

## 9.2

### Les fichiers csv

NSI 1ÈRE - JB DUTHOIT

#### 9.2.1 Les fichiers .csv

L'organisation des données en tableau est très courante et très ancienne.

Le bulletin d'un élève, les listes de classe par exemple sont organisés en table. Un fichier **csv** (pour comma separated values, soit en français valeurs séparées par des virgules) est un format très pratique pour représenter des données structurées.

1. C'est un fichier texte

- Chaque ligne du texte correspond à une ligne du tableau et les virgules correspondent aux séparations entre les colonnes.
- Toutes les lignes du fichier ont le même nombre de champs (colonnes).

#### Exercice 9.163

Créer un fichier `NotesEleves.txt` avec le bloc-notes du type :

```
*Sans titre - Bloc-notes
Fichier Edition Format Affichage
Nom,maths,nsi,anglais
Langlet,12,20,14
Steppe,15,17,19
Klein,15,11,19
```

#### Exercice 9.164

Visiter le site <https://www.data.gouv.fr/>, puis trouver le fichier qui donne la liste des pays du monde et leur capitale respective. Télécharger-le sur votre ordinateur. Ouvrir votre fichier avec le bloc-notes.

```
Fichier Edition Format Affichage Aide
NOM;NOM_ALPHA;CODE;ARTICLE;NOM_LONG;CAPITALE
Afghanistan;Afghanistan;AFG;l';République islamique d'Afghanistan;Kaboul
Afrique du Sud;Afrique du Sud;ZAF;l';République d'Afrique du Sud;Prétoria
Albanie;Albanie;ALB;l';République d'Albanie;Tirana
Algérie;Algérie;DZA;l';République algérienne démocratique et populaire;Alger
Allemagne;Allemagne;DEU;l';République fédérale d'Allemagne;Berlin
Andorre;Andorre;AND;l';Principauté d'Andorre;Andorre-la-Vieille
Angola;Angola;AGO;l';République d'Angola;Luanda
Antigua-et-Barbuda;Antigua-et-Barbuda;ATG;;Antigua-et-Barbuda;Saint John's
Arabie saoudite;Arabie saoudite;SAU;l';Royaume d'Arabie saoudite;Riyad
Argentine:Argentine:ARG:l':République argentine:Buenos Aires
```

→ Ici, c'est le "point virgule" qui a été utilisé comme séparateur.



## 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.

"NOM", "NOM\_ALPHA", "CODE", "ARTICLE" ... sont appelés des **descripteurs** alors que, par exemple, "Afghanistant", "Afghanistan" et "AFG", "1" sont les **valeurs du descripteur** "NOM".

### 9.2.2 Importer des données depuis un fichier csv avec Python

- Méthode 1 :

```
import csv
pays = []
with open("pays.csv","r") as fichier:
    table = csv.reader(fichier, delimiter = ';')
    for ligne in table:
        pays.append(ligne)
```

