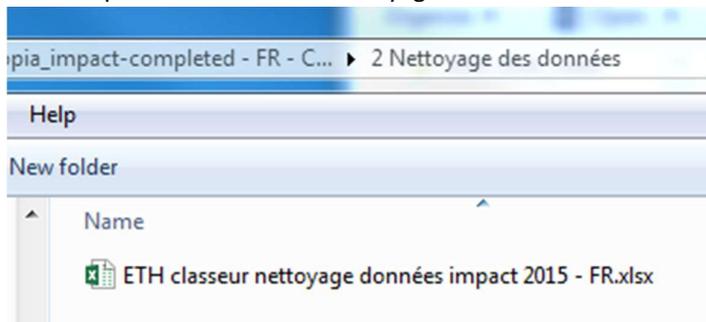
2.5 Enregistrer le classeur de nettoyage des données

1. Procéder comme précédemment pour copier le classeur de nettoyage des données dans le dossier '2 Nettoyage des données'

2. Vérifier que votre dossier '2 Nettoyage des données' ressemble à ceci :



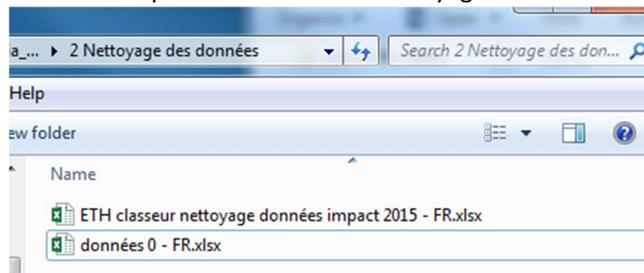
3. Renommer le 'Modèle classeur nettoyage des données' de façon plus précise, par exemple 'classeur nettoyage des données pays impact année', où l'on remplace pays et année par le pays et l'année concernés par les données recueillies. L'exemple de données contient des noms d'Éthiopie, mais les données sont mélangées. Nous appellerons ce classeur 'classeur nettoyage des données ETH impact 2015'
4. Vérifier que votre dossier '2 Nettoyage des données' ressemble à ceci :



2.6 Enregistrer les données de travail

À présent, il nous faut enregistrer les données brutes ailleurs de sorte à pouvoir travailler sur celles-ci sans modifier les données initiales.

1. Ouvrir les données brutes
2. Enregistrer les données brutes dans le dossier '2 Nettoyage des données' Tout d'abord, appuyer sur fichier – enregistrer sous
3. Puis enregistrer dans '2 Nettoyage des données' sous le nom 'données 0'
4. Fermer tous les fichiers sauf 'Classeur nettoyage des données ETH impact 2015' et 'données 0'
5. Vérifier que votre dossier '2 Nettoyage des données' ressemble à ceci :



2.7 Consulter le classeur de nettoyage des données

Ouvrir le classeur de nettoyage des données et le consulter. Vous verrez que les feuilles à compléter sont nombreuses. Ces feuilles servent à veiller à n'oublier aucune vérification. Il est également plus facile de retrouver les fichiers de nettoyage de données si tout est toujours formaté de manière similaire.

2.8 Consulter les données

Prendre un moment pour consulter les données 0. Que remarquez-vous ?

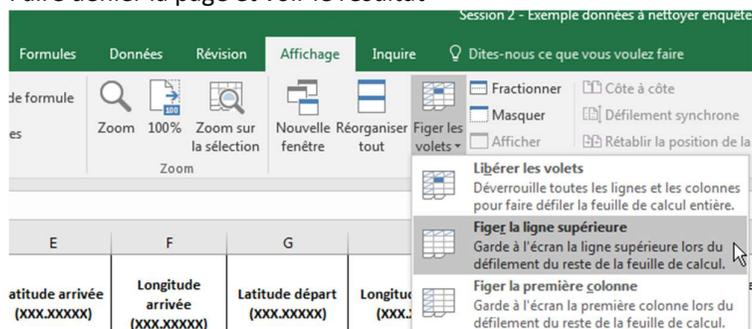
J'ai remarqué que :

- Le fichier contient trois feuilles
 - Deux feuilles contiennent des données
 - Une feuille contient le protocole et la description des variables
- Chaque feuille contient des données disposées en **lignes** et **colonnes**
- Chaque colonne n'apparaît qu'une seule fois et contient des informations sur une seule variable
- Chaque ligne concerne une école ou un élève différent(e). Toutes les informations sur cette école ou cet élève sont contenues dans cette ligne

2.9 Figurer les volets

Pour parcourir les données, il est pratique de toujours voir les noms des colonnes. Pour cela, il faut « figer les volets ».

1. Sélectionner la feuille 'école'
2. Cliquer sur Affichage -> Figer les volets -> Figer la ligne supérieure
3. Faire défiler la page et voir le résultat

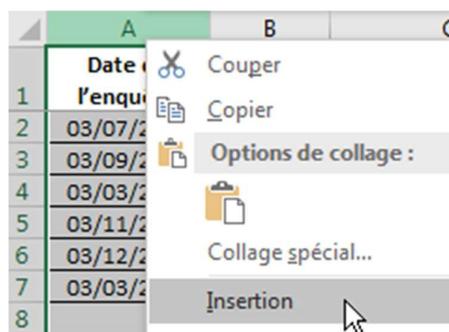


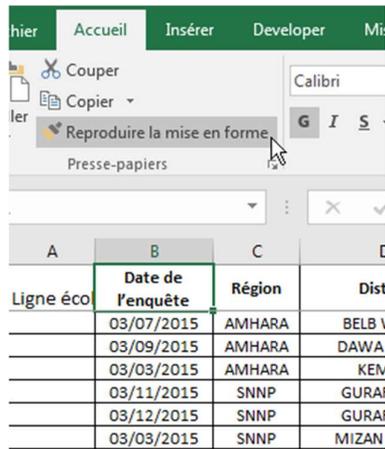
4. Faire de même pour la feuille 'élève'

2.10 Créer des numéros de ligne

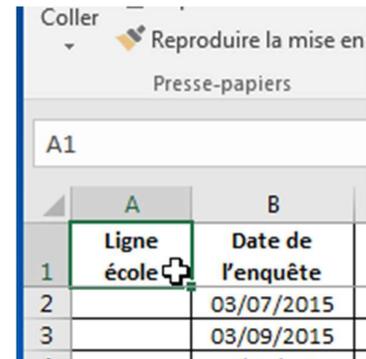
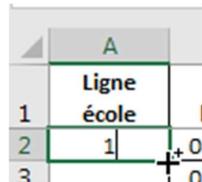
Créer des numéros de ligne peut être utile. Ainsi, vous pourrez toujours rétablir l'ordre initial des données. Pour créer des numéros de ligne dans l'ensemble de données école :

1. Prendre l'ensemble de données 'école'
2. Faire un clic droit sur la colonne A et sélectionner 'Insérer'
3. Saisir 'Ligne école' dans la cellule A1. Pour appliquer la même mise en forme que les autres cellules de la ligne 1 à A1, cliquer une fois sur la cellule B1 et sélectionner 'Reproduire la mise en forme'

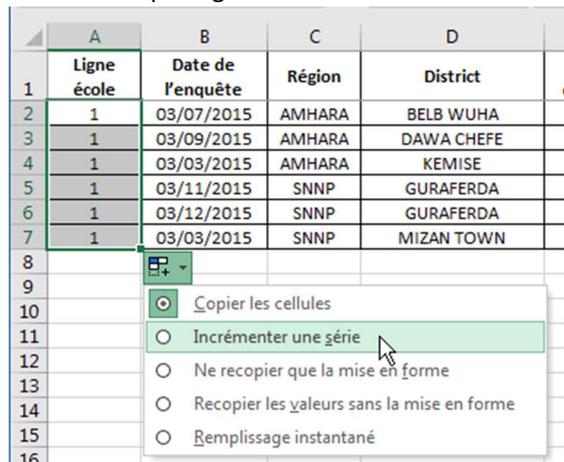




4. Cliquer une fois sur la cellule A1 pour copier uniquement la mise en forme dans cette colonne
5. Pour créer le premier numéro de ligne, saisir '1' dans la cellule A2
6. Passer sur la croix au coin inférieur droit de la cellule



7. Faire un double clic pour ajouter les numéros vers le bas
8. Si vous n'avez que des 1, cliquer sur les options de remplissage automatique et sélectionner Série de remplissage



9. Utiliser Ctrl-flèche bas pour aller en bas et vérifier que toutes les lignes possèdent un numéro différent. Pour reproduire la mise en forme, copier la mise en forme d'une autre cellule de numéro à proximité.
10. Faire de même pour la feuille 'élève' et intituler la colonne 'Ligne élève'.

2.11 Vérifier les identifiants

Les identifiants (ID) ou codes sont très importants dans l'analyse des données. Chaque ligne d'un tableau de données doit toujours posséder un identifiant unique afin de pouvoir les différencier en permanence.

Quel est l'ID de :

- l'ensemble de données école ? _____
- l'ensemble de données élève ? _____

La première chose à faire consiste à vérifier que les identifiants sont uniques. Pour cela, on utilisera un tableau croisé dynamique.

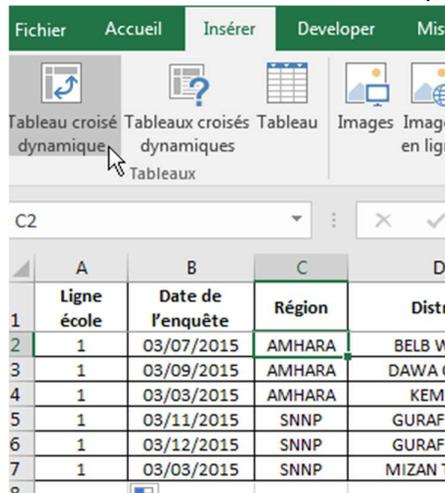
2.11.1 Tableaux croisés dynamiques

Les tableaux croisés dynamiques sont un moyen de synthétiser des données dans Excel. Ils sont très efficaces et peuvent être utilisés de nombreuses façons différentes. Pour créer un tableau croisé dynamique, il faut suivre plusieurs étapes. Nous allons faire beaucoup de tableaux croisés dynamiques dans ce cours, vous serez bientôt un expert !

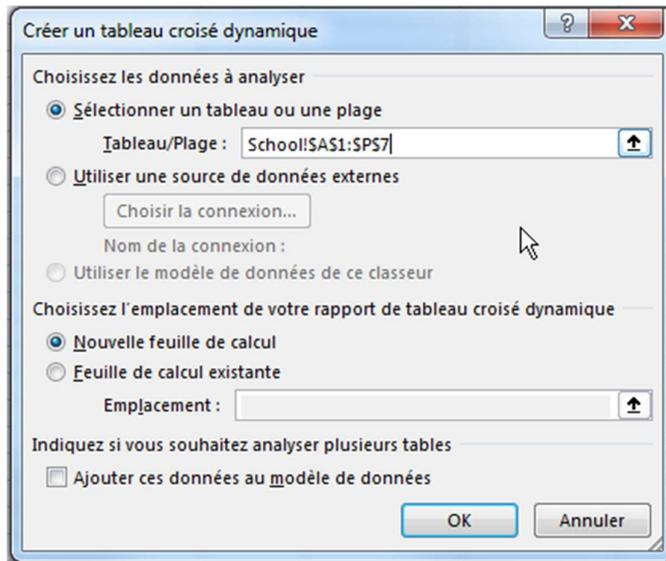
2.11.2 Vérifier que l'ID de l'école est unique.

Pour vérifier que l'ID de l'école est unique, nous créerons un tableau croisé dynamique contenant le nombre d'écoles avec chaque code d'école. Nous placerons ce tableau dans la feuille 'vérification ID élève' du classeur de nettoyage des données.

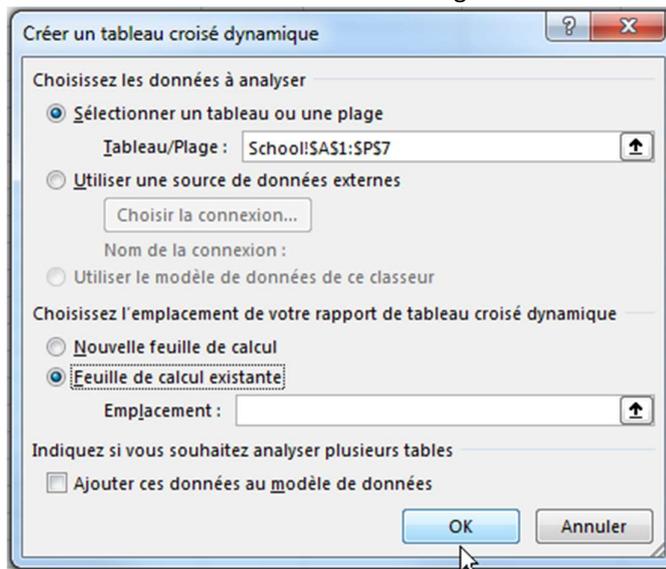
1. Prendre la feuille des écoles et cliquer une fois dans les données, par exemple la cellule C3
2. Aller sur Insertion -> Tableau croisé dynamique



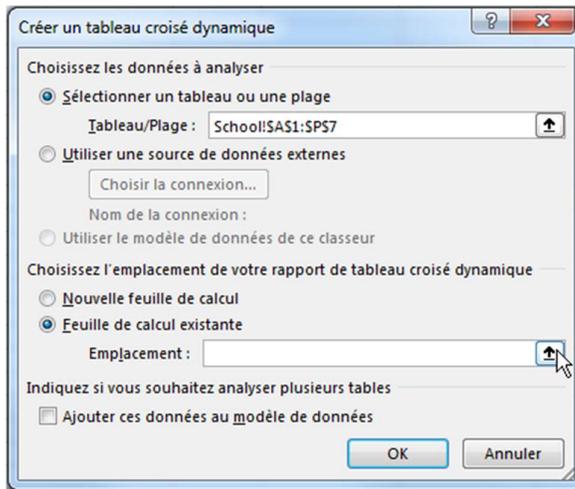
3. Vérifier que toutes les données sont incluses dans la section Tableau/Plage de la fenêtre de dialogue. *Cette étape est primordiale !!!* On doit y voir apparaître la feuille de données École puis les cellules A1 à P7.



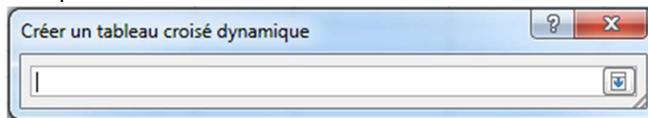
4. Nous allons placer le tableau dans 'Vérification ID école'. Pour cela, cliquer sur 'Feuille de calcul existante' dans la fenêtre de dialogue



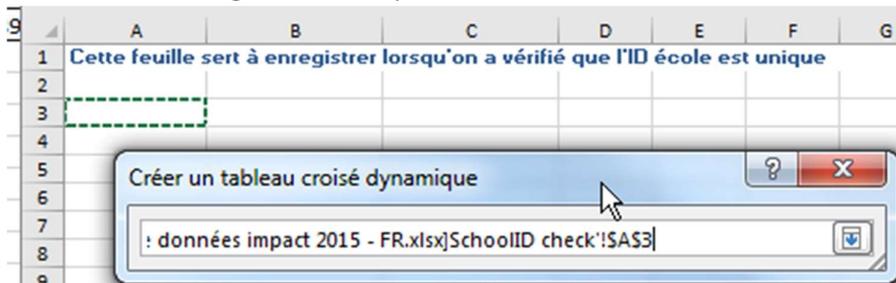
5. Ensuite, cliquer sur le petit bouton à droite du champ Emplacement



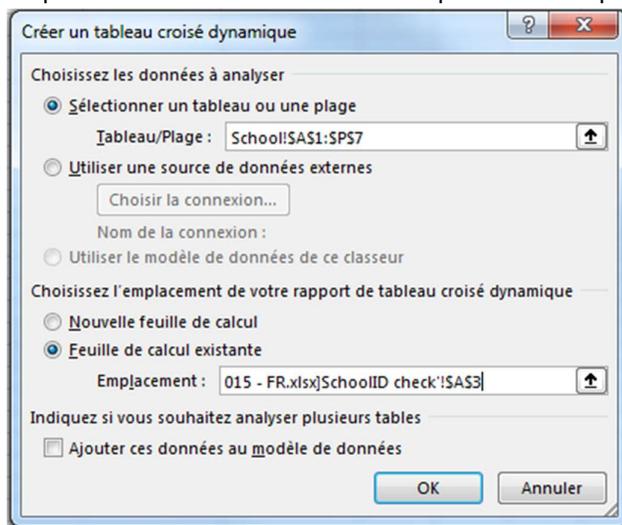
6. Une petite fenêtre comme celle-ci doit s'afficher



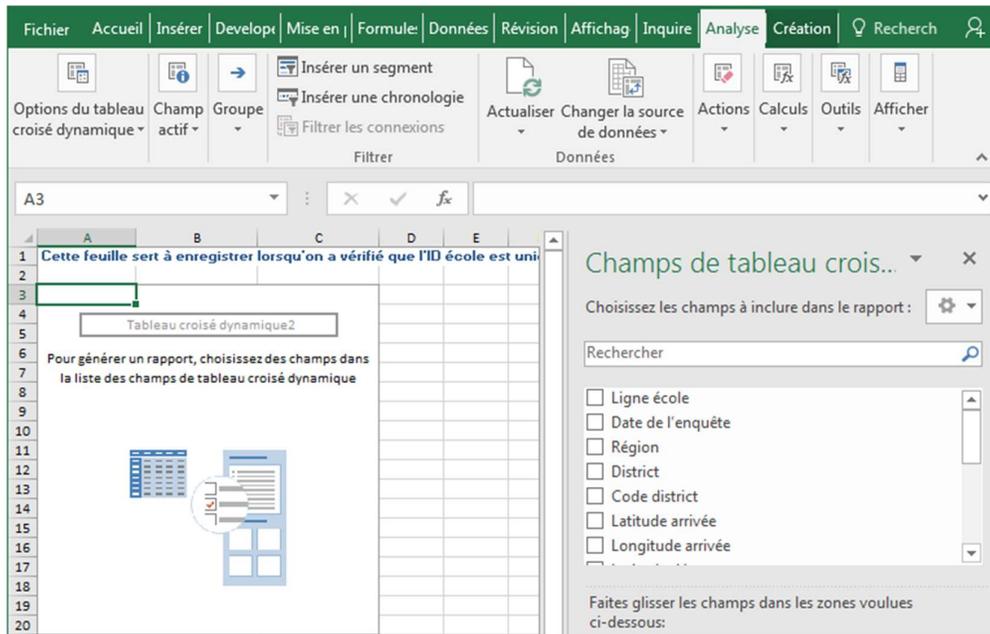
7. Cliquer sur le 'Classeur de nettoyage des données' puis aller sur la feuille 'Vérification ID école'. Cliquer sur la cellule A3. Vérifier que des pointillés entourent la cellule A3 et que la feuille et la cellule figurent dans la petite fenêtre comme ci-dessous



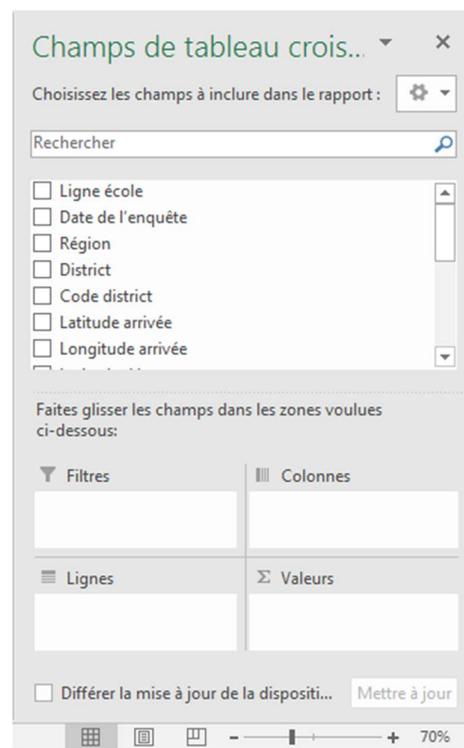
8. Cliquer sur le bouton à droite de cette petite fenêtre pour rouvrir la fenêtre précédente



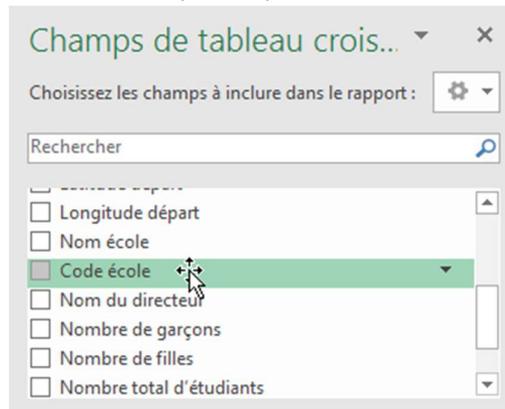
9. Cliquer sur 'OK' dans la fenêtre de dialogue
10. À présent, nous allons travailler dans le fichier 'modèle classeur de nettoyage des données'.
Il doit ressembler à ceci



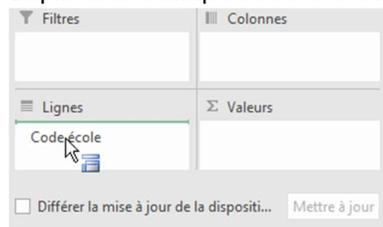
11. Cliquer dans la fenêtre du tableau croisé dynamique et une fenêtre de commande s'ouvrira à droite. La fenêtre de commande contient tous les noms de variable de l'ensemble de données école en haut et des cases vierges en bas.



