Assignment 2

Programming in R for Analytics - Fall 2018

Carnegie Mellon University, Heinz College
Instructor: Prahlad G Menon, PhD

Instructions [100 pts]

In your final project, your core R-programming skills are oriented around a specific dataset available from the world of reproducible research. Your goal will be to produce an R/Shiny application as well as an RMarkdown report (implementing specific user inputs from your R/Shiny application, say) that demonstrates your understanding of variables, data structures, program flow (e.g., conditional execution, looping) and functional programming, and applies these skills to answer interesting statistical questions involving the comparison of groups, which is core to statistical practice.

This assignment is specifically focused on:

1) Establishing your project team;
2) Identifying a project dataset and the associated published literature around it, with data definitions, etc.;
3) Establishing a hypothesis or a set of hypotheses to be tested from the data and your strategy for testing these hypotheses and visualizing the same;
4) Submission of:
    a. A preliminary RMD report loading in your data and performing some basic exploratory analysis that describes it in the context of your overarching hypotheses, including useful plots that may demonstrate your understanding of the data (Note: you don't need to complete your entire project! Just show that you've made a start.);
    b. Your dataset(s) and their data definitions;
    c. A report of your hypotheses along with relevant literature citations particular to the selected dataset.

Please ensure that your proposal is complete with a list of packages that you plan to leverage as well as that it identifies "at least one" avenue by way of which you will go beyond the classroom learnings of R / Shiny programming in order to complete your analysis. Please ensure also that you cite literature relevant to your dataset and your likely analysis approach(es).

Some potential sources of data are listed below:

1) https://www.kaggle.com
2) http://archive.ics.uci.edu/ml/datasets.html?sort=nameUp&view=list
3) https://www.census.gov/data/data-tools.html