

Statistiques Descriptives

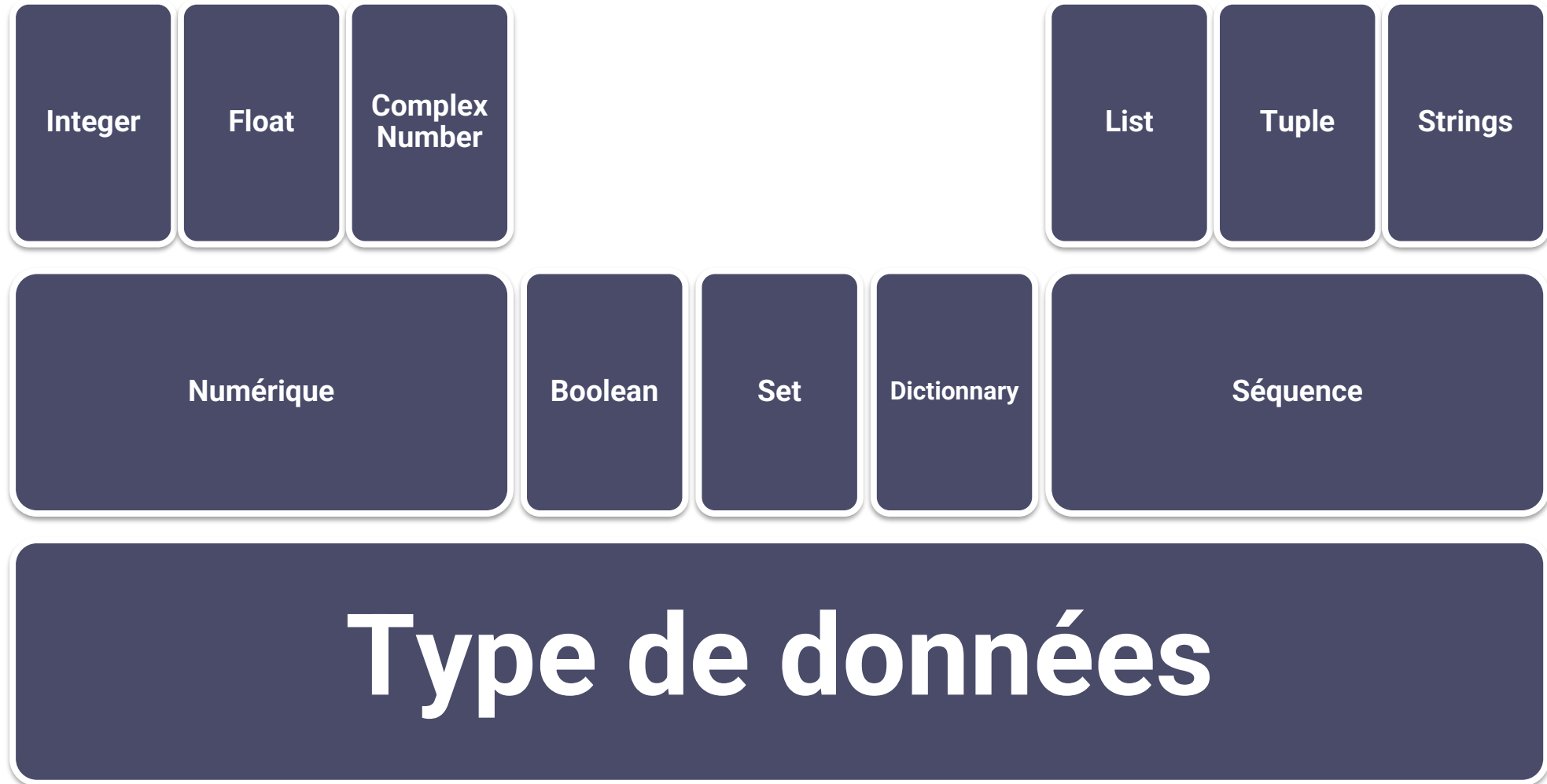
Abdellah Madane
madane@lipn.univ-paris13.fr

2022/2023

Quick facts

- Langage généraliste.
- Milliers de librairies (appelées modules) disponibles pour à peu près n'importe quelle application
- Langage le plus utilisé pour l'analyse de données et l'apprentissage (aux côté de R, Matlab, Scala...)
- Itératif, orienté objet
- IPortable/multi-plateforme
- Fortement typé
- Dynamiquement typé

Types de données



Principaux Opérateurs

Arithmétiques

+

-

*

**

/

//

%

Relationnels

==

!=

<

<=

>

>=

Booléens

and

or

not

Structures conditionnelles

SI	Si Sinon	Si Sinon si Sinon
if condition : instuction 1 instuction 2	if condition : instuction 1 instuction 2 else: instuction 1	if condition : instuction 1 instuction 2 elif condition : instuction 1 else : instuction 1

Les boucles

Les boucles sont des instructions qui permettent de répéter un bloc d'instructions plusieurs fois.

- ❑ **Boucle déterministe « for »** : répétition en précisant la valeur initiale et la valeur finale d'incrément.
- ❑ **Boucle non déterministe « while »** : Définition d'une condition d'arrêt pour interrompre la boucle.

Boucle « for »	Boucle « while »
<pre>for x in « suite de valeur » : instuction 1 instuction 2 . .</pre>	<pre>while condition : instuction 1 instuction 2 . .</pre>

La fonction «range» est une suite finie permettant de répéter un élément.

