

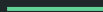


# Introduction to Pandas

BILD 62

# Objectives for today

- **Create & manipulate Pandas dataframes**



**Pandas** is a useful module that creates “data frames”

- great for real-world, heterogeneous data
- similar to Excel spreadsheets (but way faster!)
- “numpy with labels”
- Smartly deals with missing data

## Numpy:

|   | 0 | 1 | 2 |
|---|---|---|---|
| 0 |   |   |   |
| 1 |   |   |   |
| 2 |   |   |   |

## Pandas:

|          | Height | Weight | Age |
|----------|--------|--------|-----|
| Amy      |        |        |     |
| Brad     |        |        |     |
| Caroline |        |        |     |

# Useful Pandas methods

`df.mean()` Returns the mean of all columns

`df.corr()` Returns the correlation between columns in a data frame

`df.count()` Returns the number of non-null values in each data frame column

`df.max()` Returns the highest value in each column

`df.min()` Returns the lowest value in each column

`df.median()` Returns the median of each column

`df.std()` Returns the standard deviation of each column

For more useful functions, see [this overview](#).

Let's dig in!

# Resources

[A Quick Introduction to the “Pandas” Python Library](#)

[10 minutes to pandas — pandas 1.0.5 documentation](#)

[Python Data Science with pandas](#)