APPM2720 Take home project (Quiz4)

Due Friday April 29, 2016

This project reinforces the central theme of this course: data analysis. Students should formulate a concise question that might be answer by a specific dataset, conduct an analysis of the data, and submit the results using the tools in RStudio for report writing. Students can work individually or in pairs. Any team will receive a single grade. Please contact the instructor (nychka@ucar.edu) if you have questions. (Including your data set and R code in your email can often be helpful in getting a quick answer to a problem or a strange error in R.)

Sources of questions and datasets

An important part of this project is coming up with a clear question or project and trying to answer it using your data analysis skills. Some places for data are kaggle, the **dataWorkshop** R package and financial data at Yahoo. Determine your problem early and check in with your instructor to make sure your project is appropriate. Collecting your own data is highly encouraged.

Contents of your report

The project report be less than 4 pages of text (but can also include figures and R outout) and have the following sections:

1) **Project Goal** The specific question or questions that the analysis should address.

2) **Description of the data**. Keep in mind that good data analysis has the background and source of the data set in mind. You should do some reading and research to be sure you understand how the data was generated and any unusual features that are not obvious. This should be detailed enough so that someone reading your report can reproduce or recover the *exact* data set that you are using. (If the data set is not too large, e.g. less than 1 Mb, please attach to your email along with your report.) Also give a clear description of any variables or other features that are important for your analysis. In particular make sure that you report the units of the variables.

3) **Details of the analysis** Include the methods used and a summary of the R code used. Typically there will be a small amount of output included and only a few key plots. Only include material that supports your analysis or is important. For example, include a residual plot if there is something strange. Omit residual plots if

there is nothing special to note and just add a comment in your text. Only include lm summaries if if your best model or if there is something interesting. All figures should have labels for the x and y axes, a title and legend or caption that explains different symbols or line types.

4) **Conclusions** What are your findings based on your analysis? It is OK to conclude that from the analysis the data is not adequate to give a clear answer. In that case you might describe how to collect additional data to perhaps give a better answer.

5) **Future work** Suggest additional directions to improve this analysis, assemble a more informative data set, or modify your question.