Some Guidelines to the Final Project

In the final project, you will work on a data set that is interesting to you and poses interesting questions from data mining and machine learning perspective. Such data sets can arise from your research area or can be found from on-line resource, such as "UCI Machine Learning Repository" and "Datasets for Data Mining, Analysis and Knowledge Discovery". You are expected to exam the data carefully before applying statistical learning methods for interpretation and prediction. You are encouraged to use methods we have discussed in class. You are also encouraged to learn and apply new methods that are suitable for the data. For a data set arising from your research, your analysis should be orientated towards the specific research questions. If your data has been studied in the literature, it would be good to find one or two related papers to see what has been done and why it makes sense.

The final report will be due by 5 pm on Apr 25, Monday. The report should be sent to lisha.chen@yale.edu. A one to two page proposal will be due on Apr 6, Wednesday in which you will describe the data set for your project and your plan for analysis. The proposal should be handed in on paper before class.

The final report should be within 15 pages (single spaced, with front size 11 or 12), including figures and tables. Exception can be given in the case where figures and tables are intensively used. Please acknowledge the resource of the data and attach all the code in the appendix (the appendix does not count for the pages).