

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 :

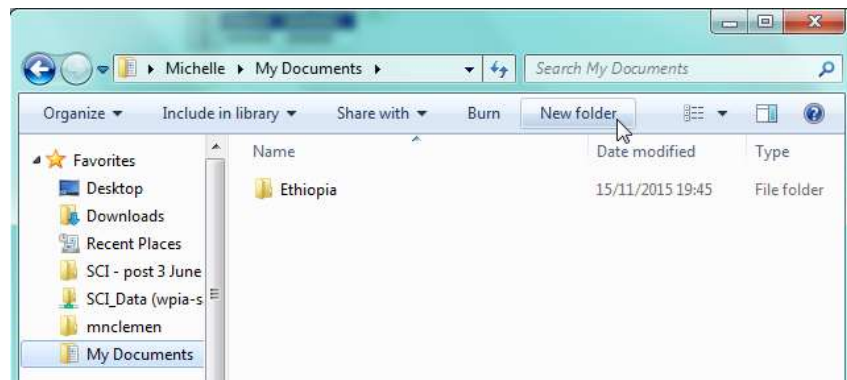
1. Ouvrir l'Explorateur Windows en faisant un clic droit sur le bouton Windows en bas à gauche



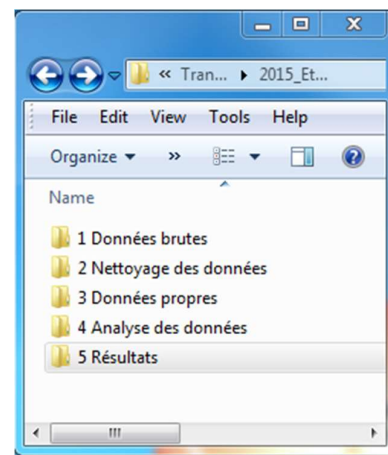
de l'écran puis cliquer sur « Ouvrir l'Explorateur Windows »



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 :

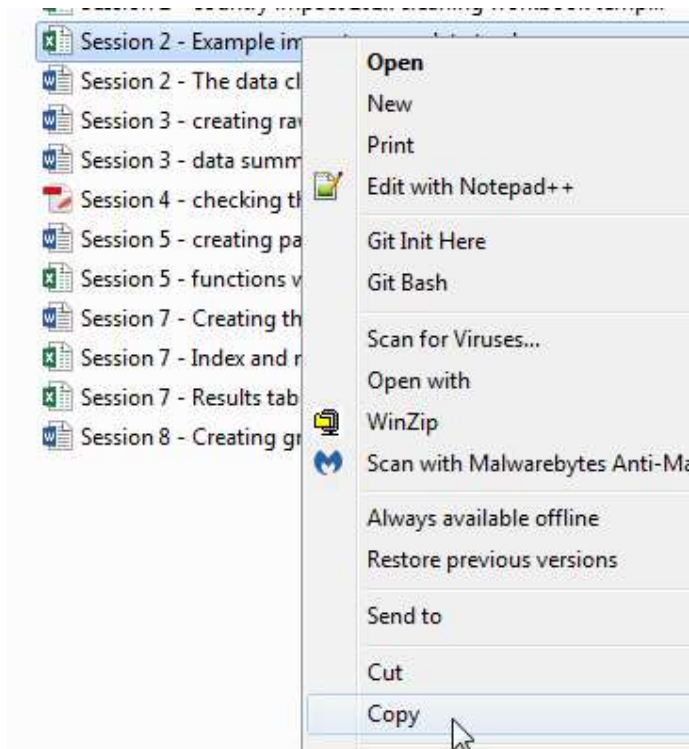


2.4 Enregistrer des données brutes

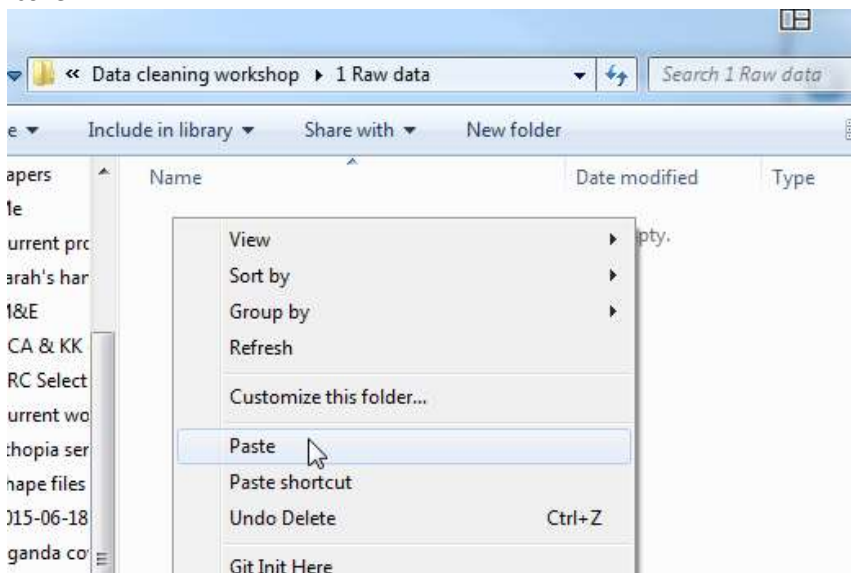
Chez SCI, nous veillons toujours à enregistrer une copie des données brutes reçues pour pouvoir les consulter ultérieurement si nécessaire. Après enregistrement, on ne touche plus à ces données.

