sqldf -> arbeitet nur mit Data Frames

sqldf(«SELECT COUNT(\*) from DF2 WHERE STATE = ‘ca’»)

restult as data frame!

**Vorbereitung:**

String lesen,

Zahl der Elemente bestimmen

in seine Komponenten zerlegen

je nach Maximum durch pentagram etc. gehen (die n-gramme müssen nach häufigkeit sortiert sein, Stemming wird nicht genutzt)

**Auswertung**:

Auf Gleichheit prüfen, höchster Match gewinnt

Kein Match – bigram-best match

**Return-Wert:**

Check auf Schimpfworte --> nicht ausgeben

Statistiken anreichern (Treffer, aus welchem n-gram)

Werte angeben

This is part one of your peer grading assignments - a report to show your fellow students how far you have got in order to (as stated in the instructions):

1. Demonstrate that you've downloaded the data and have successfully loaded it in.

2. Create a basic report of summary statistics about the data sets.

3. Report any interesting findings that you amassed so far.

4. Get feedback on your plans for creating a prediction algorithm and Shiny app.

Your report MUST be uploaded to RPubs and a link provided so students can mark your work.

Please do not worry if you submit late as the course mentors can mark your work if other learners are not available. However if learners have finished the entire course maybe you would like to return to marking other students report.