- ❑ `range(a,b) = [a,b[` avec $(a < b)$
- ❑ `range(1,5) = 1, 2, 3, 4` - pas = 1
- ❑ `range(1, 9, 2) = 1, 3, 5, 7` - pas = 2

Les fonctions

- Une fonction est une tâche partielle du programme globale qui peut être subdivisée en petites unités.
- Une fonction est réutilisable. Elle permet donc un gain de temps et sa réutilisation dans différents endroits du programme.
- Une fonction doit être définie avant son appel.
- Les fonctions qui n'utilisent pas "return" sont des fonctions qui n'ont pas de valeur. Elles sont appelées aussi "procédure". Elles ne peuvent pas être utilisées dans une formule ou comme paramètre d'une autre fonction.

Structure d'une fonction

```
def NomFonction ( Les paramètres ) :  
    Instruction 1  
    Instruction 2  
    .  
    .  
    .
```

Les structures de données

Liste

- Suite d'éléments qui peuvent être de types différents représentés par une seule variable. Chaque élément a un indice compris entre 0 et le nombre d'éléments -1
- 0 est l'indice du premier élément de la liste.
- -1 est l'indice du dernier élément de la liste.
- La fonction « len » permet de connaître la longueur d'une liste.

$L = [22, "A", 0.7], \text{ len}(L) = 3$

- Des méthodes à savoir : `L.copy()` , `L.append(x)` , `L.index(x)`, `L.insert(i, x)`, `L.pop(i)`, `L.remove(x)`

Les structures de données

Tuple

- Suite ordonnée de valeurs.
- Structure de données non modifiable que ce soit les valeurs ou la taille.
- On peut dire qu'un tuple est une liste non modifiable.
- `P = tuple(range(1,5)) -> p = (1, 2, 3, 4) , P[0] = 1`
- `L = list(range(1,5)) -> L = [1, 2, 3, 4]`

Les structures de données

Ensemble

- Structure de données où l'ordre des éléments n'est pas important et où il n'y a pas de redondances.
- set ou {}
- $E = \text{set}((1,2,3,2,5,3)) \rightarrow E = \{1,2,3,5\}$
- Absence d'indexation.
- Parcours : for x in E
- Méthodes à savoir : E.add(x) , E.remove() , E.discard() , E.clear, E.isdisjoint(F)
- Opérateurs entre sets : & (intersection) , | (union), - (exclusion), ^ (Différence symétrique), <, > .

Les structures de données

Dictionnaire

- Un dictionnaire est une collection modifiable de couples (clé non modifiable, valeur modifiable) permettant un accès à la valeur si on fournit la clé.
- On peut le voir comme une liste dans laquelle l'accès à un élément se fait par un code au lieu d'un indice.
- La clé est une chaîne de caractère mais la valeur est de n'importe quel type.
- $D = \{ \text{'clé1':valeur1, ... , 'clé N' : valeur N} \}$
- Les méthodes à savoir : `D.keys()` , `D.items()` , `D.values()`

Les matrices

- Une liste à deux dimensions est une liste où chaque élément est une liste à une dimension. Si toutes les sous-listes ont la même longueur, on parle d'une matrice.
- $M = [[1,2,3], [4,5,6]]$ est une matrice de dimension 2×3
- $M[0] = [1,2,3]$
- $M[0][0] = 1$

Les modules/bibliothèques

Numpy

- Calcul sur des objets multidimensionnels : vecteurs, matrices, ...
- NumPy peut être utilisé pour créer un array (ndarrays) et manipuler les données à l'aide de plusieurs fonctions mathématiques.
- Contrairement aux listes Python, les éléments d'un array sont homogènes.
- Utilisation similaire à Matlab
- Utilise des libraires en C compilées d'algèbre linéaire performant
- Compatible avec de très nombreuses libraires de data science/machine learning/deep learning : pandas, scikit-learn, Tensorflow, MXNet, etc.

Les modules/bibliothèques

Pandas

- La bibliothèque pandas est utile pour traiter les données structurées comme les données stockées dans des tableaux, comme les fichiers CSV, les feuilles de calcul Excel ou les tableaux de bases de données.
- Les données sont traitées sous forme d'une abstraction appelée DataFrame : données organisées par colonnes nommées (table relationnelle).
- Une donnée sous forme de colonne ou ligne (vecteur) a pour type 'Series'.
- <https://pandas.pydata.org/pandas-docs/stable/reference/frame.html>
- <https://pandas.pydata.org/pandas-docs/stable/reference/series.html>

Les modules/bibliothèques

Pandas

Columns (Axis 1)

←

→

	t	EGT_SEL	FLIGHT_MOD	FMV_SEL	HPTC_SEL	LPTC_SEL	N1_SEL	N2_ACTSEL	OIL_P	OIL_TEMP	...	T25_SEL	T3_SEL	VBV_SEL	VIB_CN1	VIB_CN2	VIB_TN1	VIB_
1	15/09/2011 14:25:58.125	NaN	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	15/09/2011 14:25:58.375	NaN	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	15/09/2011 14:25:58.625	NaN	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	15/09/2011 14:25:58.875	NaN	NaN	NaN	NaN	NaN	NaN	0	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN
5	15/09/2011 14:25:59.125	NaN	NaN	NaN	NaN	NaN	0	0	647.998	NaN	...	NaN	NaN	0	0	0	0	0

Index

5 rows × 26 columns

Valeurs manquantes

Données

Rows (Axis 0)