## Première vision d’ensemble + premier test lancement flux

* Votre projet est très volumineux (beaucoup de petites boites dans la même fenêtre). Deux possibilités :
  + Vous séparez vos différents flux. Par exemple : faire un flux pour chaque nettoyage de fichier, puis un dernier où vous faites vos jointures pour le fichier final. (en pratique c’est ce que j’ai le plus vu si mes souvenirs sont bons)
  + Vous pouvez mettre des commentaires dans votre flux aussi, pour regrouper, je vous ai fait un exemple de façon de faire, à ma manière pour que je me repère mieux.
* Vous avez nommé les boites de manière compréhensible quand c’était nécessaire : ++
* Chose qui pourrait être intéressante à creuser : pour l’import du fichier, vous pouvez peut-être créer une variable qui défini le dossier dans lequel se trouve le fichier, que vous n’ayez que ça à changer et pas le lien complet à chaque fois. A creuser, ou pas !
* Dans le flux, mettez bien vos jointures dans l’ordre où elles sont faites. Là par exemple la stabilité politique est en première ligne alors qu’on ne la joint pas en premier. Le mieux étant quand même de faire différents flux je pense.

Flux stabilité politique

* Le fichier « StabilitePolitique.csv » n’existe pas dans le Git, attention ça ne devrait pas tourner chez vous non plus : dans les derniers fichiers il y a un accent (je ne sais plus si le même problème est présent pour mortalité et température)

Flux eau potable

Flux population

* Pourquoi ici vous utilisez un *file reader* et pas un *csv reader* ?
* Pourquoi le *ruleset editor* est après l’export du fichier propre ?

Flux mortalité

* Idem que pour population, pourquoi il y a encore des manipulations du fichier après l’export csv ? **[Fanny] C’est modification ne sont utiles que dans le cas d’un jointure (par exemple ajouter une colonne contenant « Both sexe » uniquement n’apporte pas d’information au fichier quand il n’est pas au sein d’une jointure.**

Flux température

## Eau potable

* Mettez plutôt un tiret du bas pour le nom de la variable « drinking-water-services », c’est ce qui est fait généralement et ce sera plus cohérent avec « Drinking\_water\_access\_by\_population\_type ». Essayez de garder une homogénéité, la même norme dans le nom de vos variables. **[Fanny] Ok. Marie a mis à jour le nom de la variable dans Knime**
* Jolie manipulation des données pour les remettre dans le bon sens pour tableau 😊
* On est sur de la population en K ? Il faut faire \*1000 pour avoir la vraie valeur ?

**[Fanny] La population est en % dans Eau Potable**

* **[Fanny] Modification de « Total » en « Both type »**

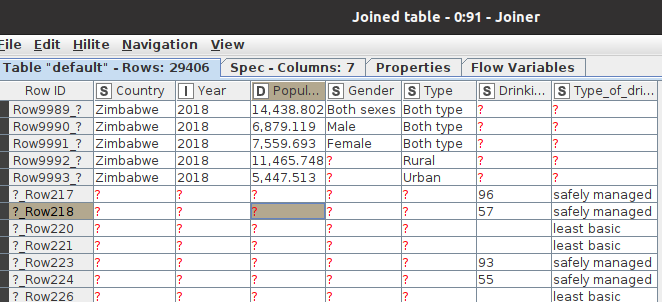
## Population

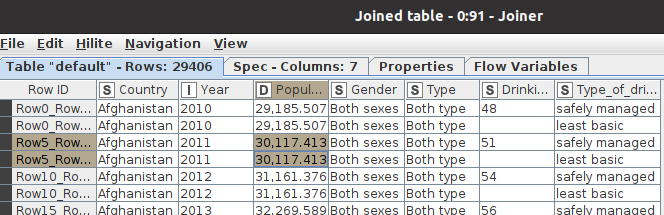
* Je n’ai pas l’impression que le pivot soit nécessaire ici, si ? Ce n’est pas grave en soit, mais en termes de performance, il vaut mieux renommer la variable (en plus vous pouvez mettre un nom plus parlant comme « population ») que de faire un pivot qui est plus compliqué pour le logiciel. **[Fanny] Effectivement, c’est pris en compte désormais**
* Jolies manipulations sur les strings !!! 😉
* Je ne sais pas si vous avez fait attention, mais il y a un écart entre le total urbain/rural vs hommes/femmes, par exemple quand je regarde l’Afghanistan en 2014 **[Fanny] Entends tu par là que pour Urban et Rural il n’y a pas le détail Female/Male ?**
* Je renommerais le « Total » en « Both type » pour plus de clarté ici **[Fanny] Effectivement, c’est pris en compte désormais**
* Votre variable « gender » n’est pas homogène (il me semble que c’est ce que Bilel avait signalé ?) **[Fanny] Effectivement, c’est pris en compte désormais**
* Idem c’est population \* 1000 ? **[Fanny] unité : 100K : population\*100 000**