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'

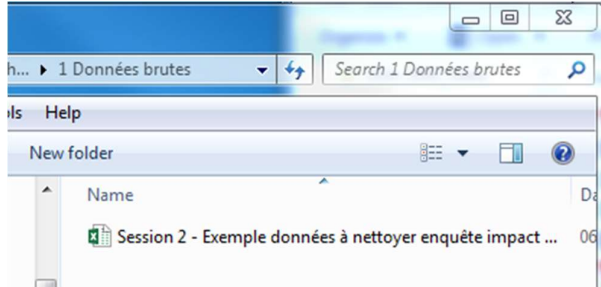


6. Vérifier que votre dossier '1 Données brutes' ressemble à ceci :

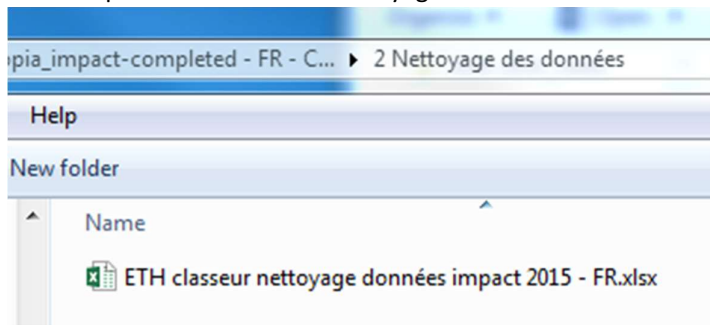
2.5 Enregistrer le classeur de nettoyage des données

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

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



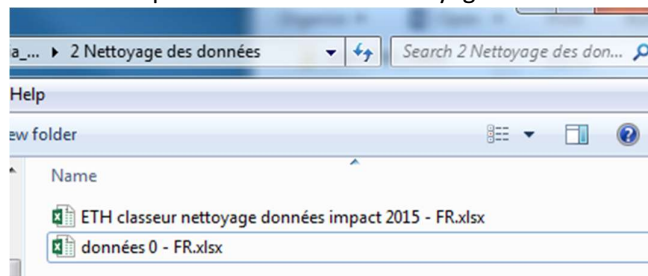
- Renommer le 'Modèle classeur nettoyage des données' de façon plus précise, par exemple 'classeur nettoyage des données pays impact année', où l'on remplace pays et année par le pays et l'année concernés par les données recueillies. L'exemple de données contient des noms d'Éthiopie, mais les données sont mélangées. Nous appellerons ce classeur 'classeur nettoyage des données ETH impact 2015'
- 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.

- Ouvrir les données brutes
- Enregistrer les données brutes dans le dossier '2 Nettoyage des données' Tout d'abord, appuyer sur fichier – enregistrer sous
- Puis enregistrer dans '2 Nettoyage des données' sous le nom 'données 0'
- Fermer tous les fichiers sauf 'Classeur nettoyage des données ETH impact 2015' et 'données 0'
- 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 ?

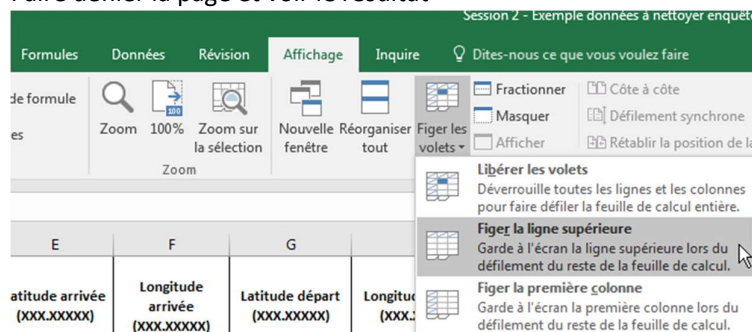
J'ai remarqué que :

- 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 Figer 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

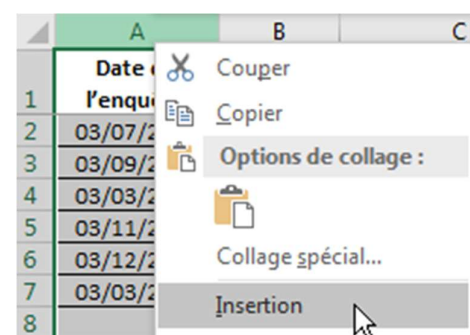


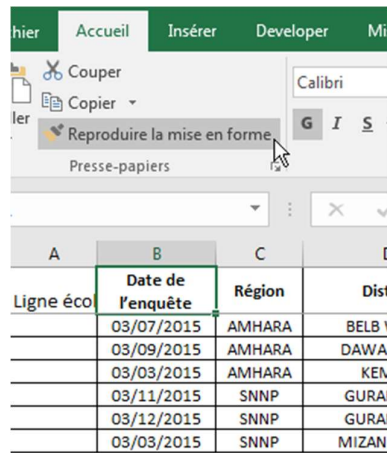
4. Faire de même pour la feuille 'élève'

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'

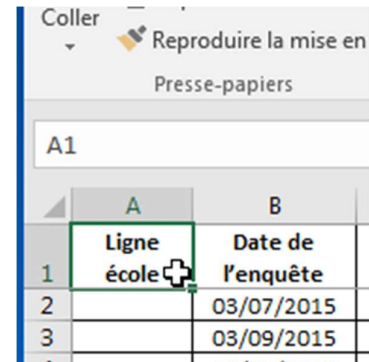
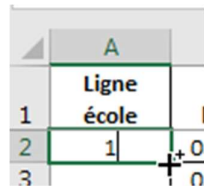




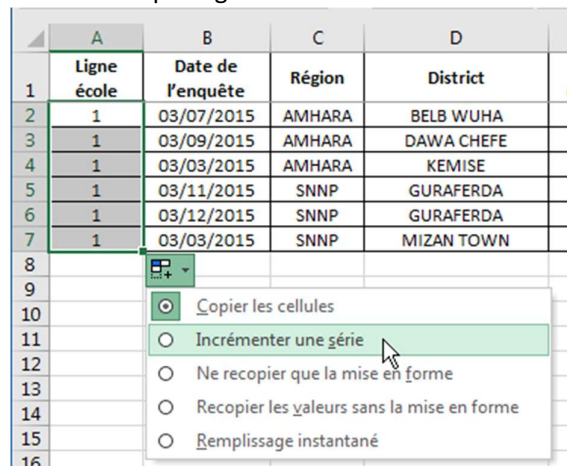
4. Cliquer une fois sur la cellule A1 pour copier uniquement la mise en forme dans cette colonne

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.

