main

Introduction to Pandas¶

Pandas is a Python library that plays a pivotal role in data science. It is used both for data wrangling and for calculations, and merges well with machine learning libraries, too. It has plenty of applications, covering a lot of the lost ground that Python had versus R in the past. You need to install it with the following command in the terminal.

```
python3 -m pip install pandas
```

Pandas uses *numpy* under the hood. *numpy* is a numerical library that enhances Python's computational capabilities. Pandas is well thought so that we do not need to explicitly invoke *numpy* often, but it will nevertheless appear every now and then.

We will mentioned *arrays* sometimes, and we will be referring to numpy's *ndarray* object. You may think of it loosely as a homogeneus list of numbers.

In [ ]:

```
from pprint import pprint
import pandas as pd
```

Series¶

Series and *DataFrame* are the two workhorses of pandas. *Series* is a one-dimensional object containing a sequence of values and an associated array of data labels called *index*.

Let's define our first *series* (the *pprint* is not necessary).

In [ ]:

```
from pprint import pprint
import pandas as pd
obj = pd.Series([1, 10, 5, 2])
pprint(obj)
```

```
0     1
1    10
```

```
2     5
3     2
dtype: int64
```

We can access the *array* and *index* attributes easily.

In [ ]:

```
obj.array
obj.index
```

Out[ ]:

```
RangeIndex(start=0, stop=4, step=1)
```