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

Laboratory work 1

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

Open R and read about data file Cars93 located in package MASS. Open the data set and inspect the variables that it contains. Run the following:

```
library(MASS)
df1=aggregate(Price~Type, data=Cars93, FUN=mean)
barplot(df1$Price, names.arg=df1$Type)
```

1. Save the plot emerged by running this code as PDF and open it in Inkscape
2. Enhance the plot as follows:
   a. Increase fonts
   b. Make proper axis labels
   c. Rotate Y axis values 90 degrees
   d. Add caption
   e. Change the color of the bars
   f. Add some graphics that would emphasize that mid-sized cars cost much more than small cars (see for example lecture slide 19)
3. Add the resulting plot to your report.

Assignment 2

Data set SENIC describes the results of measurements taken at different US hospitals. The description of the variables is given in the accompanying document **SENIC.pdf.** In this assignment, you are assumed to perform univariate analyses of data set SENIC stored in **SENIC.csv** by using **ggplot2**.

1. Read data from SENIC.txt into R.
2. For each qualitative variable produce a publication quality bar chart, and group plots into one frame (larger plot). Analyze this plot and report some interesting findings

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

3. For each quantitative variable, produce a publication quality boxplot and group plots into one frame (larger plot). Analyze this plot and report some interesting findings. If there were outlying observations present, are they the same for several plots? Can this be interpreted?

Assignment 3

The same data set as in the assignment 2 is analyzed here.

1. Write an R code that finds out in which hospital the ratio "Number of nurses/Number of beds" is the lowest. What does this ratio show?
2. Use **ggplot2** to plot "Infection risk" versus "Identification number", mark the hospital with the lowest ratio from step 1, and add a corresponding annotation. Conclusions?
3. Use the following code to compute the predicted values for a smoother and the corresponding standard errors:

```
library(fANCOVA)
mod=loess.as(data$Obs, data$X3, criterion="gcv", degree=2)
result=predict(mod, se=TRUE)
```

Use "result" object in order to compute confidence limits as fitted values plus or minus the two standard errors. Create a new data frame containing the observation number, fitted values, upper limits and lower limits. Use this data frame, geom_line() and geom_ribbon() to add the fitted curve and the confidence band to the existing plot. Does it seem that a straight horizontal line fits this band? How can this be interpreted?

- **Hint**: if you would like to see some sample code, google "confidence bounds ggplot2"

Submission procedure

Assume that X is the current lab 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.

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

- 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 *Speakers X→Documents* and opens file *Password X.txt*. Then the student should put your group report into ZIP file *Lab X.zip* and protect it with a password you found in Password X.txt
  - Uploads the file to *Collaborative workspace* 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 folder and download *Lab X.zip*. 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.