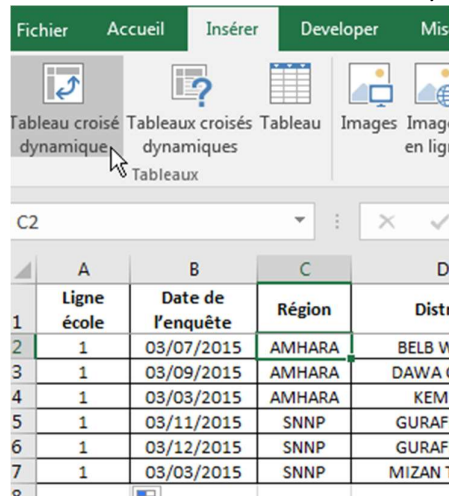
2.11.1 Tableaux croisés dynamiques

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 !

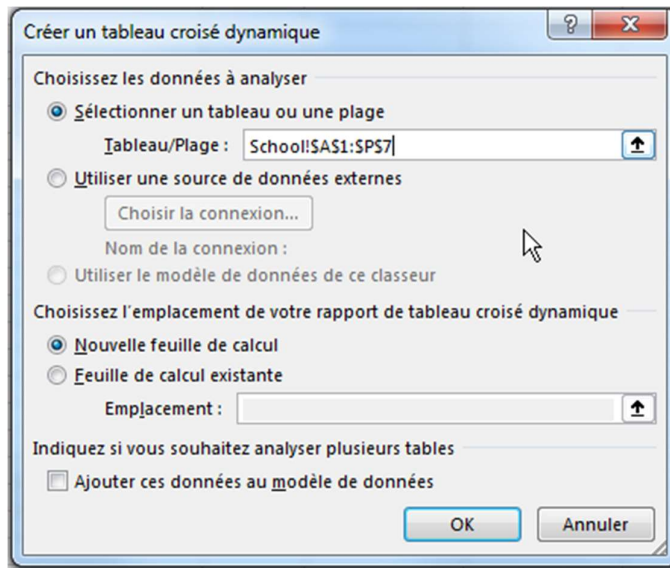
2.11.2 Vérifier que l'ID de l'école est unique.

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.

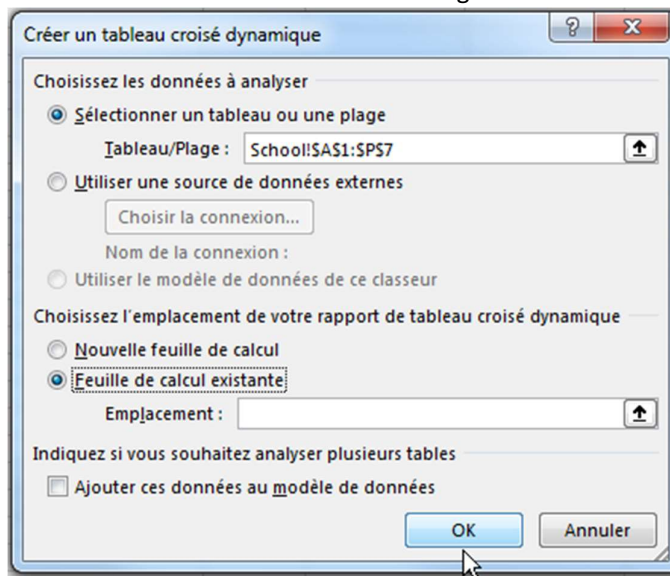
1. Prendre la feuille des écoles et cliquer une fois dans les données, par exemple la cellule C3
2. Aller sur Insertion -> Tableau croisé dynamique



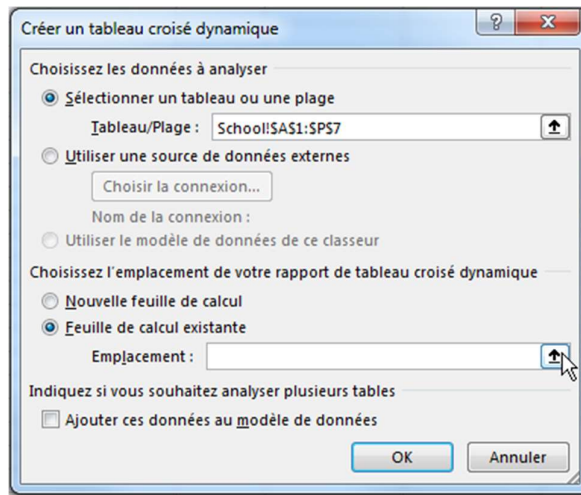
3. Vérifier que toutes les données sont incluses dans la section Tableau/Plage de la fenêtre de dialogue. *Cette étape est primordiale !!!* On doit y voir apparaître la feuille de données École puis les cellules A1 à P7.



