IST 687 Extra Credit Assignment

This extra credit assignment will focus on updating a data science project – a project that you did *not* help create.

Specifically, you will be updating a different group's final project (i.e., updating someone else's analysis). You will not be offering suggestions for improvements back to the other team, rather, you will just be trying to make a small update in their analysis (that I will describe below).

You can earn up to a maximum of 3 extra points for doing this assignment.

If you want to earn this extra credit, please complete this form (https://syracuseuniversity.qualtrics.com/jfe/form/SV_6stdj3W2YfcCEqF) by 11:30pm on Dec 10th. The form will just ask you to provide some basic info (ex. your name and your lab section date/time and instructor's name).

You will then get an email from me (on Dec 11th), providing whom to contact – to get the existing project that needs to be updated.

The assignment needs to be completed by Dec 20th. See below for details.

Please feel free to reach out to me if you have any questions at any time.

Jeff

You will have two roles for this optional assignment.

1) Share your team's code with someone who will update your team's group project.
2) Update a different team's project

1) Sharing the code

In terms of (1), sharing your team's code, someone will reach out to you (as you will to someone else, based on my email on Dec 11th). This person will be in a different lab section.

You will provide that other person with what you think is the easiest way for that other person to be able to recreate (i.e., re-run) what your team has created, in terms of data cleaning, data analysis and data visualizations. For example, you can email the files directly or copy a project folder.

You can only share code/analysis that you can run yourself. So, if you can't get all of your project code to execute (e.g., you can't run one of your teammate's analysis), then that is OK, but you need to note that fact, and provide the other person (who requested your project code) with just the code that you could get to run. You need to provide an estimate of how much of the total project analysis they received (ex. just your code, a couple of team members code, all the code).

You will get partial extra credit if you shared your code. The amount of extra credit is not dependent on the percentage of project code shared (i.e., you will not be penalized for only being able to share your part of the project, as long as you explain why that occurred, for example, you have a copy of the code, but can't run that code).

2) Updating someone else's project
If you do not get any response (after asking a classmate for their team's code), please send me an email (copying the person to whom you asked for their project code).

In terms of (2), the updating a different project, your goal is to re-run all the analysis done in that other project (not your team's project, but the "new" project that you have been given). Then, you need re-run the analysis with a random sample of 75% of the data, which you will randomly sample ***after the dataset was cleaned (using the code cleaning that was done in the other project, not via your dataset cleaning)***. That is to say, you need to sample 75% the dataset (after it has been cleaned), and then run all the analysis on that slightly smaller subset.

So, you should be able to generate two sets of results (images and analytical output) - the initial analysis, and the analysis after you created the subset of data.

You need to submit a report to me (via a survey link that I will provide in the email on Dec 11th), containing information such as:

- *For task 1*:
    - How long it took (ex. how many minutes) it took to provide this information)
    - How you actually shared the R code.
    - How hard it was for you to provide a version of the code that ran the full analysis.
    - If there was part of the analysis you could not get running (and hence, did not share with the requesting person), explain what was the issue – you will not be deducted points if you clearly explain the issue.

- *For task 2:*
    - With respect to the initial (original) analysis:
        - Were you given all the code from the other team's work (based on your thoughts, and what the person giving you the code told you)?
        - Were you able to get all of the code given to you to run correctly (why/why not)?
        - How long did this work take (getting the other group's R analysis to run), in terms of the number of minutes?
        - Provide a detailed description of what challenges you had (if any) to re-create the other team's analysis (using their code).
    - With respect to the analysis to be run on the random subset of the dataset:
        - Were you able to create the 75% sample of the dataset – after the dataset was cleaned?
        - Were able to get the analysis to be run on the random subset of 75% of the dataset?
        - How long (number of minutes) did it take you to perform this task?
        - What challenges (if any) did you encounter?
        - What did you update with respect to the R code (how much code was changed, where was the code that was changed, etc).
    - Generate a report on the results of doing the analysis, that includes:
        - The actual results of the analysis using the full data set
        - The actual results of the analysis with the 75% subset
        - Note the difference in the analysis