

Computerphysik Programmiertutorial 8a

Prof. Dr. Matteo Rizzi und Dr. Markus Schmitt – Institut für Theoretische Physik, Universität zu Köln

Github: <https://github.com/markusschmitt/compphys2022>

Inhalt dieses Notebooks: Tabulare Datenstruktur

Hantieren mit Daten: DataFrames und CSV/CSVFiles

`DataFrames` stellt eine Datenstruktur für tabulare Daten zur Verfügung ([Dokumentation](#)).

Mit `CSV` und `CSVFiles` können tabulare Daten importiert und exportiert werden.

```
In [1]: using DataFrames
using CSV
using CSVFiles
```

Ein `DataFrame` kann zum Beispiel so erstellt werden:

```
In [2]: data=DataFrame(Name=["Hans", "Anna", "Klaus", "Petra"], Alter=[55,23,124,3],Geschlecht=["M","W","M","W"])
```

Out[2]: 4 rows × 3 columns

	Name	Alter	Geschlecht
	String	Int64	String
1	Hans	55	M
2	Anna	23	W
3	Klaus	124	M
4	Petra	3	W

Wir können weitere Zeilen mit `push!` hinzufügen

```
In [3]: push!(data,["Goofy",93,"M"])
```

Out[3]: 5 rows × 3 columns

	Name	Alter	Geschlecht
	String	Int64	String
1	Hans	55	M
2	Anna	23	W
3	Klaus	124	M
4	Petra	3	W
5	Goofy	93	M

Es sind viele Funktionen definiert um die Daten in der Tabelle zu verarbeiten (siehe [Dokumentation](#)). Z.B. können wir nur die Frauen aus der Liste heraussuchen:

```
In [4]: filter(row->row.Geschlecht=="W", data)
```

Out[4]: 2 rows × 3 columns

	Name	Alter	Geschlecht
	String	Int64	String
1	Anna	23	W
2	Petra	3	W

Die Tabelle kann mit `CSV.write()` exportiert werden.

```
In [5]: CSV.write("test.csv", data)
```

Out[5]: "test.csv"

```
In [6]: newdata = DataFrame(load("test.csv"))
```

Out[6]: 5 rows × 3 columns

	Name	Alter	Geschlecht
	String	Int64	String
1	Hans	55	M
2	Anna	23	W
3	Klaus	124	M
4	Petra	3	W
5	Goofy	93	M