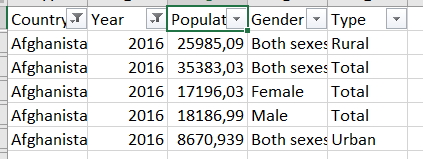
## Jointure eau potable / population

* Je vous ai simplifié l’étape : dans le joiner vous avez un onglet column selection où vous pouvez sélectionner les variables que vous voulez garder après la jointure 😉  
  **[Fanny] J’avais essayé cette technique mais il me semble que cela ne garde pas toutes les infos voir détail ici :**  
  [full outer join issue](https://forum.knime.com/t/joiner-full-outer-join-issue/15390) **+ voir image  
  Egalement, lea façon dont tu fais la jointure créé des doublons de valeur par exemple dans la colonne Population pour la premiere jointure. Ce qui, on l’a vu sur mes figures Tableau pose problème par la suite.**
* **Images**

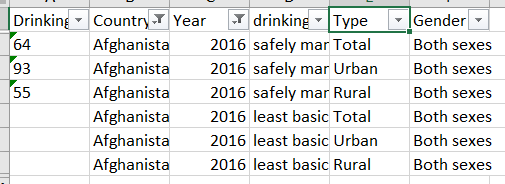
**Problème 1 : disparition d’info**

**Problème 2 : doublon**

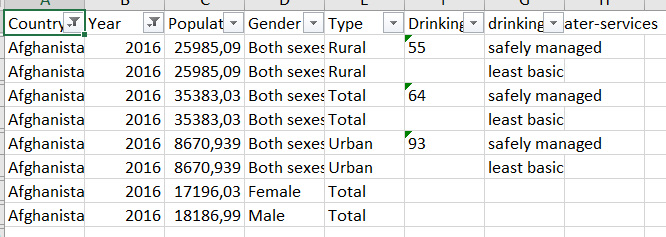
* 
* J’ai fait la modification pour chaque jointure
* J’ai changé le type en « Both sexes » dans les deux fichiers pour que ça fonctionne et que je puisse vérifier les jointures
* J’ai inversé le sens de la jointure : je trouve que ça a plus de sens de partir du fichier qui a le plus de modalités (genre et type)
* Jointure OK :
  + Population :



* + Eau potable :



* + Jointure :

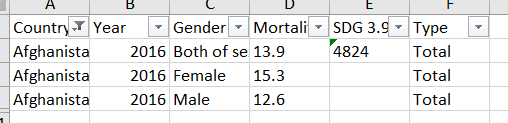


## Mortalité

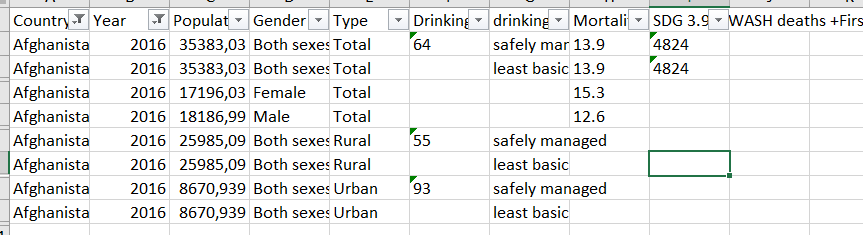
* De manière générale soyez vigilant sur les fautes de frappe, fautes de français
* Renommer « SDG 3.9.2 WASH deaths +First(Value) » avec un nom de variable plus parlant
* Pareil j’ai homogénéisé « Both of Sexe » en Both sexes

## Jointure 1 & mortalité

* Jointure ok :
  + Mortalité :



* + Jointure 2 :



Par contre un truc me chafouine, je vous conseille de mettre en colonne votre variable drinking… : une colonne least basic et une autre safely managed. Ce serait plus cohérent, vous allez déjà avoir pas mal de choses à garder en tête avec les genre / type pour éviter les doublons

## Température

* Qu’est ce que la variation de température ?
* Est-ce intéressant pour vous de garder la granularité au niveau du mois ? Comment est-ce que vous pensez exploiter ça ?
* Peut-être qu’il aurait été plus intéressant d’avoir la température moyenne sur l’année ? Et la température la plus élevée sur l’année ? Comme ça ça fait deux indicateurs et ça vous évite de dédoubler vos lignes pour quelque chose que vous n’allez pas exploiter

## Jointure 2 & température

* Je la coupe pour le moment

## Stabilité politique

* Il n’y a pas les données des 3 dernières variables après 2015, je vous laisse regarder d’où vient le souci, je pense que c’est avec les années 2015-2016 qui ne doivent pas ressortir. Je vous laisse regarder ça, ainsi que la jointure après.  
  **[Fanny] Modification faite → ajout des lignes 2016 et 2017**