Barbora Hladká

Martin Holub

{Hladka|Holub}@ufal.mff.cuni.cz

Charles University,
Faculty of Mathematics and Physics,
Institute of Formal and Applied Linguistics
Programming questions

• **Feature frequency**
  • Implement a function that receives a vector of 1s and 0s and returns the number of 1s.

• **MOV data set**
  • Run this script
  • Work with movies data frame. For each genre-related feature compute its feature frequency. Plot all the feature frequency values.
Programming questions

- **USArrests data set**
  - `d <- USArrests`
  - Print a vector of the state names from the highest Assault rate to the lowest Assault rate.
  - Produce a scatter plot of Rape and Murder.
  - Compute Pearson correlation coefficient for Murder and Rape, Rape and UrbanPop, Murder and UrbanPop.
  - Run K-Means clustering algorithm for K=3 and experiment with Assault and Rape
  - Run K-Means clustering algorithm for K=3 and experiment with all the four features.
  - Use the Elbow Method to find an optimal value of K

- **Titanic data set**
  - `d <- read.csv("https://ufal.mff.cuni.cz/ hladka/2021/docs/train.csv")`