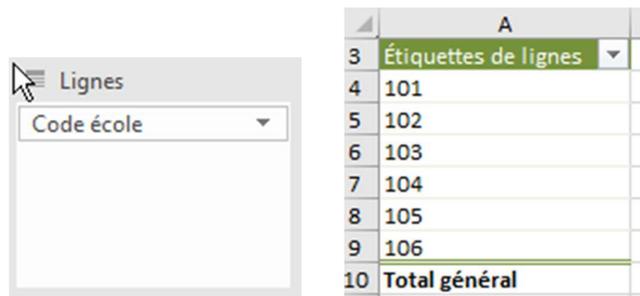
12. Nous souhaitons créer un tableau avec chaque Code d'école dans une ligne différente avec le nombre de fois où le code d'école est présent dans l'ensemble de données. Pour cela, passer le curseur sur Code d'école dans la fenêtre supérieure de sorte que la sélection soit en surbrillance puis cliquer et maintenir la souris.



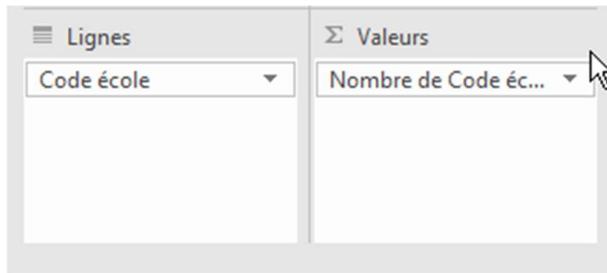
13. Déplacer le champ Code d'école vers les 'étiquettes de ligne' et relâcher la souris



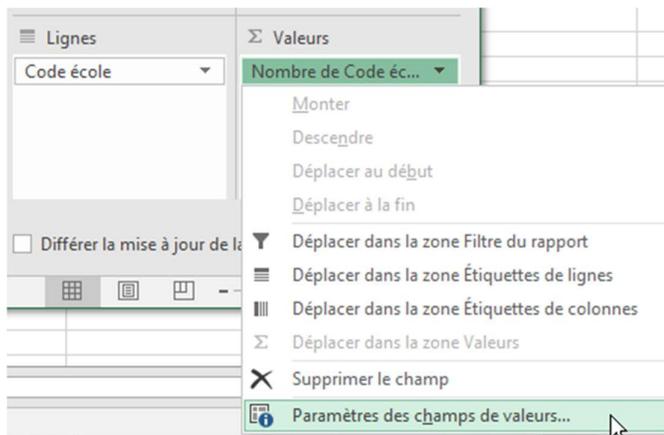
14. Vérifier que 'Code d'école' se situe dans la case Lignes et que les écoles apparaissent dans la colonne A de la feuille



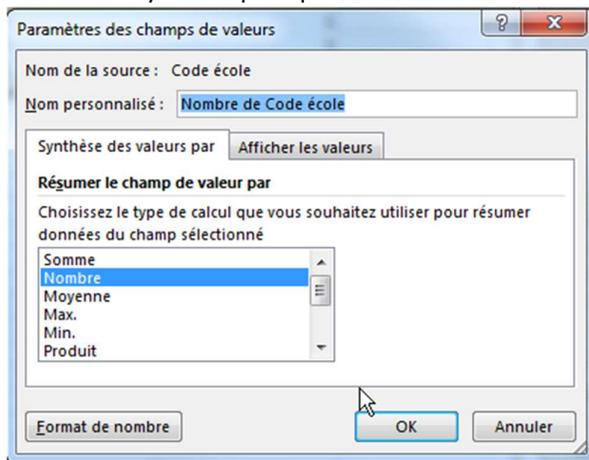
15. La colonne A contient tous les Codes d'école de l'ensemble de données École. À présent, il nous faut une autre colonne indiquant le nombre de fois où chaque Code d'école a été enregistré. Déplacer la Ligne école de la fenêtre en haut à droite pour qu'elle figure également dans le champ 'Valeurs'. Il est important de toujours utiliser une colonne dans laquelle il ne manque aucune valeur pour effectuer des décomptes dans les tableaux croisés dynamiques, sans quoi les valeurs manquantes peuvent poser des problèmes. Vérifier que votre fenêtre Étiquettes de ligne et Valeurs ressemble à ceci :



16. Pour l'instant, la fenêtre des valeurs affiche la somme de la Ligne école qui n'est pas correcte. Il nous faut regarder le décompte de plus près. Pour cela, cliquer sur la flèche 'bas' dans la fenêtre indiquant 'Somme des codes école' et cliquer sur 'Paramètres du champ valeur'.



17. Modifier la synthèse pour passer de 'somme' à 'nombre' et cliquer sur 'ok'



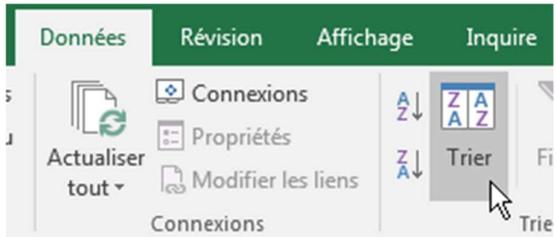
18. Consulter le tableau (colonnes A et B) pour vérifier qu'il ressemble au tableau ci-dessous. Le nombre de Codes école qui s'affiche principalement est 1, ce qui veut dire que ces Codes école n'ont été enregistrés qu'une seule fois, comme prévu. Toutefois, le Code école 104 a été enregistré deux fois et le Code école 105 n'a pas été enregistré du tout. Il nous faut regarder cela de plus près.

Étiquettes de lignes	Nombre de Code de l'école (XXX)
101	1
102	1
103	1
104	2
106	1
Total général	6

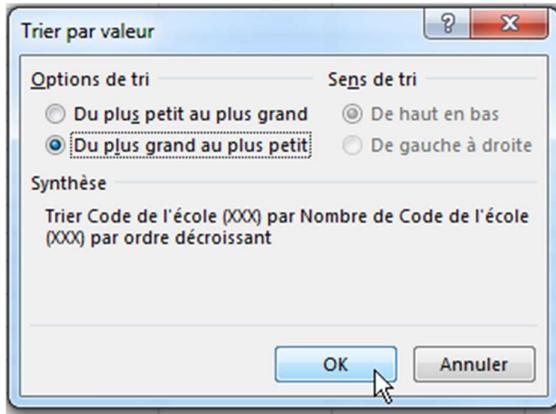
19. Si les écoles étaient nombreuses, on pourrait avoir du mal à voir que tous les chiffres ne sont pas des 1. On pourrait donc trier ces chiffres de sorte que les plus grands apparaissent en haut du tableau. Pour cela, cliquer dans la colonne nombre (n'importe quelle cellule entre B4 et B9)

Étiquettes de lignes	Nombre de Code de l'école (XXX)
101	1
102	1
103	1
104	2
106	1
Total général	6

20. Puis aller sur Données et cliquer sur Trier



21. Dans la fenêtre de dialogue, sélectionner 'Du plus grand au plus petit' puis OK



22. Vérifier que le tableau ressemble à ceci :

Étiquettes de lignes	Nombre de Code de l'école (XXX)
104	2
101	1
106	1
102	1
103	1
Total général	6

23. On voit clairement que le Code d'école 104 apparaît 2 fois. D'autres recherches dans les données révèlent que le Code école 105 manque et que 'ORENITA' aurait dû être l'école

104. Il faut apporter ces modifications dans les données. Naviguer vers 'nettoyage des données 1' puis modifier le Code d'école d'ORENITA pour 105. Vos données doivent ressembler à ceci :

J	K
Nom école	Code école
BELB W	101
BILLECHA	102
SERTIE	103
OTÉWA No 2	104
ORENITA	105
MESEGANA AKADAME	106

24. Ajouter une explication dans 'Vérification ID école' du classeur de nettoyage des données pour préciser quel changement a été effectué. Voici un exemple :

Étiquettes de lignes	Nombre de Code de l'école (XXX)	
101	1	L'école 104 est apparue deux fois
102	1	Une erreur a été faite avec l'école Orenita
103	1	Modification école Orenita 105
104	2	
106	1	
Total général	6	

2.11.3 Vérifier que l'ID de l'élève est unique.

L'ID de l'élève se compose du code du district, du code de l'école et d'un numéro unique pour chaque élève. Consulter l'onglet 'Protocole' de la feuille pour comprendre comment le Code élève a été créé.

Utiliser les étapes précédentes pour vérifier que les ID des élèves sont uniques. Saisir vos résultats dans l'onglet 'Vérifier ID élève' du classeur de nettoyage des données.

2.11.4 Enregistrer votre travail.

Enregistrer l'ensemble de données actuel sous le nom 'données 1' dans '2 nettoyage des données'. Ainsi, vous pourrez voir ce que vous avez modifié lors de ces étapes.

2.12 Relier les ensembles de données

La prochaine chose à faire consiste à vérifier que les ensembles de données sont correctement reliés. Les ensembles de données peuvent être reliés par une variable commune. Dans ce cas, la variable commune est le code de l'école. Cependant, le code de l'école n'est pas encore dans l'ensemble de données élèves. Nous devons le créer.

2.13 Créer un code d'école dans l'ensemble de données élèves

Nous allons créer un Code d'école dans le fichier élèves, en colonne T. Le Code élève est composé du district puis du numéro de l'école puis du numéro de l'élève. Il nous faut extraire les trois chiffres qui indiquent le numéro de l'école.

Pour extraire les chiffres 5 à 7, nous allons d'abord extraire les chiffres 1 à 7 puis prendre les 3 derniers chiffres. Pour cela, nous utiliserons des fonctions dans Excel. Les fonctions sont intégrées dans Excel et nous permettent de manipuler des données rapidement. Pour utiliser une fonction

dans Excel, vous devez utiliser le signe '=' au début de la cellule, de sorte qu'Excel sache qu'il s'agit d'une fonction. Pour extraire les chiffres de 5 à 7 dans le code élève :

1. Aller tout à droite de l'ensemble de données élèves (ctrl + flèche droite) pour voir la colonne T
2. Saisir 'Code école' dans la cellule T1 puis taper Entrée

	R	S	T
Numéro	Volume d'urine	S. haem Lame 1	Code école
	10	0	
	10	14	
		861	
	10	0	
	10	6	
	10	164	
	10	33	
	10	314	

3. Nous allons commencer par prendre les 7 premiers chiffres en partant de la gauche. Dans la cellule T2, écrire '=gauche('. À noter que 'GAUCHE' apparaît sous ce texte, ce qui montre qu'Excel a reconnu que la fonction d'extraction de caractères en partant de la gauche a été activée. La fonction attend à présent qu'on lui indique le texte où extraire les données (en gras) puis le nombre de caractères à extraire.

S	T	U	V
S. haem Lame 1	Code école		
0	=GAUCHE(
14	GAUCHE(texte, [no_car])		

4. Pour dire à Excel quel texte utiliser, cliquer sur la cellule B2 de la colonne Code élève sur la même ligne. Vérifier que la cellule B2 est entourée de poignées de la même couleur que dans la cellule T2, comme illustré ci-dessous.

B	T	U	V
Numéro d'identification (DDD.SSS.NNN)	Code école		
117.102.001	=GAUCHE(B2		
	GAUCHE(texte, [no_car])		

5. Ajouter ', 7)' pour dire à Excel de prendre les 7 premiers chiffres en partant de la gauche

T	U
Code école	
=GAUCHE(B2,7)	

6. Appuyer sur Entrée
7. Vérifier que la cellule T2 ressemble à ceci :

	T
em	Code
e 1	école
	117.102

8. On doit maintenant prendre les trois derniers chiffres en partant de la droite dans le code ci-dessus
9. Pour cela, on peut utiliser une autre fonction appelée 'STXT'
10. Faire un double clic sur la cellule T2 pour voir les formules. Ajouter du texte pour obtenir '=STXT(B2,5,3)' comme suit :

	T	U	V
em	Code		
e 1	école		
	=STXT(B2,5,3)		

STXT(texte, no_départ, no_car)

11. Vérifier que la cellule T2 ressemble à ceci :

	T
	Code
	école
	102

12. Faire un double clic sur la croix en bas à droite de la cellule T2 pour recopier vers le bas

	T
	Code
	école
	102

13. Aller jusqu'en bas (ctrl + flèche bas) pour vérifier que les toutes cellules s'affichent correctement.
14. Faire un double clic sur la cellule en bas et vérifier la formule pour voir si elle est correcte. Vérifier que la cellule T298 ressemble à ceci :

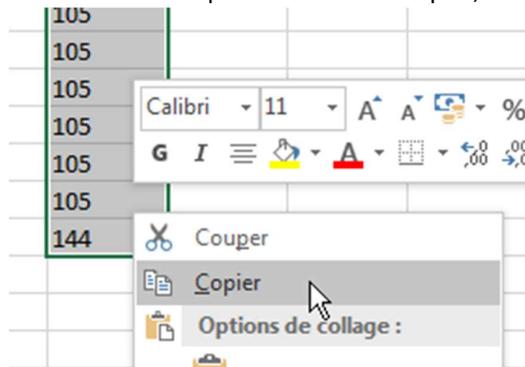
	105		
	105		
	=STXT(B298,5,3)		

STXT(texte, no_départ, no_car)

15. À présent, il nous faut retirer les formules mais conserver les nombres. Pour cela, on peut copier les données et coller uniquement les valeurs sans les formules. Aller en haut de l'ensemble de données et cliquer une fois sur T2 puis glisser vers le bas pour mettre en surbrillance ou appuyer sur ctrl + shift + flèche bas. Vérifier que les cellules jusqu'à la cellule T298 sont sélectionnées.

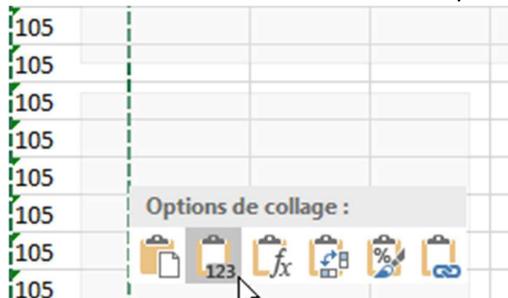
292	0	105
293	177	105
294	35	105
295	102	105
296	45	105
297	106	105
298	0	144

16. Faire un clic droit puis sélectionner copier, ou faire ctrl+c pour copier

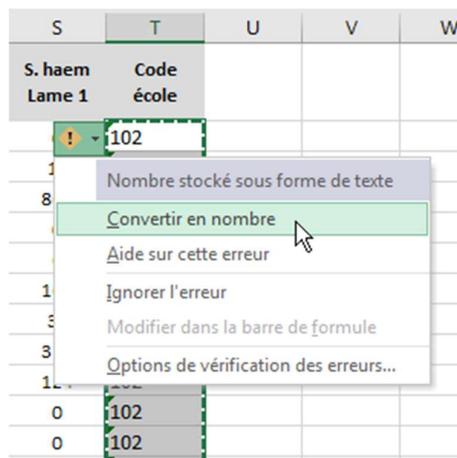


17. Veiller à être sur la cellule T2 en cliquant une fois sur la cellule T1 puis en appuyant sur la flèche bas.

18. Coller les valeurs en faisant un clic droit puis coller 123



19. Les cellules affichent à présent des triangles verts pour montrer qu'elles sont stockées sous forme de texte et non de nombre. Convertir le texte en nombre en cliquant sur le point d'exclamation et en sélectionnant 'Convertir en nombre'

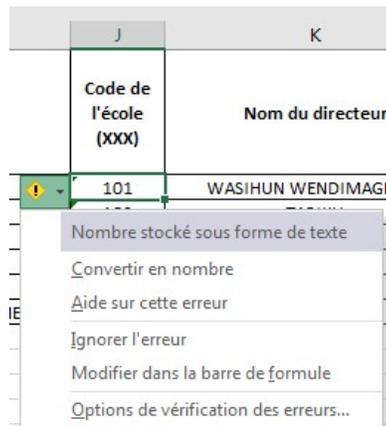


20. Les nombres seront à présent stockés uniquement comme des nombres. Pour vérifier ce point, faire un double clic sur l'une des cellules de la colonne T et voir si elle ne contient qu'un nombre.

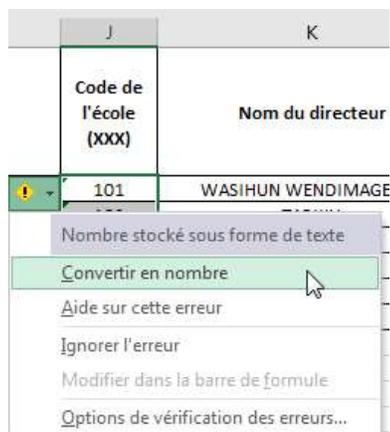
2.14 Vérifier que les mêmes codes sont contenus dans les ensembles de données écoles et élèves

Il nous faut vérifier si l'on a des informations provenant des mêmes écoles dans les deux ensembles de données. Pour cela, nous allons placer une liste des codes d'école de chaque fichier dans la même feuille Excel puis utiliser des fonctions pour les comparer.

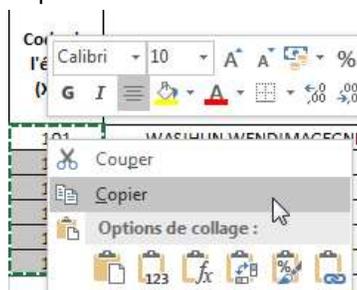
1. Pour commencer, copier les codes d'école de l'ensemble de données École dans la colonne 'ID école de l'ensemble de données école du fichier 'Vérifier ID élèves et écoles' du classeur de nettoyage des données. Pour cela, aller dans l'ensemble de données 'École' et trouver la colonne Code d'école. Des triangles verts apparaissent sur les nombres indiquant qu'Excel ne sait pas s'ils sont stockés correctement. Cliquer sur une cellule puis sur le point d'exclamation de cette cellule pour visualiser le problème.



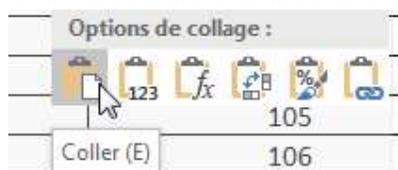
2. Les nombres sont stockés sous forme de texte, ce qui signifie qu'Excel ne pourra pas les trier correctement. Pour y remédier, mettre en surbrillance toutes les cellules concernées en cliquant sur la première cellule du haut puis en descendant. Vous pouvez également cliquer sur la première cellule (K2) puis appuyer sur ctrl + shift + flèche bas pour sélectionner toutes les cellules situées en dessous. Les cellules en surbrillance doivent apparaître comme ci-après. À présent, cliquer sur le point d'exclamation de la première cellule puis sélectionner 'Convertir en nombre'



3. Les triangles verts doivent avoir disparu de toute les cellules.
4. Il nous faut désormais copier les Codes d'école de l'ensemble de données École dans la colonne 'ID école de l'ensemble de données école du fichier 'Vérifier ID élèves et écoles' du classeur de nettoyage des données. Pour cela, veiller à ce que l'ensemble des Code d'école dans l'ensemble de données École soient toujours en surbrillance. Faire une copie avec un clic droit puis Copier, ou ctrl+c.
5. Copier les données en les sélectionnant, puis clic droit et 'copier', ou ctrl+c.



