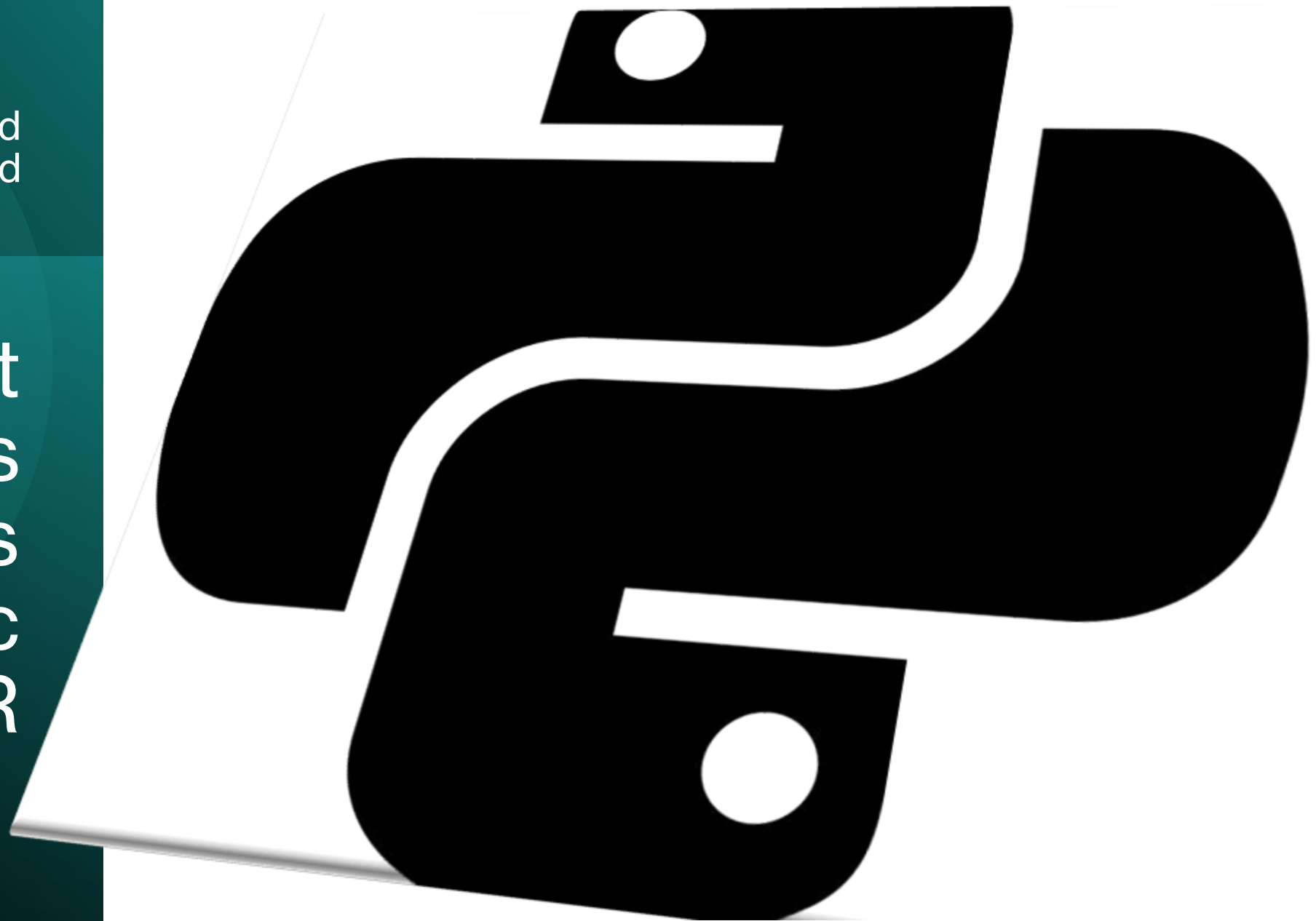
CHAIB Mohamed
Belaid

Export et Import des Données avec Python et R



Exporting and importing data with python and R

Etudiant :

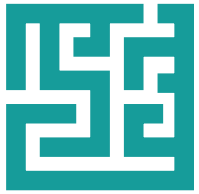
CHAIB Mohamed Belaid

Introduction

When working with data analysis or probability models, you may need to export the results to a file for future use or sharing with others.

Import/Export File Formats

Note that usage files vary according to the size, structure, and accessibility of the data. Therefore, it is necessary to identify the main file types used in Python and R.



First :Python and R files

1- CSV (Comma-Separated Values) files

Application:

- When it comes to data storage and management, it's common practice to organize data in tables with rows and columns.
- When managing data, it is often necessary to keep files lightweight and readable.
- This type of representation is particularly well-suited for data generated by probability and descriptive statistics, such as frequency tables and conditional probabilities.