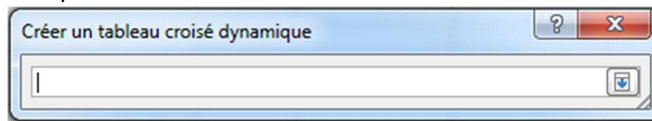
4. Nous allons placer le tableau dans 'Vérification ID école'. Pour cela, cliquer sur 'Feuille de calcul existante' dans la fenêtre de dialogue



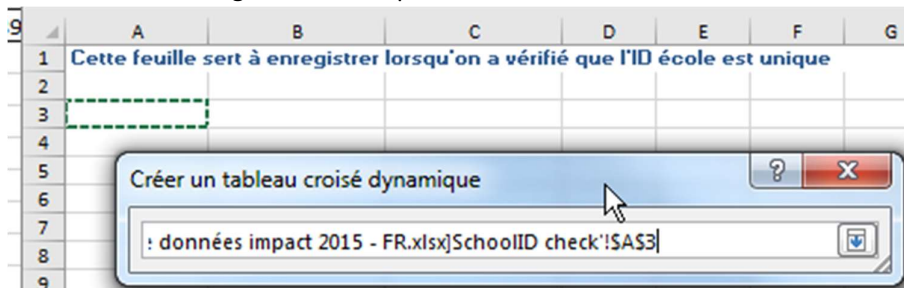
5. Ensuite, cliquer sur le petit bouton à droite du champ Emplacement



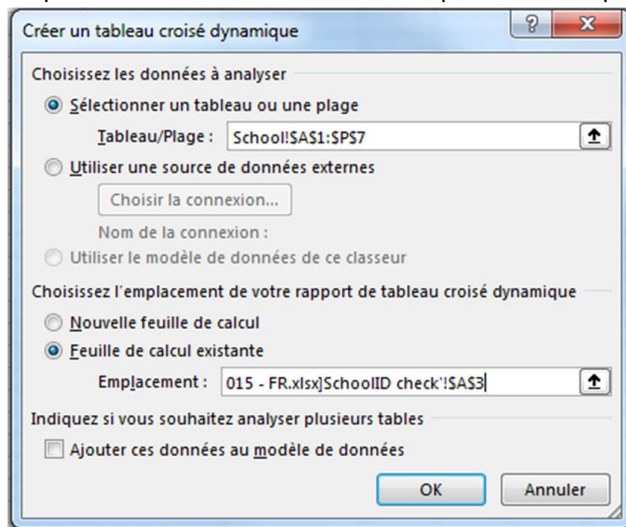
6. Une petite fenêtre comme celle-ci doit s'afficher



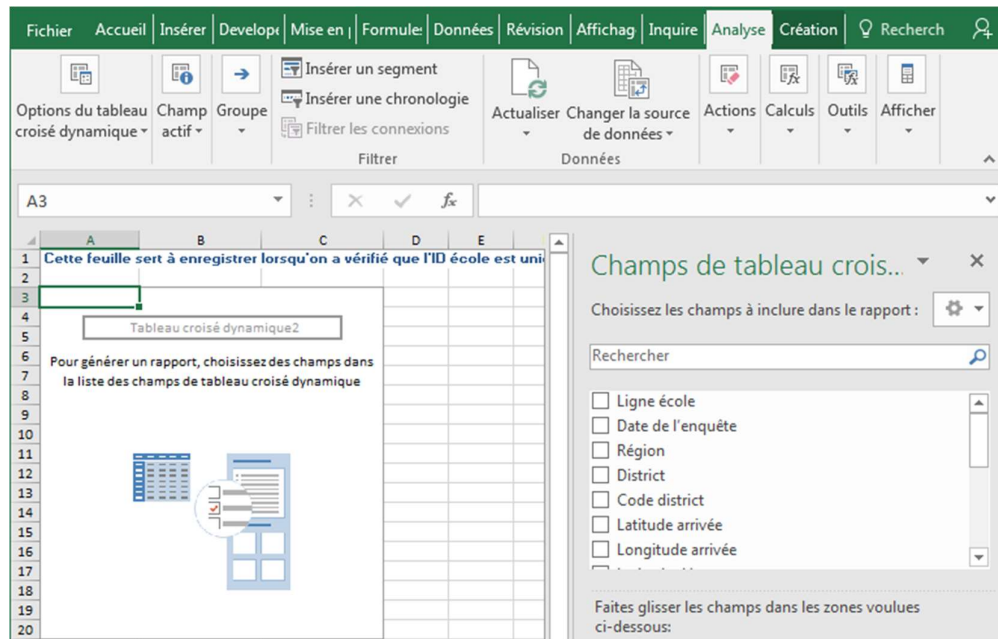
7. Cliquer sur le 'Classeur de nettoyage des données' puis aller sur la feuille 'Vérification ID école'. Cliquer sur la cellule A3. Vérifier que des pointillés entourent la cellule A3 et que la feuille et la cellule figurent dans la petite fenêtre comme ci-dessous



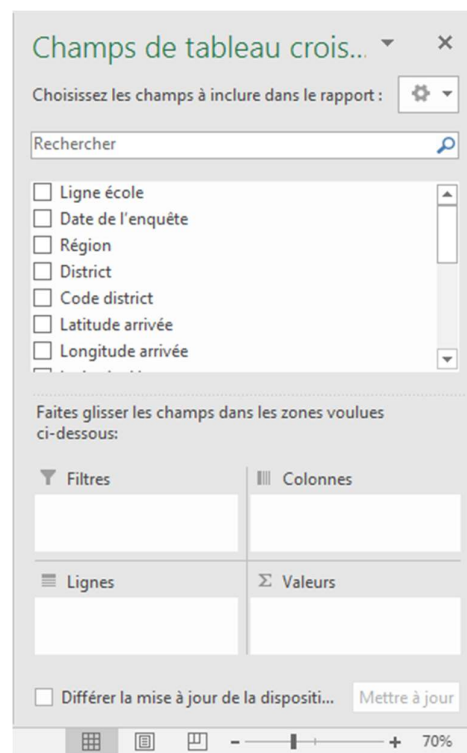
8. Cliquer sur le bouton à droite de cette petite fenêtre pour rouvrir la fenêtre précédente



