Stat 6021: Project 2

You will be working in assigned groups of 3-4 students. Each group will work on the same dataset. Each group is free to come up with their own research hypotheses and questions.

The datasets that you will be working with describes more than 6,000 red and white vinho verde wines from the north of Portugal. The original data can be found here. One dataset contains information about white wines, and the other contains information about red wines. The variables are:

- fixed acidity
- volatile acidity
- citric acid
- residual sugar
- chlorides
- free sulfur dioxide
- total sulfur dioxide
- density
- pH
- sulphates
- alcohol
- quality (score between 0 and 10)

Each project should feature:

- Clear central analytic goals and/or questions to answer. The more practical, interesting, and challenging, the better.
- Statistical methods learned in this course so far (from modules 1 to 10), including exploratory data analysis.

- A "substantial" computing component for the analysis.

Your group is to submit the following via Assignments in Collab (one submission per group):

- A report (.pdf file)
- An R script containing your code (.R file)
- A 15-minute presentation (recorded via zoom)
- The slides used in the presentation

Part 1: Presentation

Each group will give a 15-minute presentation. This presentation should be designed to be understandable by anyone familiar with the course material, but who has not read the project report. Each group is free to organize who talks about which topics during the presentation. Not everyone need talk, but all group members should be active contributors to the presentation materials. **The zoom recording link and slides is due Friday, August 7, 12pm, via Assignments. One group member will submit the video presentation and the slides on behalf of the group. Please note that it may take a while for zoom to process your recording. The 12pm deadline is for availability of the zoom recording link, not when your are done recording.**

Part 2: Feedback

Each student will be randomly assigned to view another group's project and provide feedback. **The feedback is due Friday, August 7, via Assignments. Each student will have a submission. The zoom links for the presentations will be up by 12pm on August 7.**

Part 3: Report

The report should be no more than 20 pages, and should include the following:

1. **Executive Summary**: This section should describe the high-level goals/questions of the project, the nature/characteristics of the data used in the analysis, and the results of the analysis, including any recommendations (maximum 3 pages). Graphs that enhance the executive summary may be included. This section should be written in a way that can be understood by anyone without any statistical background (avoid terms such as p-value, null hypothesis, etc).
2. **Exploratory Data Analysis**: This section should include basic data exploration, using relevant graphical and numerical summaries, as well as interpreting the summaries.

3. **Detailed Analysis**: This section should include the goals and/or questions to answer, and how your group used regression methods to answer them, as well as relevant conclusions in context.

Please note the following principles:

- If relevant, your group may add additional sections or subsections.
- Your report should not include any R code. Relevant output from R should be included only if the output is referenced to in the report.
- Your report should include correct grammar, clear explanations, and professional presentation.
- The audience for the executive summary is a client who has little experience in analyzing data. The audience for the rest of the report is another classmate your client may hire to give a second opinion for your report. Conclusions must always be made contextually.

The report is due Friday, August 7, via Assignments. One group member will submit their report (.pdf file) and the R script (.R file) on behalf of the group.

Part 4: Evaluation

Each group member will submit a brief assessment of their own and their group members' performances on the project. **The evaluation is due Friday, August 7, via Test & Quizzes.**