6. Naviguer dans la feuille 'Vérification des ID élève et école' du classeur de nettoyage des données et cliquer une fois sur la cellule A6 de la colonne 'ID école dans l'ensemble de données école'. Coller les données ici avec un clic droit puis 'coller', ou ctrl+v



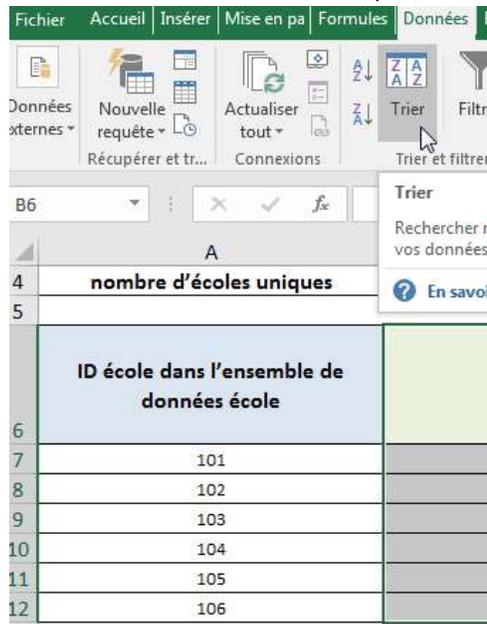
7. Vérifier que les données collées ressemblent à ceci :

ID école dans l'ensemble de données école
101
102
103
104
105
106

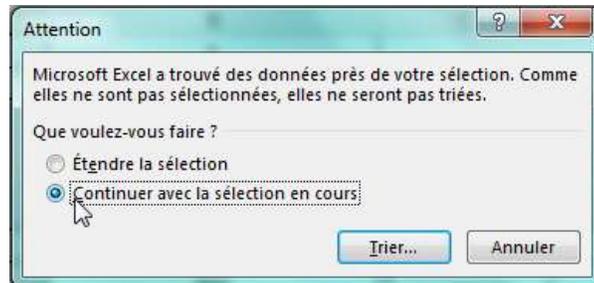
8. Compléter les informations en haut de la page pour montrer les ensembles de données que nous utilisons et le nombre d'écoles que l'on prévoit de recevoir du protocole

nom ensemble de données école	école
nom ensemble de données élèves	élève
nombre d'écoles uniques	6

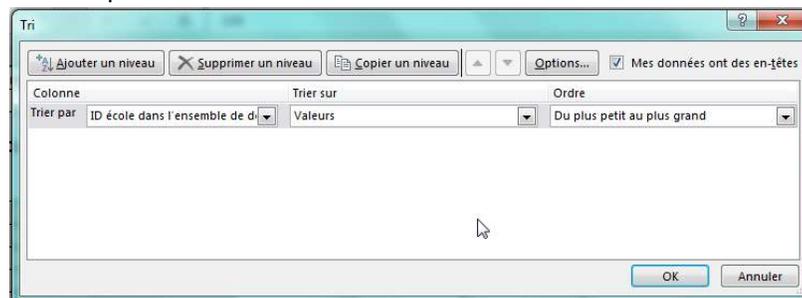
9. Vérifier que les Codes école sont triés correctement en sélectionnant toutes les données de la colonne de la cellule A6 à A12 puis aller dans Données -> Trier



10. Si un avertissement s'affiche, sélectionner Continuer avec la sélection actuelle et appuyer sur Trier



11. Vérifier que vous trie bien les ID d'école dans l'ensemble de données école puis OK



12. À présent, il nous faut importer les codes école depuis l'ensemble de données élève. Pour cela, mettre en surbrillance toutes les données entre la cellule T2 et la cellule T298 avec ctrl + shift + flèche bas puis faire ctrl + v

	S	T
288	171	105
289	2	105
290	28	105
291	642	105
292	0	105
293	177	105
294	35	105
295	102	105
296	45	105
297	106	105
298	0	144

13. Coller les données dans la colonne 'ID école dans ensemble de données élèves' de la feuille 'Vérification ID élève et école' du classeur de nettoyage des données avec ctrl + v

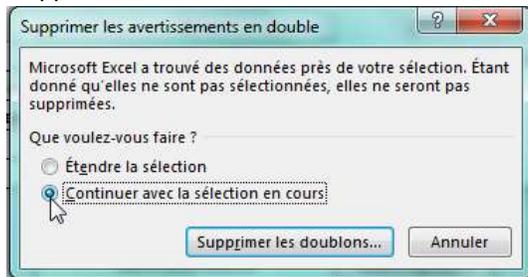
ID école dans l'ensemble de données élève	
	102
	102
	102
	102
	102
	102

14. Supprimer les doublons de codes école en sélectionnant Données -> Supprimer les doublons

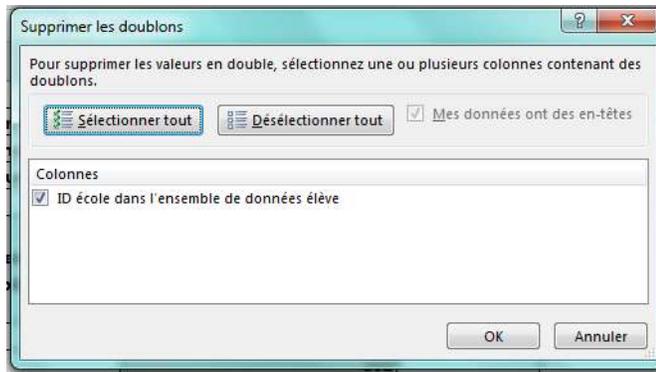
The screenshot shows the 'Supprimer les doublons' (Remove Duplicates) button on the 'Outils de données' (Data Tools) ribbon. Below it, a dialog box provides instructions: 'Supprimer les doublons' - 'Supprimer les lignes en double dans une feuille de données.' It also notes that users can select columns to verify for duplicates.

ID école dans l'ensemble de données élève	Les ID école sont-ils identiques ? (O/N)	Remarques

15. Si un avertissement s'affiche, sélectionner 'Continuer avec la sélection actuelle' puis 'supprimer les doublons'



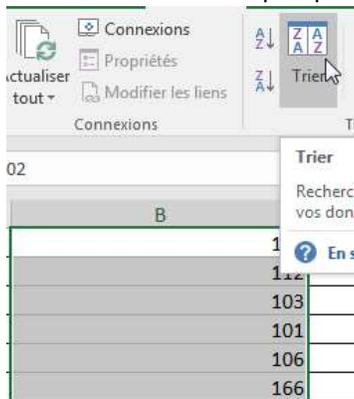
16. Vérifier que seul 'ID école dans ensemble de données élève' est sélectionné et appuyer sur OK



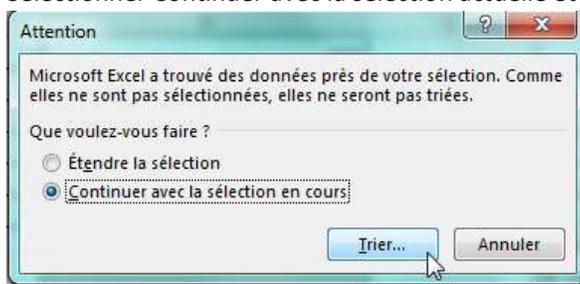
17. Vérifier la fenêtre de notification pour savoir combien il reste de valeurs uniques. Est-ce ce que l'on attend ?



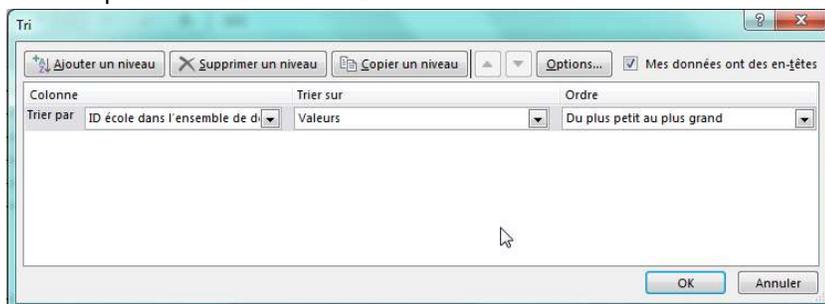
18. Trier les ID d'école uniques parmi les élèves en sélectionnant Données - Trier



19. Sélectionner Continuer avec la sélection actuelle et appuyer sur Trier



20. Vérifier que vous triez bien les ID d'école dans l'ensemble de données élèves puis OK



21. Observer la liste des Codes école de chaque ensemble de données. Sont-ils identiques ? On peut créer une fonction dans la colonne C qui indique O lorsque les codes sont identiques et N s'ils ne le sont pas. Pour cela, cliquer sur la cellule C7

ID école dans l'ensemble de données école	ID école dans l'ensemble de données élève	Les ID école sont-ils identiques ? (O/N)
101	101	

22. Saisir dans la cellule C7 '=SI(A7=B7, "O", "N")' puis Entrée

ID école dans l'ensemble de données école	ID école dans l'ensemble de données élève	Les ID école sont-ils identiques ? (O/N)
101	101	=SI(A7=B7, "O", "N")
102	102	

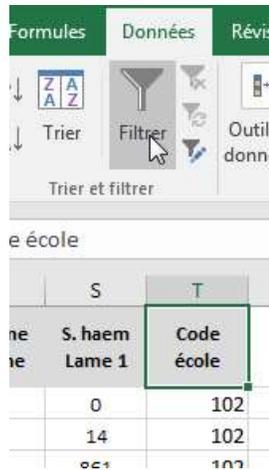
23. Faire un double clic sur la flèche en bas à droite de la cellule pour copier vers le bas

ID école dans l'ensemble de données élève	Les ID école sont-ils identiques ? (O/N)
101	S
102	

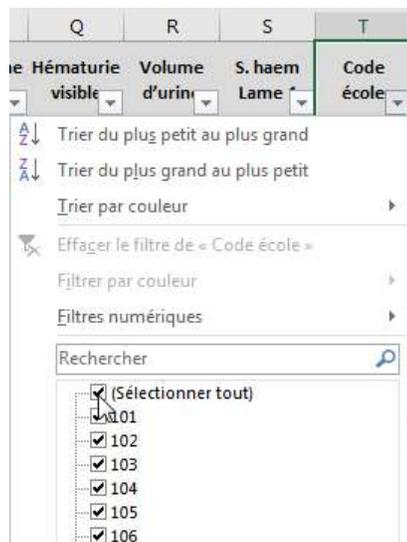
24. On voit alors quelles lignes indiquent un N signifiant que les ID ne sont pas identiques.

ID école dans l'ensemble de données école	ID école dans l'ensemble de données élève	Les ID école sont-ils identiques ? (O/N)
101	101	S
102	102	S
103	103	S
104	104	S
105	105	S
106	106	S
	112	N
	144	N
	166	N

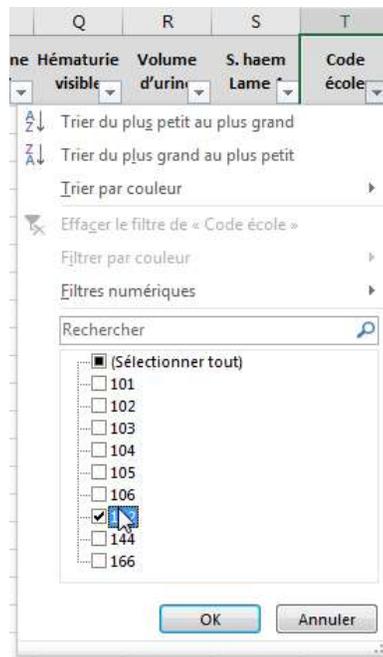
25. Toutes les écoles incluses dans l'ensemble de données école ont été également incluses dans l'ensemble de données élève. Néanmoins, certaines écoles de l'ensemble de données élèves n'ont aucune correspondance dans l'ensemble de données école. Il nous faut regarder ça de plus près pour savoir ce qu'il s'est passé. Nous allons appliquer un filtre pour visualiser les lignes qui ne sont pas identiques. Reprendre l'ensemble de données élèves et cliquer sur la ligne du haut de l'ensemble de données puis sélectionner Données - Filtrer.



26. Une petite flèche vers le bas apparaît désormais dans toutes les cellules pour indiquer que des filtres existent. Cliquer sur la flèche du filtre dans Code école de la colonne T. Cliquer sur Tout sélectionner ou tout désélectionner.



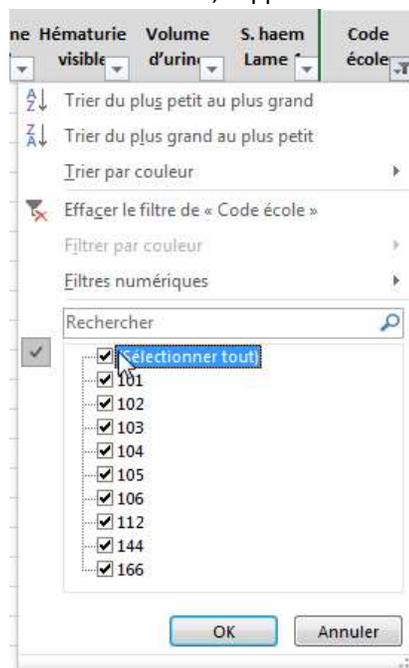
27. Puis cliquer sur 112 pour sélectionner uniquement les lignes où le code d'école est 112



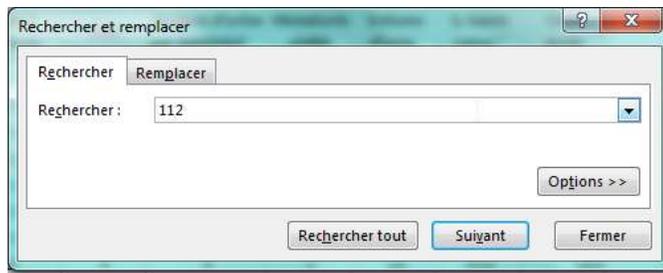
28. On voit à présent qu'il n'existe qu'une seule ligne avec le Code école 112

L	M	N	O	P	Q	R	S	T
Lame 2 Ascari	Lame 2 Ankylostomose	Lame 2 Trichuris	CCA	Analyse d'urine par bandelette	Hématurie visible	Volume d'urine	S. haem Lame	Code école
0		0	2	3		10	12	112

29. Il nous faut désormais regarder plus en détail l'élève dont le Code école est 112. Retourner à la fenêtre du filtre, supprimer les filtres en cliquant sur '(sélectionner tout)' puis OK



30. On recherche donc le code 112 dans la colonne T. Tout d'abord, mettre en surbrillance la colonne T en cliquant sur le T en haut puis appuyer sur ctrl+f pour 'find' (trouver). Ajouter 112 dans la fenêtre de recherche puis appuyer sur 'Suivant'



31. Le seul élève dont le Code école est 112 se trouve en ligne 51

49	102
50	102
51	112
52	103
53	103

32. Faire défiler sur la gauche pour voir le Code élève et le Nom de l'école. On dirait que l'élève appartient à l'école 102, Billecha.

33. Le Code élève termine par 016. Faire défiler vers le haut pour vérifier s'il existe un Code élève 117.102.116

49		117.102.049	03/09/2015	BILLECHA
50		117.102.050	03/09/2015	BILLECHA
51		117.112.016	03/09/2015	BILLECHA
52		118.103.001	03/03/2015	SERTIE
53		118.103.002	03/03/2015	SERTIE
54		118.103.003	03/03/2015	SERTIE

34. Ce Code élève n'apparaît pas dans le fichier, ce qui suggère que le Code élève 117.102.116 a été remplacé par 117.112.116 par erreur lors de la saisie. Si vous disposiez des documents papier, vous pourriez les consulter pour voir si les données du fichier et du document correspondent. Pour l'instant, nous supposons que le Code école devrait être 102 et que le Code élève devrait être 117.102.116.

35. Modifier le Code élève dans la Ligne élève 51 pour obtenir 117.102.116

	Numéro d'identification (DDD.SSS.NN)	Date de l'enquête (JJ-MM-AAAA)	Nom de l'école
1			
47	117.102.047	03/09/2015	BILLECHA
48	117.102.048	03/09/2015	BILLECHA
49	117.102.049	03/09/2015	BILLECHA
50	117.102.050	03/09/2015	BILLECHA
51	117.102.016	03/09/2015	BILLECHA
52	118.103.001	03/03/2015	SERTIE
53	118.103.002	03/03/2015	SERTIE
54	118.103.003	03/03/2015	SERTIE

36. Modifier le Code école dans la Ligne élève 50 pour obtenir 102

37. Ajouter des informations dans les Remarques de 'Vérification ID élève et école' pour décrire les modifications apportées. Si vous souhaitez créer une nouvelle ligne de texte dans la même case, utiliser alt + entrée.

	ID école dans l'ensemble de données école	ID école dans l'ensemble de données élève	Les ID école sont-ils identiques ? (O/N)	Remarques
6				
7	101		101 S	
8	102		102 S	
9	103		103 S	
10	104		104 S	
11	105		105 S	
12	106		106 S	
				Un élève avec le code école 112 aurait dû avoir le code 102
13			112 N	Modification code élève 117.102.116 et code école 102
14			144 N	
15			166 N	

38. Suivre des étapes similaires pour déterminer quoi faire des Codes école 114 et 166 en veillant à consigner les modifications effectuées

Les ID doivent à présent être corrects. Maintenant que nous savons à quelle école appartient chaque élève, nous allons consulter les données recueillies qui seront abordées à la prochaine session.

2.14.1 Enregistrer votre travail.

Enregistrer l'ensemble de données actuel sous le nom 'données 2' dans '2 nettoyage des données'. Ainsi, vous pourrez voir ce que vous avez modifié lors de ces étapes.

Session 3 Synthèses de données initiales

Il est capital d'utiliser une procédure systématique pour nettoyer les données. Cela permet de suivre précisément les éventuelles modifications apportées et permet aux autres de comprendre comment vous êtes passé des données brutes aux données nettoyées que l'on utilisera pour l'analyse.

Pour faire suite à la section précédente, maintenant que vous avez enregistré une copie des données d'origine et vérifié les ID, nous allons produire des synthèses des données brutes. Cela vous permettra de comprendre les données et d'organiser et stocker les éventuelles modifications apportées.

Dans cette optique, une série d'étapes va vous être présentée. Vous pouvez suivre ces étapes en utilisant les données d'entraînement fournies.

NB Nous utiliserons le 'classeur de nettoyage de données' et 'données 2' pour débiter cette session.

3.1 Procédure de production des synthèses de données

Le Classeur de nettoyage des données contient des modèles de production des synthèses de données dans les feuilles intitulées « Synthèse école brute » et « Synthèse élève brute ».

NB Vous pouvez copier et renommer les feuilles si vous disposez d'un plus grand nombre d'ensembles de données.

Ici nous allons montrer comment compléter la feuille de synthèse à l'aide des données d'école.

