

# Data Mining

## Homework Assignment #3

Dmytro Fishman, Anna Leontjeva and Jaak Vilo

March 6, 2014

### Task 1

Choose your favorite visualization tool, play around with it and report few fancy pictures produced by this tool, e.g. ggplot2 (in R), Excel (hope you manage to find it yourself), Google Code Playground, GNU plot, matplotlib (for Python).

### Task 2

Read in the data file: IrisData.txt. For each variable, calculate mean, standard deviation, median, minimum and maximum. Describe what type of data you have (continuous or discrete etc.). There are some obvious outliers in the data. Find a way how to detect and remove rows which contain them (hint: if you use R, function “boxplot” could be useful). Explain why some of these statistics have changed and some of them not much. In the next exercises, use the data where outliers are removed.

### Task 3

### Task 4

### Task 5

Implement a density estimation function using a triangular kernel. Use that to plot the density. Compare to histogram and a smooth kernel using the same IrisData.txt as in the second task.

### Task 6 (2pt)