COURSE SYLLABUS  
Tools for Analytics Lab - R-track

Instructor  
Name: Gergely Daroczi  
Department: Department of Economics / CEU Business School  
Semester/term, year: Winter 2015  
Course level: MA  
Credits: 2 credits

Course Description

This is an introductory course on how to use the R programming language and software environment for data manipulations, exploratory data analysis, statistical modeling, machine learning, data visualizations, automated reporting and self-service dashboards.

Course Requirements

Assessment is through a final exam (50%) and assignments (50%).

Tentative outline

1. Introduction to R (downloading/installing base R and R packages, CRAN, R ecosystem, R community, RStudio, language syntax)  
2. Loading data (text files, Excel, databases)  
3. Data transformations (filtering and summarizing data with dplyr and data.table, wide and long tables with reshape2, dirty data problems)  
4. Data visualization (special plots like beanplot, grammar of graphics via ggplot2, and interactive plots via htmlwidgets)  
5. Modeling (generalized linear models, goodness of fit, overfitting, confounders)  
6. Multivariate statistical methods (PCA, FA, MDS)  
7. Clustering and classification (supervised and unsupervised methods, hierarchical and k-means cluster analysis, logistic regression, training and test datasets, k-NN, cross validation, confusion matrix)  
8. Machine learning (decision trees, pruning, random forest, gradient boosting, support vector machines, accuracy, AUC)  
9. Dynamic and reproducible reports (markdown, knitr, pander, version control, pandoc)  
10. Interactive data analysis (shiny, dashboards, htmlwidgets)  
11. Special data types (time-series, spatial and network data)