Le processus :

1. Commencer par saisir les informations basiques de l'ensemble de données en haut à gauche de la feuille de calcul.

	nom de l'ensemble de données	
2	Nombre de lignes	
3	Nombre de colonnes	
4	Identifiant unique	

2. Pour compter rapidement le nombre de lignes, mettre en surbrillance la première colonne de données et noter le nombre qui s'affiche en bas à droite de la page :

Ligne école	Date de l'enquête	Région	District	Code district	Latitude arrivée	Longitude arrivée	Latitude départ	Longitude départ
1	03/07/2015	AMHARA	BELB WUHA	120	12.18.550	037.45.04	12.118.542	037.45.075
2	03/09/2015	AMHARA	DAWA CHEFE	117	10.46.084	039.51.800	10.46.084	039.05.799
3	03/03/2015	AMHARA	KEMISE	117	10.52.371	039.51.607	10.52.370	039.05.670
4	03/11/2015	SNNP	GURAFERDA	123	06.83.725	035.29.902	06.83.725	35.29.872
5	03/12/2015	SNNP	GURAFERDA	123	06.86.817	035.34.801	06.87.925	035.25.734
6	03/03/2015	SNNP	MIZAN TOWN	122	06.99.501	035.58.899	06.99.506	035.58.890

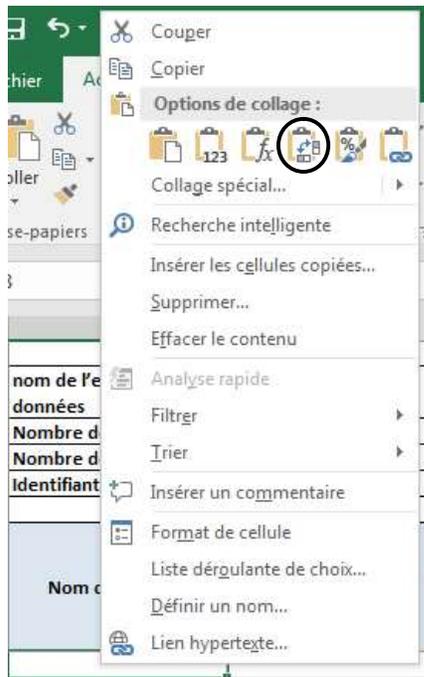
- Vous pouvez compter le nombre de variables de façon similaire, mettre en surbrillance la première colonne de données et noter le nombre qui s'affiche en bas à droite de la page :

Longitude départ	Nom école	Code école	Nom du directeur	Nombre de garçons	Nombre de filles
037.45.075	BELB W	101	WASHUN WENDIMAGEGNE	129	104
039.05.799	BILLECHA	102	TARIKU	1002	137
039.05.670	SERTIE	103	REGASSA WAKIYA	78	127
35.29.872	OTAWA No 2	104	Mengistu Erose	112	
035.25.734	ORENITA	105	AYENEV BERELE	146	108
035.58.890	MESEGANA AKADAME	106	TATEK KEBEDE	99	116

- Saisir le nom de la variable de la colonne d'identifiant unique dans la case d'identifiant unique, dans ce cas il s'agit de « Code école (XXX) ».
- Noms de variable** – Copier les en-têtes de variables de l'ensemble de données vers la feuille de synthèse brute :
Mettre en surbrillance la ligne qui contient les noms de variable et copier (ctrl+c ou clic droit puis copier).

Région	District	Code école	Longitude arrivée	Latitude départ	Longitude départ	Nom de l'école	Code de l'école	Nom du directeur	Nombre des filles	Nombre des garçons	Nombre total des élèves	Les élèves de votre établissement ont-ils reçu un traitement vermifuge l'année dernière (0 = N, 1 = O)
AMHARA	BELB WUHA	120	12.18.550	037.45.04	12.118.542	BELB W	101	WASHUN WENDIMAGEGNE	129	104	233	0
AMHARA	DAWA CHEFE	117	10.46.084	039.51.800	10.46.084	BILLECHA	102	TARIKU	1002	137	239	1
AMHARA	KEMISE	117	10.52.371	039.51.607	10.52.370	SERTIE	103	REGASSA WAKIYA	78	127	205	0
SNNP	GURAFERDA	123	06.83.725	035.29.902	06.83.725	OTAWA No 2	104	Mengistu Erose	112		112	1
SNNP	GURAFERDA	123	06.86.817	035.34.801	06.87.925	ORENITA	105	AYENEV BERELE	146	108	254	1
SNNP	MIZAN TOWN	122	06.99.501	035.58.899	06.99.506	MESEGANA AKADAME	106	TATEK KEBEDE	99	116	215	1

Faire un clic droit sur la cellule ci-dessous « Nom de variable » dans la feuille de synthèse (A8) et dans « Options de collage » sélectionner l'option Transposer.



6. **Facteur ou nombre ?** – Cette colonne doit être remplie à la main. En général, une variable est un facteur si elle possède un choix limité d'options pour classer quelque chose (par ex., oui/non ou fille/garçon) et un nombre sert à quantifier des données types (par ex., œufs de *S. mansoni*).

7.

Facteur : nombre de niveaux Nombre : minimum	Facteur : laisser vierge. Nombre : moyenne	Facteur : laisser vierge. Nombre : maximum
---	---	---

Prendre chaque variable de votre ensemble de données et produire des synthèses des informations qu'elles contiennent.

Si la variable est un facteur, remplir uniquement la première de ces trois colonnes : « Facteur : nombre de niveaux ». S'il s'agit d'un nombre, remplir les trois colonnes, minimum, moyenne et maximum.

Niveaux

Pour trouver le nombre de niveaux, utiliser un tableau croisé dynamique :
Mettre en surbrillance vos données dans chaque colonne, sélectionner

l'onglet Insertion dans la barre d'outils en haut et cliquer sur « Tableau croisé

Date de l'enquête	Région	District	Code du district (XXX)	Latitude arrivée (XXX.XXXXX)	Longitude arrivée (XXX.XXXXX)	Latitude départ (XXX.XXXXX)	Longitude départ (XXX.XXXXX)	Nom de l'école	Code de l'école (XXX)	Nom du directeur	Nombre des filles	Nombre des garçons	Nombre total des élèves	Les élèves de votre établissement ont-ils reçu un traitement vermifuge l'année dernière (O = N, 1 = O)?
03/07/2015	AMHARA	REB WUHA	120	12.18.550	037.45.04	12.118.542	037.45.075	REB W	101	WASHUN WENDIMAGEGNE	129	104	233	O
03/09/2015	AMHARA	DAWA CHEFE	117	10.46.084	039.51.800	10.46.084	039.05.799	BILLECHA	102	TARIKU	1002	197	239	I
03/03/2015	AMHARA	KEMISE	119	10.52.371	039.51.607	10.52.370	039.05.870	SERTIE	103	REGASSA WAKIYA	78	127	205	O
03/11/2015	SNMP	GURAFERDA	123	06.88.725	035.28.902	06.88.725	35.29.872	OTENA No 2	104	Mengistu Fisse	112	112	224	I
03/12/2015	SNMP	GURAFERDA	123	06.88.817	035.34.801	06.87.925	035.25.734	ORIENTA	104	AYENEW BERELE	146	108	254	I
03/03/2015	SNMP	MIZAN TOWN	122	06.99.501	035.58.899	06.99.506	035.58.890	MESEGANA AKADAME	106	TATEX KEBEDE	99	116	215	I

dynamique » puis « OK » (normalement c'est la première option) :

Vous obtiendrez une nouvelle feuille de calcul qui ressemble à ceci :

Options du tableau croisé dynamique

Champ actif

Groupe

Insérer un segment

Insérer une chronologie

Actualiser

Changer la source de données

Actions

Calculs

Outils

Afficher

Tableau croisé dynamique1

Pour générer un rapport, choisissez des champs dans la liste des champs de tableau croisé dynamique

Champs de tableau croisé...

Choisissez les champs à inclure dans le rapport :

Rechercher

- Date de l'enquête
- Région
- District
- Code du district (XXX)
- Latitude arrivée (XXX.XXXXX)
- Longitude arrivée (XXX.XXXXX)
- Latitude départ (XXX.XXXXX)
- Longitude départ (XXX.XXXXX)
- Nom de l'école
- Code de l'école (XXX)
- Nom du directeur

Faites glisser les champs dans les zones voulues ci-dessous:

Filtres

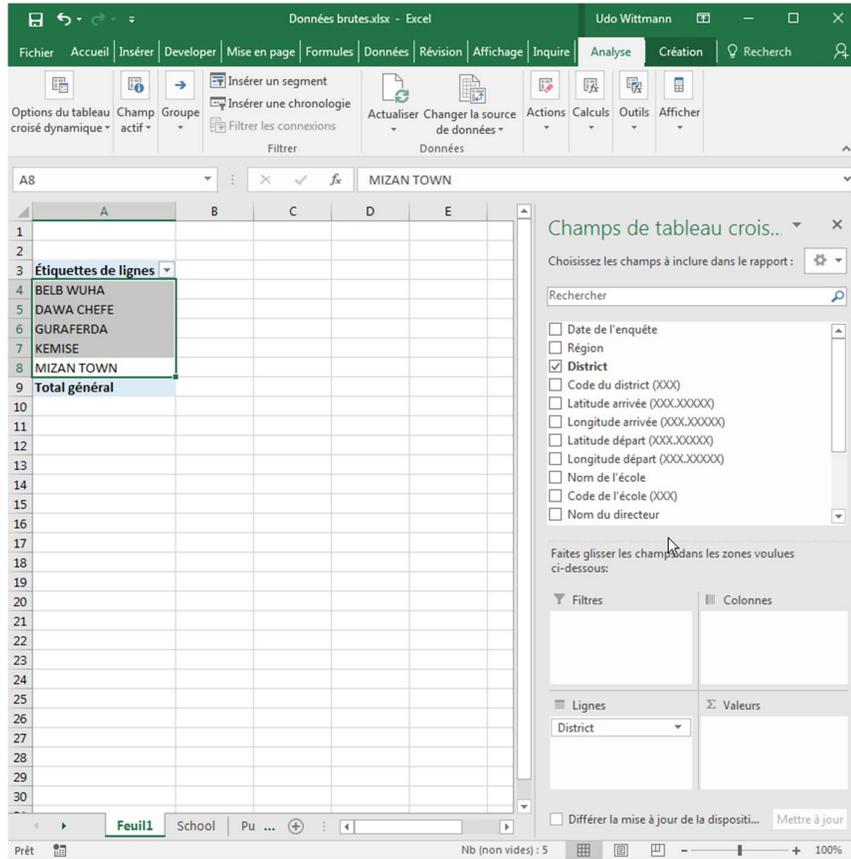
Colonnes

Lignes

Valeurs

Différer la mise à jour de la dispositi... Mettre à jour

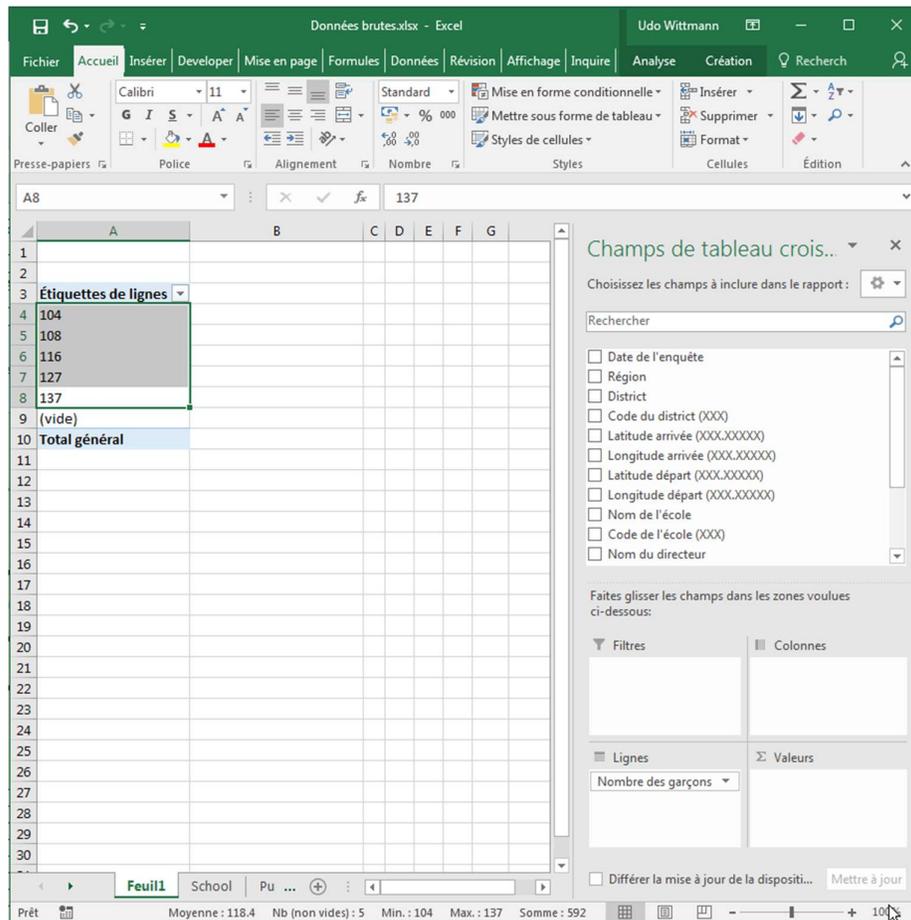
Pour voir combien de valeurs uniques existent pour chaque variable, sélectionner la variable qui vous intéresse et la faire glisser dans la fenêtre « Étiquettes de ligne ». Les valeurs uniques seront listées dans la zone de la feuille étiquetée « Tableau croisé dynamique 1 ». Vous pouvez ensuite mettre en surbrillance les valeurs uniques et noter le nombre qui s'affiche en bas à droite de la page :



Remplir la colonne appropriée de la feuille de synthèse des données. Le nombre de valeurs uniques est-il celui prévu ?

Comptes

Évaluer la variable numérique de la même manière. Sélectionner la variable et la faire glisser dans la fenêtre « Étiquettes de ligne ». Mettre en surbrillance les valeurs indiquées dans la zone Tableau croisé dynamique. Les informations en bas à droite de la page doivent afficher Moyenne, Compte (le nombre de valeurs est-il celui prévu ?), Min et Max :



Vous pouvez utiliser les valeurs en surbrillance pour remplir la feuille de synthèse des données. Utiliser la valeur Compte pour vérifier s'il manque des valeurs.

[Avant l'étape suivante, nous prendrons les données d'école et compléterons le tableau de synthèse.]

8. Nombre de valeurs manquantes

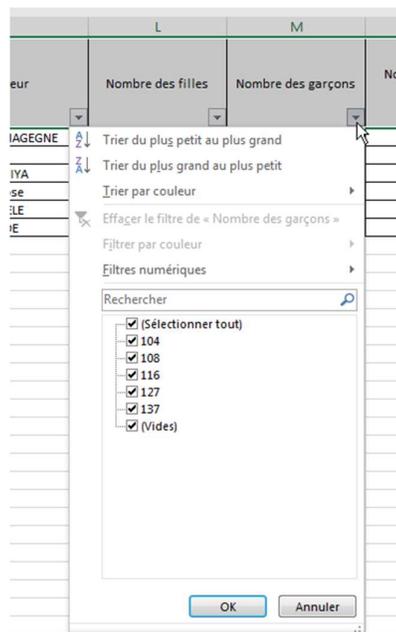
Le nombre de lignes remplies à l'étape 1 vous permet de savoir combien de valeurs il doit y avoir pour chaque variable. Utiliser le Compte indiqué précédemment pour déterminer si des valeurs manquent et saisir ce nombre dans la colonne concernée.

Méthode alternative vérifier les valeurs manquantes

Vous pouvez également utiliser l'outil « Filtre » pour voir s'il manque des valeurs.

Cliquer sur le numéro de ligne de la ligne qui contient les noms de variable (dans ce cas, la ligne 1). Sélectionner le menu Données dans le ruban en haut puis « Filtrer » :

Date de l'enquête	Région	District	Code du district (XXX)	Latitude arrivée (XXX.XXXXX)	Longitude arrivée (XXX.XXXXX)	Latitude départ (XXX.XXXXX)	Longitude départ (XXX.XXXXX)	Nom de l'école	Code de l'école (XXX)	Nom du directeur	Nombre des filles	Nombre des garçons	Nombre total des élèves	Les élèves de votre établissement ont-ils reçu un traitement vermifuge l'année dernière (0 = N, 1 = O)
09/07/2015	AMHARA	BELB WUHA	120	12.18.550	037.45.04	12.118.542	037.45.075	BELB W	101	WASHUN WENDIMAGÉNE	129	104	233	0
09/09/2015	AMHARA	DAWA CHEFE	117	10.46.084	039.51.800	10.46.084	039.05.799	BILLECHA	102	TARIKU	1002	137	239	1
09/09/2015	AMHARA	EMESE	117	10.52.371	039.51.607	10.52.370	039.05.670	SEITE	103	REGASSA WAKIYA	78	127	205	0
09/11/2015	SNMP	GUAFERDA	123	06.83.725	035.29.902	06.83.725	35.29.872	OTEWIA No 2	104	Mengistu Erose	112	112	112	1
09/12/2015	SNMP	GUAFERDA	123	06.86.817	035.34.801	06.87.925	035.25.734	ORENITA	104	AYENEW BERELE	146	108	254	1
09/09/2015	SNMP	MIZAN TOWN	122	06.99.301	035.58.899	06.99.306	035.38.890	MESGANA AKADJANE	106	TATEK ABEDE	99	116	215	1



9. Est-ce ce que l'on attend ?

Vérifier la synthèse de chaque variable et déterminer si elle est raisonnable pour les données recueillies (par ex., le nombre de districts est-il correct, la plage du nombre d'enfants est-elle raisonnable ?). Cette étape vous donnera une idée des points sur lesquels vous concentrer.

10. Maintenant que vous avez synthétisé l'état actuel des données, vous pouvez prendre chaque variable une à une et décider si des modifications doivent y être apportées. Veiller à noter chaque modification faite.

Si vous devez faire de nombreuses modifications, utiliser la feuille « Enquête école » ou « Enquête élève » pour les stocker. Il est recommandé d'inclure la référence de ligne ou l'ID unique de chaque variable modifiée de sorte que l'on puisse facilement retracer vos actions.

Maintenant, à vous : préparer une synthèse des données d'élève brutes

3.2 Préparer une synthèse des données d'élève.**3.3 Enregistrer vos données**

Après avoir examiné vos données et apporté les éventuelles modifications, enregistrer sous « données 3 ».

Enregistrer votre « classeur de nettoyage de données ».

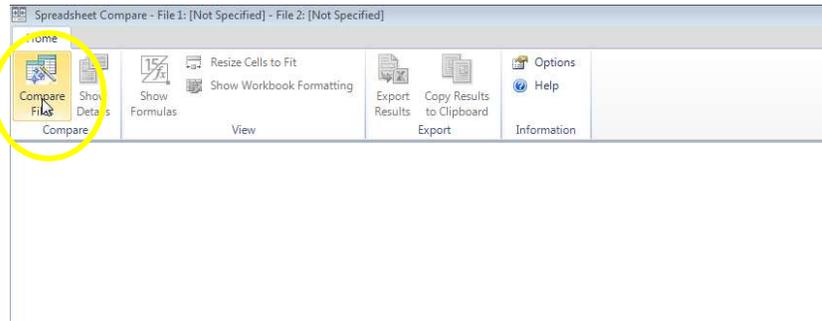
3.4 Comparer des feuilles de calcul

Pour voir facilement les modifications apportées, vous pouvez utiliser le programme « Comparer des feuilles de calcul » de MS Office.

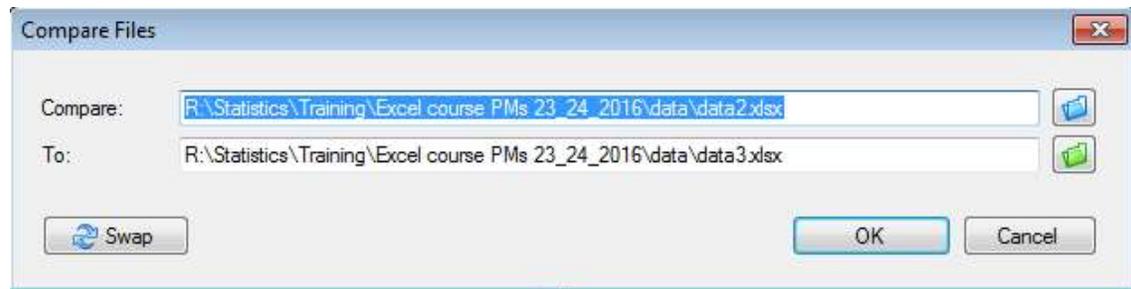


Cliquer sur le bouton Windows (entouré) et taper « spreadsheet compare » dans la case Rechercher.

