|  |  |  |
| --- | --- | --- |
| **C:\Users\annes\OneDrive\Bureau\images.png** | **Extraction et traitement de données** |  |

## Récupérer et exploiter des fichiers CSV et JSON

**On souhaite exploiter des données démographiques liées aux Pays de la Loire.**

**Pour cela, nous allons dans un premier temps regarder cette vidéo :**

[**ActivitéTraitementDeDonnéesCsvViaUnTableur.avi**](ActivitéTraitementDeDonnéesCsvViaUnTableur.avi)

**Puis rendez-vous sur le site suivant :**

<https://data.paysdelaloire.fr/explore/dataset/234400034_population-des-pays-de-la-loire-2005-2017/table/>

1. **Fichier CSV**
2. Ouvrez le fichier CSV à l’aide d’un éditeur de texte simple (notepad++).
3. Quel est le séparateur de champs ? ……………
4. Listez les descripteurs.

……………………………………………………………………………………………………………

……………………………………………………………………………………………………………

……………………………………………………………………………………………………………

1. Retrouvez le total de la population de la Sarthe pour l'année 2015.

……………………………………………………………………………………………………………

1. Vérifiez que ce total correspond à la somme des populations des tranches d'âge de la Sarthe pour cette année-là.

……………………………………………………………………………………………………………

1. **Fichier JSON**
2. Ouvrez le fichier JSON à l’aide du navigateur Firefox.
3. Listez les descripteurs.

……………………………………………………………………………………………………………

……………………………………………………………………………………………………………

……………………………………………………………………………………………………………

1. Retrouvez le total de la population de la Sarthe pour l'année 2015.

……………………………………………………………………………………………………………

1. Est-ce le même que pour le fichier CSV ? ………
2. Cliquez sur l'onglet "Données brutes" afin de voir le contenu "réel" du fichier.

Auriez-vous trouvé aussi facilement les informations demandées précédemment si la présentation du fichier avait été celle-ci ? ……….

1. **Comparaison**
2. Auriez-vous pu exploiter les données du fichier CSV si la première ligne avait été manquante ?

………….

1. Même question pour le fichier JSON. …………………..
2. Qu'est-ce qui justifie que le fichier JSON soit beaucoup plus volumineux que le fichier CSV ?

……………………………………………………………………………………………………………

## Générer des graphiques à partir de données provenant d’un fichier CSV

1. **Importer un fichier CSV dans un tableur**

On souhaite exploiter des données démographiques liées aux Pays de la Loire.

Pour cela, on utilise le fichier 234400034\_population-des-pays-de-la-loire-2005-2017.csv.

Faire clic-droit sur le fichier et choisir "Ouvrir avec LibreOffice Calc".

L'écran suivant devrait apparaître:



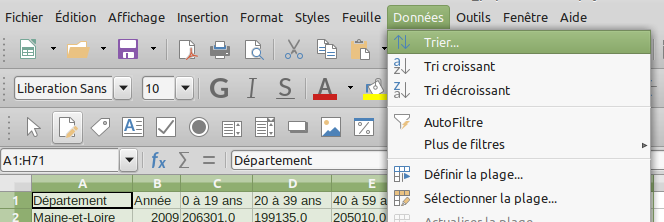
Vérifiez bien dans les options de séparateur que le point-virgule est coché, puis cliquez sur "OK".

1. **Trier des données**

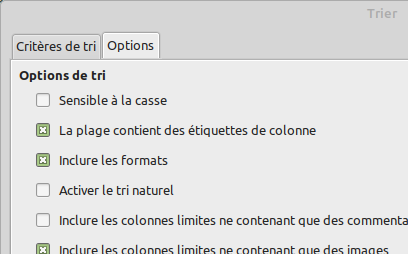
Les données importées ne sont pas du tout triées, et sont donc difficilement exploitables.

Pour faire un traitement par département, il faudrait regrouper les données selon ce critère.

