Applied Data Analytics

Pandas basics

Concatenating Series / DataFrames

Hans-Martin von Gaudecker and Aapo Stenhammar

Similar data, different sources

- Imports for each country
- The same survey questions for different years
- ...

→ Stack data on top of each other to obtain a single Series or DataFrame

Concatenating Series / DataFrames

country	year	lifeExp
Cuba	2002	77.158
Cuba	2007	78.273

country	year	lifeExp
Spain	2002	79.78
Spain	2007	80.941

country	year	lifeExp
Cuba	2002	77.158
Cuba	2007	78.273
Spain	2002	79.78
Spain	2007	80.941

Syntax

- `cuba` holds a Series or DataFrame with data from Cuba
- `spain` holds a Series or DataFrame with data from Spain

```
stacked = pd.concat([cuba, spain])
```

Careful with the index!

	country	year	lifeExp
0	Cuba	2002	77.158
1	Cuba	2007	78.273

	country	year	lifeExp
0	Spain	2002	79.78
1	Spain	2007	80.941

	country	year	lifeExp
0	Cuba	2002	77.158
1	Cuba	2007	78.273
0	Spain	2002	79.78
1	Spain	2007	80.941