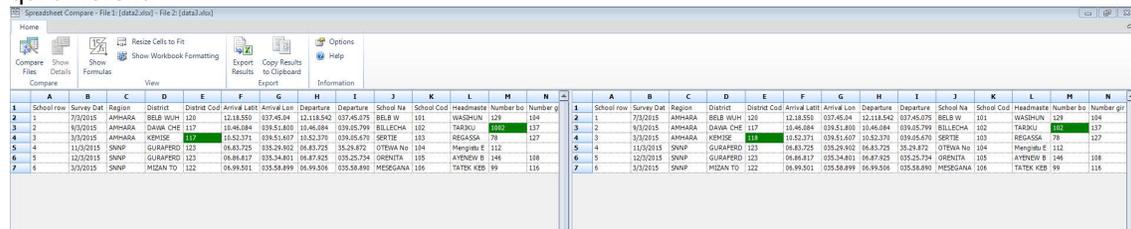
Cliquer sur le bouton « Comparer des fichiers » à gauche dans le ruban menu :



Sélectionner les deux fichiers que vous souhaitez comparer dans la boîte de dialogue et cliquer sur OK.



Spreadsheet Compare affiche les deux feuilles de calcul côte à côte et met en surbrillance les cellules qui diffèrent.



Session 4 : Vérifier si le protocole a été respecté : sexe et âge

Dans le classeur de nettoyage des données, on trouve deux feuilles de calcul qui constituent la base de nos vérifications concernant le Sexe et l'Âge, *Vérification sexe* et *Vérification âge*.

On commence par *Vérification sexe*. Le protocole impose qu'un nombre de garçons et de filles à peu près égal soit testé dans chaque école. Les écarts à ce sujet doivent être stockés dans le classeur.

	A	B	C	D	E	F	G	H
1	Cette feuille sert à vérifier que le protocole concernant les sexes à échantillonner a été respecté dans chaque école							
2	nom de l'ensemble de données							
3								
4								
5		Sexe						
6	École	Nombre de garçons	Nombre de filles	Nombre de sexes manquants	Nombre total d'élèves	% de filles	Le protocole a-t-il été respecté dans cette école ? (O/N)	Remarques
7	101	25	25	0	50	50%	O	
8	102	25	25	0	50	50%	O	
9	103	24	26	0	50	52%	O	
10	104	50	0	0	50	0%	O	École de garçons, 2 élèves enregistrés sous '0' au lieu de '1'

Ci-dessus, on trouve un exemple de tableau de *Vérification sexe* pour une enquête de cartographie. Dans chaque école, 50 élèves ont été testés. La première école était une école mixte dans laquelle 25 élèves de chaque sexe ont été échantillonnés. La deuxième école était une école pour filles dans laquelle 50 filles ont été échantillonnées. Dans les deux cas, le protocole a été respecté (en espérant qu'une école pour garçons viendra équilibrer le bilan).

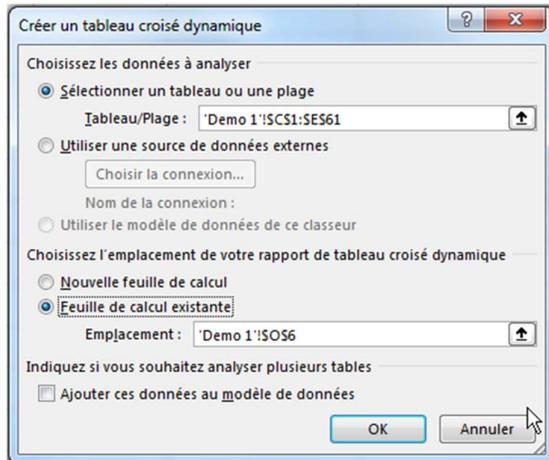
Démonstration Vérification sexe

- Ouvrir le fichier *Session_4.xlsx* et cliquer sur l'onglet de la première feuille *Demo 1*.
- Préparation du tableau : masquer les colonnes superflues, mettre en forme les en-têtes de la première ligne du tableau, insérer une colonne avec les numéros d'observation.
- En guise de petite vérification préliminaire, placer le curseur dans la première ligne de la variable *Sexe* et appuyer sur **Ctrl + ↓** pour vérifier les valeurs manquantes. Ce raccourci placera le curseur à la fin d'un bloc. **[MAJ + Ctrl + ↓** mettra en surbrillance la colonne jusqu'à la fin du bloc]
- À présent, on crée un tableau croisé dynamique qui indique le nombre de garçons et de filles dans chaque école. Mettre

C	D	E
Sexe	Âge	Code école
1	13	1
0	10	1
1		1
0	11	1
1	13	1
0	13	1
1	15	1
0	12	1
1	11	1
0	11	1
1	12	1
0	10	1
1	12	1
0	10	1
1	11	1

en surbrillance les colonnes Sexe, Âge et Code école comme illustré à droite.

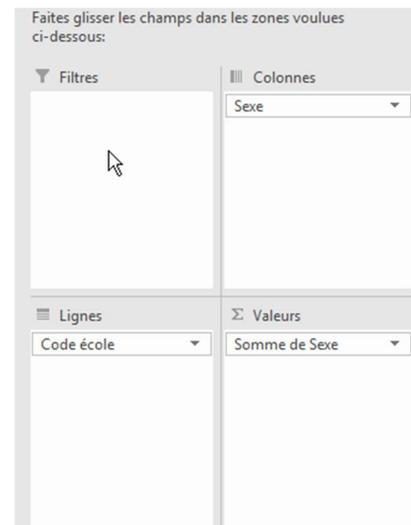
5. Cliquer sur *Tableau croisé dynamique* dans le ruban *Insertion*.
6. La fenêtre suivante apparaît :



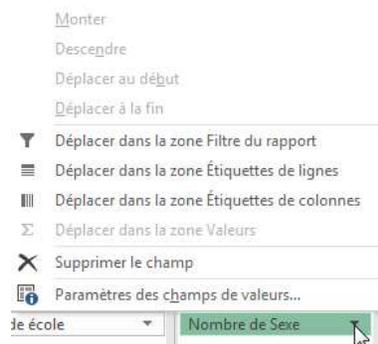
Le champ *Tableau/Plage* : désigne la plage de cellules sélectionnée à l'étape 2.

Dans la moitié inférieure de la fenêtre, un emplacement peut être choisi pour le tableau croisé dynamique. Dans cet exemple, j'ai cliqué sur la cellule O6 de la première feuille *Demo 1*. Cliquer sur OK.

7. La fenêtre de droite apparaît.
Faire glisser les noms de variable dans le tableau des paramètres du Tableau croisé dynamique comme illustré à droite.
Cliquer sur le triangle à côté de *Somme de sexe* et cliquer sur *Paramètres des champs de valeurs...*



La fenêtre suivante apparaît. Dans *Résumer le champ de valeur par*, mettre en surbrillance *Compte* et cliquer sur OK.



Le tableau croisé dynamique doit à présent indiquer le nombre de garçons et de filles de chaque école (1, 2, 3) ainsi que le nombre total d'élèves.

ID école →	Nombre de Sexe		Étiquettes de colonnes		Total général
	Étiquettes de lignes	0	1 (vide)		
1	10	10		20	
2	10	10		20	
3	11	8		19	
Total général	31	28		59	

Copier les trois premières colonnes dans le Classeur de nettoyage des données.

8. Pour déterminer le nombre total d'élèves et le % des élèves filles dans le Classeur de nettoyage des données, utiliser des formules :

B	C	D	E	F
Sexe				
Nombre de garçons	Nombre de filles	Nombre de sexes manquants	Nombre total d'élèves	% de filles
10	10	0	=SOMME(B7:D7)	=C7/E7

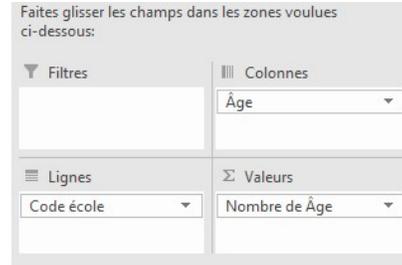
Maintenant, à vous

Ouvrir le fichier *data3.xlsx* de la dernière session et vérifier la variable Sexe. Transférer vos résultats dans le Classeur de nettoyage des données et enregistrer. Enregistrer les données nettoyées sous *data4.xlsx*.

Démonstration Vérification âge

9. Ouvrir le fichier Session_4.xlsx et aller sur la feuille Demo 2.<

10. Mettre en surbrillance les colonnes Âge et Code école et créer un tableau croisé dynamique avec les paramètres indiqués à droite.



Vous devez obtenir le tableau ci-dessous

Count of Age	Column Labels	0	7	8	9	10	11	12	13	14	15	99	12345	(blank)	Grand Total
1	Row Labels					3	5	4	4		2	1			19
2			1		2	5	3	2	2	2	3				20
3		1		1		4	4	1	1	3	3		1		19
Grand Total		1	1	1	2	12	12	7	7	5	8	1	1		58

Nombre de Âge	Étiquettes de colonnes	0	7	8	9	10	11	12	13	14	15	99	(vide)	Total général
Étiquettes de lignes														
1						3	5	4	4		2	1		19
2							1	2	5	3	2	2	3	20
3							1	1	4	4	1	2	3	19
Total général						1	1	1	2	12	12	7	8	58

11. L’observation contenant l’âge 12345 semble être une donnée factice. Vérifier les autres valeurs de cette observation et déterminer si vous souhaitez retirer la totalité de cette observation.

Les âges 0 et 99 pourraient être factices également ou une erreur de frappe. Vérifier les autres valeurs de cette observation et déterminer si vous souhaitez supprimer les valeurs âge et les marquer comme valeurs manquantes ou si vous souhaitez supprimer la totalité de l’observation.

12. Copier les parties respectives du tableau croisé dynamique dans le Classeur de nettoyage des données. Si nécessaire, ajuster la plage d’âge dans le classeur par rapport à la plage d’âge des données.

- Nombre total d’élèves = somme(< la totalité de la plage d’âge échantillonnée incluant les valeurs manquantes >)
- Nombre d’élèves dans le protocole = somme(< la plage d’âge du protocole >)
- % d’élèves dans le protocole = <nombre d’élèves dans le protocole > / <nombre total d’élèves >

Maintenant, à vous

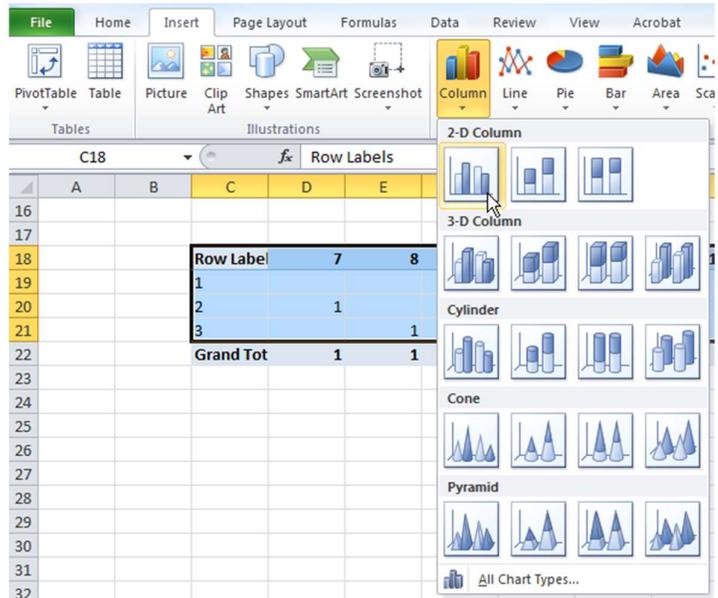
Poursuivre votre travail sur le fichier *data4.xlsx* en vérifiant la variable Âge. Transférer vos résultats dans le Classeur de nettoyage des données. À la fin de votre travail, enregistrer vos données sous 'données 4' et enregistrer le classeur de nettoyage des données.

Exercice facultatif

Pour vérifier les âges, on peut également créer une visualisation de la répartition des valeurs d'âge, par ex., un histogramme. Pour cela, vous avez besoin d'un tableau croisé dynamique qui incluent les âges comme noms de ligne

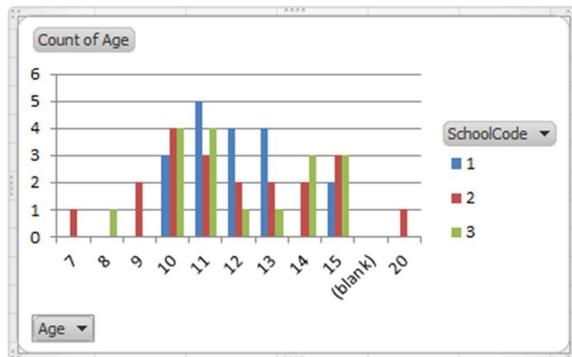
- Créer un tableau croisé dynamique similaire à celui-ci.

Count of Age				
Column Labels	1	2	3	Grand Total
Row Labels				
7		1		1
8			1	1
9		2		2
10	3	4	4	11
11	5	3	4	12
12	4	2	1	7
13	4	2	1	7
14		2	3	5
15		2	3	5
(blank)				
20		1		1
Grand Total	18	20	17	55



- Mettre en surbrillance la totalité du tableau croisé dynamique. Activer le ruban Insertion et cliquer sur Colonne. Sélectionner le premier *Histogramme 2-D* :

Vous devez obtenir un histogramme similaire à l'illustration de droite. Cet histogramme indique le nombre d'élèves pour chaque âge. Pour chaque école, une couleur différente est utilisée. Cliquer sur le triangle dans le bouton Code école à droite et sélectionner une école. Faire de même avec le bouton Âge en bas à gauche.



Session 5 Créer des variables parasitologiques

5.1 Introduction

Maintenant que nous en savons davantage sur les données et savons repérer les valeurs manquantes, nous pouvons commencer à créer des variables parasitologiques. Ces variables dépendent de l'espèce étudiée et du test utilisé :

Pour *S. mansoni* et les helminthes transmis par le sol (STH) évalués à l'aide de Kato-Katz :

1. Créer un décompte des œufs en œufs par gramme
2. Créer des variables d'intensité moyenne de l'infection
3. Créer des variables binaires (0/1) pour indiquer un résultat positif ou négatif
4. Créer des variables binaires pour indiquer si une personne présentait ou non une infection massive

Pour *S. haematobium* :

1. Créer un décompte d'œufs en œufs par 10 ml d'urine
2. Créer une variable binaire pour indiquer un résultat positif ou négatif
3. Créer une variable binaire pour indiquer si une personne présentait ou non une infection massive

Pour les autres variables parasitologiques (CCA, analyse d'urine par bandelettes, hématurie visible) :

1. Créer des variables binaires pour indiquer un résultat positif ou négatif

5.1.1 Classeur de fonctions

Les fonctions sont un concept important d'Excel, et mieux vaut les découvrir à partir d'un exemple.

1. Ouvrir le classeur nommé 'Session 5 – Classeur de fonctions'.
2. Résoudre les problèmes de la feuille 'à compléter' à l'aide des fonctions indiquées.
3. Si vous avez besoin d'aide, les réponses se trouvent dans la feuille 'Réponses'.

5.2 Créer des variables parasitologiques pour *S. mansoni* et les STH évalués par Kato-Katz

5.2.1 Créer un décompte des œufs en œufs par gramme

S. mansoni et les STH sont souvent évalués à l'aide de Kato-Katz où un ou plusieurs petits échantillons de selles sont examinés sous microscope pour détecter des œufs. Chaque échantillon de selles pèse environ $1/24^{\text{ème}}$ de gramme et on multiplie en général le nombre d'œufs trouvés par 24 pour exprimer le résultat en œufs par gramme (opg). Pour cela :

1. Ouvrir 'données 4' dans le dossier '2 Nettoyage des données'.
2. Faire défiler jusqu'à la première colonne vide à droite – colonne U
3. Masquer les lignes O à T qui ne contiennent pas des données Kato-Katz. Conserver *S. mansoni*, *Ascaris*, *Ankylostomiase* et *Trichuris*.
4. Créer une nouvelle colonne appelée 'opg S. mans 1'. Vérifier que vos résultats ressemblent à ceci :

V
opg S.mans 1

5. On ne peut pas se contenter de multiplier le nombre d'œufs trouvés par 24 car Excel indiquera que les valeurs manquantes multipliées par 24 donnent zéro (faire un test sur un ensemble de données vide). Il nous faut utiliser une fonction Si. Dans la cellule U2, saisir la formule '=SI(ESTVIDE(G2),"", G2 * 24)'

V	W	X
opg S. mans 1		
=SI(ESTVIDE(I2), "", I2*24)		

6. Remplir les informations pour l'ensemble des cellules de la colonne à l'aide de la poignée de recopie.
7. Utiliser Ctrl + v pour aller en bas et vérifier que la cellule finale est correcte
8. Trouver une valeur manquante et vérifier que le résultat est correct.
9. Ajouter les noms de variable pour les autres variables évaluées par Kato-Katz. Procéder dans le même ordre que les colonnes W à AC ci-dessous

V	W	X	Y	Z	AA	AB	AC
opg S. mans 1	opg Ascaris 1	opg Ankylo 1	opg Trichuris 1	opg S. mans 2	opg Ascaris 2	opg Ankylo 2	opg Trichuris 2

10. Dans la cellule V2, utiliser Ctrl + shift + flèche bas pour mettre en surbrillance les cellules du bas. Vérifier que vous êtes bien à la ligne inférieure (298). Utiliser la poignée de recopie pour faire glisser les formules sur 5 colonnes. Faire défiler vers le haut pour vérifier que vos données sont conformes à vos attentes. Faire un double clic dans quelques cellules pour vérifier que tout est ok.

V	W	X	Y	Z	AA	AB	AC
opg S. mans 1	opg Ascaris 1	opg Ankylo 1	opg Trichuris 1	opg S. mans 2	opg Ascaris 2	opg Ankylo 2	opg Trichuris 2
0	0		0	24	0		0
0	0		0	0	0		0
312	0		0	192	0		0
48	0		0	0	0		0
216	0		0	120	0		0

5.2.2 Créer des variables d'intensité moyenne de l'infection

Les décomptes d'œufs en œufs par gramme sont souvent traduits ensuite en intensité moyenne d'infection afin d'avoir un nombre qui représente l'intensité de l'infection. Nous utiliserons la moyenne arithmétique. Pour cela :

1. Créer une nouvelle variable dans la colonne AC appelée 'opg S. mans'.
2. On ne peut pas se contenter de calculer la moyenne des opg avec la moyenne des deux opg car Excel pensera que la moyenne de deux valeurs manquantes est une erreur alors qu'elle est juste manquante. Cependant, Excel prendra la moyenne d'une valeur manquante et

d'une valeur non manquante comme égale à la valeur non manquante. Le tableau ci-dessous indique comment Excel gère les valeurs manquantes pour calculer des moyennes.

Var 1	Var 2	Moyenne(Var1, Var 2)
6	4	5
6		6
	4	4
		#DIV/0!

