



Introduction to Pandas

BILD 62

Objectives for today

- Create & manipulate Pandas dataframes
 - Load a tabular file of genetic data into a Pandas dataframe
 - Index and subset Pandas dataframes
 - Generate descriptive statistics for Pandas dataframes
-

Name A Better Trio. I'll Wait 🤔



```
1 import numpy as np  
2 import pandas as pd  
3 import matplotlib.pyplot as plt
```

That's it.

[Source](#)



```
1 import numpy|
```



```
1 import numpy as np
```

There is no other way



**learning
numpy axis
rules**



**print output
array's
shape until
one of the
the axis
values
works out**

Credit to u/thatbrguy_

Pandas is a useful module that creates “data frames”

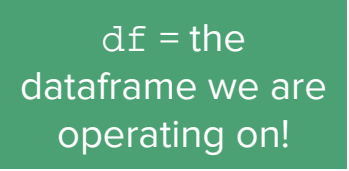
- great for real-world, heterogeneous data
- similar to Excel spreadsheets (but way faster!)
- “numpy with labels”
- Smartly deals with missing data

Numpy:

	0	1	2
0			
1			
2			

Pandas:

	Height	Weight	Age
Amy			
Brad			
Caroline			



df = the
dataframe we are
operating on!

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](#).

Resources

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

[10 minutes to pandas — pandas 1.5.3 documentation](#)

[Python Data Science with pandas](#)