Applied Data Analytics

Pandas basics

Introduction to DataFrames

Hans-Martin von Gaudecker and Aapo Stenhammar

Series / Vectors: 1-dimensional data

- GDP for one unit over many time periods ("homogenous data")
- GDP for many units at one point in time ("homogenous data")
- Many variables for one unit at one point in time ("heterogenous data")

DataFrames: 2-dimensional data

- GDP for many units over many time periods ("homogenous data")
- Many variables for many units at one point in time ("heterogenous data")
- Many variables for one unit over many time periods ("heterogenous data")

Two series for Egypt over time

```
pop = pd.Series(
  [22.2, 38.8, 73.3],
  index=[1952, 1977, 2002],
  name="pop"
)
```

```
gdp_pc = pd.Series(
  [1418.8, 2785.5, 4754.6],
  index=[1952, 1977, 2002],
  name="gdpPercap"
)
```

year	pop
1952	22.2
1977	38.8
2002	73.3

year	gdpPercap
1952	1418.8
1977	2785.5
2002	4754.6

One DataFrame for Egypt over time

```
egypt = pd.DataFrame(
    {
        "pop": pop,
        "gdpPercap": gdp_pc
    }
)
```

year	pop	gdpPercap
1952	22.2	1418.8
1977	38.8	2785.5
2002	73.3	4754.6

Syntax for DataFrame construction

```
egypt = pd.DataFrame(
  {
    "col_0": series_0,
    "col_1": series_1,
    ...
  }
)
```

- You will find lots of other ways on the web
- Stick with this one for now
- Most of the time, read in the data anyhow