3. Pour contourner ce problème, on doit utiliser une fonction SI indiquant
 - a. Si les deux valeurs manquent, afficher manquante
 - b. Sinon, afficher la moyenne des deux valeurs
4. Dans la cellule AC2, saisir la formule '=SI(ET(V2="","",Z2=""), "", MOYENNE(V2,Z2))' et appuyer sur Entrée

V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
opg S. mans 1	opg Ascaris 1	opg Ankylo 1	opg Trichuris 1	opg S. mans 2	opg Ascaris 2	opg Ankylo 2	opg Trichuris 2	S. mans opg moyenne			
0	0		0	24	0		0	=SI(ET(V2="","",Z2=""), "", MOYENNE(V2,Z2))			
0	0		0	0	0		0	SI(test_logique, [valeur_si_vrai], [valeur_si_faux])			

5. Copier la formule jusqu'en bas de la page et vérifier que les cellules sont correctement remplies.
6. Trouver quelques élèves dont les données *S. mansoni* manquent et vérifier que la formule fait ce qu'il faut.
7. Nommer les colonnes AE à AG pour les autres STH dans l'ordre qui convient puis utiliser la poignée de recopie pour compléter les colonnes. Vérifier qu'Excel fait ce qu'il faut. Créer des variables binaires pour indiquer un résultat positif ou négatif

V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
opg S. mans 1	opg Ascaris 1	opg Ankylo 1	opg Trichuris 1	opg S. mans 2	opg Ascaris 2	opg Ankylo 2	opg Trichuris 2	S. mans opg moyenne	Ascaris opg moyenne	Ankylo opg moyenne	Trichuris opg moyenne
0	0		0	24	0		0	12	0		0
0	0		0	0	0		0	0	0		0
312	0		0	192	0		0	252	0		0

Une fois l'intensité moyenne d'infection mesurée, on peut créer une variable indiquant le résultat positif ou négatif de chaque personne. On utilise 1 pour indiquer si le résultat était positif et 0 s'il était négatif. Ainsi, nous pourrions simplement prendre la moyenne des 1 et des 0 pour obtenir la prévalence. Nous verrons cela en détail ultérieurement. Pour créer les indicateurs de prévalence :

1. Créer une nouvelle variable dans la colonne AG2 appelée 'S. mans inf'.
2. Nous allons utiliser la fonction Excel 'SIGNE'. SIGNE affiche 1 lorsqu'un résultat est positif, 0 lorsqu'un résultat est 0 et -1 lorsqu'un résultat est négatif. Comme précédemment, il va nous falloir gérer les valeurs manquantes. Saisir la formule '=SI(AD2="","",SIGNE(AD2))'

AH	AI	AJ	AK
S. mans inf			
=SI(AD2="","",SIGNE(AD2))			
SI(test_logique, [valeur_si_vrai], [valeur_si_faux])			

3. Utiliser la poignée de recopie pour compléter toutes les lignes
4. Vérifier que les données sont telles que prévues. Vérifier en particulier les valeurs manquantes
5. Procéder de la même manière pour les STH dans les colonnes suivantes

5.2.3 Créer des variables binaires pour indiquer un résultat d'infection massive

Nous utiliserons 1 pour indiquer qu'une personne présente une infection massive et 0 pour indiquer qu'il n'y a pas d'infection massive (pas d'infection ou infection légère ou modérée). Les infections massives sont déterminées selon les directives de l'OMS :

Tableau 2.3 *Seuils pour le classement des infections par Ascaris lumbricoides, Trichuris trichiura, Ancylostomes et Schistosoma : infections légères, modérées ou d'intensité élevée*

Helminthe	Densité de l'infection		
	Légère	Modérée	Massive
<i>A. lumbricoides</i>	1-4999 oeufs/g	5000-49999 oeufs/g	≥50000 oeufs/g
<i>T. trichiura</i>	1-999 oeufs/g	1000-9999 oeufs/g	≥10000 oeufs/g
Ancylostomes	1-1999 oeufs/g	2000-3999 oeufs/g	≥4000 oeufs/g
<i>S. mansoni</i>	1-99 oeufs/g	100-399 oeufs/g	≥400 oeufs/g
<i>S. haematobium</i>	1-50 oeufs/10ml d'urine		≥50 oeufs/10ml d'urine ou hématurie visible
<i>S. japonicum</i>	Toute infection est considérée comme massive		

Pour créer une variable d'intensité massive pour *S. mansoni* :

1. Créer une nouvelle variable dans la colonne AK appelée 'S. mans massive'.
2. Saisir la formule '=SI(AC2="","",SI(AC2>400,1,0))'

AD	AE	AF	AG	AH	AI	AJ	AK	AL	AM	AN
S. mans opg moyenne	Ascaris opg moyenne	Ankylo opg moyenne	Trichuris opg moyenne	S. mans inf	Ascaris inf	Anylo inf	Trichuris inf	S. mans massive		
12	0		0	1	0		0=SI(AD2="","",SI(AD2>=400,1,0))			

3. Utiliser la poignée de recopie pour compléter toutes les lignes
4. Vérifier que les données sont telles que prévues. Vérifier en particulier les valeurs manquantes
5. Procéder de la même manière pour les STH dans les colonnes suivantes en modifiant la définition d'infection massive pour :
 - a. Ascaris 50 000
 - b. Ankylostomiase 4 000
 - c. Trichuris 10 000
6. Vérifier que les données sont telles que prévues, en particulier les valeurs manquantes

5.3 Créer des variables parasitologiques pour *S. haematobium* par filtration de l'urine

5.3.1 Créer des variables d'intensité moyenne de l'infection par *S. haematobium*

Lorsque l'on teste *S. haematobium* par filtration de l'urine, on filtre en général 10 ml d'urine. Toutefois, si moins de 10 ml d'urine est disponible, on estime le nombre d'œufs présents dans 10 ml avec le calcul : $\text{oeufs par 10 ml d'urine} = \frac{\text{nombre d'oeufs détectés}}{\text{volume d'urine filtré en ml}} * 10$

Pour estimer le nombre d'œufs par 10 ml d'urine :

1. Faire défiler jusqu'à la prochaine colonne disponible (AP). Masquer toutes les variables entre AP et les informations sur la filtration d'urine (entre T et AN). Vous aurez peut-être à afficher

les données masquées dans la section précédente. Intituler la colonne AP 'S. haem op10ml' et vérifier que vos données ressemblent à ceci.

S	T	U	AP
maturie visible	Volume d'urine	S. haem lame	S. haem op10ml
0	10	0	

2. Nous allons ajouter une formule qui dit :
 - a. Si le nombre d'œufs manque, afficher manquant
 - b. Si le nombre d'œufs manque ou est 10, afficher le nombre d'œufs (supposer que le volume d'urine recueilli était de 10 ml)
 - c. Si moins de 10 ml d'urine a été recueilli, utiliser la formule ci-dessus
 - d. Noter que cette formule n'inclut pas les fois où le volume d'urine a été stocké mais où le nombre d'œufs manquait, il faudrait donc étudier ça plus en détail
3. Essayer de trouver par vous-même quelle formule utiliser
4. On peut utiliser la formule suivante
 '=SI(ET(ESTVIDE(R2),ESTVIDE(S2)),"",SI(OU(R2=10,ESTVIDE(R2)),S2,10*S2/R2))'

	T	U	AP	AQ	AR	AS	AT	AU	AV	A
rie	Volume d'urine	S. haem lame	S. haem op10ml							
	10	0	=SI(ET(ESTVIDE(T2),ESTVIDE(U2)),"",SI(OU(T2=10,ESTVIDE(T2)),U2,10*U2/T2))							

5. Utiliser la poignée de recopie vers le bas et vérifier que les données sont telles que prévues.

5.3.2 Créer des variables binaires pour indiquer un résultat positif ou négatif pour S. haematobium

Créer une variable 'S. haem inf' dans la colonne AP. Utiliser la fonction SIGNE de la section 2.3.3 pour créer une variable binaire pour S. haematobium et vérifier qu'elle ressemble à ceci

AP	AQ
S. haem op10ml	S. haem inf
0	0
14	1
861	1

5.3.3 Créer un indicateur d'infection massive pour S. haematobium

Suivre les étapes de la section 2.2.4 pour créer un indicateur d'infection massive pour S. haematobium. Les décomptes d'œufs de 50 et plus par 10 ml d'urine sont considérés comme des infections massives. Vérifier que vos résultats ressemblent à ceci :

U	AP	AQ	AR
S. haem Lame <input type="text"/>	S.haem op10ml	S.haem inf	S. haem massive
0	0	0	0
14	14	1	0
861	861	1	1
0	0	0	0

5.4 Créer des indicateurs d'infection pour CCA, analyse d'urine par bandelettes et hématurie visible

Utiliser la fonction SIGNE pour créer des indicateurs pour CCA, analyse d'urine par bandelettes et hématurie visible. Vérifier que vos résultats ressemblent à ceci :

O	P	Q	AR	AS	AT
CCA <input type="text"/>	Dipstick <input type="text"/>	Visible Haematuri a <input type="text"/>	CCA pos <input type="text"/>	Dipstick pos <input type="text"/>	Visible haematuri a pos <input type="text"/>
1	0	0	1	0	0
2	3	0	1	1	0
2	5	0	1	1	0
	0	0		0	0
3	2	0	1	1	0
0	4	0	0	1	0

5.4.1 Enregistrer votre travail.

Enregistrer l'ensemble de données actuel sous le nom 'données 5' dans '2 nettoyage des données'. Ainsi, vous pourrez voir ce que vous avez modifié lors de ces étapes.

Session 6 Synthèses de données nettoyées

Démarrer cette session avec l'ensemble de données « données 5 ». Vous utiliserez également le « classeur de nettoyage des données ».

6.1 Compléter la synthèse de données nettoyées

Suivre les étapes décrites dans la Session 3 – Créer des synthèses de données brutes pour compléter l'onglet « synthèse de données nettoyées » dans le « classeur de nettoyage des données ». Normalement vous n'avez aucune modification à apporter aux données cette fois, mais c'est un bon moyen de synthétiser le contenu des données et de vérifier une seconde fois que tout est en ordre.

Vous pouvez également consulter « l'antisèche sur la synthèse de données ».

6.2 Enregistrer les données

Enregistrer les données sous « données 6 » et enregistrer les mises à jour dans le « classeur de nettoyage des données ».

Lorsque les données nettoyées vous conviennent, les enregistrer dans le dossier « 3 Données nettoyées » et leur attribuer un nom pertinent. Chez SCI nous utilisons le format suivant : *Code pays_année_type d'enquête_"données nettoyées"_date de nettoyage*

Par exemple : "ETH_2015_impact_données_nettoyées_25-10-2016

Session 7 Créer des tableaux de résultats

7.1 Synthèse

Maintenant que les données sont nettoyées, il est temps de créer un tableau de résultats.

Ce sont les informations que la plupart des personnes verront en regardant les données.

7.2 Préparer les données

1. Ouvrir les données nettoyées dans le dossier '3 Données nettoyées'. Si vous n'êtes pas sûr des données que vous avez créées, vous trouverez les données nettoyées dans le dossier '2015_Ethiopia_impact-completed'.
2. Enregistrer les données nettoyées dans '4 Analyses des données' sous 'Création du tableau de résultats'. Ainsi, vous ne pourrez faire aucune modification accidentelle des données nettoyées.

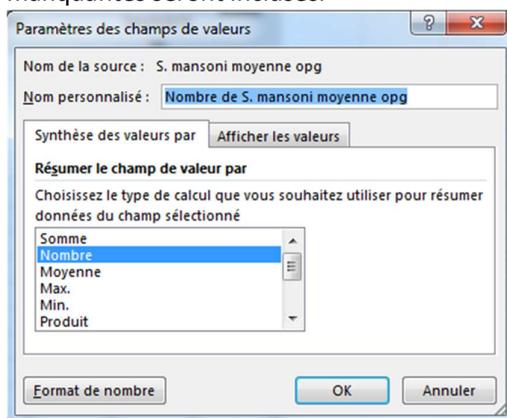
7.3 Préparer le tableau des résultats

1. Ouvrir le fichier appelé 'Session 7 – Modèle de tableau de résultats'
2. L'enregistrer dans '5 Résultats' sous un nom pertinent. Inclure la date du jour.

7.4 Compléter 'SCH global'

1. Ouvrir le fichier des résultats et cliquer sur la feuille 'SCH global'
2. Compléter l'ensemble des informations de cette feuille. Pour cela :
3. Mettre en surbrillance toutes les données dans élèves et créer un tableau croisé dynamique avec Insertion -> Tableau croisé dynamique
4. Trouver les âges minimum et maximum en plaçant âge deux fois dans 'Valeurs', un minimum et un maximum
5. Trouver le nombre d'enfants avec des données sur *S. mansoni* en plaçant 'Smans moyenne opg' dans valeurs et en sélectionnant 'Compte nombres'. Si vous sélectionnez 'Compte', les données manquantes seront incluses.

Σ Valeurs	
Min. de Âge (années)	▼
Max. de Âge (années)2	▼



6. Trouver la prévalence de *S. mansoni* en trouvant la moyenne de 'S mansinf'. Pour calculer le pourcentage positif d'une variable 0/1, il suffit de prendre la moyenne :
 - Par ex., si 100 personnes ont été testées, et que 40 étaient positives, la prévalence est de $40/100 = 0,4 = 40\%$
 - C'est pareil que $(40 * 1 + 60 * 0) / 100$

- Ce qui revient à faire la moyenne de la colonne des 0/1

Σ Valeurs	
Min. de Âge (années)	▼
Max. de Âge (années)2	▼
Nombre de S. manso...	▼
Moyenne de S. mans...	▼

7. Trouver la prévalence de l'infection 'S. mansoni massive' en calculant la moyenne de 'S mans massive'

Σ Valeurs	
Min. de Âge (années)	▼
Max. de Âge (années)2	▼
Nombre de S. mansoni moyenne opg	▼
Moyenne de S. mansoni inf	▼
Moyenne de S. mansoni infection massive	▼

8. Trouver la moyenne des opg

Σ Valeurs	
Min. de Âge (années)	▼
Max. de Âge (années)2	▼
Nombre de S. mansoni moyenne opg	▼
Moyenne de S. mansoni inf	▼
Moyenne de S. mansoni infection massive	▼
Moyenne de S. mansoni moyenne opg	▼

9. Faire de même pour les variables *S. haematobium*

Σ Valeurs	
Min. de Âge (années)	▼
Max. de Âge (années)2	▼
Nombre de S. mansoni moyenne opg	▼
Moyenne de S. mansoni inf	▼
Moyenne de S. mansoni infection massive	▼
Moyenne de S. mansoni moyenne opg	▼
Nombre de S. haem. op10ml	▼
Moyenne de S. haem inf	▼
Moyenne de S. haem infection majeure	▼
Moyenne de S. haem. op10ml	▼

10. Copier et coller le tableau de résultats dans le modèle de résultats, en utilisant 'coller valeurs'.
11. Vérifier que tous les noms de colonne sont corrects avant de supprimer les données non nécessaires
12. Mettre en forme les cellules pour faciliter leur lecture. Vérifier que vos données ressemblent à ceci :

	A	B	C	D	E	F	G	H	I
1	Plage d'âges des élèves échantillonnés	S. mans par KK Nb élèves	S. mans par KK % positifs	S. mans par KK % infection massive	S. mans par KK moyenne opg	S. haem par FU Nb élèves	S. haem par FU % positifs	S. haem par FU % infection majeure	S. haem par FU moyenne opg
2		297	40.7%	4.4%	73.58	297	39.7%	10.8%	35.50

13. L'étape suivante consiste à créer la plage d'âges des élèves échantillonnés. Pour l'instant, nous avons le minimum et le maximum dans deux colonnes, mais nous souhaitons les regrouper dans une seule. Pour les regrouper, on peut utiliser la fonction 'concaténer'.
14. Dans une cellule sous les âges minimum et maximum, taper '= concatener'. Puis cliquer sur l'âge minimum, ajouter « - » avec les guillemets et cliquer sur l'âge maximum. Vérifier que votre formule ressemble à ceci :

Min. de Âge (années)	Max. de Âge (années)2	Nombre de S. mansoni moyenne opg	Moyenne de S. mansoni inf	Moyenne de S. mansoni infection mass
8	12	297	0.407407407	0.0437
=CONCATENER(LIREDONNEESTABCROISDYNAMIQUE("Min. de Âge (années)",\$BC\$4), "- ", LIREDONNEESTABCROISDYNAMIQUE("Max. de Âge (années)2", \$BC\$4))				

15. Appuyer sur Entrée et vérifier que la cellule ressemble à ceci :

Min. de Âge (années)	Max. de Âge (années)2
8	12
8 - 12	

16. Copier et coller dans le tableau de résultats avec 'coller valeurs' et vérifier que votre tableau de résultats ressemble à ceci :

Plage d'âges des élèves échantillonnés	S. mans par KK Nb élèves	S. mans par KK % positifs	S. mans par KK % infection massive	S. mans par KK moyenne opg	S. haem par FU Nb élèves	S. haem par FU % positifs	S. haem par FU % infection majeure	S. haem par FU moyenne opg
8 - 12	297	40.7%	4.4%	73.58	297	39.7%	10.8%	35.50

17. Enregistrer le tableau de résultats

7.5 Compléter 'SCH par école

1. Pour compléter SCH par école, il vous suffit de placer le Code école dans les lignes du tableau croisé dynamique en affichant les données globales

Lignes	Σ Valeurs
Code école	Min. de Âge (années)
	Max. de Âge (années)2
	Nombre de S. mansoni moyenne opg
	Moyenne de S. mansoni inf
	Moyenne de S. mansoni infection massive
	Moyenne de S. mansoni moyenne opg
	Nombre de S. haem. op10ml
	Moyenne de S. haem inf
	Moyenne de S. haem infection majeure
	Moyenne de S. haem. op10ml

2. Copier les données dans le tableau de résultats, mettre en forme les cellules et utiliser Concaténer pour rassembler les âges. Vérifier que vos données ressemblent à ceci :

Region	District	School Name	School Code	Age range of pupil sampled	S. mans by KK N pupils	S. mans by KK % positive	S. mans by KK % heavy	S. mans by KK mean epg	S. haem by UF N pupils	S. haem by UF % positive	S. haem by UF % heavy	S. haem by UF mean ep10ml
			101	8-12	50	12.0%	2.0%	44.40	50	16.0%	4.0%	6.34
			102	8-12	50	50.0%	2.0%	63.84	50	70.0%	22.0%	94.04
			103	8-12	50	84.0%	8.0%	112.80	50	46.0%	2.0%	8.42
			104	8-12	50	24.0%	2.0%	50.16	50	2.0%	2.0%	1.78
			105	8-12	50	62.0%	10.0%	146.16	50	80.0%	34.0%	98.22
			106	8-12	47	10.6%	2.1%	20.94	47	23.4%	0.0%	2.21

3. La prochaine étape consiste à importer la région, le district et le nom d'école de chaque code d'école. Pour cela, nous allons utiliser 'index' et 'equiv'. Index et equiv sont des fonctions traitées dans le classeur Excel 'Session 7 – classeur index et equiv'. Faire les exercices de la feuille 'à compléter' du classeur. Si vous avez besoin d'aide, les réponses se trouvent dans la feuille 'Réponses'.
4. Pour importer les informations des écoles, tout d'abord copier et coller les informations de l'ensemble de données sous le tableau de résultats. Vérifier que vos données ressemblent à ceci :

Région	District	Nom école	Code école	Plage d'âges des élèves échantillonnés	S. mans par KK Nb élèves	S. mans par KK % positifs	S. mans par KK % infection majeure	S. mans par KK moyenne opg	S. haem par N élè
			101	8-12	50	12.0%	2.0%	44.40	5
			102	8-12	50	50.0%	2.0%	63.84	5
			103	8-12	50	84.0%	8.0%	112.80	5
			104	8-12	50	24.0%	2.0%	50.16	5
			105	8-12	50	62.0%	10.0%	146.16	5
			106	8-12	47	10.6%	2.1%	20.94	4
Région	District	Code district	Latitude arrivée	Longitude arrivée	Latitude départ	Longitude départ	Nom école	Code école	
AMHARA	BELB WUHA	120	12.18.550	037.45.04	12.118.542	037.45.075	BELB W	101	
AMHARA	DAWA CHEFE	117	10.46.084	039.51.800	10.46.084	039.05.799	BILLECHA	102	
AMHARA	KEMISE	118	10.52.371	039.51.607	10.52.370	039.05.670	SERTIE	103	
SNNP	GURAFERDA	123	06.83.725	035.29.902	06.83.725	35.29.872	OTewa No 2	104	
SNNP	GURAFERDA	123	06.86.817	035.34.801	06.87.925	035.25.734	ORENITA	105	
SNNP	MIZAN TOWN	122	06.99.501	035.58.899	06.99.506	035.58.890	MESEGANA AKADAME	106	

- Dans la cellule C2 pour Nom d'école, taper '=INDEX(\$H\$10:\$H\$15, EQUIV(D2, \$I\$10:\$I\$15, 0))'. Veiller à ce que EQUIV finisse par zéro et que le symbole dollar entoure la colonne Nom d'école et Code d'école. Les symboles dollar permettent qu'il soit toujours fait référence à cette donnée lorsque vous vous déplacez dans une autre cellule. Vous pouvez utiliser la touche F4 comme raccourci.

