Final Project

Description. This project is designed to provide an opportunity to apply data mining data analysis techniques to analyze a data set of your choice. Given a data set, **please pose two (or more) interesting questions about the data and try to answer them using methods you have learned.** A good way to do this is the following steps:

1. do as much data exploration as possible by visualization, summary statistics and a dimension reduction technique (if applicable)
2. apply some of the methods learned in the class (classification, clustering, regression, etc.) You are also welcome to consider more advanced methods not learned in the class, if the data suggests so.

Where to get the data? There are many data repositories online. You may use data from one of these repositories. In general, I would like to see teams taking on new and challenging data sets, and formulating novel inferential questions. The project can be open-ended. Please avoid older data sets that have already been tortured to death.

This is a team project consisting of three or four people. The number of participants in the group shall be three or four. Not five, not two. Between 3 and 4, inclusive. The project will consist of two parts: (1) **the write-up** and (2) **the presentation**.

(1) **The write-up.** Please write up a paper describing your data analysis and findings. You should discuss your results, the implication of various choices for parameters selection, and your conclusions. Otherwise the paper is free-form and can be written however you think best to present the material. **Please limit the paper to no longer than 5 pages.**

(1) **The presentation.** In lieu of a final exam, we will be having presentations of the data analysis during the final exam period. Each presentation will be about 15 minutes with some time for questions.