On a ainsi une liste de listes

- Quel est le nombre de pays présents dans cette liste ?
  - Afficher le premier élément de la liste pays.
  - Et afficher le 110 ème

```
>>> pays  
[['countryCode', 'countryName', 'area', 'population', 'continent', 'currencyCode', 'currentPopulation', 'capital'], ['AD', 'Andorra', '468', '84000', 'EU', 'EUR', 'Euro', 'Andorra la Vella'], ['AE', 'United Arab Emirates', '82880', '4975593', 'AS', 'AED', 'Dirham', 'Abu Dhabi'], ['AF', 'Afghanistan', '647500', '29121286', 'AS', 'AFN', 'Afghani', 'Kabul'], ['AG', 'Antigua and Barbuda', '443', '81000', 'NA', 'XCD', 'Dollar', 'St. John's'], ['AI', 'Anguilla', '166', '14000', 'NA', 'XCD', 'Dollar', 'The Valley'], ['AL', 'Albania', '28742', '3000000', 'EU', 'ALL', 'Lek', 'Tirana'], ['AM', 'Armenia', '29743', '3000000', 'EU', 'AMD', 'Dram', 'Yerevan'], ['AO', 'Angola', '1246700', '20000000', 'AF', 'AOA', 'Kwanza', 'Luanda'], ['AQ', 'Antarctica', '13400000', '0', 'EU', 'AOA', 'Kwanza', 'Punta Arenas'], ['AR', 'Argentina', '2780400', '40000000', 'SA', 'ARS', 'Peso', 'Buenos Aires'], ['AT', 'Austria', '83855', '8000000', 'EU', 'EUR', 'Euro', 'Vienna'], ['AU', 'Australia', '7686850', '22000000', 'OC', 'AUD', 'Dollar', 'Canberra'], ['AZ', 'Azerbaijan', '85200', '9000000', 'EU', 'AZN', 'Manat', 'Bakı'], ['BA', 'Bosnia and Herzegovina', '51129', '3000000', 'EU', 'BAM', 'Convertible Mark', 'Sarajevo'], ['BH', 'Bahrain', '7343', '1000000', 'EU', 'BHD', 'Dinar', 'Manama'], ['BD', 'Bangladesh', '1440000', '14000000', 'SA', 'BDT', 'Taka', 'Dhaka'], ['BF', 'Burkina Faso', '2742000', '16000000', 'AF', 'XOF', 'CFA Franc BEAC', 'Ouagadougou'], ['BG', 'Bulgaria', '110993', '7000000', 'EU', 'BGN', 'Leva', 'Sofia'], ['BH', 'Bahrain', '7343', '1000000', 'EU', 'BHD', 'Dinar', 'Manama'], ['BI', 'Burundi', '28246', '1000000', 'AF', 'BIF', 'Franc', 'Bujumbura'], ['BT', 'Bhutan', '38394', '700000', 'AS', 'BTN', 'Ngultrum', 'Thimphu'], ['BW', 'Botswana', '2227000', '2000000', 'AF', 'BWP', 'Pula', 'Gaborone'], ['BY', 'Belarus', '2000000', '1000000', 'EU', 'BYN', 'Belarussian Ruble', 'Minsk'], ['BZ', 'Belize', '22966', '300000', 'NA', 'BZD', 'Belize Dollar', 'Belmopan'], ['CA', 'Canada', '9904700', '35000000', 'NA', 'CAD', 'Dollar', 'Ottawa'], ['CF', 'Central African Republic', '622940', '4000000', 'AF', 'XAF', 'CFA Franc CEMAC', 'Bangui'], ['CG', 'Congo', '342000', '3000000', 'AF', 'XAF', 'CFA Franc CEMAC', 'Kinshasa'], ['CI', 'Ivory Coast', '321240', '2000000', 'AF', 'XAF', 'CFA Franc CEMAC', 'Yamoussoukro'], ['CR', 'Costa Rica', '51020', '4000000', 'NA', 'CRC', 'Colón', 'San José'], ['CY', 'Cyprus', '9245', '1000000', 'EU', 'EUR', 'Euro', 'Nicosia'], ['CZ', 'Czech Republic', '77660', '10000000', 'EU', 'CZK', 'Crown', 'Prague'], ['DE', 'Germany', '357020', '80000000', 'EU', 'EUR', 'Euro', 'Berlin'], ['DK', 'Djibouti', '22966', '500000', 'AF', 'DJF', 'Djiboutian Franc', 'Djibouti'], ['DO', 'Dominican Republic', '48440', '8000000', 'NA', 'DOP', 'Peso Dominicano', 'Santo Domingo'], ['DZ', 'Algeria', '2381700', '35000000', 'AF', 'DZD', 'Dinar', 'Algiers'], ['EC', 'Ecuador', '252470', '15000000', 'SA', 'ECU', 'Sucre', 'Quito'], ['EE', 'Estonia', '14313', '1000000', 'EU', 'EUR', 'Euro', 'Tallinn'], ['EG', 'Egypt', '1001440', '80000000', 'AF', 'EGP', 'Pound', 'Cairo'], ['ES', 