District	Nom école	Code école	des échant
=INDEX(\$H\$10:\$H\$15,EQUIV(D2,\$I\$10:\$I\$15, 0))			8-

- Appuyer sur Entrée et vérifier que le nom d'école est 'BELB W'. Noter que les données n'ont pas besoin d'être dans un ordre particulier pour que ça fonctionne.
- Déplacer la poignée de recopie vers le bas et faire un double clic sur la cellule du bas pour voir si tout fonctionne correctement. Vérifier que vos données ressemblent à ceci :

Région	District	Nom école	Code école	Plage d'âges des élèves échantillonnés	S. mans par KK Nb élèves	S. mans par KK % positifs	S. mans par KK % infection majeure	S. mans par KK moyenne opg	S. haem par FU Nb élèves	S. haem par FU % positifs	S. haem par FU % infection majeure	S. haem par FU moyenne op10ml
		BELB W	101	8-12	50	12.0%	2.0%	44.40	50	16.0%	4.0%	6.34
		BILLECHA	102	8-12	50	50.0%	2.0%	63.84	50	70.0%	22.0%	94.04
		SERTIE	103	8-12	50	84.0%	8.0%	112.80	50	46.0%	2.0%	8.42
		OTEWA No 2	104	8-12	50	24.0%	2.0%	50.16	50	2.0%	2.0%	1.78
		ORENITA	105	8-12	50	62.0%	10.0%	146.16	50	80.0%	34.0%	98.22
		=INDEX(\$H\$10:\$H\$15,EQUIV(D7,\$I\$10:\$I\$15, 0))		8-12	47	10.6%	2.1%	20.94	47	23.4%	0.0%	2.21

Région	District	Code district	Latitude arrivée	Longitude arrivée	Latitude départ	Longitude départ	Nom école	Code école
AMHARA	BELB WUHA	120	12.18.550	037.45.04	12.118.542	037.45.075	BELB W	101
AMHARA	DAWA CHEFE	117	10.46.084	039.51.800	10.46.084	039.05.799	BILLECHA	102
AMHARA	KEMISE	118	10.52.371	039.51.607	10.52.370	039.05.670	SERTIE	103
SNNP	GURAFERDA	123	06.83.725	035.29.902	06.83.725	35.29.872	OTEWA No 2	104
SNNP	GURAFERDA	123	06.86.817	035.34.801	06.87.925	035.25.734	ORENITA	105
SNNP	MIZAN TOWN	122	06.99.501	035.58.899	06.99.506	035.58.890	MESEGANA AKADAME	106

- Procéder de même pour le District et la Région. Vous pouvez utiliser un seul dollar sur le code école dans la fonction EQUIV. Par ex., EQUIV(\$D2...). Ainsi vous verrouillez la colonne sur D mais vous pouvez utiliser la poignée de recopie pour les lignes. Vérifier que vos données ressemblent à ceci :

Region	District	School Name	School Code	Age range of pupil sampled	S. mans by KK N pupils	S. mans by KK % positive	S. mans by KK % heavy	S. mans by KK mean epg	S. haem by UF N pupils	S. haem by UF % positive	S. haem by UF % heavy	S. haem by UF mean ep10ml
AMHARA	BELB WUHA	BELB W	101	8-12	50	12.0%	2.0%	44.40	50	16.0%	4.0%	6.34
AMHARA	DAWA CHEFE	BILLECHA	102	8-12	50	50.0%	2.0%	63.84	50	70.0%	22.0%	94.04
AMHARA	KEMISE	SERTIE	103	8-12	50	84.0%	8.0%	112.80	50	46.0%	2.0%	8.42
SNNP	GURAFERDA	OTEWA No 2	104	8-12	50	24.0%	2.0%	50.16	50	2.0%	2.0%	1.78
SNNP	GURAFERDA	ORENITA	105	8-12	50	62.0%	10.0%	146.16	50	80.0%	34.0%	98.22
SNNP	MIZAN TOWN	MESEGANA AKADAME	106	8-12	47	10.6%	2.1%	20.94	47	23.4%	0.0%	2.21

Region	District	District Code	Arrival Latitude	Arrival Longitude	Departure Latitude	Departure Longitude	School Name	School Code
AMHARA	BELB WUHA	120	12.18.550	037.45.04	12.118.542	037.45.075	BELB W	101
AMHARA	DAWA CHEFE	117	10.46.084	039.51.800	10.46.084	039.05.799	BILLECHA	102
AMHARA	KEMISE	118	10.52.371	039.51.607	10.52.370	039.05.670	SERTIE	103
SNNP	GURAFERDA	123	06.83.725	035.29.902	06.83.725	35.29.872	OTEWA No 2	104
SNNP	GURAFERDA	123	06.86.817	035.34.801	06.87.925	035.25.734	ORENITA	105
SNNP	MIZAN TOWN	122	06.99.501	035.58.899	06.99.506	035.58.890	MESEGANA AKADAME	106

9. Copier et coller Région, District et Nom d'école en tant que valeurs et supprimer le tableau des écoles que vous avez utilisé pour importer les données. Vérifier que votre tableau de résultats ressemble à ceci :

	A	B	C	D	E	F	G	H	I	J	K	L	M
	Région	District	Nom école	Code école	Plage d'âges des élèves échantillonnés	S. mans par KK Nb élèves	S. mans par KK % positifs	S. mans par KK % infection majeure	S. mans par KK moyenne opg	S. haem par FU Nb élèves	S. haem par FU % positifs	S. haem par FU % infection majeure	S. haem par FU moyenne op10ml
1													
2	AMHARA	BELB WUHA	BELB W	101	8-12	50	12.0%	2.0%	44.40	50	16.0%	4.0%	6.34
3	AMHARA	DAWA CHEFE	BILLECHA	102	8-12	50	50.0%	2.0%	63.84	50	70.0%	22.0%	94.04
4	AMHARA	KEMISE	SERTIE	103	8-12	50	84.0%	8.0%	112.80	50	46.0%	2.0%	8.42
5	SNNP	GURAFERDA	OTEWA No 2	104	8-12	50	24.0%	2.0%	50.16	50	2.0%	2.0%	1.78
6	SNNP	GURAFERDA	ORENITA	105	8-12	50	62.0%	10.0%	146.16	50	80.0%	34.0%	98.22
7	SNNP	MIZAN TOWN	MESEGANA AKADAME	106	8-12	47	10.6%	2.1%	20.94	47	23.4%	0.0%	2.21
8													
9													
10													
11													
12													

10. Vous pouvez également indexer les colonnes dans une fonction INDEX EQUIV. Ce qui veut dire que vous n'êtes pas obligé de préciser quelle colonne vous cherchez. Si vous souhaitez le faire, consulter la rubrique Aide sur les fonctions et utiliser Google pour comprendre !

7.6 Compléter 'STH global'

1. Compléter l'ensemble des informations de STH global.
2. Il vous faut une seule valeur pour chaque colonne
3. Les outils dont vous pourrez avoir besoin sont les mêmes que précédemment.

7.7 Compléter 'STH par école'

4. Ouvrir la feuille 'STH par école'
5. Compléter l'ensemble des informations de STH par école.
6. Vous avez besoin d'une ligne différente pour chaque école
7. Les outils dont vous pourrez avoir besoin sont les mêmes que précédemment.

Veiller à enregistrer les fichiers sur lesquels vous avez travaillé.

Session 8 Créer des tableaux de résultats

8.1 Synthèse

Maintenant que nous avons un tableau des résultats, nous pouvons créer des graphiques !

C'est comme ça que la plupart des personnes verront les résultats.

Il existe de nombreuses manières de créer des graphiques dans Excel. Le mieux est d'apprendre par la pratique mélangée à un peu de Google. Nous allons décortiquer les points principaux pour vous aider.

Tout d'abord ouvrir votre fichier résultats (par ex., ETH Impact 2015 Résultats 15-10-2016) et l'enregistrer dans le dossier "4 Analyse des données" sous le nom "créer graphiques" par ex : "ETH Impact 2015 Créer graphiques 15-10-2016".

8.2 Créer un histogramme de la prévalence de *S. haematobium* et *S. mansoni*

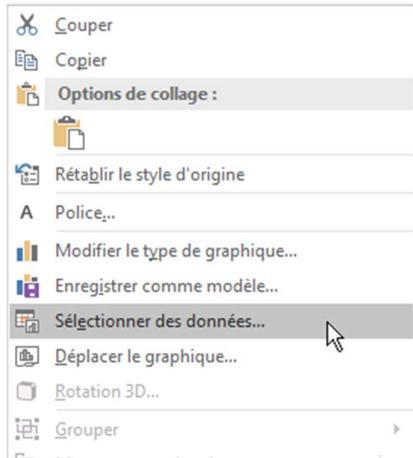
1. Cliquer dans la feuille 'SCH global' en dehors des données

	B	C
S. mans par KK	Nb élèves	S. mans par KK % positifs
	297	40.74%

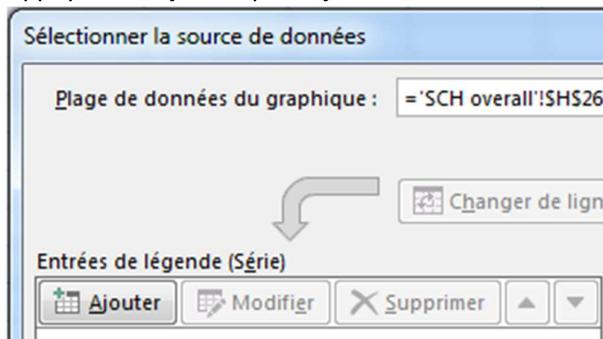
2. Aller sur Insérer -> Graphiques -> Colonnes -> Histogramme 2D

	B	C	D	E	F	G	H	I
S. mans par KK	S. mans par KK	S. mans par KK %	S. mans par KK %	S. haem par FU	S. haem par FU	S. haem par FU %	S. haem par FU %	S. haem par FU %
Nb élèves	% positifs	infection massive	moyenne opg	Nb élèves	% positifs	infection majeure	moyenne opg	
297	40.74%	4.38%						

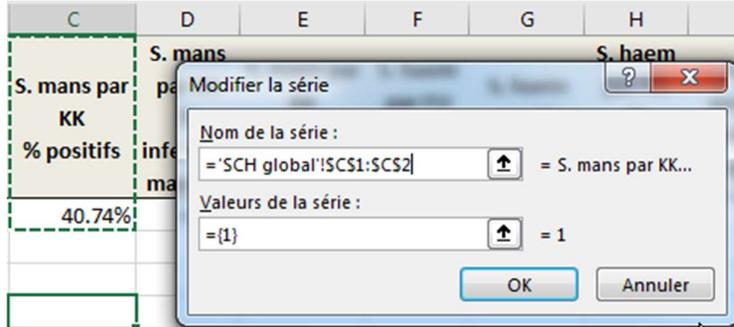
3. Vérifier qu'il existe désormais un graphique vierge dans le classeur.
4. Placer les données dans le graphique en faisant un clic droit dans le graphique et en cliquant sur 'Sélectionner des données'



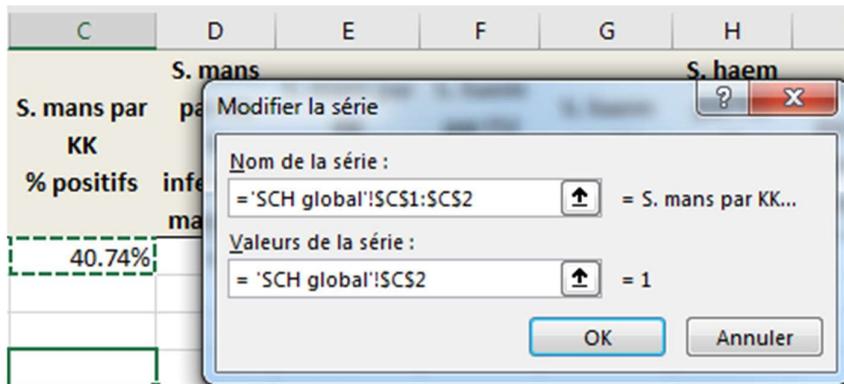
5. Appuyer sur 'Ajouter' pour ajouter des données



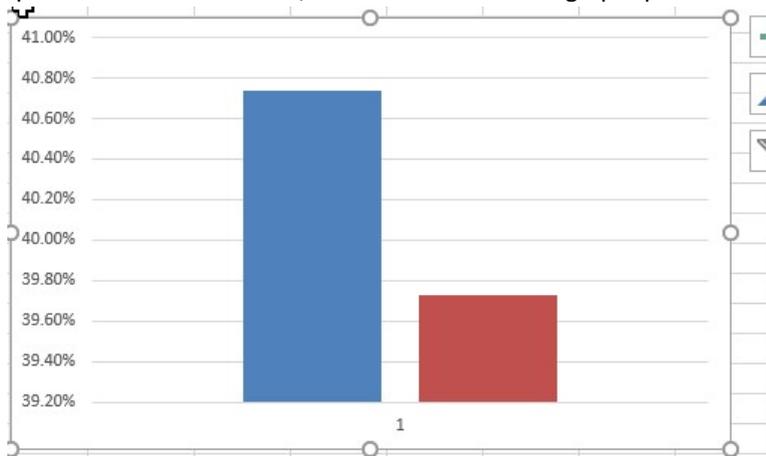
6. Pour le nom de la série, cliquer sur 'S. mans par KK % positif'



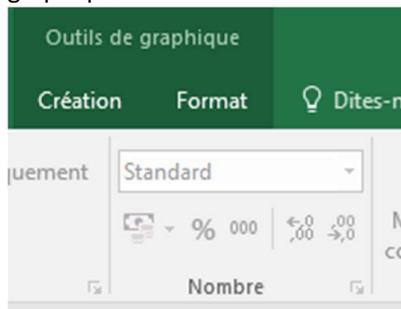
7. Pour la valeur de la série, cliquer sur le nombre correspondant de la prévalence de S. mans



8. Appuyer sur OK et faire de même pour S. haematobium. Quand vous appuyez sur OK et quittez toutes les fenêtres, vous devez obtenir un graphique comme celui-ci :



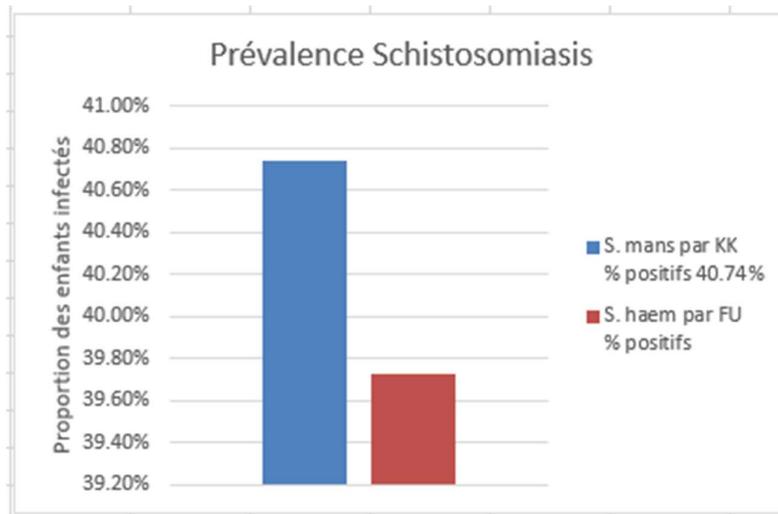
9. Il doit être mis en forme. Pour cela, cliquer dans le graphique puis aller dans Outils de graphique



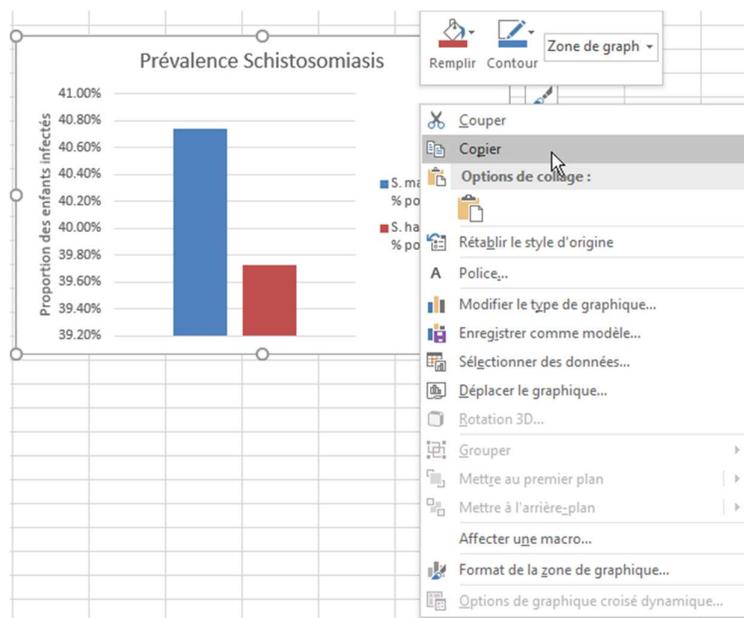
10. L'onglet « Création » est le plus utile pour l'instant. Ajouter des éléments au graphique à l'aide de 'Ajouter un élément graphique' ou consulter les options de 'Disposition rapide'

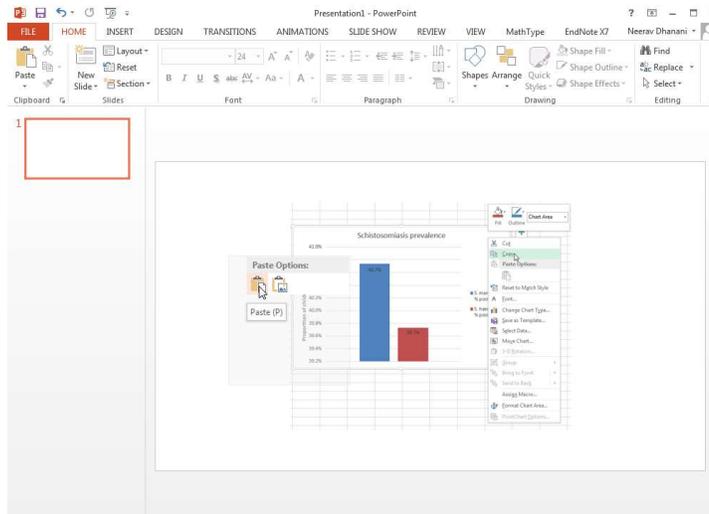


11. Pour modifier les éléments individuels dans le graphique, faire un double clic sur les éléments et utiliser les formats à droite.
12. Essayer de mettre votre graphique en forme comme ceci :



13. Une fois votre graphique terminé, faire un copier-coller dans MS Powerpoint :
Faire un clic droit dans le graphique puis Copier, ensuite clic droit dans la diapo de
Powerpoint puis Coller.



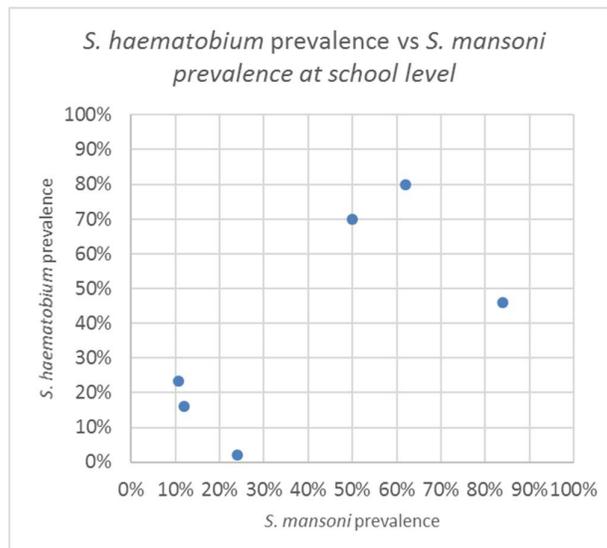


14. Enregistrer la diapo Powerpoint dans le dossier des résultats '5 Résultats' et lui donner un nom pertinent

Par ex., 'ETH_2015_prévalence_SCH_globale_25-10-2016'

8.3 Comparer la prévalence de *S. mansoni* et *S. haematobium* dans chaque école

Créer le graphique suivant :



8.4 Créer d'autres graphiques

Explorer les divers types de graphiques et créer d'autres modélisations de données. Par exemple, vous pouvez modéliser la prévalence de SCH par école ou district.