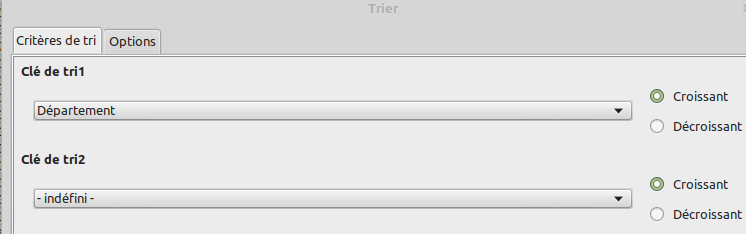
Pour cela, allez dans le menu "Données" et choisissez "Trier".



Dans l'onglet "Options", vérifiez que l'option "La plage contient des étiquettes de colonne" est cochée.



Allez dans l'onglet "Critère de tri" et choisissez "Département" comme clé de tri 1, et croissant.



Les données sont à présent regroupées par département, dans l'ordre croissant de noms de département.

Pour exploiter de façon efficace, il faudrait également que les données soient triées par date.

Trier les données par département ordre croissant (tri1) ET année(tri2) ordre croissant.

1. **Données numériques**

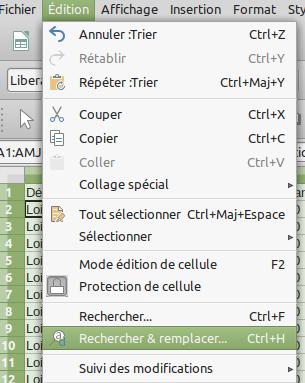
Un séparateur décimal est un symbole utilisé pour partager la partie décimale de la partie entière d'un nombre décimal. Ce symbole dépend des conventions régionales du système de numération.

Communément, il est représenté par un point dans les systèmes anglo-saxons et par une virgule dans les autres systèmes.

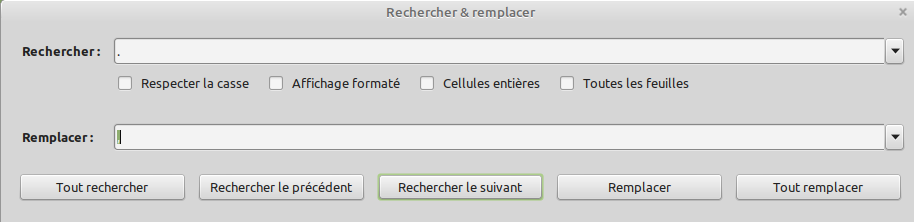
Les données importées via le fichier CSV sont au format anglo-saxon.

Il faut donc remplacer les points par des virgules afin que les données numériques soient interprétées comme telles.

Pour cela sélectionnez l'ensemble des données, allez dans le menu "Édition" et sélectionnez "Rechercher & remplacer".



Dans la zone "Rechercher", mettez un point, et dans la zone "Remplacer", mettez une virgule, puis cliquez sur "Tout remplacer".



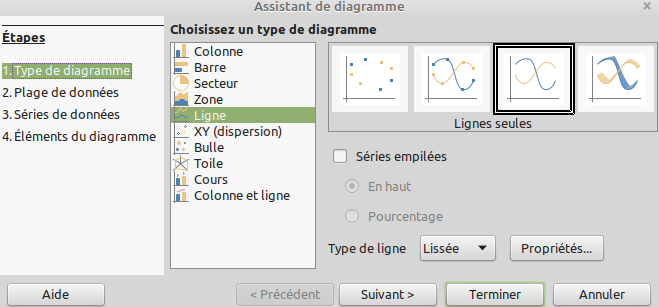
Les données numériques sont à présent exploitables.

1. **Génération de diagramme**

Nous voulons voir l'évolution de la population par tranche d'âge pour le département de la Loire atlantique.

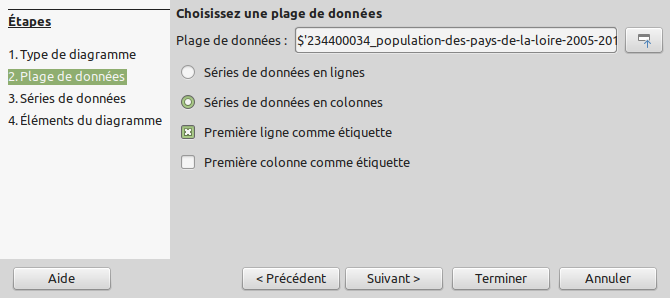
Sélectionner la première ligne ainsi que les lignes correspondant au département de la Loire atlantique, puis cliquez sur l'icône "insérer un diagramme" (ou menu "insertion" et "Diagramme")

Choisir diagramme en ligne, lignes seules et type de ligne lissée.