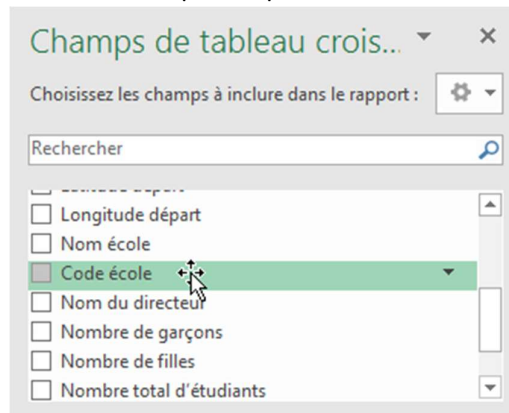
9. Cliquer sur 'OK' dans la fenêtre de dialogue
10. À présent, nous allons travailler dans le fichier 'modèle classeur de nettoyage des données'.
Il doit ressembler à ceci



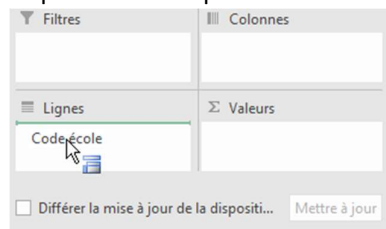
11. Cliquer dans la fenêtre du tableau croisé dynamique et une fenêtre de commande s'ouvrira à droite. La fenêtre de commande contient tous les noms de variable de l'ensemble de données école en haut et des cases vides en bas.



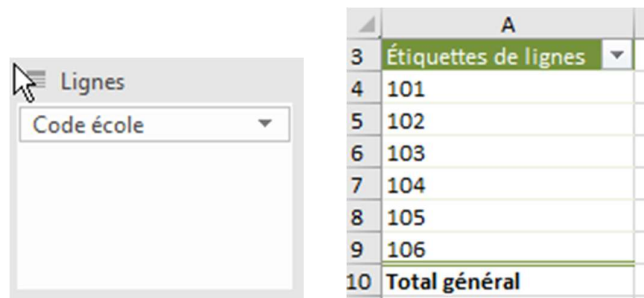
12. Nous souhaitons créer un tableau avec chaque Code d'école dans une ligne différente avec le nombre de fois où le code d'école est présent dans l'ensemble de données. Pour cela, passer le curseur sur Code d'école dans la fenêtre supérieure de sorte que la sélection soit en surbrillance puis cliquer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



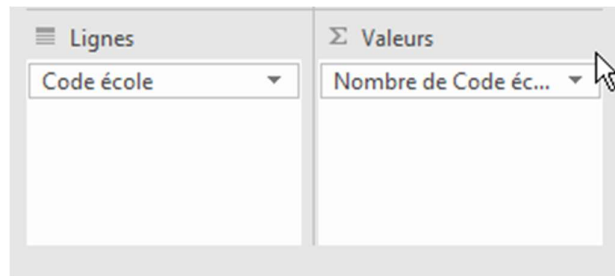
13. Déplacer le champ Code d'école vers les 'étiquettes de ligne' et relâcher la souris



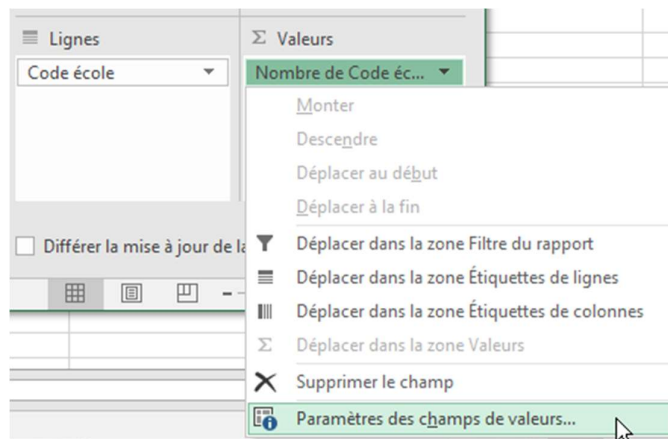
14. Vérifier que 'Code d'école' se situe dans la case Lignes et que les écoles apparaissent dans la colonne A de la feuille



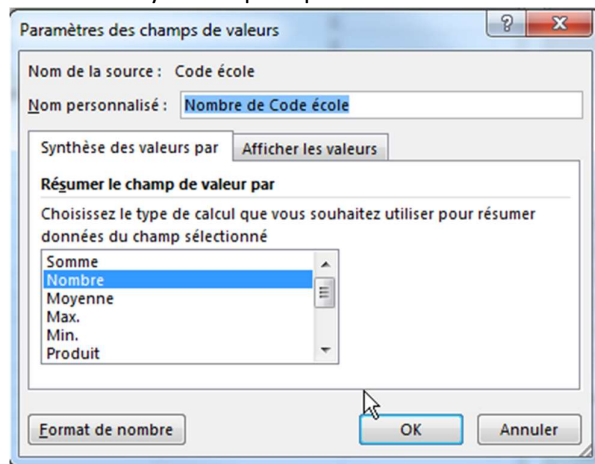
15. La colonne A contient tous les Codes d'école de l'ensemble de données École. À présent, il nous faut une autre colonne indiquant le nombre de fois où chaque Code d'école a été enregistré. Déplacer la Ligne école de la fenêtre en haut à droite pour qu'elle figure également dans le champ 'Valeurs'. 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 ressemblent 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'