'Spain', '499900', '40000000', 'EU', 'EUR', 'Euro', 'Madrid'], ['FI', 'Finland', '134120', '5000000', 'EU', 'EUR', 'Euro', 'Helsinki'], ['FO', 'Faroe Islands', '1393', '50000', 'EU', 'ISK', 'Krona', 'Tórshavn'], ['FR', 'France', '643800', '60000000', 'EU', 'EUR', 'Euro', 'Paris'], ['GE', 'Georgia', '69700', '5000000', 'EU', 'GEL', 'Lari', 'Tbilisi'], ['GH', 'Ghana', '238530', '20000000', 'AF', 'GHS', 'Cedi', 'Accra'], ['GR', 'Greece', '131940', '10000000', 'EU', 'EUR', 'Euro', 'Athens'], ['GT', 'Guatemala', '108890', '15000000', 'NA', 'GTQ', 'Quetzal', 'Guatemala City'], ['HN', 'Honduras', '112080', '5000000', 'NA', 'HNL', 'Lempira', 'Tegucigalpa'], ['ID', 'Indonesia', '1904500', '20000000', 'AF', 'IDR', 'Rupiah', 'Jakarta'], ['IE', 'Ireland', '321300', '5000000', 'EU', 'EUR', 'Euro', 'Dublin'], ['IL', 'Israel', '205000', '5000000', 'AS', 'ILS', 'Shekel', 'Tel Aviv'], ['IQ', 'Iraq', '460000', '3000000', 'AF', 'IQD', 'Dinar', 'Baghdad'], ['IS', 'Iceland', '103000', '1000000', 'EU', 'ISK', 'Krona', 'Reykjavík'], ['IT', 'Italy', '301330', '60000000', 'EU', 'EUR', 'Euro', 'Rome'], ['JO', 'Jordan', '89240', '2000000', 'AF', 'JOD', 'Dinar', 'Amman'], ['JP', 'Japan', '3779700', '120000000', 'AS', 'JPY', 'Yen', 'Tokyo'], ['LR', 'Liberia', '430000', '1000000', 'AF', 'LRD', 'Dollar', 'Monrovia'], ['LT', 'Lithuania', '25210', '1000000', 'EU', 'EUR', 'Euro', 'Vilnius'], ['LU', 'Luxembourg', '25800', '1000000', 'EU', 'EUR', 'Euro', 'Luxembourg'], ['LY', 'Libya', '1750000', '1000000', 'AF', 'LYD', 'Dinar', 'Tripoli'], ['MD', 'Moldova', '33840', '1000000', 'EU', 'MDL', 'Leu', 'Chisinau'], ['ME', 'Montenegro', '14120', '1000000', 'EU', 'EUR', 'Euro', 'Podgorica'], ['MK', 'Macedonia', '250000', '1000000', 'EU', 'MKD', 'Denar', 'Skopje'], ['ML', 'Mali', '1240000', '1000000', 'AF', 'XOF', 'CFA Franc BEAC', 'Kidal'], ['MT', 'Malta', '31400', '1000000', 'EU', 'EUR', 'Euro', 'Valletta'], ['MU', 'Mauritius', '204000', '1000000', 'AF', 'MUR', 'Rupee', 'Port Louis'], ['MV', 'Maldives', '29800', '1000000', 'AF', 'MVR', 'Rufiyaa', 'Male'], ['MZ', 'Mozambique', '802000', '1000000', 'AF', 'MZN', 'Metical', 'Maputo'], ['NP', 'Nepal', '265000', '1000000', 'AF', 'NPR', 'Rupee', 'Kathmandu'], ['NL', 'Netherlands', '166500', '10000000', 'EU', 'EUR', 'Euro', 'The Hague'], ['NO', 'Norway', '242200', '5000000', 'EU', 'NOK', 'Krone', 'Oslo'], ['NZ', 'New Zealand', '450000', '1000000', 'AF', 'NZD', 'Dollar', 'Wellington'], ['PE', 'Peru', '280000', '1000000', 'AF', 'PEN', 'Sol', 'Lima'], ['PS', 'Palestinian Territories', '460000', '1000000', 'AF', 'XJD', 'Jordanian Dinar', 'Ramallah'], ['PT', 'Portugal', '175200', '10000000', 'EU', 'EUR', 'Euro', 'Lisbon'], ['RO', 'Romania', '233300', '10000000', 'EU', 'RON', 'Leu', 'Bucharest'], ['RS', 'Serbia', '1000000', '1000000', 'EU', 'EUR', 'Euro', 'Belgrade'], ['SI', 'Slovenia', '20200', '1000000', 'EU', 'EUR', 'Euro', 'Ljubljana'], ['SK', 'Slovakia', '51000', '1000000', 'EU', 'EUR', 'Euro', 'Bratislava'], ['SV', 'El Salvador', '1000000', '1000000', 'NA', 'USD', 'US Dollar', 'San Salvador'], ['TR', 'Turkey', '750000', '10000000', 'EU', 'TRY', 'Turkish Lira', 'Ankara'], ['VA', 'Vatican City', '1000000', '1000000', 'EU', 'EUR', 'Euro', 'Vatican City'], ['YE', 'Yemen', '2140000', '1000000', 'AF', 'YER', 'Rial', 'Sana'a']]
```

