

# Einführung in die Statistik für Digital Humanists

## - Sommersemester 2022

### Aufgabenpaket 2

*Abgabe bis Mittwoch 04.05.2022 um 23:59 Uhr*

#### 2.1 Beschreibung des Datensatzes “Drama”

In Ihrer Tabelle beschreiben Sie ein Theaterstück in 6 Spalten. Beschreiben Sie den Inhalt jeder Spalte so umfangreich wie möglich. Welche qualitativen Unterschiede sehen Sie zwischen den einzelnen Spalten, welche sind sich in bestimmten Aspekten ähnlich?

#### 2.2 Numpy und Pandas

- (A) Die Pakete Numpy und Pandas sollten in Ihrem Anaconda-Python schon installiert sein. Machen Sie sich mit den Grundfunktionen der beiden Pakete vertraut. Gute Einführungen finden Sie unter anderem unter diesen beiden Links:

- <https://www.python-kurs.eu/numpy.php>
- [https://pandas.pydata.org/pandas-docs/stable/getting\\_started/10min.html](https://pandas.pydata.org/pandas-docs/stable/getting_started/10min.html)

- (B) Laden Sie Ihren Theaterstück-Datensatz in Pandas. Beispielcode:

```
import pandas as pd
sp_novels = pd.read_csv("../Datensaetze/spanish-novels.tsv", sep="\t")
```

Viel Erfolg!