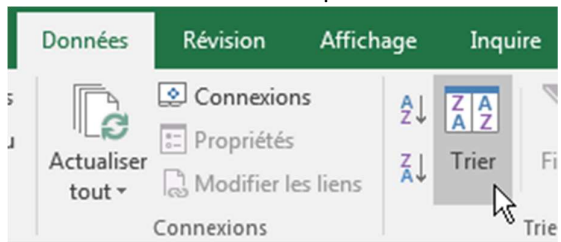
18. Consulter le tableau (colonnes A et B) pour vérifier qu'il ressemble au tableau ci-dessous. Le nombre de Codes école qui s'affiche principalement est 1, ce qui veut dire que ces Codes école n'ont été enregistrés qu'une seule fois, comme prévu. Toutefois, le Code école 104 a été enregistré deux fois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

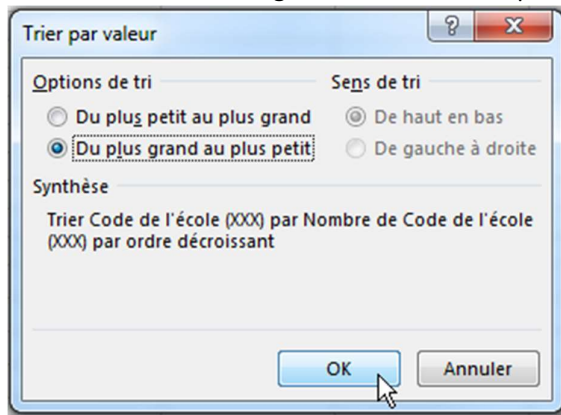
19. Si les écoles étaient nombreuses, on pourrait avoir du mal à voir que tous les chiffres ne sont pas des 1. On pourrait donc trier ces chiffres de sorte que les plus grands apparaissent en haut du tableau. Pour cela, cliquer dans la colonne nombre (n'importe quelle cellule entre B4 et B9)

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
OTewa 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
Volume	Volume d'urine	S. haem	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	Code école		
Lame 1			
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

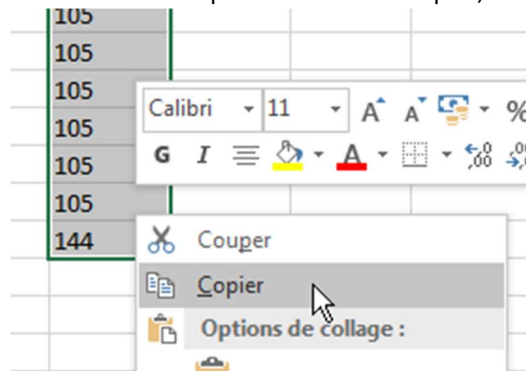
13. Aller jusqu'en bas (ctrl + flèche bas) pour vérifier que les toutes cellules s'affichent correctement.
14. Faire un double clic sur la cellule en bas et vérifier la formule pour voir si elle est correcte. Vérifier que la cellule T298 ressemble à ceci :

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.

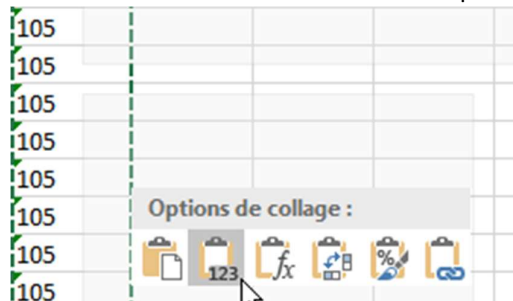
292	0	105
293	177	105
294	35	105
295	102	105
296	45	105
297	106	105
298	0	144

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

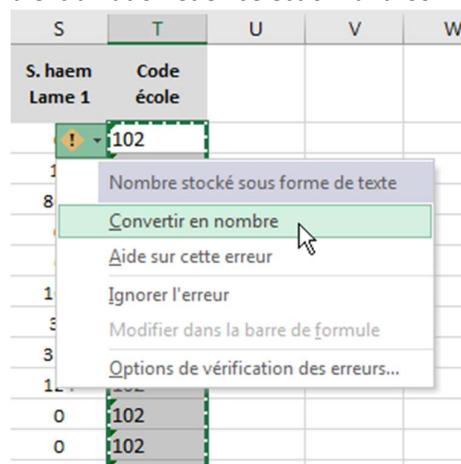


17. Veiller à être sur la cellule T2 en cliquant une fois sur la cellule T1 puis en appuyant sur la flèche bas.

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'

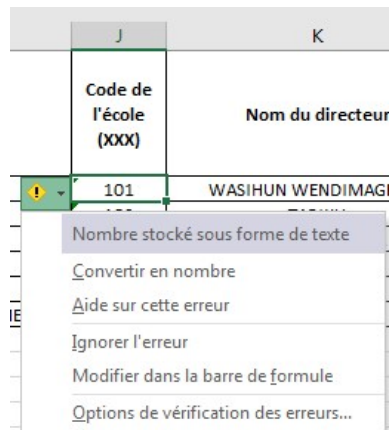


20. Les nombres seront à présent stockés uniquement comme des nombres. Pour vérifier ce point, faire un double clic sur l'une des cellules de la colonne T et voir si elle ne contient qu'un nombre.

