

Assignment 5: Data Analytics (Spring 2016) (5% oral)
Due: TUESDAY March 22, 2016 (by 2pm ET)

Submission method: presentation (after you present it) by email to pfox@cs.rpi.edu and Rahul Divekar divekr@rpi.edu

Please use the following file naming for electronic submission of your **presentation**:
DataAnalytics2016_A5_YOURFIRSTNAME_YOURLASTNAME.xxx

Late submission policy: if you do not present on Tuesday you must complete the presentation by Friday 25 March. If you do not, you receive 0% for this assignment and your project is not approved (and you still need to schedule this presentation). This may delay your start on your project and affect its completion.

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 presentations before class.

General assignment: Term project proposal. This is a chance to get feedback on what project (questions, data, methods, etc.) you currently plan for your term project.

Guidance: 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. 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 and summary (interpretation).

Note: What you present in this assignment does NOT have to be what you eventually conduct your project on. This is to get you to start thinking about how an end-to-end project would look. You do not have to come up with a positive result, i.e. disproving the hypothesis is just as good, and you will just need to explain it.

Note: There may be methods you use that we have not covered – that is okay.

1. Oral presentation (5 mins). Suggest these slides (~ 1 min. each after title slide):
 - a. Title (with your name)
 - b. Problem area – why it is of interest (in general or to you), what you might want to predict? This could be a hypothesis.
 - c. The data – where it might come from (or might come from), why it may be applicable and any preliminary assessment you’ve made?
 - d. How you plan to conduct your analysis: distribution, pattern/ relationship and model construction. What techniques do you think you will use?
 - e. How to plan to apply the model. What are the possible uncertainties?
 - f. What do you want to predict and what decisions (prescriptions) may be possible. What would a good outcome be?