Assignment 6 (term): Data Analytics (Spring 2014) (25% written+5%oral)  
Due: TUESDAY May 6, 2014 (by 2pm ET)

Submission method: written and presentation (after you present it) by email to pfox@cs.rpi.edu and Lakshmi Chenicheri chenil@rpi.edu

Please use the following file naming for electronic submission: DataAnalytics2014_A6_YOURFIRSTNAME_YOURLASTNAME.xxx, etc.

Late submission policy: this assignment is due at the end of term. If you are more than a week late it is likely that you will not have your grade for this assignment included in your final grade before they need to be submitted.

Note: Your presentation for this assignment should be the result of your own individual work. Take care to avoid plagiarism (“copying”), and include references to all web resources, texts, and class presentations. You may discuss the project with other students, but do not take written notes during these discussions, and do not share your written assignment or presentation before the class they are presented in.

General assignment: Your term projects should fall within the scope of a data analytics problem of the type you have worked with in class/ labs, or know of yourself – the bigger the data the better. This means that the work must go beyond just making lots of figures. You should develop the project to indicate you are thinking of and exploring the relationships and distributions within your data to lead to predictive models. Start with a hypothesis, think of a way to model and use the hypothesis, find or collect the necessary data, and do both preliminary analysis, detailed modeling, summary (interpretation) and possible resulting decisions.

Note: You do not have to come up with a positive result, i.e. disproving the hypothesis is just as good. Please use the section numbering below for your written submission for this assignment.

Guidance: Topics, scope and general nature – please use the opportunity in Assignment 5 (project proposals) and seek feedback from the instructor.

1. Introduction (2%)
   Describe your motivation, initial hypothesis/ idea that you wanted to investigate, and if applicable any prior work, interest in the topic (like an intro for a paper). Min. 1/2 page.

2. Data Description (3%)
   Describe how you determined which datasets you used in this project, the criteria, source, data and information-types in detail, associated documentation and any other supporting materials. Min. 1/2 page text (+graphics if applicable).

3. Analysis (8%)
   Explore the statistical aspects of your datasets. Analyze the distributions and provide summaries of the relevant statistics. Perform any transformations, interpolations,
smoothing, etc. required on the data, to begin to explore your hypothesis/ questions. Perform uni- and/ or multivariate or similar analysis. Min. 1.5 page text + graphics.

4. Model Development (8%)
Discuss and specify or estimate possible sources of error, uncertainty or bias in the data you used (or did not use). Indicate how you calculated or estimated the propagation of errors through the various stages of your analyses. Discuss the confidence in your results including any statistic measures. Min. 1.5 page text + graphics.

5. Conclusions and Discussion (4%)
Provide a description of your conclusions, and add some summary discussion about your results, what changed as you went through the project (data, analysis, etc.), what you would do next, or do differently in a subsequent exploration. Min. 1 page text + graphics (optional).

References – websites, papers, packages, data refs, etc. should be included at the end.

6. Oral presentation (5%). Suggest these slides (8 mins):
   a. Title (with your name)
   b. Problem area – what you wanted to explore/ solve/ predict and why, and what you wanted to predict?
   c. The data – where it came from, why it was applicable and the preliminary assessments you made.
   d. How you conducted your analysis: distribution, pattern/ relationship and model construction. What techniques did you use/ not use and why?
   e. How did you apply the model? How did you optimize, account for uncertainties?
   f. What did you predict and what decisions (prescriptions) were possible. What was the outcome?

Graphical Representations
Provide graphical representations related to each of questions 2, 3, and 4, at least. Ensure all figures are numbered, legible, fully explained and annotated.

The final document should be a minimum of 5-8 pages of writing (but can be more). All graphics should be within your written assignment unless they are very large. Large graphics files should be sent as a separate attachment (e.g. in a zip file).