CSV Files with python

EXPORT

```python
import pandas as pd

data = {'Nom': ['Ali', 'Omar', 'Sara'], 'Probabilité': [0.2, 0.5, 0.3]}
df = pd.DataFrame(data)

# export data to CSV
df.to_csv('donnees.csv', index=False)

```

IMPORT

```python
df_imported = pd.read_csv('donnees.csv')
print(df_imported)

```

CSV Files in R

EXPORT

```r
write.csv(data, "export_donnees.csv", row.names=FALSE)
```

IMPORT

```r
data <- read.csv("donnees.csv", header=TRUE, sep=",")
head(data)
```

2-Excel files (.xlsx)

Usage:

- When the information is complicated or covers several sheets.
- When the data must be analyzed directly in Excel.
- Suitable for storing tables of random variables and probability matrices.

Excel Files with python

EXPORT

```
1   df.to_excel('donnees.xlsx', index=False)
2
```

IMPORT

```
1   df_imported = pd.read_excel('donnees.xlsx')
2
```

Excel Files in R

EXPORT

```
write_xlsx(data, "export_donnees.xlsx")

```

IMPORT

```
data <- read_excel("donnees.xlsx", sheet = 1)
head(data)

```

3 - Java Script Object Notation (JSON) files

Usage:

- When data is not organized in a table, such as nested data or relationships between random variables..
- When data needs to be sent via APIs or stored on the web.
- Suitable for storing the results of probabilistic simulations or Markov models.

JSON Files with python

EXPORT

```
df.to_json('donnees.json', orient='records', lines=True)

```

IMPORT

```
df_imported = pd.read_json('donnees.json', lines=True)

```

JSON Files in R

EXPORT

```r
toJSON(data, pretty=TRUE, auto_unbox=TRUE, file="export_donnees.json")

```

IMPORT

```r
data <- fromJSON("donnees.json")
print(data)

```

4-Text (.txt) files

Usage:

- When data is not organized in a table, such as nested data or relationships between random variables.
- When data needs to be sent via APIs or stored on the web.
- For storing the results of probabilistic simulations or Markov models.

Txt Files with python

EXPORT

```
1    data_loaded = np.loadtxt('data.txt')
2    print(data_loaded)
3
```

IMPORT

```
1    import numpy as np
2    data = np.array([0.1, 0.5, 0.9, 1.2])
3    np.savetxt('data.txt', data)
4
```

Txt Files in R

EXPORT

```
write.table(data, "export_donnees.txt", sep="\t", row.names=FALSE)

```

IMPORT

```
data <- read.table("donnees.txt", header=TRUE, sep="\t")
print(data)

```

5 - Pickle Files (.pkl)

Usage:

To store statistical or machine-learning models.

When you need to store complex data such as dictionaries or multidimensional matrices.

Suitable for storing results of probabilistic analysis or Bayesian models.

Pickle Files with python

EXPORT

```python
import pickle
modele = {'moyenne': 5.6, 'variance': 2.3}
with open('modele.pkl', 'wb') as f:
    pickle.dump(modele, f)

```

IMPORT

```python
with open('modele.pkl', 'rb') as f:
    modele_importe = pickle.load(f)
print(modele_importe)

```

6- HDF5 (Hierarchical Data Format) files

Usage :

- Suitable for processing Markov chain data or analyzing data from large statistical experiments.
- When working with large data that cannot be fully loaded.

HDF5 Files in Python

```
1    df.to_hdf('donnees.h5', key='data', mode='w')
2    df_imported = pd.read_hdf('donnees.h5', 'data')
3
```

نوع الملف	متى يُستخدم؟	الأمر في R
CSV	بيانات جدوليّة صغيرة أو متوسطة	`read.csv()` و `write.csv()`
Excel	جداول متعددة الأوراق	`read_excel()` و `write_xlsx()`
JSON	بيانات غير منظمة أو JSON API	`fromJSON()` و `toJSON()`
TXT	بيانات نصية أو عددية	`read.table()` و `write.table()`
RData	حفظ الجلسات في R	`save()` و `load()`
Pickle	مشاركة البيانات مع Python	`reticulate::py_run_string()`

النوع	Python (Pandas) في	R في
CSV	`pd.read_csv("data.csv")`	`read.csv("data.csv")` / `write.csv(df, "data.csv")`
JSON	`pd.read_json("data.json")`	`jsonlite::fromJSON()` / `write_json()`
Excel	`pd.read_excel("data.xlsx")`	`readxl::read_excel()` / `writexl::write_xlsx()`
RDS / Pickle	`pd.read_pickle("data.pkl")`	`readRDS()` / `saveRDS()`
HDF5	`pd.read_hdf("data.h5")`	`rhdf5::h5read()` / `h5write()`
Feather	`pd.read_feather("data.feather")`	`arrow::read_feather()` / `write_feather()`

Format	File	Import commant (Python)	Export command (R)
CSV	data csv	pd.read_csv(‘data json’)	write.csv (df, ‘data.csv’)
JSON	data xlsx	pd.read_json (‘data.json)	jsonlite::write_json (df, “data.json’)
Pickle or RDS	pd. hdf	pd.read_excel (data xlsx’)	readxl:: write_excel (df, ‘data.xlsx’)
HDF5	pd. f5	pd.read_hdf (data.h 5’)	saveRDS (df, ‘data.h5’)
Feather	pd. feath	pd.read_feather (‘data.feather)	arrow::write_feather (df, ‘data.feather’)