



INTRO TO PYTHON FOR DATA SCIENCE

Pandas



Overview

- Huge amounts of data are common
- 2D Numpy array?
 - Only one type possible
- Pandas
 - High-level data manipulation
 - DataFrame



brics

```
In [1]: brics = ... # declaration left out
```

```
In [2]: brics
```

```
Out[2]:
```

column labels

	country	population	area	capital
BR	Brazil	200	8515767	Brasilia
RU	Russia	144	17098242	Moscow
IN	India	1252	3287590	New Delhi
CH	China	1357	9596961	Beijing
SA	South Africa	55	1221037	Pretoria

row labels



CSV file

 brics.csv

```
,country,population,area,capital  
BR,Brazil,200,8515767,Brasilia  
RU,Russia,144,17098242,Moscow  
IN,India,1252,3287590,New Delhi  
CH,China,1357,9596961,Beijing  
SA, South Africa,55,1221037,Pretoria
```



CSV file → DataFrame

```
In [3]: import pandas as pd
```

```
In [4]: brics = pd.read_csv("path/to/brics.csv")
```

```
In [5]: brics
```

```
Out[5]:
```

	Unnamed: 0	country	population	area	capital
0	BR	Brazil	200	8515767	Brasilia
1	RU	Russia	144	17098242	Moscow
2	IN	India	1252	3287590	New Delhi
3	CH	China	1357	9596961	Beijing
4	SA	South Africa	55	1221037	Pretoria



CSV file → DataFrame

```
In [6]: brics = pd.read_csv("path/to/brics.csv", index_col = 0)
```

```
In [7]: brics
```

```
Out[7]:
```

	country	population	area	capital
BR	Brazil	200	8515767	Brasilia
RU	Russia	144	17098242	Moscow
IN	India	1252	3287590	New Delhi
CH	China	1357	9596961	Beijing
SA	South Africa	55	1221037	Pretoria



Column access

```
In [8]: brics["country"]
```

```
Out[8]:
```

```
BR          Brazil
RU          Russia
IN          India
CH          China
SA          South Africa
Name: country, dtype: object
```

```
In [9]: brics.country
```

```
Out[9]:
```

```
BR          Brazil
RU          Russia
IN          India
CH          China
SA          South Africa
Name: country, dtype: object
```



Add Column

```
In [10]: brics["on_earth"] = [True, True, True, True, True]
```

```
In [11]: brics
```

```
Out[11]:
```

	country	population	area	capital	on_earth
BR	Brazil	200	8515767	Brasilia	True
RU	Russia	144	17098242	Moscow	True
IN	India	1252	3287590	New Delhi	True
CH	China	1357	9596961	Beijing	True
SA	South Africa	55	1221037	Pretoria	True



Add Column (2)

```
In [12]: brics["density"] = brics["population"] / brics["area"] * 1000000
```

```
In [13]: brics
```

```
Out[13]:
```

	country	population	area	capital	on_earth	density
BR	Brazil	200	8515767	Brasilia	True	23.485847
RU	Russia	144	17098242	Moscow	True	8.421918
IN	India	1252	3287590	New Delhi	True	380.826076
CH	China	1357	9596961	Beijing	True	141.398928
SA	South Africa	55	1221037	Pretoria	True	45.043680



Row access

```
In [14]: brics.loc["BR"]
```

```
Out[14]:
```

```
country          Brazil
population        200
area             8515767
capital          Brasilia
density          23.48585
on earth         True
Name: BR, dtype: object
```



Element access

```
In [15]: brics.loc["CH", "capital"]
```

```
Out[15]: Beijing
```

```
In [16]: brics["capital"].loc["CH"]
```

```
Out[16]: Beijing
```

```
In [17]: brics.loc["CH"]["capital"]
```

```
Out[17]: Beijing
```



INTRO TO PYTHON FOR DATA SCIENCE

Let's practice!