Les données structurées au format CSV:

On trouve énormément de données sur internet. Une partie de ces données sont publiques, par exemple le site data.gouv.fr récence un grand nombre de données publiques. Ces données sont librement réutilisables.

Exercice 1: Afin de découvrir ce qu'est "l'open data", allez sur le site data.gouv.fr. En haut et à gauche de la page d'acceuil, cliquez sur "Découvrez L'OpenData". Résumez en quelques lignes ce que vous aurez appris en lisant cette page.

Exercice 2: Explorez pendant quelques minutes le site data.gouv.fr. Recherchez les données "Opérations coordonnées par les CROSS" à l'aide du moteur de recherche proposé par le site. Vous pouvez constater que ces données sont au format CSV. Le format CSV est très courant sur internet. Voici ce que nous dit Wikipédia sur le format CSV :

Comma-separated values, connu sous le sigle CSV, est un format informatique ouvert représentant des données tabulaires sous forme de valeurs séparées par des virgules.

Un fichier CSV est un fichier texte, par opposition aux formats dits binaires. Chaque ligne du texte correspond à une ligne du tableau et les virgules correspondent aux séparations entre les colonnes. Les portions de texte séparées par une virgule correspondent ainsi aux contenus des cellules du tableau.

Voici un exemple du contenu d'un fichier CSV :

```
nom,prenom,date_naissance
Durand,Jean-Pierre,23/05/1985
Dupont,Christophe,15/12/1967
Terta,Henry,12/06/1978
```

Je pense qu'il est évident pour vous que nous avons ici 3 personnes :

- Jean-Pierre Durand qui est né le 23/05/1985
- Christophe Dupont qui est né le 15/12/1967
- Henry Terta qui est né le 12/06/1978

"nom", "prenom" et "date_naissance" sont appelés des descripteurs alors que, par exemple, "Durand", "Dupont" et "Terta" sont les valeurs du descripteur "nom".

Exercice 3: Donnez les différentes valeurs du descripteur "date_naissance"

ATTENTION: La virgule est un standard pour les données anglo-saxonnes, mais pas pour les données aux normes françaises. En effet, en français, la virgule est le séparateur des chiffres décimaux.

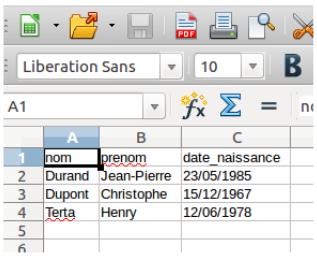
Il serait impossible de différencier les virgules des décimaux et les virgules de séparation des informations. C'est pourquoi on utilise un autre séparateur : le point-virgule (;). Dans certains cas cela peut engendrer quelques problèmes, vous devrez donc rester vigilants sur le type de séparateur utilisé.

Les tableurs, tels que "Calc" (Libre Office), sont normalement capables de lire les fichiers au format CSV. J'ai précisé "normalement" car certains tableurs gèrent mal le séparateur CSV "point-virgule" et le séparateur des chiffres décimaux "virgule".

Exercice 4: Après avoir téléchargé le fichier ident_pointVirgule.csv, ouvrez ce dernier à l'aide d'un tableur. Si par hasard votre tableur ne gère pas correctement le fichier avec le séparateur "point-virgule", voici une version "séparateur virgule" du fichier : ident_virgule.csv

Dans la suite, gardez toujours cet éventuel problème à l'esprit (surtout avec des données "made in France")

Vous devriez obtenir ceci :



Vous pouvez constater que les données sont bien "rangées" dans un tableau avec des lignes et des colonnes (voilà pourquoi on parle de données tabulaires.

Il est possible de trouver sur le web des données beaucoup plus intéressantes à traiter que celles contenues dans le fichier "ident_pointVirgule.csv" (ou "ident_virgule.csv"). Par exemple, le site sql.sh, propose un fichier csv contenant des informations sur l'ensemble des communes françaises.

Exercice 5: Ouvrez le fichier ville_point_virgule.csv à l'aide d'un tableur (c'est une version légèrement modifiée de celle disponible sur le site sql.sh, j'y ai notamment ajouté des entêtes). En cas de problème avec votre tableur, voici une version "séparateur virgule" : ville_virgule.csv (attention le séparateur "décimal" est ici le point) Comme vous pouvez le constater, nous avons 12 colonnes (et 36700 lignes si on ne compte pas l'entête !), voici la signification de ces colonnes :

- dep : numéro de département
- nom : nom de la commune
- cp : code postal
- nb_hab_2010 : nombre d'habitants en 2010
- nb_hab_1999 : nombre d'habitants en 1999
- nb_hab_2012 : nombre d'habitants en 2012 (approximatif)
- dens : densité de la population (habitants par kilomètre carré)

- surf : superficie de la commune en kilomètre carré
- long : longitude
- lat : latitude
- alt_min : altitude minimale de la commune (il manque des données pour certains territoires d'outre-mer)
- alt_max : altitude maximale de la commune (il manque des données pour certains territoires d'outre-mer)

Exercice 6: En vous aidant du fichier ouvert dans le "À faire vous-même 4", déterminez l'altitude maximale et l'altitude minimale de votre commune.