Stat 6021: Project 1

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 dataset that you will be working with describes more than 210,000 different diamonds that are for sale on bluenile.com. A .csv file of the data will be provided to you. The variables are:

- carat

- clarity

- color

- cut

- price

Detailed descriptions of these variables can be found on the database for Blue Nile. 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 6).

- 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)

The report should be no more than 15 pages, and should include the following:

1. An executive summary that describes 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.

2. A detailed description of the data, the analysis (especially this), and the conclusions reached, making use of graphics wherever appropriate.

3. Correct grammar, clear explanations, and professional presentation.

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.

The audience for the executive summary is a client who has little experience in analyzing data. The audience for the detailed description is another classmate your client may hire to give a second opinion for your report.

Please also complete the **Project 1 Evaluation Questions**, which is also **due the same day as the report**. Please note that this is specifically for the project.