

# 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. 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), 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)

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

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 2 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.

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