Cette solution présente des inconvénients

- La 1ère ligne, celle des attributs, a été chargée comme une ligne de données.
  - Toutes les données ont été converties en chaînes de caractères même les entiers.
  - Plus gênant, le lien entre les valeurs du tableau pays[1] et le nom des enregistrements, contenus dans pays[0], n'est pas direct.
  - Nous allons donc utiliser une seconde méthode !

- #### • Méthode 2 :

```

1 import csv
2 pays = []
3 with open("pays.csv","r") as fichier:
4     table = csv.DictReader(fichier, delimiter = ';')
5     for ligne in table:
6         pays.append(ligne)

```

```

>>> pays
[OrderedDict([('countryCode', 'AD'), ('countryName', 'Andorra'), ('area', '468'), ('population', '84000'), ('continent', 'EU'), ('currencyCode', 'EUR'), ('currencyName', 'Euro'), ('ital', 'Andorra la Vella')]), OrderedDict([('countryCode', 'AE'), ('countryName', 'United ab Emirates'), ('area', '82880'), ('population', '4975593'), ('continent', 'AS'), ('curre Code', 'AED'), ('currencyName', 'Dirham'), ('capital', 'Abu Dhabi')]), OrderedDict([('couyCode', 'AF'), ('countryName', 'Afghanistan'), ('area', '647500'), ('population', '291212'), ('continent', 'AS'), ('currencyCode', 'AFN'), ('currencyName', 'Afghani'), ('capital', 'abul')]), OrderedDict([('countryCode', 'AG'), ('countryName', 'Antigua and Barbuda'), ('a

```

Examinons les avantages de cette solution :

- La 1ère ligne, celle des attributs, n'a pas été chargée comme une ligne de données.
- Le lien entre les valeurs du tableau pays[1] et le nom des enregistrements, contenus dans pays[0], est direct.
- Quel est la longueur de la liste ? Est-ce en accord avec le nombre de pays ? Expliquer
- .

Mais :

- toutes les données restent converties en chaînes de caractères même les entiers.
- Le format d'affichage peut déconcerter

☛ Nous allons donc utiliser une autre amélioration !

- Méthode 3 :

```

import csv
pays = []
with open("pays.csv","r") as fichier:
    table = csv.DictReader(fichier, delimiter = ';')
    for ligne in table:
        pays.append(dict(ligne))

```

```
>>> pays
[{'countryCode': 'AD', 'countryName': 'Andorra', 'area': '468', 'population': '84000', 'continent': 'EU', 'currencyCode': 'EUR', 'currencyName': 'Euro', 'capital': 'Andorra la Vella'}, {'countryCode': 'AE', 'countryName': 'United Arab Emirates', 'area': '82880', 'population': '4975593', 'continent': 'AS', 'currencyCode': 'AED', 'currencyName': 'Dirham', 'capital': 'Abu Dhabi'}, {"countryCode": "AF", "countryName": "Afghanistan", "area": "647500", "population": "29121286", "continent": "AS", "currencyCode": "AFN", "currencyName": "Afghani", "capital": "Kabul"}, {"countryCode": "AG", "countryName": "Antigua and Barbuda", "area": "443", "population": "86754", "continent": "NA", "currencyCode": "XCD", "currencyName": "Dollar", "capital": "St. John's"}, {"countryCode": "AI", "countryName": "Anguilla", "area": "102", "population": "102", "continent": "NA", "currencyCode": "XCD", "currencyName": "Dollar", "capital": "The Valley"}]
```

Une liste de dico, c'est beaucoup plus lisible!

### Exercice 9.165

Ajouter quelques lignes au code précédent afin de convertir les str en entiers ou en flottants