SESSION 9 : Saisie des données

9.1 Introduction et principes généraux de saisie des données avec Excel

Excel est un outil idéal pour extraire des informations importantes des données que vous avez recueillies (lors d'enquêtes, tests, expérimentations, etc.)

Néanmoins, pour que cela soit pertinent, il est essentiel que les données soient saisies de façon aussi exacte que possible. Dans cette session, nous vous fournirons quelques lignes directrices pour vous aider à saisir vos données aussi précisément et utilement que possible.

L'élément le plus important de la saisie de données est la personne qui saisit les données : la tâche peut être répétitive et ennuyeuse, il faut faire de gros efforts de concentration dans la durée. C'est un facteur que vous ne maîtrisez peut-être pas. Nous allons donc vous présenter des outils Excel pour vous aider. Nous vous donnerons également quelques lignes directrices d'ordre général concernant la mise en page de vos données de sorte à en tirer le meilleur lorsque vous les analyserez.

NB. Pour réduire les risques d'erreur lors du processus de saisie des données, cela vaut vraiment la peine de prendre le temps et de faire l'effort de bien concevoir vos formulaires de saisie de données Excel.

9.2 Mise en page et formats

Lignes :

Une ligne doit contenir toutes les informations disponibles pour une unité expérimentale. Par exemple, si vous recueillez des données parasitologiques sur les enfants de l'école, une ligne doit contenir toutes les informations pour un enfant. Si vous recueillez des données de niveau de l'école (emplacement, nombre d'élèves, etc.), une ligne doit contenir toutes les informations pour une école.

Colonnes :

Une colonne doit contenir toutes les données disponibles pour une variable (par ex. âge, nombre d'œufs d'ankylostomiase). Le format des données doit être homogène ; si une variable est un nombre, toutes les valeurs saisies dans cette colonne doivent être des nombres.

	C	D	E	F	G	H	I	J	K
	Distric	Code du district (XXX)	Latitude arrivée (XXX.XXXXX)	Longitude arrivée (XXX.XXXXX)	Latitude départ (XXX.XXXXX)	Longitude départ (XXX.XXXXX)	Nom de l'école	Code de l'école (XXX)	Nom du c
1									
2	BELB WUHA	120	12.18.550	037.45.04	12.118.542	037.45.075	BELB W	101	WASHUN WE
3	DAWA CHEFE	117	10.46.084	039.51.800	10.46.084	039.05.799	BILLECHA	102	TAR
4	KEMISE	117	10.52.371	039.51.607	10.52.370	039.05.670	SERTIE	103	REGASSA
5	GURAFERDA	123	06.83.725	035.29.902	06.83.725	35.29.872	OTEWA No 2	104	Mengist
6	GURAFERDA	123	06.86.817	035.34.801	06.87.925	035.25.734	ORENITA	104	AYENEW
7	MIZAN TOWN	122	06.99.501	035.58.899	06.99.506	035.58.890	MESEGAN AKADAME	106	TATEK H
8									
9									

1 Unités expérimentales (écoles dans ce cas)

2 Variables

9.3 Fonctions de validations des données

Qu'est-ce que la validation de données ?

Dans Excel, la validation de données désigne une série de méthodes qui vous permet de placer des restrictions sur le type de données qui peut être saisi dans les cellules. Vous pouvez configurer ce paramètre de sorte que des données incorrectes ne puissent pas être enregistrées, ou qu'un avertissement s'affiche si l'utilisateur saisit des données incorrectes.

Excel propose un large éventail d'options pour la validation des données. Dans cette session, nous ne fournissons pas une liste exhaustive des nombreuses fonctions disponibles, mais nous vous montrerons comment trouver les options et vous donnerons deux exemples de technique. Nous vous montrerons également où trouver de l'aide concernant les différentes options.

Dans cette démo, nous ferons référence à des variables de l'ensemble de données pratique mais ces techniques sont évidemment applicables à n'importe quel type de données dont vous disposez.

Voici quelques exemples de ce que vous pouvez faire avec la validation de données (extrait des pages d'aide Excel) :

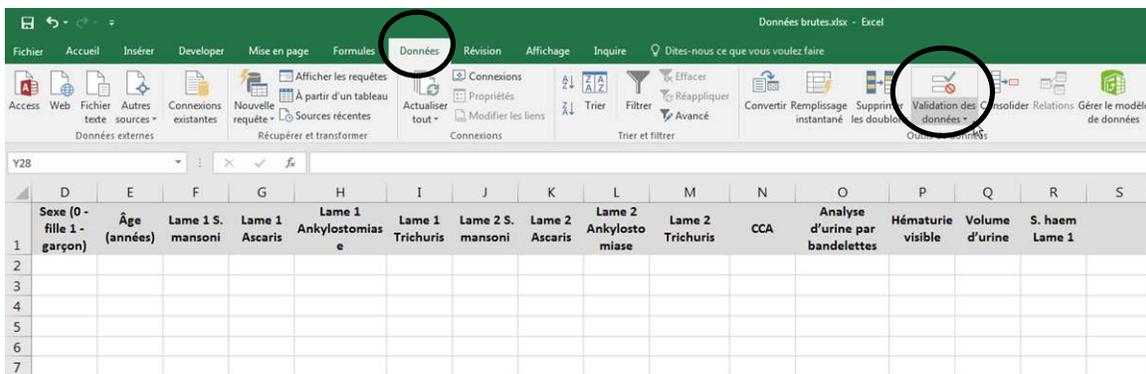
- **Restreindre les nombre en dehors d'une plage précise** Par exemple, vous pouvez définir un seuil de déduction de deux fois le nombre d'enfants dans une cellule précise.
- **Restreindre les dates en dehors d'un certain intervalle** Par exemple, vous pouvez définir un intervalle entre aujourd'hui et dans 3 trois jours.
- **Restreindre les horaires en dehors d'un certain intervalle** Par exemple, vous pouvez définir un intervalle pour servir le petit déjeuner entre l'heure d'ouverture du restaurant et 5 heures après.

- **Limiter le nombre de caractères** Par exemple, vous pouvez limiter le texte autorisé dans une cellule à 10 caractères maximum. De même, vous pouvez définir une longueur précise pour un champ nom complet (C1) égal à la longueur actuelle d'un champ prénom (A1) et d'un champ nom (B1), plus 10 caractères.
- **Valider des données sur la base de formules ou de valeurs dans d'autres cellules**

Comment utiliser la validation de données

Pour utiliser la validation de données dans une colonne, mettre en surbrillance une colonne en cliquant dans la description de la colonne en haut de la feuille (A, B, C etc.)

Puis aller dans le ruban 'Données' et cliquer sur 'Validation des données'



Cette étape vous ouvre un menu semblable à celui-ci :

Validation des données

Options | Message de saisie | Alerte d'erreur

Critères de validation

Autoriser : Tout Ignorer si vide

Données : comprise entre

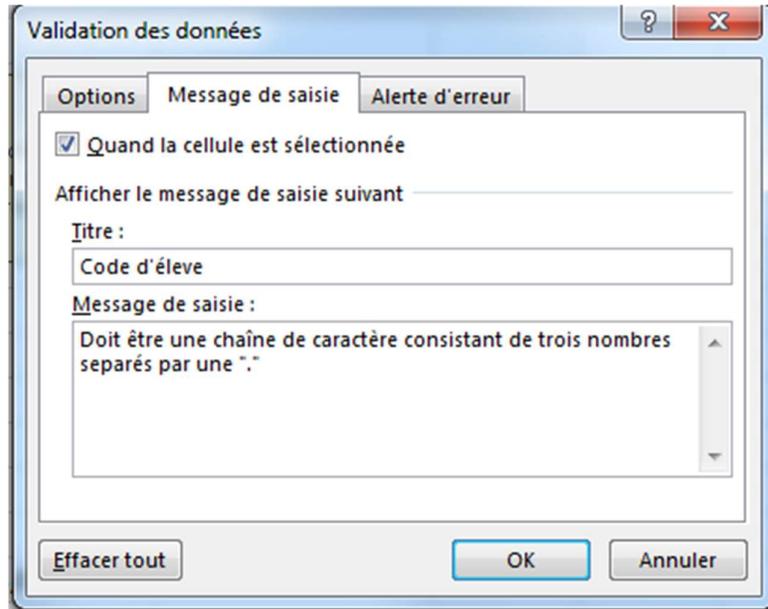
Appliquer ces modifications aux cellules de paramètres identiques

Effacer tout OK Annuler

Vous permet de choisir le type de données que vous souhaitez saisir – par ex., nombre entier, texte, date, heure. Vous pouvez également préciser les types de données avec des fonctions.

Ici vous pouvez choisir d'ignorer les cellules vides

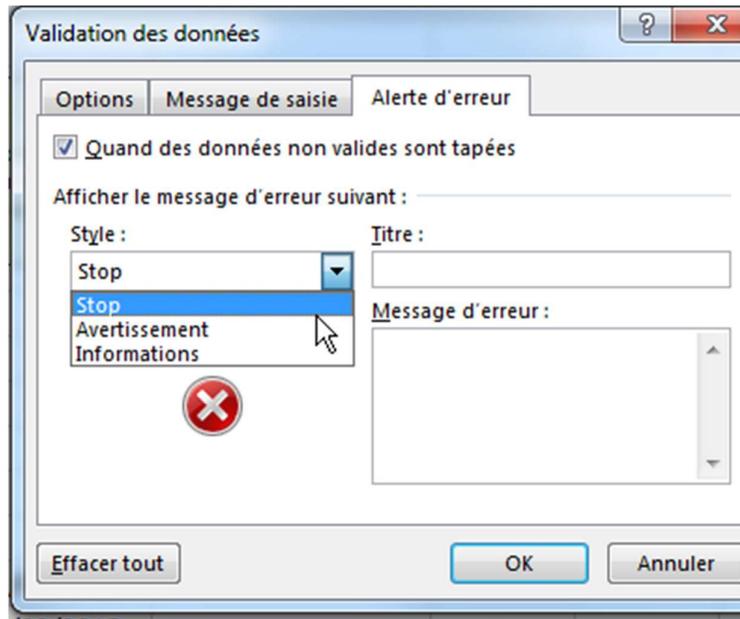
Ici vous définissez quelle plage de données est autorisée (par ex., entre 1 et 10 uniquement)



L'option 'Message de saisie' vous permet de créer un message qui s'affiche lorsqu'une cellule est sélectionnée. Cette option peut être utile pour rappeler à la personne saisissant les données quel type de données doit être saisi.

	Numéro d'identificatio n (DDD.SSS.NNN)	Date de l'enquête MM-AAA
1		
2		
3		
4		
5		
6		
7		

Code d'élève
Doit être une chaîne
de caractère
consistant de trois
nombres séparés par
une "."

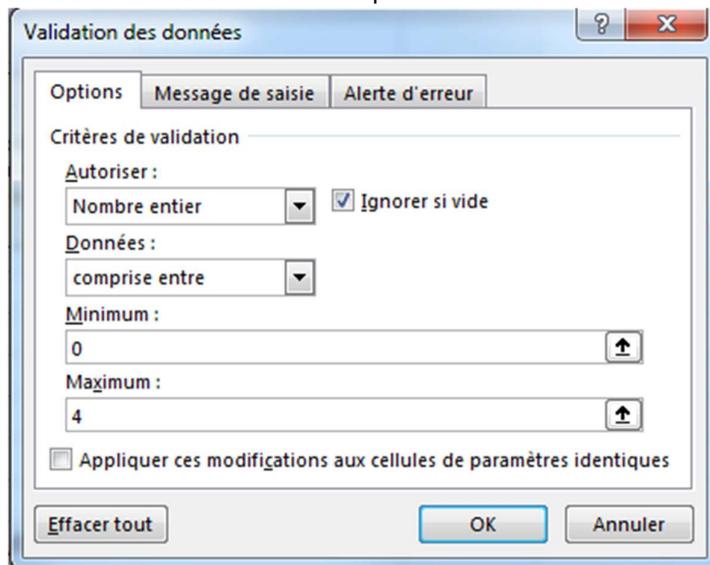


La page « alerte d'erreur » vous permet de paramétrer ce qui doit se passer en cas de saisie de données incorrectes. Vous pouvez interrompre la saisie des données ou vous pouvez créer un message d'avertissement en laissant la saisie de données se poursuivre.

Démo : utiliser la validation de données dans les informations CCA

Nous savons que les résultats CCA doivent contenir un nombre entier compris entre 0 et 4. Nous allons appliquer ces restrictions à la colonne CCA.

1. Aller sur la première cellule sous l'en-tête de la variable 'CCA' (cellule N2) et appuyer sur Ctrl+Shift+flèche bas pour mettre en surbrillance toute la colonne.
2. Aller dans le ruban 'Données' et sélectionner 'Validation des données'
3. Sélectionner Nombre entier compris entre 0 et 4 :



4. Saisir un message de saisie adapté



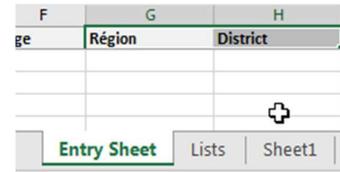
5. Saisir un message d'erreur



6. Cliquer sur OK et essayer de saisir des données dans la colonne CCA. Que se passe-t-il si vous tapez des données incorrectes ?

9.4 Listes dépendant de la valeur d'une autre cellule

Il peut arriver que les valeurs possibles d'une cellule dépendent de l'option choisie dans une autre cellule. Par exemple, si l'on a deux colonnes, une pour le nom de la région et une pour le nom du district, il pourrait être judicieux de n'autoriser que les districts présents dans la région choisie dans la colonne Région.

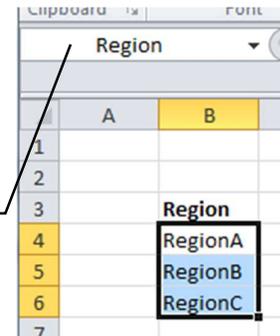


C'est possible grâce à la fonction INDIRECT.

Pour définir une validation des données dans la feuille de saisie décrite précédemment, créer une feuille séparée intitulée *listes*.

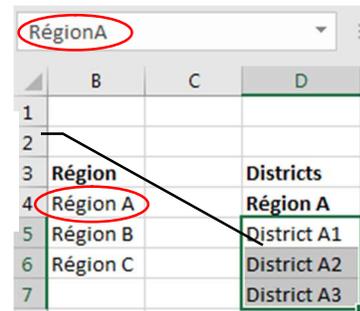
1. À présent, aller dans la feuille *listes* et créer une liste des noms de régions possibles comme illustré à droite. Ensuite, mettre en surbrillance la région contenant les valeurs de la liste et lui attribuer le nom 'Région'.

Saisir le nom 'Région' dans ce champ.



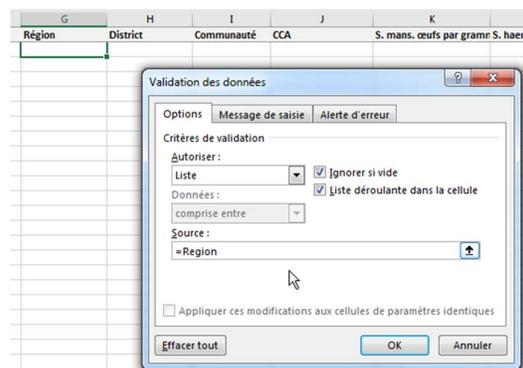
2. Ensuite, créer une liste des noms de districts de la Région A, en lui attribuant exactement le même nom que l'entrée correspondante dans la liste de régions, à savoirs 'Région A'.

C'est la liste des districts de la Région A. Cette liste s'appelle Région A.



3. Faire de même pour la Région B et la Région C, là encore en leur attribuant le nom de la région concernée.

4. Ensuite, aller dans la feuille de saisie et mettre en surbrillance la cellule située sous 'Région'. Cliquer sur Validation des données dans le ruban de Données et choisir 'Liste' dans le champ 'Autoriser' de l'onglet Options. Placer le curseur dans le champ Source, cliquer sur le bouton Réduire et surligner la zone de cellules contenant le nom 'Région'

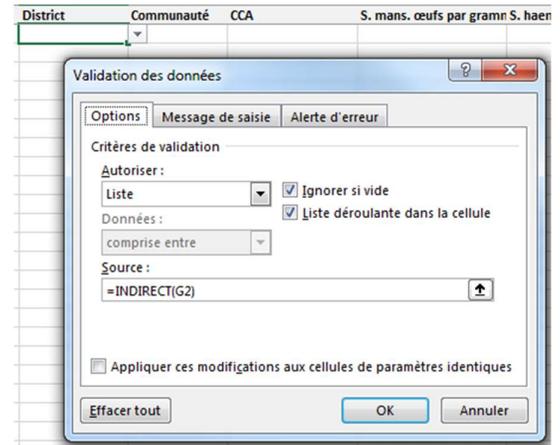


dans la feuille Listes qui contient les noms de région.

5. Pour définir une liste d'options de saisie pour le District, choisir à nouveau l'option Liste dans la boîte de dialogue de Validation des données. Mais dans le champ Source, saisir =INDIRECT(A2) A2 faisant référence au district choisi pour cette ligne.

Région	Districts
Région A	Région A
Région B	District A1
Région C	District A2
	District A3

Feuille de saisie **Listes**



6. Les options du District doivent désormais dépendre de la valeur choisie pour la Région comme illustré à droite.

A	B	C
Region	District	
RegionC	<div style="border: 1px solid black; padding: 2px;"> DistrictC1 DistrictC2 </div>	

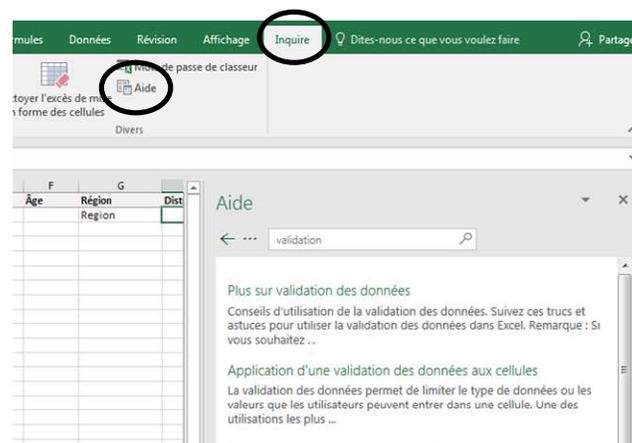
Maintenant, à vous : Explorer les options

Explorer les options avec les autres variables et essayer de définir des restrictions de validation des données adaptées.

Pour en savoir plus, vous pouvez utiliser la fonction intégrée Aide :

Ouvrir la boîte Aide et taper « validation des données » puis cliquer sur la première option 'Application d'une validation des données aux cellules'.

Vous obtiendrez des consignes détaillées sur l'ensemble des options disponibles.



Créer des synthèses de données - antisèche

0. Sélectionner la feuille de synthèse concernée dans le Classeur de nettoyage des données ou faire une copie et lui attribuer un nom adapté (par ex. Synthèse brute élève).

Fenêtre des informations basiques (en haut à gauche de la feuille)

1. Saisir le nom de l'ensemble de données.
2. Saisir le nombre de lignes que contient votre ensemble de données (nombre d'observations).
3. Saisir le nombre de colonnes que contient votre ensemble de données (nombre de variables).
4. Saisir le nom de la variable de l'identifiant unique pour votre ensemble de données.

Informations de la synthèse (le principal tableau de la feuille)

5. Copier les noms des variables de votre ensemble de données vers la colonne « Nom de variable » (utiliser copier et coller + transposer).
6. Noter si chaque variable est un facteur ou un nombre.
7. Pour les facteurs – noter le nombre de niveaux uniques (valeurs), vous pouvez utiliser des tableaux dynamiques croisés ou la fonction filtrer pour trouver rapidement.
Pour les nombres – noter les valeurs minimum, moyenne et maximum.
8. Rechercher les valeurs manquantes (utiliser la fonction filtrer) et les stocker.
9. Consulter la synthèse de chaque variable pour déterminer si elle est raisonnable et conforme à vos attentes.
10. Apporter les modifications qui vous semblent nécessaires et surtout ne pas oublier de noter précisément les modifications apportées !