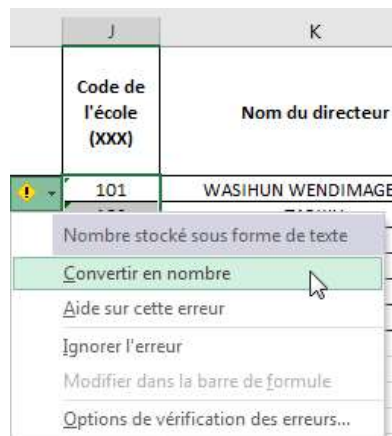
2.14 Vérifier que les mêmes codes sont contenus dans les ensembles de données écoles et élèves

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.

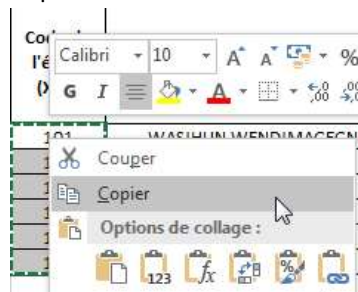
1. Pour commencer, copier les codes d'école de l'ensemble de données École dans la colonne 'ID école de l'ensemble de données école du fichier 'Vérifier ID élèves et écoles' du classeur de nettoyage des données. Pour cela, aller dans l'ensemble de données 'École' et trouver la colonne Code d'école. Des 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



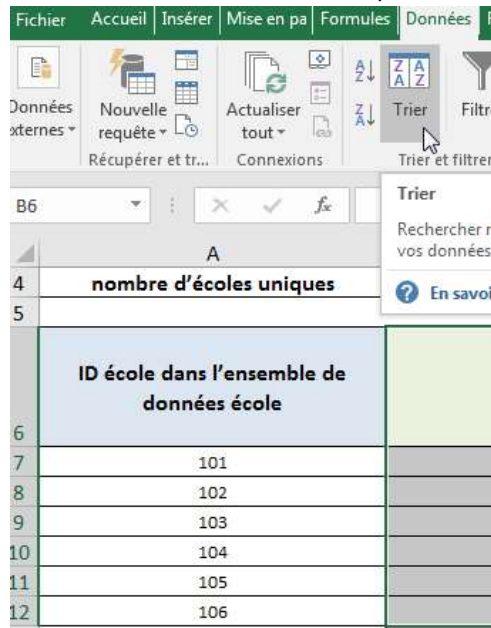
7. Vérifier que les données collées ressemblent à ceci :

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

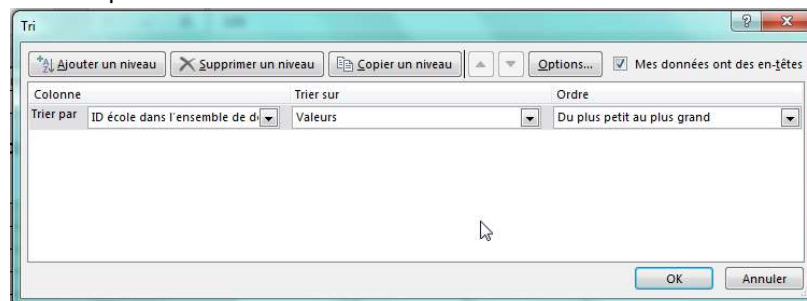
9. Vérifier que les Codes école sont triés correctement en sélectionnant toutes les données de la colonne de la cellule A6 à A12 puis aller dans Données -> Trier



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



11. Vérifier que vous triez 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

- 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

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

Supprimer les doublons

Supprimer les lignes en double dans une feuille de données.

Cette feuille sert à vérifier que les codes école dans les données école

école

élève

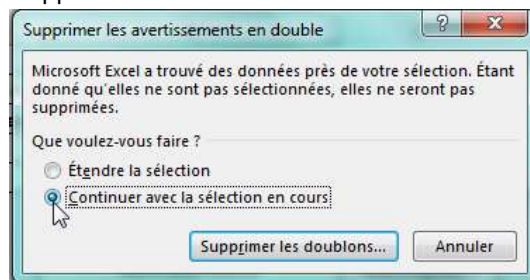
G

Supprimer les doublons

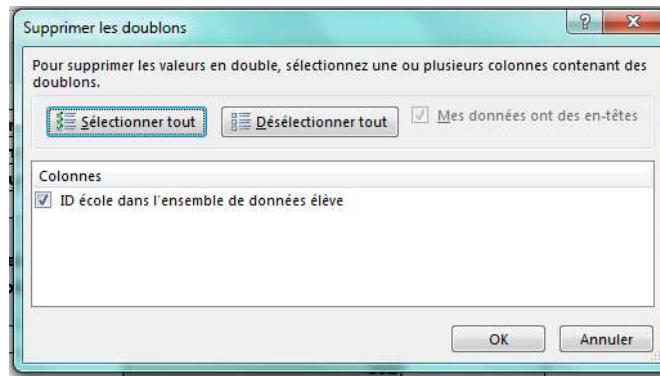
Vous pouvez sélectionner les colonnes dans lesquelles vous souhaitez vérifier la présence d'informations en double.

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

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



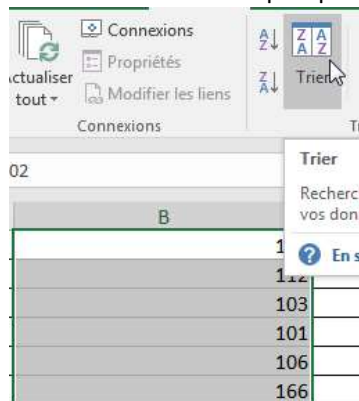
- Vérifier que seul 'ID école dans ensemble de données élève' est sélectionné et appuyer sur OK



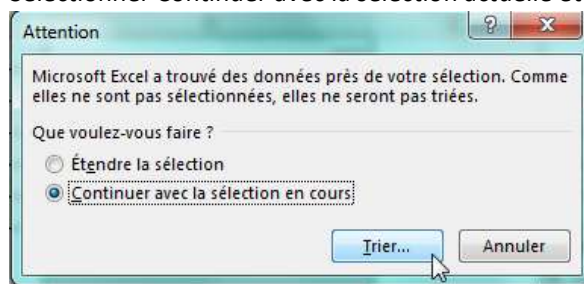
17. Vérifier la fenêtre de notification pour savoir combien il reste de valeurs uniques. Est-ce ce que l'on attend ?