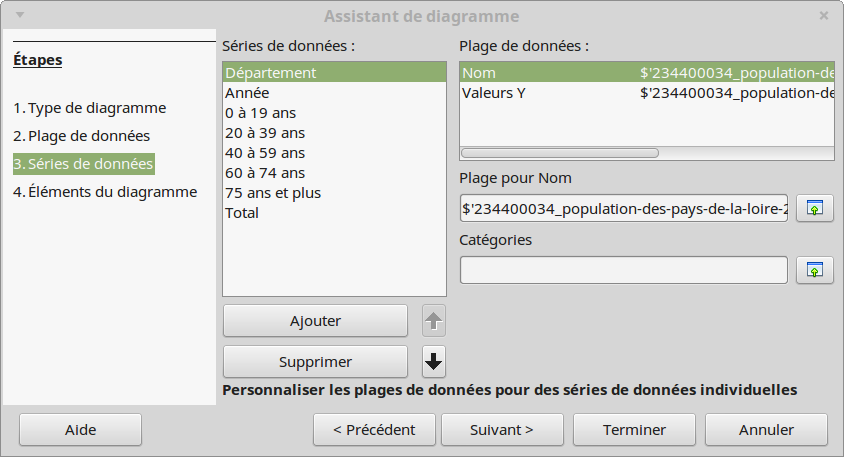


Cliquez sur "Suivant".

Sélectionnez "Série de données en colonnes" et "Première ligne comme étiquette"



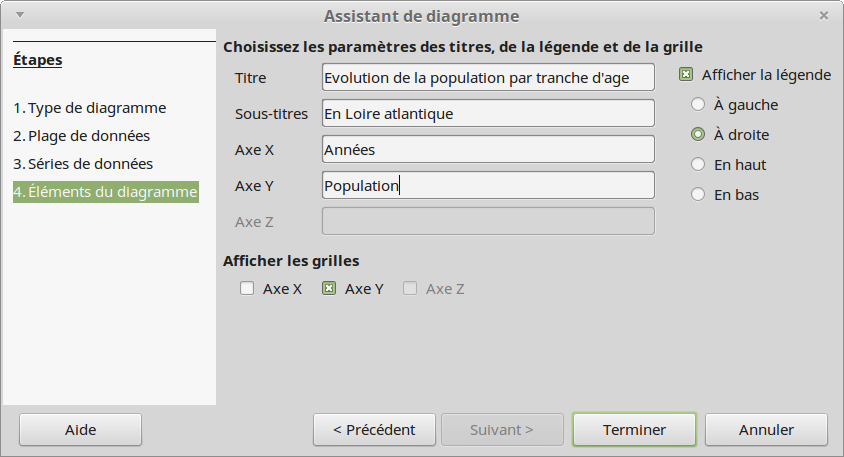
Cliquez sur "Suivant"



Supprimer Département, Année et Total

Cliquez sur l'icône C:\Users\annes\AppData\Local\Temp\lu1218833i33.tmp\lu1218833i5u_tmp_b87b5d8e81875208.pngde Catégories, puis sélectionnez les années de 2005 à 2018 pour le département de la Loire atlantique.

Cliquez sur "Suivant".



Remplissez les différents champs comme indiqué ci-dessus et cliquez sur "Terminer".

Le diagramme ci-dessous est généré.



Quelles tendances pouvez-vous déduire quant à la population selon les tranches d'âge pour le département de la Loire atlantique ?

Générez le même type de diagramme pour le Maine-et-Loire, la Sarthe, la Mayenne et la Vendée.

Peut-on généraliser les tendances de la Loire atlantique à l'ensemble des départements des Pays de la Loire ?

…………………………..