732A98 Visualization
Division of Statistics and Machine Learning
Department of Computer and Information Science

Laboratory work 3

Instructions

- Create a report to the lab solutions in PDF.
- Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments.
- **Include all your codes as an appendix into your report.**
- A typical lab report should 2-4 pages of text plus some amount of figures plus appendix with codes.
- The lab report should be submitted via LISAM before the deadline.

Assignment 1

Data set "Protein" contains information about consumption of different products in different countries of Europe (the quantitative measure is the amount of protein consumed). The variables are:

```
Country: Country name
RdMeat: Red meat
WhMeat: White meat
Eggs: Eggs
Milk: Milk
Fish: Fish
Cereal: Cereals
Starch: Starchy foods
Nuts: Pulses, nuts, and oil-seeds
FrVeg: Fruits and vegetables
```

1. Visualize the data by use of the scatter plot matrix containing regression curves showing trends. Which variables seem to have some relation to each other? Can you interpret this relationship?
2. Plot a heatmap for this data where rows and column are not permuted. Is it possible to discover groups of similar countries (clusters)? Are there outliers? Which variables seem to be correlated?
3. Plot a heatmap for this data obtained by two-way hierarchical clustering. Adjust the colors by using a different color palette and remove the dendrograms. Analyze the resulting plot: find clusters, analyze which countries belong to the same countries, which variables are the most important to define these clusters. Make the interpretation of the clusters.
4. Perform Anti-robinson unweighted seriation and PCA seriations, present the corresponding heatmaps and comment about the clusters found.
5. Is there anything common in the information you obtained by producing heatmaps and the information obtained from the scatter plot matrix.

732A98 Visualization
Division of Statistics and Machine Learning
Department of Computer and Information Science

Assignment 2

In this assignment you will analyze feedbacks given by customers for watches Casio AMW320R-1EV bought at www.amazon.com . Files Five.txt and OneTwo.txt contain feedbacks of the customers who were pleased and not pleased with their buy, respectively.

1. Use R tools to create a word cloud corresponding to Five.txt and OneTwo.txt and adjust the colors in the way you like. Analyze the graphs.
2. Run Phrase Nets as follows:
    a. If you are using your own computer, follow the instructions https://www.cg.tuwien.ac.at/courses/InfoVis/HallOfFame/2011/Gruppe08/Homepage/
    b. If you are using university computers
        i. Use the file phrase-net.zip provided with this lab and unpack it somewhere
        ii. Open the file run.txt and change the search path (first two lines) appropriately. The current search path assumes that run.txt is located in z:\732A98\phrase-nets
        iii. Launch the command line environment in Windows by opening the start menu in Windows and typing *cmd* in the search field
        iv. In the command line environment, copy and paste the contents of run.txt

Create the phrase nets for Five.Txt and One.Txt with connector words

- am,is,are,was,were
- a,the
- at
- of

where you choose ”Filter stopwords” option.

3. When you find an interesting connection between some words, use Word Trees https://www.jasondavies.com/wordtree/ to understand the context better. Note that this link does not work properly in Microsoft Edge (if you are using Windows 10) so use other browsers.

Analyze the graphs obtained and comment on the most interesting findings, like:

- Which properties of this watch are mentioned mostly often?
- What are satisfied customers talking about?
- What are unsatisfied customers talking about?
- What are good and bad properties of the watch mentioned by both groups?

- Can you understand watch characteristics (like type of display, features of the watches) by observing these graphs?

Submission procedure

Assume that X is the current lab number, Y is your group number.

If you are neither speaker nor opponent for this lab,

- Submit your report using *Lab X* item in the *Submissions* folder before the deadline.
- Make sure that you or some of your group members submits the group report using *Lab X group report* in the *Submissions* folder before the deadline

If you are a speaker for this lab,

- Submit your report using *Lab X* item in the *Submissions* folder before the deadline.
- Make sure that you or some of your group members does the following before the deadline:
  - submits the group report using *Lab X group report* in the *Submissions* folder before the deadline
  - Goes to Study room *Group Y→Documents* and opens file *Password X.txt*. Then the student should put your group report into ZIP file *Lab X_Group Y.zip* and protect it with a password you found in Password X.txt
  - Uploads the file to *Collaborative workspace →Lab X* folder

If you are opponent for this lab,

- Submit your report using *Lab X* item in the *Submissions* folder before the deadline.
- Make sure that you or some of your group members submits the group report using *Lab X group report* in the *Submissions* folder before the deadline
- After the deadline for the lab has passed, go to Collaborative workspace→*Lab X* folder and download the appropriate ZIP file. Open the PDF in this ZIP file by using the password available in *Course Documents →Password X.txt,* read it carefully and prepare (in cooperation with other group members) **at least three questions/comments/improvement suggestions per lab assignment** in order to put them at the seminar.