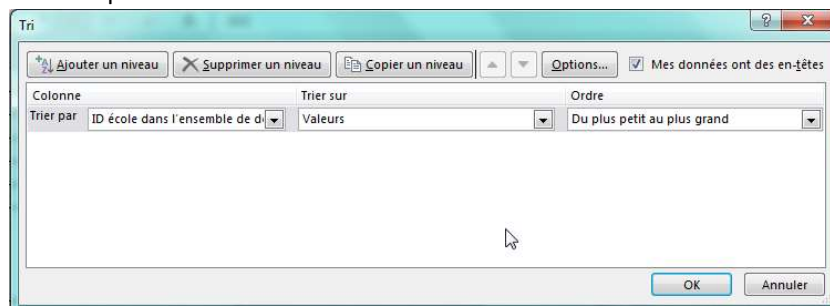
18. Trier les ID d'école uniques parmi les élèves en sélectionnant Données - Trier



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	

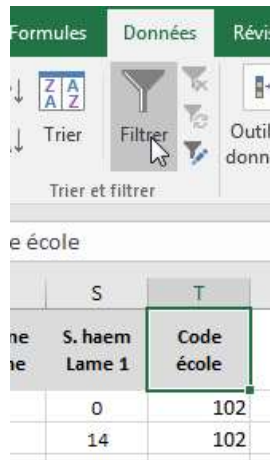
23. Faire un double clic sur la flèche en bas à droite de la cellule pour copier vers le bas

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

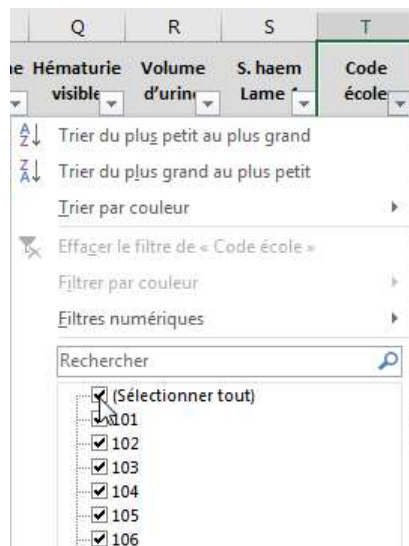
24. On voit alors quelles lignes indiquent un N signifiant que les ID ne sont pas identiques.

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

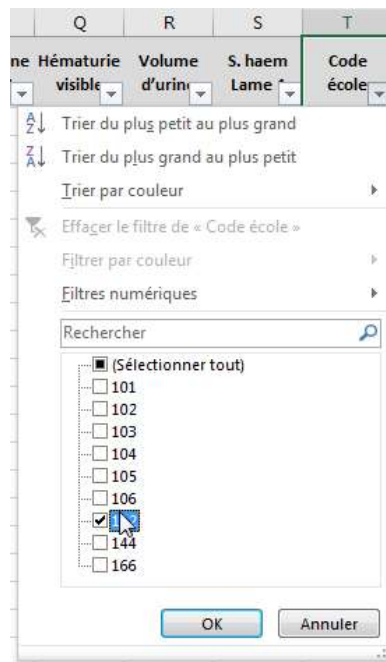
25. Toutes les écoles incluses dans l'ensemble de données école ont été également incluses dans l'ensemble de données élève. Néanmoins, certaines écoles de l'ensemble de données élèves n'ont aucune correspondance dans l'ensemble de données école. Il nous faut regarder ça de plus près pour savoir ce qu'il 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.



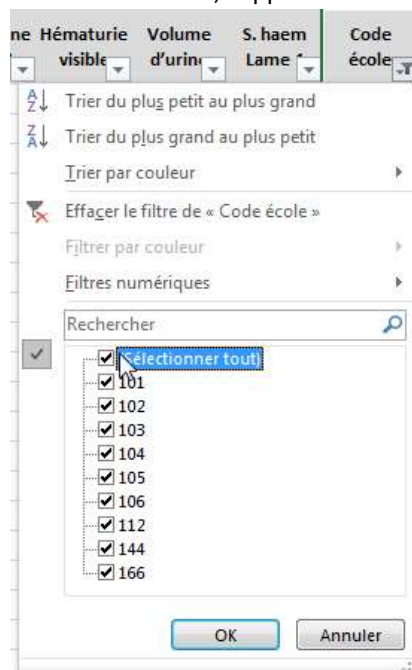
27. Puis cliquer sur 112 pour sélectionner uniquement les lignes où le code d'école est 112



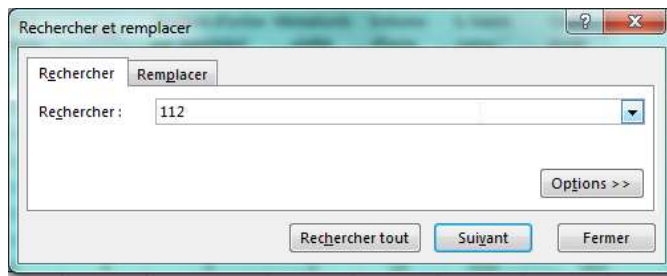
28. On voit à présent qu'il n'existe qu'une seule ligne avec le Code école 112

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.